EQUILIBRIUM FLOW MAPS

Anonymous authors
Paper under double-blind review

ABSTRACT

Flow matching has emerged as a powerful framework for generative modeling, but sampling typically requires numerically integrating a learned time-dependent vector field, requiring many network evaluations. Flow maps accelerate generation by learning the ODE solution operator directly, replacing many small numerical integration steps with a few large advances along the generative trajectory. Separately, recent methods such as Equilibrium Matching replace the time-dependent dynamics of flow matching with an autonomous vector field with data as equilibria, yielding dynamics with a natural infinite time horizon. These two directions suggest complementary goals: efficient generation with few network evaluations and principled iterative refinement as additional inference-time compute becomes available. To address these two goals simultaneously, we introduce *Equilibrium Flow Maps* (EqFM), which combine flow map amortization with autonomous equilibrium dynamics in a single model. EqFM compactifies the infinite-horizon autonomous dynamics onto a bounded domain, so the same network can advance a sample by any duration along the flow, from short refinement steps to a direct jump to equilibrium. We construct an autonomous field through infinite-time stochastic interpolants and derive complementary Lagrangian, Eulerian, and semigroup characterizations that lead to practical flow-map objectives for both direct training and distillation from an autonomous teacher. EqFM supports inference from one-step generation to iterative refinement without changing the underlying model. On ImageNet-256, EqFM improves from FID 4.28 at 1 NFE to 2.09 at 8 NFE, approaching its 500 NFE teacher at 2.00.

1 INTRODUCTION

Flow- and diffusion-based generative models produce high-quality samples across domains ranging from protein design [1], to image and video synthesis [2–5], to language generation [6]. They transport a simple noise distribution to data through learned time-dependent dynamics [4, 7]. These models follow a prescribed trajectory over a fixed horizon. While highly expressive, these methods are expensive at inference time due to the large number of integration steps. To address this, recent efforts aim to learn the flow map, which amortizes this computation into a learned solution operator, allowing one- or few-step generation [8–10].

Surprisingly, recent work has shown that strong generative performance can be achieved without explicit time or noise conditioning, using instead time-independent or autonomous generative dynamics [11–15]. These two directions are not immediately compatible in their standard formulations. Flow maps amortize the integration of finite generative dynamics, but equilibrium-based models are defined over an infinite time horizon, with the terminal state reached only asymptotically.

This raises the central question of our work:

Can we amortize infinite-horizon autonomous generative dynamics into one or a few learned updates?

We answer this question affirmatively, introducing *Equilibrium Flow Maps* (EqFM), a new class of flow maps for autonomous equilibrium dynamics. Conventional flow maps require separate start and end times because the underlying vector field is explicitly time-dependent. For an autonomous system, we show that the analogous solution operator depends only on a single duration specifying how long to advance the dynamics. The same model can therefore evaluate its terminal map directly for

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

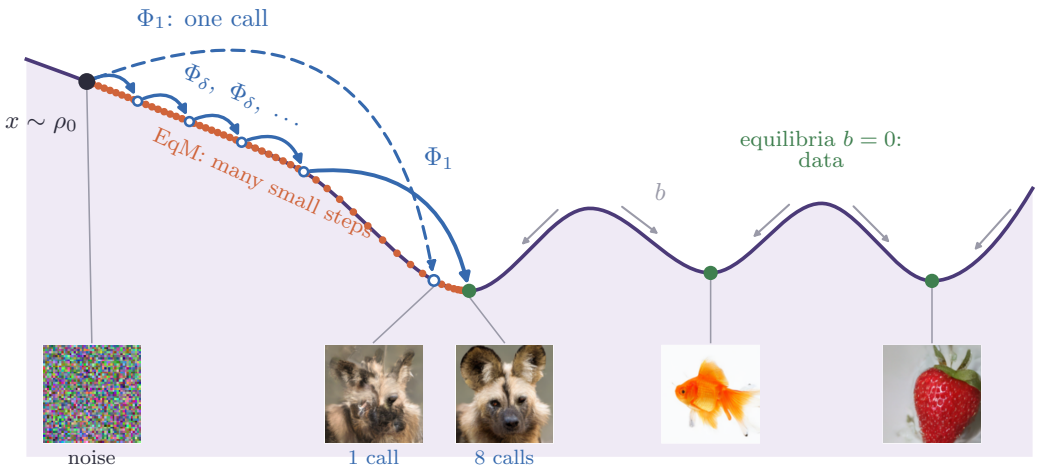


Figure 1: **Equilibrium Flow Maps (EqFM).** Autonomous field b whose equilibria ($b = 0$) are data. Equilibrium Matching (EqM, orange) descends with many small steps $x \leftarrow x + \eta\, b(x)$, which shrink near equilibrium. EqFM instead learns the flow map Φ_σ, which jumps along the same descent by any duration $\sigma \in [0, 1]$. Short advances Φ_δ compose and are followed by the terminal map Φ_1. A single learned call Φ_1 (dashed) lands near equilibrium.

one-step generation or compose finite-duration maps for few-step sampling and iterative refinement. Figure 1 illustrates the autonomous field and how EqFM advances samples along its trajectories.

Standard flow map integration of the autonomous field approaches the data manifold but reaches the terminal state only in the infinite-time limit. As a result, there is no natural finite time at which to define the terminal map. We therefore compactify the unbounded evolution onto a bounded interval while preserving the autonomous composition structure. We then construct infinite-time stochastic interpolants and show that, under suitable assumptions, the induced autonomous dynamics transport the source distribution to the target. This yields practical objectives for training EqFM from scratch or distilling it from an autonomous teacher.

Our contributions are:

- We develop an infinite-horizon stochastic interpolant formulation of autonomous generative transport and establish correct source-to-target transport in the infinite-time limit.
- We derive Lagrangian, Eulerian, and semigroup characterizations of EqFM that yield practical objectives for training from scratch or by distillation.
- We introduce stabilized classifier-free guidance for autonomous fields and distill the resulting guided dynamics into our flow map, enabling guided generation from finite-duration refinement to direct sampling.
- We validate EqFM on CIFAR-10 and ImageNet-256 and show that the learned maps can be composed over many network calls. On ImageNet-256, EqFM improves from FID 4.28 at 1 NFE to 2.09 at 8 NFE and 2.05 at 32 NFE, compared with 2.00 for its 500-NFE teacher.

2 Background

Stochastic interpolants and flow matching. Let $\rho_1 \in \mathcal{P}(\mathbb{R}^d)$ denote the data density, and let $\rho_0 \in \mathcal{P}(\mathbb{R}^d)$ denote a simple prior density, such as a Gaussian. Given independent $x_0 \sim \rho_0$ and $x_1 \sim \rho_1$, flow matching defines the stochastic interpolant

$$I_t = \alpha_t x_0 + \beta_t x_1, \qquad t \in [0, 1], \tag{1}$$

where $\alpha, \beta : [0, 1] \to [0, 1]$ are differentiable with $\alpha_0 = \beta_1 = 1$ and $\alpha_1 = \beta_0 = 0$. Its marginal density $\rho_t = \mathrm{Law}(I_t)$ satisfies the continuity equation $\partial_t \rho_t + \nabla \cdot (\rho_t b_t) = 0$ with velocity

$$b_t(x) = \mathbb{E}[\dot{I}_t \mid I_t = x] = \mathbb{E}[\dot{\alpha}_t x_0 + \dot{\beta}_t x_1 \mid I_t = x]. \tag{2}$$

Flow matching [4, 7] learns this velocity by solving the regression problem

$$b = \arg\min_{\hat{b}} \mathcal{L}_{\mathrm{FM}}(\hat{b}), \qquad \mathcal{L}_{\mathrm{FM}}(\hat{b}) = \int_0^1 \mathbb{E}\Big[\|\hat{b}_t(I_t) - \dot{I}_t\|^2\Big]\, dt. \tag{3}$$

Samples are then generated by integrating the flow

$$\dot{x}_t = b_t(x_t), \qquad x_0 \sim \rho_0, \tag{4}$$

from $t = 0$ to $t = 1$, which gives $x_1 \sim \rho_1$. Throughout, we use the linear choice $\alpha_t = 1 - t$ and $\beta_t = t$, for which $I_t = (1-t)x_0 + tx_1$ and $\dot{I}_t = x_1 - x_0$.

Flow maps. Solving (4) numerically typically requires many network evaluations. To avoid this cost, recent work learns the *flow map* $X : [0,1]^2 \times \mathbb{R}^d \to \mathbb{R}^d$, the solution operator of (4) [8–10, 16]. By definition, the flow map satisfies the jump condition along trajectories of (4),

$$X_{s,t}(x_s) = x_t, \qquad (s,t) \in [0,1]^2. \tag{5}$$

Given $X_{s,t}$, a sample from ρ_1 can be generated in a single evaluation $x_1 = X_{0,1}(x_0)$. More generally, the semigroup property

$$X_{s,t} = X_{u,t} \circ X_{s,u}, \qquad s,u,t \in [0,1], \tag{6}$$

enables multi-step sampling on an arbitrary time grid, trading additional evaluations for accuracy. There are many ways to learn the flow map, either directly from data or by distillation from a pretrained velocity field [8, 9]. This framework contains consistency models [10] and MeanFlow [16] as special cases. Because b_t depends on time, the flow map must be indexed by both a start time s and an end time t.

Equilibrium matching. Equilibrium Matching (EqM) [11] instead learns a single time-independent field $g : \mathbb{R}^d \to \mathbb{R}^d$. Intuitively, EqM drops the time input from flow matching and rescales the interpolant velocity $\dot{I}_t = x_1 - x_0$ by a factor $c(t)$ that vanishes at $t = 1$, so that the target is zero on the data. In our sign convention, its objective is

$$g = \arg\min_{\hat{g}} \mathcal{L}_{\mathrm{EqM}}(\hat{g}), \qquad \mathcal{L}_{\mathrm{EqM}}(\hat{g}) = \int_0^1 \mathbb{E}\big[\|\hat{g}(I_t) - c(t)(x_1 - x_0)\|^2\big]\, dt, \tag{7}$$

where $c : [0,1] \to [0,\infty)$ satisfies $c(1) = 0$. This differs from Wang & Du [11] by an overall sign. Samples are then generated by integrating the resulting autonomous flow $\dot{x}_\tau = g(x_\tau)$ from $x_0 \sim \rho_0$ until it settles near the data, where g vanishes. In Section 3, we show that this construction arises as a particular case of our framework: the rescaling c corresponds to a choice of clock for an infinite-horizon interpolant, which also determines the correct weighting of the loss.

