SEMIGROUP RIEMANNIAN TRANSPORT FLOW FOR FEW-STEP CRYSTAL GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

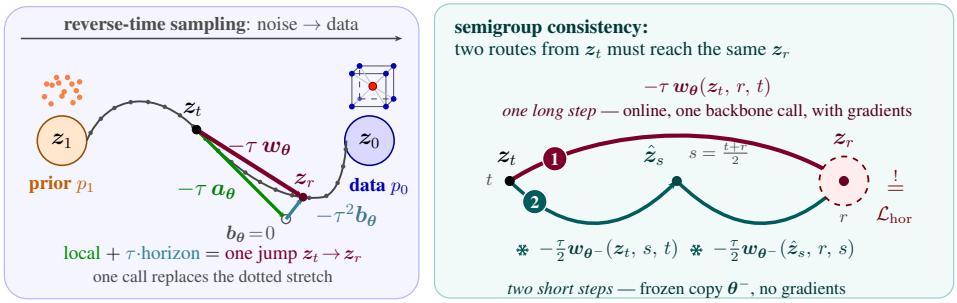
Deep generative models are the default engine for inorganic crystal design, but diffusion and flow models spend 50–5000 neural function evaluations per structure, which dominates the cost of screening campaigns, closed-loop laboratories and inverse problems that need many samples per query. The obstacle is what the model learns: flow matching predicts an *instantaneous* velocity, whereas one large step needs the *finite-horizon* velocity—the average of that velocity over the interval the step covers. Supervising it is the hard part: distillation needs a converged teacher, and derivative-based self-consistency objectives differentiate through the model, requiring a Jacobian–vector product (JVP). The finite-horizon velocity is instead characterized by a composition law—one long transport step equals two composed shorter steps—together with a boundary condition that pins the field down as the step length goes to zero, and on the flat product manifold used for crystal generation the composition law holds *exactly*. It can be imposed with ordinary forward evaluations of a frozen copy of the model, with no teacher and no derivative through the model. This yields **Semigroup Riemannian Transport Flow (SRTF)**: a semigroup consistency loss supplies the composition law, a standard flow-matching loss supplies the boundary condition, and the field is parameterized as a local velocity plus a correction proportional to the step length, so the two can be read and tested separately. We prove that the objective's population fixed point recovers the transport coefficients to leading order, so the correction is identified rather than merely fitted, and on trained networks the measured and predicted residual coefficients agree within 3%. On the four standard crystal-structure-prediction benchmarks SRTF exceeds the single-sample match rate of the published baselines considered here with only two to five sampling steps, 10–$25\times$ fewer than the fastest of them, at flow-matching training memory, whereas our JVP implementation trains at only a third of the batch size on the same GPU memory.

1 INTRODUCTION

Designing new inorganic crystals is a search over an enormous combinatorial space, and deep generative models have become the default engine for it. The scale of deployment is what makes their cost matter: screening campaigns generate structures by the million, autonomous laboratories propose and synthesize candidates continuously (Szymanski et al., 2023), and inverse problems such as solving a structure from a diffraction pattern need many samples per query (Guo et al., 2025; Lai et al., 2025; Li et al., 2025; Johansen et al., 2026). Per-structure sampling cost, not model quality alone, sets what is practically reachable in large-scale computational materials discovery.

Generative models for crystals work well. Diffusion on the crystal product manifold (Xie et al., 2022; Jiao et al., 2023; Zeni et al., 2025), Riemannian flow matching (Miller et al., 2024; Luo et al., 2025) and constraint-aware variants (Okabe et al., 2026; Jiao et al., 2024; Ruple et al., 2026) recover known structures at high rates, as do transformer- and LLM-based generators (Cao et al., 2025; Gruver et al., 2024; Lu et al., 2025; Bone et al., 2025), whose autoregressive decoding has a different cost profile. Diffusion and flow-based crystal generators also share a practical bottleneck: each sample can require 50–5000 neural function evaluations (NFEs), because what the model learns is an *instantaneous* field that must be integrated along a trajectory in many small steps to keep discretization error low.

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.



(a) Geometric intuition. (b) Semigroup supervision.

Figure 1: **SRTF framework.** *(a)* A flow-matching sampler walks the dotted stretch one small step at a time; SRTF crosses it in one jump. One backbone call at z_t returns a local velocity a_θ and a horizon correction b_θ: over a horizon τ the tangent term alone leaves the flow line (hollow marker), and the displacement term $-\tau^2 b_\theta$ bends it back onto z_r. *(b)* The horizon loss $\mathcal{L}_{\text{hor}}$ imposes the composition law as two routes out of one state—one long step of the online model against two short steps of a frozen copy (teal) via $\hat{z}_s$—Eq. (7); the local loss $\mathcal{L}_{\text{loc}}$ supplies the boundary condition.

The open challenge is learning what a large step actually needs. Taking one large reverse-time step from t down to r requires not the instantaneous velocity at t but the *average* velocity along the generative flow over the whole interval that step covers,

$$u_\star(z_t, r, t) \;=\; \frac{1}{t-r}\int_r^t v_\star(z_s, s)\,\mathrm{d}s, \tag{1}$$

where z_t is the state at time t, $\{z_s\}$ the trajectory through it, and $v_\star$ the instantaneous velocity field that flow matching learns. We call $u_\star$ the *finite-horizon velocity* and $\tau = t - r$ the *horizon*; as $\tau \to 0$, $u_\star$ reduces to $v_\star$. Learning $u_\star$ rather than $v_\star$ is what makes few-step sampling possible, and supervising it is the hard part, because it is defined by an integral along a trajectory unavailable at training time. Distillation-style methods (Song et al., 2023; Geng et al., 2025b; Lu & Song, 2025) typically need a pretrained teacher and a separate stage; teacher-free alternatives (Frans et al., 2025; Zhou et al., 2025; Woo et al., 2026) avoid that. Derivative-based self-consistency methods, of which MeanFlow (Geng et al., 2025a) is the representative, avoid the teacher but differentiate *through* the model, requiring a Jacobian–vector product (JVP) at every step. The JVP costs time, which faster kernels can mitigate (Zheng et al., 2026), but in our implementation it also roughly triples training memory, which they cannot—and memory is what bounds the model and batch size that fit on a device. On the largest crystal benchmark we study, our JVP implementation fails to converge, while the same backbone trains without incident under a JVP-free objective.

We supervise the composition law instead: as in shortcut models and semigroup Riemannian MeanFlow (Frans et al., 2025; Woo et al., 2026), the finite-horizon velocity is characterized by an algebraic identity that can be imposed with ordinary forward evaluations. On the flat product manifold used for crystal generation, transport displacements over adjacent intervals simply add, $z_t - z_r = (z_t - z_s) + (z_s - z_r)$ for $r < s < t$, so one long transport step must equal the composition of two shorter ones—a *semigroup* property of the transport map. Together with the zero-horizon boundary condition $u_\star(z, t, t) = v_\star(z, t)$, it pins $u_\star$ down completely:

The SRTF principle: a finite-horizon transport field is characterized by its semigroup composition law together with its zero-horizon boundary condition.

Both conditions are imposable from ordinary forward passes—the composition against a frozen copy of the model, the boundary condition by a standard flow-matching loss (Figure 1)—so neither a JVP nor a teacher appears anywhere. To make the objective cheap and to expose what the horizon supervision learns, we parameterize the learned stand-in for $u_\star$—the field w_θ the sampler actually calls—as a local velocity plus a first-order correction in the horizon,

$$w_\theta(z_t, r, t) \;=\; a_\theta(z_t, t) \;+\; \tau\, b_\theta(z_t, t), \tag{2}$$

where a_θ and b_θ are two lightweight readouts, with parameters θ, on one shared CSPNet backbone (Miller et al., 2024), adding $+0.45\%$ parameters: a_θ carries the local velocity, b_θ the correction

whose size grows with the horizon. Because r enters only through the scalar τ and never through the backbone, one forward pass serves both the model output and part of its own training target. This yields **Semigroup Riemannian Transport Flow (SRTF)**.

Contributions. **(1)** An identifiable JVP-free semigroup objective: the semigroup law plus its boundary condition characterizes $\boldsymbol{u}_\star$, holds *exactly* on the flat crystal manifold, and has a population fixed point that recovers the first-order transport coefficients to leading order, so the horizon head is identified rather than merely fitted (Proposition 1). **(2)** A structured first-order representation, Eq. (2), that makes the objective efficient and interpretable. **(3)** Few-step crystal generation with the source of the gain isolated: on four crystal structure prediction (CSP) benchmarks SRTF exceeds the single-sample match rates of the published baselines considered here in two to five steps, 10–$25\times$ fewer than the fastest of them, a matched-backbone factorial attributes the gain to the objective, and the objective trains at the same batch and width as plain flow matching, whereas our JVP implementation trains at only a third of the batch size.

2 Related work

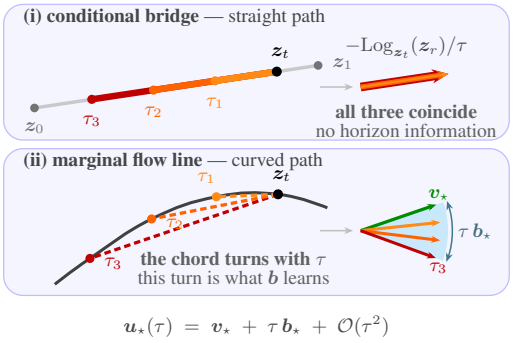
Crystal generative models. Periodic crystals are generated by diffusion, flow and transformer/LLM models—CDVAE (Xie et al., 2022), DiffCSP (Jiao et al., 2023), MatterGen (Zeni et al., 2025), FlowMM (Miller et al., 2024)—with symmetry-aware, space-group-informed and constraint-aware extensions (Luo et al., 2025; Cao et al., 2025; Gruver et al., 2024; Sriram et al., 2024; Ruple et al., 2026; Jiao et al., 2024; Okabe et al., 2026); the diffusion and flow models among them need 50–5000 evaluations per sample. **Few-step generation.** Cutting that budget in Euclidean image space splits into distillation, which typically needs a teacher (Song et al., 2023; Geng et al., 2025b; Lu & Song, 2025), and distillation-free methods: shortcut models (Frans et al., 2025), which impose a self-composition consistency on a monolithic flow map, and MeanFlow (Geng et al., 2025a) and its successors (Zhang et al., 2025; Huang et al., 2026; Lu et al., 2026), which parameterize the finite-horizon average velocity through an identity requiring a Jacobian–vector product through the model; Riemannian MeanFlow (Woo et al., 2026; Zhong et al., 2026) carries this family to manifolds, with a JVP-free semigroup form in Woo et al. (2026). **Positioning.** SRTF shares the composition principle of shortcut models and semigroup Riemannian MeanFlow and differs in two ways: it targets the crystal manifold, where composing through the exponential and logarithm maps $\mathrm{Exp}/\mathrm{Log}$ is displacement addition, and it acts on a structured field whose fixed point identifies, to leading order, the Taylor coefficients of the ideal horizon-averaged velocity, rather than on a monolithic map; its JVP-free horizon loss keeps the training graph first-order. Appendix B and Table 3 give the full discussion, comparing learned target, manifold, JVP use and head identifiability.

3 Semigroup Riemannian Transport Flow

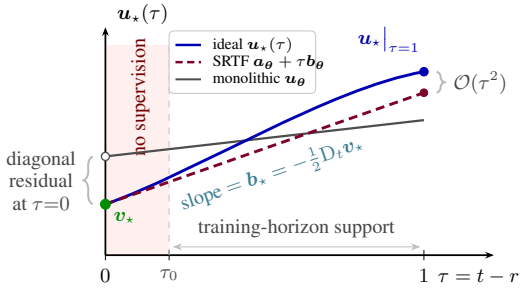
Setting. A crystal with n atoms—chemical identities, fractional coordinates and lattice parameters—is a point $\boldsymbol{z}$ of the product manifold $\mathcal{M} = \mathcal{A} \times \mathcal{F} \times \mathcal{L}^{\varphi}$ of Miller et al. (2024): analog-bit types in $\mathcal{A} = \mathbb{R}^{7n}$ (Chen et al., 2023), coordinates on the flat $3n$-torus $\mathcal{F}$, and the lattice in an unconstrained chart φ under which $\mathcal{L}^{\varphi}$ is isometric to $\mathbb{R}^6$. Every factor is flat, so Christoffel symbols vanish, $\mathrm{Exp}_{\boldsymbol{z}}(\boldsymbol{v}) = \boldsymbol{z} + \boldsymbol{v}$ (mod 1 on $\mathcal{F}$), and—the property the objective rests on—**finite jumps compose exactly**, with no parallel-transport or curvature correction. Flow matching (Lipman et al., 2023; Chen & Lipman, 2024) learns the marginal velocity $\boldsymbol{v}_\star(\boldsymbol{z}_t, t) = \mathbb{E}[\boldsymbol{v}^{\mathrm{cond}} \mid \boldsymbol{z}_t]$, the conditional average of the velocity $\boldsymbol{v}^{\mathrm{cond}}$ of the sampled bridge path, and integrates it in 50–200 steps; few-step sampling instead needs $\boldsymbol{u}_\star$, its average along the flow line over $[r, t]$. MeanFlow (Geng et al., 2025a) targets $\boldsymbol{u}_\star$ through the identity $\boldsymbol{u}_\star = \boldsymbol{v}_\star - \tau\, \mathrm{D}_t \boldsymbol{u}_\star$, where $\mathrm{D}_t = \partial_t + \boldsymbol{v}_\star \cdot \nabla$ is the material derivative along the flow; it requires a JVP, and SRTF does not. Training draws $\boldsymbol{z}_0 \sim p_0$ (data) and $\boldsymbol{z}_1 \sim p_1$ (prior) along $t\colon\ 0 \to 1$ and generation integrates $-\boldsymbol{w}_{\boldsymbol{\theta}}$ backward from $t{=}1$, with t the endpoint time, $r \le t$ the start of the interval and $\tau = t - r$ its horizon (Appendix A.1).

3.1 Transport expansion on the crystal manifold

On the flat $\mathcal{M}$ the covariant derivative reduces to that material derivative and the $\mathrm{Exp}/\mathrm{Log}$ maps act componentwise. Taylor-expanding $\boldsymbol{v}_\star$ along the backward flow line ending at $\boldsymbol{z}_t$ gives $\boldsymbol{v}_\star(\boldsymbol{z}_s, s) = \boldsymbol{v}_\star(\boldsymbol{z}_t, t) - (t-s)\mathrm{D}_t \boldsymbol{v}_\star(\boldsymbol{z}_t, t) + \mathcal{O}((t-s)^2)$; integrating over $s \in [r, t]$ with $\int_r^t (t-s)\,\mathrm{d}s = \tau^2/2$



(a) Does the chord depend on the horizon?



(b) One field, two regimes.

Figure 2: **Why the horizon correction needs flow-based supervision.** *(a)* Normalized chords $-\mathrm{Log}_{z_t}(z_r)/\tau$ at three growing horizons, redrawn from a common origin (gray arrow). **(i)** On the conditional bridge all three *coincide*: the target carries no information about τ, and a head supervised by it collapses to $b \equiv 0$ (Lemma 2). **(ii)** Along the curved marginal flow line they fan out from the tangent $v_\star$ (green); that rotation, to first order the shaded wedge $\tau b_\star$, is what the semigroup objective supervises. *(b)* In scalar form: SRTF's expansion (purple) matches $u_\star$ in value *and* slope at τ=0, where a monolithic field fitted only on horizons $\tau \geq \tau_0$ (gray) mis-specifies the diagonal.

yields the structured *transport expansion*:

$$u_\star(z_t, r, t) = \underbrace{a_\star(z_t, t) + \tau\, b_\star(z_t, t)}_{=:\, \bar{u}_{\mathrm{SRTF}}(z_t, r, t)} + \mathcal{O}(\tau^2), \tag{3}$$

with *local coefficient* $a_\star(z_t, t) = v_\star(z_t, t)$ and *first-order horizon coefficient* $b_\star(z_t, t) = -\frac{1}{2}\mathrm{D}_t v_\star(z_t, t)$, both in $T_{z_t}\mathcal{M}$ and neither depending on r. We write $\bar{u}_{\mathrm{SRTF}}$ for this first-order truncation, which the implemented field w_θ parameterizes and regresses against; its residual $u_\star - \bar{u}_{\mathrm{SRTF}} = \mathcal{O}(\tau^2)$ is the truncation term in Theorem 1 below (Proposition 2). The expansion constrains the scale profile γ multiplying $b_\star$ only through its value and slope at zero, so any smooth γ with $\gamma(0) = 0, \gamma'(0) = 1$ gives the same first-order expansion; we use $\gamma(\tau) = \tau$ and ablate alternatives in Appendix F.4. The general-manifold version is in Appendix C.7.

3.2 Factor-wise instantiation and parameterization

We instantiate the expansion componentwise across the three factors, writing $v_A^{\mathrm{cond}}, v_F^{\mathrm{cond}}, v_L^{\mathrm{cond}}$ for their conditional velocities along the interpolant z_t; the only factor-specific subtlety is translation invariance on $\mathcal{F}$, whose targets are mean-subtracted. Interpolants, the torus logarithm and the Exp maps are tabulated in Appendix D.1.[1]

Coefficient-field parameterization. A shared CSPNet backbone Φ_θ (Miller et al., 2024; Satorras et al., 2021) carries two lightweight readouts a_θ, b_θ of $\Phi_\theta(z_t, t)$, both tangent vectors in $T_{z_t}\mathcal{M}$ (mean-subtracted on the $\mathcal{F}$ block), giving the SRTF field

$$w_\theta(z_t, r, t) = a_\theta(z_t, t) + \tau\, b_\theta(z_t, t), \qquad \tau = t - r. \tag{4}$$

The start time r enters only through the scalar τ and never through the backbone, so one shared evaluation serves both coefficients; the second readout adds $+0.45\%$ parameters.

Relation to baselines. Setting $b_\theta \equiv 0$ recovers Riemannian flow matching (Chen & Lipman, 2024; Miller et al., 2024); replacing the pair by a monolithic $u_\theta(z, r, t)$ and the semigroup target by the JVP identity recovers a Riemannian MeanFlow. Section 4.3 separates the two differences.

[1] Like FlowMM (Miller et al., 2024), SRTF supervises a_θ with the conditional velocity along the sampled interpolant rather than the marginal target; the proxy is exact in expectation (Lipman et al., 2023).

3.3 TRAINING AND SAMPLING

SRTF is trained with two regression losses on the shared backbone: a local flow-matching loss on $\boldsymbol{a}_{\boldsymbol{\theta}}$ and a semigroup consistency loss on $\boldsymbol{w}_{\boldsymbol{\theta}}$. Neither differentiates through the network to build its target—both are ordinary tangent vectors, from bridge samples and from *gradient-free evaluations* of a frozen copy of the model, which is the practical difference from MeanFlow-style training.

(L1) Local flow-matching loss on $\boldsymbol{a}_{\boldsymbol{\theta}}$. The local target is the conditional velocity $\bar{\boldsymbol{v}}_c^{\mathrm{cond}}$ of factor c (mean-subtracted on $\mathcal{F}$; Table 4), measured in the tangent space:

$$\mathcal{L}_{\mathrm{loc}}(\boldsymbol{\theta}) \;=\; \mathbb{E}_{t,\boldsymbol{z}_0,\boldsymbol{z}_1} \sum_{c\in\{A,F,L\}} \tfrac{\omega_c}{\dim_c} \left\| \boldsymbol{a}_{\boldsymbol{\theta}}^c(\boldsymbol{z}_t,t) - \bar{\boldsymbol{v}}_c^{\mathrm{cond}} \right\|_c^2, \tag{5}$$

where $\dim_c \in \{7n, 3n, 6\}$ is the factor dimension, $\omega_c > 0$ are the per-factor weights (code release), and the norm is induced by the factor metric. This fixes the zeroth-order coefficient of (3).

What can supervise the horizon coefficient? $\boldsymbol{b}_{\boldsymbol{\theta}}$ needs a signal that varies with τ, and the obvious JVP-free candidate does not provide one. On the constant-speed conditional bridge the chord $-\mathrm{Log}_{\boldsymbol{z}_t}(\boldsymbol{z}_r)/\tau$ equals $\boldsymbol{z}_1 - \boldsymbol{z}_0$ for every sample, so its conditional expectation is $\boldsymbol{v}_\star(\boldsymbol{z}_t,t)$ whatever the horizon, and any horizon loss regressing on it is optimized at $\boldsymbol{b} \equiv \boldsymbol{0}$ (Lemma 2). Figure 2 illustrates this: normalized chords at three horizons coincide. A usable target must depend on the marginal flow itself while still avoiding a JVP, and the semigroup law is exactly such a target.

(L2) Horizon semigroup loss on $\boldsymbol{w}_{\boldsymbol{\theta}}$. Let $\Psi_{r,t}(\boldsymbol{z}) = \mathrm{Exp}_{\boldsymbol{z}}[-(t-r)\boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z},r,t)]$ be the jump map the field induces over horizon $\tau = t - r$. On the flat manifold Exp-composition is displacement addition at any step size, so for the *ideal* field $\boldsymbol{u}_\star$ the jump maps compose exactly, $\Psi^\star_{r,t} = \Psi^\star_{r,s} \circ \Psi^\star_{s,t}$: one long jump equals two shorter ones (Lemma 1, Remark 1). A learned field has no such property for free—imposing it is what $\mathcal{L}_{\mathrm{hor}}$ does. Dividing each displacement by its interval length writes this *semigroup* property in velocities,

$$(t-r)\,\boldsymbol{u}_\star(\boldsymbol{z}_t,r,t) \;=\; (t-s)\,\boldsymbol{u}_\star(\boldsymbol{z}_t,s,t) \;+\; (s-r)\,\boldsymbol{u}_\star(\boldsymbol{z}_s,r,s), \tag{6}$$

where $\boldsymbol{z}_s = \mathrm{Exp}_{\boldsymbol{z}_t}\big(-(t-s)\,\boldsymbol{u}_\star(\boldsymbol{z}_t,s,t)\big)$ is the state reached by the first leg. Conversely, under the regularity assumptions of Proposition 1, $\boldsymbol{u}_\star$ is the unique field satisfying (6) together with the boundary condition $\boldsymbol{u}_\star(\boldsymbol{z},t,t) = \boldsymbol{v}_\star(\boldsymbol{z},t)$. This is the SRTF principle of Section 1, and the two losses impose exactly its two conditions.

To turn the composition law into a regression target, let $\boldsymbol{\theta}^-$ be a stop-gradient copy of the parameters, take the midpoint $s = (t+r)/2$, $h = \tau/2$, and evaluate two half-steps under `no_grad`: $\boldsymbol{w}_1 = \boldsymbol{w}_{\boldsymbol{\theta}^-}(\boldsymbol{z}_t,s,t)$, then $\boldsymbol{w}_2 = \boldsymbol{w}_{\boldsymbol{\theta}^-}(\hat{\boldsymbol{z}}_s,r,s)$ at $\hat{\boldsymbol{z}}_s = \mathrm{Exp}_{\boldsymbol{z}_t}(-h\,\boldsymbol{w}_1)$. Their average is the target,

$$\bar{\boldsymbol{v}}_{\mathrm{sg}}(\boldsymbol{z}_t,r,t) \;=\; \tfrac{1}{2}\big(\boldsymbol{w}_1 + \boldsymbol{w}_2\big), \tag{7}$$

the factor $\frac{1}{2}$ being interval arithmetic; a general split $t - s = \xi\tau$ gives $\xi\boldsymbol{w}_1 + (1-\xi)\boldsymbol{w}_2$. The target is JVP-free by construction—two forward evaluations, no activation storage, the first reusing the backbone state already computed. As $\boldsymbol{\theta}^-$ equals $\boldsymbol{\theta}$ in value, the population fixed points of the bootstrap objective are those of Proposition 1 (ii), or, within the SRTF class, those of (iii) to leading order in the horizon law (Remark 3); the stop-gradient changes the optimization path, not this fixed-point set. The horizon loss regresses the long jump onto it,

$$\mathcal{L}_{\mathrm{hor}}(\boldsymbol{\theta}) \;=\; \mathbb{E}_{r,t,\boldsymbol{z}_0,\boldsymbol{z}_1} \sum_{c\in\{A,F,L\}} \tfrac{\omega_c}{\dim_c} \left\| \boldsymbol{w}_{\boldsymbol{\theta}}^c(\boldsymbol{z}_t,r,t) - \bar{\boldsymbol{v}}_{\mathrm{sg}}^c(\boldsymbol{z}_t,r,t) \right\|_c^2. \tag{8}$$

What this objective is worth rests on the following, which is the paper's central theoretical statement. Write the *semigroup residual*—the displacement by which one long jump fails to equal two shorter ones—as $\mathcal{R}[\boldsymbol{w}] := (t-r)\boldsymbol{w}(\boldsymbol{z}_t,r,t) - (t-s)\boldsymbol{w}(\boldsymbol{z}_t,s,t) - (s-r)\boldsymbol{w}(\hat{\boldsymbol{z}}_s,r,s)$, with $\hat{\boldsymbol{z}}_s = \mathrm{Exp}_{\boldsymbol{z}_t}(-(t-s)\boldsymbol{w}(\boldsymbol{z}_t,s,t))$ the state the first leg reaches.

Proposition 1 (Semigroup identification)**.** *Let $\boldsymbol{v}_\star \in C^3$ be Lipschitz in $\boldsymbol{z}$ on the region of interest.* ***(i)*** *$\mathcal{R}[\boldsymbol{u}_\star] \equiv \boldsymbol{0}$ and $\boldsymbol{u}_\star(\boldsymbol{z},t,t) = \boldsymbol{v}_\star(\boldsymbol{z},t)$, so $\boldsymbol{u}_\star$ is a zero of $\mathcal{L}_{\mathrm{hor}}$ under any sampling of states, times and split points.* ***(ii)*** *Conversely, a continuous field that is uniformly Lipschitz in its state,*

*satisfies $\mathcal{R}[\boldsymbol{w}] = \mathbf{0}$ at the midpoint split for all $r < t$, and matches the boundary condition equals $\boldsymbol{u}_\star$. **(iii)** Restricted to the SRTF class $\boldsymbol{w} = \boldsymbol{a} + \tau\boldsymbol{b}$ with $\boldsymbol{a}, \boldsymbol{b} \in C^2$, and writing the split as $t - s = \xi\tau$,*

$$\mathcal{R}[\boldsymbol{w}] \;=\; \xi(1-\xi)\,\tau^2\big(2\boldsymbol{b} + \mathrm{D}_t\boldsymbol{a}\big) + \mathcal{O}(\tau^3), \qquad \mathrm{D}_t = \partial_t + \boldsymbol{a}\cdot\nabla,$$

so the τ^2 term vanishes for one—equivalently every—interior split exactly when $\boldsymbol{b} = -\frac{1}{2}\mathrm{D}_t\boldsymbol{a}$. With $\boldsymbol{a} = \boldsymbol{v}_\star$ pinned by $\mathcal{L}_{\mathrm{loc}}$, the unique first-order-consistent solution in the class is $(\boldsymbol{v}_\star, -\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star)$: the Taylor coefficients of Eq. (3).

Parts (i)–(ii) give the objective the right population fixed point and no other. Part (iii) is the sharper one: the structured class cannot drive $\mathcal{R}$ to zero, but the leading coefficient it *can* cancel identifies $\boldsymbol{b}_\star$ uniquely—so the horizon head is identified rather than merely fitted, and the bridge-chord collapse $\boldsymbol{b} \equiv \mathbf{0}$ is excluded at second order. Since the residual is $\xi(1-\xi)$ times a split-independent bracket, that identification does not depend on where the composition is split. Proofs and the remaining regularity conditions are in Appendix C.3.

What the trained field buys at sampling time is controlled by a Grönwall estimate on the sampler.

Theorem 1 (Sampler error)**.** *Assume $\boldsymbol{v}_\star \in C^3$ and that sampler and flow paths stay in a compact region on which $\boldsymbol{u}_\star$ is L_z-Lipschitz in $\boldsymbol{z}$ and $\|\mathrm{D}_t^2\boldsymbol{v}_\star\|$ is bounded. If the trained field obeys $\sup_{\boldsymbol{z},r,t}\|\boldsymbol{w}_{\boldsymbol{\theta}} - \bar{\boldsymbol{u}}_{\mathrm{SRTF}}\| \le \epsilon_{\mathrm{hor}}$, then the K-step sampler* (11) *with uniform step $\tau = 1/K$, started from a prior draw and returning $\hat{\boldsymbol{z}}_0$, satisfies*

$$\mathbb{E}\|\hat{\boldsymbol{z}}_0 - \tilde{\boldsymbol{z}}_0\| \;\le\; e^{L_z}\big(\epsilon_{\mathrm{hor}} + C_{\boldsymbol{v}_\star}\tau^2\big), \qquad C_{\boldsymbol{v}_\star} = \tfrac{1}{6}\sup\|\mathrm{D}_t^2\boldsymbol{v}_\star\|,$$

where $\tilde{\boldsymbol{z}}_0$ is the image of the same draw under the exact reverse flow of $\boldsymbol{v}_\star$.

