VALIDITY-AWARE FLOW MATCHING ON LIE GROUPS: WHAT TRANSFERS TO TARGET-FREE GENERATORS?

Anonymous authors
Paper under double-blind review

ABSTRACT

Flow matching learns a marginal vector field from conditional training paths, yet many notions of validity concern the trajectory itself. Recent methods improve those conditional paths through geometry-aware or cost-aware interpolation, or impose constraints during fitting and inference. None of them settles which path properties survive marginalisation into a target-free generator. We study that gap on compact Lie groups, accounting for population marginalisation, finite-capacity regression and discrete sampling. At the population optimum, conditional flow matching preserves every one-time marginal, including terminal validity under state-only criteria, and expected additive state-time costs. The path law is another matter. Conditional paths and the population flow they induce can agree on every one-time marginal and still differ in region hitting, path length and endpoint coupling. Integrated cost alone cannot control endpoint violations. Control becomes available once the deployed path satisfies a temporal modulus, of the kind a speed cap or a pathwise kinetic-energy bound supplies. We trace these effects on tori, rotation groups and molecular conformations using bounded, endpoint-pinned Lie-group paths with exact left-trivialised velocities. The experiments separate endpoint coupling, path design, regression and numerical sampling, and show where a conditional-path improvement survives deployment and where a validity claim needs evidence from the deployed generator itself.

1 INTRODUCTION

Flow matching turns generative modelling into velocity regression. One chooses conditional paths between source and target samples, regresses their velocities and integrates the learned field at test time (Lipman et al., 2022; Liu et al., 2022; Albergo & Vanden-Eijnden, 2022; Albergo et al., 2025). This exposes the conditional trajectory as a design variable. Recent work improves endpoint couplings, adapts interpolation to geometry, or optimises paths for kinetic, physical, or task-specific costs (Tong et al., 2023; Pooladian et al., 2023; Chen & Lipman, 2024; Kapuśniak et al., 2024; Liu et al., 2024; Shaul et al., 2023; Du et al., 2026). Yet the deployed generator does not observe these endpoint-conditioned paths, which we call teacher paths. Squared-error CFM replaces them by a marginal vector field obtained by averaging conditional velocities at each state and time. This raises a distinct transfer question. *Which properties engineered into conditional paths remain properties of the target-free generator?* The distinction matters for path length, hitting, winding, accumulated costs, and endpoint coupling, all of which depend on temporal linkage rather than on isolated states. Without an explicit transfer argument, an improvement measured on conditional teachers can therefore be read too strongly as evidence about trajectories that the deployed model actually generates. We separate this population marginalisation step from finite regression and numerical sampling so that failures caused by averaging are not conflated with downstream approximation error. Broader comparisons and stage-by-stage discussion are deferred to Sections A and B.

Two gaps separate a conditional construction from the generator it produces. At the population optimum every one-time marginal transfers, and with it the terminal law and expected additive state–time costs (Lipman et al., 2022; Chen & Lipman, 2024). The full path law does not, and Figure 1 gives the minimal counterexample. The second gap concerns what an integrated path cost implies about the endpoint, since a small integral cannot locate cost in time. Section 3 settles both.

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

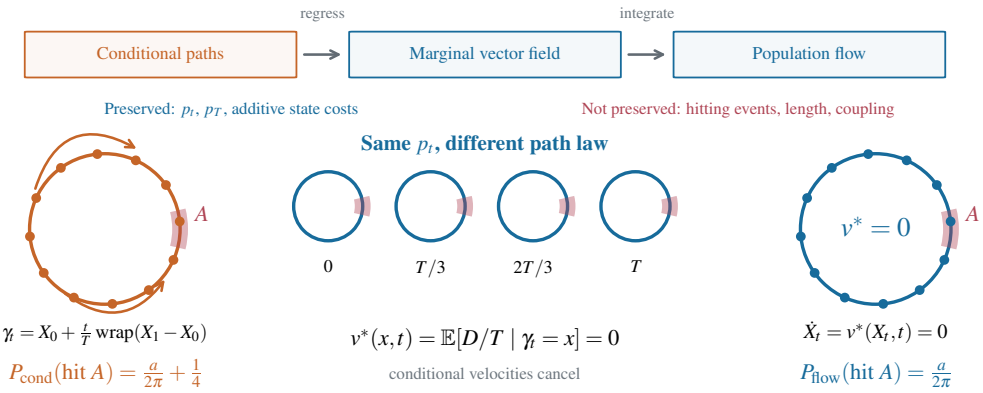


Figure 1: Matched one-time marginals do not determine the path law. Shortest-geodesic conditional paths between independent uniform endpoints on $\mathbb{S}^1$, the zero-correction member of our path family, have uniform marginals at every time. Their conditional velocities average to $v^* = 0$, so the population CFM flow is stationary. The marginals agree for all t, but arc-hitting probabilities and path lengths differ (Proposition 1).

We use compact Lie groups as a controlled non-Euclidean setting where conditional paths, exact velocities, population marginalisation, and numerical integration can be separated cleanly. Lie-group structure is already important in generative modelling of rotations, molecular conformations, and protein backbones (De Bortoli et al., 2022; Klein et al., 2023; Hassan et al., 2024; Yim et al., 2023b; Bose et al., 2024). Here it serves as a diagnostic laboratory rather than the source of the main claim. We construct bounded endpoint-pinned geodesic corrections and intervene separately on pairing, conditional path design, finite-capacity fitting, and sampling. Tori expose population non-transfer, $\mathrm{SO}(3)$ exposes representational limits, and molecular reconstruction tests whether state-space validity survives decoding.

2 Geodesic-anchored flow matching

We study flow matching through a path family whose endpoints are fixed but whose interior geometry can be changed deliberately. We use as our running example a pair of periodic angles, a point on the torus $\mathbb{T}^2$. Let $g_0 \sim \mu_0$ and $g_1 \sim \mu_1$ be paired by a coupling π,[1] and let $V \geq 0$ assign high cost to undesirable states. For a path $g : [0,T] \to G$, we measure the accumulated state cost

$$I(g) = \int_0^T V(g_t)\,dt, \qquad \overline{V}(g) = I(g)/T. \tag{1}$$

The method has two deliberately separate stages. First, we modify the endpoint-conditioned path while preserving its paired endpoints. Second, after that path family is fixed, we regress its conditional velocities into a state–time vector field and deploy only that fitted field. This separation is central to the transfer question studied in Section 3, because the conditional path network and target endpoint are available during path construction but neither is available at generation time. Population statements condition on a fixed context c, which we omit from the notation. Figure 2 places these stages alongside related interventions in flow matching.

2.1 Geodesic-anchored path family

The construction is easiest to see on a torus. For paired endpoints (g_0, g_1), let $\xi = \mathrm{wrap}(g_1 - g_0)$ be the shortest wrapped displacement and write $s = t/T$. We perturb the corresponding geodesic by an endpoint-pinned correction,

$$\gamma_\psi(t) = g_0 + s\xi + \eta_\psi(s) \pmod{2\pi}, \qquad \eta_\psi(0) = \eta_\psi(1) = 0, \tag{2}$$

[1] A coupling is a joint distribution on endpoint pairs with fixed marginals μ_0 and μ_1. Changing it changes the pairing while preserving the two endpoint distributions.

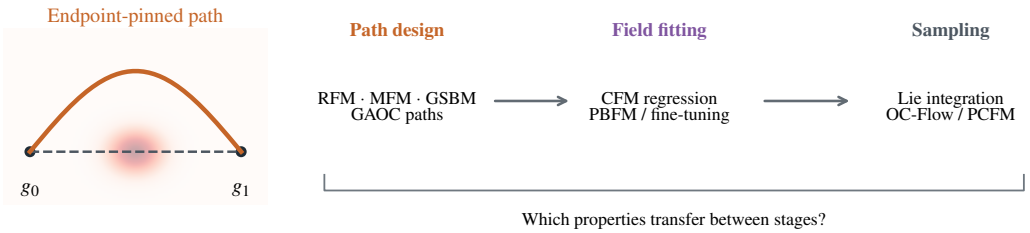


Figure 2: Interventions act at different stages of flow matching. Left, a schematic torus detour fixes both endpoints. Right, methods act on conditional paths, field fitting or inference. We analyse which properties transfer between stages. Only the fitted field and integrator are used for target-free generation. RFM is Riemannian flow matching, MFM metric flow matching, GSBM generalised Schrödinger bridge matching, PBFM physics-based flow matching, OC-Flow optimal-control flow and PCFM physics-constrained flow matching. GAOC is defined below and the mechanism controls in Section D.

so the route can change without changing the training pair. A network maps (g_0, g_1, c) to Bernstein control points,

$$\eta_\psi(s) = \sum_{j=0}^{m} B_{j,m}(s) c_j, \qquad c_0 = c_m = 0, \qquad c_j = R\, z_j / \sqrt{1 + \|z_j\|^2}. \tag{3}$$

The endpoint coefficients enforce pinning exactly, while the radial map keeps $\|c_j\| < R$. Differentiating the Bernstein basis gives $\|\dot\eta_\psi\| \le 2mR/T$, so the same parameterisation that controls the representable detour also supplies a pathwise rate bound. On the torus the conditional angular velocity is simply $u_\psi = \xi/T + \dot\eta_\psi$.

The same idea extends to a compact connected Lie group G by replacing addition with group multiplication and displacements with Lie-algebra coordinates. With a bi-invariant metric and a fixed deterministic logarithm branch, we use

$$\gamma_\psi(t) = g_0 \exp_G(s\xi) \exp_G\big(\eta_\psi(s)\big), \qquad \xi = \log_G(g_0^{-1} g_1). \tag{4}$$

The experiments use $\mathbb{T}^d$ with its flat metric and $\mathrm{SO}(3)^m$ with $\langle a, b\rangle = \frac{1}{2}\operatorname{tr}(a^\top b)$ per factor. Under the chosen minimal branch, $g_0 \exp_G(s\xi)$ is the shortest geodesic and the correction changes only its interior. For a matrix group, the left-trivialised velocity $u = \gamma^{-1}\dot\gamma$ expresses motion in a common Lie algebra and has norm equal to the path speed. Differentiating Equation (4) yields

$$u_\psi(t) = \mathrm{Ad}_{\exp_G(-\eta_\psi)}\Big(\frac{\xi}{T}\Big) + J_r(\eta_\psi)\dot\eta_\psi, \qquad J_r(\eta) = \int_0^1 \exp(-r\ \mathrm{ad}_\eta)\, dr. \tag{5}$$

The adjoint transports the geodesic velocity through the correction, while the right Jacobian differentiates the correction. Both reduce to the identity on the torus and both are needed on $\mathrm{SO}(3)$. Bi-invariance together with Equation (3) gives the conditional speed bound

$$\|u_\psi\| \le B_u = (\mathrm{diam}(G) + 2mR)/T. \tag{6}$$

2.2 Path objective and field fitting

We choose the detour by minimising an action that balances state cost against motion,

$$\mathcal{A}(\psi) = \mathbb{E}_\pi \int_0^T \Big[\lambda_V V(\gamma_\psi(t)) + \frac{\lambda_K}{2}\|u_\psi(t)\|^2 + \lambda_A \|D_t \dot\gamma_\psi(t)\|^2\Big]\, dt, \tag{7}$$

where D_t is the covariant derivative along the path. We estimate the integral with stratified times and report the unweighted state cost separately from the training objective. Setting the correction to zero recovers the geodesic control, and the principal experiments use $\lambda_A = 0$. This stage optimises endpoint-conditioned trajectories. An expected kinetic-energy penalty does not by itself impose a temporal modulus on the vector field used at deployment.

After optimising ψ, we freeze the path family and fit a target-free vector field with conditional flow matching,

$$\ell_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t\sim U[0,T],\,\pi} \|v_\theta(\gamma_\psi(t), t, c) - u_\psi(t)\|^2. \tag{8}$$

We call this controlled construction GAOC-FM (Geodesic-Anchored Optimal Control flow matching). GAOC-FM (OT) changes only the endpoint coupling, using minibatch optimal transport implemented by balanced assignment under squared geodesic distance instead of independent pairing. Both variants preserve the empirical endpoint marginals. At generation time, $g_0 \sim \mu_0$ is advanced by Lie–Euler or Lie–Heun updates $g_{k+1} = g_k \exp_G(ha_k)$. The deployed field receives state, time and context, but neither the target endpoint nor the conditional-path network.

For the regularity intervention, we also cap the deployed speed by parameterising $v_\theta = Bz_\theta/\sqrt{1+\|z_\theta\|^2}$. In exact arithmetic, both integrators then satisfy $d(g_{k+1}, g_k) \le Bh$, since each update uses one field value or an average of two capped values. The principal comparisons use unconstrained fields. The cap appears only in the regularity experiment, where we evaluate the resulting validity–fidelity trade-off directly.

