ANCESTRA:REFORMING STREAMABLE VOLUMETRIC REPRESENTATIONS THROUGH CANONICAL TRANSPORT LINEAGES AND ASYMMETRIC BASIS LIFTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Long-horizon volumetric video reconstruction requires compact and temporal modeling that remains faithful as scene evolution grows increasingly complex. Existing streamable representations either recursively propagate reconstructed states, coupling temporal continuity with accumulated errors, or rely on persistent references whose validity deteriorates as scenes diverge; symmetric Gaussian bases further limit local expressivity near discontinuities. We introduce **AncesTrA**, a streamable volumetric representation that recasts long-horizon evolution as *structured inheritance along canonical transport lineages*. **Canonical Transport Inheritance with Entropy-Coded Residuals (CiTER)** separates reconstruction from temporal propagation by evolving compact transport states over a shared canonical representation, enabling efficient temporal **compression**. **Transport-regime-aware Canonical Expiration and Reanchoring (TraCER)** regulates canonical validity by reanchoring evolution when transport dynamics indicate expiration, preserving long-horizon **stability**. **Conditional Asymmetric Basis Lifting for Expressivity (CaBLE)** selectively lifts symmetric Gaussian bases into asymmetric forms around geometric and appearance discontinuities, increasing local expressivity without increasing anchor density. Extensive experiments demonstrate state-of-the-art reconstruction fidelity: **AncesTrA** outperforms Recon-GS by approximately 2 dB with 9× lower storage on short sequences, and by approximately 5 dB with 6× lower storage on long-horizon sequences. The widening advantage with temporal extent substantiates canonical transport lineages as a principled alternative to recursive scene-state propagation, positioning **AncesTrA** as a substantive reformulation of how temporal evolution is organized in streamable volumetric representations.

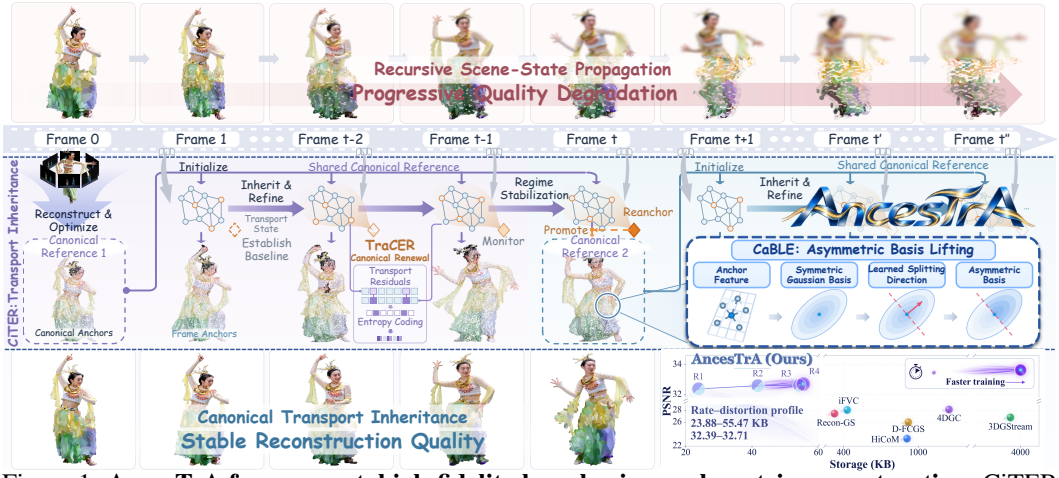

Figure 1: **AncesTrA for compact, high-fidelity long-horizon volumetric reconstruction.** CiTER decouples transport inheritance from scene reconstruction and entropy-codes transport residuals. TraCER reanchors when transport dynamics indicate canonical expiration. CaBLE conditionally lifts symmetric Gaussian bases into asymmetric forms for local expressivity.

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.

1 INTRODUCTION

Volumetric video reconstructs dynamic scenes from multi-view observations by modeling time-varying geometry and appearance, enabling applications such as digital humans, immersive communication, and autonomous driving (Wu et al., 2024; Zhang et al., 2025b). Yet its growing spatio-temporal complexity imposes a fundamental tension between *reconstruction fidelity* and *representation compactness*, which becomes increasingly severe as sequences extend over time. Resolving this tension is central to real-time volumetric video. Although offline approaches can achieve high-fidelity reconstruction, their reliance on complete sequences during optimization precludes latency-sensitive applications such as streaming media and immersive live broadcasting (Wu et al., 2025; Yoon et al., 2025). Recent methods adopt streamable or on-the-fly paradigms that incrementally encode scene evolution through compact inter-frame representations, exploiting temporal redundancy to reduce frame-wise overhead while preserving reconstruction fidelity (Tang et al., 2025b).

Despite this progress, existing streamable representations remain vulnerable to substantial degradation over long-horizon volumetric videos (Xu et al., 2024). Independently optimized frames discard temporal coherence, whereas recursive transformations preserve it by propagating reconstructed scene states across time (Tang et al., 2025b). This continuity, however, comes with an intrinsic coupling: each reconstructed state serves simultaneously as the current scene estimate and the predictive reference for future evolution, causing informative temporal structure and accumulated reconstruction imperfections to traverse the same inheritance pathway. The resulting bottleneck is therefore more fundamental than error accumulation alone; recursive state inheritance makes temporal continuity inseparable from the propagation of preceding approximation errors. A persistent reference can break this coupling, but introduces a complementary limitation: reference validity is inherently time-bounded. As the scene progressively departs from its canonical state, transport from a fixed reference must explain increasingly complex geometric and appearance variation, progressively enlarging the representational burden on temporally distant frames and undermining compact, high-fidelity reconstruction (Zhang et al., 2025a). This temporal bottleneck is further compounded by limited spatial expressivity in compact anchor-based representations. Shared latent features and centrally symmetric Gaussian bases efficiently characterize smoothly varying regions, yet provide insufficient flexibility around sharp geometric boundaries and appearance discontinuities (Kerbl et al., 2023; Lu et al., 2024). Collectively, these limitations reveal a deeper representational deficiency: existing streamable formulations conflate temporal persistence with reconstructed-state inheritance, implicitly assume time-invariant reference validity, and rely on locally constrained bases to model intrinsically asymmetric spatial structure.

To address these coupled deficiencies, we introduce **AncesTrA**, a streamable volumetric representation that reorganizes long-horizon scene modeling around *canonical transport lineages* and *conditional asymmetric basis lifting*. At its core, **Canonical Transport Inheritance with Entropy-Coded Residuals (CiTER)** resolves state-coupled temporal inheritance by separating scene reconstruction from temporal propagation: each frame is decoded from a shared canonical representation, while cross-frame dependency is carried exclusively by recursively evolving transport states. Their evolution is compactly conveyed through entropy-coded residuals, preserving temporal continuity without inheriting reconstructed scene states and their associated approximation errors. This lineage, however, cannot rely on a single canonical ancestor indefinitely. **Transport-regime-aware Canonical Expiration and Reanchoring (TraCER)** therefore monitors the evolving transport regime and renews the canonical state once its validity deteriorates, preventing temporally distant frames from being forced through an increasingly obsolete reference. Complementing this temporal organization, **Conditional Asymmetric Basis Lifting for Expressivity (CaBLE)** addresses the spatial limitation of symmetric local bases by selectively lifting Gaussian primitives into asymmetric counterparts around structural discontinuities, expanding local function capacity only where additional expressivity is required. Together, these components form a unified representational reformulation for streamable volumetric video, as illustrated in Fig. 1. CiTER establishes transport lineages, TraCER maintains their temporal validity, and CaBLE equips each lineage state with adaptive local expressivity. Rather than propagating reconstructed scenes, AncesTrA organizes temporal evolution around a persistent canonical ancestry and successive compact transport updates, while selectively enriching the spatial basis where symmetric primitives become insufficient. In this sense, **ances**try defines the representational origin and **tra**nsport encodes its evolution, yielding a compact, expressive, and sustainable formulation for long-horizon volumetric reconstruction under continuous scene change.

Extensive experiments demonstrate that **AncesTrA** achieves state-of-the-art (SOTA) reconstruction fidelity with highly compact representations, while sustaining its advantage as the temporal horizon extends. Compared with Recon-GS, **AncesTrA** improves reconstruction quality by about 2 dB on short sequences while reducing storage by 9×. On long-horizon sequences, this margin further widens to approximately 5 dB, together with a 6× reduction in storage overhead. The increasingly pronounced advantage over longer temporal horizons supports a shift from recursive scene-state propagation toward lineage-structured inheritance, positioning **AncesTrA** as a substantive reformulation of how temporal evolution is organized in streamable volumetric representations. Free-viewpoint demos are provided in Appendix A. Our main contributions are summarized as follows:

- We introduce **AncesTrA**, a unified streamable volumetric representation that recasts long-horizon scene evolution as *controlled inheritance along canonical transport lineages*, jointly organizing temporal propagation, canonical validity, and local basis expressivity.
- We propose **CiTER**, which decouples temporal dependency from reconstructed-state propagation by evolving successive transport states over a shared canonical representation and encoding their evolution through compact entropy-coded residuals.
- We develop **TraCER** to regulate the *validity horizon* of canonical ancestry by detecting transport-regime changes, expiring stale references, and reanchoring the lineage before they impose increasing representational burden on distant frames.
- We present **CaBLE**, which conditionally lifts symmetric Gaussian bases into an asymmetric function family around geometric and appearance discontinuities, selectively expanding local expressivity without indiscriminately increasing anchor density.
- Extensive evaluations demonstrate that **AncesTrA** sustains favorable quality–compactness trade-offs across both short and long horizons, with its widening advantage under prolonged scene evolution empirically validating structured canonical inheritance as a principled representation strategy for streamable volumetric video.

2 Related Work

2.1 Static Scene Reconstruction

3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) has emerged as a representation paradigm for static scene reconstruction, enabling high-fidelity rendering through explicit Gaussian primitives. However, the smooth and centrally symmetric nature of the Gaussian basis constrains its ability to represent sharp geometric boundaries and abrupt appearance transitions (Qu et al., 2024), often necessitating a dense set of primitives to faithfully preserve fine-scale structure. To improve representation efficiency, Scaffold-GS (Lu et al., 2024) organizes Gaussian primitives around sparse anchors and predicts their attributes from compact anchor features through a lightweight Gaussian generation network $\mathbf{N}_G$. Although this anchor-based formulation substantially reduces parameter redundancy, it primarily improves the parameterization and organization of Gaussian primitives rather than their intrinsic expressivity, leaving the underlying symmetric basis unchanged. In contrast, CaBLE conditionally lifts symmetric Gaussian bases into asymmetric forms around discontinuities, enhancing local expressivity while preserving compact anchor-based decoding.

2.2 Streamable Volumetric Video Reconstruction

Streamable volumetric video reconstruction incrementally models dynamic scenes from incoming multi-view observations, enabling interactive novel-view rendering under practical storage and transmission constraints (Wu et al., 2024; Zhang et al., 2025b; Ke et al., 2026). Tang et al. (2025b) exemplifies this paradigm by modeling Intra-coded frame (I-frame) with an anchor-based representation and encoding Predictive-frame (P-frame) through a compact Binary Transformation Cache (BTC), which propagates anchor positions and features from the preceding reconstructed state to exploit temporal redundancy. While effective for short-horizon redundancy reduction, this frame-to-frame formulation relies on recursive state propagation, causing reconstruction errors to propagate and progressively accumulate over extended sequences. AncesTrA transports scene evolution from a canonical representation rather than recursively propagating reconstructed states, and reanchors evolution when the canonical reference expires to preserve long-horizon reconstruction stability.

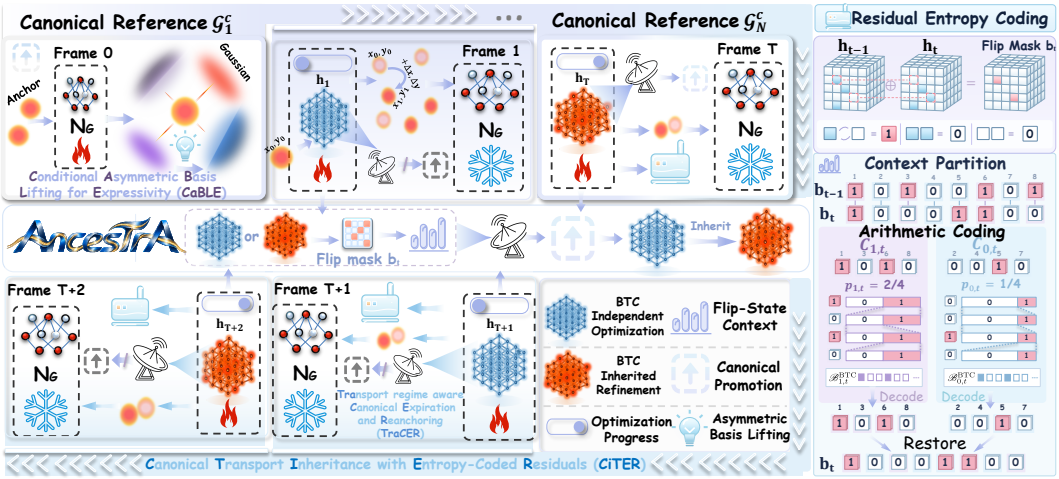

Figure 2: **Overview of AncesTrA.** Starting from the initial canonical representation $\mathcal{G}_1^c$, the Gaussian generation network $\mathbf{N}_G$ is optimized and subsequently frozen. Within each canonical transport lineage, CiTER independently optimizes the first predictive BTC to establish the transport trajectory, while subsequent BTCs inherit and refine their decoded predecessors; each frame is nevertheless reconstructed directly from the shared canonical state. For compact transmission, successive binary BTC embeddings are represented by flip residuals $\mathbf{b}_t$, partitioned into conditional contexts $\mathcal{C}_{k,t}$ and entropy-coded using transition probabilities $p_{k,t}$. When TraCER detects transport dynamics indicative of canonical expiration, the current reconstructed state is promoted to a new canonical reference, resetting the transport lineage and residual-coding context. Complementarily, CaBLE conditionally lifts symmetric Gaussian bases into asymmetric forms through learned spatial splitting and bilateral modulation, increasing local expressivity around discontinuities without increasing anchor density.

3 Method

3.1 Overview

Given a multi-view video sequence, AncesTrA performs on-the-fly volumetric reconstruction and compression by organizing long-horizon scene evolution into *canonical transport lineages*. Within each lineage, a shared canonical representation serves as a persistent reference, from which subsequent frames are reconstructed through evolving transport states. **Canonical Transport Inheritance with Entropy-Coded Residuals (CiTER)** encodes successive transport updates as compact entropy-coded residuals, separating scene reconstruction from temporal propagation. As the scene progressively departs from its canonical state, **Transport-regime-aware Canonical Expiration and Reanchoring (TraCER)** monitors the evolving transport regime, detects when the canonical reference approaches its representational limit, and establishes a renewed canonical reference. Complementing this temporal organization, **Conditional Asymmetric Basis Lifting for Expressivity (CaBLE)** selectively lifts symmetric Gaussian bases into asymmetric counterparts around geometric and appearance discontinuities that shared anchor features and symmetric bases cannot adequately represent. Together, these components organize temporal inheritance, canonical validity, and local expressivity for compact, high-fidelity long-horizon volumetric reconstruction (Fig. 2).

3.2 Canonical Transport Inheritance with Entropy-Coded Residuals (CiTER)

Existing streamable representations preserve temporal continuity through recursive transformations of reconstructed scene states, coupling temporal dependency with reconstruction or quantization errors along the same inheritance pathway. To break this coupling, we introduce **Canonical Transport Inheritance with Entropy-Coded Residuals** (CiTER), which separates scene reconstruction from temporal propagation. CiTER anchors reconstruction to a persistent canonical representation while carrying cross-frame dependency through an evolving transport trajectory, establishing *canonical transport lineages* as the basic units of streamable representation.

Canonical Transport Lineages. Let $0 = c_1 < c_2 < \cdots < c_K < N$ denote the canonical frames induced by reanchoring, and define $c_{K+1} = N$. The m-th lineage spans

$$\mathcal{S}_m = \{c_m, \ldots, \min(c_{m+1}, N-1)\}, \qquad \mathcal{P}_m = \mathcal{S}_m \setminus \{c_m\}, \tag{1}$$

$$\mathcal{L}_m = \left(\mathcal{G}_m^c, \{\hat{\boldsymbol{\theta}}_t\}_{t\in\mathcal{P}_m}\right), \tag{2}$$

$$\hat{\mathcal{G}}_t = \begin{cases} \mathcal{G}_m^c, & t = c_m, \\ \mathcal{T}_{\hat{\boldsymbol{\theta}}_t}(\mathcal{G}_m^c), & t \in \mathcal{P}_m, \end{cases} \tag{3}$$

where $\mathcal{G}_m^c$ denotes the shared canonical representation and $\hat{\boldsymbol{\theta}}_t$ denotes the decoded transport parameters. We instantiate $\mathcal{T}$ with a compact Binary Transformation Cache (BTC) (Tang et al., 2025b), which predicts anchor-position and feature transformations relative to the canonical representation. CiTER separates scene reconstruction from temporal propagation: each frame within a lineage is decoded from the shared canonical representation, while temporal continuity is maintained through the evolving transport trajectory. Accordingly, the first predictive transport is optimized independently, whereas subsequent transports inherit their decoded predecessors as initialization:

$$\boldsymbol{\theta}_t = \mathrm{Refine}\left(\hat{\boldsymbol{\theta}}_{t-1}; \mathcal{I}_t, \mathcal{G}_m^c\right), \qquad t \in \mathcal{P}_m,\ t > c_m + 1, \tag{4}$$

where $\mathcal{I}_t$ denotes the current multi-view observations. This inherited initialization preserves the temporal evolution carried by previous transport states while enabling adaptation to current observations through refinement at each optimization stage. CiTER therefore preserves temporal continuity without propagating approximation errors through reconstructed scene states.

Entropy-Coded Transport Residuals. The inherited transport trajectory provides a temporal reference for compact transmission. For the binary BTC hash embedding $\mathbf{h}_t \in \{-1, +1\}^D$, we represent temporal variation by a flip mask $\mathbf{b}_t \in \{0, 1\}^D$, partition its entries according to their preceding flip states, and model the resulting transitions with a conditional Bernoulli distribution:

$$b_{t,j} = \mathbb{I}[h_{t,j} \neq h_{t-1,j}], \qquad h_{t,j} = (1 - 2b_{t,j})h_{t-1,j}, \tag{5}$$

$$\mathcal{C}_{k,t} = \{j \mid b_{t-1,j} = k\}, \qquad k \in \{0, 1\}, \tag{6}$$

$$q_t(\mathbf{b}_t \mid \mathbf{b}_{t-1}) = \prod_{k=0}^{1} \prod_{j\in\mathcal{C}_{k,t}} p_{k,t}^{b_{t,j}} (1 - p_{k,t})^{1-b_{t,j}}, \tag{7}$$

where $b_{t,j}$ denotes the j-th entry of $\mathbf{b}_t$, and $p_{k,t}$ denotes the flip probability conditioned on the preceding flip state k. Since $\mathbf{b}_{t-1}$ is already available at the decoder, $\mathcal{C}_{k,t}$ requires no additional signaling, allowing arithmetic coding to exploit distinct transition statistics for previously changed and unchanged entries and compactly recover $\mathbf{h}_t$ from its decoded predecessor. CiTER thus concentrates optimization inheritance and residual compression within a decoder-consistent transport trajectory, while retaining the shared canonical representation as the sole reconstruction reference, preserving temporal continuity without recursive scene-state error propagation.

3.3 Transport-regime-aware Canonical Expiration and Reanchoring (TraCER)

CiTER preserves temporal inheritance without recursive scene-state propagation by decoding each lineage from its persistent canonical reference. Yet no canonical reference remains valid indefinitely: as scene content evolves, transport must reconcile the current observation with an increasingly distant canonical ancestor, imposing growing representational burden. To regulate this validity horizon, we introduce **Transport-regime-aware Canonical Expiration and Reanchoring** (TraCER), which characterizes lineage evolution from transport dynamics and renews canonical ancestry when transport enters a stabilized regime. TraCER thereby transforms fixed-reference inheritance into an adaptive sequence of canonically anchored lineages across long-horizon scene evolution.

Transport-Regime Characterization. Within lineage $\mathcal{L}_m$, transport measures the scene's departure from its canonical ancestry. Let $a_m = c_m + 1$ be the first predictive frame, with binary BTC embedding $\mathbf{h}_{a_m}$. Rather than using absolute transport magnitude, we quantify this evolution using canonical-relative transport departure $\mathcal{D}_t$ and its normalized trend:

$$\mathcal{D}_t = \frac{1}{D}\sum_{j=1}^{D} \mathbb{I}[h_{t,j} \neq h_{a_m,j}] = \frac{1}{2}\left(1 - \frac{\mathbf{h}_t^\top \mathbf{h}_{a_m}}{D}\right), \tag{8}$$

$$\widehat{\beta}_t = \frac{\mathbf{u}^\top \mathbf{d}_t}{\mathbf{u}^\top \mathbf{u}}, \qquad \mathbf{d}_t = [\mathcal{D}_{t-W}, \ldots, \mathcal{D}_t]^\top, \quad u_i = i - \frac{W}{2}, \tag{9}$$

$$\sigma_t^2 = \frac{1}{W}\sum_{i=1}^{W} \left(\Delta d_{t,i} - \overline{\Delta d}_t\right)^2, \qquad \Delta d_{t,i} = \mathcal{D}_{t-W+i} - \mathcal{D}_{t-W+i-1}, \tag{10}$$

$$\rho_t = \begin{cases} |\widehat{\beta}_t|/\sigma_t, & \sigma_t \geq \varepsilon, \\ +\infty, & \sigma_t < \varepsilon,\ |\widehat{\beta}_t| > \varepsilon, \\ 0, & \sigma_t < \varepsilon,\ |\widehat{\beta}_t| \leq \varepsilon, \end{cases} \tag{11}$$

where u_i is the centered temporal coordinate for $i = 0, \ldots, W$, and $\overline{\Delta d}_t = W^{-1} \sum_{i=1}^{W} \Delta d_{t,i}$ is the mean increment. The slope $\widehat{\beta}_t$ captures the local evolution rate of transport departure, while σ_t tracks per-frame fluctuation. Their ratio ρ_t compares systematic trend against temporal variation, with $\varepsilon = 10^{-8}$ handling the near-zero-variance case. Instead of directly thresholding drift magnitude, TraCER identifies canonical expiration from transport-regime stabilization: a small ρ_t indicates weak systematic evolution relative to local fluctuation, marking the effective validity boundary of the current canonical reference for this stage of evolution.

Canonical Expiration and Reanchoring. To prevent transient fluctuations from prematurely terminating a lineage, expiration is triggered only once the stabilized regime persists for C consecutive estimates and the lineage reaches a minimum duration $T_{\min}$. We define

$$\mathcal{A}_m = \{t \mid t - c_m \geq T_{\min},\ t \geq a_m + W + C\}, \tag{12}$$

$$e_m = \min \left\{ t \in \mathcal{A}_m \;\middle|\; \max_{0 \leq k < C} \rho_{t-k} < \tau \right\}, \tag{13}$$

$$c_{m+1} = e_m, \qquad \mathcal{G}^c_{m+1} = \hat{\mathcal{G}}_{e_m}, \tag{14}$$

$$a_{m+1} = c_{m+1} + 1, \qquad \mathcal{D}_{a_{m+1}} = 0, \tag{15}$$

where $\tau = 2$ is the dimensionless stabilization threshold (see Appendix D.5). Once the criterion is satisfied, the decoded scene state at frame e_m is promoted to the canonical reference of the next lineage, and the subsequent transport trajectory evolves from this renewed origin. The first predictive frame after reanchoring initializes an independent BTC, resets the residual-coding context, and establishes a transport baseline. TraCER thus converts transport dynamics into an adaptive mechanism for regulating canonical ancestry: it renews canonical origins only at stabilized regime boundaries, allowing CiTER to sustain compact transport inheritance across long-horizon scene evolution.