The bound separates what training controls from what the budget controls: a learned residual ϵ_{hor} that does not shrink with K, and an $\mathcal{O}(\tau^2)$ truncation that does. It bounds deviation from the exact $\boldsymbol{v}_\star$-flow, not distributional error. Per step, the residual is the local one at small τ and acquires a horizon-proportional part as τ grows; since $\tau \le t$ and the training law gives $\mathbb{E}[\tau \mid t] = t/4$, the horizons a model must represent grow with t. *Motivated* by this, we weight each loss where its residual binds:

$$\lambda_{\mathrm{loc}}(t) \;=\; 1 + \kappa_{\mathrm{loc}}\cdot(1-t), \qquad \lambda_{\mathrm{hor}}(t) \;=\; 1 + \kappa_{\mathrm{hor}}\cdot t, \tag{9}$$

with $\kappa_{\mathrm{loc}} = \kappa_{\mathrm{hor}} = 1$ throughout, never tuned per dataset, so the schedule adds no hyperparameter search. We sample t from a logit-normal law and $r \sim \mathrm{Unif}(0,t)$, reset to $r = t$ with probability 0.5 so that half the batch supplies a local-only gradient.

Full objective, and sampling. With $\ell_{\mathrm{loc}}, \ell_{\mathrm{hor}}$ the per-sample losses of (5) and (8), the weights act inside the expectation:

$$\mathcal{L}_{\mathrm{SRTF}}(\boldsymbol{\theta}) \;=\; \mathbb{E}_{t,r,\boldsymbol{z}_0,\boldsymbol{z}_1}\Big[\,\lambda_{\mathrm{loc}}(t)\,\ell_{\mathrm{loc}} \;+\; \mathbf{1}_{\{r<t\}}\,\lambda_{\mathrm{hor}}(t)\,\ell_{\mathrm{hor}}\,\Big]. \tag{10}$$

Only the online branch backpropagates, so SRTF keeps flow-matching activation memory and never materializes a forward-mode tangent buffer, at the price of at most two extra gradient-free evaluations per step (Algorithm 1); at sampling time, each step costs a single network evaluation. Sampling chooses $0 = t_0 < \cdots < t_K = 1$ and iterates backward from $\boldsymbol{z}_{t_K} \sim p_1$:

$$\boldsymbol{z}_{t_{k-1}} \;=\; \mathrm{Exp}_{\boldsymbol{z}_{t_k}}\Big(-(t_k - t_{k-1})\,\boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z}_{t_k}, t_{k-1}, t_k)\Big). \tag{11}$$

At $K{=}1$ the whole transport is one jump at $\tau{=}1$, where the $\mathcal{O}(\tau^2)$ truncation of (3) need not be small, so one-step quality is an empirical question (Section 4); at $r = t$ the field collapses to $\boldsymbol{a}_{\boldsymbol{\theta}}$ and the sampler is Riemannian flow matching. Intermediate budgets interpolate between these two regimes: with uniform steps each jump spans $\tau = 1/K$, so $\boldsymbol{b}_{\boldsymbol{\theta}}$ contributes in proportion to the step length. Theorem 1 analyzes this pure sampler; experiments use a hybrid of leading SRTF steps (possibly none) and $\boldsymbol{a}_{\boldsymbol{\theta}}$-only refinements, switched on validation (Algorithm 2). Since one evaluation of the shared backbone yields both coefficients, a K-step sampler uses exactly K NFEs.

4 Experiments

We ask whether SRTF reduces NFE, and whether the gain comes from the objective or the parameterization. Section 4.2 answers the first on four crystal structure prediction benchmarks; Section 4.3 answers the second with a controlled factorial, head-zeroing and residual-coefficient measurements. Seed reliability, diagnostics and the full ablation set are in Appendices E–F.4.

Table 1: **CSP on the four standard benchmarks**, full test splits, match rate MR_k in % with $k{=}1$ and $k{=}20$ samples per structure and matched-pair RMSE at $k{=}1$. Baselines are as published, at their authors' own NFE, except the FlowMM $k{=}20$ cells, which its paper does not report and which we measure in this pipeline (Appendix D); SRTF rows are trained and evaluated in this pipeline and average three training seeds, with the sampler configuration re-selected per seed and per NFE (Section 4.1). Appendix E.1 gives their across-seed standard deviations. Per column: **best** in bold, second-best underlined. SRTF rows are shaded; the green factor is each budget's NFE reduction relative to FlowMM's 50 steps, the fastest baseline.

		Perov-5			Carbon-24			MP-20			MPTS-52		
Method	NFE	$\mathrm{MR}_1\uparrow$	$\mathrm{MR}_{20}\uparrow$	RMSE↓	$\mathrm{MR}_1\uparrow$	$\mathrm{MR}_{20}\uparrow$	RMSE↓	$\mathrm{MR}_1\uparrow$	$\mathrm{MR}_{20}\uparrow$	RMSE↓	$\mathrm{MR}_1\uparrow$	$\mathrm{MR}_{20}\uparrow$	RMSE↓
CDVAE	5000	45.31	88.51	0.114	17.09	88.37	0.297	33.90	66.95	0.105	5.34	20.79	0.211
DiffCSP	1000	52.02	98.60	0.076	17.54	88.47	**0.276**	51.49	77.93	0.063	12.19	34.02	0.179
FlowMM	50	53.15	**100.00**	0.099	23.47	**89.51**	0.412	61.39	81.19	**0.057**	17.54	**39.12**	**0.173**
SRTF (ours)	2 (25×)	**70.82**	**100.00**	0.363	**25.85**	76.35	0.411	64.02	83.40	0.179	16.36	31.13	0.307
SRTF (ours)	5 (10×)	65.05	**100.00**	0.286	22.25	85.94	0.360	67.20	84.08	0.141	19.08	35.33	0.252
SRTF (ours)	10 (5×)	51.84	99.91	0.075	20.48	88.52	0.333	68.27	**84.41**	0.097	20.73	36.79	0.210
SRTF (ours)	20 (2.5×)	51.55	99.83	**0.068**	19.16	88.10	0.322	**68.45**	84.33	0.078	**21.25**	38.26	0.185

4.1 Setup

Datasets, metrics, and baselines. We use the four CSP benchmarks of Miller et al. (2024)—Perov-5, Carbon-24, MP-20, MPTS-52 (Castelli et al., 2012; Pickard, 2020; Jain et al., 2013; Baird et al., 2024)—reporting Match Rate (MR) and matched-pair RMSE via `pymatgen.StructureMatcher` (Ong et al., 2013) with $k{=}1, 20$ samples per structure on the **full test splits**, against published CDVAE (Xie et al., 2022), DiffCSP (Jiao et al., 2023) and FlowMM (Miller et al., 2024). Two same-backbone references are trained here, a matched flow-matching baseline and a monolithic Riemannian MeanFlow head (MF-Riem) supervised through the JVP identity (Geng et al., 2025a). All arms share the CSPNet backbone, to which SRTF's second readout adds under half a percent of parameters, and all run on a single NVIDIA RTX PRO 6000 Blackwell GPU. Appendix D specifies the arms and the budgets; the code release has the rest.

Sampler-configuration selection. For each NFE, the sampler configuration—the number of leading transport steps and an anti-annealing strength on the coordinate block, 12–15 configurations in all—is selected on a validation split and evaluated on the held-out test split. The same selection grid is used for SRTF and the matched flow-matching baseline; published rows use their authors' settings.

4.2 Few-step crystal structure prediction

The highest $k{=}1$ match rate among compared methods, at a fraction of the budget. Among the methods compared in Table 1, SRTF attains the highest $k{=}1$ match rate on all four datasets, using only 2–5 steps—10–25× fewer evaluations than FlowMM's 50. The optimum depends on the manifold: two steps suffice on the small, highly symmetric benchmarks, where further integration trades match rate for precision (Appendix E.2), while on the multi-element benchmarks accuracy rises with the budget and saturates. Matched-pair RMSE improves monotonically throughout.

A never-swept lattice knob turns Carbon-24's weakest column into its strongest. The sampler anti-anneals the coordinate and lattice blocks separately, but every cell above sweeps only the coordinate knob. Carbon-24 is pure carbon, so its matched-pair RMSE is almost entirely lattice error. Opening the grid to the lattice knob improves RMSE at every budget, and at fifty steps overtakes all three published baselines on a cell Table 1 records as a loss. The gain belongs to the wider grid rather than the knob alone, and survives held-out validation.

The gap is concentrated where the theory predicts the horizon correction matters most (Figure 3). Against the same backbone trained with plain flow matching, the advantage is more than forty points at one step and decays to nothing by fifty—consistent with the expansion, since the horizon correction is $\mathcal{O}(\tau)$ and vanishes as steps shorten. At the largest budget SRTF's search switches the transport branch off and the curves meet, consistent with the vanishing correction; the harder the manifold, the longer the correction pays off, and MPTS-52 still shows a margin at fifty steps. Figure 4 shows what the steps buy geometrically.

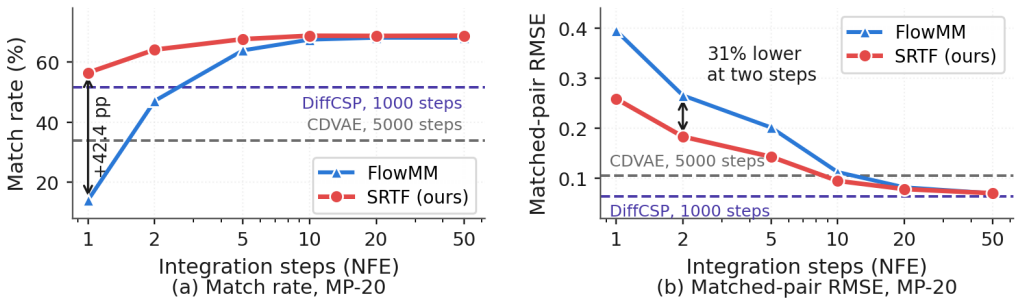


Figure 3: **MP-20 CSP against the step budget**, k=1, full test split. SRTF (red) against a matched flow-matching baseline (blue) that shares the backbone, data and budget, differs only in the objective, and selects its sampler on the same grid; the arrow in (a) is the one-step gap. Against DiffCSP and CDVAE (dashed, at their own budgets), one SRTF step passes DiffCSP's 1000-step plateau. Remaining benchmarks: Appendix E.2.

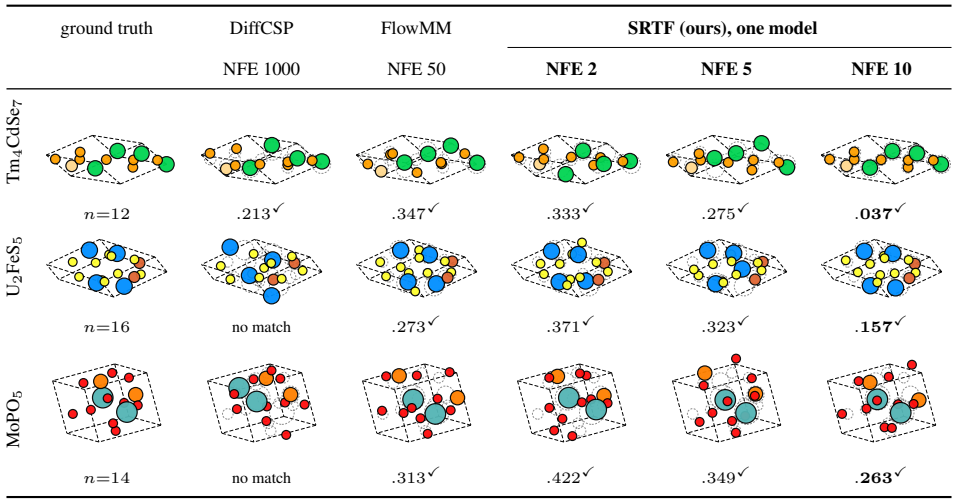


Figure 4: **Same model, more steps.** Qualitative CSP on MP-20, one test composition per row. Each structure is drawn in the ground truth's frame with the ground truth underlaid as hollow dashed circles, so the visible displacement is the RMSE printed beneath it (✓ marks a match); the last three panels of a row are the *same* checkpoint at two, five and ten evaluations. Rows are examples chosen to show the step-budget effect clearly, with a fourth in Figure 7 and the selection criterion in Appendix E.6; Table 1 carries the average over the full split.

4.3 Why it works

Relative to a monolithic JVP-trained MeanFlow, SRTF changes the *supervision* and the *representation*; Figure 5a varies them independently in a matched-budget two-by-two factorial on MP-20.

In this factorial, the supervision change has the larger effect. Holding the representation fixed and swapping the semigroup target for the JVP identity costs twelve to seventeen points, and the JVP arm *plateaus* from five steps onward—consistent with a mis-specified small-horizon field, which refinement at sampling time cannot repair. The representation makes that supervision affordable: because the horizon enters SRTF only through a scalar, its first target evaluation reuses the backbone features of the differentiated forward, whereas a monolithic field taking the horizon as a backbone input runs two extra forwards at roughly double the peak memory. It also edges out the monolithic field by a few points at one step and at twenty, and matches it at five.

The head is what carries the gain, and it aligns with the coefficient the theory names (Figure 5b,c). First, zeroing the horizon head at sampling time—which reduces the field to its local part and makes

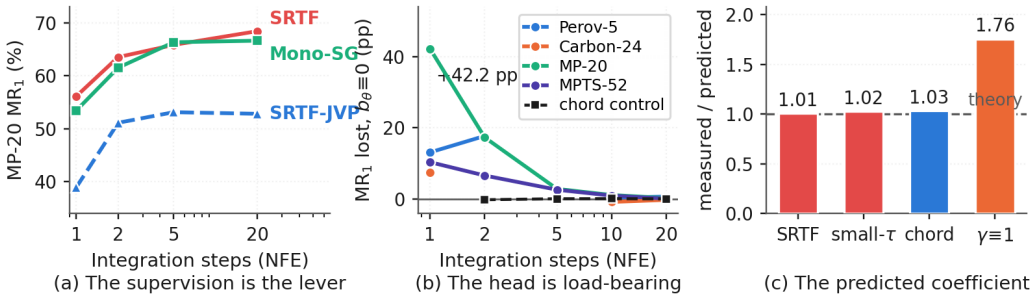


Figure 5: **The objective is the lever, and the head aligns with the coefficient the theory names.** *(a)* Decomposition × supervision factorial on MP-20 (matched budget, 2000-structure subset): the semigroup arms (SRTF, Mono-SG) isolate the representation, the structured arms (SRTF, SRTF-JVP) the supervision. *(b)* Match rate lost when $\boldsymbol{b_\theta}$ is zeroed at sampling time. *(c)* Measured vs. predicted leading coefficient; the prediction does not apply to $\gamma \equiv 1$. Appendix F.

the sampler plain flow matching on the same weights—costs more than forty points on MP-20 and seventeen on Perov-5, against nothing for a chord-supervised control. Second, the residual has the *predicted* form: Proposition 1 (iii) fixes its leading coefficient, and two measurements of it on a trained model—fitting the residual at small horizon, and evaluating the prediction directly—agree to within 3%, while the $\gamma \equiv 1$ arm, to which it does not apply, misses by 76%. **We numerically verify the predicted residual expansion on trained networks** (Appendix F.1). Identification shows in orientation: $\boldsymbol{b_\theta}$ aligns with $-\frac{1}{2}\mathrm{D}_t \boldsymbol{a_\theta}$ (median cosine $+0.65$), whereas the chord control's inert head does not (-0.11). SRTF is thus a *training principle*, not an architectural trick.

Finite-horizon supervision at flow-matching memory. The semigroup target costs about $1.3\times$ the per-step wall-clock of plain flow matching, about 30% less than even the fused JVP, and *no* additional memory. Memory bounds model and batch size: the semigroup objective leaves both unchanged, while the JVP arm gives up two thirds of the batch and half the width, and under a fixed wall-clock budget sees about 30% fewer structures. On MPTS-52 our JVP implementation *fails to converge* at both learning rates tried in its own configuration, where SRTF trains (Appendices E.4, F.7, F.8).

Two ablations anchor the parameterization: removing the horizon factor, which destroys the zero-horizon boundary condition, costs as much at one step as removing the head, while changing the profile's *shape* at fixed slope leaves accuracy unmoved, as the first-order expansion predicts: only the profile's slope at zero horizon enters it (Appendix F.4).

5 Conclusion

Few-step generation needs a finite-horizon velocity, and that field is characterized by a composition law together with its zero-horizon boundary condition—a law that holds exactly on the flat crystal product manifold and can be enforced with ordinary forward evaluations, removing the Jacobian–vector product MeanFlow-style training requires. On four CSP benchmarks SRTF exceeds the published baselines' single-sample match rates in two to five steps, 10–$25\times$ fewer than the fastest, and a matched-backbone factorial shows the objective rather than the decomposition is the larger source of the gain; the decomposition makes it memory-efficient and interpretable. The theory identifies the horizon head to leading order, and trained networks follow the predicted expansion.

Several directions follow. The composition law is exact on any flat product manifold, so the construction carries to other domains whose state space factorizes that way; on curved geometries the transport expansion still generalizes, but the semigroup objective would need additional curvature corrections, since exact jump composition no longer holds. Within crystals, symmetry-constrained integrators and richer atom-type parameterizations are natural next steps. Extending the horizon further is natural: the same composition law holds at any jump size, and how best to exploit it for the longest jumps remains open. Finally, the identification results concern population fixed points, and the optimization dynamics of the bootstrap that reaches them remain an open question for analysis.

ETHICS STATEMENT

This work uses only public benchmark datasets derived from published materials databases; it involves no human subjects, no personal data and no new data collection, and we declare no conflicts of interest. The main consideration is dual use: cheaper generation makes screening for any objective cheaper, and generated structures carry no guarantee of synthesizability or stability, so they should be treated as candidates for validation rather than as results (Appendix G).

REPRODUCIBILITY STATEMENT

Every formal statement carries its own assumptions rather than leaving them to surrounding prose, and every proof is in Appendix C; where a result is limited in scope we say so in the statement itself (Remark 3, Proposition 3, Remark 4). Section 4.1 and Appendix D give datasets, splits, backbone, readout architecture, target construction and sampler, and the code release specifies the optimizer settings, schedule constants and per-dataset epoch budgets. Two details are stated wherever a number appears: which evaluation protocol produced it—the full test split, or the fixed 2000- and 500-structure subsets used for the MF-Riem head-to-head and the controlled MP-20 studies—and which training budget applies. Table 5 reports three seeds on all four benchmarks at every NFE of Table 1, and says what the error bars include. Code, trained checkpoints, and the machine-readable results behind every table and figure will be released; all experiments run on a single GPU.

AI USE STATEMENT

In this work, we used generative AI tools (Anthropic's Claude; OpenAI's ChatGPT for manuscript review and feedback) for the following tasks. **Writing and editing:** revising and restructuring the manuscript, drafting and rewriting prose in the experimental sections and appendices, and copy-editing throughout. **Experiment execution and analysis:** orchestrating benchmark runs and evaluation sweeps from author-specified experimental designs, aggregating results into machine-readable form, and generating figures from those results. **Verification:** auditing the manuscript for internal numerical consistency against the recorded experimental outputs, and independently re-deriving and numerically checking the theoretical results of Appendix C, including proposing corrections to proofs and statements that the authors then reviewed.

We have reviewed all AI-assisted work. Every number reported in this paper is traceable to a recorded experimental artifact rather than to model-generated text, and we verified this by an automated check of the compiled manuscript against those artifacts; the theoretical statements and proofs were re-derived and checked independently of their drafting. Any issues surfaced by that process were resolved by the authors before submission. We take responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Sterling G. Baird, Hasan M. Sayeed, Joseph Montoya, and Taylor D. Sparks. matbench-genmetrics: A Python library for benchmarking crystal structure generative models using time-based splits of Materials Project structures. *Journal of Open Source Software*, 9(97):5618, 2024. doi: 10.21105/joss.05618.

Jan-Hendrik Bastek, WaiChing Sun, and Dennis M. Kochmann. Physics-informed diffusion models. In *International Conference on Learning Representations (ICLR)*, 2025.

Ilyes Batatia, Dávid Péter Kovács, Gregor Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Neural Information Processing Systems (NeurIPS)*, 2022.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, 2022.

Cyprien Bone, Matthew Walker, Kuangdai Leng, Luis M. Antunes, Ricardo Grau-Crespo, Amil Aligayev, Javier Dominguez, and Keith T. Butler. Discovery and recovery of crystalline materials with property-conditioned transformers, 2025. URL https://arxiv.org/abs/2511.21299.

Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J. Bekkers, and Max Welling. Geometric and physical quantities improve E(3) equivariant message passing. In *International Conference on Learning Representations (ICLR)*, 2022.

Zhendong Cao, Xiaoshan Luo, Jian Lv, and Lei Wang. Space group informed transformer for crystalline materials generation. *Science Bulletin*, 70(21):3522–3533, 2025.

Ivano E. Castelli et al. New cubic perovskites for one- and two-photon water splitting using the computational materials repository. *Energy & Environmental Science*, 5(10):9034–9043, 2012.

Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 2022.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations (ICLR)*, 2024.

Ting Chen, Ruixiang Zhang, and Geoffrey Hinton. Analog bits: Generating discrete data using diffusion models with self-conditioning. In *International Conference on Learning Representations (ICLR)*, 2023.

Mouyang Cheng, Weiliang Luo, Hao Tang, Bowen Yu, Yongqiang Cheng, Weiwei Xie, Ju Li, Heather J. Kulik, and Mingda Li. Enhancing materials discovery with valence-constrained design in generative modeling. *Nature Computational Science*, 2026. doi: 10.1038/s43588-026-01037-2.

Hitarth Choubisa et al. Interpretable discovery of semiconductors with machine learning. *npj Computational Materials*, 9(1):117, 2023.

Quan Dao, Hao Phung, Binh Nguyen, and Anh Tran. Flow matching in latent space. *arXiv preprint arXiv:2307.08698*, 2023.

Pierre-Paul De Breuck, Hai-Chen Wang, Gian-Marco Rignanese, Silvana Botti, and Miguel AL Marques. Generative AI for crystal structures: a review. *npj Computational Materials*, 11:370, 2025. doi: 10.1038/s41524-025-01881-2.

Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J. Bartel, and Gerbrand Ceder. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023.

Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. In *International Conference on Learning Representations (ICLR)*, 2025.

Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *Neural Information Processing Systems (NeurIPS)*, 2025a.

Zhengyang Geng, Ashwini Pokle, William Luo, Justin Lin, and J. Zico Kolter. Consistency models made easy. In *International Conference on Learning Representations (ICLR)*, 2025b.

Zhengyang Geng, Yiyang Lu, Zongze Wu, Eli Shechtman, J. Zico Kolter, and Kaiming He. Improved mean flows: On the challenges of fastforward generative models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Nate Gruver, Anuroop Sriram, Andrea Madotto, Andrew Gordon Wilson, C. Lawrence Zitnick, and Zachary Ulissi. Fine-tuned language models generate stable inorganic materials as text. In *International Conference on Learning Representations (ICLR)*, 2024.

Gabe Guo, Tristan Luca Saidi, Maxwell W. Terban, Michele Valsecchi, Simon J. L. Billinge, and Hod Lipson. Ab initio structure solutions from nanocrystalline powder diffraction data via diffusion models. *Nature Materials*, 24:1726–1734, 2025. doi: 10.1038/s41563-025-02220-y.

Emiel Hoogeboom, Víctor Garcia Satorras, Clement Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3D. In *International Conference on Machine Learning (ICML)*, 2022.

Yuhao Huang, Shih-Hsin Wang, Andrea L. Bertozzi, and Bao Wang. RMFlow: Refined mean flow by a noise-injection step for multimodal generation. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://openreview.net/forum?id=p072J56yo4.

Anubhav Jain et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. In *Neural Information Processing Systems (NeurIPS)*, 2023.

Rui Jiao, Wenbing Huang, Yu Liu, Deli Zhao, and Yang Liu. Space group constrained crystal generation. In *International Conference on Learning Representations (ICLR)*, 2024.

Frederik Lizak Johansen, Ulrik Friis-Jensen, Erik Bjørnager Dam, Kirsten M. Ø. Jensen, Rocío Mercado, and Raghavendra Selvan. deCIFer: Crystal structure prediction from powder diffraction data using autoregressive language models. *Transactions on Machine Learning Research (TMLR)*, 2026.

Qingsi Lai, Fanjie Xu, Lin Yao, Zhifeng Gao, Siyuan Liu, Hongshuai Wang, Shuqi Lu, Di He, Liwei Wang, Linfeng Zhang, Cheng Wang, and Guolin Ke. End-to-end crystal structure prediction from powder X-ray diffraction. *Advanced Science*, 12:2410722, 2025. doi: 10.1002/advs.202410722.

Qi Li, Rui Jiao, Liming Wu, Tiannian Zhu, Wenbing Huang, Shifeng Jin, Yang Liu, Hongming Weng, and Xiaolong Chen. Powder diffraction crystal structure determination using generative models. *Nature Communications*, 16:7428, 2025. doi: 10.1038/s41467-025-62708-8.

Tianhong Li and Kaiming He. Back to basics: Let denoising generative models denoise. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Chen Ling. A review of the recent progress in battery informatics. *npj Computational Materials*, 8 (1):33, 2022.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)*, 2023.

Cheng Lu and Yang Song. Simplifying, stabilizing and scaling continuous-time consistency models. In *International Conference on Learning Representations (ICLR)*, 2025.

Shuqi Lu, Haowei Lin, Lin Yao, Zhifeng Gao, Xiaohong Ji, Weinan E, Linfeng Zhang, and Guolin Ke. Uni-3DAR: Unified 3D generation and understanding via autoregression on compressed spatial tokens, 2025. URL https://arxiv.org/abs/2503.16278.

Yiyang Lu, Susie Lu, Qiao Sun, Hanhong Zhao, Zhicheng Jiang, Xianbang Wang, Tianhong Li, Zhengyang Geng, and Kaiming He. One-step latent-free image generation with pixel mean flows, 2026. URL https://arxiv.org/abs/2601.22158.

Xiaoshan Luo, Zhenyu Wang, Qingchang Wang, Xuechen Shao, Jian Lv, Lei Wang, Yanchao Wang, and Yanming Ma. CrystalFlow: a flow-based generative model for crystalline materials. *Nature Communications*, 16:9267, 2025. doi: 10.1038/s41467-025-64364-4.

Houssam Metni, Laura Ruple, Lauren N Walters, Luca Torresi, Jonas Teufel, Henrik Schopmans, Jona Östreicher, Yumeng Zhang, Marlen Neubert, Yuri Koide, et al. Generative models for crystalline materials. *Advanced Materials*, 38(18):e23620, 2026.

Benjamin Kurt Miller, Ricky T. Q. Chen, Anuroop Sriram, and Brandon M. Wood. FlowMM: Generating materials with Riemannian flow matching. In *International Conference on Machine Learning (ICML)*, 2024.

Ryotaro Okabe, Mouyang Cheng, Abhijatmedhi Chotrattanapituk, Manasi Mandal, Kiran Mak, Denisse Córdova Carrizales, Nguyen Tuan Hung, Xiang Fu, Bowen Han, Yao Wang, et al. Structural constraint integration in a generative model for the discovery of quantum materials. *Nature Materials*, 25(2):223–230, 2026. doi: 10.1038/s41563-025-02355-y.

Shyue Ping Ong, William D. Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

Teerachote Pakornchote, Natthaphon Choomphon-anomakhun, Sorrjit Arrerut, Chayanon Atthapak, Sakarn Khamkaeo, Thiparat Chotibut, and Thiti Bovornratanaraks. Diffusion probabilistic models enhance variational autoencoder for crystal structure generative modeling. *Scientific Reports*, 14: 1275, 2024.

Chris J. Pickard. AIRSS data for carbon at 10GPa and the C+N+H+O system at 1GPa. Materials Cloud Archive, 2020.

Laura Ruple, Luca Torresi, Henrik Schopmans, and Pascal Friederich. Symmetry-aware Bayesian flow networks for crystal generation. *npj Computational Materials*, 12:182, 2026. doi: 10.1038/s41524-026-02140-8.

Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *International Conference on Machine Learning (ICML)*, 2021.

Kristof T. Schütt, Oliver T. Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning (ICML)*, 2021.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *International Conference on Machine Learning (ICML)*, 2023.

Anuroop Sriram, Benjamin Kurt Miller, Ricky T. Q. Chen, and Brandon M. Wood. FlowLLM: Flow matching for material generation with large language models as base distributions. In *Neural Information Processing Systems (NeurIPS)*, 2024.

Nathan J. Szymanski, Bernardus Rendy, Yuxing Fei, Rishi E. Kumar, Tanjin He, David Milsted, Matthew J. McDermott, Max Gallant, Ekin Dogus Cubuk, Amil Merchant, et al. An autonomous laboratory for the accelerated synthesis of inorganic materials. *Nature*, 624:86–91, 2023.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

Utkarsh Utkarsh, Pengfei Cai, Alan Edelman, Rafael Gomez-Bombarelli, and Christopher Vincent Rackauckas. Physics-constrained flow matching: Sampling generative models with hard constraints. In *Neural Information Processing Systems (NeurIPS)*, 2025. URL https://openreview.net/forum?id=cf4etwjY7n.

Dongyeop Woo, Marta Skreta, Seonghyun Park, Kirill Neklyudov, and Sungsoo Ahn. Riemannian MeanFlow. In *International Conference on Machine Learning (ICML)*, 2026.

Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi S. Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. In *International Conference on Learning Representations (ICLR)*, 2022.

Shihang Xu, Shibing Chu, Rami Mrad, Zhejun Zhang, Zhelin Li, Runxian Jiao, and Yuanping Chen. Discovery of 2D materials via symmetry-constrained diffusion model. *The Journal of Physical Chemistry C*, 129(14):6794–6802, 2025. doi: 10.1021/acs.jpcc.4c08681. URL https://doi.org/10.1021/acs.jpcc.4c08681.

Ziduo Yang, Xian Wang, Yifan Li, Qiujie Lv, Calvin Yu-Chian Chen, and Lei Shen. Efficient equivariant model for machine learning interatomic potentials. *npj Computational Materials*, 11 (1):49, 2025.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? In *Neural Information Processing Systems (NeurIPS)*, 2021.

