IMPACTFLOW: GUIDING FLOW MATCHING WITH GEOMETRY FROM FUTURE OUTCOMES

Anonymous authors
Paper under double-blind review

ABSTRACT

Generative models are trained through local prediction signals, yet their quality is judged on completed outputs. Under limited capacity and computation, how should learning prioritize local errors according to their future consequences? In flow matching, equal-norm velocity errors receive equal local cost even when the remaining dynamics carry them into different output changes. We introduce **terminal-transported residual geometry**, which relates error directions at the current state to their predicted effects after the remaining generation process. Output structure specifies which changes matter, and the remaining dynamics determine how local directions contribute to those changes. **IMPACTFLOW** instantiates this view as directional residual weighting for flow matching, preserving the original conditional velocity labels and the deployment interface. Experiments on low-step HumanML3D text-to-motion generation show that transported geometry predicts independently measured output changes better than decoder-only geometry. Controlled training comparisons demonstrate improved generation quality over endpoint geometry after matching scalar training strength, supporting directional future impact as an effective learning signal. These results support a learning principle: local errors should be prioritized by their future impact on the generated outcome.

1 INTRODUCTION

Modern generative models acquire complex generation capabilities through local learning signals. They learn denoising, scores, velocities, or other prediction targets at intermediate states (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021; Lipman et al., 2023), while quality is ultimately assessed on the completed output. This raises a basic question for generative learning: **how should local prediction errors be assigned importance according to their consequences for the final outcome?** When model capacity, optimization budgets, and sampling computation are limited, this question also determines which errors learning should prioritize.

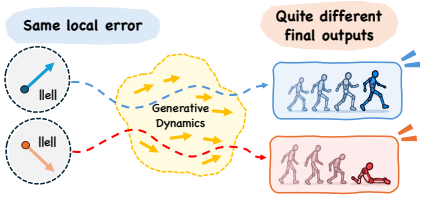


Figure 1: Equal-norm errors, different outcomes.

Flow matching (FM) makes this question particularly concrete (Lipman et al., 2023; Liu et al., 2023). As illustrated in Figure 1, two local errors at the same generation state can have equal magnitude, and therefore equal cost under standard Euclidean regression, yet lead to very different terminal effects. The remaining generation dynamics transform perturbations, rotating, mixing, amplifying, or attenuating their directions. For structured outputs such as motion, some errors may affect only local details, while others propagate to global structure, temporal relations, or contact patterns. Standard FM assigns equal cost to equal-norm velocity residuals, so this difference is not directly represented in its local training signal.

The resulting learning problem extends beyond local regression accuracy: **generation errors have future consequences, and those consequences are directional.** A local direction should matter according to how much it can change the final output after passing through the remaining generator. This perspective is especially relevant when only a few field evaluations are available at deployment.

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.

Recent progress in consistency models, shortcut models, and mean flows highlights the importance of learning under such small sampling budgets (Song et al., 2023; Frans et al., 2025; Geng et al., 2025). Low-step generation therefore provides a representative setting for studying the connection between local training errors and terminal quality.

A natural approach is to make local learning sensitive to the output. Perceptual objectives and geometry-aware motion models show how output features can give task-relevant meaning to changes in an intermediate representation (Johnson et al., 2016; Tevet et al., 2023; Berrada Ifriqi et al., 2025; Zhu et al., 2026). Decoder-induced latent metrics formalize this connection geometrically (Arvanitidis et al., 2018). For a generative process, this geometry is further shaped by the dynamics that remain between the current state and the final output. Those dynamics determine how current perturbations propagate, combine, and eventually appear in the generated result. Understanding a local error therefore requires considering **where it occurs, which direction it points in, and which generation process it will subsequently traverse**. Goal-oriented adjoints provide a classical connection between local residuals and output error (Becker & Rannacher, 2001; Cao & Petzold, 2004); recent decision-weighted FM also connects terminal task sensitivity to velocity regression through scalar weights (Xie et al., 2026). These connections motivate examining directional importance within the residual space at a fixed generation state.

We introduce **terminal-transported residual geometry** to describe the relative terminal effects of current error directions after continuation through the remaining generator. The output space specifies which changes matter, and the generation dynamics transport that importance back to the current state, inducing a directional geometry on local residuals. This yields a general training principle: **local errors should be prioritized by their future impact on the generated outcome.** Such a geometry can reallocate finite model capacity toward errors that matter more to the final result.

We instantiate this principle as **IMPACTFLOW**. During FM training, IMPACTFLOW uses terminal-transported residual geometry to make the model sensitive to the downstream importance of different local error directions while learning the original generative field. The original conditional velocity still defines the regression task; terminal geometry determines the additional emphasis assigned to residual directions. The additional geometry is used only during training, so the student architecture, decoder, and deployment procedure remain unchanged.

We study this perspective at several levels. Our diagnostics show that transported geometry predicts changes measured by an independent output evaluator better than decoder-only geometry and captures interactions across motion tokens. At fixed low-step deployment, controlled training comparisons show that transported orientation improves HumanML3D text-to-motion FID beyond endpoint geometry with matched scalar training strength. A separate equal-time comparison at a Standard-FM-selected common deployment also yields a lower observed mean FID under matched continuation time.

Together, these results support using the remaining generation process to assign geometric meaning to current training errors. By connecting terminal outputs, generation dynamics, and local optimization, IMPACTFLOW illustrates a route to outcome-aware learning while preserving the original form of generation at deployment.

Our main contributions are:
1. **Geometry connecting local errors to terminal effects.** We introduce terminal-transported residual geometry to connect local error directions with their effects on the final generated output.
2. **Outcome-aware residual weighting for flow matching.** We propose IMPACTFLOW, which uses output-induced directional geometry to reallocate emphasis among FM residual directions while retaining the original regression labels and deployment interface.
3. **Directional analysis and controlled low-step evaluation.** We investigate output-effect prediction, directional controls, and low-step HumanML3D generation to assess the information carried by transported geometry and its practical utility.

2 RELATED WORK

Local learning and efficient generation. Learning a generative process from tractable local predictions has roots in score matching and denoising autoencoders (Hyvärinen, 2005; Vincent, 2011), and underlies modern diffusion and flow models (Ho et al., 2020; Song et al., 2021; Lipman

et al., 2023). These formulations connect intermediate prediction tasks to a complete generation process, while finite capacity and computation make the allocation of learning effort consequential. Perception-prioritized training and Min-SNR weighting address this allocation by changing the emphasis across noise levels (Choi et al., 2022; Hang et al., 2023). In parallel, consistency models, shortcut models, and mean flows learn predictions suited to generation with very few evaluations (Song et al., 2023; Frans et al., 2025; Geng et al., 2025). These advances motivate a complementary question: **which error directions at the same state deserve more learning capacity because of their future consequences?** Our focus is the relative importance of local errors within a given generative task, a question that remains relevant alongside improvements to prediction targets, timestep weighting, and sampling efficiency.

Output geometry and perceptual quality. The significance of a prediction error depends on how it changes the output. Perceptual losses and learned similarity metrics established that feature distances can better reflect visual structure than uniform pixel distances (Johnson et al., 2016; Zhang et al., 2018), while decoder-induced Riemannian metrics formalized the nonuniform geometry of latent spaces (Arvanitidis et al., 2018). Recent generative models use decoder-feature objectives to supervise latent predictions (Berrada Ifriqi et al., 2025) and representation alignment to improve diffusion training (Yu et al., 2025). For motion, geometric supervision connects predictions to physical structure (Tevet et al., 2023); MoRAE further studies directional decoder sensitivity and shapes motion latents for flow-based generation (Zhu et al., 2026). These works show that output structure and learned representations determine which deviations matter. For an intermediate generation state, however, **that importance also depends on the dynamics still to come**: an error traverses the remaining generator before its decoded effect is observed. This motivates transporting output geometry back to the current state when assessing local residuals.

Learning from final outcomes. Connecting local errors to final consequences has a classical foundation in goal-oriented error estimation, where adjoints relate residuals to errors in quantities of interest (Becker & Rannacher, 2001; Cao & Petzold, 2004). In generative learning, differentiable reward fine-tuning and Adjoint Matching propagate information from final samples to improve task outcomes (Clark et al., 2024; Domingo-Enrich et al., 2025). Decision-Weighted FM more directly connects terminal decision sensitivity to the emphasis assigned to velocity regression through scalar weights (Xie et al., 2026). Together, these lines establish the value of outcome-aware learning. **Our focus is the directional structure of future impact within the local residual space.** IMPACTFLOW uses the remaining generation process to distinguish equal-norm errors at the same state while retaining the original conditional velocity labels and generation interface. Appendix H expands these connections, including recent work on representation geometry, generative control, and motion generation, and clarifies where each approach introduces output information.

3 TERMINAL-TRANSPORTED RESIDUAL GEOMETRY

IMPACTFLOW turns the future effects of local errors into a training signal (Figure 2). We derive a residual geometry from terminal sensitivity, explain its effect on learning priorities, and estimate it through a frozen reference process. The conditional velocity supplies the label; the reference process supplies the sensitivity.

3.1 FROM LOCAL ERRORS TO TERMINAL EFFECTS

Let $z_0 \sim p_0$ be noise, $z_1 \sim p_1(\cdot \mid c)$ a data latent, and c the generation condition. Flow matching on the linear conditional path uses

$$z_t = (1-t)z_0 + tz_1, \qquad u_t = z_1 - z_0, \qquad r_\theta = v_\theta(z_t, t, c) - u_t. \tag{1}$$

Ordinary FM penalizes the Euclidean magnitude of r_θ. To connect this local error to its future consequences, let $\Phi_{t\to 1}$ denote continuation under a frozen reference field. A decoder D maps its terminal latent to the output, and a readout H extracts the features whose changes matter to the task. Their composition is

$$F_t(z_t, c) = H(D(\Phi_{t\to 1}(z_t, c))) \,. \tag{2}$$

The reference field is provided by a pretrained FM teacher. The choice of H and a fixed metric $W \succeq 0$ specifies the output changes to emphasize; $\|y\|_W^2 = y^\top W y$. For a small valid-coordinate

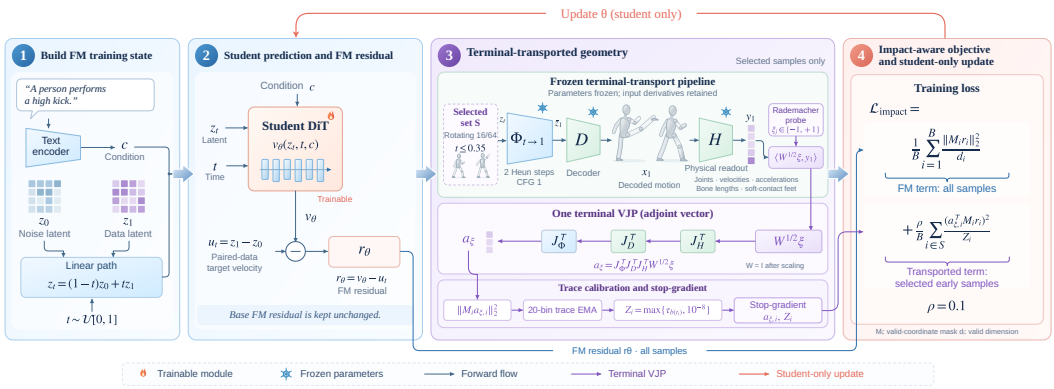


Figure 2: Four-stage reading key: (1) paired data define the FM state; (2) the student predicts velocity against the original FM residual target; (3) a frozen teacher transports output geometry from the physical readout through one terminal VJP; and (4) detached geometry weights guide student-only updates through the auxiliary residual.

perturbation δ, define its **terminal impact** as

$$\mathcal{I}_t(\delta) = \|F_t(z_t + \delta, c) - F_t(z_t, c)\|_W^2. \tag{3}$$

This is sensitivity of the chosen output features under the reference continuation.

3.2 Geometry induced by the remaining generator

With bounded first and second derivatives of F_t in a neighborhood, Taylor expansion gives

$$\mathcal{I}_t(\delta) = \delta^\top G_t^{\text{out}} \delta + O(\|\delta\|^3), \qquad G_t^{\text{out}} = J_{F_t}^\top W J_{F_t}. \tag{4}$$

Applying the chain rule reveals how the remaining dynamics shape this **terminal-transported residual geometry**:

$$G_t^{\text{out}} = J_\Phi^\top \underbrace{J_D^\top J_H^\top W J_H J_D}_{\text{terminal output geometry}} J_\Phi. \tag{5}$$

All Jacobians are evaluated along the reference continuation. The output geometry identifies consequential terminal changes; J_Φ determines which current directions produce them.

The resulting cost is both **directional and joint**. Equal-norm perturbations can align with different eigendirections of G_t^{out} and incur different terminal effects. Moreover, output changes can depend on several latent coordinates together. For a residual partitioned into motion tokens,

$$r^\top G_t^{\text{out}} r = \sum_k r_k^\top G_{kk} r_k + 2 \sum_{k<l} r_k^\top G_{kl} r_l. \tag{6}$$

The cross-token terms capture perturbations that reinforce or cancel after continuation; token position is distinct from generative time t. Appendix A.2 gives an explicit example.

Over a short interval h, a velocity residual r produces a first-order state displacement hr. Consequently, $\mathcal{I}_t(hr)/h^2 = r^\top G_t^{\text{out}} r + O(h\|r\|^3)$. This connects the geometry to the errors that FM actually learns to reduce. The quadratic measures a local displacement in the chosen output features; whether emphasizing that sensitivity improves sample quality is a separate learning question. Perturbations introduced at different generation times can reinforce or cancel, so their combined terminal effect need not equal the sum of per-state costs. Appendix A.1 relates the integrated local geometry to an upper bound on the leading-order squared terminal feature change under a smooth reference flow. Controlled training comparisons then assess its value for generation quality.