3 Equilibrium Flow Maps

In this section, we introduce an infinite-horizon stochastic interpolant that defines an autonomous equilibrium field, and we show that this field yields correct source-to-target transport in the infinite-time limit. We then prove that the induced dynamics define a one-parameter flow map. To make this infinite-time map learnable, we compactify it into a bounded family of solution operators that can be learned directly and efficiently.

3.1 Equilibrium flows on an infinite horizon

Throughout, we use $t \in [0,1]$ to denote a bounded interpolation time and $\tau \in [0,\infty)$ to denote the infinite-horizon autonomous time. We first extend the stochastic interpolant of Section 2 to an infinite horizon. To do so, we map autonomous time to interpolation time with a clock.

Definition 3.1 (Clock)**.** *A* clock *is a continuously differentiable function* $r : [0,\infty) \to [0,1)$ *satisfying*

$$r_0 = 0, \qquad \dot{r}_0 = 1, \qquad \dot{r}_\tau > 0, \qquad r_\tau \to 1 \quad \textit{as } \tau \to \infty.$$

In particular, a clock is a bijection from $[0, \infty)$ onto $[0, 1)$. For independent $x_0 \sim \rho_0$ and $x_1 \sim \rho_1$, we define the infinite-horizon interpolant and its velocity as

$$I_\tau = x_0 + r_\tau(x_1 - x_0), \qquad \dot{I}_\tau = \dot{r}_\tau(x_1 - x_0). \tag{8}$$

Thus $I_0 = x_0$ and $I_\tau \to x_1$, while r_τ determines how quickly the interpolant traverses the fixed noise–data segment. A canonical example is the exponential clock $r_\tau = 1 - e^{-\tau}$, for which $I_\tau = e^{-\tau}x_0 + (1 - e^{-\tau})x_1$ and $\dot{I}_\tau = e^{-\tau}(x_1 - x_0)$, so the interpolant relaxes exponentially toward the data point x_1.

We propose to learn an autonomous field by direct regression onto the velocity of this infinite-horizon interpolant. We call this approach *autonomous flow matching* (AFM), with objective

$$b = \arg\min_{\hat{b}} \mathcal{L}_{\mathrm{AFM}}(\hat{b}), \qquad \mathcal{L}_{\mathrm{AFM}}(\hat{b}) = \int_0^\infty \mathbb{E}\Big[\|\hat{b}(I_\tau) - \dot{I}_\tau\|^2\Big]\, d\tau. \tag{9}$$

Because b plays the same role as the flow matching velocity b_t in (4), we use the same notation without the time subscript. The field b defines the autonomous flow

$$\dot{x}_\tau = b(x_\tau), \qquad x_0 \sim \rho_0, \qquad \tau \in [0, \infty), \tag{10}$$

in which the same vector field is used at every elapsed time. The following result shows that this flow transports ρ_0 to ρ_1 in the infinite-time limit.

Theorem 3.2 (Infinite-time transport)**.** *Let x_τ solve the autonomous flow (10). Under suitable regularity assumptions, x_τ converges almost surely as $\tau \to \infty$ to a limit x_∞ on the data manifold $\mathcal{M}$ with $x_\infty \sim \rho_1$.*

The regularity assumptions and the proof are given in Appendix A.2. We show that the minimizer b is unique almost everywhere with respect to the occupation measure $\mu(y) = \int_0^\infty \rho_\tau(y)\, d\tau$ of the infinite-horizon interpolant and give its explicit population representation. After establishing the regularity and asymptotic properties of the infinite-horizon field, we apply the flux-based transport criterion of Lee et al. [14] to show that its terminal point is distributed as ρ_1.

3.2 Equilibrium flow maps

In this section, we introduce flow maps for the autonomous dynamics (10). Our key observation is that much of the flow map machinery from standard flow matching carries over to autonomous fields, with the additional simplification that the resulting map depends only on a single time. The flow map $X : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d$ of (10) sends an initial point x to the state $X_\tau(x)$ reached after evolving the dynamics for duration τ. It can be written explicitly as

$$X_\tau(x) = x + \int_0^\tau b(X_u(x))\, du, \qquad X_0(x) = x. \tag{11}$$

Analogous to the jump condition (5), along any trajectory of (10) the flow map satisfies

$$X_\tau(x_s) = x_{s+\tau}, \qquad s, \tau \geq 0. \tag{12}$$

As a direct consequence, advancing by τ_1 and then by τ_2 is the same as advancing by $\tau_1 + \tau_2$, which gives the semigroup property

$$X_{\tau_1+\tau_2} = X_{\tau_2} \circ X_{\tau_1}, \qquad \tau_1, \tau_2 \geq 0. \tag{13}$$

Because the autonomous flow (10) reaches the data only in infinite time, learning this map directly would require estimation over the unbounded horizon $\tau \in [0, \infty)$. To avoid this difficulty, we observe that the clock r of Definition 3.1 can be used to compactify time, mapping the infinite-time autonomous dynamics onto the finite interval $\sigma \in [0, 1]$, with the terminal point $\tau = \infty$ sent to $\sigma = 1$. Concretely, we change variables to the bounded duration $\sigma = r_\tau \in [0, 1)$ and define the compactified flow map $\Phi : [0, 1) \times \mathbb{R}^d \to \mathbb{R}^d$ by

$$\Phi_\sigma = X_{r^{-1}(\sigma)}, \qquad X_\tau = \Phi_{r_\tau}. \tag{14}$$

This gives an exact correspondence between infinite autonomous time and compactified time without changing the underlying flow.

Just as for the flow map of a time-dependent field [8, 9], the compactified map satisfies three identities, each of which we show in Section 4 leads to an efficient learning algorithm.

Theorem 3.3 (Compactified flow map)**.** *Let r be a clock as in Definition 3.1 with clock speed $c(\sigma) = d\sigma/d\tau$ along $\sigma = r_\tau$, and let Φ denote the compactified flow map (14). Under standard regularity assumptions, Φ satisfies the following conditions:*

(i) (Lagrangian condition): for all $\sigma \in [0, 1)$ and $x \in \mathbb{R}^d$,

$$c(\sigma)\partial_\sigma \Phi_\sigma(x) = b(\Phi_\sigma(x)). \tag{15}$$

(ii) (Eulerian condition): for all $\sigma \in [0, 1)$ and $x \in \mathbb{R}^d$,

$$c(\sigma)\partial_\sigma \Phi_\sigma(x) = \nabla_x \Phi_\sigma(x) b(x). \tag{16}$$

(iii) (Semigroup condition): for all $\tau_1, \tau_2 \geq 0$,

$$\Phi_{r_{\tau_1+\tau_2}} = \Phi_{r_{\tau_2}} \circ \Phi_{r_{\tau_1}}. \tag{17}$$

The regularity assumptions and the proof are given in Appendix A.4.

The clock speed $c : [0, 1) \to (0, \infty)$ is the rate at which compactified time advances per unit of autonomous time; explicitly, $c(\sigma) = \dot{r}\big(r^{-1}(\sigma)\big)$. As the clock slows near $\sigma = 1$, a bounded interval of σ covers an unbounded stretch of τ. For the exponential clock $r_\tau = 1 - e^{-\tau}$, we have $\dot{r}_\tau = e^{-\tau} = 1 - r_\tau$, so the clock speed is simply $c(\sigma) = 1 - \sigma$.

The Lagrangian and Eulerian conditions characterize the compactified map as the solution of an ODE along trajectories or of a transport PDE, each describing transport along the autonomous flow in compactified time. The semigroup condition states that any two jumps can be replaced by a single jump, and specifies how to do so in compactified time: add the two durations in autonomous time, then map the sum back through the clock.

Conversely, as we show in Proposition A.4, each of these conditions uniquely determines Φ. This can be used to derive self-distillation algorithms for direct training. Critical to this is the following tangent condition, which can be derived by evaluating the Lagrangian condition at $\sigma = 0$.

Corollary 3.4 (Tangent condition)**.** *Under the assumptions of Theorem 3.3,*

$$\partial_\sigma \Phi_\sigma(x)|_{\sigma=0} = b(x). \tag{18}$$

In particular, if $\Phi_\sigma(x) = x + \sigma v_\sigma(x)$ with v continuous at $\sigma = 0$, then $v_0 = b$.

The proof is given in Appendix A.4.

Compactified sampling. The terminal map $\Phi_1(x) := \lim_{\tau\to\infty} X_\tau(x)$ sends ρ_0-almost every starting point directly to its equilibrium x_∞, so under Theorem 3.2 a single evaluation $\Phi_1(x_0)$ with $x_0 \sim \rho_0$ produces a sample from ρ_1. More generally, for any durations $\sigma_1, \ldots, \sigma_K \in [0, 1]$ with $\sigma_K = 1$, EqFM iterates

$$x^{(k)} = \Phi_{\sigma_k}(x^{(k-1)}), \qquad k = 1, \ldots, K, \qquad x^{(0)} \sim \rho_0. \tag{19}$$

By the semigroup condition (17), each intermediate step advances along the same autonomous trajectory, so additional steps refine a sample rather than restart it. In practice, we use $K - 1$ finite durations followed by the terminal map Φ_1, where $K = 1$ recovers one-step generation. For the exact map, the state after Φ_1 is already at equilibrium, so further evolution leaves it unchanged.

The sampler (19) naturally admits a constant-duration refinement loop in which the same finite-duration map is applied repeatedly to increase inference-time computation. This parallels looped, equilibrium, and recurrent-depth architectures that repeatedly apply a shared transformation to deepen computation [17–21]. In EqFM, however, the shared transformation has explicit generative semantics:

each iteration advances the sample further along the same autonomous flow toward equilibrium. Thus, iterative refinement and generative transport are realized by the same operator. Recent work has explored related fixed-point and iterative mechanisms for flow models [22, 23].

Relation to Equilibrium Matching. EqM [11] learns an autonomous field by regressing onto the rescaled velocity $c(t)(x_1 - x_0)$ uniformly over interpolation time, as in (7). In our framework, changing variables to interpolation time $t = r_\tau$ gives $I_t = (1-t)x_0 + tx_1$, $\dot{I}_\tau = c(t)(x_1 - x_0)$, and $dt = c(t)\, d\tau$ for any clock, which turns the AFM objective (9) into

$$\mathcal{L}_{\mathrm{AFM}}(\hat{b}) = \int_0^1 \mathbb{E}\Big[\|\hat{b}(I_t) - c(t)(x_1 - x_0)\|^2\Big] \, \tfrac{dt}{c(t)}. \tag{20}$$

EqM thus has the same pointwise target as AFM, but omits the change-of-variables weight $1/c(t)$. This difference can bias the learned field, as independently observed by Lee et al. [14]. Whereas EqM learns only the field and samples by integrating it, EqFM also learns its compactified flow map.