Claudio Zeni, Robert Pinsler, Daniel Zügner, et al. A generative model for inorganic materials design. *Nature*, 639:624–632, 2025.

Xinxi Zhang, Shiwei Tan, Quang Nguyen, Quan Dao, Ligong Han, Xiaoxiao He, Tunyu Zhang, Chengzhi Mao, Dimitris Metaxas, and Vladimir Pavlovic. Overcoming the curvature bottleneck in MeanFlow, 2025. URL https://arxiv.org/abs/2511.23342.

Kaiwen Zheng, Yuji Wang, Qianli Ma, Huayu Chen, Jintao Zhang, Yogesh Balaji, Jianfei Chen, Ming-Yu Liu, Jun Zhu, and Qinsheng Zhang. Large scale diffusion distillation via score-regularized continuous-time consistency. In *International Conference on Learning Representations (ICLR)*, 2026.

Peichen Zhong, Xinzhe Dai, Bowen Deng, Gerbrand Ceder, and Kristin A. Persson. Crystal structure prediction with host-guided inpainting generation and foundation potentials. *Mater. Horiz.*, 12 (22):9669–9678, 2025. doi: 10.1039/D5MH00774G. URL http://dx.doi.org/10.1039/D5MH00774G.

Zichen Zhong, Haoliang Sun, Yukun Zhao, Yongshun Gong, and Yilong Yin. Riemannian MeanFlow for one-step generation on manifolds. In *International Conference on Machine Learning (ICML)*, 2026.

Linqi Zhou, Stefano Ermon, and Jiaming Song. Inductive moment matching. In *International Conference on Machine Learning (ICML)*, 2025.

APPENDIX ROADMAP

The appendix follows the order of the main text. Appendix A fixes notation and gives in full the setting Section 3 compresses to a paragraph, and Appendix B the full related work. Appendix C contains every proof and derivation behind Section 3; Appendix D the implementation and sampler details behind Section 4.1. The two experimental appendices mirror the two experimental subsections: Appendix E holds the CSP results in full, in the order of Section 4.2—seed reliability, the remaining NFE curves, the two same-backbone head-to-head comparisons and the lattice lever—and Appendix F holds the mechanism, ablation and cost results behind Section 4.3. Appendix G discusses broader impacts.

A NOTATION

A.1 THE SETTING IN FULL: REPRESENTATION, FLATNESS, AND TIME CONVENTION

Section 3 compresses this to a paragraph.

Time convention. Training draws $\boldsymbol{z}_0 \sim p_0$ (data) and $\boldsymbol{z}_1 \sim p_1$ (prior) and interpolates along $t: 0 \to 1$. Generation integrates the learned field *backward*, from t=1 to t=0, using $-\boldsymbol{w}_{\boldsymbol{\theta}}$. Throughout, t is the endpoint time at which the field is evaluated, $r \leq t$ the start of the transport interval, and $\tau = t - r$ its horizon.

Crystal representation. A periodic crystal with n atoms is a triple $\boldsymbol{z} = (\boldsymbol{z}^A, \boldsymbol{f}, \boldsymbol{l})$ combining discrete chemical identities, periodic coordinates, and continuous lattice parameters; our generative representation embeds all three into a product manifold $\mathcal{M} = \mathcal{A} \times \mathcal{F} \times \mathcal{L}^{\varphi}$. Atom identities use the 7-bit continuous analog-bit encoding of Chen et al. (2023) adopted by FlowMM, giving $\boldsymbol{z}^A \in \mathcal{A} = \mathbb{R}^{7n}$. Fractional coordinates $\boldsymbol{f} \in \mathcal{F} = (\mathbb{R}/\mathbb{Z})^{3n}$ form a flat $3n$-torus. The lattice is always handled through an unconstrained chart $\boldsymbol{l}^{\varphi} = \varphi(\boldsymbol{l}) \in \mathcal{L}^{\varphi} = \mathbb{R}^6$ (inverse softplus on the three lengths, logit on the three angles over the angle range of Miller et al. (2024); Appendix C.9), whose inverse is applied once at the end of sampling.

Why this representation is flat, and why that matters. We endow the lattice with the Euclidean metric in the φ-chart, under which the lattice factor is isometric to $\mathbb{R}^6$, while fractional coordinates

Table 2: **Notation and acronyms used throughout the paper.** Tangent vectors carry a bold script; ideal (target) quantities carry a subscript $\star$, network outputs carry a subscript $\boldsymbol{\theta}$.

Symbol / acronym	Meaning
$\mathcal{M} = \mathcal{A} \times \mathcal{F} \times \mathcal{L}^{\varphi}$	Crystal product manifold: atom types, atomic coords (fractional), lattice
$\boldsymbol{z}_t = (\boldsymbol{z}_t^A, \boldsymbol{f}_t, \boldsymbol{l}_t)$	A point on $\mathcal{M}$ at time $t \in [0,1]$ (data at t=0, prior at t=1)
$\phi_t(\boldsymbol{z}_0, \boldsymbol{z}_1)$	Conditional interpolation bridge with data $\boldsymbol{z}_0 \sim p_0$, prior $\boldsymbol{z}_1 \sim p_1$
$\boldsymbol{v}_\star(\boldsymbol{z}, t)$	Ideal instantaneous velocity field (flow-matching target)
$\boldsymbol{u}_\star(\boldsymbol{z}_t, r, t)$	Ideal finite-horizon average velocity over $[r, t]$
$\bar{\boldsymbol{v}}_{\mathrm{sg}}(\boldsymbol{z}_t, r, t)$	Bootstrapped semigroup horizon target of Eq. (7) (two no-grad half-jumps)
$\boldsymbol{\theta}^-$	Gradient-free target copy of the parameters, $\boldsymbol{\theta}^- = \mathrm{sg}(\boldsymbol{\theta})$
$\boldsymbol{a}_\star, \boldsymbol{b}_\star$	Ideal local coefficient $\boldsymbol{v}_\star$ and first-order horizon coefficient $-\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star$
$\boldsymbol{a}_{\boldsymbol{\theta}}, \boldsymbol{b}_{\boldsymbol{\theta}}$	Network readouts of the two coefficient fields
$\boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z}_t, r, t)$	SRTF field $\boldsymbol{a}_{\boldsymbol{\theta}} + \tau\, \boldsymbol{b}_{\boldsymbol{\theta}}$
$\tau = t - r$	Horizon length; t endpoint time, $r \le t$ start of the transport interval
$\gamma(\tau)$	Scale profile; the main text uses the canonical linear choice $\gamma(\tau) = \tau$
ω_c; $\lambda_{\mathrm{loc}}(t), \lambda_{\mathrm{hor}}(t)$	Per-factor loss weights; local/horizon schedule weights
$\Psi_{r,t}$	Jump map induced by $\boldsymbol{w}_{\boldsymbol{\theta}}$ over $[r, t]$ (Section 3.3)
$\mathrm{Exp}_{\boldsymbol{z}}, \mathrm{Log}_{\boldsymbol{z}}$	Riemannian exponential and logarithm maps at $\boldsymbol{z}$
$\mathcal{L}_{\mathrm{loc}}, \mathcal{L}_{\mathrm{hor}}$	Local flow-matching loss, horizon semigroup consistency loss
NFE	Neural function evaluations per sample
FM, MF	Flow matching, MeanFlow (horizon-averaged FM)
CSP	Crystal structure prediction
MR, RMSE	Match rate, root-mean-square error of matched pairs
JVP	Jacobian–vector product (forward-mode tangent through the model)
CSPNet	EGNN-style equivariant backbone of Miller et al. (2024)

form a flat torus and analog-bit types are Euclidean. In these coordinates Christoffel symbols vanish and $\mathrm{Exp}_{\boldsymbol{z}}(\boldsymbol{v}) = \boldsymbol{z} + \boldsymbol{v}$ on $\mathcal{A}, \mathcal{L}^{\varphi}$ and $(\boldsymbol{z} + \boldsymbol{v}) \bmod 1$ on $\mathcal{F}$. Two consequences are used throughout: finite jumps compose *exactly* through the factor-wise $\mathrm{Exp}/\mathrm{Log}$ maps, with no parallel-transport or curvature corrections—this is what makes the semigroup target a plain average of forward evaluations, with no Log—and the covariant material derivative reduces to $\mathrm{D}_t f = \partial_t f + (\boldsymbol{v}_\star \cdot \nabla) f$. We restrict attention to this flat representation; curved feasible sets would require covariant corrections (Appendix C.7). On $\mathcal{F}$, $\boldsymbol{v}_\star$ denotes the mean-subtracted field; by translation invariance its flow differs from the unprojected one by a global shift, so all statements on $\mathcal{F}$ hold modulo translation.

Flow matching and the finite-horizon velocity. Flow matching (Lipman et al., 2023; Chen & Lipman, 2024) learns the marginal velocity $\boldsymbol{v}_\star(\boldsymbol{z}_t, t) = \mathbb{E}[\boldsymbol{v}^{\mathrm{cond}} \mid \boldsymbol{z}_t]$ along a conditional bridge $\boldsymbol{z}_t = \phi_t(\boldsymbol{z}_0, \boldsymbol{z}_1)$, where $\boldsymbol{v}^{\mathrm{cond}}$ is the velocity of the sampled bridge path; we use the canonical linear bridge $(1-t)\boldsymbol{z}_0 + t\boldsymbol{z}_1$ on each Euclidean factor and on the lifted cover of $\mathcal{F}$. The *marginal flow line* through a state is the trajectory of $\dot{\boldsymbol{z}}_s = \boldsymbol{v}_\star(\boldsymbol{z}_s, s)$, and integrating it in reverse time takes 50–200 steps in practice. Few-step sampling instead needs the finite-horizon velocity $\boldsymbol{u}_\star$ of Eq. (1), the average of $\boldsymbol{v}_\star$ along that flow line over $[r, t]$. MeanFlow (Geng et al., 2025a) targets the same object through the self-consistency identity $\boldsymbol{u}_\star = \boldsymbol{v}_\star - \tau\, \mathrm{D}_t \boldsymbol{u}_\star$, which requires a JVP through the model. SRTF targets $\boldsymbol{u}_\star$ with no JVP. A full symbol list is in Table 2.

B Extended related work and conceptual comparison

Section 2 states the positioning in one paragraph; this appendix gives it in full.

Generative models for crystalline materials. Deep generative modeling for periodic crystals spans variational, diffusion, flow-based and transformer/LLM-based approaches. The core baselines are CDVAE (Xie et al., 2022), DiffCSP (Jiao et al., 2023), MatterGen (Zeni et al., 2025) and FlowMM (Miller et al., 2024), and several variants share their product-manifold or graph formulation: DP-CDVAE (Pakornchote et al., 2024) replaces CDVAE's Langevin corrector with a diffusion probabilistic model, and CrystalFlow (Luo et al., 2025) combines continuous normalizing flows with symmetry-aware graph inputs. Transformer- and LLM-based generators (Gruver et al., 2024; Cao et al., 2025) tokenize crystals or condition on space-group labels, FlowLLM (Sriram et al., 2024)

Table 3: **Where SRTF sits.** "JVP?" asks whether training differentiates through the model to build its own target; "Identified head?" asks whether the objective identifies an explicit horizon coefficient of the learned field.

Method	Learned target	Manifold	JVP?	Identified head?	Few-step?
Flow matching (Lipman et al., 2023; Chen & Lipman, 2024)	instantaneous $v_\star$	Riemannian	No	–	No
MeanFlow (Geng et al., 2025a)	finite-horizon $u_\star$	Euclidean	**Yes**	No	Yes
Shortcut models (Frans et al., 2025)	flow-map composition	Euclidean	No	No	Yes
Riemannian MeanFlow (Woo et al., 2026; Zhong et al., 2026)	finite-horizon $u_\star$	Riemannian	Yes; No (semigroup form)	No	Yes
SRTF (ours)	structured $\boldsymbol{a} + \tau \boldsymbol{b}$ via semigroup	flat Riemannian	**No**	**Yes** (Prop. 1 (iii))	Yes

uses a fine-tuned LLM as the base distribution for a FlowMM refinement stage, and symmetry-aware Bayesian flow networks (Ruple et al., 2026) reproduce the empirical distribution of space groups. A separate line imposes physical or symmetry constraints on the generator: DiffCSP++ (Jiao et al., 2024) reduces the space-group constraint to a basis constraint on the lattice and a Wyckoff constraint on the coordinates; SCIGEN (Okabe et al., 2026) and CHGGen (Zhong et al., 2025) enforce lattice-template constraints through conditional sampling and inpainting; CrysVCD (Cheng et al., 2026) guarantees valence balance; SCDM (Xu et al., 2025) extends symmetry-constrained diffusion to 2D materials; and PIDM (Bastek et al., 2025) couples generative modeling to PDE fulfillment. All of them require tens to thousands of function evaluations per sample, and all constrain *what* is generated. SRTF targets the same product manifold, backbone and evaluation protocol as FlowMM but learns a *finite-horizon average* velocity decomposed into a local flow-matching anchor and a horizon correction, changing *how many steps* generation takes; it is therefore composable with the constrained-generation mechanisms above. Recent surveys (Metni et al., 2026; De Breuck et al., 2025) give broader taxonomies.

Few-step generation: flow matching, Riemannian extensions and MeanFlow. Flow matching (Lipman et al., 2023; Liu et al., 2023) trains an instantaneous velocity field, and Riemannian flow matching (Chen & Lipman, 2024) extends it to manifolds through $\mathrm{Exp}/\mathrm{Log}$ maps, which is what powers FlowMM. Few-step extensions in Euclidean image space split into consistency-based methods (Song et al., 2023; Geng et al., 2025b; Lu & Song, 2025), which typically distill from or initialize with a pretrained model, and distillation-free ones: shortcut models (Frans et al., 2025) close a training signal across step sizes, while MeanFlow (Geng et al., 2025a) parameterizes the finite-horizon average velocity through a self-consistent identity requiring a Jacobian–vector product through the model. A growing line refines that objective—sharpening the regression target (Geng et al., 2026), applying it in pixel space (Lu et al., 2026), addressing the curvature bottleneck (Zhang et al., 2025), or adding a noise-injection refinement step (Huang et al., 2026)—alongside work that revisits the prediction target of pixel-space denoisers (Li & He, 2026); complementary work reduces cost through latent-space computation (Dao et al., 2023) or enforces hard constraints through inference-time corrections (Utkarsh et al., 2025), both of which act on the pipeline rather than on the field and are orthogonal to the decomposition here. Riemannian MeanFlow (Woo et al., 2026; Zhong et al., 2026) carries the average-velocity objective to manifolds; its derivative-based forms use JVPs, while the semigroup form of Woo et al. (2026) composes two model jumps through $\mathrm{Exp}/\mathrm{Log}$ without one. Relative to the Euclidean methods, SRTF is defined on the crystal product manifold; relative to all of them, it addresses one-step quality through a structured coefficient-field decomposition with a linear scale profile, trained with a JVP-free *semigroup* target that keeps the training graph first-order. It is closest in spirit to the self-composition consistency of shortcut models and to semigroup Riemannian MeanFlow, differing in that on the flat crystal manifold composition reduces to displacement addition (no Log in the target) and that the object composed is a structured field $\boldsymbol{a}_{\boldsymbol{\theta}} + \tau \boldsymbol{b}_{\boldsymbol{\theta}}$ whose fixed point identifies, to leading order, Taylor coefficients of the ideal horizon-averaged velocity (Proposition 1 (iii)), rather than a monolithic flow map.

Backbones and evaluators. The methods above rely on geometric backbones such as steerable tensor-field networks (Thomas et al., 2018; Geiger & Smidt, 2022), EGNN/SEGNN (Satorras et al., 2021; Brandstetter et al., 2022) (as in EDM (Hoogeboom et al., 2022)), PaiNN (Schütt et al., 2021), E2GNN (Yang et al., 2025) or graph transformers (Ying et al., 2021), and generated structures are evaluated with universal machine-learning interatomic potentials including NequIP (Batzner et al., 2022), M3GNet (Chen & Ong, 2022), MACE (Batatia et al., 2022) and CHGNet (Deng et al., 2023). SRTF reuses FlowMM's CSPNet backbone, so these choices are orthogonal to the decomposition.

C THEORY: PROOFS AND DERIVATIONS

C.1 FINITE-HORIZON TRANSPORT EQUALS NORMALIZED FLOW DISPLACEMENT

The following identity is elementary—it is the fundamental theorem of calculus together with the fact that the Riemannian logarithm on a flat manifold is a componentwise difference—but it is used repeatedly below, so we record it once.

Lemma 1 (Finite-horizon transport equals normalized flow displacement)**.** *Let $\mathcal{M}$ be a product of Euclidean spaces and unit tori $(\mathbb{R}/\mathbb{Z})^k$, such as the crystal manifold, and let $\boldsymbol{v}_\star \in C^1(\mathcal{M} \times [0,1])$ be a time-dependent velocity field whose flow $\dot{\boldsymbol{z}}_s = \boldsymbol{v}_\star(\boldsymbol{z}_s, s)$ exists and is unique on $[0,1]$ (this holds whenever $\boldsymbol{v}_\star$ is locally Lipschitz in $\boldsymbol{z}$ and bounded on the relevant compact region). Restrict to flow segments whose unwrapped displacement satisfies $\|\boldsymbol{z}_t - \boldsymbol{z}_r\|_\infty < \frac{1}{2}$ on every torus factor, so that $\mathrm{Log}_{\boldsymbol{z}_t}(\boldsymbol{z}_r)$ is the unique shortest geodesic and the integral form does not cross the cut locus. (The bound must be strict: at exactly $\frac{1}{2}$ two minimizing geodesics compete and* Log *is not single-valued.) For any such flow line $\{\boldsymbol{z}_s\}$ of $\boldsymbol{v}_\star$ and any $0 \le r < t \le 1$,*

$$-\tfrac{1}{t-r}\,\mathrm{Log}_{\boldsymbol{z}_t}(\boldsymbol{z}_r) \;=\; \tfrac{1}{t-r}\int_r^t \boldsymbol{v}_\star(\boldsymbol{z}_s, s)\,\mathrm{d}s \;=\; \boldsymbol{u}_\star(\boldsymbol{z}_t, r, t). \tag{12}$$

Proof. Let $\{\boldsymbol{z}_s\}_{s\in[r,t]}$ denote the unique flow line ending at $\boldsymbol{z}_t$ at time t. Throughout the proof we use that each factor of $\mathcal{M}$ is a quotient of a Euclidean space by translations, so holonomy is trivial and parallel transport along any path is path-independent and provides a canonical isomorphism between tangent spaces along the flow. We implicitly invoke this isomorphism when writing $\int_r^t \boldsymbol{v}_\star(\boldsymbol{z}_s, s)\,\mathrm{d}s$ as a single element of $T_{\boldsymbol{z}_t}\mathcal{M}$ and when interpreting $\boldsymbol{z}_t - \boldsymbol{z}_r$ as a tangent vector; on the torus factor $\mathcal{F}$, the identification is carried out in the lifted cover, and the resulting tangent vector is the representative of the shortest geodesic joining $\boldsymbol{z}_r$ and $\boldsymbol{z}_t$ (equivalently, the mod-1 signed difference).

Step 1 (definition). By the definition of the finite-horizon average velocity,

$$(t-r)\,\boldsymbol{u}_\star(\boldsymbol{z}_t, r, t) \;=\; \int_r^t \boldsymbol{v}_\star(\boldsymbol{z}_s, s)\,\mathrm{d}s.$$

This is the middle equality in (12).

Step 2 (flow-line integral). Because $\boldsymbol{z}_s$ is a flow line of $\boldsymbol{v}_\star$, $\dot{\boldsymbol{z}}_s = \boldsymbol{v}_\star(\boldsymbol{z}_s, s)$. Under the canonical flat-manifold identification of tangent spaces just described, the Fundamental Theorem of Calculus gives

$$\int_r^t \boldsymbol{v}_\star(\boldsymbol{z}_s, s)\,\mathrm{d}s \;=\; \int_r^t \dot{\boldsymbol{z}}_s\,\mathrm{d}s \;=\; \boldsymbol{z}_t - \boldsymbol{z}_r,$$

where the difference $\boldsymbol{z}_t - \boldsymbol{z}_r$ is taken in the canonical chart (e.g., componentwise on $\mathcal{A}$, $\mathcal{L}^\varphi$, and the lifted cover of $\mathcal{F}$).

Step 3 (log-map chord). On a flat manifold, the Riemannian logarithm in the canonical chart is the componentwise difference,

$$\mathrm{Log}_{\boldsymbol{z}_t}(\boldsymbol{z}_r) \;=\; \boldsymbol{z}_r - \boldsymbol{z}_t \;=\; -(\boldsymbol{z}_t - \boldsymbol{z}_r).$$

Combining Steps 1–3,

$$\boldsymbol{u}_\star(\boldsymbol{z}_t, r, t) \;=\; \frac{\boldsymbol{z}_t - \boldsymbol{z}_r}{t-r} \;=\; -\frac{1}{t-r}\,\mathrm{Log}_{\boldsymbol{z}_t}(\boldsymbol{z}_r),$$

which is the three-way identity (12). □

Remark 1 (Exact composition of jumps, at any size). Lemma 1 is a statement about the *chord*, and its cut-locus hypothesis is needed only because Log appears. The composition property used by the objective is separate and unconditional: on each factor of $\mathcal{M}$ the exponential map is translation in the canonical chart ($\boldsymbol{z} \mapsto \boldsymbol{z} + \boldsymbol{w}$ on $\mathcal{A}$ and $\mathcal{L}^\varphi$, $\boldsymbol{z} \mapsto (\boldsymbol{z} + \boldsymbol{w}) \bmod 1$ on $\mathcal{F}$), and translations form a group, so

$$\mathrm{Exp}_{\mathrm{Exp}_{\boldsymbol{z}}(\boldsymbol{w}_1)}(\boldsymbol{w}_2) \;=\; \mathrm{Exp}_{\boldsymbol{z}}(\boldsymbol{w}_1 + \boldsymbol{w}_2)$$

for *any* $\boldsymbol{w}_1, \boldsymbol{w}_2$, however large. Consecutive jumps therefore compose exactly at any step size, including the one-step regime $\tau = 1$ where the chord identity does not apply.

C.2 First-order consistency of the transport expansion

Proposition 2 (First-order consistency of the transport expansion)**.** *Let $\boldsymbol{v}_\star \in C^3(\mathcal{M} \times [0,1])$, let the backward flow segment $\{\boldsymbol{z}_s\}_{s\in[r,t]}$ through $\boldsymbol{z}_t$ exist, be unique, and remain in a compact set on which the C^3 bounds hold, let $\boldsymbol{a}_\star(\boldsymbol{z}_t,t) = \boldsymbol{v}_\star(\boldsymbol{z}_t,t)$, and let $\boldsymbol{b}_\star(\boldsymbol{z}_t,t) = -\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star(\boldsymbol{z}_t,t)$, where $\mathrm{D}_t = \partial_t + \boldsymbol{v}_\star\cdot\nabla$ denotes the material derivative along the flow. Define*

$$\bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z}_t,r,t) \;=\; \boldsymbol{a}_\star(\boldsymbol{z}_t,t) + \gamma(\tau)\boldsymbol{b}_\star(\boldsymbol{z}_t,t), \qquad \gamma(\tau)=\tau.$$

Then, uniformly over $(\boldsymbol{z}_t,t,r)$ in any compact set,

$$\boldsymbol{u}_\star(\boldsymbol{z}_t,r,t) - \bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z}_t,r,t) \;=\; \mathcal{O}(\tau^2), \qquad \tau = t-r.$$

In particular $\bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z}_t,t,t) = \boldsymbol{v}_\star(\boldsymbol{z}_t,t)$ exactly, and the leading remainder is a quadratic function of τ whose coefficient depends only on $\mathrm{D}_t^2\boldsymbol{v}_\star$.

Proposition 2 is a statement about the *form* of the SRTF field and does not depend on the training procedure. Any smooth profile with $\gamma(0)=0$ and $\gamma'(0)=1$ matches the local velocity $\boldsymbol{v}_\star$ at $\tau{=}0$; the linear choice $\gamma(\tau){=}\tau$ is the simplest representative and additionally satisfies $\gamma''(0)=0$, so the second-order residual depends only on the intrinsic flow quantity $\mathrm{D}_t^2\boldsymbol{v}_\star$. Profiles with $\gamma''(0)\neq 0$ shift the second-order residual by an extra term proportional to $\gamma''(0)\,\boldsymbol{b}_\star$; the ablation of Appendix F.4 tests one such profile, $\tau(1-\tau/2)$, alongside $\frac{2}{\pi}\sin(\pi\tau/2)$, which is normalized so that $\gamma'(0)=1$ and agrees with the linear profile to second order. Both leave accuracy unmoved, as the first-order analysis predicts. A profile with $\gamma'(0)\neq 1$—the unnormalized $\sin(\pi\tau/2)$, say—is a different matter: it rescales the identified coefficient to $-\frac{1}{2\gamma'(0)}\mathrm{D}_t\boldsymbol{v}_\star$ and is not first-order equivalent at all.

Proof. Fix $(\boldsymbol{z}_t,t)$ and write $\tau = t-r$. Parameterize the flow by $s = t-u$ with $u\in[0,\tau]$ so that $\boldsymbol{z}_{t-u}$ is the backward flow of $\boldsymbol{v}_\star$ starting at $\boldsymbol{z}_t$. A standard third-order Taylor expansion of $\boldsymbol{v}_\star$ along the flow gives

$$\boldsymbol{v}_\star(\boldsymbol{z}_{t-u},t-u) \;=\; \boldsymbol{v}_\star(\boldsymbol{z}_t,t) - u\cdot\mathrm{D}_t\boldsymbol{v}_\star(\boldsymbol{z}_t,t) + \tfrac{u^2}{2}\,\mathrm{D}_t^2\boldsymbol{v}_\star(\boldsymbol{z}_t,t) + \mathcal{O}(u^3),$$

where $\mathrm{D}_t = \partial_t + \boldsymbol{v}_\star\cdot\nabla$ is the material derivative along the flow. Integrating from $u=0$ to $u=\tau$ and dividing by τ,

$$\begin{aligned}\boldsymbol{u}_\star(\boldsymbol{z}_t,r,t) &= \frac{1}{\tau}\int_0^\tau \boldsymbol{v}_\star(\boldsymbol{z}_{t-u},t-u)\,\mathrm{d}u \\ &= \boldsymbol{v}_\star(\boldsymbol{z}_t,t) - \tfrac{\tau}{2}\,\mathrm{D}_t\boldsymbol{v}_\star(\boldsymbol{z}_t,t) + \tfrac{\tau^2}{6}\,\mathrm{D}_t^2\boldsymbol{v}_\star(\boldsymbol{z}_t,t) + \mathcal{O}(\tau^3).\end{aligned}$$

Plugging $\boldsymbol{a}_\star(\boldsymbol{z}_t,t) = \boldsymbol{v}_\star(\boldsymbol{z}_t,t)$, $\boldsymbol{b}_\star(\boldsymbol{z}_t,t) = -\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star(\boldsymbol{z}_t,t)$, and $\gamma(\tau)=\tau$ into $\bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z}_t,r,t) = \boldsymbol{a}_\star + \gamma(\tau)\boldsymbol{b}_\star$ yields

$$\boldsymbol{u}_\star(\boldsymbol{z}_t,r,t) - \bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z}_t,r,t) \;=\; \tfrac{\tau^2}{6}\,\mathrm{D}_t^2\boldsymbol{v}_\star(\boldsymbol{z}_t,t) + \mathcal{O}(\tau^3) \;=\; \mathcal{O}(\tau^2),$$

uniformly in any compact set, under the C^3 assumption on $\boldsymbol{v}_\star$. The two special cases $\bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z}_t,t,t) = \boldsymbol{a}_\star(\boldsymbol{z}_t,t) = \boldsymbol{v}_\star(\boldsymbol{z}_t,t)$ and the quadratic form of the leading remainder follow directly. □

C.3 Identification: bridge-chord degeneracy and the semigroup fixed point

This subsection contains the two results that justify the SRTF training objective of Section 3.3: a *negative* result (Lemma 2) showing that the natural JVP-free alternative—a log-map chord read from two samples of the *same* conditional bridge—cannot supervise the horizon correction; and a *positive* result (Proposition 1) showing that the semigroup objective of Eq. (7) has the ideal horizon-averaged velocity $\boldsymbol{u}_\star$ as its unique fixed point and recovers the SRTF coefficients $(\boldsymbol{a}_\star,\boldsymbol{b}_\star) = (\boldsymbol{v}_\star, -\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star)$ at first order. Remark 4 then re-derives the same coefficient as the solution of a $2{\times}2$ moment system, under the stronger hypothesis of an oracle horizon target whose conditional expectation equals $\boldsymbol{u}_\star$.