3.3 Learning priorities within the original FM task

We retain ordinary FM on every sample and add geometric emphasis on a selected subset. On valid coordinates, write the per-sample objective as $r_\theta^\top A_t r_\theta$, with

$$A_t = d^{-1} I + \rho s_t \widetilde{G}_t, \qquad \widetilde{G}_t \succeq 0. \tag{7}$$

Here d is the valid dimension, $\rho \geq 0$, s_t indicates selection, and $\widetilde{G}_t$ is a scale-calibrated, possibly randomized realization of terminal geometry. Section 3.4 constructs this weight without forming a dense matrix. The Euclidean term makes A_t positive definite and keeps every valid coordinate supervised.

The learning role of this weight follows from a conditional risk decomposition. Let $x = (z_t, t, c)$, including any length and mask information, $\mu(x) = \mathbb{E}[u_t \mid x]$, $\Sigma(x) = \mathrm{Cov}(u_t \mid x)$, and $\bar{A}(x) = \mathbb{E}[A_t \mid x]$.

Proposition 1 (Conditional target and approximation error)**.** Assume the displayed conditional expectations are finite. For $t < 1$, suppose $A_t \succ 0$ is prediction-independent and conditionally independent of the paired endpoint z_1 given x. Then

$$\mathbb{E}[r_\theta^\top A_t r_\theta \mid x] = (v_\theta - \mu)^\top \bar{A}(x)(v_\theta - \mu) + \mathrm{Tr}\big(\bar{A}(x)\Sigma(x)\big)\,, \tag{8}$$

whose unique optimum over predictions is $v = \mu(x)$.

The independence requirement applies jointly to the metric, selection, and normalization, including their randomness. Appendix B.1 provides the proof and the general expression when this requirement fails. The covariance term is independent of the student. For a finite-capacity model, $\bar{A}(x)$ therefore changes the cost of its approximation errors around the same conditional mean velocity. **Geometry reallocates learning effort within the original regression task**: directions with larger transported impact receive additional emphasis, while the conditional labels remain unchanged.

With a detached weight, the residual-space gradient is $2A_t r_\theta$. The geometric term can change both its magnitude and its direction, including how corrections are distributed across motion tokens. Its pullback through the student Jacobian gives the parameter gradient. A scalar weight on the same residual only rescales its Euclidean gradient. The scalar-matched controls in Section 4.2.2 test whether transported orientation contributes beyond changes in update strength.

For a frozen reference satisfying the stated positivity and conditional independence conditions, target preservation and transported sensitivity play complementary roles: the former retains the original regression target, while the latter allocates learning effort across residual directions. Appendix B.2 connects field approximation, reference sensitivity, and training/reference state laws to terminal sensitivity. The proposition establishes consistency of the regression target, and the experiments evaluate the transported geometry used in training.

3.4 Efficient estimation and student updates

Differentiating the reference continuation. At fixed (t, c), a K-step numerical teacher defines $\widehat{F}_K = H \circ D \circ \widehat{\Phi}_K$ and $\widehat{G}_K = J_{\widehat{F}_K}^\top W J_{\widehat{F}_K}$. We differentiate this discrete map with respect to its input state; teacher, decoder, and readout parameters remain frozen. Appendix C.2 relates this geometry to the continuous reference. The student and teacher start from the same pretrained field.

A random output direction supplies the quadratic. Draw ξ independently of the state, residual, and frozen map, with $\mathbb{E}\xi = 0$ and $\mathbb{E}[\xi\xi^\top] = I$. One terminal vector–Jacobian product (VJP) yields

$$a_\xi = J_{\widehat{F}_K}^\top W^{1/2}\xi = \nabla_{z_t}\langle W^{1/2}\xi, \widehat{F}_K(z_t, c)\rangle. \tag{9}$$

For fixed r and the numerical map,

$$\mathbb{E}_\xi[(a_\xi^\top r)^2] = r^\top \widehat{G}_K r, \qquad \mathbb{E}_\xi\|a_\xi\|^2 = \mathrm{Tr}(\widehat{G}_K). \tag{10}$$

Thus a single random rank-one estimate provides both the directional cost and a sensitivity-scale sample. We use one Rademacher probe per selected sample; Appendix C.1 gives its moments.

Calibrating scale and defining the update. Let $M_i = M_i^\top = M_i^2$ mask padding and $d_i = \mathrm{Tr}(M_i) > 0$. Masked trace samples $\|M_i a_{\xi,i}\|^2$ define a positive, timestep-conditioned scale Z_i that calibrates typical sensitivity. The unnormalized quadratic and trace estimates are unbiased; normalization with a probe-dependent scale defines its own random metric, as analyzed in Appendix C.2. For a batch of B samples and $\mathcal{S} \subseteq \{1, \ldots, B\}$, the implemented loss is (sg denotes stop-gradient)

$$\mathcal{L}_{\text{impact}} = \frac{1}{B}\sum_{i=1}^{B} \frac{\|M_i r_i\|_2^2}{d_i} + \frac{\rho}{B}\sum_{i\in\mathcal{S}} \frac{[\mathrm{sg}(a_{\xi,i})^\top M_i r_i]^2}{\mathrm{sg}(Z_i)}. \tag{11}$$

The subset $\mathcal{S}$ is selected without using the paired endpoint beyond x. On valid coordinates, $\widetilde{G}_i = \mathrm{sg}(M_i a_{\xi,i} a_{\xi,i}^\top M_i / Z_i)$ realizes equation 7. Detachment holds the geometric weight fixed during the student update, avoiding higher derivatives through the reference process.

Algorithm 1 One IMPACTFLOW training update

1: Construct FM states, conditional labels, and student residuals.
2: Select $\mathcal{S}$ for geometry evaluation using the endpoint-independent selection rule.
3: **for** each $i \in \mathcal{S}$ **do**
4: Continue the frozen teacher and evaluate $\widehat{F}_K(z_{t_i}, c_i)$.
5: Draw ξ_i; compute $a_{\xi,i}$ and its masked trace sample by one VJP.
6: **end for**
7: Set Z_i from the timestep-conditioned trace statistics; update those statistics.
8: Detach $a_{\xi,i}, Z_i$ and minimize equation 11 over θ only.

Motion instantiation. For text-to-motion generation, H combines joint positions, temporal differences, bone lengths, and contact-related foot displacement. We scale each feature group by its training-set standard deviation and set $W = I$ after scaling, independently of the evaluation embedding. Appendix C specifies the readout, sample selection, solver, and trace calibration. All geometry estimation is confined to training; generation uses the original student–decoder interface.

4 EXPERIMENTS

We evaluate IMPACTFLOW through four questions. **RQ1**: Does remaining-flow transport lower HumanML3D FID relative to scalar-matched endpoint geometry at a fixed low-step deployment? **RQ2**: Does the complete objective yield lower mean FID under equal continuation time at a Standard-selected common deployment? **RQ3**: Does transported geometry predict output effects, and does its direction matter after scalar training strength is matched? **RQ4**: How should a fixed continuation budget be allocated between geometry estimation and student updates? RQ1 isolates the transport effect; RQ2 evaluates the complete objective under a compute budget. RQ3 and RQ4 examine its mechanism and cost.

Setup. We use HumanML3D's official split and 263D representation (Guo et al., 2022); a shared frozen autoencoder maps at most 196 frames to 49 tokens of width 64. All methods share the 39.1M-parameter AdaLN DiT and seed-specific 600-epoch Standard FM parents. Standard FM uses the Euclidean residual objective; IMPACTFLOW denotes the complete proposed objective. A **fixed-deployment** comparison holds solver, NFE, and classifier-free guidance (CFG) constant. An **equal-time common-deployment** comparison selects solver and CFG using only Standard FM's mean FID, and compares IMPACTFLOW at the same solver, NFE, and CFG under the same continuation-time budget. Generation comparisons share text inputs, evaluation population, and the independent HumanML3D evaluator; training seeds are the inferential units. Appendices C and D give the paired comparisons, statistical details, and authoritative frozen configuration names.

4.1 MAIN RESULTS

4.1.1 FID AT FIXED DEPLOYMENT

Fixed-deployment transport effect (RQ1). At fixed Euler NFE8 / CFG 4.5, IMPACTFLOW lowers HumanML3D FID relative to Endpoint-matched after scalar-strength matching. Across ten fresh training seeds, the mean seedwise reduction is 13.23% $[5.24, 21.22]$ and the absolute paired improvement is 0.03686 $[0.01387, 0.05984]$; nine of ten effects favor IMPACTFLOW. For this absolute effect, two-sided exact paired sign-flip $p = 0.0098$. The fresh-seed result supports transported residual orientation beyond scalar training strength (Appendix D.3).

IMPACTFLOW is also compared with Standard FM and LPL-style under the same fixed-update evaluation. LPL-style applies physical decoder-feature supervision to one-shot clean predictions and paired endpoints.

Table 1: Fixed-deployment HumanML3D FID at Euler NFE8 / CFG 4.5. All arms use 9,100 updates. Reductions average paired seedwise effects; Panel B reports the separate five-seed objective comparison.

Comparator	IMPACTFLOW FID	Comparator FID	Mean reduction (%) [95% CI]
A. Fresh-seed confirmation ($n = 10$)			
Endpoint-matched	**0.22875**	0.26560	13.23% $[5.24, 21.22]$
B. Separate objective comparison ($n = 5$)			
Standard FM	**0.22171**	0.24674	10.39% $[2.43, 18.34]$
LPL-style	**0.22171**	0.24474	9.66% $[1.67, 17.65]$

Scope: Panel A is a prespecified ten-seed comparison with four replications per seed. Panel B uses a separate five-seed cohort. Intervals and multiplicity are specified in Appendix D.2.

Standard and endpoint-feature objectives. IMPACTFLOW also has lower observed mean FID than Standard FM and LPL-style in the separate fixed-update comparison (Table 1, Panel B). The objective definitions and corresponding statistical analysis are given in Appendices C.4 and D.2. We next test the complete objective under a shared time cap.

4.1.2 EQUAL-TIME FID AT A STANDARD-SELECTED COMMON DEPLOYMENT

Equal-time effectiveness (RQ2). At the Standard-selected midpoint NFE8 / CFG 3 deployment, equal-time IMPACTFLOW continuations have observed mean FID 0.08890, compared with 0.10026 for Standard FM. The mean seedwise relative reduction is 10.87% $[-1.86, 23.60]$, with seven of ten paired training seeds favorable. The selected deployment is held fixed across the ten paired seeds, yielding the reported common-deployment effect.

Table 2: Observed HumanML3D FID under the equal-time common deployment. Both objectives use the Standard-selected midpoint NFE8 / CFG 3 deployment; reductions average paired seedwise percentages.

Objective	Solver / CFG	FID	Mean reduction (%) [95% CI]
Standard FM	Midpoint / 3.0	0.10026	–
IMPACTFLOW	Midpoint / 3.0	**0.08890**	10.87% $[-1.86, 23.60]$

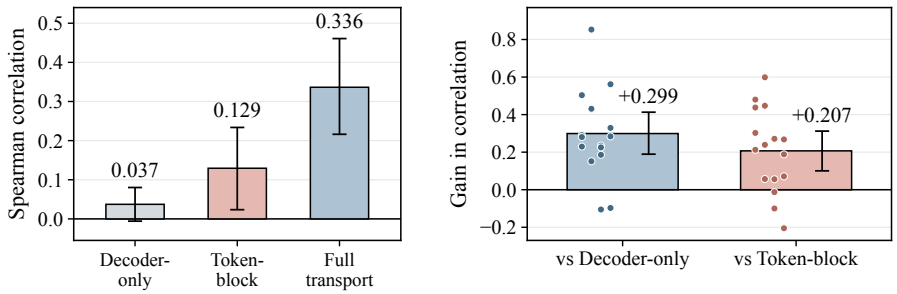
Scope: Ten paired continuations share the same 1,919-second four-GPU cap, parents, and selected deployment. Intervals are nominal paired-seed t intervals with selection held fixed. Appendix D.4 gives the scan, replications, absolute effect, and exact tests.

This comparison includes geometry computation in the training budget, measuring the complete objective when sensitivity estimation and student updates compete for time. The remaining experiments examine what the transported direction contributes and how precisely it needs to be estimated.

4.2 MECHANISM AND ANALYSIS

4.2.1 TERMINAL TRANSPORT EXPOSES DIRECTIONAL STRUCTURE

Independent-evaluator prediction (RQ3). Across 16 early motion-FM states, transported geometry has mean statewise Spearman correlation 0.336 $[0.216, 0.461]$. Its paired advantage over Decoder-only geometry is 0.299 $[0.189, 0.413]$. **Evaluator damage** is the normalized squared change in a frozen evaluator embedding under a small centered latent perturbation. Decoder-only geometry uses the same terminal state and readout but omits remaining-flow differentiation. The paired gap therefore supports output-relevant directional information from the remaining flow (Appendix E.1). The Token-block comparator in Figure 3 retains the IMPACTFLOW within-token blocks and removes the cross-token terms.



(a) Mean predictive agreement (b) Paired gains across states

Figure 3: Independent output-effect prediction in 16 states at $t = 0.1$ and 0.2, where for one state agreement is the Spearman correlation across latent perturbation directions between a physical-readout geometry score and evaluator damage, the normalized squared change in the frozen evaluator embedding. Decoder-only omits remaining-flow transport; Token-block retains only within-token blocks. (a) Mean statewise correlations. (b) Paired within-state gains; dots show all 16 states, bars show means, and whiskers show 95% state-cluster bootstrap intervals.