3 PATH-TO-FLOW TRANSFER

3.1 POPULATION MARGINALS AND PATH LAWS

At a state visited by several conditional paths, squared-error regression averages their velocities. Its population optimum is

$$v^*(g,t) = \mathbb{E}[u_\psi(t) \mid \gamma_\psi(t) = g].$$

Write p_t^ψ for the law of $\gamma_\psi(t)$. Under the continuity-equation and flow-uniqueness assumptions in Section C, the flow of v^* reproduces p_t^ψ (Lipman et al., 2022; Chen & Lipman, 2024). Let g^* denote its random trajectory and q_t^* its law at time t. Then

$$q_t^* = p_t^\psi \quad \forall t, \qquad \mathbb{E}\int_0^T \phi(t, g_t^*)\, dt = \mathbb{E}\int_0^T \phi(t, \gamma_\psi(t))\, dt \tag{9}$$

for every bounded measurable state–time observable ϕ. Hence $q_T^* = \mu_1$, $\mathbb{E}I(g^*) = \mathbb{E}I(\gamma_\psi)$, and every terminal criterion depending only on state transfers exactly at the population optimum. More generally, path families pinned to the same target law have the same population endpoint statistics regardless of their path costs. Optimising the conditional paths can instead change intermediate marginals and additive state–time quantities. Endpoint differences between finite generators must therefore come from fitting or sampling error, or from changing the target itself. This preservation result is strong but deliberately limited, because equality of one-time marginals does not identify the joint law across time, so it cannot by itself transfer hitting events, winding, path length, or endpoint pairing.

Proposition 1 (Identical marginals can hide different path events)**.** *On $\mathbb{S}^1$, let X_0, X_1 be independent uniform endpoints and $D = \mathrm{wrap}(X_1 - X_0) \in [-\pi, \pi)$. The conditional geodesic $\gamma_t = X_0 + tD/T \pmod{2\pi}$ has uniform marginals and population CFM field $v^* \equiv 0$. For a closed arc A of length $0 < a \le \pi$, its hitting probability is $a/(2\pi) + 1/4$, while the stationary population flow's is $a/(2\pi)$. Their expected path lengths are $\pi/2$ and zero, respectively.*

This construction is the zero-correction member of Equation (2) and uses no randomness beyond its endpoints. The conditional geodesics and the population flow agree at every fixed time, but marginalisation replaces the independent endpoint coupling by the identity coupling of a stationary flow. In particular, a start outside A can enter the arc with positive probability under the conditional path while the population trajectory never moves (Section C.2). Table 1 summarises the resulting boundary. Marginal and additive state–time statements transfer at the population optimum, while genuinely path-dependent statements require separate control of the deployed dynamics.

Finite approximation introduces a second, distinct issue. Small regression error does not restore uniform dynamical control, since smooth fitted fields on $\mathbb{S}^1$ can have L^2 error tending to zero while their maximum speed diverges (Section C.3). This does not contradict the population speed bound inherited by averaging bounded conditional velocities. Acceleration behaves differently even at population level because averaging introduces a covariance term into the marginal field's acceleration (Equation (14) in Section C.3). Conditional smoothness alone therefore does not bound deployed acceleration. We consequently treat population transfer, finite fitting, and numerical deployment as separate stages rather than interpreting a well-behaved teacher path as a guarantee about the sampled generator.

Table 1: Which conditional-path properties reach the generated flow? Exact statements use the population regularity assumptions. A finite fit or numerical sampler requires additional control of its own errors.

Conditional-path property	Population flow	Fitted and sampled flow
Marginals, including at T	Preserved	Subject to fitting and sampling error
Additive state–time expectations	Preserved	Subject to fitting and sampling error
Hitting, maxima, length and winding	Not preserved in general	No automatic recovery
Endpoint pairing and individual paths	Not preserved in general	No automatic recovery
Acceleration bound	Not inherited in general	Must assess deployed dynamics
Speed bound B_u	Inherited by averaging	Can enforce a field cap B

3.2 Endpoint validity certificates

The remaining question is what an integrated path cost can certify about the endpoint of the deployed trajectory. An integral controls occupancy rather than timing. Nonnegativity gives $T^{-1}\int_0^T \mathbf{1}\{V(g_t) \geq \tau\}\, dt \leq I(g)/(\tau T)$, but it does not say whether a violation occurs near the endpoint. On the circle, take $V(\theta) = (1-\cos\theta)/2$ and let a path remain at $\theta = 0$ until $T - \delta$ before moving linearly to π. Its integrated cost is $\delta/2$ while its terminal cost is one (Figure 3, left), so the integral can vanish without improving terminal validity. More generally, even with a fixed smooth potential and fixed endpoints, no bound that vanishes with $\mathbb{E}I(g)$ can control the terminal violation probability over all absolutely continuous path laws (Proposition C.2). The obstruction is temporal, because every path in the construction has finite speed, but the family has no common speed bound. This is compatible with Equation (9). Population marginal preservation transfers an already specified terminal law, whereas an integral alone does not determine that law.

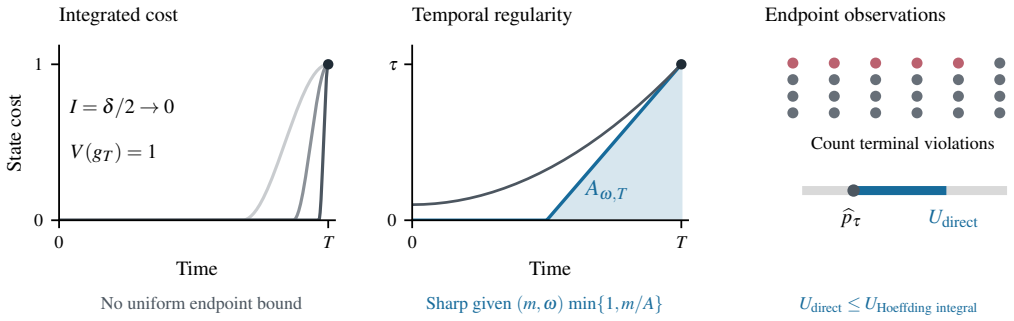


Figure 3: More information supports a stronger endpoint claim. Left, a vanishing integral can coexist with terminal cost one. Centre, a temporal modulus forces an unavoidable area $A_{\omega,T}$ and a bound that is information-theoretically sharp given only the mean and modulus. A speed cap gives the triangle shown. Right, sampled endpoints give a direct band that is at least as tight as the Hoeffding integral bound at equal confidence and sample size, as shown below. The endpoint dots and bar are schematic.

A common temporal modulus closes this gap. Suppose the cost cannot decrease by more than $\omega(s)$ when we look back s units of time from the endpoint. Then $V(g_T) \geq \tau$ forces the preceding path to occupy the deterministic floor $V(g_{T-s}) \geq (\tau - \omega(s))_+$. Integrating that floor yields an unavoidable area $A_{\omega,T}(\tau)$, turning the expected path cost into a terminal probability certificate.

Theorem 2 (From integrated cost to terminal validity)**.** *Let g be a random absolutely continuous path on $[0,T]$ and $V \geq 0$. Suppose a nondecreasing $\omega : [0,T] \to [0,\infty)$ satisfies $V(g_T) - V(g_{T-s}) \leq \omega(s)$ for all $s \in [0,T]$, almost surely. Then for every $\tau > 0$,*

$$\mathbb{P}\{V(g_T) \geq \tau\} \leq \min\left\{1, \frac{\mathbb{E}I(g)}{A_{\omega,T}(\tau)}\right\}, \qquad A_{\omega,T}(\tau) = \int_0^T (\tau - \omega(s))_+\, ds. \tag{10}$$

A zero denominator gives the uninformative bound of one.

The proof simply integrates the pathwise floor and takes expectations, and two practical moduli follow from path motion. If V is Lipschitz, a uniform speed cap yields a linear modulus. An energy bound

yields a square-root modulus. Both assumptions are pathwise, so an expected-energy penalty does not establish either one. The resulting mean-plus-modulus certificate is information-theoretically sharp given only these ingredients (Proposition C.3), although this does not claim optimality within a fixed target law, path family, or potential.

Corollary 3 (Speed and energy bounds)**.** *Let V be L_V-Lipschitz. A pathwise speed cap $\|\dot{g}_t\| \le B$ gives $\omega(s) = Ks$ with $K = L_V B$, and*

$$A_{K,T}(\tau) = \begin{cases} \tau^2/(2K), & K > 0,\ \tau \le KT, \\ \tau T - KT^2/2, & K > 0,\ \tau > KT, \\ \tau T, & K = 0. \end{cases} \tag{11}$$

A pathwise energy bound $\int_0^T \|\dot{g}_t\|^2 dt \le E$ instead gives $\omega(s) = L_V\sqrt{Es}$.

For the sampled generator, the same logic can be made discrete without inserting a quadrature approximation. A bounded conditional velocity implies $\|v^*\| \le B_u$ by conditional Jensen, but finite regression need not retain that bound. The field output map in Section 2 therefore enforces $d(g_{k+1}, g_k) \le Bh$ for $h = T/n$. If $V(g_n) \ge \tau$, the resulting cost floor along the stored trajectory gives

$$I_h^{\mathrm{trap}} \ge A_h^{\mathrm{trap}}(\tau) = h\left[\frac{(\tau - KT)_+}{2} + \sum_{j=1}^{n-1}(\tau - Kjh)_+ + \frac{\tau}{2}\right]. \tag{12}$$

For N independent generated paths from a fixed model and $0 \le V \le V_{\max}$, Hoeffding's inequality then gives

$$\mathbb{E}I_h^{\mathrm{trap}} \le U = \min\left\{TV_{\max},\ \widehat{I}_h^{\mathrm{trap}} + TV_{\max}\sqrt{\frac{\log(1/\alpha)}{2N}}\right\} \tag{13}$$

with confidence at least $1 - \alpha$, so $\min\{1, U/A_h^{\mathrm{trap}}(\tau)\}$ bounds the endpoint violation probability simultaneously for every $\tau > 0$. For M fixed models we use $\alpha = \delta/M$ for joint confidence $1 - \delta$.

When endpoints are stored, they provide a simpler direct certificate. Let $\hat{p}(\tau)$ be the fraction with $V(g_n) \ge \tau$. The one-sided DKW–Massart inequality[2] gives $\mathbb{P}\{V(g_n) \ge \tau\} \le \hat{p}(\tau) + \epsilon$ for every threshold at once, where $\epsilon = \sqrt{\log(1/\alpha)/(2N)}$. It requires no modulus, Lipschitz constant, or cost range, and for capped sampled paths it is pointwise at least as tight as the Hoeffding integral band at equal N and α (Proposition C.4). The integral certificate is therefore most useful when endpoints are unavailable but an independently justified bound on deployed path cost is available. A low training objective alone does not provide such a bound. Likewise, an observed maximum speed can replace a deterministic cap only when accompanied by a uniform tail bound for exceeding it (Equation (15)), and every certificate remains specific to the stated potential V rather than automatically transferring to other physical validity criteria.

4 Experiments

We test the transitions in Figure 1 with target-free generators that share priors, data, field architectures, and declared budgets. We fix hyperparameters from a validation pilot and evaluate path cost, occupation, endpoint violations, and distributional fidelity separately. We use intrinsic empirical W_2 as the primary fidelity measure because it follows the group's geodesic distance without introducing a kernel bandwidth. Section D details the protocols and mechanism controls.

4.1 Cost gains survive without a resolved fidelity gain

We first separate pairing from path optimisation using paired changes over ten training seeds (Figure 4). At fixed OT coupling, GAOC-FM (OT) reduces cost by 65%, 73%, and 36% on the wide ridge, narrow ridge, and coupled torus. The descriptive 95% cost intervals exclude zero, whereas the W_2 intervals cross zero on all four tasks. We treat these marginal intervals as exploratory, not as a multiplicity-adjusted family.

[2]The Dvoretzky–Kiefer–Wolfowitz inequality bounds the deviation of an empirical cumulative distribution from its population distribution uniformly over thresholds. Massart gives its sharp constant.

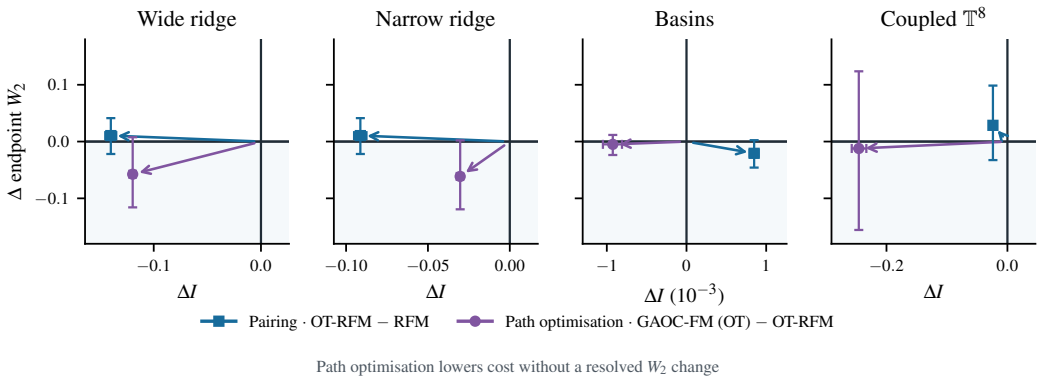


Figure 4: Cost reductions have unresolved fidelity effects. Each arrow starts at zero and ends at a paired mean difference from the stated parent method. Error bars are marginal 95% bootstrap intervals over ten matched seeds, not joint confidence regions. Left and below zero are preferable. All four path-optimisation intervals for endpoint W_2 cross zero. The basins panel uses a different horizontal scale.