3.4 Conditional Asymmetric Basis Lifting for Expressivity (CaBLE)

While CiTER and TraCER organize temporal inheritance and canonical evolution, compact anchor-based representations remain constrained by a local expressivity bottleneck. Anchor-based decoding improves parameter efficiency by generating multiple Gaussian primitives from shared anchor features (Tang et al., 2025b), but this sharing restricts each primitive to a prescribed function family. Conventional Gaussian primitives are spatially smooth and centrally symmetric, limiting their ability to model sharp geometric boundaries and localized appearance discontinuities. To address this spatial bottleneck, we introduce **Conditional Asymmetric Basis Lifting for Expressivity** (CaBLE), which enables anchor features to selectively lift symmetric Gaussian bases into asymmetric counterparts only where additional degrees of freedom are needed for local modeling. Because the lifted family retains the original Gaussian as a special case, CaBLE expands the effective local function space without indiscriminately increasing representation capacity. Formally, given a Gaussian primitive centered at μ with covariance Σ, its spatial kernel is formulated as:

$$G(\mathbf{p}; \mu, \Sigma) = \exp\left(-\frac{1}{2} (\mathbf{p} - \mu)^{\mathsf{T}} \Sigma^{-1} (\mathbf{p} - \mu) \right). \tag{16}$$

CaBLE conditionally lifts this symmetric basis by explicitly introducing an anchor-conditioned splitting direction and bilateral opacity modulation. Specifically, the Gaussian generation network $\mathbf{N}_G$ additionally predicts a directional parameter $\widetilde{\mathbf{n}}$, from which the normalized splitting normal $\mathbf{n}$ and signed coordinate $h_{\mathbf{n}}(\mathbf{p})$ are derived to parameterize the resulting spatial partition:

$$\mathbf{n} = \frac{\widetilde{\mathbf{n}} + \mathbf{n}_0}{\|\widetilde{\mathbf{n}} + \mathbf{n}_0\|_2}, \qquad h_{\mathbf{n}}(\mathbf{p}) = \mathbf{n}^{\mathsf{T}} (\mathbf{p} - \mu), \tag{17}$$

where $\mathbf{n}_0 = [0, 0, 1]^{\mathsf{T}}$ specifies a valid initialization direction. The plane $h_{\mathbf{n}}(\mathbf{p}) = 0$ partitions the primitive into two complementary half-spaces oriented by the learned normal $\mathbf{n}$:

$$G^+_{\mathbf{n}}(\mathbf{p}) = G(\mathbf{p}; \mu, \Sigma) \mathbb{I}[h_{\mathbf{n}}(\mathbf{p}) \geq 0], \tag{18}$$

$$G^-_{\mathbf{n}}(\mathbf{p}) = G(\mathbf{p}; \mu, \Sigma) \mathbb{I}[h_{\mathbf{n}}(\mathbf{p}) < 0], \tag{19}$$

where the two lifted bases share identical geometry and jointly recover the original Gaussian support. To instantiate this asymmetric modulation, the Gaussian generation network $\mathbf{N}_G$ outputs two opacity coefficients α^+ and α^- for the positive and negative half-spaces, respectively:

$$\alpha(\mathbf{p})G(\mathbf{p};\mu,\Sigma) = \alpha^+ G^+_{\mathbf{n}}(\mathbf{p}) + \alpha^- G^-_{\mathbf{n}}(\mathbf{p}). \tag{20}$$

This formulation allows the anchor feature to determine whether additional local expressivity is warranted. When $\alpha^+ = \alpha^-$, the lifted basis exactly recovers the conventional symmetric Gaussian; otherwise, the primitive acquires asymmetric modulation across the learned splitting plane. Equivalently, with $a = (\alpha^+ + \alpha^-)/2$ and $\delta = (\alpha^+ - \alpha^-)/2$, the lifted representation becomes:

$$\alpha(\mathbf{p})G(\mathbf{p}) = aG(\mathbf{p}) + \delta\left(G^+_{\mathbf{n}}(\mathbf{p}) - G^-_{\mathbf{n}}(\mathbf{p})\right). \tag{21}$$

Here, the first term preserves the original symmetric Gaussian basis, whereas the second introduces an adaptive asymmetric component governed by δ. Therefore, CaBLE retains conventional Gaussian splatting as a special case while selectively expanding primitive-level function space wherever anchor-conditioned prediction identifies the need for additional local expressivity.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate AncesTrA on short- and long-horizon volumetric video benchmarks: (1) N3DV (Li et al., 2022b), six 300-frame scenes at 2704×2028; (2) Meeting Room (Li et al., 2022a), three 300-frame scenes at 1280×720; and (3) Bar (Xu et al., 2024), a 3,540-frame 4K sequence captured by 22 synchronized cameras at 60 FPS, with a resolution of 2110×3760. For all datasets, one viewpoint is held out for testing and the remaining are used for training.

Baselines. Baselines include offline methods Swift4D (Wu et al., 2025) and SplineGS (Yoon et al., 2025), and streamable methods 3DGStream (Sun et al., 2024), HiCoM (Gao et al., 2024), 4DGC (Hu et al., 2025), 4DGCPro (Zheng et al., 2025), ComGS-l (Chen et al., 2025a), D-FCGS (Zhang et al., 2026), MoRGS-l (Lee et al., 2026), StreamSTGS (Ke et al., 2026), ClipGStream (Liang et al., 2026), Recon-GS (Fu et al., 2025), and iFVC (Tang et al., 2025b). Together, they provide a broad comparison across representative reconstruction paradigms.

Metrics. We report frame-averaged PSNR (dB), SSIM, encoded storage per frame (KB), training time (min), and rendering speed (FPS). Storage includes full canonical states at canonical frames and entropy-coded transport residuals with reanchoring overhead at predictive frames. We use published results under matching protocols and otherwise reproduce official implementations. AncesTrA and all reproduced baselines are evaluated under fair settings.

Implementation Details. For the I-frame, AncesTrA optimizes the anchor-based representation for 8K iterations. For the first predictive frame within each canonical lineage, the transport state is independently optimized for 5K iterations, while subsequent P-frames inherit the decoded transport state from the preceding frame and are refined for 300 iterations. For TraCER, we use $W = 8$, $\tau = 2$, $C = 3$, and a minimum lineage duration of $T_{\min} = 10$ frames. We train models using Adam with a learning rate of 5×10^{-3} on a single NVIDIA A800 GPU and instantiate transport with BTC.

4.2 Experimental Results

Quantitative Results. AncesTrA consistently achieves state-of-the-art reconstruction fidelity with compact storage across standard and long-horizon settings. As shown in Tab. 1, AncesTrA surpasses the recent SOTA D-FCGS by 3.12 dB on Meeting Room and attains the highest SSIM of 0.965, showing gains in both fidelity and structural similarity. On N3DV, it further reduces storage by over 200× relative to 3DGStream, highlighting representation compactness. This advantage becomes more pronounced over extended temporal evolution. As shown in Tab. 2, on the 3,540-frame Bar sequence, AncesTrA surpasses iFVC by 4.68 dB while accelerating per-frame training by approximately 5×. Fig. 3 further demonstrates that AncesTrA consistently maintains the highest reconstruction fidelity throughout the Bar sequence and exhibits remarkable temporal stability, whereas competing methods progressively degrade as the sequence evolves. These trends are consistent with CiTER enabling compact temporal inheritance and TraCER renewing the canonical reference as scene evolution extends. Taken together, these results establish AncesTrA as a compact, efficient, and temporally robust representation for streamable volumetric video reconstruction.

Table 1: **Quantitative comparison on Meeting Room and N3DV datasets.** A dash (–) indicates that the corresponding metric is not reported in the original work.

Category	Method	Meeting Room				N3DV			
		PSNR↑	SSIM↑	Storage (KB)↓	FPS↑	PSNR↑	SSIM↑	Storage (KB)↓	FPS↑
Offline	Swift4D (ICLR'25)	32.05	–	133.12	–	32.23	–	409.60	125
	SplineGS (ICLR'25)	–	–	–	–	32.60	0.950	–	76
Online	3DGStream (CVPR'24)	29.30	–	4198.40	260	31.35	0.948	7987.20	245
	HiCoM (NeurIPS'24)	26.73	–	614.40	284	32.08	0.953	706.56	255
	4DGC (CVPR'25)	28.08	0.922	430.08	213	31.58	0.943	512.00	168
	4DGCPro (NeurIPS'25)	–	–	–	–	31.64	0.944	655.36	–
	ComGS-l (NeurIPS'25)	–	–	–	–	32.12	–	112.64	147
	Recon-GS (NeurIPS'25)	30.84	–	307.20	256	32.66	0.957	450.56	250
	iFVC (AAAI'25)	32.38	–	92.16	157	32.32	–	92.16	126
	D-FCGS (AAAI'26)	30.97	0.950	92.16	288	32.02	0.951	184.32	215
	MoRGS-l (CVPR'26)	31.79	0.957	–	–	32.53	0.950	3891.20	200
	StreamSTGS (AAAI'26)	27.41	0.920	142.00	126	32.30	0.943	173.60	100
	ClipGStream (CVPR'26)	–	–	–	–	32.01	0.946	334.51	106
	AncesTrA	34.09	0.965	45.84	246	33.82	0.955	38.35	189

Table 2: **Quantitative comparison on Bar.** Our method achieves the best reconstruction quality and efficiency, with the lowest storage and fastest training.

Method	PSNR↑	Storage (KB)↓	Train (min)↓
3DGStream	26.78	3891.20	1.127
4DGC	28.11	1249.28	0.867
HiCoM	23.24	911.36	0.879
D-FCGS	26.00	921.60	-
Recon-GS	27.43	327.68	0.936
iFVC	28.03	430.08	0.906
AncesTrA	32.71	55.47	0.197

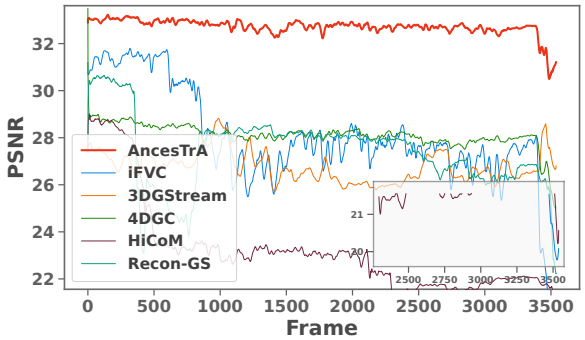

Figure 3: **Long-horizon stability on Bar.**

Qualitative Results. Qualitative comparisons corroborate AncesTrA's reconstruction advantage. As shown in Fig. 4, across Cook Spinach, Trimming, and VR Headset, competing methods blur contours, distort local geometry, and lose fine appearance details around hands, manipulated objects, thin structures, and object boundaries. In contrast, AncesTrA preserves clearer hand shapes, sharper object boundaries, and coherent local structures. Cross-sequence gains demonstrate robust local reconstruction across diverse dynamic interactions. The widening Bar gap reveals compounding difficulty with temporal distance. In the final row of Fig. 4, prolonged scene evolution causes competing methods to blur or distort the face, hands, and manipulated objects while losing fine details. By contrast, AncesTrA retains clearer facial structure, finger boundaries, and object geometry. These local gains align with CaBLE selectively expanding basis expressivity at geometric and appearance discontinuities, while the sustained Bar quality corroborates CiTER and TraCER maintaining temporal inheritance and canonical validity. Overall, AncesTrA preserves local structure and appearance across short- and long-horizon scene evolution. Supplementary free-viewpoint videos are provided in Appendix A to demonstrate temporal stability and canonical reanchoring.

4.3 Ablation Study

We conduct ablations on Bar to isolate the complementary roles of CiTER, TraCER, and CaBLE in temporal inheritance, canonical renewal, and local representation, respectively. Tab. 3 reports reconstruction fidelity, storage, rendering throughput, and per-frame training time.

Table 3: **Ablation study on Bar.**

Method	PSNR↑	Storage (KB)↓	FPS↑	Train (min)↓
w/o All	31.69	125.42	242	1.640
w/o CiTER & TraCER	32.23	110.70	229	2.313
w/o TraCER & CaBLE	31.18	16.53	243	0.107
w/o CaBLE	32.26	50.27	240	0.120
w/o TraCER	31.50	16.04	228	0.146
AncesTrA	32.71	55.47	226	0.197

❶ **Effect of Canonical Transport Inheritance.** CiTER contributes the primary efficiency gain by replacing recursive scene-state inheritance with compact transport evolution. Starting from *w/o All*, enabling CiTER alone in *w/o TraCER & CaBLE* reduces storage by 86.8% and accelerates per-frame

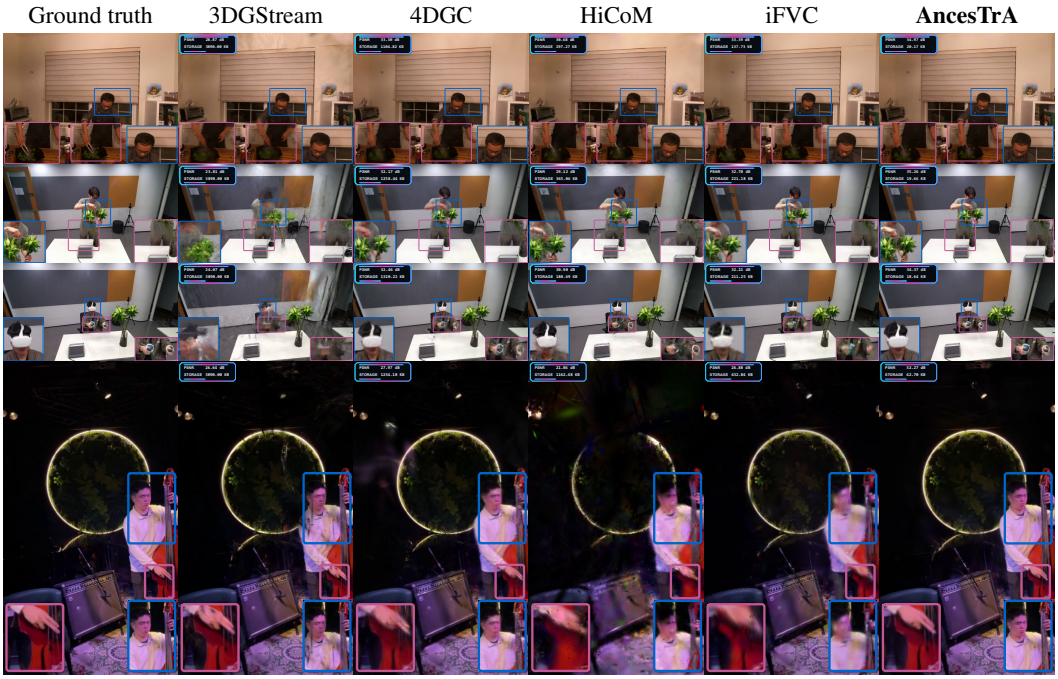

Figure 4: **Qualitative comparison. AncesTrA** consistently reconstructs finer details and delivers the best visual quality across all evaluated datasets.

training by 15.3×. This advantage remains pronounced in the complete lineage design. Relative to *w/o CiTER & TraCER*, AncesTrA improves PSNR from 32.23 to 32.71 dB while reducing storage from 110.70 to 55.47 KB and accelerating training by 11.7×. These results highlight canonical transport inheritance as the key mechanism for concentrating temporal adaptation into a compact and efficiently optimized parameter pathway throughout extended scene evolution. **❷ Effect of Canonical Reanchoring.** TraCER provides a complementary gain by renewing the canonical reference as scene evolution extends over long temporal horizons. Removing TraCER from the full model reduces PSNR by 1.21 dB. This result demonstrates that adaptive reanchoring restores canonical validity and reduces the representational burden on transport evolution. **❸ Effect of Conditional Asymmetric Basis Lifting.** CaBLE consistently strengthens reconstruction fidelity under different temporal organizations. Even without CiTER and TraCER, introducing CaBLE improves PSNR from 31.69 to 32.23 dB. Its contribution persists after canonical transport and adaptive reanchoring are established, where removing CaBLE reduces PSNR from 32.71 to 32.26 dB. This consistent improvement supports the role of asymmetric basis lifting in enriching local representation within each lineage beyond the temporal organization itself. **Taken together**, ablations reveal a division of roles: CiTER establishes compact and efficient temporal inheritance, TraCER maintains the relevance of the canonical reference, and CaBLE strengthens local reconstruction expressivity. Their combination yields the highest reconstruction fidelity while retaining the storage and optimization efficiency enabled by canonical transport inheritance. Further analysis is provided in Appendix H.

5 Conclusion

We presented AncesTrA, which reformulates long-horizon scene evolution as structured inheritance along canonical transport lineages. CiTER enables compact transport evolution from a shared canonical origin, TraCER renews this origin as its validity decays, and CaBLE selectively enhances local Gaussian expressivity around discontinuities. Together, they unify compact temporal inheritance, adaptive canonical renewal, and local expressivity within a coherent streamable framework. Extensive experiments demonstrate high reconstruction fidelity, compact storage, efficient optimization, and excellent long-horizon stability, with increasingly pronounced gains as temporal extent grows. More broadly, these results support a shift from recursive scene-state propagation toward lineage-structured inheritance, positioning AncesTrA as a substantive reformulation of how temporal evolution is organized in streamable volumetric representations.

LIMITATIONS AND FUTURE WORK

Although AncesTrA substantially advances long-horizon dynamic scene reconstruction through canonical transport lineages, further opportunities remain to enrich the motion awareness of canonical renewal. TraCER currently governs reanchoring primarily through observed transport-regime dynamics, providing an effective signal for canonical validity without explicitly modeling richer motion characteristics. A promising direction is therefore to incorporate richer motion-aware cues into the renewal criterion, enabling canonical ancestry to respond more precisely to diverse long-term temporal dynamics while preserving the compact lineage-based formulation.

AI USE STATEMENT

Generative AI tools were used solely for minor language polishing and literature summarization. AI-assisted editing was limited to improving the clarity, readability, and presentation of text originally written by the authors, without generating scientific content or substantive manuscript material. AI-generated literature summaries were used only as preliminary reading aids and were carefully verified against the original sources. AI tools were not used for literature retrieval, reference generation, experimental design, data analysis, scientific interpretation, or formulation of the paper's claims and conclusions. All references were selected and independently verified by the authors, and no AI-generated or fabricated citations were incorporated. The authors reviewed all AI-assisted edits and retain full responsibility for the final manuscript and its scientific content.

ETHICS STATEMENT

AncesTrA advances high-fidelity, storage-efficient, and temporally stable volumetric video reconstruction using publicly released research datasets under their respective terms of use. The study involves neither new data collection nor human participant recruitment, and its experimental objectives are confined to evaluating reconstruction fidelity, representation efficiency, and compression performance. The proposed framework operates on scene geometry and appearance without introducing mechanisms for identity recognition, behavioral profiling, or sensitive-attribute inference. Nevertheless, its ability to faithfully preserve fine-grained appearance and motion may retain identifiable characteristics present in the source material. Such representational fidelity underscores the importance of responsible data governance, particularly regarding informed consent, data provenance, licensing, and privacy protection. Although these considerations do not alter the reconstruction-oriented objectives of the proposed research, downstream applications involving identifiable individuals or publicly distributed content require independent assessment of their intended use and associated risks. Appropriate safeguards include verified data authorization, controlled access, proportionate retention, and measures against unauthorized reproduction or misleading redistribution. AncesTrA is intended to advance efficient dynamic-scene representation; its deployment should respect applicable data-protection requirements and maintain safeguards commensurate with the sensitivity of the reconstructed content.

REPRODUCIBILITY STATEMENT

We provide detailed methodological, implementation, and experimental specifications to facilitate faithful reproduction of AncesTrA. Sec. 3 presents the complete formulation of its three core components: Canonical Transport Inheritance with Entropy-Coded Residuals (CiTER), Transport-regime-aware Canonical Expiration and Reanchoring (TraCER), and Conditional Asymmetric Basis Lifting for Expressivity (CaBLE). Together, these formulations specify the construction of canonical transport lineages, the inheritance and compression of temporal updates, the adaptive renewal of canonical references, and the allocation of local representation capacity. Appendices C and D further provide detailed derivations of transport inheritance, residual coding, transport-regime estimation, and canonical reanchoring.

Sec. 4.1 and Appendix F document the experimental configurations, including the evaluated datasets, training and held-out viewpoints, compared baselines, evaluation metrics, optimization settings, stage-dependent training budgets, principal hyperparameters, and computational resources. We also

specify the storage measurement protocol, which uses the actual encoded representation and reports frame-averaged storage, including canonical representations, encoded transport updates, and reanchoring overhead. These specifications establish the experimental conditions and evaluation procedures needed to reproduce the reported quantitative results.

Appendix E provides explicit algorithms for canonical initialization, predictive-frame optimization, transport encoding and decoding, and canonical lineage renewal. The supplementary experiments further document long-horizon reconstruction results (Appendix G), component ablations (Appendix H), hyperparameter sensitivity (Appendix I), rate–distortion performance (Appendix J), and additional qualitative comparisons (Appendix K). Collectively, the main paper and supplementary material specify the complete reconstruction pipeline, its implementation and optimization procedures, and the corresponding evaluation protocol, providing a detailed basis for independently reproducing and verifying the reported AncesTrA results.

REFERENCES

Jiacong Chen, Qingyu Mao, Youneng Bao, Xiandong Meng, Fanyang Meng, Ronggang Wang, and Yongsheng Liang. Motion Matters: Compact gaussian streaming for free-viewpoint video reconstruction. In *Advances in Neural Information Processing Systems*, volume 38, pp. 133458–133482, 2025a. doi: 10.52202/085713-4018.

Yihang Chen, Qianyi Wu, Weiyao Lin, Mehrtash Harandi, and Jianfei Cai. HAC: Hash-grid assisted context for 3D gaussian splatting compression. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 422–438. Springer, 2024.

Yihang Chen, Qianyi Wu, Weiyao Lin, Mehrtash Harandi, and Jianfei Cai. HAC++: Towards 100X compression of 3D gaussian splatting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(11):10210–10226, 2025b. doi: 10.1109/TPAMI.2025.3594066.

Zhiwen Fan, Kevin Wang, Kairun Wen, Zehao Zhu, Dejia Xu, and Zhangyang Wang. LightGaussian: Unbounded 3D gaussian compression with 15x reduction and 200+ FPS. In *Advances in Neural Information Processing Systems*, volume 37, pp. 140138–140158, 2024. doi: 10.52202/079017-4447.

Jiaye Fu, Qiankun Gao, Chengxiang Wen, Yanmin Wu, Siwei Ma, Jiaqi Zhang, and Jian Jun Zhang. ReCon-GS: Continuum-preserved gaussian streaming for fast and compact reconstruction of dynamic scenes. In *Advances in Neural Information Processing Systems*, volume 38, pp. 149397–149424, 2025. doi: 10.52202/085713-4496.

Qiankun Gao, Jiarui Meng, Chengxiang Wen, Jie Chen, and Jian Zhang. HiCoM: Hierarchical coherent motion for dynamic streamable scenes with 3D gaussian splatting. In *Advances in Neural Information Processing Systems*, volume 37, pp. 80609–80633, 2024.

Sharath Girish, Tianye Li, Amrita Mazumdar, Abhinav Shrivastava, David Luebke, and Shalini De Mello. QUEEN: QUantized efficient ENcoding of dynamic gaussians for streaming free-viewpoint videos. In *Advances in Neural Information Processing Systems*, volume 37, pp. 43435–43467, 2024. doi: 10.52202/079017-1376.

Qiang Hu, Zihan Zheng, Houqiang Zhong, Sihua Fu, Li Song, Xiaoyun Zhang, Guangtao Zhai, and Yanfeng Wang. 4DGC: Rate-aware 4D gaussian compression for efficient streamable free-viewpoint video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 875–885, 2025.

Zhihui Ke, Yvyang Liu, Xiaobo Zhou, and Tie Qiu. StreamSTGS: Streaming spatial and temporal gaussian grids for real-time free-viewpoint video. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 40, pp. 5629–5638, 2026.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):139:1–139:14, 2023.

Sangwoon Kwak, Joonsoo Kim, Jun Young Jeong, Won-Sik Cheong, Jihyong Oh, and Munchurl Kim. MoDec-GS: Global-to-local motion decomposition and temporal interval adjustment for compact dynamic 3D gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11338–11348, 2025.

Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3D gaussian representation for radiance field. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21719–21728, 2024.

Wonjoon Lee, Sungmin Woo, Donghyeong Kim, Jungho Lee, Sangheon Park, and Sangyoun Lee. MoRGS: Efficient per-gaussian motion reasoning for streamable dynamic 3D scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 41044–41053, 2026.

Boyan Li, Yongting Chen, Dayou Zhang, and Fangxin Wang. DeformStream: Deformation-based adaptive volumetric video streaming. *arXiv preprint arXiv:2409.16615*, 2024.

Lingzhi Li, Zhen Shen, Zhongshu Wang, Li Shen, and Ping Tan. Streaming radiance fields for 3D video synthesis. In *Advances in Neural Information Processing Systems*, volume 35, pp. 13485–13498, 2022a.

Tianye Li, Mira Slavcheva, Michael Zollhoefer, Simon Green, Christoph Lassner, Changil Kim, Tanner Schmidt, Steven Lovegrove, Michael Goesele, Richard Newcombe, and Zhaoyang Lv. Neural 3D video synthesis from multi-view video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5521–5531, 2022b.

Jie Liang, Jiahao Wu, Chao Wang, Jiayu Yang, Xiaoyun Zheng, Kaiqiang Xiong, Zhanke Wang, Jinbo Yan, Feng Gao, and Ronggang Wang. ClipGStream: Clip-stream gaussian splatting for any length and any motion multi-view dynamic scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 41022–41032, 2026.

Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-GS: Structured 3D gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20654–20664, 2024.

Simon Niedermayr, Josef Stumpfegger, and Rüdiger Westermann. Compressed 3D gaussian splatting for accelerated novel view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10349–10358, 2024.

Haoxuan Qu, Zhuoling Li, Hossein Rahmani, Yujun Cai, and Jun Liu. DisC-GS: Discontinuity-aware Gaussian Splatting. In *Advances in Neural Information Processing Systems*, volume 37, pp. 112284–112309, 2024. doi: 10.52202/079017-3566.

Jiakai Sun, Han Jiao, Guangyuan Li, Zhanjie Zhang, Lei Zhao, and Wei Xing. 3DGStream: On-the-fly training of 3D gaussians for efficient streaming of photo-realistic free-viewpoint videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20675–20685, 2024.

Haocheng Tang, Ruoke Yan, Xinhui Yin, Qi Zhang, Xinfeng Zhang, Siwei Ma, Wen Gao, and Chuanmin Jia. HGC-Avatar: Hierarchical Gaussian compression for streamable dynamic 3D avatars. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 8125–8134, 2025a.

Luyang Tang, Jiayu Yang, Rui Peng, Yongqi Zhai, Shihe Shen, and Ronggang Wang. Compressing streamable free-viewpoint videos to 0.1 MB per frame. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 39, pp. 7257–7265, 2025b.

Irene Viola and Pablo Cesar. Volumetric video streaming. In *Immersive Video Technologies*, pp. 425–443. Elsevier, 2023.

Henan Wang, Hanxin Zhu, Tianyu He, Runsen Feng, Jiajun Deng, Jiang Bian, and Zhibo Chen. End-to-end rate-distortion optimized 3D gaussian representation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 76–92. Springer, 2024a.

Liao Wang, Qiang Hu, Qihan He, Ziyu Wang, Jingyi Yu, Tinne Tuytelaars, Lan Xu, and Minye Wu. Neural residual radiance fields for streamably free-viewpoint videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 76–87, 2023.

Penghao Wang, Zhirui Zhang, Liao Wang, Kaixin Yao, Siyuan Xie, Jingyi Yu, Minye Wu, and Lan Xu. V^3: Viewing volumetric videos on mobiles via streamable 2d dynamic gaussians. *ACM Transactions on Graphics*, 43(6):1–13, 2024b.

Zhe Wang, Jinghang Li, and Yifei Zhu. AirGS: Real-time 4D gaussian streaming for free-viewpoint video experiences. In *IEEE INFOCOM 2026 - IEEE Conference on Computer Communications*, pp. 1–10, 2026. doi: 10.1109/INFOCOM59046.2026.11571182.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

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.