Lemma 2 (Degeneracy of same-path bridge chords)**.** *Let $\mathcal{M}$ be the flat product manifold with the constant-speed conditional bridge $\boldsymbol{z}_s = \phi_s(\boldsymbol{z}_0,\boldsymbol{z}_1)$ of Section 3 (linear on $\mathcal{A}$, $\mathcal{L}^\varphi$; constant-speed geodesic on the lifted cover of $\mathcal{F}$), and let $0\le r<t\le 1$.* ***(i) Sample-wise collapse.*** *The bridge chord satisfies, for every sample,*

$$\bar{\boldsymbol{v}}_{\mathrm{chord}}(\boldsymbol{z}_t,r,t) \;:=\; -\tfrac{1}{t-r}\,\mathrm{Log}_{\boldsymbol{z}_t}(\boldsymbol{z}_r) \;=\; \boldsymbol{z}_1 - \boldsymbol{z}_0 \;\equiv\; \boldsymbol{v}_{\mathrm{c}}^{\mathrm{cond}},$$

so $\mathbb{E}[\bar{\boldsymbol{v}}_{\mathrm{chord}} \mid \boldsymbol{z}_t] = \boldsymbol{v}_\star(\boldsymbol{z}_t, t)$*, with no dependence on* (r, t) *beyond the time index of* $\boldsymbol{z}_t$*.* ***(ii) Forced*** $\boldsymbol{b}^* = \mathbf{0}$***.*** *Let* $(\boldsymbol{a}^*, \boldsymbol{b}^*)$ *minimize* $\lambda_{\mathrm{loc}}\mathcal{L}_{\mathrm{loc}} + \lambda_{\mathrm{hor}}\mathcal{L}_{\mathrm{hor}}$ *over an arbitrarily expressive class* $\{\boldsymbol{a}(\boldsymbol{z}, t) + \gamma(\tau)\boldsymbol{b}(\boldsymbol{z}, t)\}$*, where* $\mathcal{L}_{\mathrm{hor}}$ *regresses* $\boldsymbol{w}$ *on* $\bar{\boldsymbol{v}}_{\mathrm{chord}}$*, with* $\lambda_{\mathrm{loc}}, \lambda_{\mathrm{hor}} > 0$ *and a* τ*-sampling law under which* γ *is non-degenerate,* $\mathrm{Var}_\tau[\gamma(\tau)] > 0$ *(equivalently* $\bar{\gamma}_2 > \bar{\gamma}_1^2$*, the same condition used in Remark 4). (Since* $\lambda_{\mathrm{loc}} > 0$ *pins* $\boldsymbol{a}^* = \boldsymbol{v}_\star$*,* $\mathbb{E}_\tau[\gamma(\tau)^2] > 0$ *in fact suffices.) Then* $\boldsymbol{a}^* = \boldsymbol{v}_\star$ *and* $\boldsymbol{b}^* = \mathbf{0}$ *almost everywhere.* ***(iii) Chord-type targets cannot carry curvature.*** *More generally, for any measurable rescaling* $\psi(r, t) > 0$*, the chord-type target* $\psi(r, t)\,(\boldsymbol{z}_t - \boldsymbol{z}_r)$ *read from two points of the same constant-speed conditional path has conditional expectation* $\psi(r, t)(t - r)\,\boldsymbol{v}_\star(\boldsymbol{z}_t, t)$*: a* state-independent scalar *times the local target. Consequently, under the same non-degeneracy condition, the population optimum of any chord-type horizon loss satisfies* $\boldsymbol{b}^*(\boldsymbol{z}, t) \in \mathrm{span}\{\boldsymbol{v}_\star(\boldsymbol{z}, t)\}$ *with a state-independent coefficient—a learned step-size schedule—and can never equal the trajectory-curvature coefficient* $\boldsymbol{b}_\star = -\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star$*, which is not parallel to* $\boldsymbol{v}_\star$ *in general.*

Proof. (i) On a constant-speed path, $\boldsymbol{z}_t - \boldsymbol{z}_r = (t - r)(\boldsymbol{z}_1 - \boldsymbol{z}_0)$ componentwise in the canonical chart (lifted cover on $\mathcal{F}$; the displacement stays within the injectivity radius for bridge samples in one fundamental domain), and $\mathrm{Log}_{\boldsymbol{z}_t}(\boldsymbol{z}_r) = \boldsymbol{z}_r - \boldsymbol{z}_t$. Dividing by $t - r$ gives the claim; taking $\mathbb{E}[\,\cdot \mid \boldsymbol{z}_t]$ yields the marginal velocity $\boldsymbol{v}_\star(\boldsymbol{z}_t, t)$ by the standard conditional-to-marginal identity (Lipman et al., 2023). *(ii)* By (i), both losses regress $\boldsymbol{w}(\boldsymbol{z}_t, r, t) = \boldsymbol{a} + \gamma(\tau)\boldsymbol{b}$ against targets with identical conditional expectation $\boldsymbol{v}_\star(\boldsymbol{z}_t, t)$ for every (r, t). The population loss therefore decomposes as an irreducible conditional variance plus the bias $\mathbb{E}_{\boldsymbol{z},t}\mathbb{E}_\tau\|\boldsymbol{a} + \gamma(\tau)\boldsymbol{b} - \boldsymbol{v}_\star\|^2$ (weighted by λ's), which is minimized iff $\boldsymbol{a}(\boldsymbol{z}, t) + \gamma(\tau)\boldsymbol{b}(\boldsymbol{z}, t) = \boldsymbol{v}_\star(\boldsymbol{z}, t)$ for a.e. τ in the support. Non-degeneracy supplies two support points with $\gamma(\tau_1) \neq \gamma(\tau_2)$, which force $\boldsymbol{b} = \mathbf{0}$, and then $\boldsymbol{a} = \boldsymbol{v}_\star$. *(iii)* Identical computation with the rescaled target: the bias term becomes $\mathbb{E}_\tau\|\boldsymbol{a} + \gamma(\tau)\boldsymbol{b} - c(\tau, t)\boldsymbol{v}_\star\|^2$ with the state-independent scalar $c(\tau, t) = \psi(t - \tau, t)\,\tau$, whose minimizer over $(\boldsymbol{a}, \boldsymbol{b})$ is a pair of state-independent *scalar multiples* of $\boldsymbol{v}_\star$ (unique because the Gram matrix $\left(\begin{smallmatrix} 1 & \bar{\gamma}_1 \\ \bar{\gamma}_1 & \bar{\gamma}_2 \end{smallmatrix}\right)$ is positive definite under $\bar{\gamma}_2 > \bar{\gamma}_1^2$; project the first-order conditions onto $\boldsymbol{v}_\star$ and its orthogonal complement: the orthogonal component of both $\boldsymbol{a}^*$ and $\boldsymbol{b}^*$ vanishes). Hence $\boldsymbol{b}^* \parallel \boldsymbol{v}_\star$ pointwise with a coefficient depending only on t and the τ-law. □

Lemma 2 explains, mechanistically, why zeroing $\boldsymbol{b}_{\boldsymbol{\theta}}$ at inference had almost no effect under the bridge-chord objective (cf. the ablation discussion in Section 4): the linear-bridge chord provides no horizon supervision at all. Escaping the degeneracy requires a target that samples the *marginal flow* rather than a conditional path: either by integrating trajectories of a learned field (self-distillation) or by bootstrapping the model's own finite jumps through the algebraic property that characterizes $\boldsymbol{u}_\star$. SRTF adopts the latter—the semigroup objective of Eq. (7)—because it is JVP-free, requires no teacher rollouts, and is exact on flat $\mathcal{M}$. A complementary requirement is recorded in Remark 7: at a single horizon the split of $\boldsymbol{w}_{\boldsymbol{\theta}}$ into $(\boldsymbol{a}_{\boldsymbol{\theta}}, \boldsymbol{b}_{\boldsymbol{\theta}})$ is defined only up to a gauge, so the horizon supervision must sample at least two horizons, as the τ-law of Eq. (8) does.

Proposition 1, restated with full hypotheses. *Let* $\boldsymbol{v}_\star \in C^3(\mathcal{M} \times [0, 1])$ *be Lipschitz in* $\boldsymbol{z}$ *uniformly in* t *on the compact region of interest (*C^1 *suffices for (i)–(ii)). No cut-locus condition is needed:* $\boldsymbol{u}_\star$ *is defined by the integral in the lifted cover of* $\mathcal{F}$*, and jumps compose exactly at any size (Remark 1). For a field* $\boldsymbol{w}(\boldsymbol{z}, r, t) \in T_{\boldsymbol{z}}\mathcal{M}$ *define the semigroup residual, for* $0 \le r < s < t \le 1$*,*

$$\mathcal{R}[\boldsymbol{w}](\boldsymbol{z}_t, r, s, t) \;:=\; (t - r)\,\boldsymbol{w}(\boldsymbol{z}_t, r, t) \;-\; (t - s)\,\boldsymbol{w}(\boldsymbol{z}_t, s, t) \;-\; (s - r)\,\boldsymbol{w}(\hat{\boldsymbol{z}}_s, r, s),$$

where $\hat{\boldsymbol{z}}_s = \mathrm{Exp}_{\boldsymbol{z}_t}\big(-(t - s)\boldsymbol{w}(\boldsymbol{z}_t, s, t)\big)$ *is the state reached by the first leg.* ***(i) Fixed point.*** *The ideal horizon-averaged velocity satisfies* $\mathcal{R}[\boldsymbol{u}_\star] \equiv \mathbf{0}$ *and the diagonal boundary condition* $\boldsymbol{u}_\star(\boldsymbol{z}, t, t) = \boldsymbol{v}_\star(\boldsymbol{z}, t)$*; hence* $\boldsymbol{u}_\star$ *is a zero of the population semigroup loss* (8) *under any sampling of states, times, and split points.* ***(ii) Uniqueness.*** *Conversely, let* $\boldsymbol{w}$ *be continuous, Lipschitz in its state argument, satisfy* $\mathcal{R}[\boldsymbol{w}] = \mathbf{0}$ *for all admissible* (r, t) *with midpoint split* $s = (t + r)/2$*, and match the boundary condition* $\boldsymbol{w}(\boldsymbol{z}, t, t) = \boldsymbol{v}_\star(\boldsymbol{z}, t)$*. Then* $\boldsymbol{w} = \boldsymbol{u}_\star$ *on the admissible region.* ***(iii) First-order identification in the SRTF class.*** *Restrict to the structured class* $\boldsymbol{w} = \boldsymbol{a}(\boldsymbol{z}, t) + \tau\,\boldsymbol{b}(\boldsymbol{z}, t)$ *with* $\boldsymbol{a}, \boldsymbol{b} \in C^2$*, and write the split as* $t - s = \xi\tau$*,* $s - r = (1 - \xi)\tau$ *with* $\xi \in (0, 1)$*. Then*

$$\mathcal{R}[\boldsymbol{w}] \;=\; \xi(1 - \xi)\,\tau^2\,\big(2\boldsymbol{b} + \mathrm{D}_t\boldsymbol{a}\big)(\boldsymbol{z}_t, t) \;+\; \mathcal{O}(\tau^3), \qquad \mathrm{D}_t = \partial_t + \boldsymbol{a}\cdot\nabla,$$

so $\mathcal{R}[\boldsymbol{w}] = \mathcal{O}(\tau^3)$ *for some (equivalently, every) interior split iff* $\boldsymbol{b} = -\frac{1}{2}\mathrm{D}_t\boldsymbol{a}$*. In particular, whenever* $\mathrm{D}_t\boldsymbol{v}_\star \not\equiv \mathbf{0}$ *the choice* $\boldsymbol{b} \equiv \mathbf{0}$ *leaves a residual* $\xi(1 - \xi)\tau^2\,\mathrm{D}_t\boldsymbol{v}_\star$ *that no amount of training can remove: the degenerate solution of Lemma 2 is excluded at second order in the displacement.*

Jointly with the anchor $\boldsymbol{a} = \boldsymbol{v}_\star$ enforced by $\mathcal{L}_{\mathrm{loc}}$, the unique first-order-consistent solution in the SRTF class is $(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{v}_\star, -\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star)$—the Taylor coefficients of the transport expansion (3)*—at which the leading residual is $\mathcal{R}[\boldsymbol{w}] = -\frac{1}{2}\xi(1-\xi)\,\tau^3\,\mathrm{D}_t^2\boldsymbol{v}_\star + \mathcal{O}(\tau^4)$, i.e. $\mathcal{O}(\tau^2)$ at velocity scale. This is the same order as the field's own truncation error in Proposition 2, though not the same quantity: the truncation constant there is $+\frac{1}{6}$ and split-independent, whereas the semigroup residual carries $-\frac{1}{2}\xi(1-\xi)$ and measures how far the* composition *fails, not how far the field is from $\boldsymbol{u}_\star$.*

Proof. (i) Let $\{\boldsymbol{z}_\sigma\}$ be the flow line of $\boldsymbol{v}_\star$ through $\boldsymbol{z}_t$. By the definition of $\boldsymbol{u}_\star$ as an integral in the lifted cover, $\boldsymbol{z}_t - \boldsymbol{z}_s = (t-s)\boldsymbol{u}_\star(\boldsymbol{z}_t, s, t)$, so $\hat{\boldsymbol{z}}_s$ computed from $\boldsymbol{u}_\star$ equals the true flow point $\boldsymbol{z}_s$; likewise $\boldsymbol{z}_s - \boldsymbol{z}_r = (s-r)\boldsymbol{u}_\star(\boldsymbol{z}_s, r, s)$ and $\boldsymbol{z}_t - \boldsymbol{z}_r = (t-r)\boldsymbol{u}_\star(\boldsymbol{z}_t, r, t)$. Substituting the three identities into $\mathcal{R}[\boldsymbol{u}_\star]$ telescopes the displacements: $(\boldsymbol{z}_t - \boldsymbol{z}_r) - (\boldsymbol{z}_t - \boldsymbol{z}_s) - (\boldsymbol{z}_s - \boldsymbol{z}_r) = \boldsymbol{0}$. The boundary condition is $\boldsymbol{u}_\star(\boldsymbol{z}, t, t) = \lim_{r\to t} \frac{1}{t-r}\int_r^t \boldsymbol{v}_\star(\boldsymbol{z}_\sigma, \sigma)\mathrm{d}\sigma = \boldsymbol{v}_\star(\boldsymbol{z}, t)$.

(ii) Define the jump map $\Psi_{r,t}(\boldsymbol{z}) := \mathrm{Exp}_{\boldsymbol{z}}\big(-(t-r)\boldsymbol{w}(\boldsymbol{z}, r, t)\big)$. The residual identity $\mathcal{R}[\boldsymbol{w}] = \boldsymbol{0}$ at the midpoint split states exactly $\Psi_{r,t} = \Psi_{r,m} \circ \Psi_{m,t}$ with $m = (t+r)/2$ (on flat $\mathcal{M}$, composition of Exp displacements is displacement addition). Iterating k times over the dyadic refinement of $[r, t]$ expresses the *fixed* map $\Psi_{r,t}$ as the composition of $N = 2^k$ equal sub-jumps of size τ/N:

$$\Psi_{r,t} \;=\; \Psi_{s_0,s_1} \circ \Psi_{s_1,s_2} \circ \cdots \circ \Psi_{s_{N-1},s_N}, \qquad s_j = r + j\tau/N \;\; (s_N = t),$$

each sub-jump being one explicit Euler step $\boldsymbol{z} \mapsto \boldsymbol{z} - (\tau/N)\,\boldsymbol{w}(\boldsymbol{z}, s_{j-1}, s_j)$. By continuity of $\boldsymbol{w}$ and the boundary condition, $\sup_{\boldsymbol{z},j} \|\boldsymbol{w}(\boldsymbol{z}, s_{j-1}, s_j) - \boldsymbol{v}_\star(\boldsymbol{z}, s_j)\| \to 0$ as $N \to \infty$ (uniformly on compacts), so the composition is a consistent one-step scheme for the reverse-time ODE $\dot{\boldsymbol{z}} = \boldsymbol{v}_\star(\boldsymbol{z}, \cdot)$ with vanishing step size; by the classical convergence of the explicit Euler method under the Lipschitz assumption (discrete Grönwall), the composition converges to the exact reverse flow map of $\boldsymbol{v}_\star$ on $[r, t]$. Since the left-hand side $\Psi_{r,t}$ does not depend on N, it *equals* the reverse flow map: $\Psi_{r,t}(\boldsymbol{z}_t) = \boldsymbol{z}_r$ on every flow line. Reading off the displacement in the lifted cover, $(t-r)\boldsymbol{w}(\boldsymbol{z}_t, r, t) - \int_r^t \boldsymbol{v}_\star(\boldsymbol{z}_\sigma, \sigma)\,\mathrm{d}\sigma \in \{\boldsymbol{0}\} \times \mathbb{Z}^{3n} \times \{\boldsymbol{0}\}$; this vector is continuous in r and vanishes as $r \to t$, hence is zero, i.e. $\boldsymbol{w}(\boldsymbol{z}_t, r, t) = \boldsymbol{u}_\star(\boldsymbol{z}_t, r, t)$.

(iii) Write $\boldsymbol{a}, \boldsymbol{b}, \mathrm{D}_t\boldsymbol{a}$ for values at $(\boldsymbol{z}_t, t)$ and expand each term of $\mathcal{R}[\boldsymbol{w}]$ in τ (all remainders uniform on compacts by $\boldsymbol{a}, \boldsymbol{b} \in C^2$). The long jump: $(t-r)\boldsymbol{w}(\boldsymbol{z}_t, r, t) = \tau\boldsymbol{a} + \tau^2\boldsymbol{b}$. The first half-jump: $(t-s)\boldsymbol{w}(\boldsymbol{z}_t, s, t) = \xi\tau\big(\boldsymbol{a} + \xi\tau\,\boldsymbol{b}\big) = \xi\tau\,\boldsymbol{a} + \xi^2\tau^2\,\boldsymbol{b}$. The intermediate state: $\hat{\boldsymbol{z}}_s = \boldsymbol{z}_t - \xi\tau\,\boldsymbol{a} + \mathcal{O}(\tau^2)$ at time $s = t - \xi\tau$, so

$$\boldsymbol{a}(\hat{\boldsymbol{z}}_s, s) = \boldsymbol{a} - \xi\tau\,\mathrm{D}_t\boldsymbol{a} + \mathcal{O}(\tau^2), \qquad \boldsymbol{b}(\hat{\boldsymbol{z}}_s, s) = \boldsymbol{b} + \mathcal{O}(\tau),$$

and the second half-jump contributes $(s-r)\boldsymbol{w}(\hat{\boldsymbol{z}}_s, r, s) = (1-\xi)\tau\big[\boldsymbol{a} - \xi\tau\,\mathrm{D}_t\boldsymbol{a} + (1-\xi)\tau\,\boldsymbol{b}\big] + \mathcal{O}(\tau^3)$. Collecting orders: the $\mathcal{O}(\tau)$ terms cancel identically ($\tau\boldsymbol{a}$ vs. $\xi\tau\boldsymbol{a} + (1-\xi)\tau\boldsymbol{a}$); the $\mathcal{O}(\tau^2)$ terms give

$$\mathcal{R}[\boldsymbol{w}] \;=\; \tau^2\Big[\boldsymbol{b}\,\big(1 - \xi^2 - (1-\xi)^2\big) + \xi(1-\xi)\,\mathrm{D}_t\boldsymbol{a}\Big] + \mathcal{O}(\tau^3) \;=\; \xi(1-\xi)\,\tau^2\,\big(2\boldsymbol{b} + \mathrm{D}_t\boldsymbol{a}\big) + \mathcal{O}(\tau^3),$$

using $1 - \xi^2 - (1-\xi)^2 = 2\xi(1-\xi)$. Since $\xi(1-\xi) > 0$ for interior splits, the τ^2 coefficient vanishes iff $\boldsymbol{b} = -\frac{1}{2}\mathrm{D}_t\boldsymbol{a}$, for every ξ simultaneously. With $\boldsymbol{a} = \boldsymbol{v}_\star$ from the anchor, $(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{v}_\star, -\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star)$. Carrying the expansion one order further at that point gives the leading non-removable residual $-\frac{1}{2}\xi(1-\xi)\tau^3\mathrm{D}_t^2\boldsymbol{v}_\star + \mathcal{O}(\tau^4)$, which is $\mathcal{O}(\tau^2)$ at velocity scale—the same order as, but not equal to, the field truncation $+\frac{\tau^2}{6}\mathrm{D}_t^2\boldsymbol{v}_\star$ of Proposition 2. □

Remark 2 (Split-point invariance). The factor $\xi(1-\xi)$ multiplies the *same* bracket $2\boldsymbol{b} + \mathrm{D}_t\boldsymbol{a}$ for every interior split, so the *first-order identification* of $\boldsymbol{b}^*$ is invariant to where the composition is split: the zero of the leading term is the same for every ξ, and the midpoint $\xi = \frac{1}{2}$ (used in Eq. (7)) merely maximizes the coefficient, and with it the gradient signal. Likewise, with N equal legs, $s_j = r + j\tau/N$, the N-leg residual $(t-r)\boldsymbol{w}(\boldsymbol{z}_t, r, t) - \sum_{j=1}^N \frac{\tau}{N}\boldsymbol{w}(\hat{\boldsymbol{z}}_{s_j}, s_{j-1}, s_j)$, with $\hat{\boldsymbol{z}}_{s_N} = \boldsymbol{z}_t$ and each $\hat{\boldsymbol{z}}_{s_{j-1}}$ reached from $\hat{\boldsymbol{z}}_{s_j}$ by one leg, equals $\frac{N-1}{2N}\,\tau^2(2\boldsymbol{b} + \mathrm{D}_t\boldsymbol{a}) + \mathcal{O}(\tau^3)$: the same bracket. The invariance is a statement about that leading term, not about the exact minimizer: at a τ-law with mass away from zero the optimum sits slightly off $\boldsymbol{b}_\star$, and the $\mathcal{O}(\tau^3)$ term of the residual is not a ξ-independent multiple of $\xi(1-\xi)$ away from $\boldsymbol{b}_\star$, so a weak ξ-dependence survives. The ablation of Appendix F.5 measures it and finds it within seed noise. Note that at a general split the composed target is the interval-weighted average $\xi\boldsymbol{w}_1 + (1-\xi)\boldsymbol{w}_2$ rather than the arithmetic mean; Eq. (7) writes the latter because we fix $\xi = \frac{1}{2}$ throughout, and the two coincide there.

Remark 3 (Fixed points vs. optimization dynamics; gauge). As in analyses of consistency-style training, Proposition 1 characterizes the population *fixed points* of the bootstrapped objective (target copy equal to the current parameters); it does not by itself guarantee convergence of the optimization dynamics to that fixed point. Within the SRTF class, which does not contain $\boldsymbol{u}_\star$, a fixed point is a stationary point of the bootstrap regression at $\boldsymbol{\theta}' = \boldsymbol{\theta}$, $\mathbb{E}\langle \partial_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}}, \mathcal{R}[\boldsymbol{w}_{\boldsymbol{\theta}}]/\tau \rangle = \mathbf{0}$; with $\mathcal{R}/\tau = \xi(1-\xi)\tau(2\boldsymbol{b} + \mathrm{D}_t \boldsymbol{a}) + \mathcal{O}(\tau^2)$ and tangent directions $\delta\boldsymbol{a} + \tau\,\delta\boldsymbol{b}$, the $\delta\boldsymbol{a}$-component gives $2\boldsymbol{b} + \mathrm{D}_t \boldsymbol{a} = \mathcal{O}(\bar{\gamma}_2/\bar{\gamma}_1)$, i.e. part (iii) to leading order in the horizon law, as in Remark 4 (b). The objectives with and without the stop-gradient share this fixed-point set but not their gradient fields, which is the sense in which the stop-gradient affects only the optimization path. The warm-up phase (Appendix D) and the gradient-free target copy of Section 3.3 are the standard stabilizers for this bootstrap. Separately, note that a single semigroup term already couples the three horizons τ, $\xi\tau$, $(1-\xi)\tau$, so the single-horizon gauge freedom of Remark 7 is broken by the fixed-point condition of a *single* semigroup term at self-consistency, before marginalizing over the τ-law; $\mathcal{L}_{\mathrm{loc}}$ remains the boundary anchor that selects $\boldsymbol{u}_\star$ among all self-consistent flow maps (part (ii) above).

Remark 4 (Finite-τ-law form of the identification). Proposition 1 (iii) identifies $\boldsymbol{b}^*$ directly from the objective SRTF optimizes. The following records the same conclusion in a second, more explicit form—as the solution of a 2×2 linear system in the moments of the τ-law—which is useful for seeing how a finite τ-law perturbs the coefficient. It is a corollary of the same expansion, not an independent result, and it rests on a strictly stronger hypothesis (an oracle target), so we state it as a remark. Suppose the horizon regression target used by $\mathcal{L}_{\mathrm{hor}}$ agrees in expectation with the ideal horizon-averaged velocity $\boldsymbol{u}_\star$ of Eq. (1) up to $\mathcal{O}(\tau^2)$ at velocity scale, and fix the canonical scale profile $\gamma(\tau) = \tau$ (so that $\gamma'(0) = 1$ and the explicit Taylor coefficients of the proof below hold). Let the SRTF hypothesis class $\mathcal{H} = \{\boldsymbol{a}_{\boldsymbol{\theta}} + \gamma(\tau)\boldsymbol{b}_{\boldsymbol{\theta}}\}$ be arbitrarily expressive, and let $(\boldsymbol{a}^*, \boldsymbol{b}^*)$ minimize $\lambda_{\mathrm{loc}}\mathcal{L}_{\mathrm{loc}} + \lambda_{\mathrm{hor}}\mathcal{L}_{\mathrm{hor}}$ with $\lambda_{\mathrm{loc}}, \lambda_{\mathrm{hor}} > 0$ and a family of τ-sampling laws that is *uniformly* non-degenerate, $\bar{\gamma}_1^2/\bar{\gamma}_2 \le 1 - c < 1$ with $\bar{\gamma}_2 := \mathbb{E}_\tau[\tau^2]$. Then the minimizer is unique at each law and, as the τ-distribution concentrates at zero, almost everywhere in $(\boldsymbol{z}, t)$,

$$\boldsymbol{a}^*(\boldsymbol{z}, t) \;=\; \boldsymbol{v}_\star(\boldsymbol{z}, t) + o(1), \qquad \boldsymbol{b}^*(\boldsymbol{z}, t) \;=\; -\tfrac{1}{2}\mathrm{D}_t \boldsymbol{v}_\star(\boldsymbol{z}, t) + o(1).$$

For a general smooth profile γ with $\gamma(0) = 0$, $\gamma'(0) = 1$, the same conclusion holds with $\boldsymbol{b}^*$ shifted by a $\gamma''(0)$-dependent scalar; we record the linear case here for explicit Taylor moments.

What Remark 4 does and does not say. Three caveats. **(a) Horizon target.** It assumes the horizon target agrees in expectation with $\boldsymbol{u}_\star$ (the marginal horizon-averaged velocity). The bridge chord of Lemma 2 *fails* this hypothesis (its conditional expectation is $\boldsymbol{v}_\star$); the semigroup target of Eq. (7) *satisfies* it at its fixed point. The $\mathcal{O}(\tau^2)$ slack in the hypothesis is not cosmetic: Proposition 1 (i)–(ii) give an exact fixed point among *unrestricted* fields, but part (iii) shows no member of the structured class $\mathcal{H}$ drives the residual to zero, so an exact-oracle hypothesis would be unsatisfiable by the very object it is meant to describe. The slack costs nothing: the deviation enters the derivation below as $\mathcal{O}(\bar{\gamma}_2)$ in $\bar{\boldsymbol{u}}_1$ and $\mathcal{O}(\bar{\gamma}_3)$ in $\bar{\boldsymbol{u}}_2$, which are the remainder orders it already carries. The remark is therefore a consistency check on Proposition 1 (iii) rather than an independent route to it. **(b) τ-distribution.** The recovery $\boldsymbol{b}^* = -\frac{1}{2}\mathrm{D}_t \boldsymbol{v}_\star$ is stated in the limit of a τ-distribution concentrating at zero; in practice we train with a logit-normal t-law that yields a non-trivial τ-distribution with mass away from zero, so the implemented minimizer differs from $\boldsymbol{b}_\star$ by terms of the order of the horizon moments, which the $o(1)$ term captures and which vanish only as the τ-law concentrates at zero: the identification is a leading-order statement. **(c) Hypothesis class.** It assumes an arbitrarily expressive hypothesis class; finite-capacity effects of the SRTF class are not modelled here and appear separately as ϵ_{SRTF} in Corollary 1. The practical takeaway is that $\mathcal{L}_{\mathrm{loc}}$ and $\mathcal{L}_{\mathrm{hor}}$ target complementary, non-redundant population objects: the boundary condition and the semigroup interior of the same functional equation.