Table 2: Selected torus controls, ten seeds per cell. Entries give means and 95% bootstrap half-widths. W_2 uses intrinsic squared-distance assignment on 2000 endpoints per side; I is multiplied by 10^3 in the basins column only. Lower values are preferable. Bold marks the lowest observed mean among methods for each task and metric; it is descriptive, not a significance claim. The reference row uses 50 independent target pairs at the same sample budget (Table 3) and is excluded from bolding. These are mechanism controls (Section D).

Method	Metric	Wide ridge	Narrow ridge	Basins	Coupled $\mathbb{T}^8$
RFM	I	0.323 ±0.007	0.133 ±0.004	1.68 ±0.08	0.704 ±0.011
	W_2	0.285 ±0.040	0.285 ±0.040	0.099 ±0.015	**0.877 ±0.091**
OT-RFM	I	0.183 ±0.003	0.041 ±0.002	2.53 ±0.11	0.680 ±0.010
	W_2	0.295 ±0.047	0.295 ±0.047	0.078 ±0.015	0.905 ±0.092
GSBM-ZN	I	**0.041 ±0.002**	**0.009 ±0.001**	1.80 ±0.13	**0.225 ±0.003**
	W_2	0.419 ±0.091	0.417 ±0.076	0.136 ±0.034	1.200 ±0.120
GAOC-FM	I	0.186 ±0.006	0.070 ±0.005	**1.32 ±0.03**	0.490 ±0.005
	W_2	0.275 ±0.054	0.265 ±0.045	0.088 ±0.017	0.921 ±0.122
GAOC-FM (OT)	I	0.064 ±0.002	0.011 ±0.001	1.60 ±0.04	0.434 ±0.007
	W_2	**0.238 ±0.037**	**0.234 ±0.038**	**0.073 ±0.012**	0.893 ±0.106
Reference–reference	W_2	0.250 ±0.022	0.250 ±0.022	0.073 ±0.008	0.817 ±0.027

We report absolute cost and the same W_2 metric in Table 2. The GSBM-ZN mechanism control attains lower mean cost than GAOC-FM (OT) on three tasks but higher mean W_2 on all four. Because that comparison also changes path capacity and amortisation, it does not isolate either effect. The cost–fidelity association likewise changes sign with the comparison: across the seven controls that share the bounded path family, mean cost and endpoint W_2 correlate at +0.74 and +0.55 on the two ridges but -0.53 on the coupled torus. We use the reference row, averaged over 50 independent target pairs at the same sample budget, only to calibrate the finite-sample scale. Proximity to it does not establish equivalence to the target distribution (Table 3). Table 4 reports secondary MMD measurements for all controls.

We then test path events directly in the checkpointed transfer study (Section G). The first declared ever-hit event included $t = 0$, so shared initially invalid states fixed the method contrast on the ridges. We therefore preregistered a confirmatory study with ten new seeds, integrated each fitted field from the starting states of its own conditional paths, and measured conditional entry: among paths initially below τ, the fraction that later reach it. For GAOC-FM (OT), 5 of eight declared tests survive Holm adjustment, and none shows the generated flow entering more often. At $\tau = 1$ on the wide ridge, entry falls from 0.024 to 0.005 ($p_H = 0.016$). No path enters on the basins (Table 9). OT-RFM, whose teachers are conditional geodesics, shows the same qualitative pattern. These gaps can combine population non-transfer with fitting and sampling error.

4.2 A VALID MODULUS BOUND NEED NOT BE INFORMATIVE

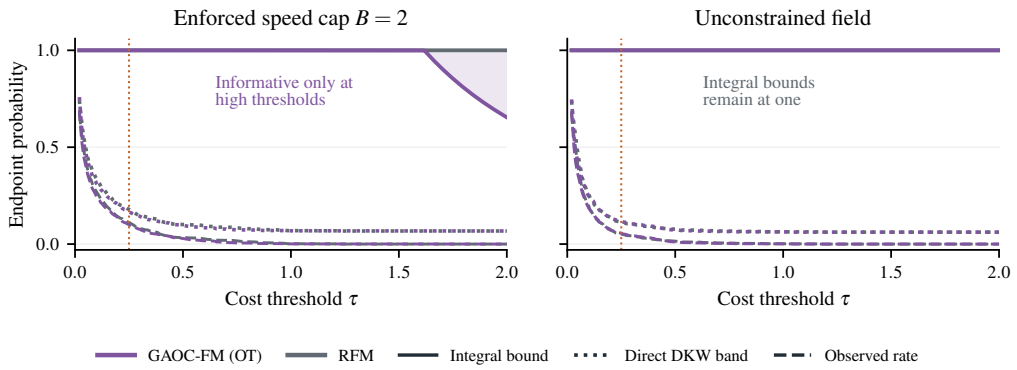


Figure 5: Endpoint observations give tighter bounds. Wide-ridge curves take the maximum over ten seeds per method. Solid curves are integral bounds, dashed curves are observed rates, and dotted curves are direct DKW bands. Each analysis uses its declared 1% family failure probability. The vertical line marks the operating threshold $\tau = 0.25$.

For GAOC-FM (OT) with cap $B = 2$, we find the first informative threshold between 1.58 and 1.62 across seeds. Every tested cap and resolution remains uninformative at $\tau = 0.25$, as do the observed-speed bounds for unconstrained fields (Figure 5 and Table 6). Tightening the cap can also raise terminal cost (Table 5).

At the operating threshold, our direct band gives 0.17 for capped GAOC-FM (OT), whereas the integral bound remains one. Across 84 000 model–threshold pairs at $N = 1000$, neither the Hoeffding nor the tested empirical-Bernstein integral estimate is tighter. Proposition C.4 gives the Hoeffding comparison. The Bernstein comparison is empirical, with protocol in Section D.

4.3 GEOMETRY LIMITS THE AVAILABLE DETOURS

At the central rotation crossing, we compute that radius $R = 0.6$ permits only $32.2°$ of deviation where $42.3°$ is required, whereas $R = 1.5$ permits $80.6°$ (Figure 7). At the smaller radius, GSBM-ZN reaches occupation 0.008 versus 0.248 for GAOC-FM (OT). The larger radius removes this local obstruction but worsens endpoint fidelity (Table 7). Because the radius changes representational capacity, we do not interpret this comparison as a general method ranking.

4.4 A COUNTEREXAMPLE TO UNCALIBRATED VALIDITY EVALUATION

Under the declared overlap rule, we observe no violations in the selected raw MD frames, whereas frozen reconstruction introduces rates of 0.041–0.376 (Figure 6). A lower generated overlap rate can therefore reward avoiding torsions that the representation cannot accommodate. It does not establish improved physical validity. Section E gives the reconstruction and frame-selection protocol, and the artefact retains the generated-conformer diagnostics.

5 RELATED WORK

Path design, geometry, and constraints. Flow matching and stochastic interpolants learn marginal velocity fields from conditional paths (Lipman et al., 2022; Liu et al., 2022; Albergo & Vanden-Eijnden, 2022; Albergo et al., 2025). Recent work improves those teachers through transport couplings (Tong et al., 2023; Pooladian et al., 2023; Chiang et al., 2026), geometry-aware interpolation (Chen & Lipman, 2024; Kapuśniak et al., 2024), and kinetic, control, or state-cost objectives (Liu et al., 2024; Shaul et al., 2023; Du et al., 2026). Manifold and molecular generators exploit non-Euclidean structure directly (De Bortoli et al., 2022; Huang et al., 2022; Klein et al., 2023; Hassan et al., 2024; Bose et al., 2024). Our question comes after all of these. Once CFM has marginalised endpoint-conditioned velocities into a target-free field, which properties of the teacher are still there? One-time

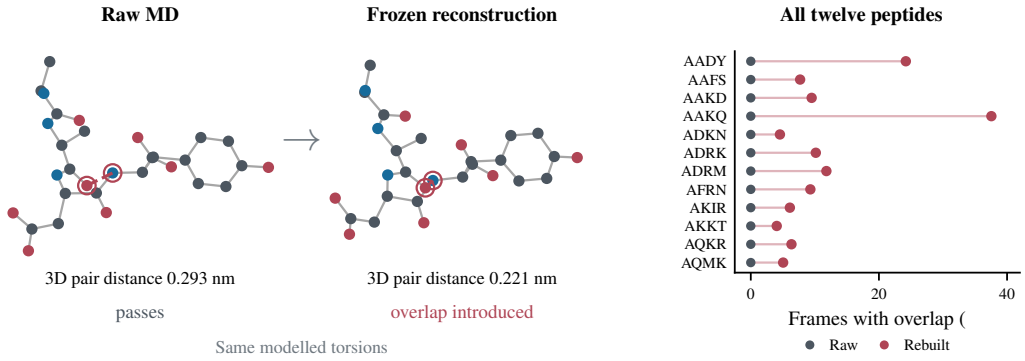


Figure 6: Reconstruction can introduce overlap at fixed torsions. Left, a held-out AADY frame and its frozen-geometry reconstruction. Circles mark one atom pair whose 3D distance crosses the 0.230 nm overlap threshold. Coordinates share an aligned projection. Right, rates on 2000 selected test frames per peptide, scored with the same rule. The frame selection and complete calibration are documented in Section E.

marginals are. Temporal coupling need not be. What we offer is an account of what a successful conditional construction implies about the deployed generator, not another path objective.

Guarantees at later stages. Other methods intervene later, regularising the fitted field or altering sampling through projection, guidance, or control (Christopher et al., 2024; Lou & Ermon, 2023; Baldan et al., 2026; Wang et al., 2024; Utkarsh et al., 2026; Liang et al., 2025; Liu et al., 2026). Approximation theory then asks how finite regression and numerical integration reproduce a specified flow (Zhou & Liu, 2025; Bansal et al., 2024; Su et al., 2026). These stages are complementary and distinct. A teacher path, a population field, a fitted field and a sampled trajectory support different claims. Our first transfer gap appears before approximation enters. The endpoint certificate runs the other way and assumes regularity of the deployed dynamics. Section A gives the broader literature map, and Section B makes the stage distinction explicit.

6 Limitations

Our endpoint bound is sharp only given a mean cost and a modulus. Fix the potential, the target law, the conditional-path family or an energy budget and a tighter result may follow. None of the integral bounds we tested is informative at the operating threshold.

The principal fitted fields are left unconstrained, so we treat the speed cap as a separate intervention rather than a property of the main models. Our mechanism controls are shared-framework implementations, not released methods run on their native benchmarks, and the main cost and fidelity intervals are descriptive.

One gap resists attribution. A difference between conditional and generated path events need not come from population non-transfer, since fitting and sampling error contribute and a finite observation grid cannot establish avoidance along a continuous path. The molecular diagnostic is narrower still: one overlap rule, a frozen reconstruction, one trajectory per peptide.

7 Conclusion

Exact flow matching preserves one-time marginals and the expected integrated state cost. It does not preserve the path law. Even conditional geodesics can lose path events once marginalised, and in our confirmatory study the generated trajectories entered costly regions less often than the teachers they were fitted to. A mean integrated cost together with a temporal modulus gives a sharp bound under exactly that information. Where endpoints can be sampled, the direct band was at least as tight in every setting we tested. State which stage a claim refers to, what was observed, and what was assumed.

REPRODUCIBILITY STATEMENT

All experiments run from a single package with a pinned environment. The analytic tasks are generated from versioned configurations with no external data. The molecular data are downloaded by an explicit manifest that records the dataset revision, the file list, byte sizes and local SHA-256 hashes, and the molecule selection rule is stated in Section E and was fixed before any model was trained. Run records store training configurations, seeds, code hashes, metric versions, devices, timings and training curves. Saved per-sample measurements support the reported recalculations. Both transfer studies save every model's weights and random-number state and verify that a reload reproduces its evaluation; the controlled comparisons store per-sample measurements for every reported statistic. Every numerical table and experimental figure is generated by an export script that fails if a primary run is missing. Conceptual figures are labelled as illustrations. Generator code and evidence dependencies are recorded in `results_manifest.json`. The test suite that guards the measurement and geometry conventions (Section D) covers numerical geometry, scoring and inference access. Upstream baseline repositories are pinned by commit in a lock file. Section D defines the scope of the mechanism controls. The repository retains the full audit history and additional results omitted from the concise appendix.