Relation to Beckmann Transport Models. Concurrently, BTM [14] learns the terminal map of an autonomous flow directly, by enforcing a stationarity condition along the field. In our framework, the semigroup property implies that the terminal map is constant along trajectories. Indeed, for any x for which $\Phi_1(x) = \lim_{\tau\to\infty} X_\tau(x)$ exists and for every $s \geq 0$,

$$\Phi_1(X_s(x)) = \lim_{\tau\to\infty} X_\tau(X_s(x)) = \lim_{\tau\to\infty} X_{\tau+s}(x) = \Phi_1(x).$$

At any such x where Φ_1 is differentiable, differentiating this identity with respect to s at $s = 0$ gives

$$\nabla_x \Phi_1(x) b(x) = 0, \tag{21}$$

which is exactly the BTM condition. Thus, BTM can be viewed as consistency training [10] for learning the endpoint map Φ_1, while EqFM learns the full flow map, which enables the multistep refinement (19). In our experiments, we show that learning the full flow map is significantly more stable and leads to much faster training than enforcing the terminal condition alone, and that multistep sampling with the full flow map significantly improves sample quality. This mirrors how, for time-dependent fields, learning the full flow map is more stable than the original consistency training objective of Song et al. [10] [8, 9, 16, 24–26].

Section 4 turns the interior characterizations into practical training objectives.

4 Method

In this section, we use Theorem 3.3 and Corollary 3.4 to derive practical objectives for learning the autonomous field and its compactified flow map, either directly from data or by distillation from a pretrained field.

Learning the autonomous field. We learn the autonomous field by minimizing the AFM objective in interpolation time (20). Because the weight $1/c(t)$ is not integrable near $t = 1$ (reflecting the infinite occupation of the interpolant as $\tau \to \infty$), we truncate the interpolation time at $1 - \varepsilon$ and set

$$\hat{b} = \arg\min_{\hat{b}} \mathcal{L}_b(\hat{b}), \qquad \mathcal{L}_b(\hat{b}) = \mathbb{E}\Big[\tfrac{1}{c(t)}\|\hat{b}(I_t) - c(t)(x_1 - x_0)\|^2\Big], \tag{22}$$

where $x_0 \sim \rho_0$, $x_1 \sim \rho_1$, $t \sim \mathrm{Unif}[0, 1-\varepsilon]$, and $I_t = (1-t)x_0 + tx_1$.

Parametrization and tangent condition. We parametrize the compactified flow map as

$$\hat{\Phi}_\sigma(x) = x + \sigma \hat{v}_\sigma(x), \tag{23}$$

which enforces $\hat{\Phi}_0(x) = x$. By Corollary 3.4, the exact map satisfies $v_0 = b$, so the network evaluated at zero duration is itself an autonomous field. We train it with a diagonal loss $\mathcal{L}_{\mathrm{diag}}$, which is either the field objective (22) with $\hat{b}$ replaced by $\hat{v}_0$, when training from data, or the distillation loss

$$\mathcal{L}_{\mathrm{diag}}(\hat{v}) = \mathbb{E}\Big[\|\hat{v}_0(I_t) - \hat{b}(I_t)\|^2\Big], \tag{24}$$

when a pretrained field $\hat{b}$ is available.

Off-diagonal loss. By Proposition A.4, the tangent condition together with any one of the three conditions of Theorem 3.3 uniquely determines Φ. Following Boffi et al. [9], we therefore combine $\mathcal{L}_{\text{diag}}$ with a loss $\mathcal{L}_{\text{off}}$ that enforces one of these conditions at positive durations,

$$\mathcal{L}(\hat{v}) = \mathcal{L}_{\text{diag}}(\hat{v}) + \alpha\,\mathcal{L}_{\text{off}}(\hat{v}), \qquad \mathcal{L}_{\text{off}} \in \{\mathcal{L}_{\text{Lag}}, \mathcal{L}_{\text{Eul}}, \mathcal{L}_{\text{Sg}}\}. \tag{25}$$

In each off-diagonal loss, b is replaced by the frozen field $\hat{b}$ for distillation, or by the model's own diagonal $\hat{v}_0$ when training from data. The semigroup loss directly penalizes the semigroup condition (17),

$$\mathcal{L}_{\text{Sg}}(\hat{v}) = \mathbb{E}\Big[\big\|\hat{\Phi}_{r_{\tau_1+\tau_2}}(I_t) - \text{sg}\left[\hat{\Phi}_{r_{\tau_2}}(\hat{\Phi}_{r_{\tau_1}}(I_t))\right]\big\|^2\Big], \tag{26}$$

over sampled durations $\tau_1, \tau_2 \geq 0$.

For the Lagrangian and Eulerian conditions, substituting the parametrization (23) into (15) and (16) gives

$$c(\sigma)\big(v_\sigma(x) + \sigma\partial_\sigma v_\sigma(x)\big) = F_\sigma(x), \tag{27}$$

where the right-hand side is

$$F_\sigma^{\text{Lag}}(x) = b(\Phi_\sigma(x)) \qquad \text{or} \qquad F_\sigma^{\text{Eul}}(x) = \nabla_x \Phi_\sigma(x) b(x). \tag{28}$$

The Lagrangian form evaluates the field at the evolved state, while the Eulerian form transports it through the map Jacobian; both agree for the exact map.

Rearranging (27) as a regression for $c(\sigma)v_\sigma$, as in MeanFlow [16], loses its signal near $\sigma = 1$, where $c(\sigma) \to 0$. We therefore add $(1 - c(\sigma))v_\sigma$ to both sides, which yields the fixed-point target

$$v_\sigma^{\text{tgt}}(x) = \hat{F}_\sigma(x) + (1 - c(\sigma))\hat{v}_\sigma(x) - \sigma c(\sigma)\partial_\sigma \hat{v}_\sigma(x), \tag{29}$$

where $\hat{F}_\sigma$ is (28) evaluated with $\hat{\Phi}$ and $\hat{b}$ (or $\hat{v}_0$). The Lagrangian and Eulerian losses then regress onto this target with a stop-gradient,

$$\mathcal{L}_{\text{Lag/Eul}}(\hat{v}) = \mathbb{E}\big[\|\hat{v}_\sigma(I_t) - \text{sg}\left[v_\sigma^{\text{tgt}}(I_t)\right]\|^2\big]. \tag{30}$$

Two design choices matter in practice. The stop-gradient avoids backpropagating through Jacobian–vector products, which reduced training cost and improved stability in our experiments. The added term $(1 - c(\sigma))\hat{v}_\sigma$ provides a direct signal near $\sigma = 1$, which improved one-step generation. Algorithms 1 and 2 summarize the procedure, and implementation details are given in Appendix C.

Algorithm 1 Field training

Require: Field $\hat{b}$; clock speed c; cutoff ε
  while not converged **do**
    Draw $x_0 \sim \rho_0$, $x_1 \sim \rho_1$, and $t \sim \text{Unif}[0, 1-\varepsilon]$
    $I_t \leftarrow (1-t)x_0 + tx_1$
    $L \leftarrow c(t)^{-1}\|\hat{b}(I_t) - c(t)(x_1 - x_0)\|^2$
    Update $\hat{b}$ by a gradient step on L
  end while

Algorithm 2 EqFM distillation

Require: Frozen field $\hat{b}$; map $\hat{\Phi}_\sigma(x) = x + \sigma\hat{v}_\sigma(x)$; losses $\mathcal{L}_{\text{diag}}, \mathcal{L}_{\text{off}}$; weight α
  while not converged **do**
    Draw $x_0 \sim \rho_0$, $x_1 \sim \rho_1$, and $t \sim \text{Unif}[0, 1]$
    $I_t \leftarrow (1-t)x_0 + tx_1$
    $L_{\text{diag}} \leftarrow \mathcal{L}_{\text{diag}}(\hat{\Phi}, \hat{b}; I_t)$
    $L_{\text{off}} \leftarrow \mathcal{L}_{\text{off}}(\hat{\Phi}, \hat{b}; I_t)$
    $L \leftarrow L_{\text{diag}} + \alpha L_{\text{off}}$
    Update $\hat{\Phi}$ by a gradient step on L
  end while

Classifier-free guidance. During sampling, classifier-free guidance [27] typically combines conditional and unconditional velocities evaluated at the same time, so that both describe the current state at the same noise level. This becomes challenging for autonomous fields, which have no time input: each field implicitly infers the noise level of a state, and conditioning can change this inference, leading to a combination of fields at different effective noise levels. Let c denote a class label and $\varnothing$ a null token, so that $b(x; c)$ and $b(x; \varnothing)$ are the conditional and unconditional autonomous fields. Naive CFG in the autonomous setting with scale w is given by

$$b_{\text{cfg}}(x; c) = w\, b(x; c) - (w - 1)\, b(x; \varnothing). \tag{31}$$

Method	1 call	2 calls	4 calls	8 calls
MeanFlow	**15.81**	11.92	11.99	11.66
Eulerian	25.70	**10.80**	**8.90**	**8.50**
Lagrangian	86.60	21.10	12.90	10.70
Semigroup	50.33	23.48	16.35	13.19

Table 1: **Compute-matched CIFAR-10 training from scratch.** FID-50k at 1, 2, 4, and 8 network calls. All methods are trained for 6 GPU-hours.

w	Naive	Norm	Perp.
1.5	2.00	2.00	2.01
5	14.69	12.35	11.44
10	43.34	10.77	11.90

Table 2: **ImageNet-256 classifier-free guidance.** FID-50k of the autonomous field teacher for each guidance rule.

A state that looks nearly denoised to the conditional field, where $b(x;c)$ is small, may look noisy to the unconditional field, where $b(x;\varnothing)$ is large. The subtraction in (31) can then cancel or even reverse the pull toward the conditional data as w grows (Appendix B).