Derivation. Let $\mathcal{H} = \{(\boldsymbol{a}, \boldsymbol{b}) : \boldsymbol{a}, \boldsymbol{b} \in L^2(\mathcal{M} \times [0,1])\}$ be the full L^2 hypothesis class, and let $(\boldsymbol{a}^*, \boldsymbol{b}^*)$ be any minimizer of $\lambda_{\mathrm{loc}}\mathcal{L}_{\mathrm{loc}} + \lambda_{\mathrm{hor}}\mathcal{L}_{\mathrm{hor}}$ with $\lambda_{\mathrm{loc}}, \lambda_{\mathrm{hor}} > 0$. The two loss functionals are

$$\mathcal{L}_{\mathrm{loc}}(\boldsymbol{a}) = \mathbb{E}_{\boldsymbol{z},t}\|\boldsymbol{a}(\boldsymbol{z}, t) - \boldsymbol{v}_\star(\boldsymbol{z}, t)\|^2,$$
$$\mathcal{L}_{\mathrm{hor}}(\boldsymbol{a}, \boldsymbol{b}) = \mathbb{E}_{\boldsymbol{z},r,t}\|\boldsymbol{a}(\boldsymbol{z}, t) + \gamma(\tau)\boldsymbol{b}(\boldsymbol{z}, t) - \boldsymbol{u}_\star(\boldsymbol{z}, r, t)\|^2,$$

where $\boldsymbol{u}_\star$ denotes the ideal horizon-averaged velocity. We carry out the computation under the hypothesis that the horizon regression target agrees with $\boldsymbol{u}_\star$ in expectation. This holds exactly along a true flow line of $\boldsymbol{v}_\star$ on flat $\mathcal{M}$ (Lemma 1) and for the semigroup target at its fixed point

(Proposition 1); it *fails* for same-path bridge chords, for which the population minimum collapses to $\boldsymbol{b}^* = \mathbf{0}$ (Lemma 2). The population first-order conditions are

$$0 = \partial_{\boldsymbol{a}}[\lambda_{\mathrm{loc}}\mathcal{L}_{\mathrm{loc}} + \lambda_{\mathrm{hor}}\mathcal{L}_{\mathrm{hor}}] \;=\; 2\lambda_{\mathrm{loc}}(\boldsymbol{a} - \boldsymbol{v}_\star) + 2\lambda_{\mathrm{hor}}\mathbb{E}_\tau[\boldsymbol{a} + \gamma(\tau)\boldsymbol{b} - \boldsymbol{u}_\star],$$
$$0 = \partial_{\boldsymbol{b}}[\lambda_{\mathrm{hor}}\mathcal{L}_{\mathrm{hor}}] \;=\; 2\lambda_{\mathrm{hor}}\mathbb{E}_\tau[\gamma(\tau)(\boldsymbol{a} + \gamma(\tau)\boldsymbol{b} - \boldsymbol{u}_\star)].$$

Writing $\bar\gamma_k := \mathbb{E}_\tau[\gamma(\tau)^k]$ and $\bar{\boldsymbol{u}}_k(\boldsymbol{z}, t) := \mathbb{E}_\tau[\gamma(\tau)^{k-1}\boldsymbol{u}_\star(\boldsymbol{z}, t-\tau, t)]$, the conditions become a 2×2 linear system in $(\boldsymbol{a}^*, \boldsymbol{b}^*)$:

$$\begin{pmatrix} \lambda_{\mathrm{loc}} + \lambda_{\mathrm{hor}} & \lambda_{\mathrm{hor}}\bar\gamma_1 \\ \lambda_{\mathrm{hor}}\bar\gamma_1 & \lambda_{\mathrm{hor}}\bar\gamma_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{a}^* \\ \boldsymbol{b}^* \end{pmatrix} = \begin{pmatrix} \lambda_{\mathrm{loc}}\boldsymbol{v}_\star + \lambda_{\mathrm{hor}}\bar{\boldsymbol{u}}_1 \\ \lambda_{\mathrm{hor}}\bar{\boldsymbol{u}}_2 \end{pmatrix}.$$

Uniqueness at each non-degenerate τ-law. The coefficient matrix has determinant $\lambda_{\mathrm{hor}}\big[\lambda_{\mathrm{loc}}\bar\gamma_2 + \lambda_{\mathrm{hor}}(\bar\gamma_2 - \bar\gamma_1^2)\big]$, which is strictly positive whenever $\lambda_{\mathrm{loc}}, \lambda_{\mathrm{hor}} > 0$ and $\bar\gamma_2 > \bar\gamma_1^2$ (non-degeneracy of the τ-law), so the minimizer $(\boldsymbol{a}^*, \boldsymbol{b}^*)$ is unique for every such τ-law.

Limit as the τ-law concentrates at zero. Taylor-expanding $\boldsymbol{u}_\star(\boldsymbol{z}, t-\tau, t) = \boldsymbol{v}_\star(\boldsymbol{z}, t) - \frac{\tau}{2}\mathrm{D}_t\boldsymbol{v}_\star(\boldsymbol{z}, t) + \mathcal{O}(\tau^2)$ (Proposition 2) and taking expectations with $\gamma(\tau) = \tau$,

$$\bar{\boldsymbol{u}}_1 \;=\; \boldsymbol{v}_\star - \tfrac{\bar\gamma_1}{2}\mathrm{D}_t\boldsymbol{v}_\star + \mathcal{O}(\bar\gamma_2), \qquad \bar{\boldsymbol{u}}_2 \;=\; \bar\gamma_1\boldsymbol{v}_\star - \tfrac{\bar\gamma_2}{2}\mathrm{D}_t\boldsymbol{v}_\star + \mathcal{O}(\bar\gamma_3).$$

We solve the 2×2 system directly by Cramer's rule. For $\boldsymbol{b}^* = (qA - pB)/\det M$ with $A = \lambda_{\mathrm{loc}} + \lambda_{\mathrm{hor}}$, $B = \lambda_{\mathrm{hor}}\bar\gamma_1$, $p = \lambda_{\mathrm{loc}}\boldsymbol{v}_\star + \lambda_{\mathrm{hor}}\bar{\boldsymbol{u}}_1$, $q = \lambda_{\mathrm{hor}}\bar{\boldsymbol{u}}_2$, the bracketed numerator expands as

$$\begin{aligned}(\lambda_{\mathrm{loc}} + \lambda_{\mathrm{hor}})\bar{\boldsymbol{u}}_2 - \bar\gamma_1\,(\lambda_{\mathrm{loc}}\boldsymbol{v}_\star + \lambda_{\mathrm{hor}}\bar{\boldsymbol{u}}_1) &= \underbrace{[(\lambda_{\mathrm{loc}} + \lambda_{\mathrm{hor}}) - (\lambda_{\mathrm{loc}} + \lambda_{\mathrm{hor}})]\bar\gamma_1\boldsymbol{v}_\star}_{=0\text{ (exact cancellation)}} \\ &\quad - \tfrac{1}{2}\big[(\lambda_{\mathrm{loc}} + \lambda_{\mathrm{hor}})\bar\gamma_2 - \lambda_{\mathrm{hor}}\bar\gamma_1^2\big]\mathrm{D}_t\boldsymbol{v}_\star \\ &\quad + \mathcal{O}(\bar\gamma_3 + \bar\gamma_1\bar\gamma_2) \\ &= -\tfrac{\det M}{2\lambda_{\mathrm{hor}}}\,\mathrm{D}_t\boldsymbol{v}_\star + \mathcal{O}(\bar\gamma_3 + \bar\gamma_1\bar\gamma_2),\end{aligned}$$

using $\det M = \lambda_{\mathrm{hor}}[\lambda_{\mathrm{loc}}\bar\gamma_2 + \lambda_{\mathrm{hor}}(\bar\gamma_2 - \bar\gamma_1^2)]$. The Cramer numerator is λ_{hor} times this bracket; dividing it by $\det M$,

$$\boldsymbol{b}^* \;=\; -\tfrac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star \;+\; \mathcal{O}\!\left(\frac{\bar\gamma_3}{\bar\gamma_2}\right) \;+\; \mathcal{O}(\bar\gamma_1),$$

where we used the non-degeneracy bound $\det M/\lambda_{\mathrm{hor}} \geq (\lambda_{\mathrm{loc}} + c\,\lambda_{\mathrm{hor}})\bar\gamma_2$ for any τ-law with $\bar\gamma_1^2/\bar\gamma_2 \leq 1 - c < 1$ (Jensen, strict under non-degeneracy). An analogous Cramer computation for $\boldsymbol{a}^*$ gives $\boldsymbol{a}^* = \boldsymbol{v}_\star + \mathcal{O}(\bar\gamma_2)$: the $\mathrm{D}_t\boldsymbol{v}_\star$ contributions cancel exactly, since $\bar\gamma_2 \cdot (-\frac{\bar\gamma_1}{2}) - \bar\gamma_1 \cdot (-\frac{\bar\gamma_2}{2}) = 0$. That cancellation is the quantitative content of the statement that the horizon loss does not perturb the diagonal at first order. As the τ-distribution concentrates at 0, $\bar\gamma_1 \to 0$ and $\bar\gamma_3/\bar\gamma_2 \to 0$ under any scaling $\tau = hX$ with X a fixed non-degenerate law on a bounded support (the natural concentration scaling), so $\boldsymbol{a}^* \to \boldsymbol{v}_\star$ and $\boldsymbol{b}^* \to -\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star$, both in L^2, with the stated $o(1)$ residual. The sequence of minimizers (unique for each non-degenerate τ-law) therefore converges to $(\boldsymbol{v}_\star, -\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star)$ as the τ-law concentrates at zero.

Corollary 1 (SRTF avoids the monolithic diagonal bias)**.** *Under the SRTF objective the semigroup loss identifies $\boldsymbol{b}^* = -\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star$ at first order (Proposition 1 (iii)) while perturbing the diagonal only at second order: the $\mathrm{D}_t\boldsymbol{v}_\star$ contributions to $\boldsymbol{a}^*$ cancel exactly in the 2×2 system of Remark 4, leaving $\boldsymbol{a}^* = \boldsymbol{v}_\star + \mathcal{O}(\bar\gamma_2)$. Writing ϵ_{SRTF} for the local head's own approximation error against $\boldsymbol{v}_\star$—a different class and a different target from the ϵ_W of Proposition 3, so the two are comparable only if the classes have comparable approximation power—the SRTF squared diagonal error is $\mathcal{O}(\epsilon_{\mathrm{SRTF}}^2 + \bar\gamma_2^2\|\mathrm{D}_t^2\boldsymbol{v}_\star\|^2)$: finite-capacity error plus a horizon-law bias that vanishes as τ^2 at velocity scale, rather than the τ_0-order representational floor of Proposition 3. The comparison is meaningful only in the regime that proposition delimits, namely a monolithic field whose training horizons are bounded away from zero, and within its regime window; it is not a claim about MeanFlow implementations that supervise the diagonal directly.*

Corollary 1 captures the structural division of labor in the SRTF class: the diagonal (τ=0) behavior is owned by $\boldsymbol{a}_{\boldsymbol{\theta}}$ and supervised by a standard flow-matching loss, so it is never asked to compromise with long-horizon targets the way a monolithic MeanFlow head is; the off-diagonal behavior is owned by $\tau\,\boldsymbol{b}_{\boldsymbol{\theta}}$ and supervised by the semigroup composition, which vanishes on the diagonal by construction. The diagonal error thereby carries only finite-capacity error and a second-order horizon-law bias, bypassing the τ_0-scaled representational residual of Proposition 3. The empirical counterpart is reported in Section 4.

C.4 Diagonal bias of a monolithic field trained away from the diagonal

Proposition 3 (Diagonal bias of a monolithic field trained away from the diagonal)**.** *Assume* $\boldsymbol{v}_\star \in C^3(\mathcal{M} \times [0,1])$. *Fix a training horizon distribution* $\tau \sim \pi$ *with* $\inf \operatorname{supp}(\pi) = \tau_0 > 0$. The hypothesis $\tau_0 > 0$ is essential and delimits the scope of this result: it describes a monolithic field that is never trained at zero horizon. MeanFlow-style implementations that sample $r{=}t$ with positive probability—including the MF-Riem baseline of Appendix D and SRTF itself (Algorithm 1)—have $\tau_0 = 0$ and are therefore *not* covered; for them the bound below is vacuous. We record the proposition because it isolates what a horizon-embedded readout must give up when the diagonal is not supervised directly, not as a claim about MeanFlow as published. *Let* $\mathcal{F}_W$ *be a parametric function class and let* ϵ_W *bound the trained minimizer's* L^2 *error against* $\boldsymbol{u}_\star$ at the edge horizon τ_0, $\mathbb{E}\|\hat{\boldsymbol{u}}_{\boldsymbol{\theta}}(\boldsymbol{z}, t-\tau_0, t) - \boldsymbol{u}_\star(\boldsymbol{z}, t-\tau_0, t)\|^2 \le \epsilon_W^2$. *This is an assumption about the trained model rather than about the class: a minimizer of the* π*-averaged loss need not be accurate at a single* τ*, least of all at the edge of the support, where a continuous* π *puts vanishing mass. Assume also (i) that* $\mathcal{F}_W$ *has bounded* L^2 *approximation error against* $\boldsymbol{u}_\star$ *on the training support, and (ii) uniformly bounded second* τ*-derivatives,* $\sup_{\hat{\boldsymbol{u}} \in \mathcal{F}_W, \tau \in [0,\tau_0]} \|\partial_\tau^2 \hat{\boldsymbol{u}}(\cdot, \cdot - \tau, \cdot)\|_{L^2} \le M_2$ *for some* $M_2 < \infty$*; and let* $\hat{\boldsymbol{u}}_{\boldsymbol{\theta}} \in \mathcal{F}_W$ *minimize the population MeanFlow loss. Define the class-representability factor*

$$\rho_W \;:=\; \sup_{\hat{\boldsymbol{u}} \in \mathcal{F}_W} \max\!\big\{0,\; -\operatorname{corr}\big(\partial_\tau \hat{\boldsymbol{u}}(\boldsymbol{z}, t-\tau, t)|_{\tau=\tau_0},\, \mathrm{D}_t \boldsymbol{v}_\star(\boldsymbol{z}, t)\big)\big\} \;\in\; [0,1],$$

where corr *is the (uncentered) cosine similarity in* L^2*, set to* 0 *when its first argument vanishes identically. Only* anti*-alignment can reduce the bound, which is why the correlation enters signed: a class whose* τ*-derivative is a* positive *multiple of* $\mathrm{D}_t \boldsymbol{v}_\star$ *has a larger diagonal error, not a smaller one. Then, provided* $\rho_W < 1$ *and the two-sided regime condition* $\tau_0^2\big(\sup \|\mathrm{D}_t^2 \boldsymbol{v}_\star\| + M_2\big) = \mathcal{O}(\epsilon_W)$ *and* $\epsilon_W = o\big(\tau_0 \sqrt{1-\rho_W^2}\, \sqrt{\mathbb{E}\|\mathrm{D}_t \boldsymbol{v}_\star\|^2}\big)$ *holds—the window in which the class error is small enough not to swamp the* τ_0*-order term, and large enough to dominate the* τ_0^2 *truncation—*

$$\mathbb{E}_{\boldsymbol{z},t}\big\|\hat{\boldsymbol{u}}_{\boldsymbol{\theta}}(\boldsymbol{z}, t, t) - \boldsymbol{v}_\star(\boldsymbol{z}, t)\big\|^2 \;\ge\; \tfrac{1}{4}\,\tau_0^2\,(1-\rho_W^2)\,\mathbb{E}\|\mathrm{D}_t \boldsymbol{v}_\star\|^2 \;-\; \mathcal{O}(\tau_0 \epsilon_W). \tag{13}$$

Outside that window the right-hand side need not be positive and the proposition says nothing: the $\Omega(\tau_0^2)$ *floor is a statement about the window, not an unconditional one. Without the first half of the condition the subtracted term picks up an extra* $\mathcal{O}(\tau_0^3)$.

Proof. Let $\hat{\boldsymbol{u}}_{\boldsymbol{\theta}} \in \mathcal{F}_W$ minimize $\mathcal{L}_{\mathrm{MF}}$ over the training support $\tau \in \operatorname{supp}(\pi)$ with $\inf \operatorname{supp}(\pi) = \tau_0 > 0$. By hypothesis the minimizer satisfies $\mathbb{E}\|\hat{\boldsymbol{u}}_{\boldsymbol{\theta}}(\boldsymbol{z}, t-\tau_0, t) - \boldsymbol{u}_\star(\boldsymbol{z}, t-\tau_0, t)\|^2 \le \epsilon_W^2$ at the edge horizon τ_0. Our strategy is to upper-bound the diagonal error of any $\hat{\boldsymbol{u}} \in \mathcal{F}_W$ in the *relaxed set*

$$\mathcal{F}_W^{\epsilon_W} \;:=\; \big\{\hat{\boldsymbol{u}} \in \mathcal{F}_W : \mathbb{E}\|\hat{\boldsymbol{u}}(\boldsymbol{z}, t-\tau_0, t) - \boldsymbol{u}_\star(\boldsymbol{z}, t-\tau_0, t)\|^2 \le \epsilon_W^2\big\},$$

which is a larger set than the MF minimizers (it only constrains behavior at $\tau = \tau_0$, not on all of $\operatorname{supp}(\pi)$). The MF minimizer $\hat{\boldsymbol{u}}_{\boldsymbol{\theta}}$ lies in $\mathcal{F}_W^{\epsilon_W}$, so

$$\mathbb{E}\|\hat{\boldsymbol{u}}_{\boldsymbol{\theta}}(\boldsymbol{z}, t, t) - \boldsymbol{v}_\star(\boldsymbol{z}, t)\|^2 \;\ge\; \min_{\hat{\boldsymbol{u}} \in \mathcal{F}_W^{\epsilon_W}} \mathbb{E}\|\hat{\boldsymbol{u}}(\boldsymbol{z}, t, t) - \boldsymbol{v}_\star(\boldsymbol{z}, t)\|^2,$$

and it suffices to lower-bound the right-hand side.

Taylor decomposition. For any $\hat{\boldsymbol{u}} \in \mathcal{F}_W^{\epsilon_W}$, write $g(\boldsymbol{z}, t) := \partial_\tau \hat{\boldsymbol{u}}(\boldsymbol{z}, t-\tau, t)\big|_{\tau=\tau_0}$. A first-order Taylor expansion of $\hat{\boldsymbol{u}}$ in τ around τ_0 gives

$$\hat{\boldsymbol{u}}(\boldsymbol{z}, t, t) \;=\; \hat{\boldsymbol{u}}(\boldsymbol{z}, t-\tau_0, t) \;-\; \tau_0\, g(\boldsymbol{z}, t) \;+\; R(\boldsymbol{z}, t),$$

with remainder R bounded in L^2 by $\frac{1}{2}\tau_0^2 \sup_{\tau \in [0,\tau_0]} \|\partial_\tau^2 \hat{\boldsymbol{u}}\|$. By the definition of $\mathcal{F}_W^{\epsilon_W}$, $\hat{\boldsymbol{u}}(\boldsymbol{z}, t-\tau_0, t) = \boldsymbol{u}_\star(\boldsymbol{z}, t-\tau_0, t) + \eta$ with $\mathbb{E}\|\eta\|^2 \le \epsilon_W^2$. Proposition 2 gives $\boldsymbol{u}_\star(\boldsymbol{z}, t-\tau_0, t) = \boldsymbol{v}_\star(\boldsymbol{z}, t) - \frac{\tau_0}{2}\mathrm{D}_t \boldsymbol{v}_\star(\boldsymbol{z}, t) + S$ with $\mathbb{E}\|S\|^2 \le C\tau_0^4$. Combining,

$$\hat{\boldsymbol{u}}(\boldsymbol{z}, t, t) - \boldsymbol{v}_\star(\boldsymbol{z}, t) \;=\; -\tau_0 \underbrace{\big(\tfrac{1}{2}\mathrm{D}_t \boldsymbol{v}_\star(\boldsymbol{z}, t) + g(\boldsymbol{z}, t)\big)}_{=:H} \;+\; \underbrace{\eta + S + R}_{=:E}. \tag{14}$$

Lower bound on the τ_0-order term. Expanding the squared L^2 norm of (14) and applying Cauchy–Schwarz to the cross term,

$$\mathbb{E}\|\hat{\boldsymbol{u}}(\boldsymbol{z},t,t)-\boldsymbol{v}_\star(\boldsymbol{z},t)\|^2 \;\geq\; \tau_0^2\,\mathbb{E}\|H\|^2 \;-\; 2\tau_0\,\sqrt{\mathbb{E}\|H\|^2}\,\sqrt{\mathbb{E}\|E\|^2}. \tag{15}$$

By construction, $\mathbb{E}\|E\|^2 = \mathcal{O}(\epsilon_W^2+\tau_0^4)$ (the τ_0^4 contribution comes from the bounded second τ-derivative assumption $M_2<\infty$ on $\partial_\tau^2\hat{\boldsymbol{u}}$), so $\sqrt{\mathbb{E}\|E\|^2} = \mathcal{O}(\epsilon_W+\tau_0^2)$. Rather than bound $\mathbb{E}\|H\|^2$ from above—the hypotheses constrain only the second τ-derivative of $\hat{\boldsymbol{u}}$, not the first, so $\|g\|$ and hence $\|H\|$ are unbounded over $\mathcal{F}_W$—we use the triangle inequality in the form $\sqrt{\mathbb{E}\|\hat{\boldsymbol{u}}(\boldsymbol{z},t,t)-\boldsymbol{v}_\star\|^2} \geq \tau_0\sqrt{\mathbb{E}\|H\|^2}-\sqrt{\mathbb{E}\|E\|^2}$, which is monotone increasing in $\sqrt{\mathbb{E}\|H\|^2}$ on the range where the right-hand side is non-negative. It therefore suffices to lower-bound $\mathbb{E}\|H\|^2$, and the conclusion follows whenever $\tau_0\sqrt{\mathbb{E}\|H\|^2}\geq 2\sqrt{\mathbb{E}\|E\|^2}$, which is the regime condition.

Best-alignment bound via ρ_W^2. For any $\hat{\boldsymbol{u}}\in\mathcal{F}_W^{\epsilon_W}$, write $g=\partial_\tau\hat{\boldsymbol{u}}(\boldsymbol{z},t-\tau,t)|_{\tau=\tau_0}$ so that $H=\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star+g$. Completing the square gives the elementary L^2 identity (no projection or linear-span argument needed)

$$\mathbb{E}\|H\|^2 \;=\; \mathbb{E}\|\tfrac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star+g\|^2 \;\geq\; \tfrac{1}{4}\,\mathbb{E}\|\mathrm{D}_t\boldsymbol{v}_\star\|^2\cdot\big(1-\max\{0,-c\}^2\big),$$

which holds for every $g\in L^2$ where $c:=\mathrm{corr}(g,\mathrm{D}_t\boldsymbol{v}_\star)$. Explicitly, writing $d:=\|\mathrm{D}_t\boldsymbol{v}_\star\|_{L^2}$ and $m:=\|g\|_{L^2}$,

$$\mathbb{E}\|\tfrac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star+g\|^2-\tfrac{1}{4}d^2(1-c^2) \;=\; \tfrac{1}{4}\big(cd+2m\big)^2 \;\geq\; 0,$$

a perfect square, so the bound is *attained* at $m=-cd/2$ and is tight for $c\leq 0$; for $c\geq 0$ the infimum is $\frac{1}{4}d^2$, attained at $g\equiv\mathbf{0}$, which is why only the negative part of c appears in ρ_W. Taking the infimum over all realizable $\hat{\boldsymbol{u}}\in\mathcal{F}_W^{\epsilon_W}$ and using the definition of ρ_W,

$$\inf_{\hat{\boldsymbol{u}}\in\mathcal{F}_W^{\epsilon_W}}\mathbb{E}\|H\|^2 \;\geq\; \tfrac{1}{4}\,(1-\rho_W^2)\,\mathbb{E}\|\mathrm{D}_t\boldsymbol{v}_\star\|^2.$$

Substituting into (15) and absorbing the cross term under the regime assumption,

$$\min_{\hat{\boldsymbol{u}}\in\mathcal{F}_W^{\epsilon_W}}\mathbb{E}\|\hat{\boldsymbol{u}}(\boldsymbol{z},t,t)-\boldsymbol{v}_\star(\boldsymbol{z},t)\|^2 \;\geq\; \tfrac{\tau_0^2}{4}(1-\rho_W^2)\,\mathbb{E}\|\mathrm{D}_t\boldsymbol{v}_\star\|^2 \;-\; \mathcal{O}(\tau_0\epsilon_W).$$

Since $\hat{\boldsymbol{u}}_{\boldsymbol{\theta}}\in\mathcal{F}_W^{\epsilon_W}$, the MF minimizer inherits this lower bound, which is (13). □

Remark 5 (Role of the hypothesis $\rho_W^2<1$). The bound is vacuous when $\rho_W=1$, i.e. when $\mathcal{F}_W$ contains an element whose τ-derivative at τ_0 is a *negative* multiple of $\mathrm{D}_t\boldsymbol{v}_\star$ up to L^2 equivalence—the class must be able to cancel the diagonal residual, not merely to correlate with it. Whether a given parametric class satisfies $\rho_W^2<1$ is a representational question that depends on the architecture, the τ-embedding, and the data distribution, and we do not prove a universal strict inequality here. Our working hypothesis is motivated by the observation that $\mathrm{D}_t\boldsymbol{v}_\star=\partial_t\boldsymbol{v}_\star+(\boldsymbol{v}_\star\cdot\nabla)\boldsymbol{v}_\star$ couples spatial gradients of $\boldsymbol{v}_\star$ with its own value, whereas ∂_τ of a horizon-embedded readout at fixed $(\boldsymbol{z},t)$ varies only the τ-coordinate of the readout; an embedding that does not expose the spatial gradient structure is unlikely to realize a scalar multiple of $\mathrm{D}_t\boldsymbol{v}_\star$ through its τ-derivative alone. SRTF circumvents this question entirely by making $\boldsymbol{b}_\star$ an explicit output whose supervision (the semigroup target of Eq. (7)) is decoupled from any τ-embedding of the backbone.

C.5 Theorem 1 (Sampler error bound)

Theorem 1, restated with full hypotheses. *Assume $\boldsymbol{v}_\star\in C^3(\mathcal{M}\times[0,1])$ and that the sampler iterates, the exact reverse flow of the prior draw and the exact flow segments through the iterates stay in a compact region of $\mathcal{M}$ on which $\boldsymbol{u}_\star$ is L_z-Lipschitz in $\boldsymbol{z}$ and $\|\mathrm{D}_t^2\boldsymbol{v}_\star\|$ is bounded (for a Gaussian prior this is a statement on that event; it holds as stated for a compactly truncated p_1). Let a trained SRTF field satisfy a uniform L^∞-in-$\boldsymbol{z}$ horizon bound $\sup_{\boldsymbol{z}\in\mathcal{M},0\leq r\leq t\leq 1}\|\boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z},r,t)-\bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z},r,t)\|\leq\epsilon_{\mathrm{hor}}$ (this uniform bound absorbs the covariate shift between training samples and sampler iterates). A K-step reverse-time SRTF sampler with uniform step $\tau=1/K$, started from a prior sample $\boldsymbol{z}_K$ at $t_K=1$ and producing the data-side iterate $\hat{\boldsymbol{z}}_0$ at $t_0=0$, satisfies*

$$\mathbb{E}\|\hat{\boldsymbol{z}}_0-\tilde{\boldsymbol{z}}_0\| \;\leq\; e^{L_z}\big(\epsilon_{\mathrm{hor}} \;+\; C_{\boldsymbol{v}_\star}\,\tau^2\big), \tag{16}$$

where $\tilde{\boldsymbol{z}}_0$ is the image of the same *prior draw under the exact reverse flow of $\boldsymbol{v}_\star$ and $C_{\boldsymbol{v}_\star}=\frac{1}{6}\sup\|\mathrm{D}_t^2\boldsymbol{v}_\star\|$. The bound controls deviation from that flow only; it is not a distributional guarantee, and it says nothing about the gap between $\boldsymbol{v}_\star$ and the true score of the data.*

Proof. Let $\{\boldsymbol{z}_k\}_{k=0}^K$ denote the SRTF sampler iterates indexed in reverse time so that $\boldsymbol{z}_K \sim p_1$ at $t_K = 1$ and $\boldsymbol{z}_0$ is the data-side endpoint at $t_0 = 0$. The reverse-time update reads $\boldsymbol{z}_{k-1} = \mathrm{Exp}_{\boldsymbol{z}_k}\big(-\tau\, \boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z}_k, t_{k-1}, t_k)\big)$ (note the spatial argument $\boldsymbol{z}_k$ at the larger time t_k, matching the parameterization of $\boldsymbol{w}_{\boldsymbol{\theta}}$ in Eq. (4)); let $\{\tilde{\boldsymbol{z}}_k\}$ denote the exact-flow counterparts $\tilde{\boldsymbol{z}}_{k-1} = \mathrm{Exp}_{\tilde{\boldsymbol{z}}_k}\big(-\tau\, \boldsymbol{u}_\star(\tilde{\boldsymbol{z}}_k, t_{k-1}, t_k)\big)$. Define the expected per-step error $e_k := \mathbb{E}\|\boldsymbol{z}_k - \tilde{\boldsymbol{z}}_k\|$ (the expectation is over the prior draw $\boldsymbol{z}_K = \tilde{\boldsymbol{z}}_K$). On the flat manifold $\mathcal{M}$, Exp is an isometry (up to the mod-1 reduction on $\mathcal{F}$, which is a contraction), so by the triangle inequality applied under expectation,

$$e_{k-1} \;\le\; e_k \;+\; \tau\, \mathbb{E}\big\|\boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z}_k, t_{k-1}, t_k) - \boldsymbol{u}_\star(\tilde{\boldsymbol{z}}_k, t_{k-1}, t_k)\big\|.$$

The per-step residual decomposes into three contributions: (i) the horizon residual at the sampler iterate, $\mathbb{E}\|\boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z}_k, \cdot, \cdot) - \bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z}_k, \cdot, \cdot)\| \le \epsilon_{\mathrm{hor}}$ by the uniform L^∞-in-$\boldsymbol{z}$ assumption; (ii) the first-order consistency remainder $\mathbb{E}\|\bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z}_k, \cdot, \cdot) - \boldsymbol{u}_\star(\boldsymbol{z}_k, \cdot, \cdot)\| \le C_2\tau^2$ with $C_2 = \frac{1}{6}\sup\|\mathrm{D}_t^2 \boldsymbol{v}_\star\|$, by the integral-form Taylor remainder along the backward flow segment from the iterate, $\frac{1}{\tau}\int_0^\tau \frac{u^2}{2}\,\mathrm{d}u = \frac{\tau^2}{6}$ (cf. Proposition 2); (iii) the Lipschitz propagation $\mathbb{E}\|\boldsymbol{u}_\star(\boldsymbol{z}_k, \cdot) - \boldsymbol{u}_\star(\tilde{\boldsymbol{z}}_k, \cdot)\| \le L_z\, e_k$. Summing,

$$e_{k-1} \;\le\; (1 + L_z\tau)\, e_k \;+\; \tau\big(\epsilon_{\mathrm{hor}} + C_2\tau^2\big).$$

Applying the discrete Grönwall inequality with $e_K = 0$ (initial condition: the sampler and the exact flow share the same prior draw at $t_K = 1$) and iterating backward to $k = 0$,

$$e_0 \;\le\; \left(\sum_{k=1}^{K}\prod_{j=1}^{k-1}(1 + L_z\tau)\right)\cdot\tau(\epsilon_{\mathrm{hor}} + C_2\tau^2) \;\le\; e^{L_z K\tau}\cdot K\tau(\epsilon_{\mathrm{hor}} + C_2\tau^2) \;=\; e^{L_z}(\epsilon_{\mathrm{hor}} + C_2\tau^2),$$

using $K\tau = 1$. The consistency constant $C_2 = \frac{1}{6}\sup\|\mathrm{D}_t^2\boldsymbol{v}_\star\|$ is finite under the C^3 assumption on $\boldsymbol{v}_\star$ (which ensures $\mathrm{D}_t^2\boldsymbol{v}_\star$ is continuous and bounded on the compact domain); concretely, $\|\mathrm{D}_t^2\boldsymbol{v}_\star\|$ can be bounded in terms of the C^2-norms of $\boldsymbol{v}_\star$ in both time and space, yielding the constant $C_{\boldsymbol{v}_\star}$ of Eq. (16). This establishes the sampler-error bound. □

Per-step resolution and the loss schedule. The main text defers to this appendix the step-resolved form of the bound that motivates the schedule (9). Retaining the per-step residual instead of the uniform bound ϵ_{hor}, the same Grönwall recursion gives

$$\mathbb{E}\|\hat{\boldsymbol{z}}_0 - \tilde{\boldsymbol{z}}_0\| \;\lesssim\; e^{L_z}\sum_{k=1}^{K}\tau\Big(\epsilon_{\mathrm{hor}}(t_k, \tau) + C_{\boldsymbol{v}_\star}\tau^2\Big),$$

where $\epsilon_{\mathrm{hor}}(t, \tau) := \sup_{\boldsymbol{z}}\|\boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z}, t-\tau, t) - \bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z}, t-\tau, t)\|$, so each step contributes in proportion to the residual *at its own horizon*. Splitting the residual as $\epsilon_{\mathrm{hor}}(t, \tau) \le \epsilon_{\boldsymbol{a}}(t) + \gamma(\tau)\,\epsilon_{\boldsymbol{b}}(t)$, with $\epsilon_{\boldsymbol{a}}(t) := \sup_{\boldsymbol{z}}\|\boldsymbol{a}_{\boldsymbol{\theta}}(\boldsymbol{z}, t) - \boldsymbol{a}_\star(\boldsymbol{z}, t)\|$ and $\epsilon_{\boldsymbol{b}}$ defined likewise, shows the local residual entering every step with weight 1 and the horizon residual entering in proportion to τ. The t-dependence comes from the *training* horizon law rather than from the sampler, whose steps are uniform: since $r \ge 0$ we always have $\tau \le t$, and the law $r \sim \mathrm{Unif}(0, t)$ gives $\mathbb{E}[\tau \mid t] = t/2$ conditional on the horizon branch being active ($t/4$ once the $r{=}t$ resets of half the batch are included, the figure quoted in Section 3.3), so the horizons a model is asked to represent grow linearly with t and the horizon residual is weighted by a factor that grows with t. That motivates λ_{hor} increasing in t, and Eq. (9) takes the simplest such choice. The companion decay of λ_{loc} does not follow from the bound—only the *relative* weight of the two losses is being shifted—and we present it as the heuristic completion of that shift rather than as a consequence. We emphasize throughout that this is a motivation, not a derivation of an optimal schedule.