ETHICS STATEMENT

This work studies generative models on compact Lie groups and evaluates them on public simulation data of short peptides. It involves no human subjects, no personally identifying data and no new data collection. Generative models of molecular conformations have dual-use potential in principle. The systems here are four-residue peptides evaluated through reconstruction diagnostics, far from any capability of concern, and we release no model weights for larger systems.

USE OF LARGE LANGUAGE MODELS

A large language model was used as an assistive tool during this work. It helped audit implementation and measurement conventions, drafted and refactored implementation, experiment and analysis code, proposed the structure of the transfer statements, and drafted and edited the text of this manuscript. All experiments were run on the authors' infrastructure, and all empirical results come from recorded runs and numerical checks. Every bibliographic record was verified by the authors against its primary source. The authors are responsible for the content, including the correctness of the proofs and the accuracy of the reported results.

REFERENCES

Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.

Giacomo Baldan, Qiang Liu, Alberto Guardone, and Nils Thuerey. Physics vs distributions: Pareto optimal flow matching with physics constraints. In *International Conference on Learning Representations*, volume 2026, pp. 101215–101243, 2026.

Vansh Bansal, Saptarshi Roy, Alessandro Rinaldo, and Purnamrita Sarkar. On the convergence and straightness of rectified flow. *arXiv preprint arXiv:2410.14949*, 2024.

A van Bondi. van der waals volumes and radii. *The Journal of Physical Chemistry*, 68(3):441–451, 1964.

Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian Fatras, Jarrid Rector-Brooks, Chenghao Liu, Andrei Nica, Maksym Korablyov, Michael Bronstein, and Alexander Tong. Se (3)-stochastic flow matching for protein backbone generation. In *International Conference on Learning Representations*, volume 2024, pp. 22590–22621, 2024.

Ricky TQ Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, volume 2024, pp. 47922–47945, 2024.

Chiensheng Chiang, Kuan-Hsun Tu, Jia-Wei Liao, Cheng-Fu Chou, and Tsung-Wei Ke. Cot-fm: Cluster-wise optimal transport flow matching. *arXiv preprint arXiv:2603.13395*, 2026.

Jacob K Christopher, Stephen Baek, and Ferdinando Fioretto. Constrained synthesis with projected diffusion models. *Advances in Neural Information Processing Systems*, 37:89307–89333, 2024.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. *Advances in Neural Information Processing Systems*, 35:2406–2422, 2022.

Shukai Du, Junzhe Zhang, and Yiming Li. Lagrangian flow matching: A least-action framework for principled path design. *arXiv preprint arXiv:2605.15419*, 2026.

Nic Fishman, Leo Klarner, Valentin De Bortoli, Emile Mathieu, and Michael Hutchinson. Diffusion models for constrained domains. *arXiv preprint arXiv:2304.05364*, 2023.

Octavian Ganea, Lagnajit Pattanaik, Connor Coley, Regina Barzilay, Klavs Jensen, William Green, and Tommi Jaakkola. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. *Advances in Neural Information Processing Systems*, 34:13757–13769, 2021.

Majdi Hassan, Nikhil Shenoy, Jungyoon Lee, Hannes Stärk, Stephan Thaler, and Dominique Beaini. Et-flow: Equivariant flow-matching for molecular conformer generation. *Advances in Neural Information Processing Systems*, 37:128798–128824, 2024.

Chin-Wei Huang, Milad Aghajohari, Joey Bose, Prakash Panangaden, and Aaron Courville. Riemannian diffusion models. *Advances in Neural Information Processing Systems*, 35:2750–2761, 2022.

Guillaume Huguet, James Vuckovic, Kilian Fatras, Eric Thibodeau-Laufer, Pablo Lemos, Riashat Islam, Cheng-Hao Liu, Jarrid Rector-Brooks, Tara Akhound-Sadegh, Michael Bronstein, et al. Sequence-augmented se (3)-flow matching for conditional protein generation. *Advances in Neural Information Processing Systems*, 37:33007–33036, 2024.

Arieh Iserles, Hans Z Munthe-Kaas, Syvert P Nørsett, and Antonella Zanna. Lie-group methods. *Acta Numerica*, 9:215–365, 2000.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *Advances in Neural Information Processing Systems*, 35:24240–24253, 2022.

Bowen Jing, Bonnie Berger, and Tommi Jaakkola. Alphafold meets flow matching for generating protein ensembles. *arXiv preprint arXiv:2402.04845*, 2024a.

Bowen Jing, Hannes Stärk, Tommi Jaakkola, and Bonnie Berger. Generative modeling of molecular dynamics trajectories. *Advances in Neural Information Processing Systems*, 37:40534–40564, 2024b.

Kacper Kapuśniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Avishek J Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold. *Advances in Neural Information Processing Systems*, 37:135011–135042, 2024.

Leon Klein, Andreas Krämer, and Frank Noé. Equivariant flow matching. *Advances in Neural Information Processing Systems*, 36:59886–59910, 2023.

Jinhao Liang, Yixuan Sun, Anirban Samaddar, Sandeep Madireddy, and Ferdinando Fioretto. Chance-constrained flow matching for high-fidelity constraint-aware generation. *arXiv preprint arXiv:2509.25157*, 2025.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Guan-Horng Liu, Tianrong Chen, Evangelos Theodorou, and Molei Tao. Mirror diffusion models for constrained and watermarked generation. *Advances in Neural Information Processing Systems*, 36: 42898–42917, 2023.

Guan-Horng Liu, Yaron Lipman, Maximilian Nickel, Brian Karrer, Evangelos Theodorou, and Ricky TQ Chen. Generalized schrödinger bridge matching. In *International Conference on Learning Representations*, volume 2024, pp. 14527–14552, 2024.

Lin Liu, Caiyan Jia, Guanyi Yu, Ziying Song, JunQiao Li, Feiyang Jia, Peiliang Wu, Xiaoshuai Hao, and Yadan Luo. Guideflow: Constraint-guided flow matching for planning in end-to-end autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3719–3728, 2026.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

Aaron Lou and Stefano Ermon. Reflected diffusion models. In *International Conference on Machine Learning*, pp. 22675–22701. PMLR, 2023.

Emile Mathieu and Maximilian Nickel. Riemannian continuous normalizing flows. *Advances in Neural Information Processing Systems*, 33:2503–2515, 2020.

Robert T McGibbon, Kyle A Beauchamp, Matthew P Harrigan, Christoph Klein, Jason M Swails, Carlos X Hernández, Christian R Schwantes, Lee-Ping Wang, Thomas J Lane, and Vijay S Pande. Mdtraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal*, 109(8):1528–1532, 2015.

Kirill Neklyudov, Rob Brekelmans, Daniel Severo, and Alireza Makhzani. Action matching: Learning stochastic dynamics from samples. In *International Conference on Machine Learning*, pp. 25858–25889. PMLR, 2023a.

Kirill Neklyudov, Rob Brekelmans, Alexander Tong, Lazar Atanackovic, Qiang Liu, and Alireza Makhzani. A computational framework for solving wasserstein lagrangian flows. *arXiv preprint arXiv:2310.10649*, 2023b.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky TQ Chen. Multisample flow matching: Straightening flows with minibatch couplings. *arXiv preprint arXiv:2304.14772*, 2023.

Stephen Prajna, Ali Jadbabaie, and George J Pappas. A framework for worst-case and stochastic safety verification using barrier certificates. *IEEE Transactions on Automatic Control*, 52(8): 1415–1428, 2007.

Neta Shaul, Ricky TQ Chen, Maximilian Nickel, Matthew Le, and Yaron Lipman. On kinetic optimal probability paths for generative models. In *International Conference on Machine Learning*, pp. 30883–30907. PMLR, 2023.

Joan Sola, Jeremie Deray, and Dinesh Atchuthan. A micro lie theory for state estimation in robotics. *arXiv preprint arXiv:1812.01537*, 2018.

Maojiang Su, Jerry Yao-Chieh Hu, Yi-Chen Lee, Ning Zhu, Jui-Hui Chung, Shang Wu, Zhao Song, Minshuo Chen, and Han Liu. High-order flow matching: Unified framework and sharp statistical rates. *Advances in Neural Information Processing Systems*, 38:45820–45932, 2026.

Jan Tauberschmidt, Sophie Fellenz, Sebastian Vollmer, and Andrew Duncan. Physics-constrained fine-tuning of flow-matching models for generation and inverse problems. In *International Conference on Learning Representations*, volume 2026, pp. 143992–144037, 2026.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *arXiv preprint arXiv:2302.00482*, 2023.

Utkarsh Utkarsh, Pengfei Cai, Alan Edelman, Rafael Gomez-Bombarelli, and Christopher Rackauckas. Physics-constrained flow matching: Sampling generative models with hard constraints. *Advances in Neural Information Processing Systems*, 38:160217–160252, 2026.

Luran Wang, Chaoran Cheng, Yizhen Liao, Yanru Qu, and Ge Liu. Training free guided flow matching with optimal control. *arXiv preprint arXiv:2410.18070*, 2024.

Yuyang Wang, Ahmed A Elhag, Navdeep Jaitly, Joshua M Susskind, and Miguel Angel Bautista. Swallowing the bitter pill: Simplified scalable conformer generation. *arXiv preprint arXiv:2311.17932*, 2023.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.

Jason Yim, Andrew Campbell, Andrew YK Foong, Michael Gastegger, José Jiménez-Luna, Sarah Lewis, Victor Garcia Satorras, Bastiaan S Veeling, Regina Barzilay, Tommi Jaakkola, et al. Fast protein backbone generation with se (3) flow matching. *arXiv preprint arXiv:2310.05297*, 2023a.

Jason Yim, Brian L Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. Se (3) diffusion model with application to protein backbone generation. *arXiv preprint arXiv:2302.02277*, 2023b.

Zhengyu Zhou and Weiwei Liu. An error analysis of flow matching for deep generative modeling. In *Forty-second International Conference on Machine Learning*, 2025.

APPENDIX CONTENTS

A. Extended related work: path design, geometry, control, constraints, approximation, and certification.

B. Broader discussion: where a trajectory-level claim lives in the flow-matching pipeline.

C. Auxiliary statements and complete proofs.

D–E. Experimental protocols and molecular representation diagnostic.

F–G. Supporting controlled results and checkpointed transfer study.

A EXTENDED RELATED WORK

In the main text we keep only the comparisons needed to establish the transfer question. Here we organise neighbouring work by the mathematical object it modifies or analyses. This distinction matters because terms such as path, constraint, physical validity, geometry, and smoothness can refer to different stages of a flow-based generator.

A.1 CONDITIONAL PATHS, ENDPOINT COUPLINGS, AND TRANSPORT DESIGN

Flow matching and stochastic interpolants reduce generative modelling to learning a time-dependent velocity field from conditional probability paths (Lipman et al., 2022; Liu et al., 2022; Albergo & Vanden-Eijnden, 2022; Albergo et al., 2025). This formulation exposes two closely related design choices, namely how source and target samples are coupled and how a conditional trajectory connects each paired endpoint. Optimal-transport conditional flow matching and multisample flow matching use coupling structure to reduce crossing and ambiguity in the regression target (Tong et al., 2023; Pooladian et al., 2023); more recent cluster-wise OT constructions further organise source–target transport to improve efficiency (Chiang et al., 2026). Rectification and kinetic formulations similarly favour straighter or lower-action transport (Liu et al., 2022; Shaul et al., 2023; Bansal et al., 2024). We take these developments as motivation, since they establish the conditional trajectory as an object deliberately engineered for desirable behaviour, not merely a proof device.

We focus on the implication after regression. Squared-error CFM takes a conditional mean velocity at each state and time. At population level this mean reproduces the prescribed family of one-time marginals under standard regularity assumptions, but it does not retain the latent endpoint label or, in general, the joint law across multiple times. Thus straightness, low action, a particular endpoint coupling, or a teacher-path event does not automatically become a property of the marginal generator. We view this as a transfer question rather than a competing path-design objective. A path optimiser can remain useful even when some path-level properties do not transfer, provided we attach the final claim to the appropriate stage.

A.2 GEOMETRY-AWARE AND MANIFOLD FLOW MATCHING

Continuous generative modelling on manifolds has developed through Riemannian continuous normalising flows, score models, and diffusion processes (Mathieu & Nickel, 2020; De Bortoli et al., 2022; Huang et al., 2022). Riemannian Flow Matching extends simulation-free velocity regression to general geometries (Chen & Lipman, 2024), while equivariant flow matching incorporates symmetry directly into the vector field (Klein et al., 2023). Metric Flow Matching goes further by learning or exploiting a metric so that conditional interpolants follow data-adapted routes rather than Euclidean straight lines (Kapuśniak et al., 2024). Molecular and structural generators use related ideas in torsional and geometric diffusion, conformer generation, and $\mathrm{SE}(3)$-structured protein models (Jing et al., 2022; Xu et al., 2022; Ganea et al., 2021; Hassan et al., 2024; Wang et al., 2023; Yim et al., 2023b;a; Bose et al., 2024; Huguet et al., 2024; Jing et al., 2024a).