To avoid this, we propose to modify the unconditional term so that guidance never weakens the pull toward the conditional data. Writing $\hat{e} = b(x;c)/\|b(x;c)\|$ for the conditional direction, we require the guided field b_w to satisfy

$$\hat{e}^\top b_w(x;c) \geq \|b(x;c)\| \qquad \text{for all } w \geq 1, \tag{32}$$

so that it pulls toward the conditional data at least as strongly as the conditional field alone. This requirement leads to two natural options.

The first rescales the unconditional field, which we call *norm matching*,

$$b_{\text{norm}}(x;c) = w\, b(x;c) - (w-1)\,\tfrac{\|b(x;c)\|}{\|b(x;\varnothing)\|}\, b(x;\varnothing). \tag{33}$$

Since $\hat{e}^\top b(x;\varnothing) \leq \|b(x;\varnothing)\|$, norm matching satisfies (32), retaining at least the original conditional pull. The second removes the component of the unconditional field along the conditional one. Letting

$$b^\perp(x;\varnothing) = b(x;\varnothing) - \big(\hat{e}^\top b(x;\varnothing)\big)\,\hat{e}$$

denote the part of the unconditional field orthogonal to $\hat{e}$, we define *perpendicular guidance*,

$$b_{\text{perp}}(x;c) = b(x;c) - (w-1)\, b^\perp(x;\varnothing). \tag{34}$$

Perpendicular guidance satisfies (32) with equality, retaining the conditional pull exactly; it is related to projected CFG in diffusion models [28].

5 Experiments

We evaluate flow map generation on CIFAR-10 [29] and ImageNet-256 [30] with FID [31].

5.1 CIFAR-10 training from scratch

We train all methods from random initialization with the same unconditional ADM UNet [32], without a teacher, pretrained checkpoint, or guidance. Because the objectives have different per-update costs, we match training compute (6 GPU-hours) rather than the number of updates (Table 1). At matched compute, the Eulerian objective outperforms MeanFlow at 2, 4, and 8 calls and generally outperforms the other objectives. All three EqFM objectives improve with additional calls.

5.2 ImageNet-256 flow map distillation

Autonomous field teacher. We first train the autonomous field b for 480k steps on ImageNet-256. Sampling the resulting field with 250 Euler steps and CFG $w = 1.5$ gives FID-50k 2.00. We use this field as the teacher for EqFM distillation.

We distill this teacher into an EqFM student with the same 675M-parameter SiT-XL/2 [33] architecture using the Eulerian loss. The student is initialized from the teacher, trained for 120k steps, and evaluated with its EMA weights. Guidance is distilled into the student, so each jump costs one

Method	NFE	FID ↓
EqFM	1	4.280
EqFM	2	2.780
EqFM	4	2.248
EqFM	8	**2.086**
BTM [14]	1	17.58
BTM	2	17.04
BTM	5	16.53
Autonomous	8	42.557
Autonomous	500	2.004

Table 3: **ImageNet-256 generation.** EqFM uses the schedule described in the text; FID-50k uses balanced labels and ADM statistics. Bold: best few-call result.

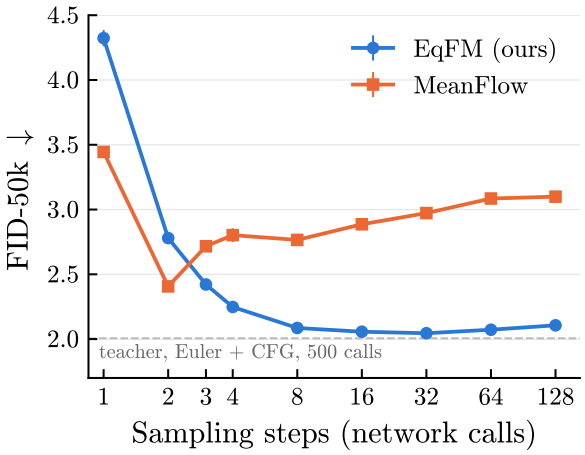


Figure 2: **FID versus network calls.** EqFM and MeanFlow-XL/2 (MeanFlow checkpoint from the AlphaFlow repository [34]); dashed line: 500-call teacher.

network call, whereas the teacher uses 500 (two per CFG step). For K calls, we split the noise level into K equal steps: under the exponential clock, a jump Φ_σ removes a fraction σ of the remaining noise, so we take $\sigma_k = 1/(K-k+1)$ for $k = 1, \dots, K$ in (19), which ends with the terminal map $\sigma_K = 1$.

With this sampler, EqFM improves from FID 4.28 at one call to 2.09 at eight calls, approaching the performance of the 500-call teacher, whereas four-step Euler sampling of the teacher with CFG at the same eight-call budget reaches 42.56. The one-call result is also competitive with MeanFlow while substantially outperforming BTM, achieving FID 4.28 compared with 17.58. We also compare with MeanFlow-XL/2 (MeanFlow checkpoint from the AlphaFlow repository [34]) under the same protocol. Their training budgets differ, so this probes behavior across calls rather than training efficiency. We find that EqFM exhibits better inference scaling than MeanFlow, reaching lower FID at three or more calls and its best FID of 2.05 at 32 calls. MeanFlow reaches its best FID of 2.41 at two calls and scores 3.10 at 128 calls.

Classifier-free guidance. We evaluate classifier-free guidance rules on the autonomous field teacher (250 Euler steps) on ImageNet-256 at several weights. At $w = 1.5$, the three rules perform similarly. At larger weights, naive CFG degrades while norm matching and perpendicular guidance retain lower FID (Table 2). We further isolate the restoring component through a smaller intervention in Appendix B, with same-noise qualitative comparisons shown in Figure 3.

6 Discussion and Limitations

Compactification makes the terminal map of equilibrium generation explicit: the same model gives finite-duration maps Φ_σ with $\sigma < 1$, which advance a sample along the flow, and the terminal map Φ_1, which sends it directly to equilibrium. Its central practical promise is to replace long autonomous integration with a small number of learned map evaluations. On ImageNet-256, eight calls with a uniform schedule recover the performance of the 500-call teacher. EqFM also overtakes MeanFlow at three calls and outperforms BTM at one call. On CIFAR-10, Eulerian EqFM outperforms MeanFlow at two to eight calls under matched training compute.

Several limitations remain. The ImageNet results come from a single training run, and EqFM and MeanFlow in the comparison across numbers of calls have different training budgets. Broader ablations should compare the Lagrangian, Eulerian, and semigroup losses at scale, as well as clock and step size choices.

ACKNOWLEDGMENTS

Omitted for double-blind review.

REPRODUCIBILITY STATEMENT

Section 4 and Algorithms 1 and 2 define the training objectives. Section 5 gives the architectures, the sampling schedule, and the evaluation protocol (FID-50k with balanced labels and ADM reference statistics). Appendix C lists the time sampling and the full training configuration for ImageNet-256. Appendix A states the assumptions and gives complete proofs of the theoretical results.

AI USE STATEMENT

We used large language models to assist with editing the text and the LaTeX source of this paper, including wording, formatting, and consistency checks. The method, theory, experiments, and results are the authors' own, and the authors reviewed and take full responsibility for all content.

REFERENCES

[1] Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. SE(3) diffusion model with application to protein backbone generation. In *International Conference on Machine Learning (ICML)*, 2023. URL https://arxiv.org/abs/2302.02277.

[2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020. URL https://arxiv.org/abs/2006.11239.

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

[4] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2210.02747.

[5] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J. Fleet. Video diffusion models. In *Advances in Neural Information Processing Systems*, 2022. URL https://arxiv.org/abs/2204.03458.

[6] Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T. Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. In *Advances in Neural Information Processing Systems*, 2024. URL https://arxiv.org/abs/2406.07524.

[7] Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209): 1–80, 2025. URL https://arxiv.org/abs/2303.08797.

[8] Nicholas M. Boffi, Michael S. Albergo, and Eric Vanden-Eijnden. Flow map matching with stochastic interpolants: A mathematical framework for consistency models. *Transactions on Machine Learning Research*, 2025. URL https://arxiv.org/abs/2406.07507.

[9] Nicholas M. Boffi, Michael S. Albergo, and Eric Vanden-Eijnden. How to build a consistency model: Learning flow maps via self-distillation. In *Advances in Neural Information Processing Systems*, 2025. URL https://arxiv.org/abs/2505.18825.

[10] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *International Conference on Machine Learning*, 2023. URL https://arxiv.org/abs/2303.01469.

[11] Runqian Wang and Yilun Du. Equilibrium matching: Generative modeling with implicit energy-based models. *arXiv preprint arXiv:2510.02300*, 2025. URL https://arxiv.org/abs/2510.02300.

[12] Qiao Sun, Zhicheng Jiang, Hanhong Zhao, and Kaiming He. Is noise conditioning necessary for denoising generative models? In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 57469–57502. PMLR, 2025. URL https://proceedings.mlr.press/v267/sun25g.html.

[13] Mojtaba Sahraee-Ardakan, Mauricio Delbracio, and Peyman Milanfar. The geometry of noise: Why diffusion models don't need noise conditioning. *arXiv preprint arXiv:2602.18428*, 2026. URL https://arxiv.org/abs/2602.18428.

[14] Cheuk-Kit Lee, Florentin Coeurdoux, Yuyuan Chen, Sophia Tang, Peter Potaptchik, Yilun Du, Michael Samuel Albergo, and Eric Vanden-Eijnden. Beckmann transport models: From autonomous flows to one-step maps. *arXiv preprint arXiv:2608.01692*, 2026. URL https://arxiv.org/abs/2608.01692.

[15] Zahra Kadkhodaie, Aram-Alexandre Pooladian, Sinho Chewi, and Eero P. Simoncelli. Blind denoising diffusion models and the blessings of dimensionality. *arXiv preprint arXiv:2602.09639*, 2026. URL https://arxiv.org/abs/2602.09639.

[16] Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *Advances in Neural Information Processing Systems*, 2025. URL https://arxiv.org/abs/2505.13447.

[17] Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, and Łukasz Kaiser. Universal transformers. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1807.03819.

[18] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems*, 2019. URL https://arxiv.org/abs/1909.01377.

[19] Angeliki Giannou, Shashank Rajput, Jy-Yong Sohn, Kangwook Lee, Jason D. Lee, and Dimitris Papailiopoulos. Looped transformers as programmable computers. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 11398–11442. PMLR, 2023. URL https://proceedings.mlr.press/v202/giannou23a.html.

[20] Liu Yang, Kangwook Lee, Robert D. Nowak, and Dimitris Papailiopoulos. Looped transformers are better at learning learning algorithms. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2311.12424.