Structure within physical geometry. On a separate frozen MoRAE generator, physical geometry is anisotropic and spectrally concentrated. Here, **physical-readout sensitivity** is the squared response to an equal-norm latent direction; its geometric-mean P90/P10 ratio is 4.81 $[3.67, 6.47]$, and the leading eight directions capture 57.7% and 53.3% of the trace at $t = 0.1$ and $t = 0.2$. Appendix E.1 gives the full protocol.

From sensitivity to training. IMPACTFLOW transport predicts evaluator changes better than Decoder-only in 14 of 16 states and Token-block in 13 of 16 (Figure 3(b)). These paired gains motivate a training-level test of transported orientation under matched scalar strength.

4.2.2 DIRECTION MATTERS BEYOND SCALAR STRENGTH

Directional attribution (RQ3). Across three separate five-seed comparisons, IMPACTFLOW has lower mean FID than each direction-changing control. Token-sign tests coherent cross-token orientation, Token-matched additionally equalizes scalar training strength, and Endpoint-matched removes remaining-flow transport. Each row uses its own independently trained IMPACTFLOW reference.

Table 3: Directional controls at fixed Euler NFE8 / CFG 4.5 and 9,100 updates. Each row is a separate paired experiment; reductions average seedwise percentages relative to the listed control.

Control	IMPACTFLOW FID	Control FID	Mean reduction (%) [95% CI]
Token-sign	**0.22154**	0.24978	11.31% $[6.72, 15.91]$
Token-matched	**0.22154**	0.23759	6.69% $[1.92, 11.46]$
Endpoint-matched	**0.20735**	0.26185	20.80% $[11.40, 30.20]$

Inference: Every row has five favorable seed effects; each primary has one-sided exact $p = 0.03125$ (two-sided $p = 0.0625$), with pointwise paired-seed t intervals. The repeated 0.22154 means come from independent arms.

Control summary. Token-sign changes cross-token orientation; Token-matched also matches total-gradient norm, auxiliary/base ratio, clipping, and FP32 update length. Endpoint-matched substitutes a norm-matched endpoint adjoint with the same scalar matching (Appendices F and F.2).

Directional interpretation. The fresh-seed confirmation and Token-sign contrast support coherent cross-token orientation beyond within-token geometry, and Token-matched retains the lower observed FID after scalar-strength matching. Endpoint-matched isolates remaining-flow transport at a common terminal state (Section 4.1.1).

The scalar-matched controls isolate directional effects under per-step budgets inherited from the paired IMPACTFLOW arm. Matching the auxiliary/base ratio, total-gradient norm, clipping, and FP32 update length attributes the resulting differences to direction while preserving realized training strength. Standard FM and LPL-style provide complementary complete-objective comparisons under the stated settings.

4.2.3 ALLOCATING COMPUTE BETWEEN GEOMETRY AND STUDENT UPDATES

Compute allocation (RQ4). At the common midpoint deployment, the one-probe, two-step estimator retains the most student updates and has the lowest observed mean FID, 0.08450. Increasing probe count or rollout depth raises fixed-checkpoint gradient cosine as high as 0.832 but leaves fewer student updates.

Table 4: Fixed-checkpoint estimator accuracy and NFE8 continuation outcomes under the shared 1,919-second cap. Cosine uses the independent 256-probe, eight-step reference $g^{\mathrm{ref}}_{8,256}$.

Teacher Heun steps K	Probes P	Gradient cosine	Updates	Euler FID	Midpoint FID
2	1	0.334	9311.7	0.21148	**0.08450**
2	4	0.548	3912.7	0.21013	0.08727
8	1	0.574	2856.3	0.27889	0.12625
8	4	0.832	1082.7	0.23820	0.09272

Protocol: Three training seeds. Gradient estimates share a fixed trace normalizer; continuations use evolving normalization. The default models also serve as the readout-ablation reference (Appendix D.1).

Under the shared cap, higher fixed-checkpoint cosine leaves fewer student updates; the economical estimator yields the lowest observed midpoint FID (Appendix E.3).

Readout directions at fixed estimator settings. In the three-seed comparison with one probe and two teacher steps fixed, the complete physical readout has the lowest observed mean midpoint FID: 0.08450, versus 0.11363 for positions and 0.09412 without contact. Changing H changes the output geometry while holding probe count and teacher depth fixed. Appendix G.3 defines the feature blocks and reports both deployments.

Implications for training design. Readout design determines which output changes receive weight, while estimator cost determines how often the student uses that information. These complementary studies favor the complete physical readout and one-probe, two-step estimator within the tested budget, with an unchanged deployment interface.

5 CONCLUSION

We have studied how local generative errors should be prioritized when their importance is determined by a completed output. Terminal-transported residual geometry connects that importance to both output structure and the dynamics that remain. IMPACTFLOW uses this connection to weight FM residual directions while keeping the original conditional velocity labels and generation interface. HumanML3D experiments show that transport adds output-relevant information beyond decoder-only geometry and yields a fixed-deployment FID benefit after scalar-strength matching; equal-time comparisons yield a lower mean FID at the shared Standard-selected deployment. Together, these findings support treating future impact as a geometric property of local residuals in controlled low-step motion generation.

AI USE STATEMENT

Generative AI tools were used only to aid or polish writing and to assist in proving mathematical claims. The authors reviewed the resulting text and proofs and take responsibility for the final manuscript.

ETHICS STATEMENT

This work uses the publicly released HumanML3D benchmark and its official data split; it does not collect new human-subject data or personal identifiers. The study evaluates aggregate text-to-motion generation behavior and makes no claims about individual people or demographic groups. The released code and provenance records support transparent research use of the reported models and evaluations.

REPRODUCIBILITY STATEMENT

Sections 3 and 4 define the training objective, deployment comparisons, and reported estimands. Appendices C and D specify the matrix-free training algorithm, discrete teacher map, dataset and evaluator setup, comparison design, and statistical procedures. The study-configuration index in Appendix D.5 records the semantic frozen configurations, seed and parent cohorts, budgets, reuse relations, and inferential units for the reported comparisons. The anonymous source package provides the implementation, frozen configurations, and launch and aggregation scripts. Complete execution parameters, random streams, precision policies, and evaluation grids are recorded in machine-readable form. Together, the manuscript and source package specify the complete protocol needed to rerun the experiments.

REFERENCES

Georgios Arvanitidis, Lars Kai Hansen, and Søren Hauberg. Latent space oddity: on the curvature of deep generative models. In International Conference on Learning Representations, 2018. URL https://arxiv.org/abs/1710.11379.

Roland Becker and Rolf Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. Acta Numerica, 10:1–102, 2001. doi: 10.1017/S0962492901000010.

Tariq Berrada Ifriqi, Pietro Astolfi, Melissa Hall, Marton Havasi, Yohann Benchetrit, Adriana Romero-Soriano, Karteek Alahari, Michal Drozdzal, and Jakob Verbeek. Boosting latent diffusion with perceptual objectives. In International Conference on Learning Representations, pp. 12128–12141, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/204fee94c982a19230c39045aa54f977-Abstract-Conference.html.

Yang Cao and Linda Petzold. A posteriori error estimation and global error control for ordinary differential equations by the adjoint method. SIAM Journal on Scientific Computing, 26(2): 359–374, 2004. doi: 10.1137/S1064827503420969.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In International Conference on Learning Representations, 2024. URL https://arxiv.org/abs/2302.03660.

Won June Cho, Daeky Jeong, Hyeongyeol Lim, and Hongjun Yoon. STREAM: Stochastic riemannian flow matching with anisotropic decoder for digital histopathology image generation. arXiv preprint arXiv:2606.07036, 2026. URL https://arxiv.org/abs/2606.07036.

Jooyoung Choi, Jungbeom Lee, Chaehun Shin, Sungwon Kim, Hyunwoo Kim, and Sungroh Yoon. Perception prioritized training of diffusion models. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11472–11481, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Choi_Perception_Prioritized_Training_of_Diffusion_Models_CVPR_2022_paper.html.

Kevin Clark, Paul Vicol, Kevin Swersky, and David J. Fleet. Directly fine-tuning diffusion models on differentiable rewards. In International Conference on Learning Representations, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/13e8be77982beb73d7ed0bbf122f9f3c-Abstract-Conference.html.

Friso de Kruiff, Erik Bekkers, Ozan Öktem, Carola-Bibiane Schönlieb, and Willem Diepeveen. Pullback flow matching on data manifolds. arXiv preprint arXiv:2410.04543, 2024. URL https://arxiv.org/abs/2410.04543.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In International Conference on Learning Representations, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/852f50969a9e523ec41d26f2f68bd456-Abstract-Conference.html.

Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. In International Conference on Learning Representations, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/559a0998fab1d19b80e7e43a5852401c-Abstract-Conference.html.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In Advances in Neural Information Processing Systems, volume 38, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/6d13e085b79d454da5910e4ca82a3d9d-Abstract-Conference.html.

Chuan Guo, Shihao Zou, Xinxin Zuo, Sen Wang, Wei Ji, Xingyu Li, and Li Cheng. Generating diverse and natural 3d human motions from text. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5152–5161, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Guo_Generating_Diverse_and_Natural_3D_Human_Motions_From_Text_CVPR_2022_paper.html.

Tiankai Hang, Shuyang Gu, Chen Li, Jianmin Bao, Dong Chen, Han Hu, Xin Geng, and Baining Guo. Efficient diffusion training via Min-SNR weighting strategy. In IEEE/CVF International Conference on Computer Vision, pp. 7441–7451, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Hang_Efficient_Diffusion_Training_via_Min-SNR_Weighting_Strategy_ICCV_2023_paper.html.

Aaron J. Havens, Benjamin Kurt Miller, Bing Yan, Carles Domingo-Enrich, Anuroop Sriram, Daniel S. Levine, Brandon M. Wood, Bin Hu, Brandon Amos, Brian Karrer, Xiang Fu, Guan-Horng Liu, and Ricky T. Q. Chen. Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. In International Conference on Machine Learning, volume 267, pp. 22204–22237, 2025. URL https://proceedings.mlr.press/v267/havens25a.html.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In Advances in Neural Information Processing Systems, volume 33, pp. 6840–6851, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

M. F. Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. Communications in Statistics - Simulation and Computation, 18(3):1059–1076, 1989. doi: 10.1080/03610918908812806.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. Journal of Machine Learning Research, 6(24):695–709, 2005. URL https://www.jmlr.org/papers/v6/hyvarinen05a.html.

Justin Johnson, Alexandre Alahi, and Fei-Fei Li. Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision, pp. 694–711, 2016. URL https://cs.stanford.edu/people/jcjohns/eccv16/.

Kacper Kapuśniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Avishek Joey Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold. In Advances in Neural Information Processing Systems,

2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/f381114cf5aba4e45552869863deaaa7-Abstract-Conference.html.

Xingjian Leng, Jaskirat Singh, Yunzhong Hou, Zhenchang Xing, Saining Xie, and Liang Zheng. REPA-E: Unlocking VAE for end-to-end tuning with latent diffusion transformers. In IEEE/CVF International Conference on Computer Vision, pp. 18262–18272, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Leng_REPA-E_Unlocking_VAE_for_End-to-End_Tuning_of_Latent_Diffusion_Transformers_ICCV_2025_paper.html.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In International Conference on Learning Representations, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds. In International Conference on Learning Representations, 2022. URL https://arxiv.org/abs/2202.09778.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In International Conference on Learning Representations, 2023. URL https://arxiv.org/abs/2209.03003.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In Advances in Neural Information Processing Systems, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/260a14acce2a89dad36adc8eefe7c59e-Abstract-Conference.html.

Zichong Meng, Yiming Xie, Xiaogang Peng, Zeyu Han, and Huaizu Jiang. Rethinking diffusion for text-driven human motion generation: Redundant representations, evaluation, and masked autoregression. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Meng_Rethinking_Diffusion_for_Text-Driven_Human_Motion_Generation_Redundant_Representations_Evaluation_CVPR_2025_paper.html.

Fangran Miao, Jian Huang, and Ting Li. Riemannian motion generation: A unified framework for human motion representation and generation via riemannian flow matching. arXiv preprint arXiv:2603.15016, 2026. URL https://arxiv.org/abs/2603.15016.

Mathis Petrovich, Michael J. Black, and Gül Varol. Action-conditioned 3d human motion synthesis with transformer VAE. In International Conference on Computer Vision, pp. 10985–10995, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Petrovich_Action-Conditioned_3D_Human_Motion_Synthesis_With_Transformer_VAE_ICCV_2021_paper.html.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In International Conference on Learning Representations, 2022. URL https://openreview.net/forum?id=TIdIXIpzhoI.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In International Conference on Machine Learning, volume 37, pp. 2256–2265, 2015. URL https://proceedings.mlr.press/v37/sohl-dickstein15.html.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In International Conference on Learning Representations, 2021. URL https://arxiv.org/abs/2011.13456.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In International Conference on Machine Learning, volume 202, pp. 32211–32252, 2023. URL https://proceedings.mlr.press/v202/song23a.html.

Zheng Tan, Weizhen Wang, Andrea L. Bertozzi, and Ernest K. Ryu. STORK: Faster diffusion and flow matching sampling by resolving both stiffness and structure-dependence. In International Conference on Learning Representations, pp. 75478–75506, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/7a641b8ec86162fc875fb9f6456a542f-Abstract-Conference.html.

Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit H. Bermano. Human motion diffusion model. In International Conference on Learning Representations, 2023. URL https://openreview.net/forum?id=SJ1kSyO2jwu.

Pascal Vincent. A connection between score matching and denoising autoencoders. Neural Computation, 23(7):1661–1674, 2011. doi: 10.1162/NECO_a_00142. URL https://doi.org/10.1162/NECO_a_00142.

Jize Xie, Haomiao Wu, Qiang Chen, Xiu Su, and Yi Chen. Decision-weighted flow matching for contextual stochastic optimization. arXiv preprint arXiv:2606.16790, 2026. URL https://arxiv.org/abs/2606.16790.

Sihyun Yu, Sangkyung Kwak, Huiwon Jang, Jongheon Jeong, Jonathan Huang, Jinwoo Shin, and Saining Xie. Representation alignment for generation: Training diffusion transformers is easier than you think. In International Conference on Learning Representations, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/d9e42b4d7163931f3689d6d6fbaa11d0-Abstract-Conference.html.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–595, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_The_Unreasonable_Effectiveness_CVPR_2018_paper.html.

Yifei Zhu, Mingyi Shi, Yangyang Cai, Miao Cheng, Yoshifumi Kitamura, and Taku Komura. MoRAE: Flow-friendly self-supervised latents for text-to-motion generation. arXiv preprint arXiv:2607.29180, 2026. URL https://arxiv.org/abs/2607.29180.

APPENDIX OVERVIEW

The appendix follows the connection from terminal effects to residual geometry, learning, implementation, and empirical evidence:

- **Terminal effects and geometry.** Appendix A proves the trajectory-sensitivity bound and illustrates interactions between residual directions.
- **Learning from the geometry.** Appendix B establishes target preservation and connects regression geometry, reference sensitivity, and training/reference state laws.
- **Estimation and implementation.** Appendix C derives the randomized estimator and numerical geometry, then specifies the readout, training algorithm, and compared objectives.
- **Experimental evidence.** Appendix D describes the paired studies and statistical inference. Appendices E, F, and G give the sensitivity diagnostics, directional controls, and compute/readout studies.
- **Broader context.** Appendix H expands the connections to related work.

A TERMINAL EFFECTS AND RESIDUAL GEOMETRY

Throughout the analysis, conditioning includes the task input and sequence length/mask, and matrices act on valid latent coordinates. Vector norms are Euclidean and matrix norms are induced spectral norms; $\|y\|_W^2 = y^\top W y$ for a fixed $W \succeq 0$.

A.1 RELATION TO TRAJECTORY PERTURBATIONS

The terminal pullback is the within-time block of a trajectory-wide sensitivity kernel. A local perturbation result makes the connection precise.

Reference and regularity. Fix c and z_0, put $g = H \circ D$, and let $\bar{z}_t$ solve $\dot{\bar{z}}_t = \bar{v}(\bar{z}_t, t, c)$ on $[0,1]$. Consider the interpolated field $v^\epsilon = \bar{v} + \epsilon w$ and the trajectory z_t^ϵ with the same initial state. Assume $\bar{v}$ and w are continuous in time and twice continuously differentiable in state, with uniformly bounded derivatives through order two on a common neighborhood of these trajectories for sufficiently small $|\epsilon|$. Assume the trajectories exist in that neighborhood and g is twice continuously differentiable near their endpoints. These local conditions give a twice differentiable dependence on ϵ and bounded Taylor remainders.

Define the state-transition matrix by

$$\partial_s \Psi(s,t) = D_z \bar{v}(\bar{z}_s, s, c)\Psi(s,t), \qquad \Psi(t,t) = I.$$

Along the reference trajectory,

$$L_t = J_g(\bar{z}_1)\Psi(1,t) = J_{F_t^{\bar{v}}}(\bar{z}_t, c), \qquad G_t = L_t^\top W L_t.$$

Proposition 2 (Local terminal sensitivity)**.** Under the reference and regularity conditions above, write $w_t = w(\bar{z}_t, t, c)$ and $\Delta y_\epsilon = g(z_1^\epsilon) - g(\bar{z}_1)$. Then

$$\Delta y_\epsilon = \epsilon \int_0^1 L_t w_t \, dt + O(\epsilon^2), \tag{12}$$

$$\|\Delta y_\epsilon\|_W^2 = \epsilon^2 \int_0^1 \int_0^1 w_t^\top K(t,s) w_s \, dt \, ds + O(|\epsilon|^3), \qquad K(t,s) = L_t^\top W L_s. \tag{13}$$

For a finite constant C independent of sufficiently small ϵ,

$$\|\Delta y_\epsilon\|_W^2 \le \epsilon^2 \int_0^1 w_t^\top G_t w_t \, dt + C|\epsilon|^3. \tag{14}$$

Proof. The first variation $\eta_t = \partial_\epsilon z_t^\epsilon|_{\epsilon=0}$ satisfies

$$\dot{\eta}_t = D_z \bar{v}(\bar{z}_t, t, c)\eta_t + w_t, \qquad \eta_0 = 0.$$

Variation of constants gives $\eta_1 = \int_0^1 \Psi(1,t) w_t \, dt$. Taylor expansion of $g(z_1^\epsilon)$ yields equation 12. Squaring in the W seminorm gives equation 13; the product of the first-order term and the bounded second-order remainder is $O(|\epsilon|^3)$. Finally, Jensen's inequality on the unit interval gives

$$\left\| \int_0^1 W^{1/2} L_t w_t \, dt \right\|_2^2 \le \int_0^1 \|W^{1/2} L_t w_t\|_2^2 \, dt = \int_0^1 w_t^\top G_t w_t \, dt.$$

Absorbing the squared-output remainder into $C|\epsilon|^3$ proves equation 14. □

The kernel satisfies $K(s,t) = K(t,s)^\top$ and $K(t,t) = G_t$. Its double integral includes reinforcement and cancellation between generation times; Jensen supplies an envelope for their combined effect. The envelope is tight at first order when $W^{1/2} L_t w_t$ is constant in time. Generation time t is distinct from motion-token position: off-diagonal entries within G_t couple latent tokens, whereas $K(t,s)$ relates perturbations at different generation times.

A.2 Interactions between residual directions

Here, t is generative time, distinct from motion-token position. Partitioning the residual by token gives

$$r^\top G_t^{\text{out}} r = \sum_k r_k^\top G_{kk} r_k + 2 \sum_{k<l} r_k^\top G_{kl} r_l.$$

Tokenwise perturbations can reinforce or cancel after continuation. Thus unchanged individual-token sensitivities can produce different joint costs through the cross-token terms. For two scalar tokens, let $r = (1,1)$, $G_{11} = G_{22} = 1$, and $G_{12} = \pm\gamma$, $0 < \gamma < 1$. Equal self-costs then give joint cost $2 \pm 2\gamma$: reinforcement for $+\gamma$, cancellation for $-\gamma$. For a general residual $r = (r_1, r_2)$, the derivative of this illustrative unnormalized full quadratic is

$$\nabla_r (r^\top G_\pm r) = 2 G_\pm r = 2(r_1 \pm \gamma r_2, \; r_2 \pm \gamma r_1). \tag{15}$$

The general-residual formula establishes coordinate-ratio and joint-cost changes. The symmetric $r = (1,1)$ illustrates reinforcement/cancellation and scalar rescaling: its correction is $2(1 \pm \gamma)(1,1)$. The one-probe estimator in Appendix C.1 is a stochastic rank-one sketch of this unnormalized quadratic; before detached trace scaling, its expected residual derivative recovers $2G_t^{\text{out}} r$.

B Geometry-weighted flow matching

We first establish when residual weighting preserves the conditional FM target, then connect its excess risk to the terminal-sensitivity bound.

B.1 Population-optimum proof

Positive random weights preserve the conditional regression target under joint endpoint independence; the selected-sample sketch fits this construction. Condition on $x = (z_t, t, c)$ and set $\mu(x) = \mathbb{E}[u_t \mid x]$, $\Sigma(x) = \mathrm{Cov}(u_t \mid x)$. Let $\widehat{A}$ be a prediction-independent symmetric random weight that is conditionally independent of u_t given x, and let $\bar{A}(x) = \mathbb{E}[\widehat{A} \mid x] \succ 0$. Assume the displayed quadratic expectations are finite. For the linear path and $t < 1$, $u_t = (z_1 - z_t)/(1-t)$, so endpoint independence implies this target-independence condition. The uniform training-time law assigns zero mass to $t = 1$. For a fixed prediction v,

$$\begin{aligned} \mathbb{E}[(v - u_t)^\top \widehat{A} (v - u_t) \mid x] &= v^\top \bar{A} v - 2 v^\top \bar{A} \mu + \mathbb{E}[u_t^\top \bar{A} u_t \mid x] \\ &= (v - \mu)^\top \bar{A} (v - \mu) + \mathrm{Tr}(\bar{A}\Sigma). \end{aligned}$$

Positive definiteness gives the unique optimum $v = \mu(x)$, proving Proposition 1 and its random-weight extension. For the implemented weight,

$$\widehat{A}_i = d_i^{-1} I + \rho \mathbf{1}[i \in \mathcal{S}] \widetilde{a}_{\xi,i} \widetilde{a}_{\xi,i}^\top / Z_i,$$

the masked adjoint $\widetilde{a}_{\xi,i} = M_i a_{\xi,i}$ is restricted to valid coordinates, $\rho \ge 0$, and $Z_i > 0$. Consequently $\widehat{A}_i \succeq d_i^{-1} I$. The target-preservation condition applies jointly to the metric, selection, and normalizer,

including their randomness and any batch or history information. Freezing the numerical teacher and readout fixes the geometry for a training state; detaching their adjoint fixes the student differentiation path. Joint conditional independence of the resulting weight supplies the population-optimum statement. The FM labels remain unchanged in the loss construction in every case.

The general normal equation makes the role of this condition explicit. For an endpoint-dependent weight with the same integrability and positive-definiteness properties, the optimum is

$$v^*_{\widehat{A}}(x) = \bar{A}(x)^{-1}\mathbb{E}[\widehat{A}u_t \mid x] = \mu(x) + \bar{A}(x)^{-1}\mathbb{E}[\widehat{A}(u_t - \mu(x)) \mid x]. \tag{16}$$

Equation 16 identifies the exact zero-cross-moment condition for retaining the original conditional target. Conditional independence guarantees this condition.

B.2 Connecting terminal geometry to the FM regression problem

The conditional regression decomposition identifies the part of the training residual that measures field approximation. For an ideal frozen-teacher metric $G(x) \succeq 0$, define

$$\mathcal{R}_G(v \mid x) = \mathbb{E}[(v - u_t)^\top G(x)(v - u_t) \mid x].$$

The proof in Appendix B.1 gives the excess-risk identity

$$\mathcal{R}_G(v_\theta \mid x) - \mathcal{R}_G(\mu \mid x) = \|v_\theta - \mu\|^2_{G(x)}. \tag{17}$$

The conditional covariance of u_t contributes the same additive term to both risks. In particular, paired-label variation is distinct from the student field's approximation error.

Reference-field decomposition. The perturbation direction relative to the reference field is

$$v_\theta - \bar{v} = \underbrace{v_\theta - \mu}_{e_\theta} + \underbrace{\mu - \bar{v}}_{b}.$$

For any $\gamma > 0$, Young's inequality in the G seminorm yields

$$\|v_\theta - \bar{v}\|^2_G \le (1+\gamma)\|e_\theta\|^2_G + (1+\gamma^{-1})\|b\|^2_G. \tag{18}$$

Thus the reference sensitivity decomposes into the student field's approximation component and the reference field's alignment with the marginal FM field.

Training-to-reference state bridge. Let P_t be the FM training-state law and Q_t the reference-flow state law, conditional on the same task input. At a time where $Q_t \ll P_t$ and $dQ_t/dP_t \le \kappa_t < \infty$, nonnegative integrands satisfy $\mathbb{E}_{Q_t} f \le \kappa_t \mathbb{E}_{P_t} f$. Combining this comparison with equations 17 and 18 gives

$$\begin{aligned}\mathbb{E}_{Q_t}\|v_\theta - \bar{v}\|^2_G \le (1+\gamma)\kappa_t \mathbb{E}_{P_t}[\mathcal{R}_G(v_\theta \mid x) - \mathcal{R}_G(\mu \mid x)] \\ + (1+\gamma^{-1})\mathbb{E}_{Q_t}\|b\|^2_G.\end{aligned} \tag{19}$$

This conditional comparison transfers training risk to the reference-trajectory integrand through the density ratio and reference-field alignment. For a matched teacher and matched state laws, $b = 0$ and $Q_t = P_t$, and the integrand's expectation equals the weighted excess regression risk. In the implementation, the selected, normalized discrete metric below defines the actual training risk.

C Geometry estimation and implementation

C.1 Moments of the randomized quadratic

Fix a differentiable terminal map, a realized residual $r \in \mathbb{R}^d$, and its output-space metric $W \in \mathbb{R}^{m\times m}$. Write $W^{1/2}$ for its symmetric positive-semidefinite square root and put $\mathcal{J} = W^{1/2}J_F \in \mathbb{R}^{m\times d}$, so $G = \mathcal{J}^\top\mathcal{J}$. The probe is drawn independently of these fixed quantities.