We use compact Lie groups more narrowly. Rather than proposing a more general manifold generator, we use Lie groups as a controlled non-Euclidean setting in which path construction and left-trivialised velocities remain explicit (Iserles et al., 2000; Sola et al., 2018). This lets us isolate whether a geometric or trajectory-level property engineered into endpoint-conditioned interpolants remains after CFM replaces them by a target-free marginal field. The circle counterexample makes the distinction

especially clean, because it needs no geometric approximation, yet temporal coupling is lost already at exact population level.

A.3 Stochastic control, least action, and state-cost path objectives

Several recent formulations optimise conditional paths through stochastic control, Schrödinger bridges, or action minimisation. Generalized Schrödinger Bridge Matching optimises controlled stochastic paths under kinetic and state costs (Liu et al., 2024); kinetic and action-based formulations penalise unnecessarily curved or energetic trajectories (Shaul et al., 2023; Neklyudov et al., 2023a;b); and Lagrangian or least-action perspectives make path design itself the optimisation problem (Du et al., 2026). These objectives are natural when the scientific goal is genuinely path-dependent, for example when a route should avoid high-energy states or follow plausible dynamics.

Our results do not argue against these objectives; they identify which conclusions need another step after optimisation. If we evaluate a state cost in expectation at each time, its time integral transfers exactly at population level because it depends only on one-time marginals. Quantities that depend on temporal coupling need not transfer. Hitting, winding, path length and a specific endpoint pairing are all of that kind. In this sense, two teacher processes can agree at every marginal time while inducing different trajectory laws. The distinction separates objectives that CFM automatically inherits from those that require deployed-trajectory evidence or additional structure.

A.4 Constraint-aware generation and intervention stage

Constraints can enter at several points in a generative pipeline. Projected, reflected, and mirror diffusion methods modify state-space dynamics to respect feasibility or geometry (Christopher et al., 2024; Fishman et al., 2023; Lou & Ermon, 2023; Liu et al., 2023). Flow-based methods can instead alter training or fine-tuning objectives, as in physics-biased or physics-constrained flow matching (Baldan et al., 2026; Tauberschmidt et al., 2026); add optimal-control guidance at inference (Wang et al., 2024); correct a pretrained flow to satisfy hard constraints (Utkarsh et al., 2026); or solve sampling-time constrained optimisation problems (Liang et al., 2025; Liu et al., 2026). We view these approaches as complementary because they intervene on different objects.

For interpretation, that intervention stage is central. A feasibility property imposed before regression describes conditional teachers; a penalty on the learned field changes fitted dynamics; a projection or controller at sampling changes the deployed process directly. Our stage-by-stage analysis makes these distinctions explicit, and for that reason we avoid treating heterogeneous constraint mechanisms as if they differed by only one modelling choice.

A.5 Approximation, discretisation, and certificates

A separate theory asks how faithfully finite models and numerical solvers reproduce a specified flow. Existing analyses bound velocity estimation, distributional error, or integration error, while high-order and rectified methods seek smoother or more easily integrated trajectories (Zhou & Liu, 2025; Bansal et al., 2024; Su et al., 2026). Our main counterexample occurs upstream of these effects, since even a perfect conditional-mean regressor integrated exactly can preserve every one-time marginal while changing the path law. Finite capacity and discretisation can add further deviations, but they do not explain this first transfer gap.

Our endpoint certificate addresses a different inference problem, namely when an expected path integral can imply a terminal-violation bound. We show that the integral alone is insufficient because cost can concentrate arbitrarily late, while a temporal modulus of the deployed cost process supplies the missing information. This logic aligns with barrier-certificate and safety-verification principles, where guarantees must refer to the dynamics actually executed rather than only to a nominal path or training objective (Prajna et al., 2007). We also compare our integral route with direct endpoint estimation; in the tested setting, direct observations are at least as informative, so we view the certificate as most useful when path cost is independently controlled but terminal events are unavailable.

B BROADER DISCUSSION: WHERE A TRAJECTORY CLAIM LIVES

The paper separates four stages that are often discussed together, namely conditional path and coupling design, exact population marginalisation, finite-capacity regression and numerical sampling. This decomposition is not intended as a taxonomy for its own sake. It identifies the object to which a claim applies and therefore the evidence needed to support it.

Conditional design. At this stage an endpoint pair and, possibly, auxiliary randomness define a teacher trajectory. Path length, action, collision avoidance, and other trajectory statistics can be measured exactly on that conditional object. Improving them is meaningful, but it does not by itself identify the corresponding statistic of the target-free generator.

Population marginalisation. Squared-error CFM replaces conditional velocities by their conditional expectation given the current state and time. Under the assumptions in Section C, this preserves the complete family of one-time marginals. Consequently, fixed-time state statistics and expected additive state–time costs transfer exactly. The circle construction in Proposition 1 shows why this statement cannot be strengthened to equality of path laws without additional assumptions, because opposite teacher motions can cancel locally even though each teacher path moves substantially.

Finite regression. A trained neural field approximates the population field only under the metric controlled by its training objective and model class. In particular, a small L^2 velocity error does not imply a uniform speed cap; Proposition C.1 gives an explicit smooth counterexample. Therefore a temporal modulus needed for endpoint certification must either be enforced or verified on the fitted/deployed dynamics, rather than inherited from bounded teacher velocities by assertion.

Numerical deployment. The actual sampled trajectory is produced by a solver. Solver caps, projection, guidance, or correction can create useful guarantees at this stage even when they were absent earlier. Conversely, coarse integration can violate properties of the continuous field. Our discrete certificate is written for the sampled path for precisely this reason.

This perspective suggests a practical reporting rule. If a paper motivates a method through a path-level property, it should state whether the evidence concerns conditional teachers, the population field, the finite learned field, or sampled trajectories. A transfer theorem is appropriate when a property can be proved to cross stages; otherwise direct measurement at the claimed stage is the clearest evidence. The rule is intentionally neutral about which stage is preferable, since conditional path design, learned-field regularisation and sampling-time correction each solve legitimate problems. The point is to prevent a guarantee attached to one object from being read as a guarantee about another.

C AUXILIARY STATEMENTS AND COMPLETE PROOFS

C.1 CONDITIONAL PATHS AND THEIR VELOCITIES

We use a compact connected Lie group with a bi-invariant metric and a deterministic minimal logarithm branch. Ties at the cut locus are resolved measurably. No smoothness of the logarithm across that locus is assumed. All time derivatives below hold for a fixed endpoint pair, so the selected logarithm ξ is constant.

Proof of Equations (5) *and* (6). Identify the Lie algebra with matrices. Differentiating the exponential series and collecting terms gives $e^{-\eta}\frac{d}{dt}e^{\eta} = \int_0^1 e^{-r\eta}\dot\eta e^{r\eta}\,dr = J_r(\eta)\dot\eta$. The product rule for $\gamma = g_0 e^{t\xi/T}e^{\eta}$ therefore gives

$$\gamma^{-1}\dot\gamma = e^{-\eta}(\xi/T)e^{\eta} + J_r(\eta)\dot\eta,$$

which is Equation (5). The Bernstein basis is nonnegative and sums to one. Since $c_0 = c_m = 0$ and $\|c_j\| < R$, it gives

$$\eta(0) = \eta(1) = 0, \qquad \|\eta(s)\| \le R[1-(1-s)^m - s^m], \qquad \eta'(s) = m\sum_{j=0}^{m-1} B_{j,m-1}(s)(c_{j+1}-c_j).$$

Hence $\|\dot\eta\| \le 2mR/T$. Bi-invariance makes every adjoint an isometry. Its average $J_r(\eta)$ has operator norm at most one, so $\|u\| \le \|\xi\|/T + 2mR/T \le (\mathrm{diam}(G) + 2mR)/T$. Endpoint pinning follows by substituting $s = 0, 1$ into the path. □

C.2 Population matching and the deterministic circle example

The marginal statement uses the following sufficient regularity assumptions. Conditional paths are absolutely continuous and $\mathbb{E}\int_0^T \|u_t\|^2 dt < \infty$. Their initial and terminal laws are μ_0, μ_1. The conditional mean v^* has a representative measurable in time and continuously differentiable in space, with $\int_0^T (\|v_t^*\|_\infty + \|\nabla v_t^*\|_\infty) dt < \infty$. These assumptions give a unique global flow $\Phi_{s,t}$ and justify the duality argument below. They are assumptions for exact population matching, not consequences of finite training loss.

Proof of Equation (9). For $f \in C^1(G)$, the chain rule, integrability and conditioning give

$$\frac{d}{dt}\mathbb{E} f(\gamma_t) = \mathbb{E}\langle \nabla f(\gamma_t), u_t\rangle = \int_G \langle \nabla f(g), v^*(g,t)\rangle \, dp_t^\psi(g)$$

for almost every t. Thus p_t^ψ solves the weak continuity equation with velocity v^* and initial law μ_0. For any fixed $b \le T$, the time-dependent test $f_t(g) = f(\Phi_{t,b}(g))$ satisfies $\partial_t f_t + \langle \nabla f_t, v^*\rangle = 0$. Testing the same equation with f_t shows $\int f\, dp_b^\psi = \int f(\Phi_{0,b}(g))\, d\mu_0(g)$. Time-dependent tests follow by approximation from the stated regularity. Since C^1 functions determine measures on compact G, this identifies $p_b^\psi = (\Phi_{0,b})_\# \mu_0 = q_b^*$, including $b = T$. For bounded measurable ϕ, Fubini now gives equality of the two expected integrals. The same conclusion holds for nonnegative ϕ by Tonelli. Finally, if $\|u_t\| \le B_u$ almost surely, conditional Jensen gives $\|v^*(g,t)\| \le B_u$ for p_t^ψ-almost every g and almost every t. □

Proof of Proposition 1. Let $D = \mathrm{wrap}(X_1 - X_0) \in [-\pi, \pi)$. Independent Haar endpoints imply that D is uniform on $[-\pi, \pi)$ and independent of X_0. Conditional on $D = d$, the translation $X_0 + td/T$ is Haar. Consequently D remains uniform after conditioning on $\gamma_t = x$, and $v^*(x,t) = \mathbb{E}[D/T \mid \gamma_t = x] = 0$. The population flow is stationary and both laws have the same marginals. For a fixed displacement d and a closed arc of length $0 < a \le \pi$, the set of starting points whose geodesic intersects the arc has length $a + |d|$. There is no saturation because $a + |d| \le 2\pi$. Since $\mathbb{E}|D| = \pi/2$, averaging over D gives

$$\mathbb{P}\{\gamma \text{ hits } A\} = \frac{a}{2\pi} + \frac{1}{4}, \qquad \mathbb{P}\{g^* \text{ hits } A\} = \frac{a}{2\pi}.$$

The conditional path's expected length is $\pi/2$, whereas the population length is zero. Given an initially valid state $X_0 \notin A$, their entry probabilities are $1/[4(1 - a/(2\pi))]$ and zero, respectively. The conditional path is exactly Equation (2) with every correction coefficient zero. It is deterministic given (X_0, X_1), apart from the immaterial antipodal tie. Its endpoint coupling is independent; the population coupling is the identity. Marginal preservation does not require these couplings to agree. □

C.3 What fitting and averaging do not control

Proposition C.1 (L^2 risk does not control uniform speed)**.** *Under the uniform law on $\mathbb{S}^1$, there are smooth fields with target field zero whose squared L^2 risk tends to zero and whose maximum speed diverges.*

Proof. Take $\varphi \in C_c^\infty((-1,1))$ with $0 \le \varphi \le 1$, maximum one, and set $v_n(\theta) = n\varphi(n^4\theta)$ in $(-\pi, \pi)$, extended periodically. Its support lies inside that chart, so the extension is smooth. Directly,

$$\|v_n\|_\infty = n, \qquad \mathbb{E}|v_n(X)|^2 = \frac{1}{2\pi n^2}\int_{-1}^{1} \varphi(z)^2\, dz \longrightarrow 0.$$

This disproves a uniform-speed implication, without asserting typical-path or endpoint error. □

Acceleration under averaging. On a flat torus, assume a positive differentiable density p and enough integrable derivatives to differentiate the velocity moments. Put $v = \mathbb{E}[u \mid x, t]$, $a = \mathbb{E}[\dot{u} \mid x, t]$ and $\Sigma = \mathbb{E}[u \otimes u \mid x, t] - v \otimes v$. Differentiating the expectations of $f(x_t)$ and $f(x_t)u_t$ gives $\partial_t p + \nabla \cdot (pv) = 0$ and $\partial_t(pv) + \nabla \cdot (p(v \otimes v + \Sigma)) = pa$. Expanding the latter and substituting the former proves

$$\partial_t v + (v \cdot \nabla)v = a - p^{-1}\nabla \cdot (p\Sigma). \tag{14}$$

For an explicit failure to inherit zero acceleration, on the circle take $\dot{\gamma} = S \in \{-1, 1\}$ with equal sign probabilities and conditional initial densities $(1 + S\epsilon \sin x)/(2\pi)$, where $0 < \epsilon < 1$. Then $p_t = (1 - \epsilon \cos x \sin t)/(2\pi)$ and $v = \epsilon \sin x \cos t/(1 - \epsilon \cos x \sin t)$. At $t = 0$, its material acceleration is $2\epsilon^2 \sin x \cos x$, although each conditional path has zero acceleration. For $T = 1$, these paths are deterministic shortest geodesics between their endpoints.