Jiahao Wu, Rui Peng, Zhiyan Wang, Lu Xiao, Luyang Tang, Jinbo Yan, Kaiqiang Xiong, and Ronggang Wang. Swift4D: Adaptive divide-and-conquer gaussian splatting for compact and efficient reconstruction of dynamic scene. In *International Conference on Learning Representations (ICLR)*, 2025.

Zhen Xu, Yinghao Xu, Zhiyuan Yu, Sida Peng, Jiaming Sun, Hujun Bao, and Xiaowei Zhou. Representing long volumetric video with temporal gaussian hierarchy. *ACM Transactions on Graphics*, 43(6):1–18, 2024.

Jinbo Yan, Rui Peng, Zhiyan Wang, Luyang Tang, Jiayu Yang, Jie Liang, Jiahao Wu, and Ronggang Wang. Instant Gaussian Stream: Fast and generalizable streaming of dynamic scene reconstruction via gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16520–16531, 2025.

Jihwan Yoon, Sangbeom Han, Jaeseok Oh, and Minsik Lee. SplineGS: Learning smooth trajectories in gaussian splatting for dynamic scene reconstruction. In *International Conference on Learning Representations (ICLR)*, 2025.

Chao Zhang, Yifeng Zhou, Shuheng Wang, Wenfa Li, Degang Wang, Yi Xu, and Shaohui Jiao. EvolvingGS: Stable volumetric video via high-fidelity evolving 3D gaussian reconstruction. In *Proceedings of the SIGGRAPH Asia 2025 Technical Communications (SIGGRAPH Asia)*, pp. 1–4. Association for Computing Machinery, 2025a. doi: 10.1145/3757376.3771397.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595, 2018.

Wenkang Zhang, Yan Zhao, Qiang Wang, Zhixin Xu, Li Song, and Zhengxue Cheng. D-FCGS: Feedforward compression of dynamic gaussian splatting for free-viewpoint videos. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 40, pp. 16361–16369, 2026.

Xinjie Zhang, Zhening Liu, Yifan Zhang, Xingtong Ge, Dailan He, Tongda Xu, Yan Wang, Zehong Lin, Shuicheng Yan, and Jun Zhang. MEGA: Memory-efficient 4D gaussian splatting for dynamic scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27828–27838, 2025b.

Zerong Zheng, Xiaochen Zhao, Hongwen Zhang, Boning Liu, and Yebin Liu. AvatarReX: Real-time expressive full-body avatars. *ACM Transactions on Graphics*, 42(4):1–19, 2023.

Zihan Zheng, Zhenlong Wu, Houqiang Zhong, Yuan Tian, Ning Cao, Lan Xu, Jiangchao Yao, Xiaoyun Zhang, Qiang Hu, and Wenjun Zhang. 4DGCPro: Efficient hierarchical 4D gaussian compression for progressive volumetric video streaming. In *Advances in Neural Information Processing Systems*, volume 38, pp. 111881–111906, 2025. doi: 10.52202/085713-3376.

AncesTrA: Reforming Streamable Volumetric Representations through Canonical Transport Lineages and Asymmetric Basis Lifting

Appendix

Anonymous submission

OVERVIEW

The submission consists of two files: `AncesTrA.pdf`, containing the main paper and this appendix, and `Supplementary Material.zip`, containing two free-viewpoint rendering demonstration videos. The PDF provides detailed methodological derivations, implementation procedures, and additional experimental results, while the videos enable direct visual assessment of reconstruction quality, temporal coherence, and canonical reanchoring.

```
[AncesTrA Submission Materials]/
  |-- AncesTrA.pdf
  |-- Supplementary Material.zip
```

This appendix provides detailed derivations of AncesTrA's canonical transport formulation, specifies its encoding and reconstruction procedures, and presents additional evidence for long-horizon reconstruction fidelity, representation compactness, and temporal stability. Specifically, it further examines the three core components: Canonical Transport Inheritance with Entropy-Coded Residuals (CiTER), Transport-regime-aware Canonical Expiration and Reanchoring (TraCER), and Conditional Asymmetric Basis Lifting for Expressivity (CaBLE).

CiTER separates scene reconstruction from temporal inheritance by reconstructing each frame from a shared canonical representation while encoding successive transport updates through compact residuals. TraCER regulates the validity of the canonical reference through transport-regime-aware expiration and reanchoring. Complementarily, CaBLE selectively enriches symmetric Gaussian bases with asymmetric modulation to improve local expressivity around structural discontinuities. Together, these mechanisms support compact and faithful streamable reconstruction over extended temporal horizons.

The appendix begins with free-viewpoint video demonstrations (Appendix A), which provide side-by-side comparisons with iFVC and explicitly visualize canonical reanchoring during continuous scene evolution. It then offers a further discussion of related streamable volumetric representations (Appendix B), clarifying the distinctions between AncesTrA and prior codec-oriented, canonical-space, deformation-based, and system-level streaming approaches. Method derivations, algorithms, and implementation details are subsequently provided in Appendices C–F, followed by additional long-horizon results, component ablations, hyperparameter sensitivity, rate–distortion analysis, and qualitative comparisons (Appendices G–K).

Contents

A. Free-Viewpoint Video Demonstrations 17
B. Further Discussion of Related Work 17
C. Detailed Derivation of CiTER 18
D. Detailed Derivation of TraCER 31
E. Algorithm Details 40
F. Implementation Details 43

G. Additional Long-Horizon Results 44
H. Supplementary Ablation Analysis 47
I. Hyperparameter Sensitivity 49
J. Threshold-Controlled Rate–Distortion Performance 50
K. Additional Qualitative Results 51

A Free-Viewpoint Video Demonstrations

To provide a direct and immersive assessment of AncesTrA, we include free-viewpoint video demonstrations on the Trimming and Flame Steak scenes in the Supplementary Material. Each demo renders the reconstructed dynamic scene along continuously varying camera trajectories and presents a side-by-side comparison with iFVC, together with frame-wise reconstruction quality and storage statistics. This setting enables viewers to jointly inspect novel-view fidelity, temporal stability, preservation of fine local structures, and representation compactness throughout continuous scene evolution. We further visualize representative canonical refresh events, making TraCER directly observable by showing when AncesTrA terminates an aging transport lineage and reanchors subsequent evolution to a renewed canonical origin. These demonstrations therefore provide continuous spatio-temporal evidence for the central design of AncesTrA: sustaining faithful free-viewpoint rendering over extended scene evolution through canonical transport inheritance and adaptive reanchoring, while retaining a compact representation.

B Further Discussion of Related Work

Codec-oriented Gaussian streaming and canonical segmentation. Recent works have explored streamable Gaussian representations from the perspective of codec compatibility and practical delivery. Wang et al. (2024b) formulate dynamic Gaussian sequences as compact 2D Gaussian videos, enabling hardware video codecs and real-time playback on commodity and mobile devices. Their reconstruction is organized into fixed-size frame groups, where the first frame is reconstructed as a keyframe and subsequent frames are obtained sequentially through motion estimation and attribute refinement. AncesTrA addresses a different bottleneck: rather than primarily making per-frame Gaussian attributes codec-friendly, it restructures how temporal information itself is inherited. CiTER reconstructs frames from a shared canonical origin while confining cross-frame dependency to evolving transport states, thereby avoiding the use of recursively reconstructed scene states as the principal carrier of temporal evolution. Moreover, instead of relying on predetermined frame grouping, TraCER determines when the current canonical ancestry should terminate from the evolution of the transport regime itself.

AirGS (Wang et al., 2026) is more closely related in that it introduces multiple canonical spaces for long-sequence 4D Gaussian streaming. It identifies frames whose reconstruction quality falls below a prescribed threshold as new keyframes, reconstructs a new canonical space, and further incorporates bandwidth-aware Gaussian pruning for adaptive transmission. AncesTrA differs both in the criterion for canonical renewal and in the dependency structure within each canonical interval. TraCER infers canonical expiration from transport-regime statistics rather than directly thresholding reconstruction fidelity, while CiTER evolves compact transport states over a shared canonical representation instead of recursively propagating the reconstructed Gaussian state itself. Thus, AirGS primarily couples quality-aware keyframe selection with network-aware delivery, whereas AncesTrA focuses on the underlying representation problem of *what should be inherited and how long that inheritance should remain valid*.

Prior-driven and deformation-based streamable dynamics. HGC-Avatar (Tang et al., 2025a) approaches compact dynamic Gaussian transmission through a human-specific hierarchical representation, separating a motion layer based on SMPL-X pose parameters from a structural layer that maps pose-conditioned inputs to Gaussian representations through a StyleUNet-based generator. This design provides semantically compact motion coding and controllable avatar rendering, but relies on strong parametric human priors. AncesTrA does not assume such a predefined human motion model; instead, temporal variation is encoded through learned transport states inherited along canonical lineages, allowing the representation to organize scene evolution directly from the reconstructed volumetric content. Its compression mechanism is therefore complementary in nature: CiTER exploits residual structure between successive transport states, while CaBLE selectively increases local basis expressivity where symmetric Gaussian primitives become insufficient.

DeformStream (Li et al., 2024) exploits temporal redundancy from a different representation domain. It reconstructs subsequent mesh frames through embedded inter-frame deformation and couples this mechanism with a QoE model and dynamic-programming-based adaptation to changing network conditions. Its central objective is efficient mesh delivery under bandwidth and client-side

reconstruction constraints. In contrast, AncesTrA operates on Gaussian volumetric representations and treats long-horizon temporal evolution as a representation-level inheritance problem. Rather than propagating mesh geometry through deformation alone, it separates a persistent canonical state from the transport process that explains its temporal evolution, while TraCER explicitly bounds the validity of that canonical ancestry and CaBLE adapts local representation capacity around geometric and appearance discontinuities.

Representation-level streamability and system-level delivery. The broader survey of volumetric video streaming by Viola & Cesar (2023) emphasizes that "streaming" encompasses multiple levels, including representation compression, sequential delivery, network and user adaptation, and complete client–server playback systems. AncesTrA primarily addresses the *representation-level* component of this stack: its goal is to construct a compact, sequentially decodable volumetric representation whose reconstruction fidelity remains stable as temporal extent increases. It is therefore complementary to system-level mechanisms such as bandwidth-aware pruning, bitrate adaptation, or QoE-driven scheduling employed by AirGS and DeformStream. Such mechanisms could in principle operate on top of AncesTrA, whereas CiTER and TraCER address the more fundamental organization of the underlying temporal representation.

Taken together, these works expose several complementary routes toward streamable volumetric video: V^3 emphasizes codec-compatible Gaussian serialization and mobile playback, AirGS combines quality-driven canonical renewal with network-adaptive transmission, HGC-Avatar exploits human-specific structural and motion priors, and DeformStream couples mesh deformation with bandwidth-aware delivery. AncesTrA instead targets a shared representational question underlying long-horizon reconstruction: *how temporal information should be inherited when neither recursively propagated scene states nor indefinitely persistent references remain reliable*. CiTER determines *what* is propagated by confining temporal dependency to compact transport states, TraCER determines *how long* a canonical ancestor remains valid, and CaBLE determines *where* additional spatial expressivity is required. This combination shifts streamable volumetric modeling from recursive scene-state propagation toward lineage-structured inheritance, distinguishing AncesTrA as a reformulation of how long-horizon temporal evolution is organized rather than merely another compression or delivery mechanism.

C Detailed Derivation of Canonical Transport Inheritance with Entropy-Coded Residuals

This section provides a detailed formulation of **Canonical Transport Inheritance with Entropy-Coded Residuals** (CiTER). Conventional streamable reconstruction commonly couples two distinct forms of temporal dependency within a single recursive propagation pathway (Wang et al., 2023; Li et al., 2022a; Sun et al., 2024; Gao et al., 2024; Hu et al., 2025; Tang et al., 2025b; Fu et al., 2025): *scene-state dependency*, which determines the representation from which the current scene is reconstructed, and *transport dependency*, which characterizes the temporal evolution of the transformation itself. This coupling allows temporal information to be inherited across frames, but simultaneously exposes subsequent reconstructions to errors accumulated in preceding scene states. CiTER reformulates temporal inheritance by separating these dependencies: scene reconstruction is anchored to a persistent canonical representation, whereas temporal continuity is preserved through an evolving sequence of compact transport states. This separation constitutes the central design principle of CiTER, enabling temporal coherence and inter-frame compression without recursive scene-state propagation.

We first formalize canonical-referenced reconstruction and the parameterization of the transport operator, followed by a derivation of decoder-consistent transport inheritance. We then elaborate on the inter-frame codec underlying CiTER, including exact binary residual representation for the Binary Transformation Cache (BTC) (Tang et al., 2025b; Girish et al., 2024; Hu et al., 2025; Zhang et al., 2026), transition-conditioned entropy modeling, and adaptive residual coding for continuous transport-head parameters. Finally, we establish the closed-loop consistency of the resulting transport trajectory and clarify its interaction with Transport-regime-aware Canonical Expiration and Reanchoring (TraCER). Together, these derivations expose how CiTER unifies temporal optimization and entropy-efficient transmission within a common transport domain while preserving canonical-referenced scene reconstruction. The notation used throughout is summarized in Tab. 4.

Table 4: Notation used in the AncesTrA formulation and algorithms.

Notation	Description
Canonical Transport Lineages and Reconstruction	
t, m, N, K	Frame index t; canonical-lineage index m; total number of frames N; total number of canonical transport lineages K, respectively. Frames are indexed by $t \in \{0, \dots, N-1\}$.
c_m, c_{m+1}	Canonical-frame indices of the current and subsequent lineages, respectively. The canonical frames satisfy $0 = c_1 < c_2 < \cdots < c_K < N$, with $c_{K+1} = N$.
$\mathcal{S}_m, \mathcal{P}_m$	Frame set of the m-th lineage $\mathcal{S}_m = \{c_m, \dots, \min(c_{m+1}, N-1)\}$ and its predictive-frame set $\mathcal{P}_m = \mathcal{S}_m \setminus \{c_m\}$.
a_m	First predictive frame of lineage m, $a_m = c_m + 1$.
$\mathcal{L}_m$	The m-th canonical transport lineage, $\mathcal{L}_m = (\mathcal{G}^c_m, \{\hat{\boldsymbol{\theta}}_t\}_{t \in \mathcal{P}_m})$.
$\mathcal{I}_t$	Multi-view observations of frame t used to optimize or refine the current transport state.
$\mathcal{G}^c_m$	Shared canonical scene representation of lineage m. It remains fixed within the lineage and serves as the reconstruction reference for all frames in $\mathcal{P}_m$.
$\hat{\mathcal{G}}_t$	Decoded scene representation at frame t. For a canonical frame, $\hat{\mathcal{G}}_{c_m} = \mathcal{G}^c_m$; for a predictive frame, $\hat{\mathcal{G}}_t = \mathcal{T}_{\hat{\boldsymbol{\theta}}_t}(\mathcal{G}^c_m)$.
$\mathcal{T}_{\boldsymbol{\theta}}$	Transport operator that maps a canonical representation to a frame-specific representation according to transport parameters $\boldsymbol{\theta}$.
$\boldsymbol{\theta}_t, \hat{\boldsymbol{\theta}}_t$	Optimized transport state at frame t before coding and its decoded transport state after coding, respectively. Subsequent predictive frames are initialized from $\hat{\boldsymbol{\theta}}_{t-1}$.
CiTER Transport Parameterization and Binary Residual Coding	
$\mathbf{h}_t, D$	Binary BTC hash embedding $\mathbf{h}_t \in \{-1, +1\}^D$ at frame t and its embedding dimension D, respectively.
$\mathbf{w}^{\text{anc}}_t, \mathbf{w}^{\text{feat}}_t$	Parameters of the anchor-motion and feature-transformation prediction heads at frame t, respectively. Together with $\mathbf{h}_t$, they constitute the transport state $\boldsymbol{\theta}_t = (\mathbf{h}_t, \mathbf{w}^{\text{anc}}_t, \mathbf{w}^{\text{feat}}_t)$.
$\mathbf{x}, \mathbf{F}^c$	Position and feature of an anchor in the canonical representation, respectively.
$\mathbf{z}_t(\mathbf{x})$	Compact transformation feature extracted by the BTC at anchor position $\mathbf{x}$: $\mathbf{z}_t(\mathbf{x}) = \text{BTC}_{\mathbf{h}_t}(\mathbf{x})$.
$g_{\text{anc}}, g_{\text{feat}}$	Anchor-motion and feature-transformation prediction heads, respectively.
$\Delta\mathbf{x}_t, \Delta\mathbf{F}_t$	Predicted anchor-position displacement and anchor-feature residual at frame t, respectively.
$s_{m,t}, s_{f,t}$	Predicted scaling coefficients applied to the anchor-position displacement and feature residual, respectively.
$\mathbf{x}_t, \mathbf{F}_t$	Transported anchor position and feature at frame t, given by $\mathbf{x}_t = \mathbf{x} + s_{m,t}\Delta\mathbf{x}_t$ and $\mathbf{F}_t = \mathbf{F}^c + s_{f,t}\Delta\mathbf{F}_t$.
$b_{t,j}, \mathbf{b}_t$	Binary flip indicator of the j-th BTC entry and the corresponding flip-mask vector. Specifically, $b_{t,j} = \mathbb{I}[h_{t,j} \neq h_{t-1,j}]$ and $\mathbf{b}_t \in \{0,1\}^D$.
$\mathcal{C}_{k,t}, k$	Temporal-context partition $\mathcal{C}_{k,t} = \{j \mid b_{t-1,j} = k\}$, where $k \in \{0,1\}$ denotes the preceding flip state.
$p_{k,t}$	Probability that the current BTC entry flips, conditioned on its preceding flip state k.
$q_t(\mathbf{b}_t \mid \mathbf{b}_{t-1})$	Factorized conditional Bernoulli model used to entropy-code the current BTC flip mask given the preceding flip mask.
$N_{k,t}$	Number of BTC entries belonging to context $\mathcal{C}_{k,t}$, i.e., $N_{k,t} = \lvert\mathcal{C}_{k,t}\rvert$.
$\bar{p}_t, H_b(\cdot)$	Global empirical BTC flip probability and binary entropy function, respectively. They are used to characterize the ideal coding rate under an unconditional Bernoulli model.
$R^{\text{single}}_t, R^{\text{ctx}}_t$	Ideal coding costs of the single-context and transition-conditioned BTC entropy models, respectively.
CiTER Coding of Continuous Transport Heads	
r, ℓ	Transport-head type $r \in \{\text{anc}, \text{feat}\}$ and parameter-tensor index ℓ, respectively.
$\mathbf{w}^r_{t,\ell}, \hat{\mathbf{w}}^r_{t,\ell}$	The ℓ-th parameter tensor of transport head r before coding and its decoded reconstruction at frame t, respectively.

Continued on next page

Table 4 (continued)

Notation	Description
$\Delta \mathbf{w}_{t,\ell}^r$	Inter-frame parameter residual relative to the preceding decoded tensor: $\Delta \mathbf{w}_{t,\ell}^r = \mathbf{w}_{t,\ell}^r - \hat{\mathbf{w}}_{t-1,\ell}^r$.
$q_{t,\ell}^r, \mathbf{z}_{t,\ell}^r$	Uniform-quantization step and quantized integer residual for the ℓ-th tensor of head r, respectively.
$s_{t,\ell}^r$	Tensor-dependent scale statistic used by the zero-centered Gaussian entropy model for continuous transport-head residuals.
$\mathcal{M}_t^r$	Coding mode selected for transport head r at frame t, chosen from *full*, *residual*, and *skip* whenever the corresponding mode is valid.
$\mathcal{V}_t^r$	Set of valid coding modes available to transport head r at frame t.
$R_t^r(\mathcal{M})$	Complete coded size associated with coding mode $\mathcal{M}$ for transport head r at frame t.
Q_{fp16}	Deterministic fp16 projection applied when reconstructing continuous-valued transport-head parameters.
$\mathscr{B}_t, \mathscr{H}_t$	Complete coded transport payload at frame t and the compact syntax that specifies coding modes and entropy parameters, respectively.
$\mathscr{B}_t^{\mathrm{BTC}}, \mathscr{B}_t^{\mathrm{anc}}, \mathscr{B}_t^{\mathrm{feat}}$	Coded payloads corresponding to the BTC embedding, anchor-motion head, and feature-transformation head, respectively.
TraCER Transport-regime-aware Canonical Expiration and Reanchoring	
$\mathcal{D}_t$	Canonical-relative transport departure at frame t, defined as the normalized Hamming distance between the current BTC embedding $\mathbf{h}_t$ and the first predictive BTC embedding $\mathbf{h}_{a_m}$ of the current lineage.
$\mathbf{d}_t$	Sliding-window vector of transport departures, $\mathbf{d}_t = [\mathcal{D}_{t-W}, \ldots, \mathcal{D}_t]^\top$.
W	Sliding-window length used for transport-regime characterization. The window contains $W+1$ departure observations.
$u_i, \mathbf{u}$	Centered temporal coordinate $u_i = i - W/2$, $i = 0, \ldots, W$, and the corresponding coordinate vector $\mathbf{u}$.
$\hat{\beta}_t$	Least-squares estimate of the local temporal slope of the canonical-relative transport departure over the current window.
$\Delta d_{t,i}, \overline{\Delta d}_t$	Frame-wise increment $\Delta d_{t,i} = \mathcal{D}_{t-W+i} - \mathcal{D}_{t-W+i-1}$ and the mean increment over the current window, respectively.
σ_t	Standard deviation of the centered frame-wise departure increments, used to characterize local transport fluctuation.
ρ_t	Dimensionless transport-regime statistic measuring local trend strength relative to increment fluctuation. Small ρ_t indicates a weakly evolving transport regime.
ε	Numerical tolerance used to handle the degenerate case $\sigma_t \approx 0$ in the definition of ρ_t; the paper uses $\varepsilon = 10^{-8}$.
τ	TraCER stabilization threshold. A transport-regime estimate is regarded as stabilized when $\rho_t < \tau$.
C	Number of consecutive stabilized regime estimates required before canonical expiration can be declared.
$T_{\min}$	Minimum admissible duration of a canonical lineage before expiration is allowed.
$\mathcal{A}_m$	Set of frames admissible for canonical-expiration testing: $\mathcal{A}_m = \{t \mid t - c_m \geq T_{\min}, t \geq a_m + W + C\}$.
e_m	Expiration frame of lineage m, defined as the first admissible frame for which the stabilization criterion has remained satisfied for C consecutive estimates.
$c_{m+1}, \mathcal{G}_{m+1}^c$	Canonical frame index and canonical representation of the renewed lineage. Upon expiration, $c_{m+1} = e_m$ and $\mathcal{G}_{m+1}^c = \hat{\mathcal{G}}_{e_m}$.
CaBLE Conditional Asymmetric Basis Lifting	
$\mathbf{p}, \mu, \Sigma$	Spatial query position, Gaussian center, and Gaussian covariance matrix, respectively.
$G(\mathbf{p}; \mu, \Sigma)$	Conventional symmetric Gaussian spatial kernel centered at μ with covariance Σ.
$\mathbf{N}_G$	Gaussian generation network that predicts Gaussian attributes and, under CaBLE, the directional splitting parameter and bilateral opacity coefficients.
$\widetilde{\mathbf{n}}, \mathbf{n}_0$	Predicted directional parameter and fixed initialization direction, respectively, with $\mathbf{n}_0 = [0, 0, 1]^\top$.
$\mathbf{n}$	Normalized splitting normal, $\mathbf{n} = (\widetilde{\mathbf{n}} + \mathbf{n}_0)/\|\widetilde{\mathbf{n}} + \mathbf{n}_0\|_2$.

Continued on next page

Table 4 (continued)

Notation	Description
$h_{\mathbf{n}}(\mathbf{p})$	Signed coordinate of spatial position $\mathbf{p}$ relative to the learned splitting plane: $h_{\mathbf{n}}(\mathbf{p}) = \mathbf{n}^\top(\mathbf{p} - \mu)$.
$G^+_{\mathbf{n}}(\mathbf{p}), G^-_{\mathbf{n}}(\mathbf{p})$	Two complementary half-space Gaussian bases obtained by restricting the original Gaussian to $h_{\mathbf{n}}(\mathbf{p}) \geq 0$ and $h_{\mathbf{n}}(\mathbf{p}) < 0$, respectively.
α^+, α^-	Anchor-conditioned opacity coefficients applied to the positive and negative half-spaces, respectively.
a, δ	Symmetric opacity component $a = (\alpha^+ + \alpha^-)/2$ and asymmetric component $\delta = (\alpha^+ - \alpha^-)/2$ in the CaBLE basis decomposition.
$\alpha(\mathbf{p})$	Spatially asymmetric opacity modulation induced by α^+ and α^-. When $\alpha^+ = \alpha^-$, CaBLE exactly degenerates to the conventional symmetric Gaussian.

C.1 Canonical-Referenced Transport Representation

As defined in Eq. 1, the m-th canonical transport lineage is associated with a persistent canonical frame c_m and the predictive-frame set

$$\mathcal{P}_m = \{c_m + 1, \ldots, \min(c_{m+1}, N-1)\}. \tag{22}$$

Consistent with the notation adopted in the TraCER derivation, the first predictive frame of this lineage is denoted by

$$a_m = c_m + 1. \tag{23}$$

The defining property of CiTER is that every predictive frame within a lineage is reconstructed directly from the same canonical representation. Specifically, for any $t \in \mathcal{P}_m$,

$$\hat{\mathcal{G}}_t = \mathcal{T}_{\hat{\boldsymbol{\theta}}_t}(\mathcal{G}^c_m), \tag{24}$$

where $\mathcal{G}^c_m$ denotes the persistent canonical representation and $\hat{\boldsymbol{\theta}}_t$ denotes the decoded transport parameters associated with frame t.

This formulation differs from conventional recursive transformation paradigms (Tang et al., 2025b), in which each reconstructed scene state serves as the reference for its successor:

$$\hat{\mathcal{G}}_t = \mathcal{T}_{\hat{\boldsymbol{\theta}}_t}(\hat{\mathcal{G}}_{t-1}). \tag{25}$$

Although such recursion efficiently exploits adjacent-frame continuity (Fu et al., 2025; Gao et al., 2024; Sun et al., 2024), it also establishes an inseparable dependency between temporal inheritance and the quality of intermediate reconstructions. Repeated application of Eq. 25 yields

$$\hat{\mathcal{G}}_t = \mathcal{T}_{\hat{\boldsymbol{\theta}}_t} \circ \mathcal{T}_{\hat{\boldsymbol{\theta}}_{t-1}} \circ \cdots \circ \mathcal{T}_{\hat{\boldsymbol{\theta}}_{a_m}}(\mathcal{G}^c_m), \tag{26}$$

where $\circ$ denotes operator composition. Under this recursive organization, every intermediate reconstruction participates in the transformation pathway of subsequent frames. Consequently, approximation or quantization errors introduced at an earlier stage may be inherited and further amplified as the sequence progresses.