[21] Jonas Geiping, Sean McLeish, Neel Jain, John Kirchenbauer, Siddharth Singh, Brian R. Bartoldson, Bhavya Kailkhura, Abhinav Bhatele, and Tom Goldstein. Scaling up test-time compute with latent reasoning: A recurrent depth approach. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://arxiv.org/abs/2502.05171.

[22] Jaehoon Yoo, Wonjung Kim, Floor Eijkelboom, Chanhyuk Lee, Nicholas M. Boffi, Seunghoon Hong, and Jinwoo Kim. Self-conditioned flow map language models via fixed-point flows. *arXiv preprint arXiv:2607.00714*, 2026. URL https://arxiv.org/abs/2607.00714.

[23] Alec Helbling, Andrey Bryutkin, Mauro Martino, Duen Horng Chau, Nima Dehmamy, and Hendrik Strobelt. Flow reasoning models: Turning flows into efficient recurrent reasoners. *arXiv preprint arXiv:2606.29150*, 2026. URL https://arxiv.org/abs/2606.29150.

[24] Yang Song and Prafulla Dhariwal. Improved techniques for training consistency models. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2310.14189.

[25] Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, and Stefano Ermon. Consistency trajectory models: Learning probability flow ODE trajectory of diffusion. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2310.02279.

[26] Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your flow: Scaling continuous-time flow map distillation. *arXiv preprint arXiv:2506.14603*, 2025. URL https://arxiv.org/abs/2506.14603.

[27] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. URL https://arxiv.org/abs/2207.12598.

[28] Seyedmorteza Sadat, Otmar Hilliges, and Romann M. Weber. Eliminating oversaturation and artifacts of high guidance scales in diffusion models. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2410.02416.

[29] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. URL https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

[30] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009. URL https://www.image-net.org/static_files/papers/imagenet_cvpr09.pdf.

[31] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in Neural Information Processing Systems*, 2017. URL https://arxiv.org/abs/1706.08500.

[32] Prafulla Dhariwal and Alex Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, 2021. URL https://arxiv.org/abs/2105.05233.

[33] Nanye Ma, Mark Goldstein, Michael S. Albergo, Nicholas M. Boffi, Eric Vanden-Eijnden, and Saining Xie. SiT: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *European Conference on Computer Vision (ECCV)*, 2024. URL https://arxiv.org/abs/2401.08740.

[34] Huijie Zhang, Aliaksandr Siarohin, Willi Menapace, Michael Vasilkovsky, Sergey Tulyakov, Qing Qu, and Ivan Skorokhodov. AlphaFlow: Understanding and improving MeanFlow models. *arXiv preprint arXiv:2510.20771*, 2025. URL https://arxiv.org/abs/2510.20771.

[35] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019. URL https://arxiv.org/abs/1711.05101.

[36] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. URL https://arxiv.org/abs/1512.03385.

A ASSUMPTIONS AND SUPPORTING RESULTS

This appendix proves the population characterization and transport result used in Section 3, and then records the standard flow identities used for the compactified flow map.

A.1 SCHEDULES AND NOTATION

Let r be a clock as in Definition 3.1. Since $r : [0, \infty) \to [0, 1)$ is strictly increasing, it has an inverse $r^{-1} : [0, 1) \to [0, \infty)$. Define the compactified clock speed

$$c(t) := \dot{r}_{r^{-1}(t)}, \qquad t \in [0, 1). \tag{35}$$

Thus, along $t = r_\tau$, $dt = \dot{r}_\tau \, d\tau = c(t) \, d\tau, \qquad d\tau = \frac{dt}{c(t)}$. Integrating the latter identity gives

$$r^{-1}(t) = \int_0^t \frac{du}{c(u)}. \tag{36}$$

Since $r^{-1}(t) \to \infty$ as $t \uparrow 1$, $\int_0^1 \frac{dt}{c(t)} = \infty$.

For independent endpoint samples (x_0, x_1), recall the infinite-horizon interpolant

$$I_\tau = (1 - r_\tau)x_0 + r_\tau x_1, \qquad \dot{I}_\tau = \dot{r}_\tau(x_1 - x_0).$$

Let ρ_τ denote $\mathrm{Law}(I_\tau)$ whenever it exists.

We also use the same straight paths parametrized by bounded interpolation time $t \in [0, 1]$; as in Section 3, the index (t or τ) indicates which parametrization is meant:

$$I_t := (1 - t)x_0 + tx_1, \qquad t \in [0, 1]. \tag{37}$$

For $t < 1$, let ρ_t denote the density of I_t and define the usual conditional velocity

$$b_t(x) := \mathbb{E}[x_1 - x_0 \mid I_t = x]. \tag{38}$$

Because the infinite-horizon interpolant at autonomous time τ equals the bounded interpolant at interpolation time $t = r_\tau$, we have

$$\rho_\tau(x) = \rho_{r_\tau}(x). \tag{39}$$

Moreover,

$$\begin{aligned} \mathbb{E}[\dot{I}_\tau \mid I_\tau = x] &= \dot{r}_\tau \mathbb{E}[x_1 - x_0 \mid I_{r_\tau} = x] \\ &= c(r_\tau) b_{r_\tau}(x). \end{aligned} \tag{40}$$

Define the infinite-time occupation density

$$\mu(x) := \int_0^\infty \rho_\tau(x) \, d\tau. \tag{41}$$

Using $t = r_\tau$ and $d\tau = dt/c(t)$,

$$\mu(x) = \int_0^1 \frac{\rho_t(x)}{c(t)} \, dt. \tag{42}$$

The same change of variables rewrites the AFM objective as

$$\begin{aligned} \mathcal{L}_{\mathrm{AFM}}(\hat{b}) &= \int_0^\infty \mathbb{E}\left[\|\hat{b}(I_\tau) - \dot{I}_\tau\|^2\right] d\tau \\ &= \int_0^1 \mathbb{E}\left[\|\hat{b}(I_t) - c(t)(x_1 - x_0)\|^2\right] \frac{dt}{c(t)}. \end{aligned} \tag{43}$$

Finally,

$$\int_0^\infty \dot{r}_\tau^2 \, d\tau = \int_0^1 c(t) \, dt. \tag{44}$$

Indeed, $dt = \dot{r}_\tau d\tau$ and $c(t) = \dot{r}_\tau$ along $t = r_\tau$.

A.2 Population regression and infinite-time transport

We now characterize the population AFM field and prove the transport statement of Theorem 3.2.

Assumptions. Let

$$x_0 \sim \rho_0 = \mathcal{N}(0, I_d), \qquad x_1 \sim \rho_1 = p_1 \, d\mathrm{vol}_{\mathcal{M}},$$

independently, where $\mathcal{M} \subset \mathbb{R}^d$ is a compact embedded C^3 manifold without boundary, of dimension k with $d - k \geq 2$, and $p_1 \in C^1(\mathcal{M}), \qquad p_1 > 0$.

We assume the finite clock-energy condition

$$\int_0^\infty \dot{r}_\tau^2 \, d\tau = \int_0^1 c(t) \, dt < \infty. \tag{45}$$

We also assume that the occupation density satisfies

$$\mu \in C^1(\mathbb{R}^d \setminus \mathcal{M}), \qquad \mu(x) < \infty \quad \text{for every } x \notin \mathcal{M}. \tag{46}$$

Under the Gaussian source, positivity of μ is automatic. Indeed, for every finite τ, conditioning on x_1 gives

$$I_\tau \mid x_1 \sim \mathcal{N}\left(r_\tau x_1, (1 - r_\tau)^2 I_d\right).$$

Since $r_\tau < 1$, this Gaussian has a strictly positive density on $\mathbb{R}^d$. Hence $\rho_\tau(x) > 0$ for every x and finite τ, and therefore

$$0 < \mu(x) < \infty, \qquad x \notin \mathcal{M}. \tag{47}$$

Reference autonomous transport. For the bounded linear interpolant (37), define

$$\nu(x) := \int_0^1 \rho_t(x) \, dt, \qquad j(x) := \int_0^1 \rho_t(x) b_t(x) \, dt, \tag{48}$$

and

$$\bar{b}(x) := \frac{j(x)}{\nu(x)}. \tag{49}$$

We use Proposition 1 of Lee et al. [14], specialized to the linear schedule $I_t = (1 - t)x_0 + tx_1$. Under the data assumptions above, their result states that, for ρ_0-almost every initial point x, there exists a finite $S(x) > 0$ such that the solution

$$\frac{dY_s}{ds} = \bar{b}(Y_s), \qquad Y_0 = x, \tag{50}$$

remains in $\mathbb{R}^d \setminus \mathcal{M}$ for $0 \leq s < S(x)$ and has a unique endpoint

$$T_{\mathrm{BTM}}(x) := \lim_{s \uparrow S(x)} Y_s \in \mathcal{M}. \tag{51}$$

Moreover,

$$(T_{\mathrm{BTM}})_{\#}\rho_0 = \rho_1. \tag{52}$$

The proof of their Proposition 1 also establishes

$$\nu > 0, \qquad j, \nu, \bar{b} \in C^1(\mathbb{R}^d \setminus \mathcal{M}). \tag{53}$$

Population AFM field. We first identify the minimizer of the infinite-horizon regression objective.

Because $\dot{I}_\tau = \dot{r}_\tau(x_1 - x_0)$, its total target energy is

$$\begin{aligned} A &:= \int_0^\infty \mathbb{E}\|\dot{I}_\tau\|^2 \, d\tau \\ &= \mathbb{E}\|x_1 - x_0\|^2 \int_0^\infty \dot{r}_\tau^2 \, d\tau < \infty. \end{aligned} \tag{54}$$

The first factor is finite because x_0 is Gaussian and x_1 is supported on the compact set $\mathcal{M}$; the second is finite by (45).

The integrated conditional current of the AFM interpolant is exactly j. Indeed, using (39) and (40),

$$\begin{aligned}\int_0^\infty \rho_\tau(x)\mathbb{E}[\dot{I}_\tau \mid I_\tau = x]\,d\tau &= \int_0^\infty \rho_{r_\tau}(x)\dot{r}_\tau b_{r_\tau}(x)\,d\tau \\ &= \int_0^1 \rho_t(x)b_t(x)\,dt \\ &= j(x),\end{aligned} \tag{55}$$

where the third line uses $dt = \dot{r}_\tau\,d\tau$.