Remark 6 (Many-step reduction). If additionally $\sup_{\boldsymbol{z},t}\|\boldsymbol{a}_{\boldsymbol{\theta}}(\boldsymbol{z}, t) - \boldsymbol{v}_\star(\boldsymbol{z}, t)\| \le \epsilon_{\mathrm{loc}}$ and $\boldsymbol{b}_{\boldsymbol{\theta}}$ is bounded, then because $\bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z}, r, t) = \boldsymbol{v}_\star(\boldsymbol{z}, t) + \tau\boldsymbol{b}_\star(\boldsymbol{z}, t)$ and $\boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z}, r, t) = \boldsymbol{a}_{\boldsymbol{\theta}}(\boldsymbol{z}, t) + \gamma(\tau)\boldsymbol{b}_{\boldsymbol{\theta}}(\boldsymbol{z}, t)$ with $\gamma(\tau) = \mathcal{O}(\tau)$,

$$\sup_{\boldsymbol{z}}\|\boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z}, r, t) - \bar{\boldsymbol{u}}_{\mathrm{SRTF}}(\boldsymbol{z}, r, t)\| \;\le\; \sup_{\boldsymbol{z}}\|\boldsymbol{a}_{\boldsymbol{\theta}} - \boldsymbol{v}_\star\| + \mathcal{O}(\tau) \;\le\; \epsilon_{\mathrm{loc}} + \mathcal{O}(\tau),$$

so $\epsilon_{\mathrm{hor}}(t, \tau) \le \epsilon_{\mathrm{loc}} + \mathcal{O}(\tau)$ and the per-step bound above becomes $e^{L_z}(\epsilon_{\mathrm{loc}} + \mathcal{O}(\tau) + C\tau^2) = e^{L_z}(\epsilon_{\mathrm{loc}} + \mathcal{O}(\tau))$. In words: at large K (small τ), the SRTF sampler degenerates gracefully to the standard flow-matching sampler driven by $\boldsymbol{a}_{\boldsymbol{\theta}}$ alone, recovering the FM-precision regime.

C.6 GAUGE STRUCTURE OF THE TWO-LOSS OBJECTIVE

Remark 7 (Gauge-breaking by a non-degenerate τ-law). Restricted to a single horizon τ_0, the parameterization $\boldsymbol{w}_{\boldsymbol{\theta}} = \boldsymbol{a}_{\boldsymbol{\theta}} + \gamma(\tau)\boldsymbol{b}_{\boldsymbol{\theta}}$ admits the gauge family $(\boldsymbol{a}_{\boldsymbol{\theta}}, \boldsymbol{b}_{\boldsymbol{\theta}}) \mapsto (\boldsymbol{a}_{\boldsymbol{\theta}} + \gamma(\tau_0)\boldsymbol{c}, \boldsymbol{b}_{\boldsymbol{\theta}} - \boldsymbol{c})$ for any field $\boldsymbol{c}(\boldsymbol{z}, t)$: any horizon regression that evaluates $\boldsymbol{w}_{\boldsymbol{\theta}}$ at τ_0 only cannot distinguish the gauge orbit. Evaluating the horizon supervision over a non-degenerate τ-law with $\bar{\gamma}_2 > \bar{\gamma}_1^2$ breaks this gauge; the local loss $\mathcal{L}_{\mathrm{loc}}$ improves identifiability and training stability but is not strictly necessary for *gauge-breaking*—which is a weaker property than selecting $\boldsymbol{u}_\star$, for which the boundary condition supplied by $\mathcal{L}_{\mathrm{loc}}$ is essential (Proposition 1 (ii)). The semigroup loss of Eq. (8) is stronger still: each individual term couples the three horizons τ, $\xi\tau$, $(1-\xi)\tau$, so the gauge is broken in the fixed-point conditions of every semigroup term, at self-consistency $\boldsymbol{\theta}^- = \boldsymbol{\theta}$ (Remark 3). At fixed target parameters a single term is a plain regression at one horizon and is gauge-invariant like any other; it is the self-consistent fixed-point set that the coupling constrains.

Direct substitution.

$$\boldsymbol{a}_{\boldsymbol{\theta}}(\boldsymbol{z}, t) + \gamma(\tau_0)\boldsymbol{c}(\boldsymbol{z}, t) + \gamma(\tau_0)\big(\boldsymbol{b}_{\boldsymbol{\theta}}(\boldsymbol{z}, t) - \boldsymbol{c}(\boldsymbol{z}, t)\big) \;=\; \boldsymbol{a}_{\boldsymbol{\theta}}(\boldsymbol{z}, t) + \gamma(\tau_0)\boldsymbol{b}_{\boldsymbol{\theta}}(\boldsymbol{z}, t),$$

so $\boldsymbol{w}_{\boldsymbol{\theta}}$ at the single horizon τ_0 is unchanged: $\mathcal{L}_{\mathrm{hor}}$ restricted to τ_0 is gauge-invariant in the direction $\boldsymbol{c}$. At a second horizon $\tau_1 \neq \tau_0$ the same transformation changes $\boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z}, t-\tau_1, t)$ by $(\gamma(\tau_0) - \gamma(\tau_1))\boldsymbol{c}$, which is nonzero unless $\boldsymbol{c} \equiv 0$, provided the pair is chosen with $\gamma(\tau_0) \neq \gamma(\tau_1)$—non-constancy of γ alone does not supply such a pair, whereas $\mathrm{Var}_\tau[\gamma(\tau)] > 0$ does. Equivalently, evaluating $\mathcal{L}_{\mathrm{hor}}$ over a non-degenerate τ-law with $\bar{\gamma}_2 > \bar{\gamma}_1^2$ produces the positive-determinant 2×2 system of Remark 4, which pins down $(\boldsymbol{a}_{\boldsymbol{\theta}}, \boldsymbol{b}_{\boldsymbol{\theta}})$ uniquely. The local loss $\mathcal{L}_{\mathrm{loc}}(\boldsymbol{a}) = \mathbb{E}\|\boldsymbol{a} - \boldsymbol{v}_\star\|^2$ provides an additional pointwise anchor of $\boldsymbol{a}_{\boldsymbol{\theta}}$ to $\boldsymbol{v}_\star$ at every $(\boldsymbol{z}, t)$ but is not strictly necessary for gauge-breaking once $\mathcal{L}_{\mathrm{hor}}$ is evaluated over a non-degenerate τ-law.

C.7 RIEMANNIAN TRANSPORT EXPANSION ON A GENERAL MANIFOLD

Section 3 states the transport expansion directly on flat $\mathcal{M}$. This subsection records where it comes from intrinsically, and—more usefully—what of the construction does *not* survive curvature.

Let $\mathcal{M}$ carry the Levi-Civita connection, with parallel transport $\mathrm{P}_{\boldsymbol{z}_s \to \boldsymbol{z}_t}$ along a curve and covariant derivative $\frac{D}{\mathrm{d}t}$ along it; on flat $\mathcal{M}$ the latter reduces to the material derivative $\partial_t + \boldsymbol{v}_\star \cdot \nabla$ with no Christoffel corrections, and we use D_t for both. Define the finite-horizon velocity by transporting velocities back to $T_{\boldsymbol{z}_t}\mathcal{M}$ before averaging, $\boldsymbol{u}_\star(\boldsymbol{z}_t, r, t) = \frac{1}{\tau}\int_r^t \mathrm{P}_{\boldsymbol{z}_s \to \boldsymbol{z}_t}\boldsymbol{v}_\star(\boldsymbol{z}_s, s)\,\mathrm{d}s$. Taylor-expanding the integrand about $s = t$, using $\mathrm{P}_{\boldsymbol{z}_t \to \boldsymbol{z}_t} = \mathrm{Id}$ and $\frac{\mathrm{d}}{\mathrm{d}s}[\mathrm{P}_{\boldsymbol{z}_s \to \boldsymbol{z}_t} X(\boldsymbol{z}_s, s)]|_{s=t} = \frac{D}{\mathrm{d}t}X$, and integrating with $\int_r^t (t-s)^k \mathrm{d}s/\tau = \tau^k/(k+1)$, gives the intrinsic series

$$\boldsymbol{u}_\star(\boldsymbol{z}_t, r, t) = \boldsymbol{v}_\star(\boldsymbol{z}_t, t) \;-\; \tfrac{\tau}{2}\tfrac{D}{\mathrm{d}t}\boldsymbol{v}_\star(\boldsymbol{z}_t, t) \;+\; \tfrac{\tau^2}{6}\tfrac{D^2}{\mathrm{d}t^2}\boldsymbol{v}_\star(\boldsymbol{z}_t, t) \;+\; \mathcal{O}(\tau^3). \tag{17}$$

Setting $\boldsymbol{a}_\star = \boldsymbol{v}_\star$, $\boldsymbol{b}_\star = -\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star$ and $\gamma(\tau) = \tau$ recovers the structured expansion of Eq. (3): the zeroth-order term is the local velocity, the first-order term the covariant trajectory curvature.

What curvature costs. The expansion generalizes; the rest of the construction does not. Two things the objective relies on hold only on flat $\mathcal{M}$. First, the chord identity of Lemma 1: on a curved manifold the transported average and the log-map displacement differ by $\mathcal{O}(K_{\max}\tau^2\|\boldsymbol{v}_\star\|^3)$, the Jacobi-field deviation of $\mathrm{d}(\mathrm{Log})$ from parallel transport over a chord of length $\lesssim \|\boldsymbol{v}_\star\|\tau$, by a standard Jacobi-field comparison estimate. Second, exact composition of jumps (Remark 1), which rests on the exponential map being translation in a global chart. The semigroup identity (6) would therefore hold only up to a curvature correction per composition, and that correction—not the expansion—is what an extension to curved geometries would have to absorb, for instance by composing jumps through the logarithm map as in the semigroup form of Riemannian MeanFlow (Woo et al., 2026). It is why the method is stated on flat $\mathcal{M}$.

Specialization. On $\mathcal{M}$ each factor is a quotient of a Euclidean space by translations, so holonomy is trivial, parallel transport is the identity under the canonical identification, $\frac{D}{\mathrm{d}t}$ carries no Christoffel terms, and Exp is componentwise translation—the identity on $\mathcal{A}$ and $\mathcal{L}^\varphi$, and mod-1 reduction on $\mathcal{F}$.

C.8 Factor-wise decoupling on the crystal product manifold

Lemma 3 (Product-manifold decoupling). *On $\mathcal{M} = \mathcal{A} \times \mathcal{F} \times \mathcal{L}^{\varphi}$ with block-diagonal product metric, the following decouple factor-wise:*

(i) the Riemannian exponential and logarithm maps: $\mathrm{Exp}_{(\boldsymbol{z}^A, \boldsymbol{z}^F, \boldsymbol{z}^L)}(\boldsymbol{v}^A, \boldsymbol{v}^F, \boldsymbol{v}^L) = (\mathrm{Exp}^{\mathcal{A}}_{\boldsymbol{z}^A} \boldsymbol{v}^A, \mathrm{Exp}^{\mathcal{F}}_{\boldsymbol{z}^F} \boldsymbol{v}^F, \mathrm{Exp}^{\mathcal{L}}_{\boldsymbol{z}^L} \boldsymbol{v}^L)$*, and similarly for* Log*;*

(ii) the semigroup target: $\bar{\boldsymbol{v}}_{\mathrm{sg}} = (\bar{\boldsymbol{v}}^A_{\mathrm{sg}}, \bar{\boldsymbol{v}}^F_{\mathrm{sg}}, \bar{\boldsymbol{v}}^L_{\mathrm{sg}})$, since both half-jump evaluations and the intermediate Exp *in Eq.* (7) *act blockwise, so the identities of Lemma 1 and Proposition 1 (i),(iii) hold componentwise (with the joint D_t);*

(iii) the SRTF expansion: $\boldsymbol{w}^c_{\boldsymbol{\theta}} = \boldsymbol{a}^c_{\boldsymbol{\theta}} + \tau\, \boldsymbol{b}^c_{\boldsymbol{\theta}}$ for each $c \in \{A, F, L\}$;

(iv) the two SRTF losses: $\mathcal{L}_{\bullet} = \mathcal{L}^A_{\bullet} + \mathcal{L}^F_{\bullet} + \mathcal{L}^L_{\bullet}$ for $\bullet \in \{\mathrm{loc}, \mathrm{hor}\}$.

Consequently Lemma 1, Proposition 2, Lemma 2 and Proposition 1 (i),(iii) hold blockwise, being componentwise vector identities. Note what this does not *say: the marginal velocity $\boldsymbol{v}^c_{\star}$ depends on the joint state, so D_t is a joint operator and $\boldsymbol{b}^c_{\star}$ is not a per-factor object—which is precisely why the two readouts sit on one shared backbone. Two results are therefore excluded. Theorem 1 does not decouple: its recursion needs $\boldsymbol{u}^c_{\star}$ to be Lipschitz in $\boldsymbol{z}^c$ alone, which is false, and per-factor errors propagate into every other factor, so the Grönwall constant is irreducibly joint. Proposition 3 is excluded for a separate reason: its representability factor is a supremum of a* normalized *correlation, and a supremum of ratios does not decompose across a product because the denominators involve the total norms. The only factor-specific subtlety is the translation-invariance on $\mathcal{F}$, handled by the mean-subtracted target convention of Section 3.*

Proof. With a block-diagonal metric, geodesics on $\mathcal{M}$ are factor-wise concatenations of geodesics on the three factors, which gives (i); applying (i) to the horizon target and to the coefficient decomposition gives (ii)–(iii); and (iv) follows from the additivity of the squared product norm. □

We record the lemma explicitly because several downstream arguments (e.g., the translation-invariance asymmetry on $\mathcal{F}$) rely on factor-wise independence rather than a global manifold argument.

C.9 Flatness under the logit lattice reparameterization

The six raw lattice parameters $(a, b, c, \alpha, \beta, \gamma_{\mathrm{ang}})$—the third angle written with a subscript to free γ for the scale profile—live in the unbounded open set $(0, \infty)^3 \times (60^\circ, 120^\circ)^3$, the angle range following Miller et al. (2024). The chart φ maps it diffeomorphically onto $\mathbb{R}^6$ componentwise: inverse-softplus on the three lengths, and a logit on each angle after rescaling its range to $(0, 1)$. Its inverse φ^{-1} is softplus on the lengths and a rescaled sigmoid on the angles.

We then *define* the metric on $\mathcal{L}$ as the push-forward of the Euclidean metric on $\mathbb{R}^6$ under φ^{-1}. By construction $(\mathcal{L}, g)$ is isometric to $\mathbb{R}^6$, hence flat, geodesically complete and free of cut locus, so Exp and Log are addition and subtraction in the chart and Lemma 1 applies without modification. It is worth being explicit that this is a modeling choice and not a property of lattice-parameter space: a diffeomorphism does not by itself flatten anything, and a different metric—one induced from the space of lattice matrices, say—would not be flat in this chart. Two consequences follow. The per-factor loss $\|\cdot\|_{\mathcal{L}}$ is a chart-coordinate norm, so the factor weight $\omega_{\mathcal{L}}$ selects a metric as well as balancing scales; and the only factor of $\mathcal{M}$ with genuinely non-Euclidean geometry is the torus $\mathcal{F}$.

D Implementation details

D.1 Factor-wise instantiation on the crystal manifold

This subsection gives the per-factor detail deferred from Section 3.2; Lemma 3 is what licenses treating the three factors separately, and Table 2 collects the symbols.

Table 4: **Factor-wise SRTF instantiation on the flat crystal manifold $\mathcal{M} = \mathcal{A} \times \mathcal{F} \times \mathcal{L}^{\varphi}$.** The local target $\bar{\boldsymbol{v}}_c^{\mathrm{cond}}$ is the bridge velocity, mean-subtracted on $\mathcal{F}$ and used as-is on $\mathcal{A}, \mathcal{L}$; the horizon target is the componentwise semigroup composition of two gradient-free half-jumps (Section 3.3). Both targets live in the tangent space $T_{\boldsymbol{z}_t}\mathcal{M}$; the torus logarithm $\mathrm{Log}^{\mathcal{F}}$ is given in Eq. (18).

Factor	Manifold	Cond. velocity $\boldsymbol{v}_c^{\mathrm{cond}}$	Regression target $\bar{\boldsymbol{v}}_c^{\mathrm{cond}}$	$\mathrm{Exp}_{\boldsymbol{z}}^{c}(\boldsymbol{w})$
Atom types $\mathcal{A}$	$\mathbb{R}^{7n}$ (flat)	$\boldsymbol{z}_1^A - \boldsymbol{z}_0^A$	$\boldsymbol{v}_A^{\mathrm{cond}}$	$\boldsymbol{z}^A + \boldsymbol{w}$
Frac. coords $\mathcal{F}$	$(\mathbb{R}/\mathbb{Z})^{3n}$ (flat torus)	$\mathrm{Log}_{\boldsymbol{f}_0}^{\mathcal{F}}(\boldsymbol{f}_1)$	$\bar{\boldsymbol{v}}_F^{\mathrm{cond}}$	$(\boldsymbol{z}^F + \boldsymbol{w}) \bmod 1$
Lattice $\mathcal{L}$	$\mathbb{R}^6$ via φ-chart (flat)	$\boldsymbol{l}_1^{\varphi} - \boldsymbol{l}_0^{\varphi}$	$\boldsymbol{v}_L^{\mathrm{cond}}$	$\boldsymbol{z}^L + \boldsymbol{w}$

Atom types and lattice parameters. Both $\mathcal{A} = \mathbb{R}^{7n}$ and $\mathcal{L}^{\varphi} = \mathbb{R}^6$ are flat Euclidean, so the interpolant is linear ($\boldsymbol{z}_t^A = (1-t)\boldsymbol{z}_0^A + t\boldsymbol{z}_1^A$; $\boldsymbol{l}_t^{\varphi} = (1-t)\boldsymbol{l}_0^{\varphi} + t\boldsymbol{l}_1^{\varphi}$) and the conditional velocity is constant along the path. The chart φ on $\mathcal{L}$ maps the six raw parameters into a flat Euclidean $\mathbb{R}^6$ by a smooth diffeomorphism built componentwise from inverse-softplus maps on the three positive lengths and logit maps on the three angles (range $[60^\circ, 120^\circ]$, following Miller et al. (2024); full chart in Appendix C.9); the inverse chart is applied only at the end of sampling.

Atomic coordinates. On the flat torus of fractional positions, tangent vectors are canonically in $\mathbb{R}^{3n}$. The geodesic interpolant uses the torus exponential and logarithm maps,

$$\boldsymbol{f}_t \;=\; \mathrm{Exp}_{\boldsymbol{f}_0}^{\mathcal{F}}\big(t \cdot \mathrm{Log}_{\boldsymbol{f}_0}^{\mathcal{F}}(\boldsymbol{f}_1)\big), \quad \mathrm{Log}_{\boldsymbol{f}_0}^{\mathcal{F}}(\boldsymbol{f}_1) \;=\; (\boldsymbol{f}_1 - \boldsymbol{f}_0 + \tfrac{1}{2}) \bmod 1 - \tfrac{1}{2}, \tag{18}$$

so the raw conditional velocity $\boldsymbol{v}_F^{\mathrm{cond}} = \mathrm{Log}_{\boldsymbol{f}_0}^{\mathcal{F}}(\boldsymbol{f}_1)$ is constant along the path. The mean-subtracted target used wherever a regression target appears has per-atom components $\bar{\boldsymbol{v}}_{F,i}^{\mathrm{cond}} = \boldsymbol{v}_{F,i}^{\mathrm{cond}} - \frac{1}{n}\sum_j \boldsymbol{v}_{F,j}^{\mathrm{cond}}$, where $\boldsymbol{v}_{F,i}^{\mathrm{cond}} \in \mathbb{R}^3$ denotes the velocity on the i-th atom.

Backbone and readouts. All arms share the CSPNet backbone of Miller et al. (2024) at its default width and depth. Time is encoded with sinusoidal features; the horizon never enters the backbone, modulating only the analytic factor $\gamma(\tau)$ in the readout, which is what lets the first target evaluation reuse the backbone features of the differentiated forward. The two readout branches are lightweight per-factor projections on the shared hidden state; the second adds under half a percent to the parameter count. On $\mathcal{F}$ we mean-subtract after the readout; on $\mathcal{A}$ and $\mathcal{L}$ the raw output is used.

Baselines. The three published baselines are evaluated at their authors' own full-NFE budgets. **FlowMM (in-pipeline)**, the "matched flow-matching baseline" of the main text, is the matched flow-matching arm used for every controlled comparison in the paper: the same backbone, data, optimizer, schedule and training budget as SRTF, differing only in the objective, and swept over the same NFE grid with its sampler configuration selected on the same grid. It isolates the effect of the objective from everything else in the pipeline; since the FlowMM paper does not report best-of-20 match rate, every FlowMM k=20 value in Table 1 is measured this way. **MF-Riem** ports the image-space MeanFlow objective of Geng et al. (2025a) to the flat crystal manifold, the flat-manifold counterpart of the derivative-based Riemannian MeanFlow objectives (Woo et al., 2026; Zhong et al., 2026): a single head trained via the self-consistent MeanFlow identity with a Jacobian–vector product through the model, and MeanFlow's adaptive weighting. It differs from SRTF in two coupled respects, the structured decomposition and the JVP-free target, which the control grid of Appendix F.3 separates. All of these arms share the same backbone, differing only in the readout and the loss.

Horizon target construction. For the horizon loss we draw a second time $r \sim \mathrm{Unif}(0, t)$, take the midpoint split $s = (t + r)/2$, and assemble the semigroup target of Eq. (7) from two forward evaluations of a gradient-free target copy, with the torus wrap on $\mathcal{F}$ and the φ-chart on $\mathcal{L}$ inside the intermediate Exp. No JVP or forward-mode tangent is required anywhere: the target is an ordinary vector, the target evaluations store no activations, and the training graph is a standard first-order graph through one backbone forward and the two readouts. A short warm-up with the horizon loss switched off brings the local head close to $\boldsymbol{v}_\star$ by pure flow matching before the bootstrap is activated.

Training and evaluation protocol. All arms use Adam with a cosine schedule, EMA weights for evaluation, and the FlowMM convention of per-factor loss multipliers balancing the scales of the

three product factors. The per-step time is drawn from a logit-normal law concentrating mass away from the bridge endpoints, and the horizon is drawn uniformly below it for half the batch and set equal to it for the other half, so that half of every batch supplies a pure flow-matching gradient for the local head. Two epoch budgets appear in this paper: the main CSP results use per-dataset budgets matched to FlowMM, while the MF-Riem head-to-head of Appendix E.4 uses a shorter flat budget on all four datasets, because the JVP baseline's cost made the longer budgets impractical. Both arms share it within that table, so the comparison is matched even though it is not the main-table budget, and every *controlled* comparison in the paper holds the budget fixed across the cells being compared. Optimizer settings, schedule constants, factor weights, epoch counts and the sampler configuration grids are listed with the code release.

Sampling. We use the componentwise exponential map on $\mathcal{M}$, which is the identity on $\mathcal{A}$ and $\mathcal{L}$ and mod-1 reduction on $\mathcal{F}$. For the hybrid sampler we integrate k_{SRTF} transport-expansion steps with $\boldsymbol{w}_{\boldsymbol{\theta}}$ followed by k_{FM} Euler steps with $\boldsymbol{a}_{\boldsymbol{\theta}}$, at uniform step sizes; the anti-annealing knob of Section 4.1 scales the coordinate-block velocity by $1 + \alpha(1-t)$ with strength α (and analogously on the lattice block).

D.2 Algorithms and hybrid sampling

This appendix writes out the SRTF training loop and the sampler in full.

Training step. Algorithm 1 summarizes one mini-batch of SRTF training. The key practical feature is lines 5–7: the horizon target is assembled from *gradient-free* forward evaluations of the target copy $\boldsymbol{\theta}^{-}$ (no activation storage, no forward-mode tangent), so the overall training graph is a standard first-order graph through one differentiated backbone forward and the two readouts. When $r = t$ the horizon path is skipped entirely, so no degenerate $0/0$ horizon computation ever occurs.

1: Sample $(\boldsymbol{z}_0, \boldsymbol{z}_1) \sim p_0 \otimes p_1$, $t \sim \mathrm{LogitNormal}$; with prob. $\frac{1}{2}$ set $r \leftarrow t$ (so half the batch supplies a pure flow-matching gradient for $\boldsymbol{a}_{\boldsymbol{\theta}}$, the boundary anchor of Proposition 1 (ii)), else $r \sim \mathrm{Unif}(0, t)$; let $\tau \leftarrow t - r$.
2: Form the bridge sample $\boldsymbol{z}_t = \phi_t(\boldsymbol{z}_0, \boldsymbol{z}_1)$ factor-wise.
3: Local target: $\bar{\boldsymbol{v}}^{\mathrm{cond}} \leftarrow (\boldsymbol{v}_A^{\mathrm{cond}}, \bar{\boldsymbol{v}}_F^{\mathrm{cond}}, \boldsymbol{v}_L^{\mathrm{cond}})$ with $\bar{\boldsymbol{v}}_F^{\mathrm{cond}}$ mean-subtracted.
4: **if** $r < t$ **then** *// horizon path active*
5: Midpoint split: $s \leftarrow (t + r)/2$, $h \leftarrow \tau/2$.
6: **No-grad target forwards ($\boldsymbol{\theta}^{-}$):** $\boldsymbol{w}_1 \leftarrow \boldsymbol{w}_{\boldsymbol{\theta}^-}(\boldsymbol{z}_t, s, t)$; $\hat{\boldsymbol{z}}_s \leftarrow \mathrm{Exp}_{\boldsymbol{z}_t}(-h\,\boldsymbol{w}_1)$ factor-wise; $\boldsymbol{w}_2 \leftarrow \boldsymbol{w}_{\boldsymbol{\theta}^-}(\hat{\boldsymbol{z}}_s, r, s)$.
7: **Semigroup target (no JVP):** $\bar{\boldsymbol{v}}_{\mathrm{sg}} \leftarrow \frac{1}{2}(\boldsymbol{w}_1 + \boldsymbol{w}_2)$, with mean subtraction on $\mathcal{F}$.
8: **else** *// $r{=}t$: pure flow-matching sample*
9: Mark sample as flow-matching-only; $\mathcal{L}_{\mathrm{hor}}$ contribution from this sample is set to 0.
10: **end if**
11: Forward pass (with gradients): $h_{\mathrm{feat}} \leftarrow \Phi_{\boldsymbol{\theta}}(\boldsymbol{z}_t, t)$, $\boldsymbol{a}_{\boldsymbol{\theta}} \leftarrow h^a_{\boldsymbol{\theta}}(h_{\mathrm{feat}})$, $\boldsymbol{b}_{\boldsymbol{\theta}} \leftarrow h^b_{\boldsymbol{\theta}}(h_{\mathrm{feat}})$; $\boldsymbol{w}_{\boldsymbol{\theta}} \leftarrow \boldsymbol{a}_{\boldsymbol{\theta}} + \gamma(\tau)\,\boldsymbol{b}_{\boldsymbol{\theta}}$ (with $\gamma(0) = 0$ when $r{=}t$, so $\boldsymbol{w}_{\boldsymbol{\theta}} = \boldsymbol{a}_{\boldsymbol{\theta}}$).
12: Compute $\mathcal{L}_{\mathrm{loc}}$ (Eq. (5)) and $\mathcal{L}_{\mathrm{hor}}$ (Eq. (8), regressing $\boldsymbol{w}_{\boldsymbol{\theta}}$ on $\bar{\boldsymbol{v}}_{\mathrm{sg}}$, summing only over $r{<}t$ samples); backprop $\mathcal{L}_{\mathrm{SRTF}}$ (Eq. (10)) and update $\boldsymbol{\theta}$ with one Adam step.