CiTER eliminates this recursive scene-state pathway within each lineage by directly associating every predictive frame with the shared canonical origin:

$$\hat{\mathcal{G}}_t = \mathcal{T}_{\hat{\boldsymbol{\theta}}_t}(\mathcal{G}^c_m), \qquad \forall t \in \mathcal{P}_m. \tag{27}$$

Unlike Eq. 26, the resulting reconstruction does not involve a composition of intermediate reconstructed scene states. Thus, for adjacent predictive frames within the same lineage, the direct scene-state dependency satisfies

$$\frac{\partial \hat{\mathcal{G}}_t}{\partial \hat{\mathcal{G}}_{t-1}} = 0, \qquad t, t-1 \in \mathcal{P}_m, \tag{28}$$

when the canonical representation and current transport are held fixed.

Eq. 28 characterizes the non-recursive structure of canonical-referenced reconstruction. Importantly, this property concerns direct dependence on the preceding reconstructed scene state; it does not imply that successive transport parameters are temporally independent. CiTER instead relocates

temporal inheritance from the reconstructed scene representation to the substantially more compact transport trajectory (Gao et al., 2024; Fu et al., 2025; Liang et al., 2026). The resulting organization preserves temporal continuity while preventing intermediate scene reconstructions from becoming obligatory references for subsequent frames.

C.2 Transport Parameterization

To instantiate canonical-referenced reconstruction, CiTER employs a Binary Transformation Cache (BTC) (Tang et al., 2025b) together with two lightweight prediction heads that characterize anchor-position and latent-feature evolution. The transport operator $\mathcal{T}_{\boldsymbol{\theta}_t}$ is parameterized by

$$\boldsymbol{\theta}_t = \left(\mathbf{h}_t, \mathbf{w}_t^{\mathrm{anc}}, \mathbf{w}_t^{\mathrm{feat}}\right), \tag{29}$$

where

$$\mathbf{h}_t \in \{-1, +1\}^D \tag{30}$$

denotes the binary BTC hash embedding, and $\mathbf{w}_t^{\mathrm{anc}}$ and $\mathbf{w}_t^{\mathrm{feat}}$ collect the parameters of the anchor-motion and feature-transformation heads, respectively.

Consider a canonical anchor with spatial position $\mathbf{x}$ and latent feature $\mathbf{F}^c$ (Lu et al., 2024; Tang et al., 2025b). The BTC first extracts a compact transformation feature conditioned on the canonical anchor position:

$$\mathbf{z}_t(\mathbf{x}) = \mathrm{BTC}_{\mathbf{h}_t}(\mathbf{x}). \tag{31}$$

Based on this shared transformation feature, the two prediction heads estimate the frame-specific geometric and latent-feature variations:

$$[\Delta\mathbf{x}_t, s_{m,t}, s_{f,t}] = g_{\mathrm{anc}}\left(\mathbf{z}_t(\mathbf{x}); \mathbf{w}_t^{\mathrm{anc}}\right), \tag{32}$$

$$\Delta\mathbf{F}_t = g_{\mathrm{feat}}\left(\mathbf{z}_t(\mathbf{x}); \mathbf{w}_t^{\mathrm{feat}}\right). \tag{33}$$

Here, $\Delta\mathbf{x}_t$ and $\Delta\mathbf{F}_t$ denote the predicted position and feature residuals, while $s_{m,t}$ and $s_{f,t}$ are the corresponding learned scaling coefficients. The transported anchor state is subsequently obtained through

$$\mathbf{x}_t = \mathbf{x} + s_{m,t}\Delta\mathbf{x}_t, \tag{34}$$

$$\mathbf{F}_t = \mathbf{F}^c + s_{f,t}\Delta\mathbf{F}_t. \tag{35}$$

This parameterization concentrates frame-specific adaptation in a compact transport operator rather than modifying the underlying canonical representation. Within each lineage, the canonical anchor state remains fixed, while the transport parameters evolve to account for the geometric and appearance variations observed at successive frames. The Gaussian-attribute decoder is further shared and fixed throughout the entire sequence (Lu et al., 2024; Kerbl et al., 2023). Consequently, the representation separates persistent scene information from frame-dependent transformation parameters, establishing the structural basis for the temporal inheritance and residual coding developed below.

C.3 Decoder-Consistent Transport Inheritance

Canonical-referenced reconstruction removes recursive dependence between scene states, but independently optimizing every transport would discard the temporal coherence that makes streamable reconstruction efficient. CiTER resolves this tension through *transport inheritance*: the decoded transport of the preceding frame provides an evolution-aware initialization for the current frame, while the canonical representation remains the sole scene-reconstruction reference within the lineage.

At the beginning of each lineage, the first predictive transport is independently optimized with respect to the newly established canonical representation:

$$\boldsymbol{\theta}_{a_m} = \mathrm{Optimize}\left(\mathcal{I}_{a_m}, \mathcal{G}_m^c\right). \tag{36}$$

This initialization establishes the transport-space origin from which subsequent temporal evolution is inherited.

For every subsequent predictive frame, CiTER follows Eq. 4 and refines the decoded transport parameters of the preceding frame (Tang et al., 2025b; Gao et al., 2024; Fu et al., 2025; Yan et al., 2025):

$$\boldsymbol{\theta}_t = \mathrm{Refine}\left(\hat{\boldsymbol{\theta}}_{t-1}; \mathcal{I}_t, \mathcal{G}_m^c\right), \qquad t > a_m. \tag{37}$$

Here, $\mathcal{I}_t$ denotes the current multi-view observations, and $\hat{\boldsymbol{\theta}}_{t-1}$ denotes the reconstructed transport state after coding and decoding. This inherited initialization transfers previously learned transformation information to the current optimization stage, allowing refinement to concentrate on the additional variation required by the incoming observations.

Crucially, inheritance operates on the *decoded* transport rather than the unquantized encoder-side parameters. This distinction establishes a common temporal reference for optimization and compression: the state used to initialize the current transport is identical to the preceding reconstruction available at the decoder. Accordingly, the transport trajectory within a lineage takes the form

$$\hat{\boldsymbol{\theta}}_{a_m} \to \hat{\boldsymbol{\theta}}_{a_m+1} \to \cdots \to \hat{\boldsymbol{\theta}}_{c_{m+1}}, \tag{38}$$

whereas every reconstructed scene state remains directly associated with the common canonical representation $\mathcal{G}_m^c$.

The resulting dependency structure is therefore asymmetric:

$$\boxed{\begin{aligned} \text{scene state:} \quad & \mathcal{G}_m^c \to \hat{\mathcal{G}}_t, \\ \text{transport state:} \quad & \hat{\boldsymbol{\theta}}_{t-1} \to \boldsymbol{\theta}_t \to \hat{\boldsymbol{\theta}}_t. \end{aligned}} \tag{39}$$

This asymmetric organization is the defining characteristic of CiTER. Persistent scene information is inherited directly from a common canonical origin, while temporal variation is inherited through successive transport states. Beyond improving optimization efficiency, the latter dependency provides a natural reference for inter-frame compression: because adjacent transports are initialized from the same evolving trajectory, their parameter differences can be represented and entropy-coded rather than repeatedly transmitting complete transport states.

C.4 Exact Binary Transport Residuals

The binary BTC embedding provides a particularly compact representation of inter-frame transport evolution (Tang et al., 2025b; Girish et al., 2024). Unlike continuous-valued parameters, whose residuals generally require quantization, the binary alphabet admits an exact transition representation that preserves the current BTC state without additional approximation.

For each embedding entry, define the binary transition indicator as in Eq. 5:

$$b_{t,j} = \mathbb{I}[h_{t,j} \neq h_{t-1,j}], \qquad b_{t,j} \in \{0, 1\}. \tag{40}$$

Since $h_{t,j}, h_{t-1,j} \in \{-1, +1\}$, the transition indicator admits the equivalent algebraic representation

$$b_{t,j} = \frac{1 - h_{t,j} h_{t-1,j}}{2}. \tag{41}$$

When the two binary states agree, $h_{t,j} h_{t-1,j} = 1$, yielding $b_{t,j} = 0$. Conversely, a sign reversal produces a product of -1 and therefore $b_{t,j} = 1$.

Given $h_{t-1,j}$ and $b_{t,j}$, the inverse mapping follows directly:

$$h_{t,j} = (1 - 2b_{t,j}) h_{t-1,j}. \tag{42}$$

Thus, conditioned on the preceding BTC state, the current embedding entry and its transition indicator are in one-to-one correspondence:

$$h_{t,j} \longleftrightarrow b_{t,j}. \tag{43}$$

Extending this identity to the complete embedding yields

$$\mathbf{h}_t \longleftrightarrow \mathbf{b}_t \quad \text{given} \quad \mathbf{h}_{t-1}. \tag{44}$$

This bijection establishes that the binary transport residual $\mathbf{b}_t$ is a lossless reparameterization of the current BTC state. Rather than transmitting the absolute binary embedding, CiTER transmits the state transitions relative to the preceding decoded transport. Each zero-valued residual preserves the corresponding binary state, while each one-valued residual specifies an exact sign reversal. Inter-frame transport evolution is thereby represented as a binary state-transition process, whose coding cost is determined by the statistical structure of the resulting flip mask.

To characterize the compressibility of this representation, consider modeling the entries of $\mathbf{b}_t$ with an unconditional Bernoulli model whose empirical flip probability is

$$\bar{p}_t = \frac{1}{D}\sum_{j=1}^{D} b_{t,j}. \tag{45}$$

Under this model, the ideal negative log-likelihood evaluated at the empirical maximum-likelihood estimate is

$$\begin{aligned} R_t^{\text{single}} &= -\sum_{j=1}^{D} \left[b_{t,j} \log_2 \bar{p}_t + (1 - b_{t,j}) \log_2 (1 - \bar{p}_t) \right] \\ &= D H_b(\bar{p}_t), \end{aligned} \tag{46}$$

where

$$H_b(p) = -p \log_2 p - (1-p) \log_2 (1-p) \tag{47}$$

denotes the binary entropy function, with the usual continuous extension at $p \in \{0, 1\}$.

Eq. 46 establishes a direct connection between temporal parameter coherence and coding efficiency. When successive transport states differ in only a small fraction of BTC entries, the empirical transition probability satisfies $\bar{p}_t \ll 1/2$, yielding

$$H_b(\bar{p}_t) \ll 1. \tag{48}$$

Consequently, the ideal coding cost of the transition mask can be substantially smaller than that of directly representing the current binary embedding. Transport inheritance therefore provides more than an optimization prior: it also induces a temporally correlated parameter trajectory whose redundancy can be exploited for entropy-efficient transmission.

C.5 Transition-Conditioned Inter-Frame Entropy Coding

Although the unconditional Bernoulli model captures the overall transition activity of the current BTC residual, it does not distinguish between entries with different temporal transition histories. In particular, entries that changed in the preceding frame may exhibit transition statistics different from those that remained unchanged. CiTER exploits this dependency through a transition-conditioned entropy model that characterizes the current flip probability using the preceding decoded flip state.

Whenever the preceding flip mask is available, the BTC indices are partitioned according to Eq. 6:

$$\mathcal{C}_{k,t} = \{ j \mid b_{t-1,j} = k \}, \qquad k \in \{0, 1\}. \tag{49}$$

The two contexts distinguish entries that remained unchanged at the preceding timestep from those that underwent a binary transition. Let

$$N_{k,t} = |\mathcal{C}_{k,t}|, \qquad N_{0,t} + N_{1,t} = D, \tag{50}$$

denote the number of embedding entries assigned to each context. For every nonempty context, the empirical conditional transition probability is estimated as

$$p_{k,t} = \frac{1}{N_{k,t}} \sum_{j \in \mathcal{C}_{k,t}} b_{t,j}. \tag{51}$$

The resulting factorized conditional Bernoulli model, consistent with the main-paper formulation, is

$$q_t(\mathbf{b}_t \mid \mathbf{b}_{t-1}) = \prod_{k=0}^{1} \prod_{j \in \mathcal{C}_{k,t}} p_{k,t}^{b_{t,j}} (1 - p_{k,t})^{1 - b_{t,j}}. \tag{52}$$

An empty context contributes an empty product and requires no associated residual symbols.

By assigning separate transition probabilities to the two contexts, the conditional model captures temporal dependencies that cannot be represented by a single marginal flip probability. Evaluating its ideal negative log-likelihood at the empirical conditional probabilities gives

$$\begin{aligned} R_t^{\text{ctx}} &= -\log_2 q_t(\mathbf{b}_t \mid \mathbf{b}_{t-1}) \\ &= \sum_{k=0}^{1} N_{k,t} H_b(p_{k,t}). \end{aligned} \tag{53}$$

The coding advantage of this temporal partition can be characterized directly through the concavity of binary entropy.

Proposition C.1 (Context-conditioned entropy does not exceed marginal entropy)**.** *Let*

$$\pi_{k,t} = \frac{N_{k,t}}{D}. \tag{54}$$

Then the ideal coding costs of the conditional and unconditional empirical Bernoulli models satisfy

$$R_t^{\mathrm{ctx}} \le R_t^{\mathrm{single}}. \tag{55}$$

Proof. The empirical marginal flip probability can be expressed as a mixture of the context-conditioned probabilities:

$$\begin{aligned} \bar{p}_t &= \frac{1}{D}\sum_{j=1}^{D} b_{t,j} \\ &= \sum_{k=0}^{1} \frac{N_{k,t}}{D}\frac{1}{N_{k,t}} \sum_{j\in\mathcal{C}_{k,t}} b_{t,j} \\ &= \sum_{k=0}^{1} \pi_{k,t} p_{k,t}, \end{aligned} \tag{56}$$

where terms corresponding to empty contexts are omitted.

The binary entropy function is strictly concave on $(0, 1)$, since

$$H_b''(p) = -\frac{1}{p(1-p)\ln 2} < 0, \qquad 0 < p < 1. \tag{57}$$

Applying Jensen's inequality therefore yields

$$\sum_{k=0}^{1} \pi_{k,t} H_b(p_{k,t}) \le H_b\left(\sum_{k=0}^{1} \pi_{k,t} p_{k,t}\right) = H_b(\bar{p}_t). \tag{58}$$

Multiplying both sides by D and substituting Eqs. 46 and 53 establishes Eq. 55. □

Because H_b is strictly concave, when both contexts are nonempty, equality holds if

$$p_{0,t} = p_{1,t}, \tag{59}$$

meaning that the preceding transition state provides no distinction between the empirical current-frame transition probabilities. If the two nonempty contexts exhibit different transition probabilities, strict concavity guarantees a strictly smaller ideal conditional entropy. When only one context is populated, the conditional and marginal models coincide.

The same result admits an information-theoretic interpretation. Consider a uniformly sampled BTC index and let B_{t-1} and B_t denote its preceding and current binary transition states, respectively. Under the empirical distribution induced by the BTC entries, Eqs. 46 and 53 correspond to

$$R_t^{\mathrm{single}} = DH(B_t), \qquad R_t^{\mathrm{ctx}} = DH(B_t \mid B_{t-1}), \tag{60}$$

and their difference becomes

$$\begin{aligned} R_t^{\mathrm{single}} - R_t^{\mathrm{ctx}} &= D\left[H(B_t) - H(B_t \mid B_{t-1})\right] \\ &= DI(B_t; B_{t-1}) \ge 0. \end{aligned} \tag{61}$$

Eq. 61 provides a precise characterization of the ideal entropy reduction: the additional coding benefit of transition conditioning is equal to the empirical mutual information between successive binary transition states, scaled by the BTC embedding dimension. Accordingly, the conditional model exploits not only the sparsity of inter-frame changes but also the temporal structure of where those changes occur.

This theoretical advantage concerns the entropy of the coded symbols under the specified empirical models. A practical bitstream additionally requires finite-precision probability signaling, arithmetic-coder termination, and stream-container metadata. Such overhead is not represented by Eq. 55 and must be accounted for when evaluating the actual coded size. The entropy analysis therefore characterizes the information available for compression, rather than guaranteeing an identical reduction in total transmitted bits.

C.6 Causal Availability of the Temporal Context

Transition-conditioned entropy coding is compatible with the causal constraints of streamable reconstruction because its context is derived entirely from previously decoded transport information. Before decoding frame t, the decoder has already recovered $\mathbf{h}_{t-1}$ and, when temporal conditioning is enabled, retained the preceding flip mask $\mathbf{b}_{t-1}$. The context partitions

$$\mathcal{C}_{0,t}, \mathcal{C}_{1,t} \tag{62}$$

are therefore deterministically available before decoding any current-frame residual symbols.

The decoding dependency can be expressed as

$$\mathbf{b}_{t-1} \longrightarrow \{\mathcal{C}_{0,t}, \mathcal{C}_{1,t}\}, \tag{63}$$

$$\mathscr{B}_t^{\mathrm{BTC}} \longrightarrow \mathbf{b}_t, \tag{64}$$

$$(\mathbf{h}_{t-1}, \mathbf{b}_t) \longrightarrow \mathbf{h}_t, \tag{65}$$

where $\mathscr{B}_t^{\mathrm{BTC}}$ denotes the transmitted BTC residual bitstream at frame t. Because both encoder and decoder construct the context partition from the same preceding flip mask, no additional index map is required to communicate the partition itself. Only the coded residual symbols and their associated entropy-model parameters need to be conveyed.

The beginning of each canonical lineage introduces a natural boundary condition. At $t = a_m$, no preceding transport exists within the newly established lineage. CiTER therefore transmits the complete BTC state $\mathbf{h}_{a_m}$ to initialize the decoder's transport reference. At the next predictive frame, $t = a_m + 1$, the preceding BTC state is available and a binary residual $\mathbf{b}_t$ can be constructed relative to $\mathbf{h}_{a_m}$. However, the lineage does not yet contain a preceding residual mask, so this first inter-frame residual is coded using the unconditional Bernoulli model.

Once $\mathbf{b}_{a_m+1}$ has been decoded, the preceding transition context becomes available for all subsequent predictive frames. The conditional model of Eq. 52 can then be applied without violating decoding causality.

The resulting coding hierarchy within each canonical lineage is

$t = a_m$:	full BTC state,
$t = a_m + 1$:	single-context BTC residual,
$t > a_m + 1$:	transition-conditioned BTC residual.

(66)

This staged initialization ensures that every temporal dependency used by the entropy model is available at the decoder before it is required. It also confines the residual history to the current canonical lineage, allowing a newly established lineage to initialize its transport and entropy contexts independently of the preceding one.

C.7 Adaptive Inter-Frame Coding of the Transport Heads

While the binary BTC embedding, one component of $\boldsymbol{\theta}_t$, admits exact transition coding, the anchor-motion and feature-transformation heads contain continuous-valued parameters whose inter-frame residuals require quantization. Applying the same binary transition model to these components would disregard their distinct statistical structure. CiTER therefore adopts a heterogeneous transport codec: exact binary residual coding for the BTC embedding and adaptive quantized residual coding for the continuous MLP heads (Wang et al., 2024a; Chen et al., 2024; 2025b; Niedermayr et al., 2024).

Let

$$r \in \{\mathrm{anc}, \mathrm{feat}\} \tag{67}$$

index the two transport heads, and let $\mathbf{w}_{t,\ell}^r$ denote the ℓ-th parameter tensor of head r at frame t. Consistent with decoder-synchronized transport inheritance, inter-frame residuals are computed relative to the preceding *decoded* tensor $\hat{\mathbf{w}}_{t-1,\ell}^r$:

$$\Delta \mathbf{w}_{t,\ell}^r = \mathbf{w}_{t,\ell}^r - \hat{\mathbf{w}}_{t-1,\ell}^r. \tag{68}$$

This choice ensures that the reference used for residual construction is identical to the state available during decoding, preventing quantization errors from introducing an encoder–decoder reference mismatch.

Because the statistical redundancy of continuous transport parameters varies across heads and timesteps, CiTER supports three coding modes: direct transmission, exact reference reuse, and entropy-coded residual transmission.

Full mode. When no compatible preceding reference is available, or when inter-frame residual coding provides no reduction in transmitted size, the current head is represented directly in fp16:

$$\mathcal{M}_t^r = \text{full}. \tag{69}$$

This mode provides an independent reconstruction of the current head and establishes a valid reference for subsequent residual coding.

Skip mode. When the current head is identical to its preceding decoded reconstruction, transmitting its parameters again is unnecessary. Specifically, if

$$\mathbf{w}_{t,\ell}^r = \hat{\mathbf{w}}_{t-1,\ell}^r, \tag{70}$$

holds for every parameter tensor ℓ, the decoder reuses the preceding head without modification:

$$\hat{\mathbf{w}}_t^r = \hat{\mathbf{w}}_{t-1}^r, \qquad \mathcal{M}_t^r = \text{skip}. \tag{71}$$

Only the corresponding mode syntax is required, allowing exact reuse of temporally unchanged transport components.

Residual mode. For a compatible preceding reference, temporal parameter differences can instead be represented through quantized residuals. For the ℓ-th parameter tensor, the uniform quantization step is defined as

$$q_{t,\ell}^r = \frac{\left\|\Delta \mathbf{w}_{t,\ell}^r\right\|_\infty}{2^{b-1}-1}, \tag{72}$$

where $b = 8$ in our implementation. For a nonzero residual tensor, the corresponding integer representation is

$$\mathbf{z}_{t,\ell}^r = \text{round}\left(\frac{\Delta \mathbf{w}_{t,\ell}^r}{q_{t,\ell}^r}\right). \tag{73}$$

The quantization step is determined by the maximum absolute residual magnitude, ensuring that the quantized values remain within the prescribed signed integer range.

Before the final fp16 projection, the scalar uniform quantizer satisfies the conventional error bound

$$\left|\Delta w - q_{t,\ell}^r \,\text{round}\left(\frac{\Delta w}{q_{t,\ell}^r}\right)\right| \le \frac{q_{t,\ell}^r}{2}. \tag{74}$$

Thus, the quantization step directly controls the maximum scalar approximation error introduced by residual discretization.

The resulting integer residuals are entropy-coded using a zero-centered Gaussian model (Wang et al., 2024a; Chen et al., 2024; 2025b; Lee et al., 2024; Fan et al., 2024). To characterize the residual distribution of each parameter tensor, CiTER constructs a tensor-dependent scale statistic

$$s_{t,\ell}^r = \max\left(\text{Std}(\Delta \mathbf{w}_{t,\ell}^r), q_{t,\ell}^r\right), \tag{75}$$

where the quantization and scale parameters required for decoding are transmitted as side information.

After entropy decoding, the reconstructed parameter tensor is obtained by adding the dequantized residual to its preceding decoded reference:

$$\hat{\mathbf{w}}_{t,\ell}^r = Q_{\text{fp16}}\left(\hat{\mathbf{w}}_{t-1,\ell}^r + q_{t,\ell}^r \mathbf{z}_{t,\ell}^r\right), \tag{76}$$

where Q_{fp16} denotes the deterministic fp16 projection adopted by the implementation. The error bound in Eq. 74 characterizes the residual quantizer before this final projection.

Importantly, residual coding is not assumed to be uniformly advantageous across all transport heads or frames. The additional cost of quantization parameters, entropy-model statistics, and coding metadata may outweigh the redundancy captured by the residual representation, particularly when the current transport undergoes substantial parameter changes. CiTER therefore selects the coding mode according to the complete transmitted size, including the required side information:

$$\mathcal{M}_t^r = \arg\min_{\mathcal{M}\in\mathcal{V}_t^r} R_t^r(\mathcal{M}), \tag{77}$$

where $\mathcal{V}_t^r \subseteq \{\text{full}, \text{residual}, \text{skip}\}$ denotes the set of valid coding modes for the current head. Skip mode is admissible only when Eq. 70 holds, whereas residual mode requires a compatible preceding decoded reference.

Through this adaptive mode-selection mechanism, CiTER accommodates heterogeneous temporal redundancy across transport components. Small parameter variations can be represented through compact entropy-coded residuals, unchanged heads can be inherited without retransmission, and substantial parameter changes can be accommodated through direct fp16 representation. This flexibility preserves the coding advantage of temporal inheritance without imposing residual coding when its complete transmission cost becomes unfavorable.

C.8 Closed-Loop Decoder Consistency

The temporal residual codec introduced above relies on previously reconstructed transport parameters as its prediction reference. While this dependency enables compact inter-frame transmission, it also imposes a strict consistency requirement: any discrepancy between the encoder-side reference and the decoder's reconstructed state would affect the interpretation of subsequent residuals and could propagate throughout the remaining lineage.

CiTER enforces this requirement through closed-loop transport reconstruction. After encoding each predictive-frame transport, the encoder applies the same decoding operations used by the receiver and commits the resulting reconstructed parameters as the reference for subsequent optimization and residual coding. The transport trajectory is therefore maintained in the decoded parameter domain rather than propagated through independently evolving encoder-side and decoder-side states.

Let

$$\mathscr{B}_t \tag{78}$$

denote the complete coded transport payload at frame t. Abstractly, the reconstruction process is expressed as

$$\hat{\boldsymbol{\theta}}_t = \mathrm{Dec}\left(\mathscr{B}_t; \hat{\boldsymbol{\theta}}_{t-1}, \mathbf{b}_{t-1}\right). \tag{79}$$

The BTC embedding is recovered exactly through Eq. 42, while the continuous transport-head parameters are reconstructed according to their selected coding modes. At the first predictive frame of a lineage, the full transport payload initializes the reference state without requiring preceding transport parameters or a preceding flip mask. For the first residual-coded frame $t = a_m + 1$, the argument $\mathbf{b}_{t-1}$ is unavailable and the decoder reduces to the unconditional residual-decoding case described in Sec. C.6; from $t \geq a_m + 2$, the preceding flip mask is available and the general form above applies directly.

The consistency of this construction follows directly from deterministic closed-loop decoding.