Proposition 3 (Rademacher quadratic moments)**.** For independent Rademacher coordinates of $\xi \in \mathbb{R}^m$, let $a_\xi = \mathcal{J}^\top \xi$, $b = \mathcal{J} r$, and $q_\xi = (a_\xi^\top r)^2$. Then

$$\mathbb{E} q_\xi = \|b\|_2^2 = r^\top G r, \tag{20}$$

$$\mathrm{Var}(q_\xi) = 2\left(\|b\|_2^4 - \sum_{k=1}^m b_k^4\right). \tag{21}$$

The average $\widehat{q}_P = P^{-1}\sum_{p=1}^P q_{\xi_p}$ of P independent probes has the same mean and

$$\mathrm{Var}(\widehat{q}_P) = \frac{\mathrm{Var}(q_\xi)}{P} \le \frac{2}{P}(r^\top G r)^2.$$

Proof. Expand $q_\xi = \sum_k b_k^2 + 2\sum_{k<l} b_k b_l \xi_k \xi_l$. Each cross term has mean zero. In its squared centered expansion, independence and zero means eliminate all pairs of distinct unordered index pairs. Hence $\mathrm{Var}(q_\xi) = 4\sum_{k<l} b_k^2 b_l^2$, which is equation 21. Independence across probes gives the $1/P$ variance law. □

The same output-space probe supplies

$$\widehat{\tau}_\xi = \|a_\xi\|_2^2 = \xi^\top \mathcal{J}\mathcal{J}^\top \xi, \qquad \mathbb{E}\widehat{\tau}_\xi = \mathrm{Tr}(\mathcal{J}\mathcal{J}^\top) = \mathrm{Tr}(G). \tag{22}$$

Writing $C = \mathcal{J}\mathcal{J}^\top$, the same expansion yields $\mathrm{Var}(\widehat{\tau}_\xi) = 2\sum_{k\neq l} C_{kl}^2$. Each $a_\xi a_\xi^\top$ has rank at most one, and $\mathbb{E}[a_\xi a_\xi^\top] = G$, so for a fixed residual

$$\nabla_r q_\xi = 2a_\xi(a_\xi^\top r), \qquad \mathbb{E}_\xi[\nabla_r q_\xi] = 2Gr. \tag{23}$$

This is the precise connection between the illustrative full-quadratic correction in equation 15 and its stochastic rank-one implementation. With padding, replace $\mathcal{J}$ by $\mathcal{J}M$ and G by MGM throughout these moment identities. These are moments of the scalar quadratic and trace at fixed inputs. The parameter-gradient accuracy measured in Appendix E.3 additionally involves the student Jacobian; it is a separate estimator diagnostic.

C.2 Discrete transport, normalization, and selection

A discrete rollout defines a concrete training geometry. Let $\widehat{\Phi}_K$ be the fixed K-step teacher map in real arithmetic. At a differentiable point, its composition $\widehat{F}_K = g \circ \widehat{\Phi}_K$ has the pullback $\widehat{G}_K = J_{\widehat{F}_K}^\top W J_{\widehat{F}_K}$, and Proposition 3 applies to this map. For the default training configuration, $K = 2$ Heun steps. With $\delta = (1-t)/K$, times $s_j = t + j\delta$, and $y_0 = z_t$, each step is

$$k_{1,j} = v_{\bar{\theta}}(y_j, s_j, c), \quad k_{2,j} = v_{\bar{\theta}}(y_j + \delta k_{1,j}, s_j + \delta, c), \quad y_{j+1} = y_j + \tfrac{\delta}{2}(k_{1,j} + k_{2,j}).$$

Thus $\widehat{\Phi}_K(z_t) = y_K$ uses $2K$ teacher field evaluations. The terminal VJP differentiates both stages of every step with t, c and the parameters fixed; it differentiates the chosen numerical map. The smooth-reference result in Proposition 2 uses the exact flow, and the comparison below relates their pullbacks. The diagnostics in Appendix E.3 evaluate this numerical implementation at the training precision.

The discretized and exact-flow metrics differ through both state transport and the terminal readout Jacobian. At a fixed input, write

$$T = J_\Phi, \quad \widehat{T} = J_{\widehat{\Phi}_K}, \qquad B = J_g(\Phi)^\top W J_g(\Phi), \quad \widehat{B} = J_g(\widehat{\Phi}_K)^\top W J_g(\widehat{\Phi}_K).$$

Then $G = T^\top B T$, $\widehat{G}_K = \widehat{T}^\top \widehat{B}\widehat{T}$, and adding and subtracting $\widehat{T}^\top B \widehat{T}$ gives

$$\|\widehat{G}_K - G\| \le \|\widehat{T}\|^2 \|\widehat{B} - B\| + \|B\|(\|\widehat{T}\| + \|T\|)\|\widehat{T} - T\|. \tag{24}$$

This algebraic bound separates the changed terminal readout Jacobian from the changed flow Jacobian. The depth diagnostics compare the resulting geometries directly at the stated checkpoints.

Positive calibration and unbiasedness have distinct roles. For a positive normalizer Z fixed independently of the current probe, conditioning on $r, \mathcal{J}, Z$ yields

$$\mathbb{E}_\xi[q_\xi/Z \mid r, \mathcal{J}, Z] = r^\top G r/Z.$$

A current-probe denominator defines a different averaged metric. For example, with

$$\mathcal{J} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad r = (1,0)^\top, \qquad \mathbb{E}_\xi[q_\xi/\widehat{\tau}_\xi] = \frac{2}{5}, \quad \frac{\mathbb{E}_\xi q_\xi}{\mathbb{E}_\xi \widehat{\tau}_\xi} = \frac{2}{3},$$

equal probe signs give $(q_\xi, \widehat{\tau}_\xi) = (4, 5)$ and opposite signs give $(0, 1)$, each with probability $1/2$. In the implementation, seen timestep bins use a pre-update EMA; an unseen bin uses the synchronized current-batch trace mean. The positive floor guarantees $Z_i > 0$ in both cases. Detachment holds the realized denominator fixed during differentiation; probe independence determines whether its expectation is the fixed-scale quadratic above. The bin scale calibrates typical sensitivity within a timestep bin while retaining variation between states in that bin.

Detached student gradient. With the mask and normalizer from equation 11, define $J_{v,i} = \partial v_\theta(z_{t_i}, t_i, c_i)/\partial\theta$. For a selected sample, the detached auxiliary quadratic contributes

$$\nabla_\theta \ell_{\mathrm{aux},i} = \frac{2\rho}{Z_i}(a_{\xi,i}^\top M_i r_i) J_{v,i}^\top M_i a_{\xi,i}.$$

Thus the terminal adjoint supplies the residual-space direction and the student Jacobian maps its signed correction to parameters.

For selection indicator $s_i = \mathbf{1}[i \in \mathcal{S}]$, the actual mean weight is

$$\bar{A}(x) = d^{-1} I + \rho\, \mathbb{E}[s_i \widetilde{a}_{\xi,i} \widetilde{a}_{\xi,i}^\top / Z_i \mid x].$$

It preserves the population target under the joint endpoint-independence condition of Appendix B.1, including when a probe-dependent denominator changes its mean geometry. Under this condition, averaging selection and sketch randomness at fixed x, r gives the expected auxiliary residual correction

$$\mathbb{E}[\nabla_r \ell_{\mathrm{aux}} \mid x, r] = 2\big(\bar{A}(x) - d^{-1} I\big) r.$$

The per-sample parameter correction is obtained by applying $J_v^\top$; equation 11 then averages these contributions with $1/B$, so the auxiliary strength includes the selected fraction. The rotating quota and early-time window allocate the auxiliary computation; the Euclidean term retains every valid residual coordinate. Together, these statements characterize the geometry and the implemented training loss. The generation experiments establish its empirical FID effects at the specified solver, guidance, and continuation budget.

C.3 Readout and training configuration

Representation and parent models. The frozen motion codec is a MARDM Conv-AE with width 512, depth 3, temporal stride 2×2, and latent width 64, trained for 100 epochs with seed 3464. Text conditioning uses precomputed FP16 embeddings from a frozen OpenAI CLIP ViT-B/32 encoder initialized from MoRAE-DiT-XL. All arms share the codec, text encoder, evaluator, and eligible test set.

Each seed-specific Standard-FM parent is trained for 600 epochs on four ranks with batch 64 per rank. Parent training uses AdamW with $(\beta_1, \beta_2) = (0.9, 0.999)$, $\epsilon = 10^{-8}$, learning rate 2×10^{-4}, zero weight decay, 1,000-step linear warmup followed by cosine decay, uniform time, standard Gaussian noise, text dropout 0.1, global clipping at 1.0, and EMA decay 0.9999. Continuations initialize the student and frozen teacher from the parent's final EMA weights; the optimizer settings in the table apply to continuation training.

Physical features. The decoder first reverses latent standardization, reconstructs the normalized 263D motion representation, and reverses the dataset motion normalization. Joint positions $p_{f,j} \in \mathbb{R}^3$ are recovered from the first 67 motion coordinates, using the 22-joint HumanML3D skeleton. For frame index f and joint index j, define

$$v_{f,j} = p_{f,j} - p_{f-1,j}, \qquad a_{f,j} = v_{f,j} - v_{f-1,j}, \qquad \ell_{f,j} = \|p_{f,j} - p_{f,\pi(j)}\|_2.$$

Table 5: Shared HumanML3D representation and continuation settings. Fixed-update comparisons use 9,100 updates; equal-time studies use Appendix G. Endpoint-matched comparisons use native FP32 arithmetic; other continuations use the default BF16 forward computation.

Component	Setting
Motion codec / latent	Frozen MARDM Conv-AE; 263D, at most 196 frames / 49 tokens, width 64
Text condition	Frozen 512D OpenAI CLIP ViT-B/32 embeddings
Student	AdaLN DiT, width 512, depth 8, eight heads, 39.1M parameters
Parent pretraining	Seed-specific Standard FM; 600 epochs, four ranks $\times$ 64, load epoch-599 EMA
Continuation	100 epochs, 9,100 updates; four ranks, batch 64/rank
Optimizer	AdamW, zero weight decay, cosine schedule; LR 2×10^{-5}
Time / noise	Uniform $t \in [0, 1]$ / standard Gaussian on valid coordinates
Text dropout / teacher CFG	0.1 / 1.0, using the student's condition
Gradient clip / student EMA	Global norm 1.0 / decay 0.9999
Production teacher map	Two Heun steps over the remaining interval; BF16 forward by default, native FP32 for Endpoint-matched comparisons
Metric sketch	One Rademacher probe; detached terminal VJP
Selection	Rotating 25% quota intersected with $t \leq 0.35$
Auxiliary / trace	$\rho = 0.1$ / 20-bin EMA, decay 0.99, floor 10^{-8}
Evaluation	Frozen HumanML3D evaluator; 4,640 common eligible entries, four replications per training seed

Here $\pi(j)$ is the parent of a non-root joint. These are per-frame differences, with no FPS multiplier. Each difference operator prepends a zero frame. During training, positions and bones require a valid current frame, velocities and foot features require two consecutive valid frames, and accelerations require three; all other entries are zero. The zero-based parent array is

$$(-1, 0, 0, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 9, 9, 12, 13, 14, 16, 17, 18, 19).$$

For foot joints $\mathcal{F} = \{10, 11\}$, let p^y denote height along coordinate 1. The soft-contact and foot-displacement features are

$$c_{f,j} = \mathrm{sigmoid}\left(\frac{0.08 - p^y_{f,j}}{0.02}\right), \qquad s_{f,j} = c_{f,j} v_{f,j}, \quad j \in \mathcal{F}.$$

Contact weights are differentiable functions of the decoded positions.

Scaling and output metric. For each group $b \in \{p, v, a, \ell, s\}$, pool its scalar entries over the training motions and compute

$$\sigma_b = \sqrt{\max\{\mathbb{E}[f_b^2] - (\mathbb{E}[f_b])^2, 10^{-12}\}}, \qquad H(x) = \mathrm{concat}_b \, \mathrm{vec}(f_b(x)/\sigma_b).$$

Statistics use the full training motions and the zero-prepended difference operators, and are frozen for all continuations. Scaling uses one scalar standard deviation per group, without mean subtraction or division by group dimension. We set $W = I$ after scaling; equivalently, the raw feature blocks have diagonal weights σ_b^{-2}. For the fixed 196-frame tensor the five groups have respectively 66, 66, 66, 21, and 6 coordinates per frame, so the concatenated readout has $m = 44{,}100$ coordinates, including masked zeros. Its statistics and definitions use no test-evaluator embeddings.

Teacher and optimization. Student and teacher load the EMA weights of the same seed-specific parent at epoch 599. The teacher stays in evaluation mode with frozen parameters; its rollout is differentiated only with respect to the selected input. It receives the student's text condition after 0.1-probability dropout, without classifier-free amplification (teacher CFG 1). Noise is standard Gaussian and t is uniform on $[0, 1]$; padded noise, latents, and residual coordinates are zero. The

25% quota rotates through batch positions with a persistent cursor, selecting 16 of the 64 local batch positions on each rank, then intersects with $t \leq 0.35$. The cursor starts at zero and advances by 16 modulo 64 on every batch, before the time-window intersection; positions rejected by that window are not refilled. Under uniform training times, the expected selected fraction is $0.25 \times 0.35 = 8.75\%$. Every sample retains the FM loss. AdamW uses $(\beta_1, \beta_2) = (0.9, 0.999)$, $\epsilon = 10^{-8}$, zero weight decay, no warmup, and global gradient clipping at 1.0. Fixed-update runs decay the learning rate from 2×10^{-5} with a cosine schedule. Student EMA decay is 0.9999; reported generation uses the final raw student weights. The equal-time schedule is specified in Appendix G.