Algorithm 1: SRTF training step (one mini-batch), semigroup horizon target. $\bar{\boldsymbol{v}}^{\mathrm{cond}}$ denotes the local flow-matching target per Table 4; $\bar{\boldsymbol{v}}_{\mathrm{sg}}$ denotes the semigroup target of Eq. (7); $\boldsymbol{\theta}^{-}$ is the gradient-free target copy $\mathrm{sg}(\boldsymbol{\theta})$, so $\boldsymbol{w}_1$ and the online forward of line 11 can share one backbone evaluation. During the warm-up epochs ($\leq E_{\mathrm{warm}}$), $\lambda_{\mathrm{hor}} \equiv 0$ and lines 5–7 are skipped.

Hybrid sampler. Because $\boldsymbol{a}_{\boldsymbol{\theta}}$ and $\boldsymbol{w}_{\boldsymbol{\theta}}$ share the same backbone and live in the same tangent space, a hybrid schedule of k_{SRTF} transport-expansion steps followed by k_{FM} local flow-matching refinements recovers many-step RMSE at a fraction of the total budget, without retraining; pure SRTF and pure flow matching are its two boundary cases. Algorithm 2 writes out the iteration. The switch k_{SRTF} is the number of leading transport steps, selected per NFE on the validation split (Section 4.1); $k_{\mathrm{SRTF}}{=}K$ is the pure $\boldsymbol{w}_{\boldsymbol{\theta}}$ sampler that Theorem 1 analyzes.

Require: Trained $\boldsymbol{w}_{\boldsymbol{\theta}} = \boldsymbol{a}_{\boldsymbol{\theta}} + \gamma(\tau)\boldsymbol{b}_{\boldsymbol{\theta}}$, schedule $1 = t_K > t_{K-1} > \cdots > t_0 = 0$, switch $k_{\mathrm{SRTF}} \le K$.

Sample $\boldsymbol{z}_{t_K} \sim p_1$: Gaussian on $\mathcal{A}$ and $\mathcal{L}^{\varphi}$, uniform on $\mathcal{F}$.
for $k = K, K-1, \ldots, 1$ **do**
  if $k > K - k_{\mathrm{SRTF}}$ **then** *// transport-expansion step*
    $\boldsymbol{w} \leftarrow \boldsymbol{w}_{\boldsymbol{\theta}}(\boldsymbol{z}_{t_k}, t_{k-1}, t_k)$
  else *// flow-matching refinement step*
    $\boldsymbol{w} \leftarrow \boldsymbol{a}_{\boldsymbol{\theta}}(\boldsymbol{z}_{t_k}, t_k)$
  end if
  $\boldsymbol{z}_{t_{k-1}} \leftarrow \mathrm{Exp}_{\boldsymbol{z}_{t_k}}\big(-(t_k - t_{k-1})\,\boldsymbol{w}\big)$ factor-wise (Table 4).
end for
return $\boldsymbol{z}_{t_0}$: apply φ^{-1} to the lattice factor; threshold atom-type coordinates componentwise at zero to recover $\{-1, +1\}^{7n}$.

Algorithm 2: SRTF hybrid sampling on the flat crystal manifold $\mathcal{M}$.

Table 5: **SRTF across three independent training seeds**, all four CSP benchmarks and the full NFE grid $\{2, 5, 10, 20\}$. Mean $\pm$ standard deviation over seeds $\{42, 43, 44\}$; identical training recipe within each dataset; only the seed differs, and the sampler configuration is re-selected per seed. Best value per dataset and column in bold. Evaluation on the full test split.

Dataset	NFE	$\mathrm{MR}_1 \uparrow$	$\mathrm{MR}_{20} \uparrow$	RMSE$\downarrow$
Perov-5	2	$\mathbf{70.82 \pm 0.56}$	$\mathbf{100.00 \pm 0.00}$	0.3631 ± 0.0234
	5	65.05 ± 0.34	$\mathbf{100.00 \pm 0.00}$	0.2861 ± 0.0030
	10	51.84 ± 0.53	99.91 ± 0.02	0.0749 ± 0.0031
	20	51.55 ± 0.44	99.83 ± 0.05	$\mathbf{0.0682 \pm 0.0034}$
Carbon-24	2	$\mathbf{25.85 \pm 0.47}$	76.35 ± 0.26	0.4108 ± 0.0016
	5	22.25 ± 0.11	85.94 ± 0.89	0.3600 ± 0.0053
	10	20.48 ± 1.36	$\mathbf{88.52 \pm 0.39}$	0.3330 ± 0.0053
	20	19.16 ± 0.58	88.10 ± 0.41	$\mathbf{0.3221 \pm 0.0040}$
MP-20	2	64.02 ± 0.50	83.40 ± 0.24	0.1788 ± 0.0012
	5	67.20 ± 0.32	84.08 ± 0.29	0.1410 ± 0.0022
	10	68.27 ± 0.23	$\mathbf{84.41 \pm 0.33}$	0.0968 ± 0.0027
	20	$\mathbf{68.45 \pm 0.33}$	84.33 ± 0.44	$\mathbf{0.0778 \pm 0.0016}$
MPTS-52	2	16.36 ± 0.31	31.13 ± 0.60	0.3070 ± 0.0060
	5	19.08 ± 0.47	35.33 ± 0.56	0.2517 ± 0.0015
	10	20.73 ± 0.60	36.79 ± 0.37	0.2102 ± 0.0028
	20	$\mathbf{21.25 \pm 0.61}$	$\mathbf{38.26 \pm 0.29}$	$\mathbf{0.1851 \pm 0.0017}$

E CRYSTAL STRUCTURE PREDICTION, IN FULL

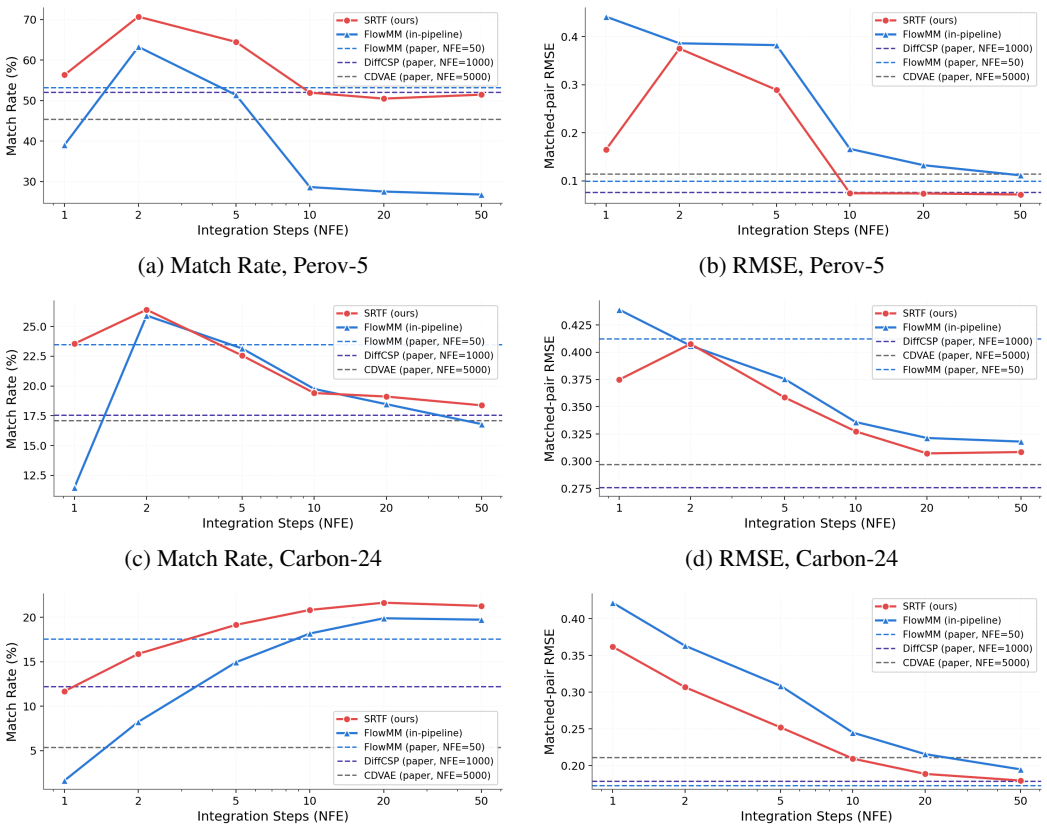
This appendix follows Section 4.2: seed statistics for every cell of Table 1, the NFE curves on the three benchmarks Figure 3 omits, the two same-backbone head-to-head comparisons, the lattice lever, and further qualitative examples on all four benchmarks. Hyperparameters, sampler grids and per-cell measurement protocols are documented with the code release.

E.1 SEED RELIABILITY ACROSS ALL FOUR BENCHMARKS AND EVERY NFE

Table 5 gives the three-seed statistics behind Table 1. Seed noise is small: the standard deviation is at most 1.4 points on the match-rate columns and 0.024 on RMSE. The $\pm$ mixes training-seed variance with the per-seed re-selection of the sampler configuration and is therefore an upper bound on pure seed noise.

E.2 NFE-SWEEP CURVES ON THE REMAINING BENCHMARKS

Figure 6 repeats the MP-20 sweep of Figure 3 on the three remaining benchmarks.



(a) Match Rate, Perov-5 (b) RMSE, Perov-5

(c) Match Rate, Carbon-24 (d) RMSE, Carbon-24

(e) Match Rate, MPTS-52 (f) RMSE, MPTS-52

Figure 6: **CSP match rate (left) and RMSE (right) vs. integration steps, k=1, on Perov-5, Carbon-24 and MPTS-52.** SRTF (red) and FlowMM (in-pipeline) (blue); dashed lines mark the published baseline plateaus. Perov-5 and Carbon-24 peak within one or two steps for both methods; curves are single-run sweeps, and Table 1 gives the three-seed means; MPTS-52 reproduces—and widens—the SRTF-vs-FlowMM gap seen on MP-20 in Figure 3.

Perov-5 and Carbon-24: two steps are enough. On both small, highly symmetric manifolds SRTF passes every published baseline plateau at two steps, clearing FlowMM's fifty-step and DiffCSP's thousand-step values on each dataset at a twenty-fifth and a five-hundredth of their budgets. Match rate then falls away with further integration—a `StructureMatcher` artifact at high symmetry rather than a property of the transport expansion, since the first step or two land near the exactly symmetric positions and further Euler refinement drifts off the matched diagonal, and the same-backbone flow matcher shows the same behavior. Against that flow matcher the two datasets behave differently. On Carbon-24 SRTF leads by twelve points at one step and the two are within one to two points of each other from two steps onward; on Perov-5 SRTF leads throughout and by a widening margin, because the flow matcher's decay is far steeper. The Carbon-24 parity is what the geometric picture predicts: a thin, highly symmetric subset of $\mathcal{M}$ that the local head already reaches within two steps leaves little for the horizon correction to do. The $\boldsymbol{b}$-zeroing measurements say the same thing directly—on Carbon-24 the MR-optimal sampler takes no transport step at two and five steps (Figure 5).

MPTS-52: the MP-20 picture, widened. On the ≤ 52-atom manifold the same comparison separates further and, unlike MP-20, does not close. SRTF passes the DiffCSP thousand-step plateau in two steps and the FlowMM fifty-step plateau in five, a tenfold reduction in budget, and it leads the in-pipeline FlowMM at every point of the sweep—sevenfold at one step, and still measurably ahead at fifty. The natural reading is that the harder and higher-dimensional the manifold, the longer

Table 6: **Head-to-head MF-Riem vs. SRTF on the four CSP benchmarks.** Match rate at k=1 and k=20, and matched-pair RMSE at k=1, paired by NFE budget. Both methods share the CSPNet backbone, the flat crystal product manifold and a matched training budget; **evaluation is on a fixed** 2000**-structure test subset per dataset**, not the full split of Table 1, so cells here are comparable to each other but not to that table. Per-cell winner in **bold**; ties shown with both entries bold.

		Perov-5			Carbon-24			MP-20			MPTS-52		
Method	NFE	$MR_1\uparrow$	$MR_{20}\uparrow$	RMSE↓	$MR_1\uparrow$	$MR_{20}\uparrow$	RMSE↓	$MR_1\uparrow$	$MR_{20}\uparrow$	RMSE↓	$MR_1\uparrow$	$MR_{20}\uparrow$	RMSE↓
MF-Riem	2	64.85	**100.00**	**0.2722**	19.65	**79.65**	**0.4059**	40.85	78.10	0.2414	10.35	28.75	0.3487
SRTF (ours)	2	**68.45**	**100.00**	0.3859	**25.15**	75.25	0.4129	**61.05**	**83.80**	**0.2098**	**20.50**	**38.00**	**0.2958**
MF-Riem	5	**60.30**	**100.00**	0.2338	20.10	**81.25**	**0.3975**	46.35	83.65	0.2053	13.35	34.65	0.2984
SRTF (ours)	5	58.45	**100.00**	**0.1493**	**22.25**	78.40	0.4066	**64.45**	**84.95**	**0.1573**	**23.25**	**41.50**	**0.2490**
MF-Riem	20	**59.70**	**100.00**	0.1945	**20.45**	84.55	0.3788	47.50	84.20	0.1839	14.95	36.30	0.2848
SRTF (ours)	20	54.40	99.95	**0.0702**	20.20	**88.80**	**0.3471**	**67.10**	**85.10**	**0.0922**	**27.75**	**46.25**	**0.1611**

the horizon correction keeps paying, which is also where our JVP implementation fails to converge (Section 4.3).

E.3 Matched-backbone comparison with flow matching

Table 1 compares SRTF against numbers published by other authors, each obtained with its own backbone, budget and sampler tuning. That situates the work in the literature but confounds the objective with everything else in the pipeline. The complementary *controlled* comparison holds all of it fixed: SRTF against FlowMM (in-pipeline), defined in Appendix D, with the sampler swept independently for each method at each budget so that neither is handicapped by tuning.

The few-step gap is where the objective shows up. Figure 3 is that comparison on MP-20: SRTF is at least as good as the matched FlowMM (in-pipeline) at *every* budget on both metrics, and the advantage is concentrated where the theory says it should be—more than forty points at one step, seventeen at two—shrinking to about a point by ten and into seed noise by fifty. As $\tau \to 0$ the horizon correction vanishes by construction and the two methods converge, which is what the curves do.

Precision on commonly matched structures. Raw matched-pair RMSE is selection-dependent: a method that matches more structures is scored on a harder set. Restricting to the structures that *both* methods match, SRTF's MP-20 RMSE is roughly half that of FlowMM (in-pipeline), and below FlowMM's published fifty-step plateau. We report the common-subset figure alongside the raw one wherever precision is compared.

The pattern across the four benchmarks. We repeated the matched-backbone comparison on all four datasets, retraining the FlowMM (in-pipeline) at each dataset's budget. SRTF matches or beats the same-backbone flow matcher on match rate across the board, to within about a point on Carbon-24 and by large margins elsewhere: a decisive few-step match-rate win plus a large common-subset precision win on MP-20, a seven-to-seventeen-point few-step win on Perov-5, and a quarter off common-subset RMSE on MPTS-52. The advantages concentrate where trajectory curvature, and therefore a non-zero $\boldsymbol{b}_\star$, is largest, which is exactly what the expansion predicts; on the small, high-symmetry manifolds, where the local head alone nearly suffices, it holds flow-matching parity.

E.4 Head-to-head comparison with MF-Riem

MF-Riem is the same-backbone MeanFlow-style few-step reference: one horizon-averaged head $\boldsymbol{u}_{\boldsymbol{\theta}}$ supervised against the JVP-based self-consistent MeanFlow target (Geng et al., 2025a), on the same CSPNet backbone, the same manifold, the same NFE sweep and the same training budget as SRTF. The two differ in two ways at once—the architectural decomposition and the target type—so we read this as a head-to-head against the standard few-step reference, not as an isolation of either axis; that is the job of the 2×2 grid of Appendix F.3. Table 6 reports the NFE-paired comparison.

Table 7: **Opening the sampler grid to the lattice knob on Carbon-24**. Configurations are selected on a validation split and evaluated on the held-out test split; "before" selects over the coordinate knob only and "after" over the wider grid including the lattice knob; each row is selected on its own metric, so the RMSE and MR_1 entries of a column need not come from the same configuration, and neither row is the Table 1 cell, which uses a different coordinate-knob setting. RMSE improves at every step budget, and the fifty-step value is better than all three published baselines. The main text's claim rests on the held-out check described below.

NFE	2	5	10	20	50
RMSE, coordinate knob only	0.4032	0.3762	0.3436	0.3147	0.2768
RMSE, both knobs	**0.3472**	**0.3043**	**0.2884**	**0.2800**	**0.2532**
gain	+0.056	+0.072	+0.055	+0.035	+0.024
MR_1 gain (pp)	+0.84	+0.25	+2.81	+0.69	+3.60

The pattern splits by dataset complexity. SRTF takes nine of the twelve $k{=}1$ match-rate pairs and nine of the twelve RMSE pairs. On the two multi-element benchmarks it wins *every* cell: MP-20 by eighteen to twenty points of match rate at every budget with better precision throughout, and MPTS-52 at roughly double MF-Riem's match rate at every budget. On the two small, high-symmetry benchmarks SRTF holds the precision edge at twenty steps, cutting Perov-5 RMSE by two thirds; the match-rate cells there are split between the two methods, a consequence of the matcher artifact described above, under which any additional integration drifts off the exactly symmetric positions. The mechanism is the division of labor between the two heads: SRTF's local head is supervised by a standard flow-matching loss and inherits its precision ceiling, whereas a monolithic head must represent the local tangent velocity and the long-horizon chord with the same output.

Cost, and where the JVP target stops working. At a matched budget MF-Riem takes roughly twice SRTF's training time and four times its memory across the four datasets. A separate experiment on MPTS-52—the largest cells—finds our JVP implementation failing to converge at its own configuration: its loss rises monotonically and reproducibly at both learning rates we tried, while SRTF's falls by two thirds on the identical setup. That run is not the one behind Table 6, whose MPTS-52 cells come from a configuration that did train; we report it because it marks where the JVP path stops being usable rather than merely expensive. Nothing overflows and no divergence rule fires, so the accurate description is stable non-convergence rather than blow-up, which is the stronger statement: an unlucky seed is ruled out. Our JVP path becomes unusable exactly where the transport problem is hardest, while the gradient-free semigroup target trains without incident.

E.5 The lattice lever: a never-swept sampler knob

The sampler applies anti-annealing separately to the coordinate block and to the lattice block. **Every cell reported in this paper before this subsection, and every cell in the pipeline we inherited, sweeps only the coordinate knob.** A knob that exists but is never varied is indistinguishable, in a results table, from a knob that does not exist, so we ask what this one is worth. Carbon-24 is where it should matter most—composition is fixed, so matched-pair RMSE is almost entirely lattice error—and Table 7 reports what opening the grid to it is worth.

Validated on held-out data. This sweep widens Carbon-24's grid roughly fivefold over the rest of the paper, so we validated it on held-out data: each configuration is chosen on the validation split, and the untouched test split serves as the held-out evaluation split. The gain survives there, and those are the numbers the main text uses. MPTS-52 gains less on match rate but reaches a fifty-step RMSE ahead of both FlowMM's and DiffCSP's published values. We did not re-run the main table under the wider grid: Table 1 reports the protocol used uniformly for every method, and this subsection reports what a knob nobody had swept is worth on top of it.

E.6 Further qualitative examples

Figure 7 carries the fourth of the four MP-20 rows of Figure 4, under the same rendering convention. The selection criterion for all four: of 171 rendered MP-20 test structures, the 15 for which SRTF at

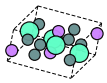
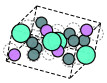
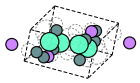
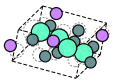
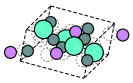
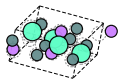
	ground truth	DiffCSP	FlowMM	**SRTF (ours), one model**		
		NFE 1000	NFE 50	**NFE 2**	**NFE 5**	**NFE 10**
$LiEuGe_2$						
	n=16	.447✓	no match	.483✓	.427✓	**.312**✓

Figure 7: **Qualitative CSP on MP-20, continued from Figure 4.** Each generated structure is drawn in the ground truth's frame with the ground truth underlaid as hollow dashed circles, so the visible displacement is the RMSE printed beneath it; the last three panels are the *same* checkpoint at two, five and ten evaluations. FlowMM does not match this composition at its own 50-step budget.

ten steps attains the lowest RMSE of all methods and improves monotonically from two steps, of which four are drawn. The criterion picks cases where the step budget is doing visible work, which is what the figure is for; Table 1 reports the average over the full split.

Figures 8, 9 and 10 repeat the comparison on the other three benchmarks with the same renderer and the same camera. Rows are again drawn from a pool of 171 test structures per benchmark, under an explicit criterion: all three SRTF budgets match, the error falls strictly from two to five to ten steps, and SRTF at ten steps is the most accurate of all methods, ranked by the fraction of the best baseline's error it removes (a baseline that does not match is scored at the matcher tolerance). The criterion is met by 36, 20 and 27 structures on Perov-5, Carbon-24 and MPTS-52, and the top four of each are drawn. The picture is the same on all three: the one checkpoint refines with the step budget and at ten steps lands on the ground-truth sites, on the small symmetric cells and on MPTS-52's largest cells alike.

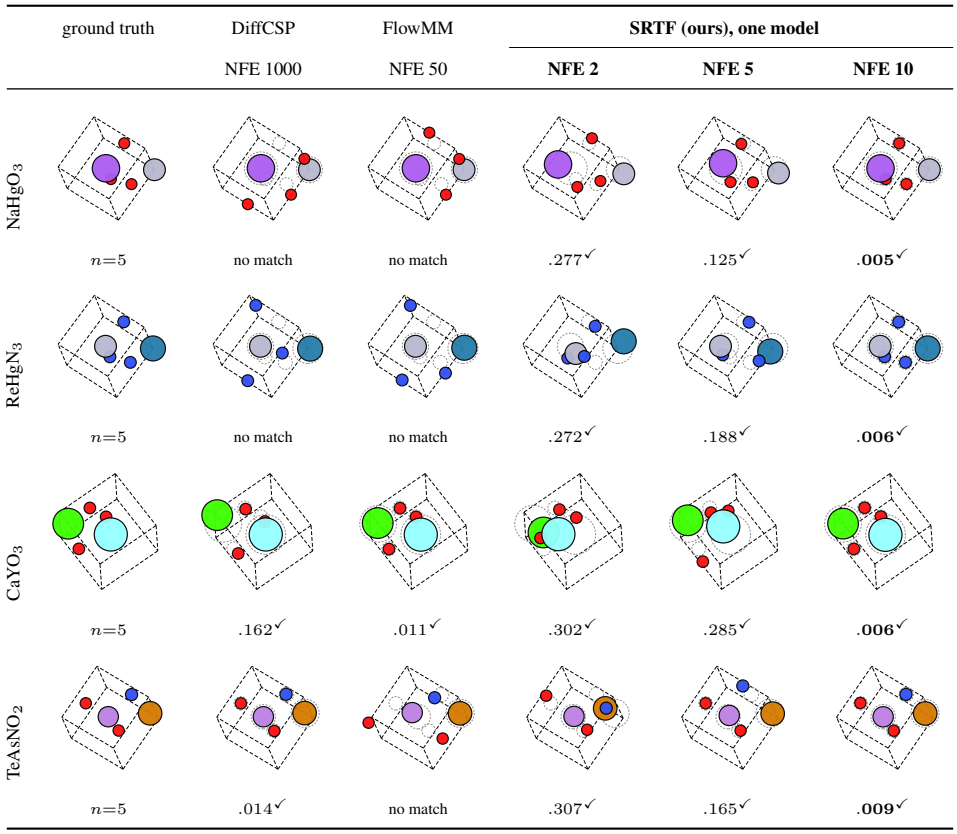


Figure 8: **Same model, more steps, on Perov-5.** Layout and conventions as in Figure 4. At ten steps the SRTF error is below 0.01 in every row, including on the two compositions that neither baseline matches at its full budget.

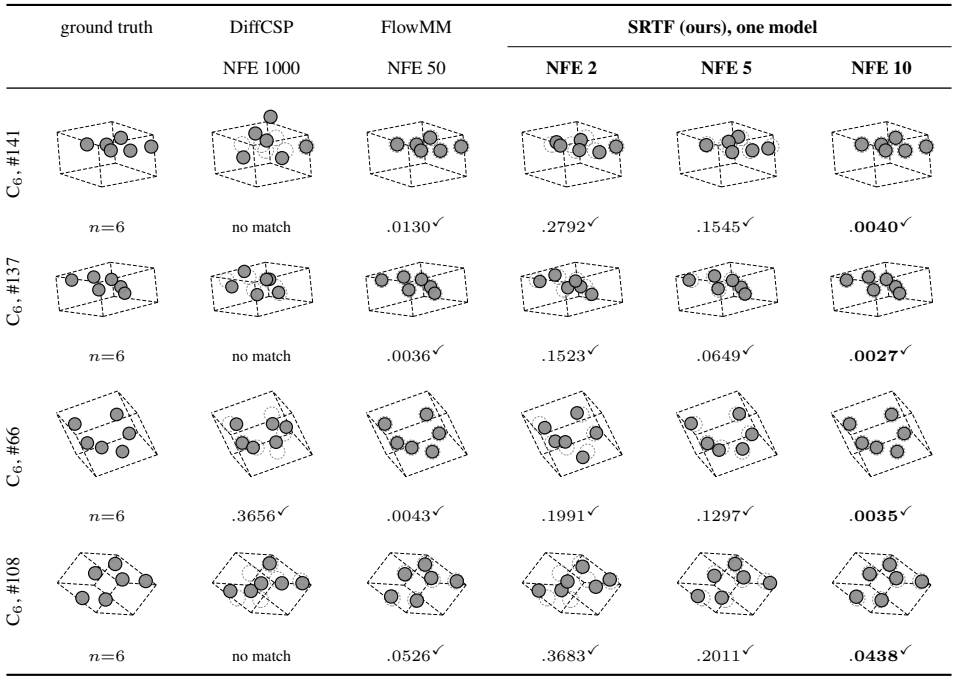


Figure 9: **Same model, more steps, on Carbon-24.** Layout and conventions as in Figure 4; every composition is elemental carbon, so rows differ only in cell and site geometry and are labeled by test index, and RMSE is given to four decimals. The ten-step SRTF structure is the closest to the ground truth in every row.

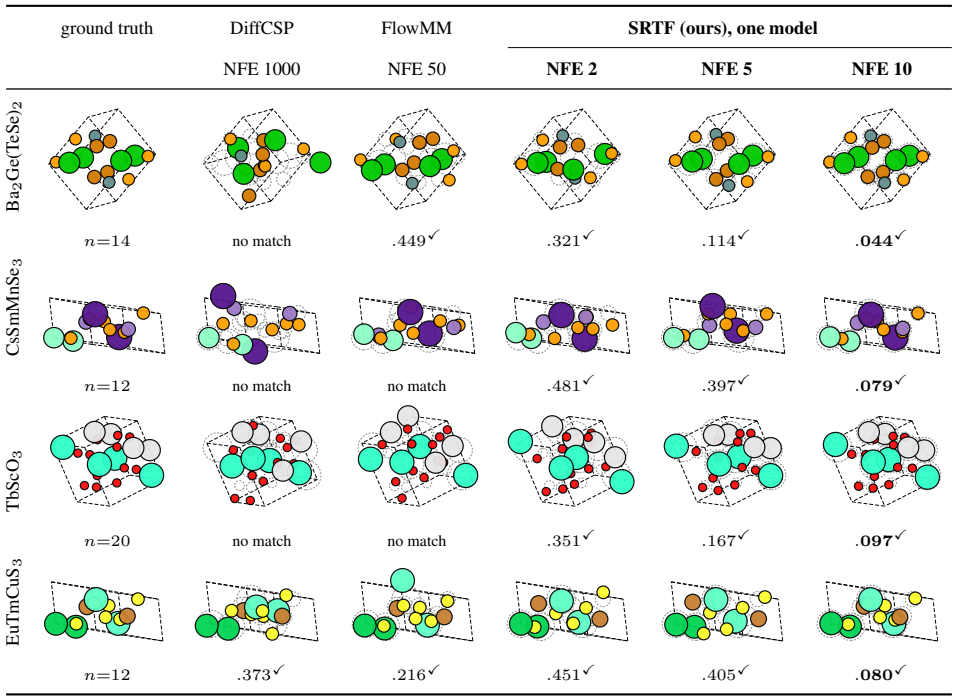


Figure 10: **Same model, more steps, on MPTS-52**, the hardest benchmark. Layout and conventions as in Figure 4. The SRTF error falls by a factor of roughly four to seven between two and ten steps, and on three of the four compositions DiffCSP does not match at its thousand-step budget.