C.4 Endpoint information from an integral

Proposition C.2 (Failure without common temporal regularity)**.** *Even for a fixed smooth nonnegative potential on a compact Lie group and fixed endpoints, no function $f(x) \to 0$ as $x \downarrow 0$ can satisfy $\mathbb{P}\{V(g_T) \geq 1\} \leq f(\mathbb{E}I(g))$ for all absolutely continuous path laws.*

Proof of Proposition C.2. On $\mathbb{S}^1$ put $V(\theta) = (1 - \cos\theta)/2$. The absolutely continuous path that stays at zero until $T - \delta$ and then moves linearly to π has integral $\delta/2$ and terminal cost one. All paths have the same endpoints. Their deterministic laws therefore have violation probability one while their expected integrals tend to zero. This contradicts every proposed bound $f(\mathbb{E}I) \to 0$. □

Proof of Theorem 2 and the occupation inequality. On the event $H = \{V(g_T) \geq \tau\}$, nonnegativity and the backward modulus give $V(g_{T-s}) \geq (\tau - \omega(s))_+$ for every s. Integrating yields $I \geq A_{\omega,T}(\tau)\mathbf{1}_H$. Taking expectations and truncating at one proves the theorem when $A > 0$; when $A = 0$ only the bound one follows. Likewise $V(g_t) \geq \tau\mathbf{1}\{V(g_t) \geq \tau\}$, integrated over time, gives the occupation inequality. □

Proof of Corollary 3. Lipschitz continuity and path length give $V(g_T) - V(g_{T-s}) \leq L_V \int_{T-s}^{T} \|\dot{g}_t\| dt$. A pathwise speed cap bounds this by $L_V B s$. Integrating $(\tau - Ks)_+$ up to $\min(T, \tau/K)$ gives Equation (11), with $K = 0$ handled directly. Under the pathwise energy bound, Cauchy–Schwarz instead gives $\int_{T-s}^{T} \|\dot{g}_t\| dt \leq \sqrt{Es}$. Writing $c = L_V\sqrt{E}$ and $s_* = \min(T, \tau^2/c^2)$ for $c > 0$,

$$A_{\omega,T}(\tau) = \tau s_* - \frac{2}{3} c s_*^{3/2}.$$

For $c = 0$ it is τT. Both arguments are pathwise before expectations are taken, so an expected-energy penalty does not establish either assumption. □

Proposition C.3 (Information-theoretic sharpness given only a mean and modulus)**.** *Fix T, $\tau > 0$, $m \geq 0$ and a nondecreasing ω. Over all nonnegative measurable cost processes $(c_t)_{t\in[0,T]}$ with $c_T - c_{T-s} \leq \omega(s)$ almost surely and $\mathbb{E}\int_0^T c_t\, dt \leq m$,*

$$\sup \mathbb{P}\{c_T \geq \tau\} = \min\{1, m/A_{\omega,T}(\tau)\},$$

and the supremum is attained, with value one when $A_{\omega,T}(\tau) = 0$. For $\omega(s) = L_V B s$ and $L_V > 0$, paths on the unit circle with cost $V(\theta) = L_V d(\theta, 0)$ and speed at most B attain it for $0 < \tau \leq L_V\pi$. The same holds on the grid for Equation (12)*.*

Proof of Proposition C.3. The preceding floor proves the upper bound for every admissible cost process. For attainment, let $r_t = (\tau - \omega(T - t))_+$ for $t < T$ and $r_T = \tau$. It has integral $A = A_{\omega,T}(\tau)$ and satisfies the backward modulus, since $r_T - r_{T-s} = \min(\tau, \omega(s)) \leq \omega(s)$ for $s > 0$. If $A = 0$, use r with probability one. Otherwise use r with probability $p = \min(1, m/A)$ and the zero process with probability $1 - p$. Its expected integral is $pA \leq m$ and its violation probability is p. Thus the bound is information-theoretically sharp given only (m, ω).

For the speed case on the circle, take $V(\theta) = L_V d(\theta, 0)$, $L_V > 0$, and $\theta_t = (\tau - L_V B(T-t))_+/L_V$. For $0 < \tau \leq L_V\pi$ this path stays in $[0, \pi]$, has speed at most B, and realises the required cost ramp.

The same mixture attains the bound. On a grid, its trapezoidal integral is exactly $A_h^{\mathrm{trap}}(\tau)$, so the discrete bound is also attained with the corresponding mixture weight. These constructions impose no fixed prior or target law. For $E, L_V > 0$, the square-root ramp need not be attainable under energy E. It would require $d(g_{T-s}, g_T) \geq \sqrt{Es}$ for all sufficiently small s, hence $\int_{T-s}^{T} \|\dot{g}_t\|^2 dt \geq E$ by Cauchy–Schwarz. This contradicts absolute continuity of that integral as $s \downarrow 0$. □

C.5 Discrete sampling and confidence bounds

Proof of Equations (12) *and* (13). For either capped integrator, $\|a_k\| \leq B$ and $d(g_{k+1}, g_k) \leq h\|a_k\| \leq Bh$. The triangle inequality gives $d(g_{n-j}, g_n) \leq Bjh$. If $V(g_n) \geq \tau$, Lipschitz continuity gives the floor $V(g_{n-j}) \geq (\tau - L_V Bjh)_+$. Multiplying by the nonnegative trapezoidal weights proves Equation (12). Its denominator is positive for $\tau > 0$ because the terminal weight contributes $h\tau/2$. Each path integral is in $[0, TV_{\max}]$. For N independent paths from a fixed trained model, Hoeffding's inequality bounds its expectation by $\widehat{I} + TV_{\max}\sqrt{\log(1/\alpha)/(2N)}$ with probability $1-\alpha$. The deterministic range permits truncation at $TV_{\max}$. This one event bounds the mean integral, so division by the deterministic denominator holds simultaneously over all thresholds. A union bound with $\alpha = \delta/M$ gives joint coverage for M fixed models. □

Proposition C.4 (Direct estimation dominates the Hoeffding integral bound)**.** *If every sampled path satisfies the enforced cap, then for every* $0 < \tau \leq V_{\max}$ *with* $A_h^{\mathrm{trap}}(\tau) > 0$, $\min\{1, U/A_h^{\mathrm{trap}}(\tau)\} \geq \min\{1, \hat{p}(\tau) + \epsilon\}$ *at equal* α *and* N.

Proof of Proposition C.4. Let $A = A_h^{\mathrm{trap}}(\tau) > 0$ and $\epsilon = \sqrt{\log(1/\alpha)/(2N)}$. The pathwise floor gives $\widehat{I} \geq A\hat{p}(\tau)$, and $A \leq \tau T \leq TV_{\max}$. If $U = TV_{\max}$, the integral band is one and the claim follows. Otherwise $U/A = \widehat{I}/A + (TV_{\max}/A)\epsilon \geq \hat{p}(\tau) + \epsilon$. Truncating preserves this inequality. The one-sided DKW bound gives $1 - F(\tau-) \leq 1 - \widehat{F}_N(\tau-) + \epsilon$ simultaneously for all τ, so the right side is the stated direct band. The comparison is pointwise for these two estimators at equal N, α; it does not compare all possible variance-adaptive estimators. □

Observed speeds. Let S be a path's largest displacement rate over the solver steps. For any $b > 0$, the floor on $\{S \leq b\}$ gives

$$\mathbb{P}\{V(g_n) \geq \tau\} \leq \min\left\{1, \mathbb{P}\{S > b\} + \frac{\mathbb{E}I_h^{\mathrm{trap}}}{A_h^{\mathrm{trap}}(\tau; L_V b)}\right\}. \tag{15}$$

Indeed, split the endpoint event according to $S \leq b$ and bound its first part by $\mathbb{E}I/A$. One-sided DKW bounds the other part by $1 - \widehat{F}_N(b) + \sqrt{\log(1/\alpha)/(2N)}$ simultaneously over b. This permits choosing b from the same sample. Combining it with Hoeffding uses two failure events per model; we allocate $\alpha = 0.01/(2M)$ to each.

D Experimental protocols

Tasks and scores. All paths use $T = 1$. Torus potentials sum periodic ridges $A_i \exp\{\kappa_i(\cos(n_i \cdot \theta - \phi_i) - 1)\}$ with integer n_i. We give the wide and narrow ridge tasks the same uniform prior and four-mode von Mises target, changing only ridge concentration from 6 to 30. We use two-mode source and target laws with two ridges for the basins task. For the eight-dimensional task we use a uniform prior, four target modes, and six coordinate or coupled ridges. For rotations we use a pointing cost $2\exp\{8(\langle Ra, b\rangle - 1)\}$ at the representative geodesic midpoint. We record every distribution and potential parameter in the versioned run record.

Intrinsic empirical W_2 is the primary endpoint-fidelity metric throughout the main text. We solve a balanced assignment with squared geodesic costs and take its square root, using 2000 samples per side on tori and 1000 on rotations. These are empirical distances, with finite-sample reference scales, not estimates of zero error for a perfect generator. Path cost, threshold occupation and terminal violation are distinct observables. The operating threshold is $\tau = 0.25$. MMD is a secondary kernel-dependent diagnostic. Its unbiased squared-MMD U-statistic may be negative, including in the reference–reference comparison in Table 4.

Reference calibration. For each replicate, we draw 4000 fresh independent target samples and split them into two sets of 2000, matching the main W_2 budget. We repeat this 50 times for each distinct target law, independently of the model evaluations. Table 3 separates uncertainty in the mean from variation between individual sample pairs. The latter is described by empirical percentiles, not a confidence interval for the mean or a test of model equivalence. Both ridge tasks use the same calibration because their target laws agree. Seeds, sample hashes and all distances are retained.

Table 3: Reference–reference intrinsic W_2, 50 independent target pairs with 2000 samples per side. The mean's 95% bootstrap interval uses 20 000 resamples. The empirical 2.5th–97.5th percentiles describe variation between individual sample pairs. The two ridge tasks share a target law and one calibration.

Target	Mean	95% CI of mean	Sample-pair percentiles
Both ridges	0.250	[0.229, 0.272]	[0.128, 0.401]
Basins	0.073	[0.066, 0.081]	[0.045, 0.140]
Coupled $\mathbb{T}^8$	0.817	[0.789, 0.844]	[0.642, 1.030]

Training and evaluation. In compact diagnostic labels, *teacher* denotes the conditional paths and *student* the fitted field or its generated trajectories. This shorthand refers to the two stages of CFM regression. For the principal analytic cells, we train ten seeds with a width-256 depth-4 SiLU student, periodic or rotation-matrix state features, and four Fourier time features. We use a width-256 depth-3 amortised teacher with degree-5 corrections and radius 0.6. We train for 3000 teacher and 12 000 student updates with batch size 256, learning rate 10^{-3}, and 16 teacher or 4 student time samples per pair. We select $(\lambda_V, \lambda_K, \lambda_A) = (1, 0.01, 0)$ after a validation pilot that searched $\lambda_V \in \{0, 0.1, 1, 5, 20\}$ and $\lambda_K \in \{0.001, 0.01, 0.1\}$ on one task and seed; we then freeze that choice for all reported cells. OT assignment and its diagnostics use blocks of 256. Principal sampling uses 200 Lie–Euler steps. Cost uses 1000 paths on tori and 800 on rotations; endpoint pools contain 10 000 and 8000 samples before metric subsampling. Radius 1.5 changes the rotation teacher capacity only.

The main effect plots report descriptive paired seed means and marginal 95% percentile intervals from 20 000 bootstrap resamples. Absolute tables use 2000 resamples and report half-widths; the reference calibration uses 20 000 resamples over target pairs. Matched training seed labels do not imply common evaluation draws. These exploratory intervals are not simultaneous tests or evidence of equivalence. The separate transfer study retains its declared sixteen-test family and Holm adjustment (Section G).