It follows that the candidate population field is

$$b(x) := \frac{j(x)}{\mu(x)}, \qquad x \notin \mathcal{M}. \tag{56}$$

We next check that b belongs to the natural regression space $L^2(\mu)$. For fixed x, Cauchy–Schwarz gives

$$\begin{aligned}\|j(x)\|^2 &= \left\| \int_0^\infty \rho_\tau(x)\mathbb{E}[\dot{I}_\tau \mid I_\tau = x]\,d\tau \right\|^2 \\ &\le \left(\int_0^\infty \rho_\tau(x)\,d\tau\right)\left(\int_0^\infty \rho_\tau(x)\left\|\mathbb{E}[\dot{I}_\tau \mid I_\tau = x]\right\|^2 d\tau\right) \\ &= \mu(x)\int_0^\infty \rho_\tau(x)\left\|\mathbb{E}[\dot{I}_\tau \mid I_\tau = x]\right\|^2 d\tau.\end{aligned} \tag{57}$$

Conditional Jensen gives

$$\left\|\mathbb{E}[\dot{I}_\tau \mid I_\tau = x]\right\|^2 \le \mathbb{E}[\|\dot{I}_\tau\|^2 \mid I_\tau = x].$$

Therefore

$$\begin{aligned}\int_{\mathbb{R}^d} \frac{\|j(x)\|^2}{\mu(x)}\,dx &\le \int_0^\infty \int_{\mathbb{R}^d} \rho_\tau(x)\mathbb{E}[\|\dot{I}_\tau\|^2 \mid I_\tau = x]\,dx\,d\tau \\ &= \int_0^\infty \mathbb{E}\|\dot{I}_\tau\|^2\,d\tau \\ &= A < \infty.\end{aligned} \tag{58}$$

Since $b = j/\mu$,

$$\int_{\mathbb{R}^d} \mu(x)\|b(x)\|^2\,dx = \int_{\mathbb{R}^d} \frac{\|j(x)\|^2}{\mu(x)}\,dx < \infty.$$

Now let $\hat{b} \in L^2(\mu)$. The cross term is integrable because

$$\begin{aligned}&\int_0^\infty \mathbb{E}\left[|\hat{b}(I_\tau)\cdot \dot{I}_\tau|\right] d\tau \\ &\qquad \le \left(\int_{\mathbb{R}^d} \mu(x)\|\hat{b}(x)\|^2\,dx\right)^{1/2} A^{1/2} < \infty.\end{aligned} \tag{59}$$

Conditioning on $I_\tau = x$ and using (55),

$$\begin{aligned}\int_0^\infty \mathbb{E}[\hat{b}(I_\tau)\cdot \dot{I}_\tau]\,d\tau &= \int_{\mathbb{R}^d} \hat{b}(x)\cdot j(x)\,dx \\ &= \int_{\mathbb{R}^d} \mu(x)\hat{b}(x)\cdot b(x)\,dx.\end{aligned} \tag{60}$$

Similarly,

$$\int_0^\infty \mathbb{E}\|\hat{b}(I_\tau)\|^2\,d\tau = \int_{\mathbb{R}^d} \mu(x)\|\hat{b}(x)\|^2\,dx.$$

Expanding the regression objective therefore gives

$$\begin{aligned}\mathcal{L}_{\mathrm{AFM}}(\hat{b}) &= A + \int \mu(x)\|\hat{b}(x)\|^2\,dx - 2\int \mu(x)\hat{b}(x)\cdot b(x)\,dx \\ &= \mathcal{L}_{\mathrm{AFM}}(b) + \int_{\mathbb{R}^d} \mu(x)\|\hat{b}(x) - b(x)\|^2\,dx. \end{aligned} \tag{61}$$

Hence $b = j/\mu$ is the unique minimizer in $L^2(\mu)$, up to $\mu(x)\,dx$-almost-everywhere equality.

Relation to the reference field. Equations (49) and (56) give

$$b(x) = \frac{\nu(x)}{\mu(x)}\bar{b}(x). \tag{62}$$

Set

$$a(x) := \frac{\nu(x)}{\mu(x)}. \tag{63}$$

By (46), (47), and (53),

$$a \in C^1(\mathbb{R}^d \setminus \mathcal{M}; (0,\infty)). \tag{64}$$

Lemma A.1 (Positive time reparameterization). *Let $\Omega \subset \mathbb{R}^d$ be open, $f \in C^1(\Omega;\mathbb{R}^d)$, and $a \in C^1(\Omega;(0,\infty))$. Set $g = af$. Suppose $\frac{dY_s}{ds} = f(Y_s), \qquad Y_0 = x$, is defined on $[0,S)$ and $Y_s \to p \in \partial\Omega \qquad$ as $s \uparrow S$. Then the g-trajectory from x follows the same oriented curve and has the same endpoint p, possibly under a different time parametrization.*

Proof. Define $\tau(s) := \int_0^s \frac{du}{a(Y_u)}$. Since $a(Y_u) > 0$, $\tau(s)$ is strictly increasing. Let $\tau^\star := \lim_{s\uparrow S}\tau(s) \in (0,\infty]$ and let $s = \eta(\tau)$ denote the inverse map from $[0,\tau^\star)$ to $[0,S)$. Differentiating $\tau(\eta(\tau)) = \tau$ gives $\eta'(\tau) = a(Y_{\eta(\tau)})$. Define $X_\tau := Y_{\eta(\tau)}$. Then

$$\begin{aligned}\frac{dX_\tau}{d\tau} &= \left.\frac{dY_s}{ds}\right|_{s=\eta(\tau)} \eta'(\tau) \\ &= f(Y_{\eta(\tau)})a(Y_{\eta(\tau)}) \\ &= g(X_\tau). \end{aligned} \tag{65}$$

Thus X is a g-trajectory from the same initial point. Since $g \in C^1$, the ODE is locally unique, so this is the g-trajectory from x.

Because η is increasing and maps $[0,\tau^\star)$ onto $[0,S)$, X and Y visit exactly the same points in the same order. Finally, $\eta(\tau) \uparrow S \qquad$ as $\tau \uparrow \tau^\star$, and therefore

$$\lim_{\tau\uparrow\tau^\star} X_\tau = \lim_{\tau\uparrow\tau^\star} Y_{\eta(\tau)} = \lim_{s\uparrow S} Y_s = p.$$

□

Proposition A.2 (Terminal transport). *Under the assumptions above, for ρ_0-almost every initial point x, the AFM field b has the same terminal point as the reference field $\bar{b}$. If $T(x)$ denotes this endpoint, then*

$$T_\# \rho_0 = \rho_1. \tag{66}$$

Proof. Fix x outside the ρ_0-null exceptional set of Proposition 1 of Lee et al. [14]. Its reference trajectory (50) satisfies

$$Y_s \longrightarrow T_{\mathrm{BTM}}(x) \in \mathcal{M} \qquad \text{as } s \uparrow S(x).$$

On $\Omega := \mathbb{R}^d \setminus \mathcal{M}$, equation (62) gives $b = a\bar{b}$ with $a > 0$ and $a \in C^1(\Omega)$. Applying Lemma A.1 with $f = \bar{b}, \qquad g = b$ shows that the AFM trajectory follows exactly the same oriented curve. Hence $T(x) = T_{\mathrm{BTM}}(x)$ for ρ_0-almost every x. Using (52), $T_\#\rho_0 = (T_{\mathrm{BTM}})_\#\rho_0 = \rho_1$. □

Infinite autonomous time. To obtain the precise infinite-time statement of Theorem 3.2, assume in addition that the AFM ODE

$$\dot{X}_\tau = b(X_\tau) \tag{67}$$

is forward complete on $\Omega = \mathbb{R}^d \setminus \mathcal{M}$: for ρ_0-almost every initial point, its solution exists and remains in Ω for every finite $\tau \geq 0$.

For such an initial point, the time reparameterization in Lemma A.1 has terminal time

$$\tau^\star = \int_0^{S(x)} \frac{ds}{a(Y_s)} = \int_0^{S(x)} \frac{\mu(Y_s)}{\nu(Y_s)}\, ds.$$

If $\tau^\star < \infty$, the AFM trajectory would approach $\mathcal{M}$ at the finite time $\tau^\star$, contradicting forward completeness in Ω. Hence $\tau^\star = \infty$. Combining this with Proposition A.2 gives

$$\lim_{\tau\to\infty} X_\tau(x) = T(x) \in \mathcal{M} \quad \text{for } \rho_0\text{-almost every } x, \qquad T_\# \rho_0 = \rho_1. \tag{68}$$

This proves Theorem 3.2 under the stated regularity and forward-completeness assumptions.

A.3 Finite-duration maps and compactification

Let $\Omega \subset \mathbb{R}^d$ be open, let $b \in C^1(\Omega; \mathbb{R}^d)$, and assume that its forward trajectories remain in Ω and exist for every finite duration. Let $X_\tau(x)$ denote the resulting autonomous flow.

Proposition A.3 (Flow identities)**.** *The flow is jointly C^1 on $[0,\infty) \times \Omega$ and satisfies*

$$X_{\tau+s}(x) = X_s(X_\tau(x)), \tag{69}$$
$$\partial_\tau X_\tau(x) = b(X_\tau(x)), \tag{70}$$
$$\partial_\tau X_\tau(x) = \nabla_x X_\tau(x) b(x). \tag{71}$$

For the clock of Appendix A.1, define $\Phi_\sigma := X_{r^{-1}(\sigma)}, \qquad \sigma \in [0,1)$. Then

$$c(\sigma)\partial_\sigma \Phi_\sigma(x) = b(\Phi_\sigma(x)), \tag{72}$$
$$c(\sigma)\partial_\sigma \Phi_\sigma(x) = \nabla_x \Phi_\sigma(x) b(x), \tag{73}$$
$$\Phi_{r_{\tau_1+\tau_2}} = \Phi_{r_{\tau_2}} \circ \Phi_{r_{\tau_1}}. \tag{74}$$

Proof. Since $b \in C^1(\Omega)$, the ODE is locally unique and solutions depend C^1-smoothly on their initial conditions. The assumed forward existence therefore defines a jointly C^1 flow.

Starting from $X_\tau(x)$ and evolving for an additional duration s solves the same initial-value problem as continuing the trajectory from x for total duration $\tau + s$. Hence $X_{\tau+s}(x) = X_s(X_\tau(x))$. Equation (70) is the defining ODE.