One-probe training algorithm. Initialize the trace bins as unseen once before continuation training. For each training batch:

1. Draw a training batch with its noise, paired endpoints, conditions, times, and length masks; construct z_t, u_t, r_θ on the linear path. Apply the endpoint-independent rotating quota and intersect it with $t \leq 0.35$.
2. On the selected states, take two frozen-teacher Heun steps to $\bar{z}_1$. Evaluate $H(D(\bar{z}_1))$, draw one output Rademacher probe, and compute the terminal VJP a_ξ through both stages of both steps (four teacher field evaluations).
3. Set padded adjoint coordinates to zero and form trace samples $\|M_i a_{\xi,i}\|_2^2$ in FP32. Assign bin $b(t) = \min\{\lfloor 20t \rfloor, 19\}$ and synchronize each bin's sum and count across ranks.
4. For each selected sample, use the bin's pre-update EMA if seen, or the synchronized current mean if unseen; clamp the denominator at 10^{-8}. After fixing these denominators, initialize each newly observed bin to its current mean and update previously seen, nonempty bins as $z_b \leftarrow 0.99 z_b + 0.01 \widehat{\tau}_b$. Here $\widehat{\tau}_b$ is the synchronized mean trace in bin b. Empty bins retain their state. All ranks participate in the reduction, including ranks with no selected samples. Joint-control studies share one trace reduction and one EMA update per step.
5. Detach the adjoint and normalizer. Use $M_i r_i$ in the quadratic and $\|M_i r_i\|^2/d_i$ in the base term of equation 11; update the student only.

For the multi-probe ablations, average the individual squared projections and trace samples over probes; the squared projection of an averaged adjoint is a different estimator. Use the averaged trace samples in the same single bin update. The default remains one probe.

The unnormalized randomized quadratic is unbiased. In the unseen-bin fallback the current probe contributes to the trace denominator and defines the positive averaged metric in Appendix C.2. The weighted-regression result requires the metric, selection, and normalizer to be jointly conditionally independent of the paired endpoint, including the mask information in the conditioning variables.

C.4 Objective and intervention definitions

Write $\mathcal{L}_0 = B^{-1} \sum_i \|M_i r_i\|^2/d_i$. Let λ_{geo} and λ_{LPL} denote the corresponding scalar weights, and define

$$Q(a, Z) = \frac{\lambda_{\text{geo}}}{B} \sum_{i \in \mathcal{S}} \frac{(\text{stopgrad}(a_i)^\top M_i r_i)^2}{\text{stopgrad}(Z_i)}.$$

At the actual teacher endpoint, let $g_i = J_{HD}(\bar{z}_{1,i})^\top W^{1/2} \xi_i$ and $a_{F,i} = J_{\Phi,i}^\top g_i$. All quadratic rows below use the same stated mask and selection rule. Adjoints and their norms are restricted to valid latent coordinates. Complete execution settings are recorded in the frozen study configurations. The **Endpoint-independent** objective is trained independently with its own trace state, optimizer, and update trajectory, using the endpoint adjoint without remaining-flow transport.

For LPL-style, the clean prediction is $\hat{z}_1 = z_t + (1-t) v_\theta(z_t, t, c)$. Both predicted and paired clean latents are decoded through the same frozen autoencoder and physical readout, using the same frame mask. The objective averages their squared feature discrepancy over readout coordinates and selected samples, with $(1-t)$ entering through the clean prediction. This supplies a decoder-feature clean-prediction comparator under the shared training and evaluation protocol. Endpoint-independent applies the endpoint quadratic with its own trace, optimizer, EMA, and update trajectory. It retains the common teacher forward rollout while providing an independently trained decoder-geometry baseline.

Table 6: Definitions of the compared objectives and controls. "Own trace" applies the same bin/EMA rule to that model's adjoint samples and maintains model-specific running statistics.

Objective / control	Loss or intervention	Trace / optimization source
Standard FM	$\mathcal{L}_0$	No geometry or trace state; own updates
IMPACTFLOW	$\mathcal{L}_0 + Q(a_F, Z_F)$; VJP through J_Φ	Own trace and updates
Token-sign	$\mathcal{L}_0 + Q(Sa_F, Z_F)$; seeded token signs after transported VJP	Shared IMPACTFLOW trace; own updates
Token-matched	Original FM base gradient and token-sign auxiliary gradient, followed by the matching in Appendix F	Shared IMPACTFLOW trace and global budgets
Endpoint-matched	$\mathcal{L}_0 + Q(\|a_F\|g/\|g\|, Z_F)$, then component-ratio and update matching; endpoint adjoint g before transport	Shared IMPACTFLOW trace, adjoint norms, and global budgets
Endpoint-independent	$\mathcal{L}_0 + Q(g, Z_D)$ at the actual teacher endpoint; endpoint adjoint g before transport	Own trace, optimizer, EMA, and updates
LPL-style	$\mathcal{L}_0 + \frac{\lambda_{\mathrm{LPL}}}{Bm}\sum_{i\in\mathcal{S}} \|\Psi(\hat{z}_{1,i}) - \Psi(z_{1,i})\|^2$, $\Psi = H \circ D$	Same selection; feature mean-square; own updates

D EXPERIMENTAL DESIGN AND STATISTICAL INFERENCE

D.1 PAIRED COMPARISONS AND SHARED MODELS

All comparisons use HumanML3D. The five-seed studies use parents trained with seeds 17, 43, 101, 137, and 211, with separate continuations for each objective comparison or directional intervention. The ten-seed fixed-deployment confirmation uses a new parent cohort disjoint from these five seeds. The ten-seed equal-time comparison starts from the same new parents, but trains separate Standard-FM and IMPACTFLOW continuations. Effects are paired within each comparison and reported separately by study.

Table 7: Experimental comparisons, training budgets, and selection rules. Fixed-update comparisons use Euler NFE8 / CFG 4.5. The shared time cap is 1,919 seconds on four GPUs per continuation.

Comparison	Seeds and budget	Deployment and inferential role
Standard / LPL-style	Five; 9,100 updates	Fixed-deployment objective comparisons
Token-sign	Five; 9,100 updates	Fixed deployment; prespecified directional contrast
Token-matched	Five; 9,100 updates	Fixed deployment; prespecified scalar-matched contrast
Endpoint-matched	Five; 9,100 updates	Fixed deployment; prespecified transport contrast
Transport confirmation	Ten new; 9,100 updates	Fixed deployment; prespecified absolute and relative effects
Equal-time common deployment	Ten; shared cap	Paired effect at the Standard-selected deployment
Independently selected objectives	Five; shared cap	Separate validation-selected deployments for Standard, LPL-style, and IMPACTFLOW
Endpoint-independent	Five; shared cap	Separate validation-selected deployments
Estimator / readout ablations	Three; shared cap	Means at two fixed deployments

Pairing and numerical precision. Within each jointly trained directional comparison, arms share batches, FM noise and targets, conditioning dropout, selection, the teacher sketch, and pre-update trace normalization, while maintaining separate student, optimizer, and EMA states. Token-matched computes base and auxiliary parameter gradients separately and uses the same FP32 Adam-update implementation in both arms. Its IMPACTFLOW reference is independently trained from the Token-sign reference; their mean FIDs, 0.221535 and 0.221543, both round to 0.22154. The five- and ten-seed Endpoint-matched comparisons use native FP32 throughout, including teacher/readout forward passes and global gradients. Independent objective comparisons maintain their own optimization and, where applicable, trace states. The Standard and LPL-style fixed-update comparisons share the same IMPACTFLOW reference.

Evaluation uses the same 4,640 eligible test motions. Endpoint-matched and Endpoint-independent comparisons explicitly pair conditions and initial noise. The other five-seed objective and token-control studies use common, disjoint replication seed blocks with stride 1,048,576. Each generation comparison averages four full-test replications per arm and training seed.

Shared ablation reference. Estimator and readout ablations use seeds 17, 43, and 101. Their default one-probe, two-step, complete-readout setting uses the corresponding IMPACTFLOW models from the Endpoint-independent comparison; other settings are separate continuations. The fixed-checkpoint gradient diagnostic uses these default final models and their parent EMAs, holding the default model's final trace normalizer fixed. The ablations therefore share a reference, and the diagnostic involves no additional training.

D.2 STATISTICAL ESTIMANDS AND INFERENCE

Let f_s and c_s be the IMPACTFLOW and comparator FIDs after averaging the four replications for training seed s. Each replication evaluates the complete eligible test set; embeddings are not pooled across replications. The training seed is the inferential unit. Arm means average seed FIDs, while the relative effect averages seedwise percentages:

$$d_s = 100\frac{c_s - f_s}{c_s}, \qquad \bar{d} = \frac{1}{n}\sum_s d_s, \qquad \mathrm{CI}_{95\%} = \bar{d} \pm t_{0.975,n-1}\frac{s_d}{\sqrt{n}},$$

where s_d is the sample standard deviation. This relative estimand differs from taking the percentage change between arm means. Absolute-effect intervals use the same paired-seed construction, with the sign stated for each contrast. Intervals are pointwise Student-t summaries with four degrees of freedom for five seeds and nine for ten seeds.

For each five-seed directional comparison, the prespecified alternative is lower FID for IMPACTFLOW and the exact statistic is $T(d) = \bar{d}$. Under joint invariance to seedwise sign flips conditional on effect magnitudes, the one-sided test is

$$p = 2^{-n}\sum_{\epsilon\in\{-1,1\}^n} \mathbf{1}[T(\epsilon \odot d) \geq T(d)]\,.$$

All three comparisons have five favorable seed effects, giving one-sided $p = 1/32 = 0.03125$ and two-sided $p = 2/32 = 0.0625$. These tests are reported separately, without cross-comparison multiplicity adjustment. The t intervals and exact tests rely on their respective Student-t and sign-symmetry assumptions.

The fixed-update objective family includes Standard FM, Token-block, Diagonal, Shuffled, nonlinear terminal (100 epochs), and nonlinear terminal with nominal compute matching (63 epochs). These objective comparisons use paired seed effects, with Holm adjustment across the fixed-update family. The LPL-style result provides an additional objective comparison. The independently selected equal-time comparisons choose each objective's deployment by its five-seed validation mean before test evaluation. The ten-seed common-deployment analysis uses the Standard-selected operating point detailed below. Geometry diagnostics use state-cluster intervals to preserve the state-level unit of analysis.

D.3 FIXED-DEPLOYMENT TRANSPORT CONFIRMATION

The confirmatory comparison pairs IMPACTFLOW with Endpoint-matched at Euler NFE8 / CFG 4.5. Both arms receive 9,100 updates on four ranks with batch 64 per rank. Ten new training seeds and four test replications per arm yield ten paired FID observations, analyzed separately from the five-seed directional study.

Define the absolute improvement $\delta_s = c_s - f_s$, positive when IMPACTFLOW has lower FID. The primary two-sided exact test enumerates all $2^{10} = 1024$ sign assignments under seedwise sign symmetry:

$$p_{\text{two}} = 2^{-10} \sum_{\epsilon \in \{-1,1\}^{10}} \mathbf{1}\left[\left|\frac{1}{10}\sum_s \epsilon_s \delta_s\right| \geq |\bar{\delta}|\right].$$

The mean absolute improvement is 0.03686 with 95% CI $[0.01387, 0.05984]$ and two-sided $p = 0.009765625$. The separately prespecified relative effect is $\bar{d} = 10^{-1}\sum_s 100\delta_s/c_s = 13.23\%$, with 95% CI $[5.24, 21.22]$. Both intervals use the paired-seed construction with nine degrees of freedom. Nine of ten seed effects favor IMPACTFLOW.

Table 8: Seedwise FID in the fixed-Euler confirmation. Each arm averages four full-test replications; the difference is Endpoint-matched minus IMPACTFLOW. All ten seeds enter inference, with nine positive differences. Displayed values are rounded independently.

Training seed	Endpoint-matched	IMPACTFLOW	Difference
73112611	0.293532	0.246785	0.046747
483261482	0.253406	0.282174	-0.028768
1305241733	0.263598	0.232224	0.031374
1458892312	0.233318	0.226323	0.006995
1487101345	0.286780	0.216626	0.070154
1518631543	0.221793	0.193905	0.027888
1732070980	0.314012	0.235002	0.079009
1910842913	0.286201	0.234193	0.052008
2021812784	0.262759	0.202170	0.060590
2066485671	0.240625	0.218064	0.022561
Mean	0.265602	0.228747	0.036856

The physical readout defines the training geometry; the independent HumanML3D evaluator defines the distributional FID endpoint.

D.4 EQUAL-TIME EFFECTIVENESS AT A COMMON DEPLOYMENT

Standard FM and IMPACTFLOW are separately continued from the ten parents used in the fixed-deployment confirmation. Each continuation has the same 1,919-second four-A100 cap, including all continuation overhead (Appendix G). Parent and representation training precede this budget; deployment evaluation is accounted separately.

The deployment scan crosses Euler, ordinary Heun, midpoint, and AB2 with CFG $\{3, 4.5, 6\}$ at NFE8 for both objectives. Selection uses only Standard-FM outcomes: midpoint / CFG 3 has the lowest ten-seed mean FID among the twelve settings and is then used for both objectives. Selection and comparison use the same training seeds and full-test replications. The reported estimand conditions on this scan-selected common operating point, which is held fixed across both objectives. Four replications per arm and seed yield ten paired observations.

With c_s denoting Standard FM and f_s denoting IMPACTFLOW, define $\delta_s = c_s - f_s$ and $d_s = 100\delta_s/c_s$. The nominal 95% paired-seed t intervals hold the selected deployment fixed. The two-sided sign-flip test uses $|\bar{\delta}|$ and the same 1,024-assignment construction as the fixed-deployment confirmation.