F MECHANISM, ABLATIONS, AND COST

Proposition 1 (iii) predicts a specific object: to leading order in the horizon, the population fixed point has horizon head $\boldsymbol{b}_\star = -\frac{1}{2}\mathrm{D}_t \boldsymbol{v}_\star$. Three cheap measurements turn that prediction into a checkable one, all reusing the trained model with no additional training. **(D1)** tracks the per-factor magnitude of the horizon readout relative to the local one over training. **(D2)** asks the sharper directional question, comparing $\boldsymbol{b}_{\boldsymbol{\theta}}$ against a finite-difference estimate of the material derivative taken along the model's own flow; the estimate is biased, since it uses $\boldsymbol{a}_{\boldsymbol{\theta}}$ in place of $\boldsymbol{v}_\star$ and truncates at first order, so we read it qualitatively—a distribution concentrated well above zero for the semigroup model and near zero for a chord-supervised control is the signature we test for, not numerical equality with $\boldsymbol{b}_\star$. **(D3)** zeroes the horizon head at sampling time, which reduces the field to its local part and turns the sampler into plain Riemannian flow matching on the same weights, and reports the resulting loss in accuracy. A head can be identified in the population sense and still contribute little at a particular budget, so we report the two facts separately rather than conflating them. The two arms are run on identical prior draws throughout, which removes sampling noise exactly and is what licenses reading sub-point entries at all.

(D1, D3) The head is driven to a stable magnitude, and it does work at inference. Table 8 reports both for the semigroup model and for a chord-supervised control trained identically except for the horizon target. Both models start training with comparable readout magnitudes and diverge within the first few hundred epochs: under semigroup supervision the ratio settles at a stable, non-trivial value and stays there, while under chord supervision it decays monotonically to roughly a ninth of that—exactly the collapse toward the gauge floor that Lemma 2 predicts for a target carrying no horizon dependence. The functional consequence follows. The chord model's match rate is essentially invariant to the number of transport steps taken at sampling: its horizon head is inert and the model behaves as flow matching with extra steps, and its accuracy is correspondingly indistinguishable from the matched flow-matching baseline. The semigroup model shows a real dependence on the transport schedule and an enormous few-step advantage over the control—more than forty points at one step—narrowing to a couple of points by twenty, which is the expected convergence, since as $\tau \to 0$ the correction vanishes by construction and both models reduce to the same local field. The regime where the head's inference-time role is unambiguous is the one- and two-step column; at moderate budgets the mechanism is that the semigroup objective, which is well-posed only because the structured field gives $\boldsymbol{b}$ somewhere to live, shapes the shared representation into one that supports long jumps.

(D2) The direction is right, and it is sharpest where few-step sampling happens. D1 and D3 establish that the horizon head is active and load-bearing; D2 asks whether it points where Proposition 1 (iii) says it should. Table 9 reports the uncentered cosine between $\boldsymbol{b}_{\boldsymbol{\theta}}$ and the finite-difference estimate on held-out MP-20 bridge states. Pooled over all times the semigroup model is positively aligned on four states in five, and the chord control is not aligned at all; the separation is stable across the finite-difference step size. The time-resolved picture is the more informative one. Semigroup supervision recovers the predicted direction most sharply at the *noise* end of the trajectory, where nearly every state is positively aligned—and under the reverse-time convention that is exactly where a one- or two-step sampler places its jumps. The chord control moves the other way over the same range, becoming *anti*-aligned at large t. The two objectives therefore do not merely differ in how strongly they excite the horizon head: they disagree about its direction precisely in the regime where few-step accuracy is decided. That is a sharper statement than either D1 or D3 alone supports, and it is the mechanism behind the head-zeroing split of Figure 5.

The estimator bounds what this can show: it is first-order and uses $\boldsymbol{a}_{\boldsymbol{\theta}}$ in place of $\boldsymbol{v}_\star$, so the claim it supports is directional consistency with $-\frac{1}{2}\mathrm{D}_t \boldsymbol{v}_\star$, not numerical equality. Appendix F.2 removes the estimator entirely on a synthetic manifold with closed-form coefficient fields, where the same separation appears far more sharply.

F.1 THE PREDICTED RESIDUAL COEFFICIENT, MEASURED ON TRAINED MODELS

Proposition 1 (iii) gives the semigroup residual as $\mathcal{R} = \xi(1-\xi)\tau^2(2\boldsymbol{b} + \mathrm{D}_t \boldsymbol{a}) + \mathcal{O}(\tau^3)$. The measurable part of that statement is the leading *coefficient*, not the exponent: the bracket vanishes only under exact pointwise identification, so any residual identification error is itself an $\mathcal{O}(\tau^2)$

Table 8: **Identification diagnostics on MP-20** (matched 2000-epoch budget; 2000-structure evaluation subset). $\overline{\|\boldsymbol{b}_{\boldsymbol{\theta}}\|}/\overline{\|\boldsymbol{a}_{\boldsymbol{\theta}}\|}$ is the terminal ratio of readout magnitudes (D1). The MR columns are $k{=}1$ match rate; "k_{SRTF}-spread" is the range of MR at $\mathrm{NFE}{=}20$ across the number of transport steps k_{SRTF} used at sampling, a direct measure of whether the horizon head is doing anything at inference (D3).

Horizon target	$\overline{\|\boldsymbol{b}\|}/\overline{\|\boldsymbol{a}\|}$	NFE=1	NFE=2	NFE=5	NFE=20	k_{SRTF}-spread @20
Semigroup (SRTF)	**0.099**	**56.1**	**63.5**	**65.8**	**68.4**	2.1 pp
Bridge chord (control)	0.011	14.1	47.3	62.0	66.7	<1 pp
Semigroup advantage	9×	+42.0	+16.2	+3.8	+1.7	—

Table 9: **(D2) Directional alignment of $\boldsymbol{b}_{\boldsymbol{\theta}}$ with $-\frac{1}{2}\mathrm{D}_t\boldsymbol{v}_\star$**, resolved by endpoint time t. Mean uncentered cosine on the fractional-coordinate block of held-out MP-20 states (Table 10 reports the median over $t \in (0.55, 0.95)$, hence its larger value); "frac>0" is the fraction of states with positive alignment. Under the reverse-time convention large t is the noise end, where a one- or two-step sampler takes its jumps. The $\mathcal{A}$ block is not applicable, since CSP conditions on composition; the $\mathcal{L}$ block is measured and does not separate the two models.

	Semigroup (SRTF)		Bridge chord (control)	
t bin	align	frac>0	align	frac>0
$[0.0, 0.2)$	+0.177	0.687	+0.032	0.540
$[0.2, 0.4)$	+0.053	0.574	+0.025	0.528
$[0.4, 0.6)$	+0.306	0.786	−0.015	0.454
$[0.6, 0.8)$	**+0.623**	**0.972**	−0.102	0.334
$[0.8, 1.0)$	+0.541	0.965	−0.194	0.210
pooled	+0.346	0.800	−0.052	0.410

contribution and a finite network has a τ^2 residual whatever its heads contain. For each trained model we therefore obtain the coefficient twice—once by fitting the residual at small horizon, once by evaluating $\|2\boldsymbol{b}_{\boldsymbol{\theta}} + \mathrm{D}_t\boldsymbol{a}_{\boldsymbol{\theta}}\|$ directly—and compare.

Three readings. First, the measured residual follows the predicted expansion quantitatively on trained networks. Second, the measurement is discriminative: on the $\gamma \equiv 1$ arm, whose field has no τ-coefficient so that the bracket evaluated on its readouts is not its leading term, the same comparison misses by a wide margin. Identification itself is read from the cosine column, where the semigroup arms align and the chord control does not. Third, the chord control also satisfies the identity, which is the correct outcome—the identity is a property of the *residual*, and the chord control satisfies it with an inert horizon head and a bracket carried by $\mathrm{D}_t\boldsymbol{a}_{\boldsymbol{\theta}}$ alone. What separates the arms is not whether the identity holds but what it holds *at*: the semigroup arms reach it with an active, correctly oriented horizon head, the chord arm with an inert, anti-aligned one.

F.2 Identification on a synthetic flat product manifold

On crystals there is no closed form for $\boldsymbol{v}_\star$, so the directional diagnostic must compare $\boldsymbol{b}_{\boldsymbol{\theta}}$ against a finite-difference estimate that itself uses $\boldsymbol{a}_{\boldsymbol{\theta}}$. To separate the method from its estimator we repeat the measurement on a synthetic flat product manifold $\mathbb{R}^2 \times (\mathbb{R}/\mathbb{Z})^2$—the same product structure as the crystal manifold, including the mod-1 factor—with a target for which the coefficient fields are available in closed form.

Table 11 reports the four arms. With the estimator removed the separation is unambiguous: the semigroup arm recovers the closed-form horizon field closely, against essentially zero for the chord control, which also explains why the crystal cosine—computed against a finite-difference estimate—reads lower in absolute terms. One-step sample quality separates by an order of magnitude on a manifold small enough that nothing is hidden by capacity. The $\gamma \equiv 1$ arm does not merely degrade but fails catastrophically, and it is the only arm whose local head is destroyed as well, which is what a broken boundary condition should do.

Table 10: **Measured versus predicted residual coefficient** (MP-20, $\xi{=}\frac{1}{2}$, small-τ fit). "Bracket" is the prediction $\|2\boldsymbol{b}_{\boldsymbol{\theta}} + \mathrm{D}_t \boldsymbol{a}_{\boldsymbol{\theta}}\|$; "measured" is the fitted τ^2-coefficient of $\|\mathcal{R}\|/[\xi(1-\xi)]$; the cosine is the median per-state value over $t \in (0.55, 0.95)$. The ratio is within about 3% of unity for every arm to which the expansion applies. The $\gamma \equiv 1$ arm is the control the prediction *should* fail on: its field has no τ-coefficient, so the bracket evaluated on its readouts is not its leading term and the prediction does not apply to it; the ratio duly misses by 76%.

Arm	$\cos(\boldsymbol{b}_{\boldsymbol{\theta}}, -\frac{1}{2}\mathrm{D}_t \boldsymbol{a}_{\boldsymbol{\theta}})$	bracket (predicted)	measured	ratio
SRTF	$+0.647$	3.142	3.165	$\mathbf{1.007}$
SRTF, small-τ law	$+0.680$	2.010	2.060	$\mathbf{1.025}$
Chord control	-0.106	2.754	2.843	$\mathbf{1.032}$
$\gamma \equiv 1$ (boundary condition broken)	-0.117	3.342	5.868	1.756

Table 11: **Synthetic manifold with exact coefficient fields** ($\mathbb{R}^2 \times (\mathbb{R}/\mathbb{Z})^2$, three seeds). $\cos(\boldsymbol{b}_{\boldsymbol{\theta}}, \boldsymbol{b}_\star)$ compares the learned horizon head against its closed-form target; W_2 is the 2-Wasserstein distance to the true density at one step.

Arm	$\cos(\boldsymbol{a}_{\boldsymbol{\theta}}, \boldsymbol{v}_\star)$	$\cos(\boldsymbol{b}_{\boldsymbol{\theta}}, \boldsymbol{b}_\star)$	W_2 at NFE$=1$ $\downarrow$
Flow matching	$+0.999$	—	0.573
Chord control	$+0.999$	$+0.022$	0.576
$\gamma \equiv 1$	$+0.102$	$+0.022$	$8.0{\times}10^6$
SRTF	$+0.998$	$\mathbf{+0.848}$	$\mathbf{0.067}$

F.3 The $2{\times}2$ decomposition–supervision control grid

The SRTF-vs-MF-Riem comparison varies two design axes at once: the architectural decomposition (structured pair vs. monolithic head) and the target formulation (gradient-free semigroup bootstrap vs. JVP-based MeanFlow identity). The two missing corners isolate them. **Mono-SG** keeps the semigroup target and replaces the structured field by a single head $\boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{z}, r, t)$ taking the horizon as a backbone input—on this flat manifold, essentially the semigroup form of Riemannian MeanFlow (Woo et al., 2026); **SRTF-JVP** keeps the SRTF architecture and replaces the semigroup target by the MeanFlow self-consistent identity, computed by forward-mode differentiation through the model exactly as in Geng et al. (2025a). All four cells share the backbone, the diagonal anchor, the time law, the weighting schedule, warm-up, EMA, batch size, epochs and seeds; the full recipe for each is in the code release.

Table 12 reports the grid.

Supervision has the larger effect. Holding the decomposition fixed and swapping the semigroup target for the JVP identity costs fifteen to seventeen points at both ends of the sweep. More telling than the size of the gap is its shape: SRTF-JVP *plateaus* from five steps onward, gaining nothing from additional refinement. A field whose small-horizon behavior is mis-specified cannot be repaired by refinement at sampling time, however many steps are taken; the plateau is consistent with that picture.

The decomposition is what makes that supervision affordable. Holding supervision fixed, the structured field leads the monolithic one by a few points at the smallest and largest budgets, and the two are level at five steps. Its decisive contribution, though, is cost. Because the horizon enters SRTF only through an analytic scalar, the first target evaluation can reuse the backbone features of the differentiated forward; a monolithic head takes the horizon as a backbone input and cannot, so it must run two genuine extra forwards, at nearly twice the peak training memory for the same objective. The structured decomposition is what makes semigroup supervision *cheap*.

Both axes matter and supervision matters more, so we frame SRTF as a training-objective contribution with a representation that makes it practical, rather than claiming the decomposition drives the few-step gains.

Table 12: **The 2×2 decomposition–supervision grid on MP-20** (matched 2000-epoch budget, identical backbone and data, 2000-structure evaluation subset, sampler configuration per cell selected on the validation split). Match rate at $k=1$, with training peak memory (batch 256, width 512; as in Table 15). The MF-Riem corner is reported at its own matched budget in Table 6.

Supervision	Decomposition	NFE=1	NFE=2	NFE=5	NFE=20	train peak (GB)
Semigroup	structured (**SRTF**)	**56.1**	**63.5**	65.8	**68.4**	**4.17**
Semigroup	monolithic (Mono-SG)	53.4	61.5	**66.3**	66.6	7.56
JVP identity	structured (SRTF-JVP)	38.9	51.1	53.1	52.8	15.31
JVP identity	monolithic (MF-Riem)	*see Table 6*				

Table 13: **Ablations on MP-20 CSP** (matched 2000-epoch budget, 2000-structure evaluation subset). Match rate at $k=1$. The chord-supervised control replaces the semigroup target with the degenerate bridge chord; the γ rows change only the scale profile, and $\gamma \equiv 1$ removes it. For the $\gamma \equiv 1$ row, k_{SRTF} is the number of transport steps the sampler's configuration search selects at that budget: it selects zero everywhere, i.e. offered the transport branch it declines to take a single step.

Variant	NFE=1	NFE=2	NFE=5	NFE=20
Chord-supervised control (a)	14.1	47.3	62.0	66.7
$-\boldsymbol{b_\theta}$ at inference, NFE=2 (b)	—	46.4	—	—
$\gamma(\tau)=\frac{2}{\pi}\sin(\pi\tau/2)$ (c)		within ±0.4 pp of the linear profile		
$\gamma(\tau)=\tau(1-\tau/2)$ (c)		within ±0.4 pp of the linear profile		
$\gamma \equiv 1$, no τ factor (d)	14.4 ($k_{\mathrm{SRTF}}=0$)	45.7 ($k_{\mathrm{SRTF}}=0$)	62.1 ($k_{\mathrm{SRTF}}=0$)	67.0 ($k_{\mathrm{SRTF}}=0$)
Full SRTF	**56.1**	**63.5**	**65.8**	**68.4**

F.4 Additional ablations

We vary four design choices in isolation on MP-20, holding everything else fixed: **(a)** the horizon supervision, replaced by the degenerate bridge chord of Lemma 2; **(b)** the horizon head, zeroed at inference so the field reduces to its local part; **(c)** the scale profile $\gamma(\tau)$, replaced by two smooth alternatives that preserve the boundary conditions $\gamma(0)=0, \gamma'(0)=1$; and **(d)** the horizon factor removed altogether, $\gamma \equiv 1$, which breaks them.

The supervision is the load-bearing choice. Table 13 collects the four variants. Replacing the semigroup target with the bridge chord costs more than forty points at one step and sixteen at two, and the resulting model is statistically indistinguishable from a matched flow-matching baseline—it is flow matching with extra steps, exactly as Lemma 2 predicts for a target carrying no horizon information. Zeroing the horizon head at inference in the trained semigroup model costs seventeen points at two steps, confirming that the head is functionally active at low budgets; Figure 5 repeats the measurement on the full split and all four datasets and agrees.

γ is structure, not a tuning knob. Changing the *shape* of γ at fixed boundary conditions barely matters: both non-linear alternatives land within half a point of the linear profile, which is what the first-order analysis predicts, since all three agree to first order. We adopt the linear profile for simplicity. Removing the factor altogether is a different kind of change, because it destroys the zero-horizon boundary condition that makes $\boldsymbol{a_\theta}$ the local field and pins the split between the heads (Proposition 1 (ii), Remark 7). The decomposition then does not degrade but collapses: the sampler declines the transport branch at every budget, one-step accuracy sits at the flow-matching floor, and the magnitude ratio $\|\boldsymbol{b_\theta}\|/\|\boldsymbol{a_\theta}\|$ never settles, swinging by more than an order of magnitude across training instead of staying in the narrow band the correctly scaled model holds. That is the mirror image of the chord-supervised control, where the same ratio collapses toward zero. Both failure modes leave the sampler with an unusable horizon head—one by starving it, the other by letting it run away—and what separates them from SRTF is whether the horizon supervision carries information and whether γ anchors the field at $\tau=0$.

Table 14: **The fixed point does not move with the split point, the leg count, or the warm-up—and does move with the horizon law** (MP-20, matched 2000-epoch recipe, 2000-structure evaluation subset, Δ in MR_1 percentage points against the reference run). The reference band for the learned decomposition $\|\boldsymbol{b_\theta}\|/\|\boldsymbol{a_\theta}\|$ is 0.09–0.17 across the runs of this paper.

Variant	NFE=1	2	5	10	20	$\|\boldsymbol{b}\|/\|\boldsymbol{a}\|$	verdict
$\xi \sim \mathrm{Unif}(0.1, 0.9)$, weighted target	-0.32	$+0.10$	$+0.52$	$+0.13$	-0.52	0.133	invariant
Three equal legs (N=3)	-0.33	-0.24	$+0.18$	-0.08	-0.52	0.150	invariant
No warm-up (E_{warm}=0)	$+1.12$	$+0.20$	-0.24	-0.61	-0.98	0.152	invariant
Small-horizon law ($\tau \leq 0.05\,t$)	$\mathbf{-42.95}$	-12.03	-0.55	-0.27	-0.06	0.147	**not free**

F.5 Invariance of the objective, and the one parameter that is not free

The semigroup objective has two free structural parameters we have fixed by fiat: the split point ξ and the number of legs N. The theory says the first-order identification depends on neither (Remark 2). Three training runs test that, and a fourth tests the one parameter that is *not* free.

Table 14 reports the four runs. **The first three rows are one structural result, not three null ablations.** The objective is a two-parameter family whose fixed point does not move: accuracy is unchanged to within about a point and, more tellingly, the learned decomposition lands in the same narrow band in every case. That is what Remark 2 predicts, and it is a robustness property worth having—the objective has no schedule to tune. The warm-up behaves the same way, so we keep it as a cheap stabilizer rather than a necessity, and three legs cost more time than two without buying anything.

The fourth row is the exception that validates a default. The expansion of Section 3.1 is a small-τ statement, so concentrating the training horizon near zero ought to help—and it does, for identification: that run has the best directional alignment and the smallest bracket of any arm we trained (Table 10). It is also more than forty points worse at one step. One-step sampling evaluates the field far beyond any horizon it saw in training, and no amount of identification quality survives that extrapolation. The default $r \sim \mathrm{Unif}(0, t)$ is therefore not an arbitrary choice but a necessary compromise between the limit the theory describes and the finite horizons the sampler actually uses, and the compromise is visible because the run that optimizes one end of it is catastrophic at the other.

F.6 Are the intermediate states physical?

The flat lattice chart is what makes finite jumps compose exactly, but the true space of unit cells is not flat, so a natural objection is that large jumps in the chart could pass through, or land on, geometrically degenerate cells. Sampling MP-20 with identical prior draws at every step budget, we logged every intermediate state and checked it for a non-positive-definite metric tensor, a cell angle outside $[60^\circ, 120^\circ]$, and an interatomic contact below 0.5 Å. **We observe no such degeneracy in the tested trajectories**: not one structure, at any step index and any step budget, produced a non-positive-definite metric—the count is exactly zero everywhere, including the two-step sampler, whose single jumps are the largest. Large displacements in the flat chart do not produce unrealizable cells. One cost is worth recording: close contacts are budget-dependent, 2.44% of two-step samples carrying a sub-0.5 Å contact against 0.59% at fifty steps and 0.00% in real MP-20 data, so fewer steps buy speed at a small price in local geometry. The trajectory data dispose of the related worry about mid-trajectory pathology: angle violations along the trajectory are an order of magnitude rarer than at the endpoint, and contact violations start high—a property of the Gaussian prior, where atoms are placed independently—and fall monotonically toward their final value, so the sampler never passes through a region worse than the prior it starts from.

F.7 Training and inference cost

Table 15 reports per-step training and inference costs on identical hardware and backbone; Figure 11(b) repeats the training measurement at width 512 and batch 256 in a separate run, so its absolute values differ slightly.

Table 15: **Measured efficiency on MP-20.** Training: per-step wall-clock and peak GPU memory on the shared backbone (batch 256, width 512), all rows measured in one session on an otherwise idle GPU and on identical batches, so both columns are comparable across every row (median of 7 interleaved rounds; interquartile range within $\pm 2\%$). Inference: per-structure latency at $\mathrm{NFE} \in \{5, 10, 20, 50\}$. The semigroup target costs additional forward passes but no additional activation memory; the JVP target costs both.

		Training (per step)		Inference (per structure, ms)			
Supervision	Representation	ms/step ↓	peak (GB) ↓	NFE=5	NFE=10	NFE=20	NFE=50
Flow matching	single head	**60.4**	**4.12**	0.58	1.30	2.32	5.94
Semigroup	structured (**SRTF**)	77.1	4.17	0.58	1.16	2.47	6.93
Semigroup	monolithic (Mono-SG)	173.9	7.56	—	—	—	—
JVP identity	structured (SRTF-JVP)	165.4	15.31	—	—	—	—
JVP identity	structured, fused[†]	108.2	12.25	—	—	—	—
JVP identity	monolithic (MF-Riem)	107.1	12.25	—	—	—	—

[†]SRTF-JVP with its auxiliary outputs $(\boldsymbol{a}_{\boldsymbol{\theta}}, \boldsymbol{b}_{\boldsymbol{\theta}})$ returned by the JVP pass itself instead of a second differentiated forward; identical loss and gradients (relative difference 8×10^{-7}). Its cost equals MF-Riem's, so the JVP overhead is a property of the identity, not of the representation.

Where the cost falls. Relative to plain flow matching, the semigroup target costs $1.28\times$ the per-step wall-clock, because it requires extra gradient-free evaluations, and $1.01\times$ the memory, because those evaluations store no activations. Every JVP variant costs both: 1.77–$2.74\times$ the time and 3.0–$3.7\times$ the memory. The fair comparison is with the fused arm, the cheapest form of the JVP identity: SRTF is still $1.40\times$ faster per step (77.1 vs. 108.2 ms) at $2.9\times$ less memory (4.17 vs. 12.25 GB). The monolithic semigroup arm is the slowest of all (173.9 ms), because its target needs additional full forward passes of the (r, t)-conditioned backbone, whereas SRTF reuses its main forward pass; the structured decomposition therefore also saves compute, on top of its accuracy edge in Table 12. Since peak memory is what bounds trainable model size and batch size, that is the axis on which the JVP-free construction pays off most. The inference columns confirm that the dual readout is free at sampling: full-field and local-only sampling differ by at most a millisecond per structure at every budget, in both directions, i.e. within measurement noise. At full-run scale the gap against MF-Riem is larger still (Appendix E.4): roughly twice the time and four times the memory.

Is the fixed epoch budget fair to both methods? A fixed budget favors whichever method converges faster, so the comparison is sound only if the budget is past convergence for both. An instrumented re-run of SRTF and the in-pipeline FlowMM arm confirms it is: both rise steeply and then plateau, and neither gains anything over the final quarter of the budget, while training loss continues to fall without buying held-out accuracy—the ordinary signature of a converged run rather than a truncated one. The same instrumentation gives the magnitude diagnostic of Appendix F over training: SRTF's ratio $\|\boldsymbol{b}_{\boldsymbol{\theta}}\|/\|\boldsymbol{a}_{\boldsymbol{\theta}}\|$ drops sharply the moment the warm-up ends and the horizon loss switches on, then holds a narrow band for the rest of training.

F.8 Scaling: what the memory headroom buys

Figure 11 shows that the semigroup objective leaves the same trainable batch and width as plain flow matching, whereas our JVP implementation does not. Table 16 asks what that headroom converts into under a fixed wall-clock budget.

Table 16 reports that comparison. **The headroom is feasibility, not throughput.** All three SRTF arms consume roughly the same number of structures in the same wall-clock, since the GPU is saturated regardless of batch, so raising SRTF's own batch does not make it faster. The result the comparison is for is the JVP row: in the identical budget the JVP arm sees about 30% fewer structures at nearly eighteen times the memory, and ends 25–32 points below every SRTF arm, more than ten times the standard error. Removing the original implementation's handicap (unfused, batch 1536: 8.7M structures, 86.0 GB, MR_1 24.2) raises the JVP arm by 9.2 points without closing that gap; the two large-batch SRTF arms are within noise of each other. These MR_1 values are not comparable to Table 1, which uses the full test split, the full training budget and other NFEs. That is the concrete

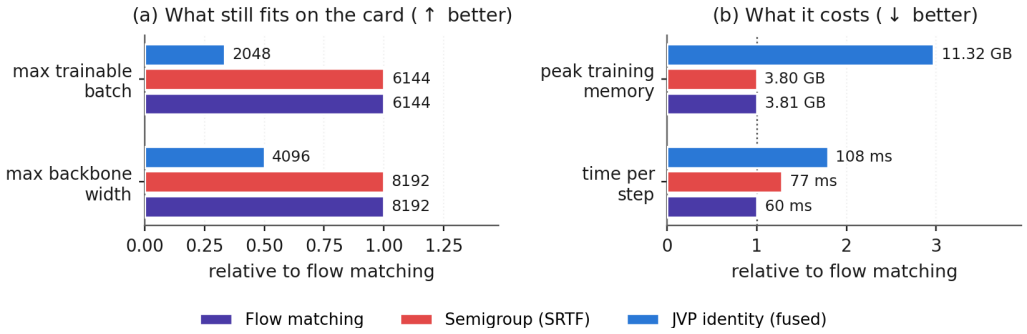


Figure 11: **The semigroup objective buys finite-horizon supervision at flow-matching memory** (MP-20, same backbone throughout, idle GPU; bars relative to flow matching, absolute values labeled). The JVP arm is the fused implementation of Table 15, the cheapest form of the JVP identity. *(a)* Largest batch trainable at width 512 and largest width at batch 256, the last rung before OOM: SRTF matches flow matching on both axes (batch 6144, width 8192, i.e. 2.5B parameters), against 2048 and 4096 (0.64B) for JVP. *(b)* Peak memory and time per step at width 512, batch 256.

Table 16: **Fixed wall-clock budget on MP-20**: 100 minutes per arm on an idle GPU, learning rate scaled as $3 \times 10^{-4}\sqrt{B/256}$, warm-up over the same 400k structures for every arm. Final MR_1 is measured at NFE 20 with EMA weights on a fixed subset of 500 MP-20 test structures, identical for every arm (standard error ≈ 2 pp). The JVP arm runs at its own memory ceiling, using the fused implementation of Table 15 (identical gradients, $\sim$20% less memory); SRTF runs at its default batch and at two batches the JVP path cannot afford.

Arm	batch	steps	structures seen	peak GB	final MR_1
JVP identity (fused)	1792	7,272	13.0M	79.7	33.4
SRTF	256	70,657	18.1M	**4.5**	**65.0**
SRTF	2048	9,146	18.7M	29.5	58.8
SRTF	4096	4,538	18.6M	57.4	60.2

form of the claim—the objective's memory profile is what makes a configuration trainable at all. A companion run at matched steps finds accuracy peaking well below the largest trainable width, so the width ceiling of Figure 11 is headroom in hand rather than a constraint being felt.

G Broader Impacts

Few-step crystal generators lower the cost of exploring new materials for energy storage, catalysis, and semiconductor applications (Ling, 2022; Choubisa et al., 2023). Faster generation can accelerate beneficial materials discovery, but it also lowers the barrier to screening large design spaces; downstream use of generated structures (e.g., for synthesis planning) should include validation pipelines to filter unstable or unsafe compositions.