Autonomous durations commute: $X_{\tau+s} = X_\tau \circ X_s$. Differentiating this identity with respect to s at $s = 0$ gives $\partial_\tau X_\tau(x) = \nabla_x X_\tau(x) b(x)$, which proves (71).

Finally, $\Phi_\sigma = X_{r^{-1}(\sigma)}$ and $\frac{d}{d\sigma} r^{-1}(\sigma) = \frac{1}{c(\sigma)}$. Applying the chain rule to (70) and (71) yields (72) and (73). The autonomous semigroup gives

$$\Phi_{r_{\tau_1+\tau_2}} = X_{\tau_1+\tau_2} = X_{\tau_2} \circ X_{\tau_1} = \Phi_{r_{\tau_2}} \circ \Phi_{r_{\tau_1}},$$

which proves (74). □

A.4 Proof of Theorem 3.3

Let r be a clock as in Definition 3.1 with $c(\sigma) = \dot{r}_{r^{-1}(\sigma)}$, and let $\Phi_\sigma = X_{r^{-1}(\sigma)}$. Suppose $\Omega \subseteq \mathbb{R}^d$ is open, $b \in C^1(\Omega; \mathbb{R}^d)$, and the forward flow is defined in Ω for every finite duration. The following restates Theorem 3.3 for the global case $\Omega = \mathbb{R}^d$; Proposition A.3 establishes the same identities on any such Ω.

Theorem 3.3 (Compactified flow map)**.** *Let r be a clock as in Definition 3.1 with clock speed $c(\sigma) = d\sigma/d\tau$ along $\sigma = r_\tau$, and let Φ denote the compactified flow map (14). Under standard regularity assumptions, Φ satisfies the following conditions:*

(i) (Lagrangian condition): for all $\sigma \in [0, 1)$ and $x \in \mathbb{R}^d$,

$$c(\sigma)\partial_\sigma \Phi_\sigma(x) = b(\Phi_\sigma(x)). \tag{15}$$

(ii) (Eulerian condition): for all $\sigma \in [0, 1)$ and $x \in \mathbb{R}^d$,

$$c(\sigma)\partial_\sigma \Phi_\sigma(x) = \nabla_x \Phi_\sigma(x) b(x). \tag{16}$$

(iii) (Semigroup condition): for all $\tau_1, \tau_2 \geq 0$,

$$\Phi_{r_{\tau_1+\tau_2}} = \Phi_{r_{\tau_2}} \circ \Phi_{r_{\tau_1}}. \tag{17}$$

Proof. The Lagrangian and Eulerian identities are exactly (72) and (73) of Proposition A.3. The semigroup identity is (74) of the same proposition. □

Uniqueness. The three characterizations above also determine the compactified flow map uniquely once $\Phi_0 = \mathrm{Id}$ and its tangent at $\sigma = 0$ is fixed to b.

Proposition A.4 (Uniqueness)**.** *Under the assumptions of Theorem 3.3, let $Y : [0, 1) \times \Omega \to \Omega$ be jointly C^1 and satisfy*

$$Y_0(x) = x, \qquad \partial_\sigma Y_\sigma(x)|_{\sigma=0} = b(x).$$

If Y satisfies any one of the following conditions,

$$c(\sigma)\partial_\sigma Y_\sigma(x) = b(Y_\sigma(x)), \tag{75}$$

$$c(\sigma)\partial_\sigma Y_\sigma(x) = \nabla_x Y_\sigma(x) b(x), \tag{76}$$

$$Y_{r_{\tau_1+\tau_2}} = Y_{r_{\tau_2}} \circ Y_{r_{\tau_1}}, \tag{77}$$

then $Y_\sigma = \Phi_\sigma$ for all $\sigma \in [0, 1)$.

Proof. Define $F_\tau(x) := Y_{r_\tau}(x)$. Since $r_0 = 0$, we have $F_0(x) = Y_0(x) = x$.

If Y satisfies the Lagrangian condition (75), then

$$\begin{aligned}\partial_\tau F_\tau(x) &= \dot{r}_\tau \, \partial_\sigma Y_\sigma(x)|_{\sigma=r_\tau} \\ &= c(r_\tau)\partial_\sigma Y_{r_\tau}(x) = b(Y_{r_\tau}(x)) = b(F_\tau(x)).\end{aligned}$$

Thus F_τ solves the same autonomous ODE as X_τ with the same initial condition, so ODE uniqueness gives $F_\tau = X_\tau$.

If Y satisfies the Eulerian condition (76), then $\partial_\tau F_\tau(x) = \nabla_x F_\tau(x) b(x)$. Fix $T \geq 0$ and $x \in \Omega$, and set $z_s := X_{T-s}(x)$ for $0 \leq s \leq T$. Then $\dot{z}_s = -b(z_s)$, and hence

$$\begin{aligned}\frac{d}{ds} F_s(z_s) &= \partial_s F_s(z_s) + \nabla_x F_s(z_s)\dot{z}_s \\ &= \nabla_x F_s(z_s) b(z_s) - \nabla_x F_s(z_s) b(z_s) = 0.\end{aligned}$$

Evaluating this constant at $s = 0$ and $s = T$ yields $F_T(x) = F_0(X_T(x)) = X_T(x)$.

Finally, if Y satisfies the semigroup condition (77), then $F_{\tau+h} = F_h \circ F_\tau$. Differentiating in h at zero gives

$$\partial_\tau F_\tau(x) = \partial_h F_h(y)|_{h=0,\, y=F_\tau(x)}.$$

Because $F_h(y) = Y_{r_h}(y)$, the clock satisfies $\dot{r}_0 = 1$, and the assumed tangent condition gives $\partial_h F_h(y)|_{h=0} = \dot{r}_0 \, \partial_\sigma Y_\sigma(y)|_{\sigma=0} = b(y)$. Thus $\partial_\tau F_\tau(x) = b(F_\tau(x))$; together with $F_0(x) = x$, ODE uniqueness again yields $F_\tau = X_\tau$. In every case, $Y_\sigma = F_{r^{-1}(\sigma)} = X_{r^{-1}(\sigma)} = \Phi_\sigma$. □

Proof of Corollary 3.4. Evaluating the Lagrangian condition (15) at $\sigma = 0$ gives

$$c(0)\ \partial_\sigma \Phi_\sigma(x)|_{\sigma=0} = b(\Phi_0(x)).$$

By Definition 3.1, $r_0 = 0$ and $\dot{r}_0 = 1$, so $r^{-1}(0) = 0$ and $c(0) = \dot{r}_{r^{-1}(0)} = \dot{r}_0 = 1$. Moreover, by (14) and (11), $\Phi_0(x) = X_{r^{-1}(0)}(x) = X_0(x) = x$. Therefore $\partial_\sigma \Phi_\sigma(x)|_{\sigma=0} = b(x)$. If $\Phi_\sigma(x) = x + \sigma v_\sigma(x)$ and $v_\sigma(x)$ is continuous at $\sigma = 0$, then for $\sigma > 0$,

$$v_\sigma(x) = \frac{\Phi_\sigma(x) - \Phi_0(x)}{\sigma}.$$

Taking $\sigma \downarrow 0$ gives $v_0(x) = \partial_\sigma \Phi_\sigma(x)|_{\sigma=0} = b(x)$. □

B CLASSIFIER-FREE GUIDANCE

Classifier-free guidance. Let $v_\tau(x)$ and $v_\tau(x \mid c)$ denote the unconditional and conditional time-dependent velocities. At each fixed time, ordinary CFG forms

$$v_\tau^w(x) = w\, v_\tau(x \mid c) + (1 - w)\, v_\tau(x).$$

The autonomous construction, however, marginalizes over time. Define

$$\pi(\tau \mid x) = \frac{\rho_\tau(x)}{\int_0^\infty \rho_s(x)\, ds}, \qquad \pi_c(\tau \mid x) = \frac{\rho_\tau(x \mid c)}{\int_0^\infty \rho_s(x \mid c)\, ds}.$$

Then

$$b_u(x) = \int_0^\infty \pi(\tau \mid x) v_\tau(x)\, d\tau, \qquad b_c(x) = \int_0^\infty \pi_c(\tau \mid x) v_\tau(x \mid c)\, d\tau.$$

Applying CFG after removing time therefore gives

$$\begin{aligned} b^w_{\text{naive}}(x) &= w b_c(x) + (1 - w) b_u(x) \\ &= w \int_0^\infty \pi_c(\tau \mid x) v_\tau(x \mid c)\, d\tau + (1 - w) \int_0^\infty \pi(\tau \mid x) v_\tau(x)\, d\tau. \end{aligned}$$

By contrast, applying CFG to the time-dependent dynamics first and only then removing time gives

$$b^w(x) = \int_0^\infty \pi_w(\tau \mid x) \big[w\, v_\tau(x \mid c) + (1 - w)\, v_\tau(x) \big]\, d\tau,$$

where $\pi_w(\tau \mid x)$ is the trajectory-time posterior induced by the guided dynamics. Hence, in general, $b^w_{\text{naive}} \neq b^w$, because the conditional, unconditional, and guided fields average velocities under different time posteriors. A sufficient compatibility condition is

$$\pi_c(\tau \mid x) = \pi(\tau \mid x) = \pi_w(\tau \mid x),$$

which need not hold in general.

High-guidance instability. The field-space rule above is nevertheless useful at moderate guidance. To understand its failure at large w, at states with $b_c \neq 0$ decompose the unconditional field relative to the conditional field as

$$b_u = \alpha b_c + b_u^\perp, \qquad \alpha = \frac{b_u^\top b_c}{\|b_c\|^2}, \qquad (b_u^\perp)^\top b_c = 0.$$

Then naive CFG becomes

$$\begin{aligned} b^w_{\text{naive}} &= w b_c - (w - 1) b_u \\ &= \underbrace{\big[1 - (w - 1)(\alpha - 1)\big]}_{\gamma(x)} b_c - (w - 1) b_u^\perp. \end{aligned}$$

Thus guidance has two effects: the orthogonal term steers the trajectory transversely, while $\gamma(x)$ rescales the conditional autonomous field itself. Since repeated autonomous evolution relies on b_c to move the state toward equilibrium, $\gamma < 1$ weakens this restoring component and $\gamma < 0$ reverses it.