At midpoint NFE8 / CFG 3, Standard FM has mean FID 0.100259 and IMPACTFLOW has mean FID 0.088898. The paired absolute reduction is 0.011361 $[-0.000921, 0.023643]$; the mean seedwise

Table 9: Authoritative configurations for the reported experimental roles. Budget and unit entries state the cohort used for each reported comparison.

Configuration	Comparison and scientific role	Cohort, budget, and unit
`confirmatory_protocol/protocol.json`	Ten-seed fixed-deployment transport confirmation and Standard-selected common-deployment equal-time comparison.	P0: 9,100 updates per arm; P2: 1,919-second complete-exit cap; training seed.
`equal_time_protocol.json`	Five-seed independently selected Standard FM, LPL-style, and IMPACTFLOW continuations.	Five paired seeds; 1,919-second complete-exit cap; training seed.
`geometry_diagnostics/independent_evaluator.json`, `geometry_diagnostics/morae_physical.json`, and `geometry_control_protocol.json`	Frozen early-time evaluator and physical state sets, followed by the production FM-residual geometry diagnostic.	Sixteen or 32 fixed states for the early-time diagnostics; five paired seeds at 100 epochs and motion/noise-pair clusters for the production diagnostic.
`directional_control_protocol.json`	Coupled IMPACTFLOW versus Token-sign directional intervention.	Five paired seeds; 100 epochs; training seed.
`strength_control_protocol.json`	IMPACTFLOW versus scalar-matched Token-matched attribution.	Five paired seeds; 100 epochs; training seed.
`lpl_baseline.json`	Fresh fixed-update LPL-style continuations compared with the shared fresh Full reference at the frozen deployments.	Five paired seeds; 9,100 updates; four evaluation replications; training seed.
`ablation_protocol.json` and `production_evaluation_protocol.json`	Endpoint-independent, estimator, and physical-readout studies.	Five-seed main cohort or three-seed ablations; elapsed-time or fixed-update budget; training seed.
`gradient_diagnostic.json`	Fixed-checkpoint probe-gradient diagnostic.	Three seeds, parent/final checkpoints, four held-out batches; seed–checkpoint pair.

relative reduction is 10.87% $[-1.86, 23.60]$. Seven of ten effects favor IMPACTFLOW. The two-sided exact sign-flip tests give $p = 0.0703125$ for the absolute effect and $p = 0.08984375$ for the relative effect. Holm adjustment over the twelve absolute deployment contrasts gives $p = 0.2109375$ for midpoint / CFG 3. These paired results quantify the lower observed mean FID at the shared Standard-selected deployment.

D.5 STUDY CONFIGURATIONS AND SCOPE

The fixed-deployment confirmation tests transported direction with matched scalar training strength. The equal-time comparison tests the complete objective when geometry and student optimization share a compute budget. Geometry diagnostics measure sensitivity to controlled perturbations; directional interventions test whether that information affects learning; estimator and readout ablations examine its computational and task-specific choices. The paired seed results and confidence intervals retain their study-specific units of analysis. Shared parents and reused ablation references are identified above, and the complementary results are synthesized according to their distinct experimental roles.

Each comparison is identified by its frozen configuration:

E GEOMETRY AND GRADIENT DIAGNOSTICS

E.1 EARLY-TIME GEOMETRY DIAGNOSTICS

Table 10: Geometry diagnostics. The physical and independent-evaluator studies use separate frozen generators and state sets. Both use 64 directions and 64 adjoint probes per state, eight Heun metric steps, and a 16-step reference rollout.

Diagnostic	Protocol
Physical geometry	32 early-time states of frozen MoRAE, CFG 8.5. Compare pullback scores with centered finite-difference physical-readout sensitivity. Report the geometric mean of statewise P90/P10 ratios and a state-cluster bootstrap interval.
Independent evaluator	16 states from the frozen motion-FM parent: eight fixed prompts at each of $t = 0.1$ and 0.2, CFG 4.5, seeds 1700–1715. Score directions with the physical pullback; measure damage in the frozen HumanML3D evaluator embedding, excluded from training. Average within-state Spearman correlations and paired IMPACTFLOW–Decoder-only geometry differences.
Perturbations / intervals	Gaussian directions normalized to unit norm on valid latent coordinates; centered perturbation size $h = 3 \times 10^{-4}\|z_t\|_2$. Percentile 95% intervals use 10,000 state-cluster bootstrap resamples, RNG seed 3407.

The resampling unit is the state on the specified frozen generator and prompt set. Each resampled state retains all its perturbation directions and comparator scores. Paired advantages bootstrap the within-state correlation differences; the 64 directions per state measure that state's association. These intervals summarize statewise variation conditional on the fixed generator and prompts.

For a unit direction q, the centered directional response is

$$d(q) = \frac{\|F_t^{\mathrm{ref}}(z_t + hq) - F_t^{\mathrm{ref}}(z_t - hq)\|_2^2}{4h^2}.$$

Using the physical readout, $d(q)$ measures physical-readout sensitivity; using independent evaluator embeddings, it measures evaluator damage. IMPACTFLOW scores use $P^{-1}\sum_{p=1}^{P}(a_p^\top q)^2$ with $P = 64$. The Decoder-only geometry diagnostic evaluates the physical readout Jacobian at the same terminal state, omitting remaining-flow transport. Token-block uses $P^{-1}\sum_p \sum_k (a_{p,k}^\top q_k)^2$, where k indexes latent tokens; it removes cross-token terms from the IMPACTFLOW score.

The physical diagnostic's geometric-mean P90/P10 ratio is 4.81 $[3.67, 6.47]$. The leading eight sensitivity directions capture 57.7% and 53.3% at $t = 0.1$ and $t = 0.2$. In the separate evaluator study, transported geometry has Spearman 0.336 $[0.216, 0.461]$ and a paired advantage of 0.299 $[0.189, 0.413]$ over Decoder-only geometry. These early-time measurements motivate the computation window used in training. Their state sets and diagnostic guidance scales are separate from the training batches and teacher CFG 1.

Figure 3(a) averages correlations computed within states; panel (b) shows paired correlation gains for those same states. Token-block removes cross-token terms from the diagnostic quadratic and tests whether those terms add prediction. Remaining-flow attribution comes from the IMPACTFLOW–Decoder-only geometry comparison at the common terminal state. Token-sign is a separate training intervention that preserves within-token blocks and changes cross-token signs.

E.2 SENSITIVITY ALONG FM RESIDUAL DIRECTIONS

Along actual FM residual directions, the teacher metric has a positive rank association with independent evaluator damage. The diagnostic evaluates 640 states from 80 held-out motion/noise pairs, with eight states per pair, along paired FM chord residuals under the training mask. The metric uses two Heun steps, BF16, and teacher CFG 1; terminal evaluator damage uses deployment CFG 4.5. The 32- and 64-step reference rollouts cover the entire remaining interval; the estimand is the rank association

across all 640 states. The one-probe score uses the IMPACTFLOW remaining-flow adjoint with the training mask and no student update.

Table 11: Rank associations with independent evaluator damage along FM residual directions, after displacement-magnitude normalization, at amplitude multiplier 1. Intervals cluster the held-out motion/noise pairs. The one- and 64-probe estimates use the same two-step BF16 metric.

Direction population / student	One probe	64 probes
FM chord / Standard	$0.217\ [0.094, 0.340]$	$0.206\ [0.080, 0.332]$
FM chord / IMPACTFLOW	$0.297\ [0.186, 0.409]$	$0.243\ [0.089, 0.396]$

These associations use all evaluated motion/noise pairs. The comparison with Decoder-only geometry and Token-block in Appendix E.1 separately tests the contribution of transport and cross-token terms; the next diagnostic measures the accuracy of the resulting auxiliary parameter gradient.

E.3 FIXED-CHECKPOINT PROBE-GRADIENT DIAGNOSTIC

Table 12: Protocol for the probe-count and teacher-depth gradient diagnostic. Entries report mean gradient agreement with the trace normalizer held fixed.

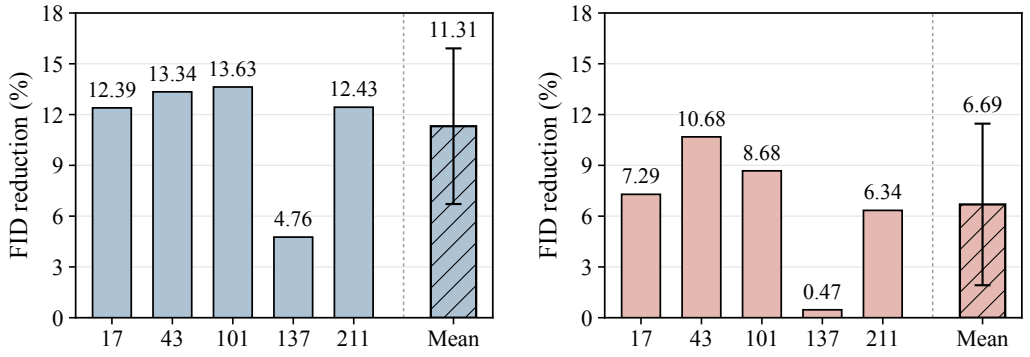
Field	Setting
Checkpoints	Final equal-time IMPACTFLOW models for seeds 17, 43, and 101; the broader diagnostic also measures their parent EMA models.
Fixed states	First 1,024 validation motions per seed, split into four global batches of 256 across four ranks. Hold states, FM targets, conditions, times, masks, selection, and student forward fixed.
Draws / reference	16 draws per batch. The one-probe draw is the first member of each four-probe block. An independent 256-probe parameter gradient is the shared within-depth reference for each two- or eight-step teacher; the eight-step reference also supplies the common cross-depth comparison. Two disjoint 128-probe halves assess reference self-agreement.
Arithmetic / trace	BF16 teacher/readout forward; FP32 parameters and globally averaged gradients. Hold each seed's final default IMPACTFLOW trace normalizer fixed across probe counts and checkpoints.
Aggregation	Compute cosine of each auxiliary-gradient draw to its 256-probe reference, average draws and the four batches within seed, then report the mean of the three seed-level averages.

For the final IMPACTFLOW checkpoints and two Heun teacher steps, the resulting cosine rises from 0.484 with one probe to 0.781 with four probes. Against the common eight-step reference, the same two-step estimates have cosines 0.334 and 0.548. The eight-step estimates have cosines 0.574 and 0.832 with one and four probes. All low-budget probes are independent of the reference stream. An independent 256-probe stream supplies the reference geometry, and two disjoint 128-probe halves measure its self-agreement. The diagnostic compares estimators at fixed checkpoints with a shared normalizer; the continuation studies report generation outcomes under evolving training normalization.

F DIRECTIONAL ATTRIBUTION CONTROLS

F.1 CROSS-TOKEN ORIENTATION AND SCALAR-STRENGTH MATCHING

Figure 4(a,b) reports separate five-seed Token-sign and Token-matched comparisons from common parents. Each has its own independently trained IMPACTFLOW reference. Token-sign shares that reference's one-probe sketch and applies a seed-specific sign pattern over temporal tokens, preserving



(a) Token-sign control

(b) Token-matched control

Figure 4: Seedwise relative FID reductions of IMPACTFLOW versus (a) Token-sign and (b) Token-matched at Euler NFE8 / CFG 4.5. Positive values favor IMPACTFLOW. Solid bars correspond to training seeds labeled on the horizontal axis; hatched bars average seedwise reductions, with 95% paired-seed t intervals. Each panel is a separate five-seed study with its own IMPACTFLOW reference.

the realized diagonal, trace, rank, eigenvalues, and within-token blocks. The seeded pattern is shared throughout each paired run.

Token-matched retains the base-gradient and token-sign auxiliary-gradient directions while matching four scalar quantities from its paired IMPACTFLOW reference: total-gradient norm, auxiliary/base gradient-norm ratio, clipping coefficient, and realized FP32 Adam update length. This construction isolates directional effects under matched scalar training strength.

Writing b_C, a_C for those globally averaged base and ρ-weighted auxiliary parameter gradients, the intervention is

$$\widetilde{b}_C = \frac{\|b_F\|}{\|b_C\|} b_C, \qquad \widetilde{a}_C = \frac{\|a_F\|}{\|a_C\|} a_C, \qquad \widehat{g}_C = \frac{\|b_F + a_F\|}{\|\widetilde{b}_C + \widetilde{a}_C\|} (\widetilde{b}_C + \widetilde{a}_C).$$

It then applies the IMPACTFLOW clipping coefficient and rescales its own Adam proposal to the realized IMPACTFLOW FP32 update length. Each arm retains its own optimizer state, while the common final gradient scale preserves the component ratio.

Table 13: Paired directional-control results at Euler NFE8. Entries under seeds 17–211 are paired IMPACTFLOW relative FID reductions (%) versus the listed control. The IMPACTFLOW and control FIDs are 0.22154 and 0.24978 for Token-sign, and 0.22154 and 0.23759 for Token-matched. Each row has its own one-sided exact $p = 0.03125$ and two-sided $p = 0.0625$; the two tests are reported separately.

Control	17	43	101	137	211	Mean [95% CI]
Token-sign	12.39	13.34	13.63	4.76	12.43	11.31 $[6.72, 15.91]$
Token-matched	7.29	10.68	8.68	0.47	6.34	6.69 $[1.92, 11.46]$

For Token-matched, global gradient-norm and component-ratio relative errors are below 6.04×10^{-8} and 1.14×10^{-7}, and clipping coefficients agree exactly. The cumulative update-norm discrepancy remains below 3.0×10^{-7} of the IMPACTFLOW update-path length. Parameter direction, coordinatewise preconditioning, layerwise norms, and gradient covariance remain free, preserving the directional contrast of interest.