Mechanism controls. We implement these mechanism controls in a shared framework; we do not present them as evaluations of released methods on their native benchmarks. RFM uses independent geodesic pairs; OT-RFM changes only pairing. GAOC adds a bounded amortised detour, with either coupling. GSBM-ZN uses an alternating, unbounded per-minibatch spline solve, so its comparison also changes capacity and amortisation. For the additional controls, we use an RBF metric-weighted kinetic objective (MFM-style), a running-cost student objective, conflict-free physical and fitting gradients (PBFM-style), inference-time quadratic control (OC-Flow-style), or stepwise projection (PCFM-style). Shared architecture does not make their compute or path families identical. The repository records their configurations and upstream commit references. Conclusions concern these mechanisms and settings.

Modulus intervention. We leave the principal students unconstrained. For the certificate study, a separate radial output map imposes $B = 2$ or 4, and the wide-ridge sweep tests $B \in \{1, 1.5, 2, 3, 4\}$ at 200 and 400 steps. We obtain certified potential constants from analytic envelopes. For a ridge, maximise $|\sin x| e^{\kappa(\cos x - 1)}$ using $c_* = (\sqrt{1 + 4\kappa^2} - 1)/(2\kappa)$; this gives gradient bound $A\kappa\|n\|\sqrt{1 - c_*^2}\, e^{\kappa(c_* - 1)}$. Sum these bounds over ridges and use $V_{\max} = \sum_i A_i$. The discrete floor uses the actual trapezoidal weights. Integral confidence events have joint failure probability 0.01 over 90 barrier and sweep models. Direct-band comparisons use their declared family of 210 capped models and the same 0.01 family failure probability. Unconstrained observed-speed bands allocate two failure events per model as in Equation (15).

Independent implementation checks. We compare analytic Lie-group velocities with symmetric finite differences, the right Jacobian with numerical integration, and circle hitting with exact segment intersections. We also check endpoint pinning, near-π exp/log round trips, quaternion sign invariance, group membership, and solver convergence on smooth fields. Measurement tests cover constant and ramp integrals, periodic seams, extremal modulus bounds and deliberate cap violations. Inference tests verify that sampling needs neither teacher nor target endpoints. These checks test the implementation independently of the reported method rankings. Audit history and exhaustive diagnostics remain in the repository, outside the submission narrative.

E MOLECULAR REPRESENTATION DIAGNOSTIC

We use this experiment to test a validity measurement, not molecular generation quality. We use the public MDGen tetrapeptide trajectories (Jing et al., 2024b), loaded with MDTraj (McGibbon et al., 2015) and pinned by dataset and file hashes. Before training, we excluded proline, cysteine and histidine, stratified the remaining four-residue peptides by side-chain torsion count, and selected twelve sequences alphabetically within the three strata. We split each trajectory into contiguous training, validation, calibration and test blocks. We cap autocorrelation-based buffers at 5% of trajectory length, so we do not claim that the test frames are independent equilibrium samples.

We freeze bond lengths, angles, and amide and arginine guanidinium torsions at a training template during reconstruction. Modelled angles are non-ring rotatable heavy-atom bonds excluding those conjugated C–N bonds. We compare the same 2000 evenly spaced held-out frames before and after torsion extraction and reconstruction. An overlap is a heavy-atom pair more than three bonds apart with distance below $0.75(r_i + r_j)$, using Bondi radii (Bondi, 1964). The rule is identical for both coordinate sets. None of the sampled raw frames fails it; reconstruction produces rates of 0.041–0.376. The mean held-out reconstruction RMSD is 0.048–0.071 nm. These observations establish a representation artefact under this rule, not a physical-validity guarantee for raw MD or a physical improvement by generated conformers.

Figure 6 selects the first qualifying AADY calibration frame, original index 787700, that passes raw and fails after reconstruction. The most overlapping pair has indices 8 and 18 and changes from 0.293 to 0.221 nm against a 0.230 nm threshold. Modelled torsions agree within 10^{-8} radians. A common aligned projection is used only for display. We save the example coordinates, selection rule, and source hashes. The full twelve-peptide calibration is shown in the main figure. Generated-conformer scores, reference-split calibration and the released torsional-diffusion distribution-shift comparison remain available in the artefact; they do not support an application-performance claim here.

F SUPPORTING CONTROLLED RESULTS

We report the secondary MMD diagnostic for all ten mechanism controls in Table 4, but we keep intrinsic W_2 as the primary fidelity measure when interpreting the main effects. Tables 5 and 6 show how enforcing the modulus changes cost and how the operating-point bound remains uninformative across the full cap and resolution sweep. Every recorded model–threshold comparison of the Hoeffding and empirical-Bernstein integral estimates with the direct band at $N = 1000$ is retained in the evidence report.

F.1 ROTATION CAPACITY

For a degree-5 correction at the midpoint, the coefficient bound in Section C gives $\|\eta(1/2)\| \leq 15R/16$. This bounds the pointing-direction change because a rotation of angle $\|\eta\|$ moves any unit vector through at most that angle. At the representative crossing, we need $\theta > \arccos(1 + \log(0.25/2)/8) = 42.3^\circ$ to reduce $2\exp\{8(\cos\theta - 1)\}$ below 0.25. The radii 0.6 and 1.5 allow at most 32.2° and 80.6°. Thus the smaller radius cannot clear this crossing, whereas the larger radius removes this local obstruction without proving avoidance for every pair. Table 7 reports all methods and both radii, with maximum teacher endpoint error $7.2 \cdot 10^{-7}$. Zero displayed occupation is a finite-grid result.

Table 4: Controlled torus tasks, ten matched training seeds per method and task. Means with 95% bootstrap half-widths from 2000 resamples. MMD_4^2 is periodic-kernel $\mathrm{MMD}^2 \times 10^4$ at $\sigma = 1$, an unbiased U-statistic that may be negative. This secondary diagnostic uses the ten shared-framework mechanism controls. Lower values are preferable; bold marks the lowest observed mean among methods in each task–metric column and does not imply statistical superiority. The independent reference comparison calibrates scale and is excluded from bolding. All generation is target-free.

Method	$\mathbb{T}^2$ bar.		$\mathbb{T}^2$ narr.		$\mathbb{T}^2$ bas.		$\mathbb{T}^8$ coup.	
	I	MMD_4^2	I	MMD_4^2	I	MMD_4^2	I	MMD_4^2
RFM	0.323 ± 0.007	3.1 ± 0.9	0.133 ± 0.004	3.1 ± 0.9	0.002 ± 0.000	8.0 ± 5.4	0.704 ± 0.011	$\mathbf{6.8 \pm 1.0}$
OT-RFM	0.183 ± 0.003	3.1 ± 1.5	0.041 ± 0.002	3.1 ± 1.5	0.003 ± 0.000	$\mathbf{2.5 \pm 1.7}$	0.680 ± 0.010	8.0 ± 1.0
MFM-style	0.197 ± 0.006	3.2 ± 1.6	0.049 ± 0.004	3.1 ± 1.6	0.003 ± 0.000	3.8 ± 2.2	0.677 ± 0.008	7.2 ± 1.4
GSBM-ZN	$\mathbf{0.041 \pm 0.002}$	27.6 ± 10.1	$\mathbf{0.009 \pm 0.001}$	23.8 ± 6.5	0.002 ± 0.000	25.2 ± 14.2	$\mathbf{0.225 \pm 0.003}$	60.6 ± 6.8
CFM+cost	0.304 ± 0.007	5.3 ± 2.4	0.123 ± 0.004	4.4 ± 1.7	0.002 ± 0.000	11.9 ± 4.3	0.657 ± 0.009	8.8 ± 1.0
PBFM-style	0.302 ± 0.010	5.0 ± 2.6	0.125 ± 0.005	5.4 ± 2.4	0.002 ± 0.000	9.9 ± 3.5	0.657 ± 0.008	8.9 ± 1.2
OC-Flow	0.229 ± 0.005	21.6 ± 3.1	0.065 ± 0.003	10.5 ± 1.9	0.002 ± 0.000	12.0 ± 4.1	0.587 ± 0.010	19.6 ± 2.6
PCFM	0.093 ± 0.002	35.8 ± 4.7	0.041 ± 0.001	7.3 ± 1.9	0.002 ± 0.000	7.2 ± 4.9	0.228 ± 0.001	410.0 ± 5.0
GAOC-FM	0.186 ± 0.006	4.5 ± 2.8	0.070 ± 0.005	3.3 ± 1.4	$\mathbf{0.001 \pm 0.000}$	4.9 ± 2.9	0.490 ± 0.005	10.5 ± 2.7
GAOC-FM (OT)	0.064 ± 0.002	$\mathbf{1.9 \pm 1.0}$	0.011 ± 0.001	$\mathbf{1.3 \pm 0.6}$	0.002 ± 0.000	$\mathbf{2.5 \pm 0.9}$	0.434 ± 0.007	11.6 ± 1.8
reference vs. reference	–	-0.21	–	-0.21	–	0.25	–	-0.03

Table 5: Wide-ridge student speed ablation, ten seeds per cell. The radial output map enforces B. A smaller cap can change both cost and endpoint fidelity. Lower I, MMD^2, and terminal cost are preferable; bold marks the lowest observed value in each of those columns (ties included), not a significance claim.

Method	enforced B	I	observed $\mathrm{speed}_{\max}$	MMD^2	$V(g_T)$
GAOC-FM (OT) (ours)	2.0	0.084 ± 0.002	2.00	$4.28 \cdot 10^{-4}$	0.0788
GAOC-FM (OT) (ours)	4.0	$\mathbf{0.064 \pm 0.002}$	3.61	$2.87 \cdot 10^{-4}$	**0.0587**
GAOC-FM (OT) (ours)	–	$\mathbf{0.064 \pm 0.002}$	3.79	$\mathbf{1.92 \cdot 10^{-4}}$	**0.0587**
RFM (geodesic CFM)	2.0	0.336 ± 0.007	2.00	$1.26 \cdot 10^{-3}$	0.0860
RFM (geodesic CFM)	4.0	0.325 ± 0.007	3.50	$2.37 \cdot 10^{-4}$	0.0630
RFM (geodesic CFM)	–	0.323 ± 0.007	3.78	$3.12 \cdot 10^{-4}$	0.0619

Table 6: Operating-point sweep, five seeds per cap and step count. $\tau^\star$ is the first grid threshold with confidence bound below one; its range is over seeds. A single 1% family failure probability covers these 50 CPU models and the 40 capped CUDA barrier models in Figure 5. Threshold resolution is approximately 0.005. No bound is informative at $\tau = 0.25$.

B	Steps	$\widehat{I}$	$\tau^\star$ range	Informative seeds
1	200	0.1564	1.290–1.320	5/5
1	400	0.1567	1.290–1.320	5/5
1.5	200	0.1086	1.449–1.479	5/5
1.5	400	0.1088	1.449–1.479	5/5
2	200	0.0841	1.593–1.613	5/5
2	400	0.0842	1.593–1.613	5/5
3	200	0.0653	1.861–1.896	5/5
3	400	0.0654	1.861–1.896	5/5
4	200	0.0637	–	0/5
4	400	0.0638	–	0/5

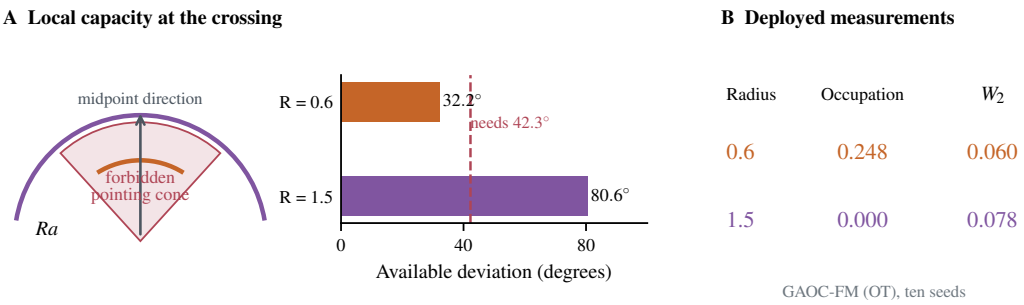


Figure 7: A larger correction removes a local capacity obstruction, but fidelity can worsen. Left, a schematic section of the forbidden pointing cone and the available angular deviations at the representative crossing. The default radius cannot supply the required 42.3°. Right, deployed ten-seed means for GAOC-FM (OT) show lower occupation but higher intrinsic endpoint W_2 at the larger radius. The printed zero is a finite-grid measurement. Full results and uncertainty appear in Table 7.

Table 7: Collision-aware rotations on the quaternion-based logarithm, ten seeds at each radius. Entries give means and standard deviations except MMD^2, which is a mean of unbiased U-statistics and may be negative. W_2 is intrinsic, using 1000 endpoints per side. All rows are mechanism controls. Only the GAOC-FM teachers depend on the correction radius, so the $R = 1.5$ block lists those two; every other method is identical at both radii. The central crossing requires 42.3° deviation; the correction can supply at most 32.2° at $R = 0.6$ and 80.6° at $R = 1.5$. GSBM-ZN has lower occupation and path cost than GAOC-FM (OT) at $R = 0.6$. Maximum teacher endpoint error is $7.2 \cdot 10^{-7}$. Lower values are preferable. Bold marks the lowest observed mean within each radius block and metric (terminal violation is not ranked because its displayed values tie); it is descriptive only. Printed zeros describe finite-grid measurements.