Proposition C.2 (Closed-loop reference consistency)**.** *Suppose the encoder and decoder establish an identical decoded transport reference at the beginning of a canonical lineage. If the encoder reconstructs every subsequent coded transport using the same deterministic decoding operations as the decoder and adopts the reconstructed state as its next temporal reference, then*

$$\hat{\boldsymbol{\theta}}_t^{\mathrm{enc}} = \hat{\boldsymbol{\theta}}_t^{\mathrm{dec}} \tag{80}$$

holds for every predictive frame within the lineage.

Proof. At the first predictive frame a_m, the transmitted full transport establishes an identical decoded reference at both sides:

$$\hat{\boldsymbol{\theta}}_{a_m}^{\mathrm{enc}} = \hat{\boldsymbol{\theta}}_{a_m}^{\mathrm{dec}}. \tag{81}$$

For the first residual-coded frame $t = a_m + 1$, both sides start from this identical decoded transport reference and apply the same unconditional residual-decoding procedure. They therefore reconstruct the same transport state and the same first flip mask $\mathbf{b}_{a_m+1}$. This establishes the base case required for transition-conditioned decoding.

For any subsequent frame $t \geq a_m + 2$, assume that encoder–decoder reference equality holds at frame $t-1$. Because the preceding BTC residual is decoded through the same deterministic binary reconstruction rule, the associated flip masks are also identical:

$$\mathbf{b}_{t-1}^{\mathrm{enc}} = \mathbf{b}_{t-1}^{\mathrm{dec}}. \tag{82}$$

Both sides therefore possess the same transport reference and temporal entropy context before processing frame t.

Given the identical transmitted payload $\mathscr{B}_t$, applying the same deterministic decoding map yields

$$\begin{aligned}\hat{\boldsymbol{\theta}}_t^{\mathrm{enc}} &= \mathrm{Dec}\left(\mathscr{B}_t; \hat{\boldsymbol{\theta}}_{t-1}^{\mathrm{enc}}, \mathbf{b}_{t-1}^{\mathrm{enc}}\right) \\ &= \mathrm{Dec}\left(\mathscr{B}_t; \hat{\boldsymbol{\theta}}_{t-1}^{\mathrm{dec}}, \mathbf{b}_{t-1}^{\mathrm{dec}}\right) \\ &= \hat{\boldsymbol{\theta}}_t^{\mathrm{dec}}.\end{aligned} \tag{83}$$

Thus, reference equality at frame $t-1$ implies equality at frame t. Together with the identical initialization at a_m and the explicitly established first residual step at $a_m + 1$, the result follows by induction. □

This proposition establishes that the transport trajectory can be inherited and entropy-coded without accumulating encoder–decoder reference mismatch. Quantization of the continuous transport heads may introduce approximation errors relative to the parameters obtained before coding, but these errors are incorporated identically into the reconstructed states maintained by both sides. They therefore do not produce divergent temporal references.

Closed-loop consistency should be distinguished from the elimination of transport approximation errors. CiTER guarantees synchronization of the decoded transport trajectory, while the fidelity of the resulting reconstruction remains dependent on the accuracy of the coded transport and the representational suitability of its canonical reference.

Accordingly, Eq. 4 is operationally realized through the following closed-loop procedure:

$$\boxed{\text{decode} \rightarrow \text{inherit} \rightarrow \text{refine} \rightarrow \text{residual-code} \rightarrow \text{decode}} \tag{84}$$

Rather than independently optimizing and transmitting successive transport networks, CiTER maintains a single decoder-synchronized temporal trajectory. This property provides the consistency required for both transport inheritance and inter-frame residual reconstruction over extended predictive sequences.

C.9 Practical Composition of the Predictive-Frame Bitstream

The complete predictive-frame bitstream integrates the discrete and continuous coding mechanisms developed above. The resulting transport payload can be represented conceptually as

$$\mathscr{B}_t = \left(\mathscr{H}_t, \mathscr{B}_t^{\mathrm{BTC}}, \mathscr{B}_t^{\mathrm{anc}}, \mathscr{B}_t^{\mathrm{feat}}\right), \tag{85}$$

where $\mathscr{H}_t$ denotes the compact syntax specifying the coding modes and entropy parameters, while the remaining components contain the BTC, anchor-motion, and feature-transformation payloads, respectively.

When transition-conditioned BTC coding is available, the binary residual payload is partitioned into two context-specific streams:

$$\mathscr{B}_t^{\mathrm{BTC}} = \left(\mathscr{B}_{0,t}^{\mathrm{BTC}}, \mathscr{B}_{1,t}^{\mathrm{BTC}}\right), \tag{86}$$

corresponding to the temporal partitions $\mathcal{C}_{0,t}$ and $\mathcal{C}_{1,t}$. Because these partitions are reconstructed deterministically from the preceding flip mask, their membership does not require additional signaling. When a preceding residual context is unavailable, BTC residual coding reduces to a single Bernoulli-coded stream.

In practice, arithmetic-coded symbols are processed in parallel chunks, introducing chunk metadata and coding overhead beyond the ideal entropy expressions derived earlier. Similarly, each continuous transport head independently selects its coding mode, with the corresponding mode syntax and residual parameters incorporated into the complete payload. The resulting codec therefore combines two complementary forms of temporal representation:

$$\boxed{\begin{array}{rl} \text{BTC:} & \text{exact binary transition coding,} \\ \text{MLP heads:} & \text{rate-adaptive closed-loop residual coding.} \end{array}} \tag{87}$$

This heterogeneous design exploits the intrinsic structure of each transport component. Binary BTC transitions are recovered exactly and modeled using temporal transition statistics, whereas continuous head parameters are quantized and adaptively coded according to their realized transmission costs. Both mechanisms operate relative to the same preceding decoded transport state, integrating parameter inheritance and compression within a unified closed-loop representation.

C.10 Connection between CiTER and TraCER

CiTER and TraCER regulate complementary temporal scales of the same canonical transport lineage. CiTER characterizes *local transport evolution* between adjacent predictive frames and exploits this dependency for optimization initialization and entropy-efficient transmission. For the binary BTC embedding, each adjacent-frame transition is represented exactly by

$$\mathbf{h}_{t-1} \xrightarrow{\mathbf{b}_t} \mathbf{h}_t. \tag{88}$$

This local representation captures which transport parameters change at each timestep without requiring the corresponding scene states to be recursively propagated.

TraCER instead characterizes the *lineage-scale evolution* of the transport trajectory relative to the first predictive transport $\mathbf{h}_{a_m}$. Its canonical-relative departure $\mathcal{D}_t$ measures the normalized Hamming distance between the current BTC state and this initial transport reference $\mathbf{h}_{a_m}$. Consequently, $\mathcal{D}_t$ reflects the net effect of the successive binary transitions generated and transmitted by CiTER, accounting for the fact that repeated sign reversals may return individual BTC entries to their initial states.

The two mechanisms therefore address distinct but closely related aspects of long-horizon temporal modeling (Xu et al., 2024; Liang et al., 2026; Wu et al., 2025). CiTER governs how the current canonical-to-frame transport can be efficiently inherited and encoded from its decoded predecessor. TraCER governs how long the current canonical representation should remain the reference for that evolving transport trajectory. Whereas CiTER exploits adjacent-frame coherence, TraCER evaluates the temporal regime of the resulting lineage and identifies when its canonical origin should be renewed.

When TraCER declares canonical expiration at frame e_m, the corresponding decoded scene state $\hat{\mathcal{G}}_{e_m}$ is promoted to the canonical reference of the next lineage. CiTER subsequently initializes a new transport trajectory from $a_{m+1} = e_m + 1$. The preceding BTC reference, residual history, and transition-conditioned entropy context are reset, ensuring that subsequent transport updates are defined relative to the renewed canonical origin.

This interaction establishes a hierarchical temporal organization:

$$\boxed{\begin{array}{ll} \textbf{within a lineage:} & \text{inherit and entropy-code transport evolution,} \\ \textbf{between lineages:} & \text{renew the canonical origin according to TraCER.} \end{array}} \tag{89}$$

Through this division of temporal roles, CiTER preserves fine-grained continuity within each canonical lineage, while TraCER regulates the validity horizon of the canonical representation supporting that continuity. Canonical renewal thus complements rather than replaces transport inheritance, allowing the framework to preserve local temporal redundancy without indefinitely relying on an increasingly distant scene reference.

Summary. CiTER reformulates streamable volumetric reconstruction as a decoder-synchronized canonical transport process, separating the persistence of scene information from the temporal evolution of its associated transformation. Within each lineage, every predictive frame is reconstructed

directly from a shared canonical representation, while successive transport states inherit their decoded predecessors to preserve temporal continuity and reduce optimization redundancy. The binary BTC embedding admits an exact flip-mask representation, allowing its inter-frame evolution to be recovered without residual quantization. Transition-conditioned entropy modeling further exploits temporal correlations between successive flip masks, reducing the ideal coding cost from $DH(B_t)$ to $DH(B_t \mid B_{t-1})$, with an entropy reduction equal to $DI(B_t; B_{t-1})$. Continuous transport-head parameters are accommodated through adaptive full, residual, and skip coding, while closed-loop reconstruction ensures that every subsequent transport is inherited from a decoder-consistent reference.

These mechanisms establish a unified transport-domain pathway for temporal optimization and inter-frame compression without recursively propagating reconstructed scene states. The resulting trajectory also provides the observable transport dynamics required by TraCER to regulate canonical expiration and renewal. Together, local transport inheritance and lineage-scale canonical reanchoring organize long-horizon scene evolution into a sequence of compact, decoder-consistent canonical transport lineages.

D Detailed Derivation of Transport-regime-aware Canonical Expiration and Reanchoring

This section presents a detailed derivation of **Transport-regime-aware Canonical Expiration and Reanchoring** (TraCER), which extends canonical transport inheritance by explicitly regulating the temporal validity of its underlying scene reference. Although CiTER eliminates recursive scene-state propagation by reconstructing predictive frames from a persistent canonical representation, the suitability of this reference cannot be assumed to remain invariant throughout prolonged scene evolution. As the scene progressively departs from its canonical state, the associated transport trajectory must accommodate increasingly distant geometric and appearance variations. The fundamental question is therefore not how frequently a canonical representation should be refreshed, but how its effective validity horizon can be inferred from the temporal behavior of the transport itself.

TraCER addresses this question by interpreting canonical-reference evolution through the dynamics of the BTC trajectory established by CiTER. Rather than imposing a predetermined lineage duration or an absolute transport-departure threshold, TraCER identifies changes in the local transport regime and uses persistent stabilization as an operational criterion for canonical renewal. We first formalize canonical-relative transport departure and establish its relationship with the entropy-coded residuals of CiTER. We then introduce a stochastic interpretation of transport-regime stabilization, motivating the use of local temporal trends rather than absolute departure magnitudes. Building on this analysis, we derive the dimensionless trend-to-fluctuation criterion and the persistent stabilization rule governing canonical expiration. Finally, we detail how the decoded representation at an expiration boundary establishes a renewed canonical origin, yielding an adaptive sequence of canonical transport lineages.

D.1 Canonical-Relative Transport Departure

Consider the m-th canonical transport lineage $\mathcal{L}_m$ defined in Sec. 3.2, with canonical frame c_m and first predictive frame

$$a_m = c_m + 1. \tag{90}$$

CiTER independently optimizes the transport associated with this first predictive frame, establishing an initial transformation between the canonical representation and the immediately subsequent scene state. Its binary BTC embedding,

$$\mathbf{h}_{a_m} = [h_{a_m,1}, \ldots, h_{a_m,D}]^\top \in \{-1, +1\}^D, \tag{91}$$

therefore provides a natural transport-space reference against which subsequent evolution within the lineage can be characterized.

To quantify this evolution, we measure the departure of the current BTC embedding from the first predictive transport. For an arbitrary frame $t \geq a_m$, define the entry-wise departure indicator

$$z_{t,j} = \mathbb{I}[h_{t,j} \neq h_{a_m,j}], \qquad j \in \{1, \ldots, D\}. \tag{92}$$

Here, $z_{t,j} = 1$ indicates that the j-th binary transport parameter differs from its initial state, whereas $z_{t,j} = 0$ indicates that it remains aligned with the lineage baseline. Aggregating these indicators across the complete BTC embedding yields the canonical-relative transport departure

$$\mathcal{D}_t = \frac{1}{D} \sum_{j=1}^{D} z_{t,j}. \tag{93}$$

By construction, $\mathcal{D}_t \in [0, 1]$ represents the normalized Hamming distance between the current transport and its lineage origin. In particular,

$$\mathcal{D}_{a_m} = 0, \tag{94}$$

since the first predictive transport is identical to the reference against which departure is measured.

The binary structure of the BTC also admits an equivalent correlation-based formulation. For any $x, y \in \{-1, +1\}$,

$$\mathbb{I}[x \neq y] = \frac{1 - xy}{2}, \tag{95}$$

because $xy = +1$ when the two binary states agree and $xy = -1$ otherwise. Applying this identity to each BTC entry gives

$$\begin{aligned} \mathcal{D}_t &= \frac{1}{2D} \sum_{j=1}^{D} (1 - h_{t,j} h_{a_m,j}) \\ &= \frac{1}{2} \left(1 - \frac{\mathbf{h}_t^\top \mathbf{h}_{a_m}}{D} \right), \end{aligned} \tag{96}$$

which recovers Eq. 8 in the main paper.

These equivalent formulations reveal two complementary properties of canonical-relative departure. Eq. 93 characterizes the fraction of binary transport parameters that have changed relative to the lineage origin, while Eq. 96 expresses the same quantity through the normalized alignment between the current and baseline BTC embeddings. Together, they establish $\mathcal{D}_t$ as an observable measure of transport-space evolution relative to a fixed reference.

Importantly, $\mathcal{D}_t$ characterizes changes in the compact transformation representation rather than measuring geometric displacement or image distortion directly. Its role is to expose how the transformation required to reconstruct the evolving scene changes relative to the transport established immediately after the canonical frame. The temporal trajectory of $\mathcal{D}_t$, rather than its absolute magnitude alone, consequently provides the basis for characterizing the evolving transport regime and regulating canonical renewal.

D.2 Relation to CiTER Transport Residuals

TraCER derives its temporal characterization from the same BTC trajectory that CiTER exploits for transport inheritance and entropy-coded transmission. The two mechanisms therefore operate on a shared representation of temporal evolution, but extract information at different temporal scales. CiTER characterizes local changes between adjacent predictive frames, whereas TraCER examines how these successive transitions collectively alter the transport state relative to its lineage origin.

Recall that CiTER represents each inter-frame BTC transition through the binary flip indicator

$$b_{t,j} = \mathbb{I}[h_{t,j} \neq h_{t-1,j}], \tag{97}$$

where $b_{t,j} = 1$ denotes a sign reversal of the j-th BTC entry between consecutive predictive frames. The corresponding frame-wise transport activity is defined as

$$\mathcal{F}_t = \frac{1}{D} \sum_{j=1}^{D} b_{t,j}. \tag{98}$$

While $\mathcal{F}_t$ quantifies the proportion of transport parameters modified during the current transition, $\mathcal{D}_t$ measures their collective departure from the fixed lineage baseline. The relationship between these quantities can be formalized through the following proposition.

Proposition D.1 (Residual control of departure variation)**.** *For every $t > a_m$,*

$$|\mathcal{D}_t - \mathcal{D}_{t-1}| \leq \mathcal{F}_t. \tag{99}$$

Proof. Consider the canonical-relative departure indicator of an arbitrary BTC entry:

$$z_{t,j} = \mathbb{I}[h_{t,j} \neq h_{a_m,j}]. \tag{100}$$

If the corresponding BTC parameter remains unchanged between consecutive frames, i.e., $h_{t,j} = h_{t-1,j}$, its relationship with the fixed lineage baseline is necessarily preserved. Therefore, $z_{t,j} = z_{t-1,j}$. Conversely, a change in the canonical-relative departure indicator requires the corresponding BTC parameter to undergo an inter-frame transition. It follows that

$$|z_{t,j} - z_{t-1,j}| \leq \mathbb{I}[h_{t,j} \neq h_{t-1,j}] = b_{t,j}. \tag{101}$$

Aggregating this relation over all BTC entries and applying the triangle inequality yields

$$\begin{aligned} |\mathcal{D}_t - \mathcal{D}_{t-1}| &= \left| \frac{1}{D} \sum_{j=1}^{D} (z_{t,j} - z_{t-1,j}) \right| \\ &\leq \frac{1}{D} \sum_{j=1}^{D} |z_{t,j} - z_{t-1,j}| \\ &\leq \frac{1}{D} \sum_{j=1}^{D} b_{t,j} = \mathcal{F}_t. \end{aligned} \tag{102}$$

This establishes Eq. 99. □

The proposition establishes that the instantaneous variation in canonical-relative departure is bounded by the transport activity encoded through the CiTER residual. However, the two quantities are not generally identical: a binary transition may either increase or decrease departure from the lineage baseline. To characterize this distinction explicitly, we decompose the current transport activity according to the direction of each transition relative to the initial BTC state.

Specifically, define

$$\mathcal{F}_t^+ = \frac{1}{D} \sum_{j=1}^{D} \mathbb{I}[z_{t-1,j} = 0,\ z_{t,j} = 1], \tag{103}$$

$$\mathcal{F}_t^- = \frac{1}{D} \sum_{j=1}^{D} \mathbb{I}[z_{t-1,j} = 1,\ z_{t,j} = 0]. \tag{104}$$

Here, $\mathcal{F}_t^+$ measures transitions away from the lineage baseline, whereas $\mathcal{F}_t^-$ measures transitions that restore agreement with it. Because each BTC entry is binary, every sign reversal necessarily belongs to exactly one of these two categories. Consequently,

$$\mathcal{F}_t = \mathcal{F}_t^+ + \mathcal{F}_t^-, \qquad \mathcal{D}_t - \mathcal{D}_{t-1} = \mathcal{F}_t^+ - \mathcal{F}_t^-. \tag{105}$$

Eq. 105 exposes a fundamental distinction between *transport activity* and *net canonical-relative evolution*. The CiTER residual magnitude $\mathcal{F}_t$ measures the total proportion of BTC entries undergoing state transitions, irrespective of their direction relative to the lineage origin. In contrast, the increment of $\mathcal{D}_t$ captures the imbalance between transitions away from and toward that origin. A small departure increment therefore does not necessarily imply that the underlying transport has become inactive: it may also arise when opposing transitions approximately cancel.

This distinction is central to the design of TraCER. The absolute magnitude of canonical-relative departure does not directly reveal whether the transport is undergoing sustained directional evolution, fluctuating around a relatively stable configuration, or changing only marginally between frames. TraCER therefore characterizes the local dynamical regime of the departure trajectory rather than treating its absolute magnitude as a direct measure of representational capacity.

The resulting interpretation also clarifies the complementary roles of CiTER and TraCER. CiTER exploits adjacent-frame transport redundancy to support compact residual transmission and efficient

optimization initialization. TraCER interprets the longer-term evolution of the same transport trajectory relative to its lineage origin, providing the temporal evidence used to determine when that origin should be renewed.

D.3 A Stochastic Interpretation of Transport-Regime Stabilization

The preceding analysis establishes a direct relationship between the local transport transitions encoded by CiTER and the canonical-relative departure observed by TraCER. We next examine why this departure may evolve from a regime of sustained directional change toward a regime characterized by a diminishing temporal trend. More importantly, we show why the departure magnitude at which stabilization occurs need not be consistent across different lineages.

To obtain an analytically tractable interpretation, we consider a first-order stochastic model of canonical-relative BTC evolution. This model serves to characterize the possible structure of the transport trajectory and motivate the subsequent regime criterion; it is not an assumption imposed on the actual transport process during reconstruction.

Let

$$\Delta = t - a_m \tag{106}$$

denote the temporal distance from the first predictive frame of the current lineage. For the j-th BTC entry, consider the binary canonical-relative state

$$Z_\Delta^{(j)} = \mathbb{I}[h_{a_m+\Delta,j} \neq h_{a_m,j}]. \tag{107}$$

The state $Z_\Delta^{(j)} = 0$ indicates agreement with the initial transport, while $Z_\Delta^{(j)} = 1$ indicates departure from that baseline. For analysis, suppose that this state follows a first-order, time-homogeneous two-state transition process with probabilities

$$p = \Pr\left(Z_{\Delta+1}^{(j)} = 1 \mid Z_\Delta^{(j)} = 0\right), \tag{108}$$

$$r = \Pr\left(Z_{\Delta+1}^{(j)} = 0 \mid Z_\Delta^{(j)} = 1\right). \tag{109}$$

The parameter p characterizes transitions away from the lineage baseline, whereas r characterizes transitions that restore agreement with it. The corresponding transition matrix is

$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ r & 1-r \end{bmatrix}. \tag{110}$$

To characterize the expected evolution of this process, define

$$e_\Delta = \mathbb{E}[Z_\Delta^{(j)}] = \Pr(Z_\Delta^{(j)} = 1), \tag{111}$$

where the equality follows because $Z_\Delta^{(j)} \in \{0, 1\}$ is a binary indicator. Applying the law of total probability to the two possible preceding states gives

$$\begin{aligned} e_{\Delta+1} = {} & \Pr(Z_{\Delta+1}^{(j)} = 1 \mid Z_\Delta^{(j)} = 0) \Pr(Z_\Delta^{(j)} = 0) \\ & + \Pr(Z_{\Delta+1}^{(j)} = 1 \mid Z_\Delta^{(j)} = 1) \Pr(Z_\Delta^{(j)} = 1). \end{aligned} \tag{112}$$

By definition,

$$\Pr(Z_{\Delta+1}^{(j)} = 1 \mid Z_\Delta^{(j)} = 0) = p, \tag{113}$$

and

$$\Pr(Z_{\Delta+1}^{(j)} = 1 \mid Z_\Delta^{(j)} = 1) = 1 - r. \tag{114}$$

Substituting these transition probabilities, together with $\Pr(Z_\Delta^{(j)} = 1) = e_\Delta$ and $\Pr(Z_\Delta^{(j)} = 0) = 1 - e_\Delta$, yields

$$\begin{aligned} e_{\Delta+1} &= p(1 - e_\Delta) + (1 - r)e_\Delta \\ &= p + (1 - p - r)e_\Delta. \end{aligned} \tag{115}$$

Introducing

$$\lambda = 1 - p - r, \tag{116}$$

the stationary departure probability q_∞ satisfies the fixed-point relation

$$q_\infty = p + \lambda q_\infty. \tag{117}$$

For $p + r > 0$, solving this equation gives

$$q_\infty = \frac{p}{p+r}. \tag{118}$$

Since the initial transport is identical to the lineage baseline, $Z_0^{(j)} = 0$ and hence $e_0 = 0$. Subtracting the fixed-point relation from Eq. 115, we obtain

$$e_{\Delta+1} - q_\infty = \lambda(e_\Delta - q_\infty). \tag{119}$$

Repeated substitution then yields

$$e_\Delta - q_\infty = \lambda^\Delta(e_0 - q_\infty) = -q_\infty \lambda^\Delta. \tag{120}$$

The expected canonical-relative departure therefore admits the closed-form expression

$$\boxed{e_\Delta = q_\infty \left(1 - \lambda^\Delta\right) = \frac{p}{p+r}\left[1 - (1-p-r)^\Delta\right].} \tag{121}$$

For the non-oscillatory regime

$$0 \le \lambda < 1 \quad \Longleftrightarrow \quad 0 < p + r \le 1, \tag{122}$$

the expected departure approaches the stationary level q_∞ monotonically. More importantly, the corresponding expected frame-wise increment satisfies

$$e_{\Delta+1} - e_\Delta = q_\infty(1-\lambda)\lambda^\Delta. \tag{123}$$

Thus, the expected directional evolution decays geometrically as the transport process approaches its stationary regime, even though its limiting departure magnitude q_∞ may remain substantial.

This result provides the central motivation for regime-aware canonical expiration. An absolute threshold on $\mathcal{D}_t$ would implicitly assume that the effective renewal boundary occurs at a common departure magnitude across different lineages. However, Eq. 118 shows that even this elementary stochastic model admits different asymptotic departure levels according to the relative transition tendencies p and r. A universal magnitude threshold would consequently conflate the characteristic departure scale of a lineage with the dynamical regime of its evolution. In contrast, Eq. 123 demonstrates that a diminishing local trend characterizes the approach toward stabilization without requiring different trajectories to converge to the same departure magnitude.

The homogeneous model can be further generalized to account for heterogeneous transport components. Suppose different BTC entries exhibit distinct transition probabilities (p_j, r_j), and define

$$q_{\infty,j} = \frac{p_j}{p_j + r_j}, \qquad \lambda_j = 1 - p_j - r_j, \tag{124}$$

for entries with $p_j + r_j > 0$. By linearity of expectation, the expected canonical-relative departure becomes

$$\mathbb{E}[\mathcal{D}_{a_m+\Delta}] = \frac{1}{D}\sum_{j=1}^{D} q_{\infty,j}\left(1 - \lambda_j^\Delta\right), \tag{125}$$

without requiring statistical independence among BTC entries. Entries that remain fixed contribute zero departure.

This formulation reveals that heterogeneous transport evolution generally produces a superposition of multiple characteristic temporal rates rather than a single exponential trajectory. Different scene dynamics and transport components may therefore exhibit distinct stabilization profiles, making it inappropriate to assume a universal temporal evolution curve. The appropriate characterization is instead the local regime exhibited by the observed transport trajectory.

We emphasize that Eqs. 121–125 provide an analytical interpretation of transport-regime stabilization rather than a parametric model fitted during reconstruction. TraCER itself does not estimate p, r, q_∞, or λ, nor does it assume that the BTC trajectory is stationary or time-homogeneous. Instead, it directly characterizes the recent evolution of the observed canonical-relative departure $\mathcal{D}_t$, enabling canonical renewal to be governed by transport dynamics without explicitly modeling their underlying stochastic process.

D.4 Local Transport Trend Estimation

Motivated by the preceding analysis, TraCER characterizes the local evolution of canonical-relative departure without fitting a global temporal model. This is important because the transport trajectory may exhibit heterogeneous transition statistics and distinct evolution regimes throughout a lineage. A sliding-window formulation allows the renewal criterion to respond to recent transport behavior while remaining independent of assumptions about the complete trajectory.