Norm-matched guidance. We remove the relative scale mismatch before extrapolation by defining

$$\widetilde{b}_u = \frac{\|b_c\|}{\|b_u\|} b_u, \qquad b_{\text{norm}} = w b_c - (w-1)\widetilde{b}_u.$$

Writing

$$\cos\theta = \frac{b_u^\top b_c}{\|b_u\|\,\|b_c\|},$$

the component of the guided field along b_c satisfies, for $w \geq 1$,

$$\frac{b_c^\top b_{\text{norm}}}{\|b_c\|^2} = w - (w-1)\cos\theta \geq 1.$$

Norm matching therefore prevents a difference in conditional and unconditional field magnitudes from weakening the conditional restoring component.

Perpendicular guidance. Alternatively, we discard the parallel part of the unconditional correction and retain only its orthogonal component,

$$b_{\text{perp}} = b_c - (w-1) b_u^\perp.$$

By construction,

$$\frac{b_c^\top b_{\text{perp}}}{\|b_c\|^2} = 1,$$

so the conditional restoring component is left exactly unchanged. This is closely related to projected CFG rules used for diffusion models [28]; here the projection has a direct interpretation in terms of preserving the autonomous dynamics toward equilibrium. At zeros, define the normalized b_u term to be zero when $b_u = 0$, and set $b_u^\perp = 0$ when $b_c = 0$.

A targeted ImageNet-256 intervention supports this interpretation. At $w = 10$, naive guidance gives FID-50k 43.3. Preventing only reversal, $\gamma \leftarrow \max(\gamma, 0)$, improves FID to 27.0, whereas preserving at least the original restoring strength, $\gamma \leftarrow \max(\gamma, 1)$, gives 10.5, comparable to norm matching (10.8) and perpendicular guidance (11.9):

rule	naive	$\gamma \geq 0$	$\gamma \geq 1$	norm	perp
FID-50k	43.3	27.0	10.5	10.8	11.9

At the standard $w = 1.5$, the guidance rules are essentially indistinguishable, indicating that the stabilization primarily matters in the high-guidance regime.

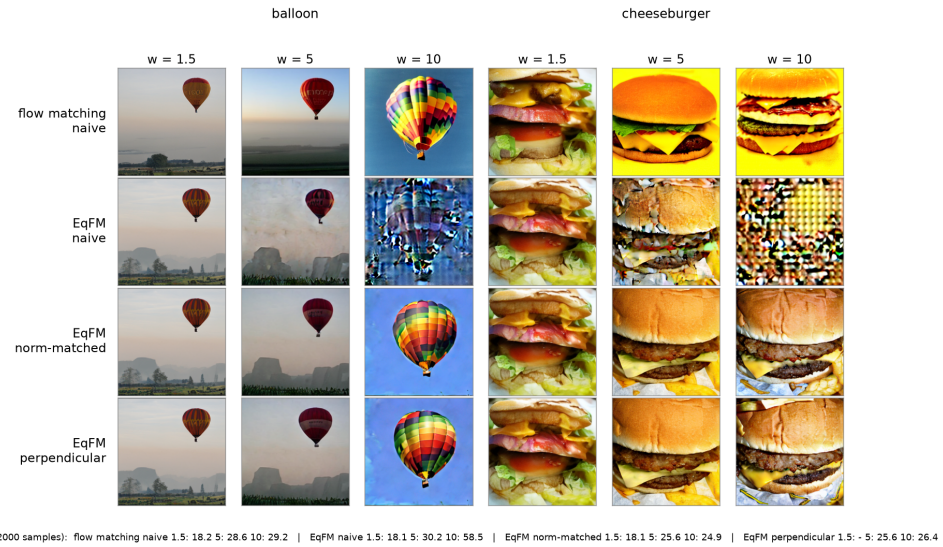


Figure 3: **High-guidance ImageNet-256 examples.** Balloon and cheeseburger samples at three guidance strengths, with the same initial noise across methods in each column. Naive CFG degrades at $w = 10$; norm matching and perpendicular guidance retain coherent samples. The FID-2k values shown in the panel use 2,000 samples.

Setting	Autonomous field	EqFM distillation
Model	SiT-XL/2, 256^2	SiT-XL/2, 256^2
Weights used	EMA + frozen raw	EMA teacher
Trainable parameters	Full network	Full network
Objective	AFM field	Eulerian EqFM
Active losses	AFM objective; data anchor, 0.5	Diagonal, 1; distillation, 1
Time sampling	Uniform-weighted sampler	Uniform interpolant and σ
σ range	—	$[10^{-4}, 1)$
CFG scale	—	1.5, guided; norm matched
Optimizer	AdamW [35] $(0.9, 0.999)$	AdamW $(0.9, 0.999)$
Weight decay	0	0
Learning rate	1×10^{-4}	1×10^{-4}
Learning-rate schedule	None	Cosine decay from 25k to 400k
Global batch size	256	256 $(8 \times 8 \times 4)$
EMA decay μ	0.9999	0.9999
Gradient clipping	None	1
Precision	FP32	FP32
Training steps	480k	120k

Table 4: ImageNet-256 training configuration of the autonomous field teacher and the EqFM student.

C Implementation details

Table 4 lists the ImageNet-256 training configuration of the autonomous field teacher and the EqFM student. Table 5 lists the CIFAR-10 training configuration.

Time sampling. For the diagonal loss, we sample the interpolant time $t \sim \mathrm{Unif}[0, 1]$ and evaluate the student at the diagonal $\sigma = 0$. For the Eulerian loss, we reuse the same interpolant sample and independently draw a compactified duration $\sigma \sim \mathrm{Unif}[10^{-4}, 1)$; the corresponding autonomous time is then determined by the clock, $\tau = -\log(1-\sigma)$, rather than sampled separately. For the semigroup loss, we first sample the total compactified duration σ with density proportional to σ on $[0, 1)$, then sample $\sigma' \sim \mathrm{Unif}[0, \sigma]$. We set

$$\sigma_- = \frac{\sigma - \sigma'}{1 - \sigma'},$$

so that the two substeps compose to the sampled total duration. Indeed, with $\tau' = r^{-1}(\sigma')$ and $\tau_- = r^{-1}(\sigma_-)$, the semigroup condition (17) gives

$$\Phi_{r_{\tau'+\tau_-}} = \Phi_{\sigma_-} \circ \Phi_{\sigma'}, \qquad r_{\tau'+\tau_-} = 1 - (1-\sigma')(1-\sigma_-) = \sigma,$$

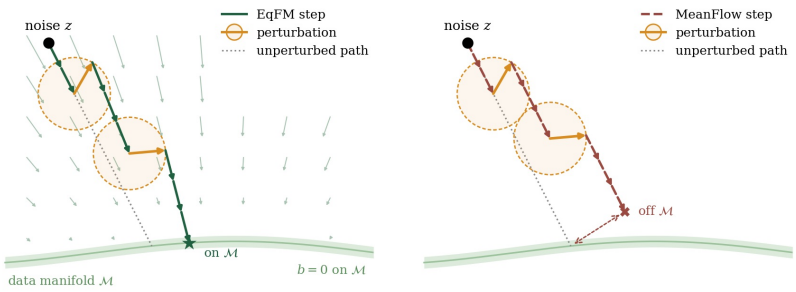
where the first equality uses the exponential clock $r_\tau = 1 - e^{-\tau}$.

Setting	MeanFlow (baseline)	EqFM
Model	ADM UNet, 32^2	ADM UNet, 32^2
Parameters	36.08 M	35.75 M
Weights used	Random init, no teacher	Random init, no teacher; $b = \mathrm{sg}[v_0]$, live
Trainable parameters	Full network	Full network
Objective	MeanFlow $u(z, r, t)$	Eulerian, Lagrangian, or semigroup
Active losses	MeanFlow only ($r = t$ on 25%)	MeanFlow only ($r = t$ on 25%)
Time sampling	Logit-normal (t, r), v1	Uniform t and σ
σ range	—	$[0,\ 1]$
CFG scale	None (unconditional)	None (unconditional)
Optimizer	Adam $(0.9, 0.999)$	AdamW $(0.9, 0.999)$
Weight decay	0	0
Learning rate	1×10^{-4}	1×10^{-4}
Learning-rate schedule	Cosine decay from 60k to 100k	Cosine decay from 60k to 100k
Global batch size	128	128
EMA decay μ	0.9999	0.9999 (warm-up ramp)
Gradient clipping	None	None
Precision	FP32	FP32
Training steps	100k (6 GPU-hours)	Compute-matched (6 GPU-hours)

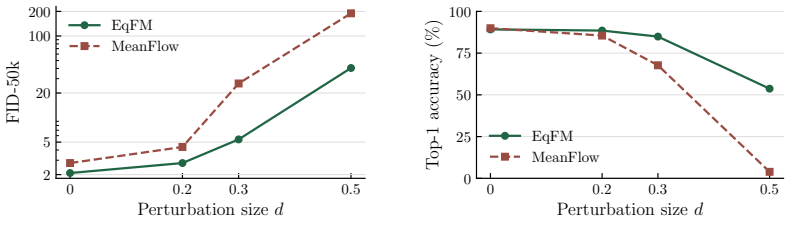
Table 5: CIFAR-10 from-scratch training configuration (1×GH200). Both arms use random initialization and no teacher. The parameter counts differ only in the time input: MeanFlow embeds two times, while EqFM embeds a single duration σ.

D ADDITIONAL EXPERIMENTS

Robustness to perturbations. Before calls $K/4$, $K/2$, and $3K/4$ of the $K = 8$-call sampler, we replace x by $x + d\|x\|v$ for a random unit direction v, without changing either sampler's schedule. We pair noises, labels, and perturbation directions across models. At $d = 0.3$, FID rises by 3.31 for EqFM versus 23.40 for MeanFlow; at $d = 0.5$, top-1 accuracy is 53.7% versus 3.9% (Figure 5c,d). This is consistent with restoring equilibrium dynamics: EqFM advances from its perturbed state without a time input, whereas MeanFlow uses the nominal time even after the state leaves its expected path, allowing errors to compound.



(a) EqFM (b) MeanFlow



(c) FID-50k (d) Top-1 accuracy

Figure 5: **Robustness to perturbations during ImageNet-256 sampling.** (a,b) Schematic responses to unannounced perturbations; two are drawn, while the experiment applies three. (c,d) FID-50k (log scale) and ResNet-50 [36] top-1 accuracy as perturbation size d increases.



Figure 4: **One versus eight EqFM calls on ImageNet-256.** Paired samples use the same initial noises. Eight calls use the schedule of Section 5. The panel was curated from 675 candidates using only the eight-call outputs; the paired one-call outputs were not inspected during selection. All displayed images belong to the corresponding scored 50k sample sets.