Radius	Method	$\mathrm{occ}_{0.25}$	term-viol	I	W_2	MMD^2
0.6	GAOC-FM	0.349 ± 0.021	0.000 ± 0.001	0.2311 ± 0.0238	0.059 ± 0.007	$1.7 \cdot 10^{-3}$
	GAOC-FM (OT)	0.248 ± 0.004	0.000 ± 0.001	0.1614 ± 0.0020	0.060 ± 0.009	$1.5 \cdot 10^{-3}$
	GSBM-ZN	$\mathbf{0.008 \pm 0.005}$	0.000 ± 0.001	$\mathbf{0.0269 \pm 0.0045}$	0.087 ± 0.022	$2.3 \cdot 10^{-3}$
	OC-Flow	0.463 ± 0.003	0.000 ± 0.000	0.4048 ± 0.0073	0.453 ± 0.012	$1.8 \cdot 10^{-1}$
	OT-RFM	0.605 ± 0.002	0.000 ± 0.000	0.6984 ± 0.0037	$\mathbf{0.052 \pm 0.004}$	$\mathbf{6.4 \cdot 10^{-4}}$
	PCFM	0.741 ± 0.016	0.528 ± 0.156	0.2356 ± 0.0018	1.547 ± 0.060	$1.1 \cdot 10^{0}$
	RFM	0.613 ± 0.003	0.000 ± 0.000	0.7243 ± 0.0026	0.057 ± 0.012	$9.3 \cdot 10^{-4}$
1.5	GAOC-FM	$\mathbf{0.000 \pm 0.000}$	0.000 ± 0.000	$\mathbf{0.0137 \pm 0.0009}$	0.088 ± 0.026	$9.1 \cdot 10^{-3}$
	GAOC-FM (OT)	$\mathbf{0.000 \pm 0.000}$	0.000 ± 0.000	0.0142 ± 0.0005	$\mathbf{0.078 \pm 0.023}$	$\mathbf{6.2 \cdot 10^{-3}}$

G CHECKPOINTED TRANSFER STUDY

We recorded the protocol and primary family before running the full grid. In Study 1 we train OT-RFM and GAOC-FM (OT) on all four torus tasks with seeds 0–9 and width 256. In Study 2 we freeze each wide-ridge teacher across student widths 64, 128, 192 and 256 while keeping the training stream and update count fixed. Equal updates do not imply equal compute. All 140 models have saved checkpoints whose reload reproduces the deployed integral evaluation.

For each model we evaluate 1024 paths on a common 800-interval observation grid. Student starting draws are shared between methods and solver variants. Teacher endpoint pairs use a separate random stream, with OT blocks of 256. Thus teacher–student differences compare seed-level means from the same initial law, not trajectories with identical initial states. Teachers are also checked on 1600 intervals. The student reference doubles Lie–Heun steps from 400 until two successive refinements change the mean endpoint by less than 10^{-3} rad, mean cost by less than 1% (or 10^{-5} near zero), and hit classification on fewer than 1% of paths at each threshold. All models satisfy these tests by 1600 steps. This controls the reported numerical comparisons, not continuous-time event detection between observation nodes. Length and open-path displacement are not winding numbers.

Our sixteen declared two-sided tests compare method means and teacher–student gaps in integral and ever-hit probability at $\tau = 0.25$ for each task. Table 8 retains every test, with Holm adjustment and 20 000-resample matched-seed bootstrap intervals. The displayed difference is student minus teacher; reversing its sign leaves the two-sided test unchanged. Ever-hit includes $t = 0$. Its method contrast is zero on both ridge tasks, where shared initially invalid student states account for the hits. The post-hoc entry statistic in the artefact is $\mathbb{P}\{V(g_0) < \tau,\ \max_k V(g_k) \geq \tau\}$, a joint probability. It is not divided by the valid-start probability. The confirmatory study below tests the conditional quantity on new seeds.

Table 8: Declared primary family of the transfer study, ten matched training seeds, student evaluated with its converged same-field reference. Differences are seed means with 20 000-resample seed-bootstrap intervals; p is the exact two-sided sign-flip test and p_H is Holm-adjusted over the sixteen tests. "Ours" is GAOC-FM (OT); student-minus-teacher rows concern GAOC-FM (OT). Student minus teacher gaps in a path event can contain population non-transfer as well as finite-fit error.

Task	Contrast	Difference [95% CI]	p	p_H
$\mathbb{T}^2$ wide	I, ours − OT-RFM	-0.1146 $[-0.1173, -0.1118]$	0.002	0.031
$\mathbb{T}^2$ wide	$\text{hit}_{0.25}$, ours − OT-RFM	$+0.0000$ $[+0.0000, +0.0000]$	1.000	1.000
$\mathbb{T}^2$ wide	I, student − teacher	$+0.0018$ $[-0.0003, +0.0040]$	0.145	1.000
$\mathbb{T}^2$ wide	$\text{hit}_{0.25}$, student − teacher	-0.0001 $[-0.0143, +0.0147]$	1.000	1.000
$\mathbb{T}^2$ narrow	I, ours − OT-RFM	-0.0294 $[-0.0308, -0.0280]$	0.002	0.031
$\mathbb{T}^2$ narrow	$\text{hit}_{0.25}$, ours − OT-RFM	$+0.0000$ $[+0.0000, +0.0000]$	1.000	1.000
$\mathbb{T}^2$ narrow	I, student − teacher	$+0.0009$ $[+0.0001, +0.0019]$	0.092	0.918
$\mathbb{T}^2$ narrow	$\text{hit}_{0.25}$, student − teacher	-0.0031 $[-0.0126, +0.0069]$	0.576	1.000
$\mathbb{T}^2$ basins	I, ours − OT-RFM	-0.0009 $[-0.0010, -0.0008]$	0.002	0.031
$\mathbb{T}^2$ basins	$\text{hit}_{0.25}$, ours − OT-RFM	-0.0001 $[-0.0004, +0.0002]$	1.000	1.000
$\mathbb{T}^2$ basins	I, student − teacher	$+0.0000$ $[-0.0001, +0.0001]$	0.496	1.000
$\mathbb{T}^2$ basins	$\text{hit}_{0.25}$, student − teacher	$+0.0002$ $[-0.0002, +0.0007]$	0.750	1.000
$\mathbb{T}^8$	I, ours − OT-RFM	-0.2395 $[-0.2427, -0.2360]$	0.002	0.031
$\mathbb{T}^8$	$\text{hit}_{0.25}$, ours − OT-RFM	-0.0021 $[-0.0043, +0.0004]$	0.135	1.000
$\mathbb{T}^8$	I, student − teacher	$+0.0139$ $[+0.0074, +0.0205]$	0.006	0.064
$\mathbb{T}^8$	$\text{hit}_{0.25}$, student − teacher	-0.0119 $[-0.0146, -0.0090]$	0.002	0.031

Confirmatory study. Before any confirmatory run, we committed a second protocol (`TRANSFER-CONFIRM-PREREG.md`). We repeat Study 1 on seeds 10–19 with the same configuration, but evaluate each student from its teacher's own 1024 starting states so that teacher and student are paired path by path. The primary quantity is conditional entry, the fraction of paths starting below τ whose cost later reaches τ on the shared grid. The declared family is the GAOC-FM (OT) student-minus-teacher difference at $\tau \in \{0.5, 1\}$ on each task, eight exact sign-flip tests with Holm adjustment. All 80 references converged and all reloads reproduced their evaluations. OT-RFM and $\tau = 0.25$ are secondary and appear in the evidence report.

Table 9: Confirmatory transfer study, declared before its runs: new seeds 10–19, each GAOC-FM (OT) student rolled out from its teacher's own starting states. Conditional entry is the fraction of paths starting below τ that later reach it. Means over ten seeds; the difference is student minus teacher with a seed-bootstrap interval; p is the exact sign-flip test and p_H is Holm-adjusted over the estimable tests.

Task	τ	Teacher	Student	Difference [95% CI]	p	p_H
$\mathbb{T}^2$ wide	0.5	0.000	0.000	-0.000 $[-0.001, +0.000]$	1.000	1.000
$\mathbb{T}^2$ wide	1	0.024	0.005	-0.019 $[-0.023, -0.015]$	0.002	0.016
$\mathbb{T}^2$ narrow	0.5	0.006	0.000	-0.006 $[-0.010, -0.003]$	0.008	0.039
$\mathbb{T}^2$ narrow	1	0.010	0.005	-0.006 $[-0.008, -0.004]$	0.002	0.016
$\mathbb{T}^2$ basins	0.5	0.000	0.000	$+0.000$ $[+0.000, +0.000]$	1.000	1.000
$\mathbb{T}^2$ basins	1	0.000	0.000	$+0.000$ $[+0.000, +0.000]$	1.000	1.000
$\mathbb{T}^8$	0.5	0.853	0.801	-0.051 $[-0.058, -0.044]$	0.002	0.016
$\mathbb{T}^8$	1	0.179	0.163	-0.016 $[-0.023, -0.008]$	0.010	0.039

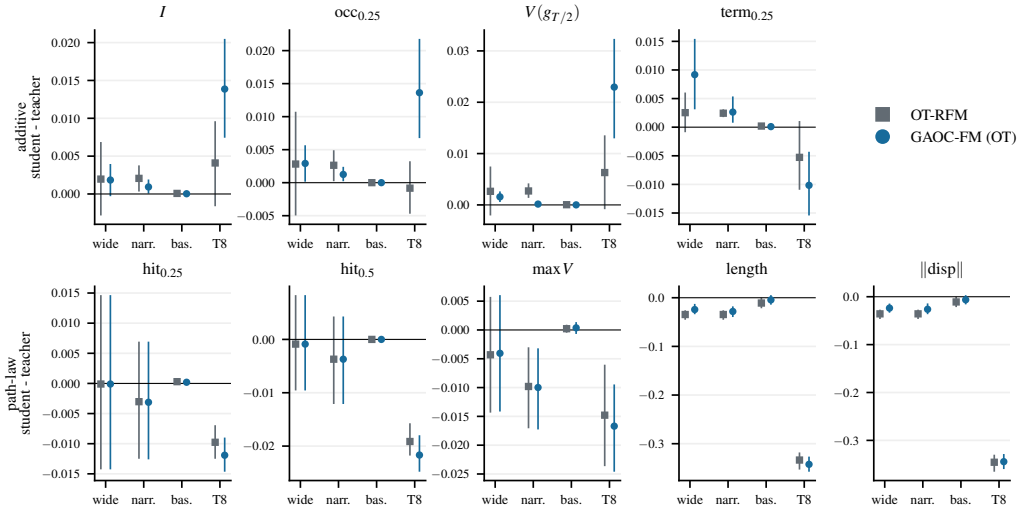


Figure 8: Exploratory student-minus-teacher differences, with marginal 95% seed-bootstrap intervals. Top, additive and fixed-time expectations preserved by exact population matching. Bottom, path-dependent quantities. The two evaluations use separate initial draws from the same prior. A path-event gap can combine population non-transfer, finite fitting and sampling error.

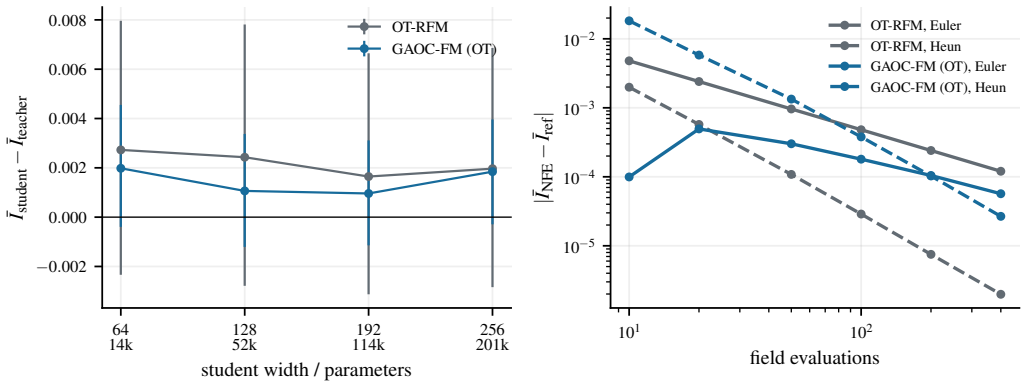


Figure 9: Wide-ridge controls. Left, cost gaps across student capacities with frozen teachers. This does not separate approximation, optimisation and generalisation. Right, error in mean cost relative to the converged field rollout, using common initial states. The plotted error is absolute; the GAOC-FM (OT) Euler error changes sign between 10 and 20 evaluations, so the small value at 10 is cancellation rather than a convergence result.