Specifically, TraCER maintains a window of $W+1$ consecutive departure observations:

$$\mathbf{d}_t = [\mathcal{D}_{t-W}, \ldots, \mathcal{D}_t]^\top. \tag{126}$$

Assigning equally spaced temporal coordinates

$$x_i = i, \qquad i = 0, \ldots, W, \tag{127}$$

their mean is

$$\bar{x} = \frac{1}{W+1} \sum_{i=0}^{W} i = \frac{W}{2}. \tag{128}$$

The corresponding centered coordinates are

$$u_i = x_i - \bar{x} = i - \frac{W}{2}, \tag{129}$$

which satisfy

$$\sum_{i=0}^{W} u_i = 0. \tag{130}$$

To estimate the systematic component of local transport evolution, consider the linear approximation

$$\mathcal{D}_{t-W+i} \approx \alpha_t + \beta_t i. \tag{131}$$

Here, α_t denotes the local intercept and β_t characterizes the directional evolution rate of canonical-relative departure within the current window. The corresponding least-squares estimates are obtained by minimizing

$$(\widehat{\alpha}_t, \widehat{\beta}_t) = \arg\min_{\alpha,\beta} \sum_{i=0}^{W} \left(\mathcal{D}_{t-W+i} - \alpha - \beta i\right)^2. \tag{132}$$

Using the centered temporal-coordinate vector $\mathbf{u}$, the ordinary-least-squares estimate of the local slope simplifies to

$$\widehat{\beta}_t = \frac{\sum_{i=0}^{W} u_i \mathcal{D}_{t-W+i}}{\sum_{i=0}^{W} u_i^2} = \frac{\mathbf{u}^\top \mathbf{d}_t}{\mathbf{u}^\top \mathbf{u}}, \tag{133}$$

recovering Eq. 9 in the main paper.

For equally spaced temporal coordinates, the denominator admits the closed-form expression

$$\mathbf{u}^\top \mathbf{u} = \frac{W(W+1)(W+2)}{12}. \tag{134}$$

The estimated slope $\widehat{\beta}_t$ measures the signed local rate of canonical-relative transport evolution. A positive value indicates that the transport is moving away from its lineage baseline on average, whereas a negative value indicates a net return toward that baseline. Since TraCER is concerned with the persistence of systematic evolution rather than its direction, the subsequent regime criterion uses the magnitude $|\widehat{\beta}_t|$.

Importantly, local slope estimation characterizes the recent trajectory rather than prescribing how transport should evolve globally. It therefore accommodates heterogeneous and nonstationary transport behavior while providing a compact statistic from which the current dynamical regime can be inferred.

D.5 Dimensionless Transport-Regime Criterion

Although the local slope characterizes systematic transport evolution, its absolute magnitude alone is insufficient to identify a stabilized regime. The significance of a particular slope depends on the characteristic variation of the trajectory within the same temporal window. A small trend may indicate meaningful directional evolution in an otherwise nearly constant trajectory, whereas a comparable trend may be negligible relative to substantial frame-wise fluctuations.

TraCER addresses this ambiguity by normalizing the local trend against the fluctuation of successive departure increments. This formulation separates systematic evolution from local irregularity and produces a dimensionless criterion that can be applied across lineages with different characteristic departure scales.

For $i = 1, \ldots, W$, define the frame-wise departure increments

$$\Delta d_{t,i} = \mathcal{D}_{t-W+i} - \mathcal{D}_{t-W+i-1}, \tag{135}$$

with empirical mean

$$\overline{\Delta d}_t = \frac{1}{W} \sum_{i=1}^{W} \Delta d_{t,i}. \tag{136}$$

The local increment variance is then defined as

$$\sigma_t^2 = \frac{1}{W} \sum_{i=1}^{W} \left(\Delta d_{t,i} - \overline{\Delta d}_t \right)^2. \tag{137}$$

The use of centered increments, rather than the variance of the departure observations themselves, is essential for separating systematic evolution from local fluctuation. To demonstrate this property, consider the following local signal-plus-fluctuation model:

$$\mathcal{D}_{t-W+i} = \alpha + \beta i + \xi_i, \tag{138}$$

where β denotes the systematic evolution rate and ξ_i represents zero-mean local variation. The corresponding frame-wise increments become

$$\Delta d_{t,i} = \beta + \xi_i - \xi_{i-1}. \tag{139}$$

Their empirical mean is

$$\begin{aligned} \overline{\Delta d}_t &= \beta + \frac{1}{W} \sum_{i=1}^{W} (\xi_i - \xi_{i-1}) \\ &= \beta + \frac{\xi_W - \xi_0}{W}, \end{aligned} \tag{140}$$

where the intermediate fluctuation terms cancel telescopically. Subtracting this mean from each increment gives

$$\Delta d_{t,i} - \overline{\Delta d}_t = (\xi_i - \xi_{i-1}) - \frac{\xi_W - \xi_0}{W}. \tag{141}$$

Eq. 141 establishes that the deterministic linear trend β cancels exactly from the centered increment fluctuations. Consequently, σ_t characterizes local variation around the mean rate of transport evolution rather than conflating that variation with the systematic trend itself. Directly computing the variance of $\mathcal{D}_t$ would not provide this separation, since the variation of departure observations generally increases with the magnitude of the underlying linear trend.

Building on this decomposition, TraCER defines the regular-case transport-regime statistic as

$$\rho_t = \frac{|\widehat{\beta}_t|}{\sigma_t}, \qquad \sigma_t \geq \varepsilon. \tag{142}$$

The ratio compares the strength of systematic directional evolution with the characteristic fluctuation of frame-wise transport departure. Since both quantities measure changes in the same normalized departure variable per frame, their ratio is dimensionless.

An additional property of this construction is its invariance to affine rescaling of the observed departure trajectory within the regular regime. Consider

$$\mathcal{D}'_t = a\mathcal{D}_t + b, \qquad a \neq 0. \tag{143}$$

The corresponding slope and increment-fluctuation statistics satisfy

$$|\widehat{\beta}'_t| = |a||\widehat{\beta}_t|, \qquad \sigma'_t = |a|\sigma_t, \tag{144}$$

and therefore, provided both trajectories remain in the regular branch,

$$\rho'_t = \rho_t. \tag{145}$$

This invariance clarifies why trend normalization is appropriate for canonical lineages with heterogeneous transport characteristics. Whereas an absolute departure or slope threshold would depend on the characteristic magnitude of the observed trajectory, the normalized ratio compares systematic evolution against the fluctuation scale intrinsic to the current lineage. TraCER can therefore characterize transport regimes without requiring different lineages to exhibit comparable absolute departure levels.

The regular formulation, however, requires separate treatment in the degenerate case $\sigma_t \approx 0$, where the increment fluctuation approaches zero. Two qualitatively distinct regimes must be distinguished in this case. If the transport follows an approximately deterministic but nonzero trend, a vanishing fluctuation should not be interpreted as evidence of stabilization. Conversely, when both the local trend and increment fluctuation are negligible, the departure trajectory is effectively constant.

To handle these cases, TraCER adopts the following piecewise definition:

$$\rho_t = \begin{cases} |\widehat{\beta}_t|/\sigma_t, & \sigma_t \geq \varepsilon, \\ +\infty, & \sigma_t < \varepsilon,\ |\widehat{\beta}_t| > \varepsilon, \\ 0, & \sigma_t < \varepsilon,\ |\widehat{\beta}_t| \leq \varepsilon, \end{cases} \tag{146}$$

where $\varepsilon = 10^{-8}$, consistent with the main paper. The infinite branch prevents nearly deterministic directional evolution from being incorrectly classified as stabilized, while the zero branch identifies an effectively constant trajectory. The numerical tolerance also defines the boundary of the regular regime in which the affine-invariance property applies.

The resulting statistic ρ_t should therefore be interpreted as a *trend-to-fluctuation ratio*. A large value indicates that the recent transport trajectory exhibits systematic directional evolution relative to its local irregularity. A small value indicates that its directional trend is weak relative to frame-wise fluctuation, or that the trajectory has become effectively constant.

We emphasize that ρ_t characterizes the observed transport regime and serves as an operational criterion for canonical renewal, rather than estimating the probability of canonical expiration. TraCER therefore identifies persistent weak-trend behavior and declares a candidate canonical-renewal boundary when

$$\rho_t < \tau. \tag{147}$$

This criterion provides a representation-level signal for determining when the current lineage should be renewed.

D.6 Persistent Regime Stabilization and Canonical Expiration

The transport-regime statistic provides a local characterization of canonical-relative evolution, but an isolated low value of ρ_t may arise from transient fluctuations rather than a persistent change in the underlying transport regime. Directly renewing the canonical reference in response to individual estimates could therefore introduce unnecessary lineage boundaries and additional representation overhead.

TraCER mitigates this sensitivity by requiring temporal persistence before declaring canonical expiration. Specifically, the renewal decision combines three complementary conditions: sufficient transport history for regime estimation, a minimum admissible lineage duration, and consecutive confirmation that the observed transport regime satisfies the stabilization criterion.

Let C denote the number of consecutive stabilized estimates required for renewal, and let $T_{\min}$ denote the minimum admissible duration of a canonical lineage. Because the regime statistic at frame t depends on the sliding window

$$\{\mathcal{D}_{t-W}, \ldots, \mathcal{D}_t\}, \tag{148}$$

meaningful estimation requires sufficient departure observations to have accumulated after the first predictive frame a_m.

TraCER therefore restricts expiration testing to the admissible candidate set

$$\mathcal{A}_m = \{t \mid t - c_m \geq T_{\min},\ t \geq a_m + W + C\}, \tag{149}$$

which is consistent with Eq. 12 in the main paper. The first constraint prevents premature renewal before the current canonical representation has reached its minimum permitted duration. The second imposes a conservative warm-up period, ensuring that sufficient lineage history is available for sliding-window estimation and repeated stabilization assessment.

Within the admissible set, canonical expiration is declared at the earliest frame for which the stabilization criterion has been satisfied over the preceding C consecutive estimates:

$$e_m = \min \left\{ t \in \mathcal{A}_m \;\middle|\; \max_{0 \leq k < C} \rho_{t-k} < \tau \right\}. \tag{150}$$

The maximum operator provides a compact formulation of temporal persistence:

$$\max_{0 \leq k < C} \rho_{t-k} < \tau \quad \Longleftrightarrow \quad \rho_t < \tau,\ \rho_{t-1} < \tau,\ \ldots,\ \rho_{t-C+1} < \tau. \tag{151}$$

Consequently, a single low regime estimate is insufficient to terminate the current lineage. Canonical expiration requires the weak-trend condition to persist throughout the prescribed confirmation interval, subject to the minimum-duration and warm-up constraints.

This construction turns local transport statistics into a temporally constrained renewal policy. The minimum-duration condition limits excessively frequent reanchoring, while consecutive confirmation reduces sensitivity to short-lived fluctuations in the departure trajectory. Unlike fixed-period refresh strategies, the resulting lineage duration is determined by the observed temporal behavior of the transport rather than by a predetermined number of frames.

The interpretation of Eq. 150 warrants an important distinction. Transport-regime stabilization is employed as an observable signal for defining a canonical-renewal boundary; it is not assumed to prove that the BTC has exhausted its representational capacity or that the canonical reference has become geometrically invalid. Different scene dynamics and optimization processes may produce weak local trends for different reasons. TraCER therefore treats persistent stabilization as an operational criterion for renewing the canonical origin, allowing the lineage structure to adapt to transport evolution without requiring a universal departure threshold or an explicit estimate of the remaining representational capacity.

D.7 Reanchoring and Lineage Renewal

Once TraCER identifies the expiration frame e_m, the current lineage is terminated and a renewed canonical reference is established from the corresponding decoded scene representation. Specifically,

$$c_{m+1} = e_m, \qquad \mathcal{G}^c_{m+1} = \hat{\mathcal{G}}_{e_m}. \tag{152}$$

The decoded representation at the expiration boundary therefore assumes a new role as the canonical origin for subsequent scene evolution. Importantly, this promotion changes its reference role rather than its original predictive-frame coding identity: the reconstructed state is inherited as the canonical representation of the next lineage without requiring a new independently optimized scene representation at the renewal boundary.

The first predictive frame of the renewed lineage is

$$a_{m+1} = c_{m+1} + 1. \tag{153}$$

CiTER independently initializes the corresponding transport relative to the renewed canonical representation, establishing a new BTC baseline $\mathbf{h}_{a_{m+1}}$. By construction, the canonical-relative departure of this first predictive transport satisfies

$$\begin{aligned} \mathcal{D}_{a_{m+1}} &= \frac{1}{D} \sum_{j=1}^{D} \mathbb{I}[h_{a_{m+1},j} \neq h_{a_{m+1},j}] \\ &= 0. \end{aligned} \tag{154}$$

Thus, the departure reset in Eq. 15 follows directly from redefining the transport-space origin and introduces no additional approximation.

Canonical renewal also requires the temporal dependencies maintained by CiTER to be reinitialized. The preceding BTC reference, inter-frame residual history, and transition-conditioned entropy context are reset at the lineage boundary. The first predictive transport of the new lineage establishes an independent reference, and subsequent frames inherit and entropy-code their transformations relative to this renewed trajectory. In this way, scene-reference renewal and transport-reference renewal remain synchronized within the decoder-consistent representation.

The resulting temporal organization can be expressed as a sequence of canonically anchored lineages (Xu et al., 2024; Wu et al., 2025; Kwak et al., 2025; Liang et al., 2026),

$$\mathcal{L}_1 \to \mathcal{L}_2 \to \cdots \to \mathcal{L}_K, \tag{155}$$

where every predictive frame within lineage $\mathcal{L}_m$ is reconstructed directly from its corresponding persistent canonical representation:

$$\hat{\mathcal{G}}_t = \mathcal{T}_{\hat{\boldsymbol{\theta}}_t}(\mathcal{G}^c_m), \qquad t \in \mathcal{P}_m. \tag{156}$$

Within each lineage, temporal continuity is maintained through the inherited transport trajectory of Eq. 4, without recursively transforming preceding reconstructed scene states. At selected lineage boundaries, TraCER renews the canonical origin by promoting a decoded representation, thereby establishing a new reference for subsequent evolution.

This organization introduces two complementary temporal scales into streamable volumetric reconstruction. CiTER preserves fine-grained temporal coherence through adjacent-frame transport inheritance and entropy-coded residuals, while TraCER governs the longer-term validity of the canonical representation supporting that inheritance. Their interaction replaces unrestricted frame-wise scene-state recursion with structured canonical ancestry and transport-regime-aware renewal, allowing temporal dependencies to remain concentrated in compact transport trajectories while periodically updating the scene reference from which those trajectories are defined.

Summary. TraCER formulates canonical renewal as an adaptive decision derived from the dynamical regime of the transport trajectory rather than from a predetermined reference lifetime. Building on the binary BTC evolution established by CiTER, canonical-relative departure $\mathcal{D}_t$ characterizes the changing relationship between the current transport and its lineage origin. The residual analysis distinguishes total transport activity from net directional departure, while the stochastic interpretation demonstrates why stabilized trajectories may exhibit heterogeneous limiting departure magnitudes. Motivated by these properties, TraCER estimates the local trend $\widehat{\beta}_t$, normalizes it by the characteristic increment fluctuation σ_t, and obtains the dimensionless regime statistic ρ_t. Persistent weak-trend behavior, subject to sufficient temporal history and the minimum lineage duration, defines the operational expiration boundary e_m. The corresponding decoded representation is subsequently promoted to establish a renewed canonical origin, with the BTC baseline and residual-coding context reinitialized for the next lineage. Through this coordinated renewal of scene and transport references, TraCER extends the local temporal inheritance of CiTER into an adaptive sequence of canonical transport lineages, providing a structured mechanism for organizing prolonged scene evolution without reverting to frame-wise recursive scene-state propagation.

E Algorithm Details

This section provides a complete algorithmic account of how AncesTrA constructs, inherits, and renews canonical transport lineages throughout streamable reconstruction. The overall pipeline consists of two coupled procedures. Algorithm 1 first specifies the residual entropy codec used by CiTER to represent successive Binary Transformation Cache (BTC) states through compact, decoder-consistent updates. Algorithm 2 then integrates this codec into the complete reconstruction pipeline, where CiTER governs transport inheritance, TraCER determines canonical expiration and reanchoring, and CaBLE enhances local Gaussian expressivity. Together, these procedures establish how AncesTrA maintains temporal continuity through evolving transport states while reconstructing each frame directly from its active canonical reference, thereby separating inter-frame parameter inheritance from recursive scene-state propagation.

Algorithm 1 details the transition-conditioned branch of binary transport residual encoding and decoding within a canonical lineage. Given the current BTC state and its preceding decoded reference, the encoder first constructs a binary flip mask that identifies the entries whose signs have changed. When the preceding flip mask is available, these entries are partitioned into two temporal contexts according to whether their corresponding BTC parameters changed in the previous frame. The encoder estimates the empirical flip probability within each context and transmits the resulting probability parameters as part of the bitstream syntax. Arithmetic coding then encodes the current flip mask under the corresponding conditional Bernoulli model, exploiting both the sparsity of parameter changes and the temporal dependence of their locations. During decoding, the same context partitions are reconstructed from the previously decoded flip mask, eliminating the need to transmit their membership explicitly. The decoder recovers the current flip mask using the transmitted probability parameters and reconstructs the BTC state through exact binary sign reversal. For the first predictive frame of a newly established lineage, the complete BTC state is transmitted to initialize the transport reference. The first subsequent residual uses a single-context model until the preceding flip mask becomes available. This initialization ensures that all temporal dependencies required by the residual codec are causally available at the decoder.

Algorithm 2 describes how residual coding is integrated into frame-wise reconstruction. The initial frame is independently optimized to establish the first canonical representation, which remains the shared reconstruction reference until the current lineage expires. For the first predictive frame of each lineage, CiTER independently initializes and optimizes a new transport state with respect to the active canonical representation. Subsequent predictive frames instead inherit the preceding decoded transport parameters and refine them using the incoming multi-view observations. This distinction allows the initial transport to establish a transformation appropriate to the current canonical state, while later frames concentrate optimization on the additional variation required by scene evolution. Following optimization, the current transport is encoded relative to its decoded predecessor: the BTC embedding is represented using Algorithm 1, whereas the continuous anchor-motion and feature-transformation heads employ the adaptive coding modes detailed in Appendix C. The encoded transport is then reconstructed through the same decoding operations used by the receiver, ensuring that subsequent optimization and residual coding inherit a common decoded reference. Importantly, the inherited transport serves only as the temporal initialization and coding reference. The current scene representation is reconstructed directly from the active canonical representation rather than recursively transforming the preceding reconstructed scene state.

The reconstruction process further incorporates TraCER to regulate the temporal validity of the active canonical reference. After each predictive frame has been reconstructed, its decoded BTC state is compared with the transport baseline established by the first predictive frame of the current lineage. The resulting canonical-relative departure is incorporated into a sliding temporal window, from which TraCER estimates the local transport trend, incremental fluctuation, and corresponding dimensionless regime statistic. Canonical expiration is considered only after sufficient temporal observations have been accumulated and the minimum lineage duration has been satisfied. The expiration criterion additionally requires the stabilized transport regime to persist for the prescribed number of consecutive estimates, preventing isolated fluctuations from triggering premature reanchoring. Once these conditions are met, the current decoded scene representation is promoted to the canonical reference of the next lineage. The promoted frame retains its predictive-frame coding identity; reanchoring changes its reference role without requiring an independently reconstructed I-frame. The transport baseline, residual-coding history, and regime statistics are subsequently reset, allowing the next predictive frame to establish a new transport trajectory from the renewed canonical representation.

CaBLE operates throughout this procedure as part of the shared Gaussian generation process. For each reconstructed frame, the transported anchor positions and features are decoded into Gaussian attributes, including the learned splitting direction and bilateral opacity coefficients introduced by CaBLE. These attributes enable asymmetric opacity modulation where additional local expressivity is required, while preserving the conventional symmetric Gaussian as a special case. The Gaussian decoder is shared across the sequence, so canonical renewal updates the anchor reference without introducing an independent attribute decoder for each lineage. In combination, Algorithms 1 and 2 realize the complementary roles of the three AncesTrA components: CiTER preserves temporal continuity through decoder-consistent transport inheritance and entropy-coded residuals, TraCER regulates canonical validity through transport-regime-aware reanchoring, and CaBLE enriches the

spatial representation used to reconstruct each frame. Their coordination organizes long-horizon scene evolution into successive canonical transport lineages, maintaining a persistent reconstruction reference within each lineage while adapting both temporal transport and local expressivity to the evolving scene.

Algorithm 1: Residual Entropy Coding of Transport States

Input: current BTC state $\mathbf{h}_t$, preceding BTC state $\mathbf{h}_{t-1}$, and preceding flip mask $\mathbf{b}_{t-1}$

Output: encoded BTC payload $\mathscr{B}_t^{\mathrm{BTC}}$ and reconstructed BTC state $\mathbf{h}_t$

Hyperparameters: BTC embedding dimension D

Encoding procedure:

$b_{t,j} \leftarrow \mathbb{I}\left[h_{t,j} \neq h_{t-1,j}\right], \qquad j = 1, \ldots, D$

$\mathcal{C}_{k,t} \leftarrow \left\{j \mid b_{t-1,j} = k\right\}, \qquad k \in \{0, 1\}$

$p_{k,t} \leftarrow \dfrac{1}{|\mathcal{C}_{k,t}|} \displaystyle\sum_{j \in \mathcal{C}_{k,t}} b_{t,j}, \qquad k \in \{0, 1\}$

$\{p_{k,t}\}_{k=0}^{1} \longrightarrow \mathscr{H}_t$

$q_t\left(\mathbf{b}_t \mid \mathbf{b}_{t-1}\right) \leftarrow \displaystyle\prod_{k=0}^{1} \prod_{j \in \mathcal{C}_{k,t}} p_{k,t}^{b_{t,j}} \left(1 - p_{k,t}\right)^{1 - b_{t,j}}$

$\mathscr{B}_t^{\mathrm{BTC}} \leftarrow \mathrm{ArithmeticEncode}\left(\mathbf{b}_t; q_t\right)$

Decoding procedure:

$\mathscr{H}_t \longrightarrow \{p_{k,t}\}_{k=0}^{1}$

$\mathcal{C}_{k,t} \leftarrow \left\{j \mid b_{t-1,j} = k\right\}, \qquad k \in \{0, 1\}$

$\mathbf{b}_t \leftarrow \mathrm{ArithmeticDecode}\left(\mathscr{B}_t^{\mathrm{BTC}}; \{\mathcal{C}_{k,t}, p_{k,t}\}_{k=0}^{1}\right)$

$h_{t,j} \leftarrow \left(1 - 2b_{t,j}\right) h_{t-1,j}, \qquad j = 1, \ldots, D$

return $\mathscr{B}_t^{\mathrm{BTC}}$ and $\mathbf{h}_t$

Algorithm 2: Reconstruction with Canonical Transport Lineages

Input: multi-view observations $\{\mathcal{I}_t\}_{t=0}^{N-1}$

Output: canonical transport lineages $\{\mathcal{L}_m\}_{m=1}^{K}$ and reconstructed representations $\{\hat{\mathcal{G}}_t\}_{t=0}^{N-1}$

Hyperparameters: $W, C, T_{\min}, \tau, \varepsilon$

$m \leftarrow 1$, $c_m \leftarrow 0$, and $\mathcal{G}_m^c \leftarrow \mathrm{OptimizeCanonical}(\mathcal{I}_0)$
$\hat{\mathcal{G}}_0 \leftarrow \mathcal{G}_m^c$; initialize empty transport-regime and residual-coding contexts
for $t = 1$ **to** $N-1$ **do**
  Transport construction and coding:
  if $t = c_m + 1$ **then**
    Independently optimize $\boldsymbol{\theta}_t$ from $(\mathcal{I}_t, \mathcal{G}_m^c)$
    Encode and decode $\boldsymbol{\theta}_t$ to obtain $\hat{\boldsymbol{\theta}}_t$
    $a_m \leftarrow t$, $\mathbf{h}_{a_m} \leftarrow \mathbf{h}_t$; initialize the lineage transport baseline
  else
    $\boldsymbol{\theta}_t \leftarrow \mathrm{Refine}(\hat{\boldsymbol{\theta}}_{t-1}; \mathcal{I}_t, \mathcal{G}_m^c)$
    Entropy-code the transport residual relative to $\hat{\boldsymbol{\theta}}_{t-1}$, using Algorithm 1 for the BTC state
    Decode the payload to obtain $\hat{\boldsymbol{\theta}}_t$
  end if
  Canonical reconstruction:
  $\hat{\mathcal{G}}_t \leftarrow \mathcal{T}_{\hat{\boldsymbol{\theta}}_t}(\mathcal{G}_m^c)$, with anchor-conditioned CaBLE decoding
  Transport-regime update:
  if $t = a_m$ **then**
    $\mathcal{D}_t \leftarrow 0$
  else
    $\mathcal{D}_t \leftarrow \dfrac{1}{D}\sum_{j=1}^{D} \mathbb{I}[h_{t,j} \neq h_{a_m,j}]$
  end if
  if $t \geq a_m + W$ **then**
    Compute $\widehat{\beta}_t$, σ_t, and ρ_t from $\{\mathcal{D}_{t-W}, \ldots, \mathcal{D}_t\}$
  end if
  Canonical expiration and reanchoring:
  if $t - c_m \geq T_{\min}$, $t \geq a_m + W + C$, and $\max_{0 \leq k < C} \rho_{t-k} < \tau$ **then**
    $c_{m+1} \leftarrow t$ and $\mathcal{G}_{m+1}^c \leftarrow \hat{\mathcal{G}}_t$
    $m \leftarrow m + 1$; reset the transport reference, residual history, and regime context
  end if
end for
$K \leftarrow m$
return $\{\mathcal{L}_m\}_{m=1}^{K}$ and $\{\hat{\mathcal{G}}_t\}_{t=0}^{N-1}$

F Implementation Details

F.1 Data Preparation and Evaluation Protocol

Datasets. We evaluate AncesTrA on short- and long-horizon volumetric video benchmarks: (1) N3DV (Li et al., 2022b), six 300-frame scenes at 2704×2028; (2) Meeting Room (Li et al., 2022a), three 300-frame scenes at 1280×720; (3) Bar (Xu et al., 2024), a 3,540-frame 4K sequence captured by 22 synchronized cameras at 60 FPS, with a resolution of 2110×3760; and (4) AvatarRex (Zheng et al., 2023), from which we select two long-horizon sequences, ZXC and LBN2, containing 1,801 and 1,871 frames, respectively. The AvatarRex sequences are captured by 16 synchronized RGB cameras at 30 FPS, with a resolution of 1500×2048. For all datasets, one viewpoint is held out for testing and the remaining are used for training.