F.2 REMAINING-FLOW TRANSPORT AT A COMMON TERMINAL STATE

The five-seed transport comparison trains IMPACTFLOW, Endpoint-matched, and a total-gradient-matched endpoint control for 9,100 updates from each parent. The ten-seed confirmation repeats the

IMPACTFLOW versus Endpoint-matched contrast with new seed-specific parents. The arms share the frozen-teacher endpoint, output probe, mask, selection, and IMPACTFLOW trace normalizer. At the endpoint, let $g = J_{HD}(\bar{z}_1)^\top W^{1/2}\xi$. IMPACTFLOW uses $J_\Phi^\top g$; each control uses g rescaled to match the per-example norm of the IMPACTFLOW adjoint, matching the realized rank-one trace and spectrum before parameter-gradient control.

For globally averaged base and auxiliary gradients b and a, the ratio-matched control forms

$$\widetilde{b}_C = \frac{\|b_F\|}{\|b_C\|} b_C, \qquad \widetilde{a}_C = \frac{\|a_F\|}{\|a_C\|} a_C,$$

then scales their sum to the IMPACTFLOW total-gradient norm. It shares the IMPACTFLOW clipping coefficient and matches the realized FP32 Adam update length, while retaining the control components' directions. The aggregate relative update-length discrepancy is at most 2.957×10^{-7}. The matched scalar budgets isolate remaining-flow direction, while the equal-time study evaluates an independently trained decoder-geometry objective.

Table 14: Prespecified Euler NFE8 / CFG 4.5 IMPACTFLOW-versus-Endpoint-matched result. The first two columns report arm-mean FIDs; seed columns report paired IMPACTFLOW relative FID reductions (%) versus Endpoint-matched.

Control mean	IMPACTFLOW mean	17	43	101	137	211	Mean reduction
0.26185	0.20735	19.90	9.83	29.01	18.50	26.76	20.80

The mean seedwise relative FID reduction is 20.80% $[11.40, 30.20]$, and the prespecified one-sided exact sign-flip test gives $p = 0.03125$ (two-sided $p = 0.0625$). The absolute IMPACTFLOW–control difference is -0.05450 $[-0.08321, -0.02579]$.

G COMPUTE ALLOCATION AND READOUT CHOICES

G.1 CONTINUATION-TIME BUDGET

All equal-time comparisons and the estimator/readout ablations use a 1,919-second cap on four NVIDIA A100-SXM4-80GB GPUs, equivalent to at most 2.132 GPUh per continuation. The budget includes the complete continuation process, including geometry estimation, student updates, and startup and finalization overhead. Shared parent and representation training precede the cap; validation and test evaluation are accounted separately.

A time-normalized cosine schedule gives each objective access to its available student updates. Let A be elapsed time measured from before launch, S the first-update time, $B = 1919$ the cap, and $R = 60$ seconds a reserve for completion. The learning rate is

$$\eta(A) = 10^{-5}\left[1 + \cos\left(\pi \ \mathrm{clip}\left(\frac{A - S}{B - R - S}, 0, 1\right)\right)\right].$$

Training stops synchronously while reserving R plus twice the longest observed iteration. The Standard-FM, LPL-style, and IMPACTFLOW comparison uses the same four devices and stays within the shared cap.

Across the five equal-time training seeds, the independently selected systems obtain mean test FID 0.12219 for Standard FM, 0.12615 for LPL-style, and **0.08538** for IMPACTFLOW. Thus, IMPACTFLOW reduces FID by 30.13% relative to Standard FM and by 32.32% relative to LPL-style.

G.2 INDEPENDENT ENDPOINT-GEOMETRY OBJECTIVE

Endpoint-independent removes remaining-flow transport while retaining the common teacher forward rollout (Table 6). It and IMPACTFLOW use separate continuations, each with its own trace, optimizer, and EMA, under the same cap and validation menu. IMPACTFLOW receives 9,383.6 student updates on average, compared with 19,760.0 for Endpoint-independent.

Validation selects midpoint / CFG 3 for IMPACTFLOW and AB2 / CFG 4.5 for Endpoint-independent. At these separately selected deployments, test FID is 0.08422 versus 0.13095, with mean seedwise relative reduction 35.74% $[13.99, 57.49]$. This measures the independently selected systems, whereas the main comparison holds deployment fixed across objectives.

G.3 READOUT ABLATION

The complete physical readout achieves the lowest mean FID among the three tested readouts at both deployments in Table 15. These continuations use seeds 17, 43, and 101, one probe, two teacher Heun steps, $\rho = 0.1$, and the common 1,919-second cap. Each setting maintains its own trace EMA. The complete-readout reference is the same three default models used in Table 4, as indexed in Appendix D.1.

Table 15: Readout ablation at equal continuation time. Each entry is the mean FID over three training seeds at NFE8.

Readout	Euler / CFG 4.5	Midpoint / CFG 3
Position only	0.26111	0.11363
Complete physical readout	**0.21148**	**0.08450**
Without contact	0.23425	0.09412

Changing the readout changes the output geometry and therefore the residual directions penalized by the auxiliary loss. Temporal differences weight relative-frame changes, skeletal features couple joints, and contact-weighted displacement emphasizes support. The complete-versus-position comparison changes several feature blocks together; the no-contact row removes only the contact feature from the complete readout.

H EXTENDED RELATED WORK

This section expands the three themes in Section 2. The organizing question is how information about a completed output can shape the learning of its generating process.

H.1 LOCAL SUPERVISION, LEARNING PRIORITIES, AND EFFICIENT GENERATION

Score matching estimates a distribution through local derivatives of its log density (Hyvärinen, 2005); the connection to denoising autoencoders links this perspective to learning from corrupted observations (Vincent, 2011). Diffusion probabilistic models turn generation into a sequence of reverse transitions (Sohl-Dickstein et al., 2015; Ho et al., 2020), and score-based stochastic differential equations unify continuous-time versions of this construction (Song et al., 2021). Flow matching learns vector fields through conditional path supervision (Lipman et al., 2023), while rectified flow studies transport with straightened trajectories (Liu et al., 2023). These foundations explain how local regression can define a generative process. Our question concerns how finite model and optimization resources should be allocated across residual directions.

Several lines of work already redistribute learning effort or alter the predictions learned for efficient sampling. Perception-prioritized training changes the weights of denoising tasks at different noise levels to emphasize perceptually relevant learning (Choi et al., 2022); Min-SNR weighting addresses conflicting optimization demands across timesteps (Hang et al., 2023). Both provide scalar timestep-dependent emphasis; IMPACTFLOW adds residual-direction resolution at each state. Progressive distillation compresses a teacher's sampling procedure into successively fewer steps (Salimans & Ho, 2022). Consistency models learn mappings that agree along a generative trajectory (Song et al., 2023); shortcut models condition predictions on step size and enforce consistency across step lengths (Frans et al., 2025); mean flows learn average velocities over time intervals (Geng et al., 2025). These few-step formulations adapt the prediction problem to efficient generation, whereas IMPACTFLOW studies how terminal effects can prioritize errors in the original conditional-FM problem.

Sampling algorithms offer another complementary route to efficiency. PNDM and DPM-Solver exploit structure in diffusion dynamics to improve numerical integration (Liu et al., 2022; Lu et al.,

2022), and STORK accelerates sampling through stabilized Runge–Kutta integration with Taylor-approximated internal stages (Tan et al., 2026). Changing the numerical solver changes how a learned field is executed; changing residual geometry changes which field errors receive additional training emphasis. The two directions address different parts of the connection between local predictions and final sample quality.

H.2 Output structure, representation learning, and generative geometry

Perceptual objectives compare images through features of a pretrained network (Johnson et al., 2016), and learned perceptual similarity measures demonstrate that deep features can align with human judgments across visual distortions (Zhang et al., 2018). Latent diffusion separates the representation and generation stages (Rombach et al., 2022), making their geometric compatibility consequential: equal distances in a compressed representation need not correspond to equal changes after decoding. Decoder-induced Riemannian metrics describe this effect as a pullback of output-space geometry (Arvanitidis et al., 2018). Latent Perceptual Loss (LPL) uses decoder features to compare an estimated clean latent with its paired clean target during diffusion or FM training (Berrada Ifriqi et al., 2025). Representation Alignment (REPA) supervises noisy-input hidden representations with features from a pretrained visual encoder (Yu et al., 2025); REPA-E extends alignment to joint tuning of the autoencoder and diffusion transformer (Leng et al., 2025). Together, these approaches connect generation to perceptual or semantic structure through prediction losses and representation learning.

Recent work also examines directional sensitivity explicitly. MoRAE analyzes motion-latent residuals through the decoder pullback metric, including residual magnitude, mean decoder gain, and directional alignment, and learns representations suited to subsequent flow training (Zhu et al., 2026). STREAM uses singular directions of a trained velocity-field Jacobian to construct anisotropic noise for decoder training (Cho et al., 2026). Thus, both residual orientation and information from generation dynamics already inform representation or decoder design. **The distinction here is where that information acts:** ImpactFlow holds the decoder fixed and uses the complete remaining teacher–decoder–readout map to assign directional costs to current FM residuals. This differs from shaping a representation for later generation or supervising a clean prediction directly with decoder features. Our LPL-style experimental control adapts the latter idea to physical motion features while preserving the fixed motion decoder and FM training setting.

Geometry can also specify the space in which generation takes place. Riemannian Flow Matching constructs target fields on manifolds (Chen & Lipman, 2024); Metric Flow Matching learns interpolants using a data-induced Riemannian metric (Kapuśniak et al., 2024); Pullback Flow Matching learns a metric-preserving diffeomorphism between the data manifold and a designable latent manifold, then performs Riemannian flow matching under the induced pullback metric (de Kruiff et al., 2024). These methods use geometry to define support, paths, or metric-preserving maps. In our construction, the path and conditional velocity labels remain fixed, and geometry describes the output consequences of errors along that path. The shared language of metrics and pullbacks therefore refers to a different role in the learning problem.

H.3 Terminal objectives and the assignment of local importance

Classical goal-oriented error estimation evaluates local residuals according to their effect on a selected quantity of interest, using an adjoint to propagate sensitivity backward (Becker & Rannacher, 2001; Cao & Petzold, 2004). This provides a conceptual precedent for linking local approximation to terminal accuracy. Generative reward optimization makes a related connection for learned samplers. DRaFT differentiates reward through a sampling process to fine-tune diffusion models (Clark et al., 2024). Adjoint Matching formulates fine-tuning as stochastic optimal control and uses detached adjoint-based targets to learn a reward-improving control (Domingo-Enrich et al., 2025). Adjoint Sampling extends this framework to sampling from unnormalized densities, using reciprocal projections and a replay buffer to reuse terminal samples and energy gradients (Havens et al., 2025). These approaches make the desired terminal distribution or reward part of the task that the generative dynamics learn to accomplish.

Decision-Weighted FM is particularly close to our learning question. It relates velocity error to downstream decision loss through transported sensitivity bounds, then uses endpoint-dependent scalar weights in FM regression (Xie et al., 2026). It retains conditional velocity labels while applying scalar

emphasis derived from downstream sensitivity. **IMPACTFLOW resolves the directional structure within each state's residual space**, assigning different costs to equal-norm errors according to their effects after propagation through the remaining generator.

The relationship to adjoint methods is similarly about the use of the transported signal. In IMPACTFLOW, adjoints estimate a quadratic cost on the existing conditional-FM residual, preserving it as the training target. The conditional-independence and positivity assumptions in Appendix B.1 establish preservation of the population optimum under the resulting directional weighting. Random probes and stochastic trace estimation (Hutchinson, 1989) make the metric practical without materializing it and provide the computational realization of this learning principle.

H.4 STRUCTURED MOTION GENERATION AND THE SCOPE OF THE COMPARISON

Motion makes output consequences concrete because errors can affect pose, temporal coordination, and interactions with the ground. ACTOR models action-conditioned motion with a transformer variational autoencoder (Petrovich et al., 2021), and HumanML3D supports the study of language-conditioned motion generation (Guo et al., 2022). MDM predicts clean motion in a diffusion process, enabling geometric losses on reconstructed motion (Tevet et al., 2023); MARDM combines masked autoregressive modeling with diffusion-based motion prediction (Meng et al., 2025). In 2026, MoRAE develops latents suited to motion flow modeling (Zhu et al., 2026), while Riemannian Motion Generation represents motion on product manifolds and learns generative dynamics with Riemannian FM (Miao et al., 2026). These works improve the representation or generative formulation used for structured outputs. Our motion study holds the student architecture and decoder interface fixed to investigate a different variable: the training importance assigned to local residual directions.

Table 16: Where output information enters closely related approaches.

Approach	Role of output information	Main object being shaped
LPL (Berrada Ifriqi et al., 2025)	Decoder-feature distance to a paired clean target	Clean-prediction supervision
REPA / REPA-E (Yu et al., 2025; Leng et al., 2025)	Pretrained visual features	Hidden representations; jointly tuned latent space in REPA-E
MoRAE (Zhu et al., 2026)	Motion structure and decoder sensitivity	Representation for subsequent flow training
STREAM (Cho et al., 2026)	Field-Jacobian directions for anisotropic noise	Decoder robustness during training
Adjoint Matching (Domingo-Enrich et al., 2025)	Terminal reward and backward sensitivity	Control fine-tuning targets
Decision-Weighted FM (Xie et al., 2026)	Downstream decision sensitivity	Scalar weights on velocity regression
IMPACTFLOW	Output geometry transported through remaining generation	Directional costs on the original FM residual