Metrics. We report frame-averaged PSNR, SSIM (Wang et al., 2004), LPIPS (Zhang et al., 2018), encoded storage, training time, and rendering speed. Storage includes the full canonical states at

canonical frames and entropy-coded transport residuals with reanchoring overhead at predictive frames (Tang et al., 2025b; Gao et al., 2024; Hu et al., 2025; Fu et al., 2025; Zhang et al., 2026).

F.2 Representation and Hyperparameters

The coding identities of I-frames and P-frames are independent of the role of a canonical reference (Lu et al., 2024; Tang et al., 2025b). In our implementation, the sequence begins with a single I-frame, which also initializes the canonical anchor reference of the first lineage. Subsequent canonical references are established by reanchoring reconstructed P-frames; these frames retain their P-frame coding identity while assuming the reference role for the following lineage. Accordingly, the canonical anchor reference is lineage-scoped and renewed only at reanchoring boundaries, whereas the Gaussian decoder is sequence-scoped: it is instantiated at the initial I-frame and remains shared and fixed thereafter (Kerbl et al., 2023; Lu et al., 2024).

AncesTrA maintains a canonical anchor reference $\mathcal{G}_m^c$ within each canonical lineage and represents each predictive frame through a compact transport state $\boldsymbol{\theta}_t$. The transport operator $\mathcal{T}_{\boldsymbol{\theta}_t}$ is instantiated using the Binary Transformation Cache (BTC) (Tang et al., 2025b), which predicts anchor-position and feature transformations relative to the canonical anchor reference. CaBLE extends Gaussian generation with a feature-dependent splitting direction and bilateral opacity modulation.

F.3 Stage-Specific Optimization

AncesTrA uses different optimization schedules for I-frame construction and predictive transport adaptation. The I-frame representation is optimized from scratch for 8K iterations. For the first predictive frame following the canonical reference of each lineage, a new transport state is independently initialized and optimized for 5K iterations. For every subsequent P-frame, optimization is initialized from the *decoded* transport state of the preceding frame, $\hat{\boldsymbol{\theta}}_{t-1}$, and refined against the current multi-view observations for 300 iterations. We train models using Adam with a learning rate of 5×10^{-3}. All experiments are conducted with a fixed random seed of 0.

F.4 State Transition at P-frames and Reanchoring

TraCER evaluates the transport-regime statistic over the sliding window defined in Sec. 3.3, using $W = 8$ and $\tau = 2$. To prevent transient fluctuations from prematurely terminating a lineage, canonical expiration requires sufficient transport history, a minimum lineage duration of $T_{\min} = 10$ frames, and persistence of the stabilized regime over $C = 3$ consecutive estimates. At the expiration boundary, the decoded anchor state is promoted to the canonical reference of the next lineage. The first predictive frame following reanchoring independently initializes a new transport state with respect to this renewed reference.

F.5 Training-Time Measurement

For the Bar sequence, we measure per-frame training time under the optimization settings corresponding to each method's reported reconstruction results. AncesTrA uses 8K iterations for the I-frame, 5K iterations for the first P-frame of each canonical lineage, and 300 refinement iterations for subsequent P-frames. The optimization schedules of reproduced baselines are configured based on their original publications and official implementations.

Under this experimental setting, AncesTrA achieves approximately 5$\times$ faster per-frame training than iFVC on the Bar sequence while delivering a substantial 4.68 dB improvement in reconstruction quality, as reported in Tab. 2. This substantial acceleration significantly alleviates the computational bottleneck of frame-wise optimization, reducing the processing overhead associated with continuously incoming frames and bringing high-fidelity volumetric video reconstruction with real-time rendering closer to practical deployment.

G Additional Long-Horizon Results

The main paper reports long-horizon reconstruction performance on Bar in Tab. 2, temporal stability in Fig. 3, and component ablations in Tab. 3. To further evaluate the generalizability of AncesTrA,

we extend the long-horizon experiments to the AvatarRex ZXC and LBN2 sequences (Zheng et al., 2023) under the protocol in Sec. 4.1. These sequences provide additional long-horizon settings for evaluating reconstruction fidelity, storage efficiency, and temporal stability beyond the Bar sequence.

G.1 Additional Results on ZXC and LBN2

Tabs. 5 and 6 present the quantitative results on ZXC and LBN2, respectively. AncesTrA achieves the highest PSNR on both sequences while requiring the lowest storage among the evaluated methods. Compared with iFVC (Tang et al., 2025b), AncesTrA improves PSNR by approximately 4 dB on ZXC and 2 dB on LBN2, while reducing storage requirements by approximately 4× and 7×, respectively. These results demonstrate that AncesTrA achieves improved reconstruction fidelity while retaining a compact representation. This joint improvement benefits from the complementary designs of CiTER and TraCER. Within each lineage, CiTER separates scene reconstruction from temporal inheritance by reconstructing predictive frames from a shared canonical representation while encoding successive transport changes as compact, entropy-coded residuals. This organization exploits adjacent-frame redundancy without recursively propagating reconstructed scene states. TraCER further renews the canonical reference according to the observed transport regime, allowing subsequent transformations to operate relative to an updated origin rather than indefinitely relying on the same reference. Together, these mechanisms support compact temporal inheritance and sustained reconstruction quality over extended sequences.

The improvement is also reflected in structural and perceptual reconstruction quality. AncesTrA achieves SSIM values of approximately 0.950 and 0.941 on ZXC and LBN2, respectively. On ZXC, it additionally obtains the lowest LPIPS. On LBN2, AncesTrA matches the lowest LPIPS among the compared methods while requiring substantially less storage. These results show that the compact temporal representation also preserves structural and perceptual quality. This performance is consistent with the complementary role of CaBLE, which enhances local Gaussian expressivity around geometric and appearance discontinuities. By selectively lifting symmetric Gaussian bases into asymmetric counterparts, CaBLE provides additional flexibility for representing local structures and appearance variations, complementing the temporal organization established by CiTER and TraCER.

Together with the Bar results, the additional experiments demonstrate that the quality and compactness advantages of canonical transport lineages extend across multiple long-horizon scenes, rather than being restricted to a particular motion sequence.

Table 5: Quantitative comparison on the long-horizon ZXC sequence.

Method	PSNR↑	SSIM↑	LPIPS↓	Storage (KB)↓
3DGStream (CVPR'24)	23.17	0.895	0.104	3891.20
4DGC (CVPR'25)	20.66	0.837	0.133	1433.60
HiCoM (NeurIPS'24)	20.67	0.756	0.132	174.08
Recon-GS (NeurIPS'25)	20.65	0.791	0.126	102.40
iFVC (AAAI'25)	24.06	0.928	0.091	71.68
AncesTrA	28.00	0.950	0.078	19.26

Table 6: Quantitative comparison on the long-horizon LBN2 sequence.

Method	PSNR↑	SSIM↑	LPIPS↓	Storage (KB)↓
3DGStream (CVPR'24)	24.97	0.906	0.090	3891.20
4DGC (CVPR'25)	21.48	0.858	0.120	1443.84
HiCoM (NeurIPS'24)	23.49	0.853	0.098	122.88
Recon-GS (NeurIPS'25)	22.51	0.828	0.135	102.40
iFVC (AAAI'25)	26.06	0.930	0.080	92.16
AncesTrA	27.98	0.941	0.080	12.65

G.2 Long-Horizon Stability on ZXC

Beyond sequence-averaged reconstruction quality, we evaluate temporal stability through the frame-wise PSNR trajectory on the 1,801-frame ZXC sequence. As shown in Fig. 5, AncesTrA generally

maintains higher reconstruction quality than the competing methods across the sequence, and its advantage remains visible during the later stages. These observations show that the sequence-level improvement is supported by sustained reconstruction fidelity rather than being concentrated in the initial frames. This sustained performance benefits from the temporal organization established by CiTER. Conventional recursive scene-state propagation makes each reconstructed scene a reference for subsequent frames, allowing intermediate reconstruction or quantization errors to enter the recursive reference chain. CiTER instead reconstructs every predictive frame within a lineage directly from the shared canonical representation, while temporal continuity is preserved through inherited transport states. In this way, temporal information can still be reused across adjacent frames without requiring intermediate reconstructed scene states to become recursive references. TraCER further complements this mechanism by regulating the validity horizon of the canonical reference. As scene evolution extends, TraCER renews the canonical origin according to the observed transport regime, allowing subsequent transport evolution to proceed from an updated reference. The ZXC results therefore support the benefit of combining compact transport inheritance within lineages with adaptive canonical renewal, maintaining reconstruction fidelity and low storage throughout long-horizon scene evolution.

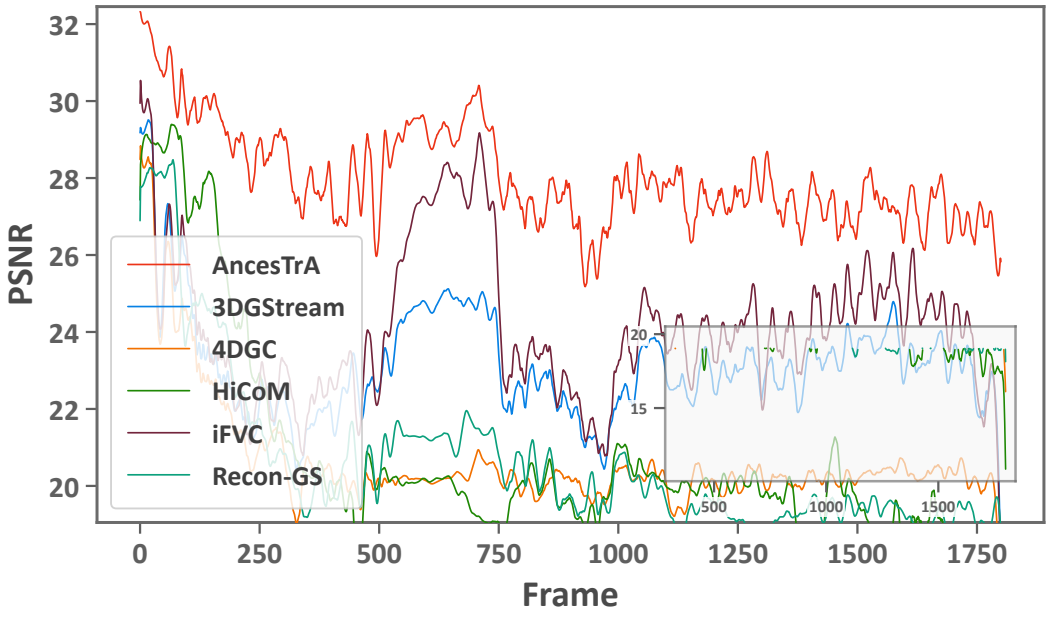

Figure 5: **Long-horizon stability on ZXC.** Frame-wise PSNR comparison over 1,801 frames. AncesTrA maintains the strongest trajectory among the plotted methods. The inset shows the portions of the PSNR trajectories below the lower limit of the main plot's y-axis.

G.3 Long-Horizon Stability on LBN2

Fig. 6 further evaluates temporal stability on the 1,871-frame LBN2 sequence. AncesTrA maintains higher reconstruction quality over the sequence. Moreover, AncesTrA achieves approximately 7× lower storage than iFVC while improving sequence-average PSNR by approximately 2 dB. These results show that the quality–storage advantage can be maintained as scene evolution extends over a long temporal horizon. This performance is consistent with the complementary roles of CiTER and TraCER. CiTER preserves efficient temporal inheritance through compact transport updates within each lineage, while TraCER regulates how long the current canonical representation remains an effective reference. When the transport regime satisfies the renewal criterion, a decoded scene state is promoted as the canonical reference of the next lineage, after which CiTER continues transport inheritance from the renewed origin. This division of roles allows AncesTrA to preserve adjacent-frame temporal coherence without indefinitely relying on a single canonical reference. Together with the ZXC results, these observations demonstrate that the proposed lineage-based representation preserves high reconstruction fidelity and substantial storage efficiency across different long-horizon motion sequences.

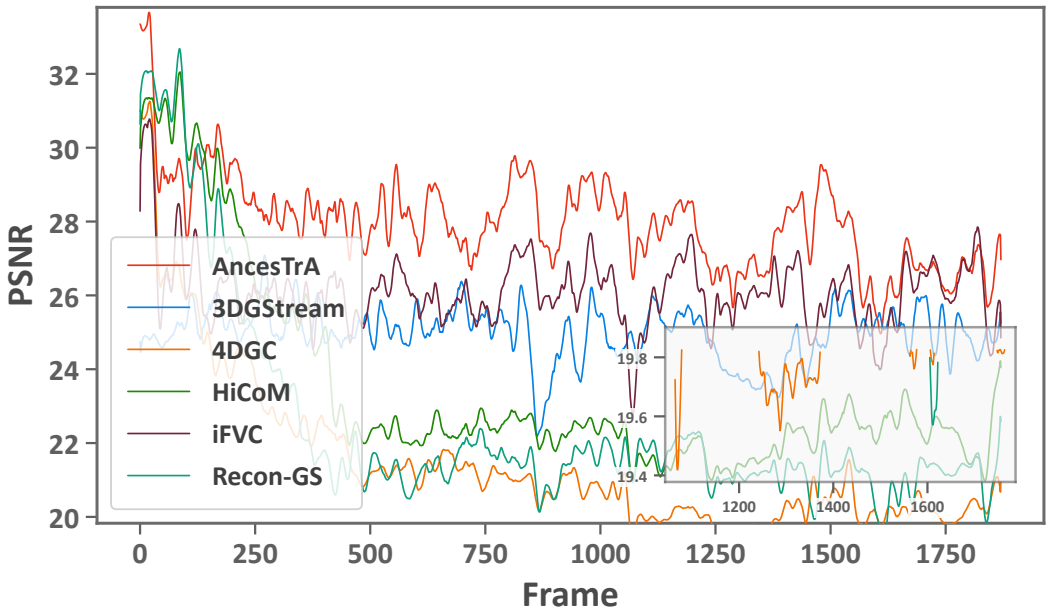

Figure 6: **Long-horizon stability on LBN2.** Frame-wise PSNR comparison over 1,871 frames. AncesTrA maintains the strongest trajectory across the sequence. The inset shows the portions of the PSNR trajectories below the lower limit of the main plot's y-axis.

Table 7: Module configurations corresponding to the main-paper ablation variants. A check mark denotes an enabled module, while a cross denotes a disabled module.

Configuration	CiTER	TraCER	CaBLE
1: w/o All	✗	✗	✗
2: w/o CiTER & TraCER	✗	✗	✓
3: w/o TraCER & CaBLE	✓	✗	✗
4: w/o CaBLE	✓	✓	✗
5: w/o TraCER	✓	✗	✓
Full: **AncesTrA**	✓	✓	✓

H Supplementary Ablation Analysis

We further investigate the contributions of CiTER, TraCER, and CaBLE on the 3,540-frame Bar sequence, whose sustained scene evolution provides a challenging setting for evaluating both temporal efficiency and long-horizon reconstruction stability. Tab. 3 summarizes the sequence-averaged reconstruction quality, storage, and computational performance of the ablation variants, while Fig. 7 presents their frame-wise PSNR trajectories. Together, these results allow us to examine how the three components contribute to compact temporal inheritance, adaptive canonical renewal, and local representation expressivity.

H.1 Variant Definitions

The configurations of the ablation variants are summarized in Tab. 7. *w/o All* removes all three components. *w/o CiTER & TraCER* retains only CaBLE, whereas *w/o TraCER & CaBLE* retains only CiTER. *w/o CaBLE* enables CiTER and TraCER, while *w/o TraCER* enables CiTER and CaBLE. The full **AncesTrA** enables all three components. These configurations allow us to examine the efficiency gains introduced by CiTER, assess the contribution of adaptive canonical renewal, and evaluate CaBLE under different temporal organizations.

H.2 Quantitative Analysis of Component Contributions

As shown in Tab. 3, enabling CiTER alone reduces storage from 125.42 to 16.53 KB and shortens per-frame training time from 1.640 to 0.107 minutes, achieving approximately 8× more compact

storage and 15× faster optimization. These results demonstrate that CiTER substantially reduces frame-wise optimization costs by inheriting and refining compact transport states. This efficiency improvement benefits from separating scene reconstruction from temporal inheritance. Within each lineage, predictive frames are reconstructed directly from a shared canonical representation, while subsequent transport states inherit their decoded predecessors as initialization and are transmitted through entropy-coded residuals. This organization concentrates temporal adaptation into a compact transport trajectory without recursively propagating reconstructed scene states. The benefits of this temporal organization extend to the complete framework, where CiTER is combined with adaptive canonical renewal. Compared with *w/o CiTER & TraCER*, AncesTrA improves PSNR from 32.23 to 32.71 dB while reducing storage from 110.70 to 55.47 KB and accelerating training by approximately 12×. By coordinating compact transport inheritance with canonical renewal, the complete framework maintains efficient temporal adaptation and high reconstruction fidelity throughout extended scene evolution. TraCER further contributes to reconstruction fidelity by maintaining the relevance of the canonical reference over long temporal horizons. Removing TraCER from the complete model reduces PSNR from 32.71 to 31.50 dB, corresponding to a 1.21 dB degradation. This result demonstrates the importance of adaptive canonical renewal even when compact transport inheritance and enhanced local expressivity are retained. Although CiTER avoids recursive scene-state propagation within each lineage, a persistent canonical reference can become progressively less representative as scene content evolves. TraCER addresses this limitation by characterizing transport dynamics relative to the current lineage and renewing the canonical reference when the transport regime satisfies its persistent stabilization criterion. Subsequent transport evolution can therefore proceed from an updated canonical origin, allowing CiTER to preserve compact temporal inheritance without indefinitely relying on an increasingly distant scene reference. CaBLE consistently improves reconstruction fidelity under different temporal organizations. Without CiTER and TraCER, introducing CaBLE increases PSNR from 31.69 to 32.23 dB, yielding a 0.54 dB improvement. After CiTER and TraCER are enabled, adding CaBLE further raises PSNR from 32.26 to 32.71 dB, providing an additional 0.45 dB gain. These consistent improvements support the contribution of asymmetric basis lifting to local representation beyond temporal organization itself. Conventional symmetric Gaussian bases have limited capacity to characterize local discontinuities and abrupt appearance transitions, even when temporal evolution is efficiently modeled. CaBLE addresses this limitation by selectively lifting symmetric Gaussian bases into asymmetric counterparts through an anchor-feature-dependent splitting direction and bilateral opacity modulation. This additional flexibility enriches local reconstruction expressivity while complementing the compact temporal inheritance and adaptive canonical renewal established by CiTER and TraCER.

H.3 Temporal Behavior of Ablation Variants

The frame-wise PSNR trajectories in Fig. 7 provide further insight into the sequence-level results. AncesTrA maintains the highest reconstruction quality among the ablation variants over most of the Bar sequence, whereas the variants without TraCER exhibit pronounced performance degradation during extended portions of the middle and late stages. In particular, removing TraCER while retaining CiTER and CaBLE produces a pronounced quality gap relative to the complete model, further supporting the contribution of adaptive canonical renewal to long-horizon reconstruction stability. This difference is consistent with the limitation of maintaining a fixed canonical origin as scene evolution extends. Although CiTER preserves temporal continuity through inherited transport states, subsequent transformations must accommodate increasingly distant scene variations when the canonical reference remains unchanged. TraCER complements this inheritance by renewing the canonical origin according to the observed transport regime, allowing subsequent transport evolution to proceed from an updated reference rather than indefinitely relying on the same canonical state. The remaining trajectories further illustrate the contribution of local representation expressivity. Compared with the complete model, *w/o CaBLE* exhibits a generally lower PSNR trajectory despite retaining both CiTER and TraCER. This persistent quality gap is consistent with the quantitative results, indicating that efficient temporal inheritance and adaptive canonical renewal do not eliminate the need for expressive local Gaussian bases. CaBLE complements the temporal organization by selectively enhancing the representation of local geometric and appearance discontinuities. Together, the quantitative and frame-wise results support the complementary roles of the three components: CiTER establishes compact and efficient temporal inheritance, TraCER maintains the relevance of the canonical reference over extended sequences, and CaBLE strengthens local reconstruction expressivity. Their combination enables AncesTrA to sustain high reconstruction fidelity

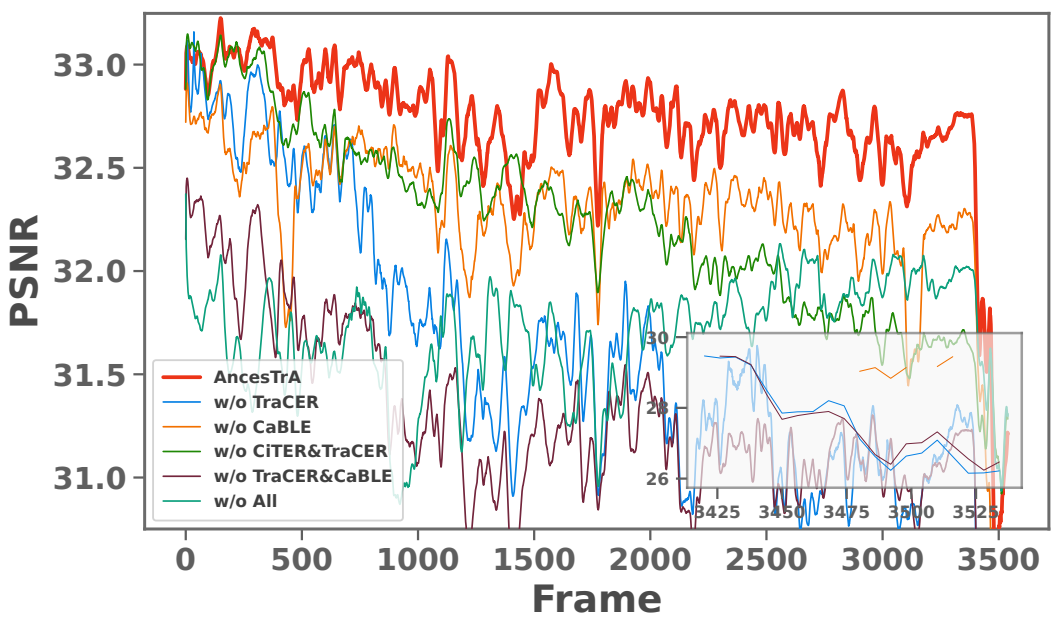

Figure 7: **Frame-wise ablation results on Bar.** AncesTrA sustains the highest reconstruction quality over most of the sequence, while variants without TraCER exhibit pronounced performance degradation during extended portions of the middle and late stages. The inset shows the portions of the PSNR trajectories below the lower limit of the main plot's y-axis.

while retaining compact storage and efficient optimization throughout long-horizon scene evolution.

I Hyperparameter Sensitivity

We further investigate the sensitivity of AncesTrA to three key hyperparameters on the 3,540-frame Bar sequence: the refinement budget for subsequent P-frames, the TraCER stabilization threshold τ, and the temporal window size W. Following a controlled one-factor-at-a-time protocol, each experiment varies a single hyperparameter while keeping all remaining settings fixed. These parameters govern different aspects of the framework: the refinement budget determines the optimization cost of temporal adaptation, τ controls the admission criterion for canonical renewal, and W determines the amount of temporal evidence used to identify stabilized transport regimes.

Table 8: Sensitivity to the CiTER refinement budget on Bar.

Iterations	PSNR↑	SSIM↑	LPIPS↓	Storage (KB)↓
50	31.94	0.904	0.144	50.01
100	32.35	0.913	0.138	60.54
300	32.71	0.913	0.134	55.47
600	32.64	0.902	0.133	50.74
1000	32.66	0.910	0.132	53.68

Refinement budget. As shown in Tab. 8, increasing the refinement budget from 50 to 300 iterations progressively improves PSNR from 31.94 to 32.71 dB. With a small refinement budget, the inherited transport state provides an effective initialization but still requires further adaptation to match the current observations. Increasing the number of iterations therefore allows the residual discrepancy between the inherited transport and the current frame to be more thoroughly resolved, leading to consistent reconstruction gains. Once the budget reaches 300 iterations, however, the inherited transport has already been sufficiently aligned with the current observations. Further increasing the budget to 600 and 1000 iterations slightly reduces PSNR to 32.64 and 32.66 dB, respectively, suggesting that excessive frame-wise refinement may begin to overfit the training observations and no longer improve generalization to held-out views. This behavior is consistent with CiTER, where subsequent P-frames inherit the decoded transport state of the preceding frame rather than starting

optimization from scratch. We therefore adopt 300 iterations, which achieves the highest PSNR among the evaluated settings while avoiding unnecessary optimization and potential overfitting from larger refinement budgets.

Table 9: Sensitivity to the TraCER stabilization threshold on Bar, with the no-renewal comparison.

Configuration	Value	PSNR↑	SSIM↑	LPIPS↓	Storage (KB)↓
TraCER	0.5	32.35	0.912	0.136	30.96
	1.0	32.39	0.912	0.139	23.88
	1.5	32.60	0.912	0.136	41.75
	2.0	32.71	0.913	0.134	55.47
	4.0	32.63	0.906	0.134	54.29
w/o TraCER	–	31.50	0.901	0.149	16.04

Stabilization threshold τ. Tab. 9 shows that increasing τ from 0.5 to 2 progressively improves PSNR from 32.35 to 32.71 dB, whereas further increasing it to 4 reduces PSNR to 32.63 dB. This trend follows directly from the role of τ in TraCER. Since canonical expiration requires the regime statistic to satisfy $\rho_t < \tau$ persistently, a small τ imposes a strict stabilization criterion. Under such settings, TraCER may continue using the current canonical reference until the transport trajectory becomes strongly stabilized, even when that reference has already become increasingly distant from the evolving scene. Increasing τ relaxes this requirement and allows moderate local transport fluctuations to be tolerated, enabling suitable stabilization regimes to be recognized more effectively. However, when τ becomes too large, the criterion becomes overly permissive and may classify transport regimes as stabilized before their evolution has sufficiently settled, reducing the effectiveness of the resulting canonical renewal. The comparison with the no-renewal setting further highlights the value of adaptive canonical renewal: removing TraCER reduces storage to 16.04 KB, but incurs a substantial PSNR drop from 32.71 to 31.50 dB. This result demonstrates that the additional representation cost introduced by selective renewal is effectively translated into improved long-horizon reconstruction fidelity. We therefore adopt $\tau = 2$, which achieves the highest PSNR among the evaluated thresholds by providing an appropriate balance between overly delayed and overly permissive canonical renewal.

Table 10: Sensitivity to the TraCER window size on Bar.

Window W	PSNR↑	SSIM↑	LPIPS↓	Storage (KB)↓
4	32.29	0.912	0.138	29.94
6	32.42	0.906	0.136	45.12
8	32.71	0.913	0.134	55.47
12	32.59	0.912	0.134	50.04
16	32.54	0.914	0.135	45.26

Window size W. As reported in Tab. 10, increasing W from 4 to 8 progressively improves PSNR from 32.29 to 32.71 dB, whereas further increasing it to 12 and 16 reduces PSNR to 32.59 and 32.54 dB, respectively. This trend reflects the importance of selecting an appropriate amount of temporal information for transport-regime characterization. When the window is too short, the available temporal evidence is limited, making the estimated transport trend and fluctuation more sensitive to incidental frame-level variations. Increasing W therefore provides richer temporal context and allows TraCER to distinguish persistent stabilization from transient fluctuations more reliably. However, an excessively long window introduces more historical observations that may be less representative of the current transport regime. Such redundant historical information can weaken the influence of recent transport changes on regime estimation, reducing responsiveness to ongoing scene evolution and consequently degrading reconstruction fidelity. We therefore select $W = 8$, which achieves the highest PSNR among the evaluated window sizes by providing sufficient temporal evidence without introducing excessive historical information.

J Threshold-Controlled Rate–Distortion Performance

We further evaluate the rate–distortion behavior of AncesTrA by varying the TraCER stabilization threshold τ while keeping all remaining settings fixed. Changing τ modifies how permissively stabilized transport regimes are admitted for canonical renewal, thereby altering the resulting balance

between reconstruction fidelity and representation cost without changing the underlying reconstruction framework. Each configuration is reconstructed and encoded independently under the same evaluation protocol.

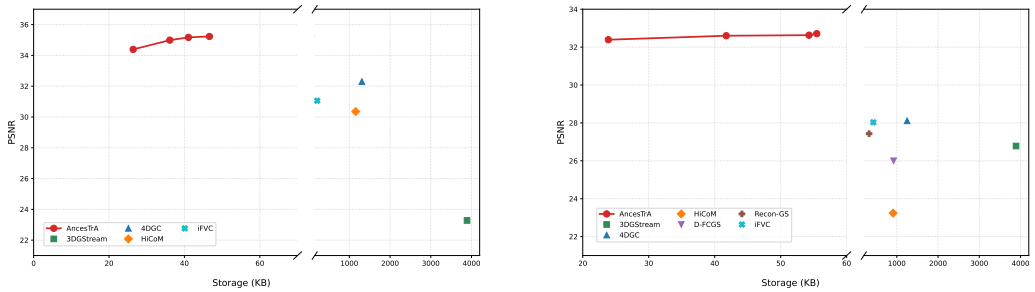

Figure 8: Threshold-controlled rate–distortion performance on *Trimming* (left) and *Bar* (right). AncesTrA operating points are obtained by varying the TraCER threshold τ.

As shown in Fig. 8, AncesTrA consistently occupies the upper-left region of the rate–distortion space on both the short-sequence *Trimming* scene and the long-sequence *Bar* scene. Across the evaluated threshold settings, its operating points remain below 60 KB while maintaining substantially higher PSNR than the representative streamable reconstruction methods. This advantage is particularly evident on *Bar*, where competing methods require storage on the order of hundreds to thousands of KB while still exhibiting considerably lower reconstruction quality. The same trend is observed on *Trimming*, demonstrating that the favorable quality–storage behavior of AncesTrA is preserved across substantially different temporal horizons.

The progression of the AncesTrA operating points further illustrates how TraCER provides a controllable quality–storage trade-off. A stricter stabilization threshold admits canonical renewal more conservatively, reducing the associated renewal overhead but potentially retaining an increasingly distant canonical reference for longer. Relaxing the threshold allows more transport regimes to qualify for renewal, which can improve reconstruction fidelity by maintaining a more representative canonical origin, at the cost of additional canonical storage. The resulting operating points therefore reflect the balance between preserving compact transport inheritance and allocating additional representation capacity to canonical renewal. Importantly, varying τ changes this balance without requiring a different reconstruction model or coding pipeline.

The consistently favorable frontier also reflects the complementary interaction between TraCER and CiTER. Canonical renewal prevents later scene evolution from being modeled indefinitely relative to an increasingly outdated reference, while CiTER continues to encode temporal evolution through compact inherited transport states within each lineage. Consequently, the additional representation cost introduced by renewal is used to maintain an effective canonical origin rather than repeatedly rebuilding complete frame-wise scene representations. The fact that this behavior persists on both *Trimming* and the substantially longer *Bar* sequence demonstrates that AncesTrA retains strong rate–distortion efficiency as the temporal horizon expands.

K Additional Qualitative Results

We provide additional qualitative comparisons on four representative scenes: Bar, Cut Roasted Beef, Discussion, and Flame Steak. These scenes present different reconstruction challenges, including prolonged scene evolution, fine-grained hand–object interactions, articulated human motion, and rapid local appearance changes. The comparisons complement the quantitative results by examining the preservation of geometric structures, local appearance details, and reconstruction consistency under different temporal conditions.

K.1 Bar

Long-horizon appearance preservation. The 3,540-frame Bar sequence provides a challenging setting for evaluating reconstruction quality over extended temporal horizons. As shown in Fig. 9, the differences between the methods are particularly visible around the subject's face, hands, and

musical instrument. The competing methods exhibit different forms of reconstruction degradation: some produce softened facial features and blurred hand contours, while others introduce more pronounced color artifacts and distortions in the surrounding scene. These differences remain apparent in the later sampled frames, where prolonged motion and changes in the subject's position place additional demands on temporal reconstruction.

In contrast, AncesTrA preserves clearer facial details, more distinct hand boundaries, and more coherent instrument geometry across the sampled stages. Even toward the end of the sequence, when the subject moves partially out of view, the reconstructed scene remains visually consistent with the ground truth. This behavior is consistent with the complementary roles of CiTER and TraCER. Within each lineage, CiTER reconstructs predictive frames from a shared canonical representation while inheriting compact transport states, avoiding recursive propagation of reconstructed scene states. TraCER further renews the canonical origin according to the observed transport regime, preventing subsequent reconstruction from relying indefinitely on an increasingly distant reference. Together with the local expressivity introduced by CaBLE, this temporal organization helps preserve detailed geometry and appearance throughout prolonged scene evolution.

K.2 Cut Roasted Beef

Fine-grained geometric reconstruction. Fig. 10 presents qualitative comparisons on the Cut Roasted Beef scene, where the subject's head, arms, and interactions with objects on the countertop provide representative regions for evaluating local reconstruction quality. The differences are particularly evident in the highlighted facial and hand–object regions. 3DGStream exhibits substantial blurring and incomplete reconstruction of the moving subject, while other competing methods show varying degrees of softened facial contours and loss of fine details around the hands and nearby objects. In comparison, AncesTrA produces more clearly defined facial structures, coherent arm contours, and detailed hand–object regions across the selected frames.

These improvements are consistent with the combination of efficient temporal adaptation and enhanced local representation expressivity. CiTER adapts inherited transport states to the evolving scene while maintaining reconstruction relative to the active canonical representation. CaBLE further increases the flexibility of the local Gaussian representation by selectively introducing asymmetric bases with feature-dependent splitting directions and bilateral opacity modulation. This additional expressivity helps accommodate local geometric boundaries and abrupt appearance transitions that are difficult to represent using symmetric Gaussian bases alone. The resulting comparisons illustrate how AncesTrA preserves fine-grained geometric and appearance details during dynamic hand–object interactions.

K.3 Discussion

Stable reconstruction under articulated motion. The Discussion sequence contains multiple moving subjects and frequent changes in body posture and relative spatial configuration. As shown in Fig. 11, these dynamics introduce visible differences in the reconstruction of facial features, subject silhouettes, and objects on the table. In particular, 3DGStream exhibits pronounced blurring around moving subjects, while some other methods produce softened facial details and less distinct local boundaries. AncesTrA maintains clearer subject contours and facial structures across the selected frames while preserving the overall spatial arrangement of the participants and surrounding objects.

This behavior is consistent with the separation of scene reconstruction and temporal inheritance in CiTER. Rather than recursively transforming reconstructed scene states, AncesTrA reconstructs each predictive frame from its active canonical representation and inherits temporal information through compact transport updates. This organization allows the transport state to adapt to changing human poses without directly using imperfections in the previous reconstructed scene as the next frame's reconstruction reference. CaBLE complements this temporal organization by providing additional flexibility for representing local geometric and appearance variations. Together, these mechanisms support coherent scene reconstruction while preserving the clarity of moving subjects and fine local structures.

K.4 FLAME STEAK

Appearance fidelity under rapid variation. The Flame Steak scene presents a distinct challenge due to the rapid changes in flame shape, brightness, and local appearance during cooking. As illustrated in Fig. 12, these variations make the flame region particularly sensitive to limitations in temporal adaptation and local representation capacity. The competing methods exhibit different reconstruction behaviors, including weakened flame appearance, blurred flame boundaries, and discrepancies in the shape and intensity of the highlighted regions. In comparison, AncesTrA more faithfully reproduces the changing flame appearance across the selected frames while maintaining clear subject contours and coherent surrounding scene geometry.

The observed results are consistent with the combined contributions of CiTER and CaBLE. CiTER maintains an efficient pathway for adapting the scene representation through inherited transport states, allowing the reconstruction to accommodate rapid changes without repeatedly rebuilding the complete scene. Meanwhile, CaBLE selectively enhances local Gaussian expressivity through asymmetric basis lifting, providing additional flexibility for representing irregular boundaries and abrupt appearance transitions. These complementary capabilities help AncesTrA reconstruct visually complex regions in which geometric structure and appearance change simultaneously, extending its qualitative performance beyond scenes dominated by articulated motion.

Taken together, these comparisons demonstrate the effectiveness of AncesTrA across a diverse range of reconstruction challenges. The long-horizon Bar results illustrate the contribution of compact transport inheritance and adaptive canonical renewal to sustained reconstruction quality, while the remaining scenes highlight the importance of temporal adaptation and expressive local Gaussian bases for dynamic geometry and appearance. By combining CiTER, TraCER, and CaBLE, AncesTrA preserves detailed structures and coherent appearance under prolonged scene evolution, articulated human motion, and rapid local variation.

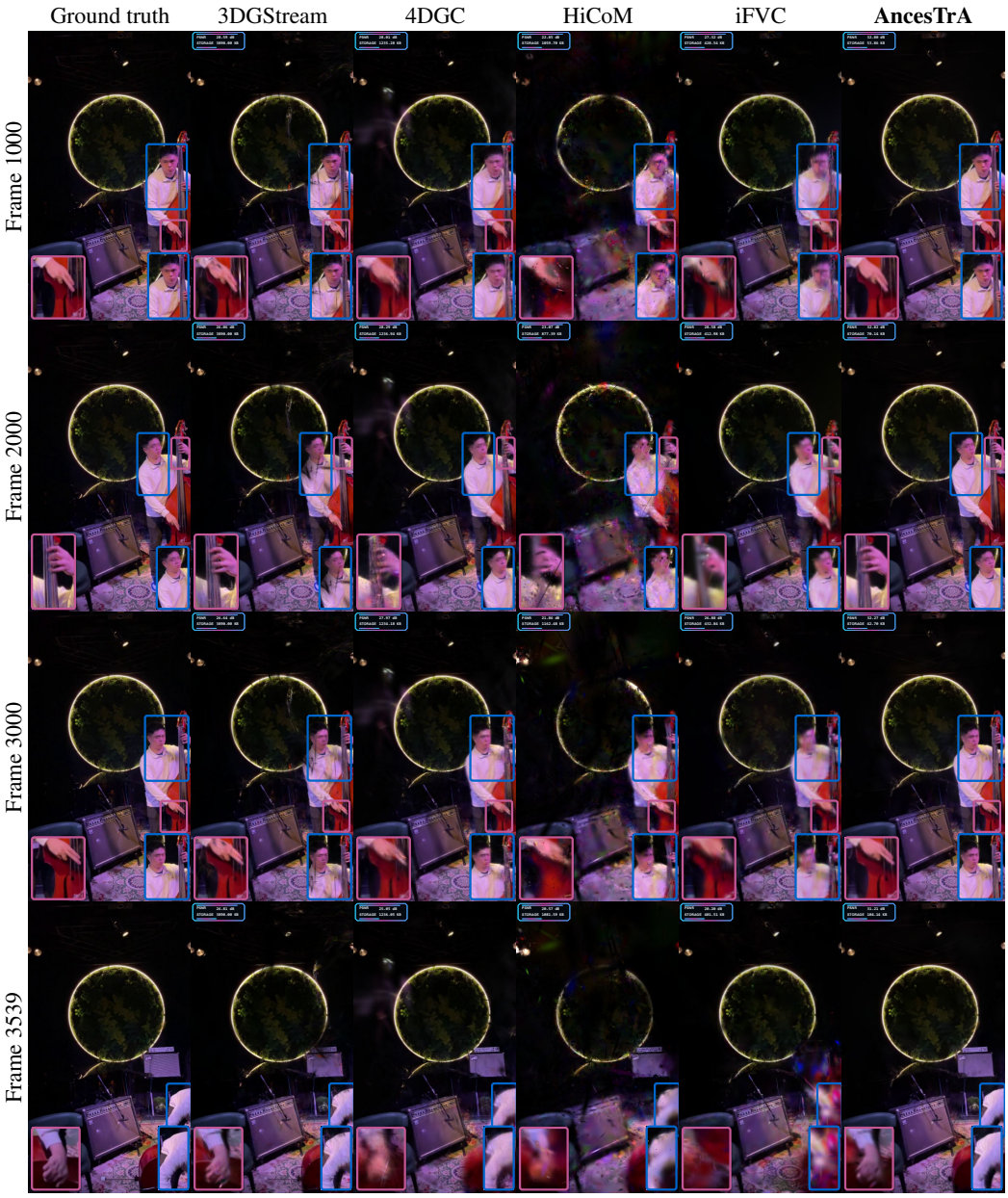

Figure 9: Long-horizon qualitative comparison on the Bar sequence. Selected frames at different temporal positions illustrate reconstruction quality over prolonged scene evolution. From left to right: Ground truth, 3DGStream, 4DGC, HiCoM, iFVC, and AncesTrA. From top to bottom: Frames 1000, 2000, 3000, and 3539.

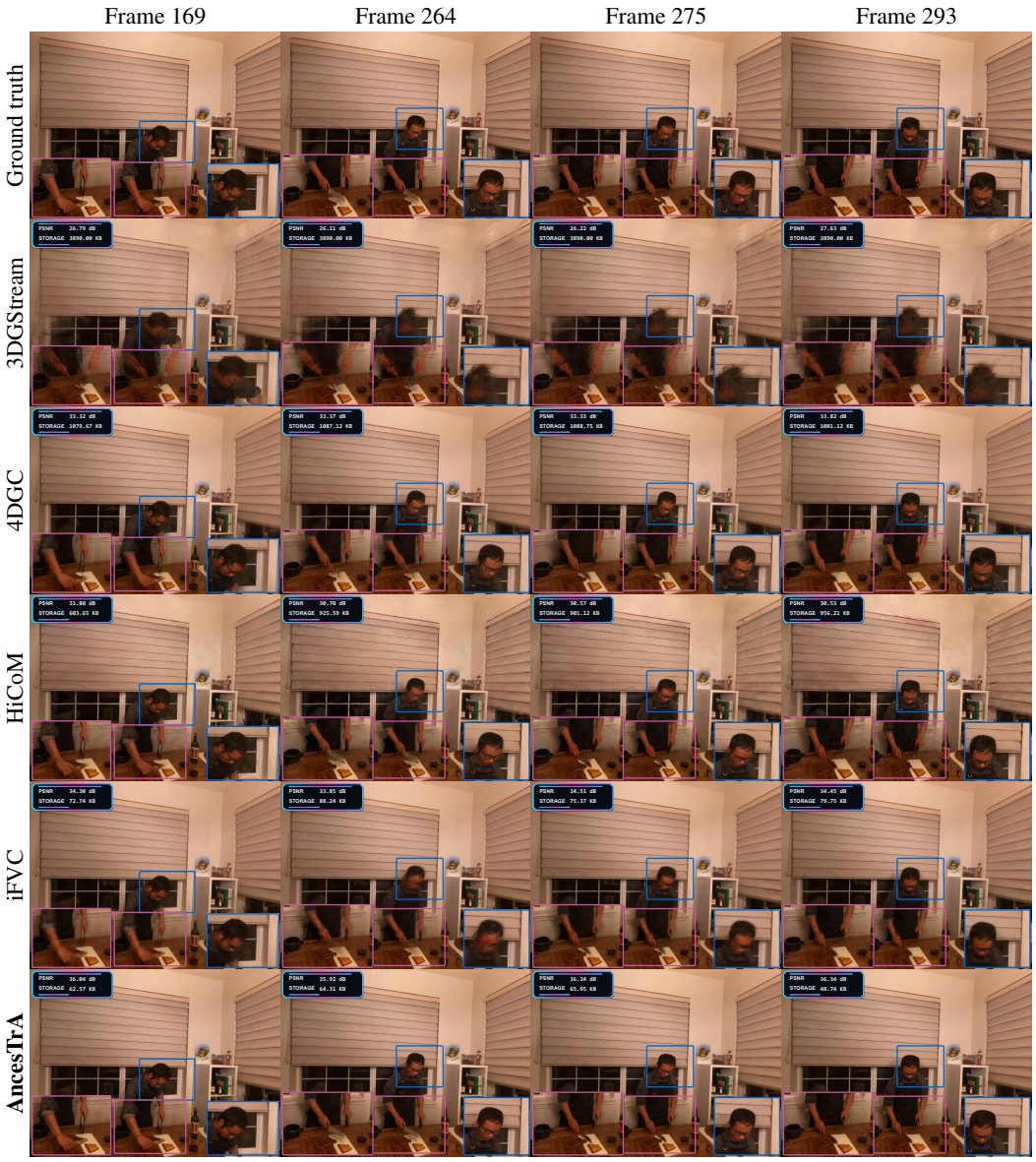

Figure 10: Qualitative comparison on the Cut Roasted Beef scene. From top to bottom: Ground truth, 3DGStream, 4DGC, HiCoM, iFVC, and AncesTrA. From left to right: Frames 169, 264, 275, and 293.

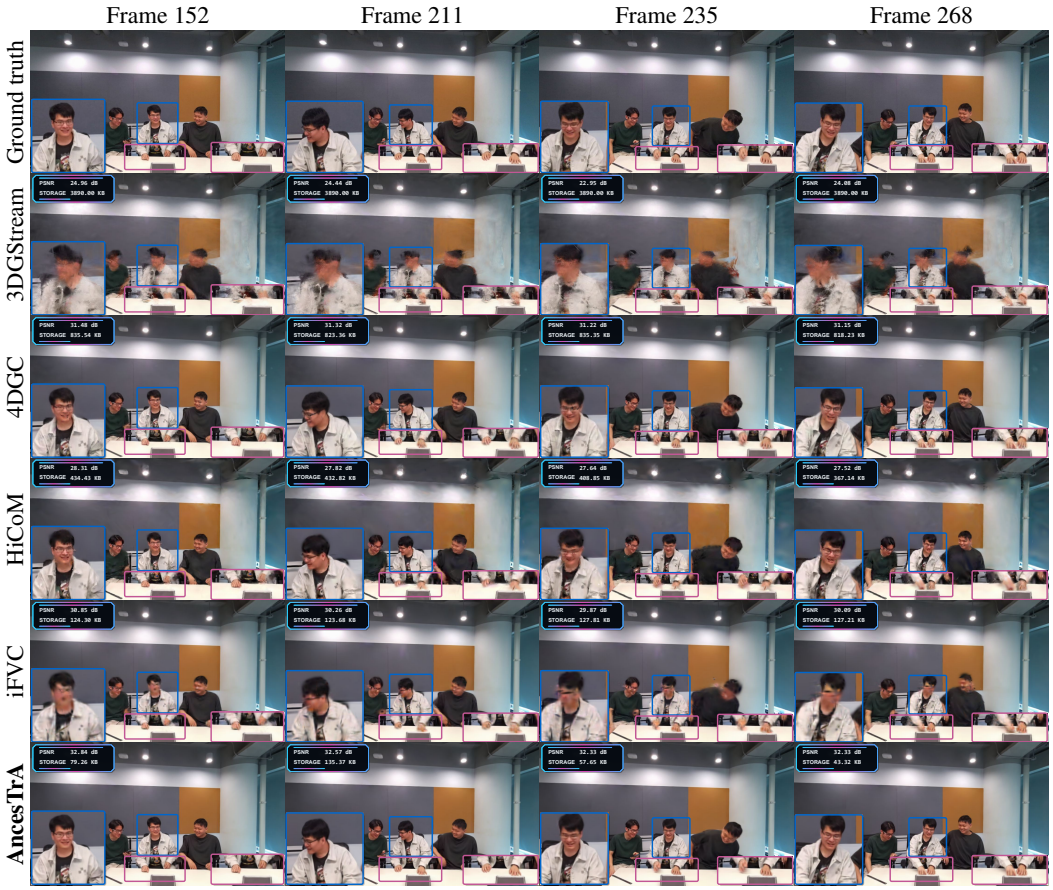

Figure 11: Qualitative comparison on the Discussion scene. From top to bottom: Ground truth, 3DGStream, 4DGC, HiCoM, iFVC, and AncesTrA. From left to right: Frames 152, 211, 235, and 268.

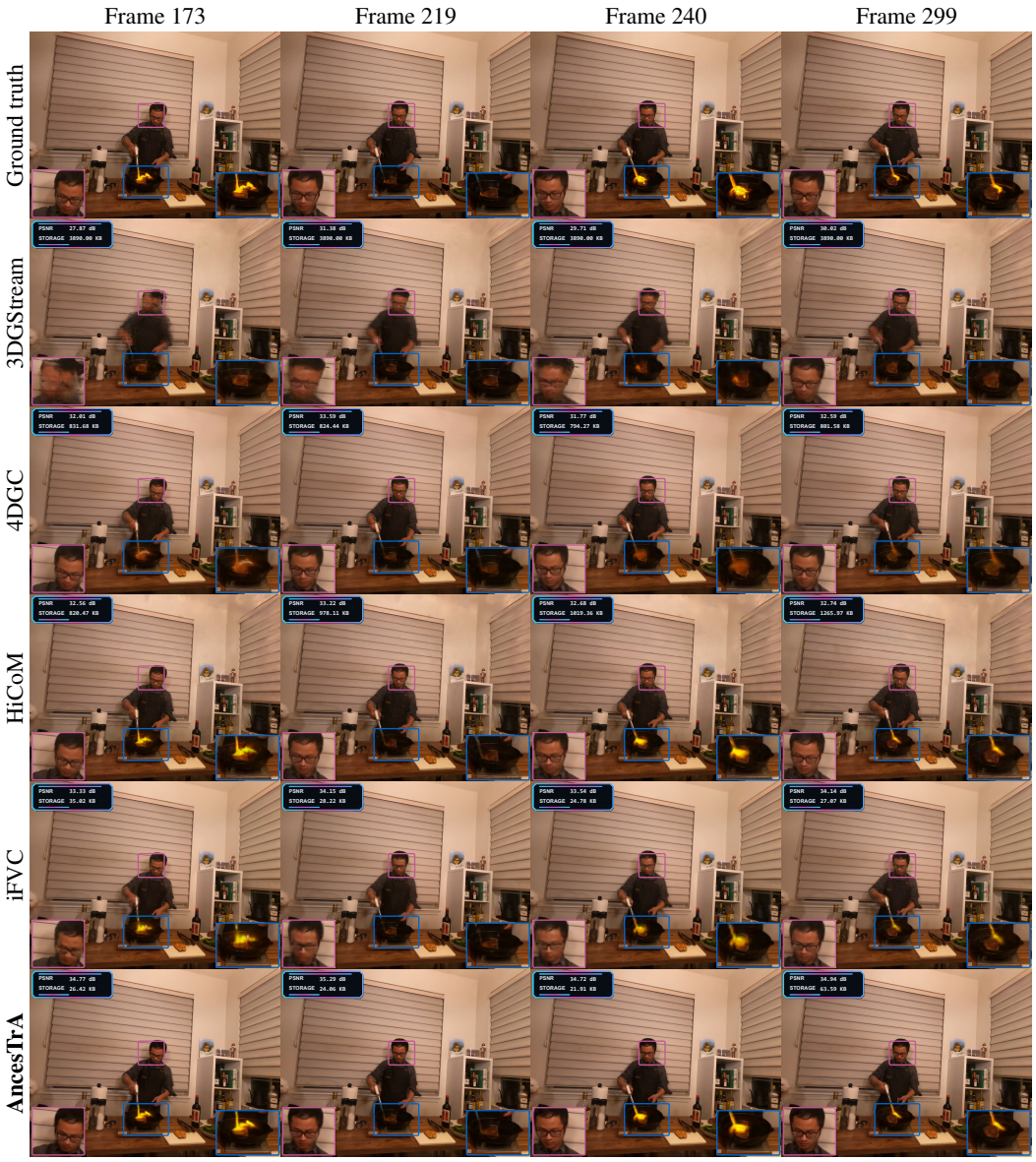

Figure 12: Qualitative comparison on the Flame Steak scene. From top to bottom: Ground truth, 3DGStream, 4DGC, HiCoM, iFVC, and AncesTrA. From left to right: Frames 173, 219, 240, and 299.