VTV-FM: Flow Matching through Variational Terminal-Velocity Closure

Haoyang Jiang* **Yuheng Li*** **Di Yang***
Yanhai Xiong **Haipeng Chen†** **Yi He†**
Department of Data Science, William & Mary
Williamsburg, VA 23185, USA
{hjiang16,yli95,dyang06,yxiong05,hchen23,yihe}@wm.edu
*Equal contribution. †Corresponding authors.

Abstract

Flow matching (FM) learns generative transport by fitting continuous-time motion from a simple source distribution to the data distribution. Most existing methods use first-order bridges: once a source and a target sample are paired, the path is a straight motion with constant velocity. FM with optimal transport (OT) improves the pairing, but the bridge itself remains linear, limiting its ability to model curved motion, acceleration, and changing directions. A natural remedy is to use second-order phase-space dynamics; however, learning the bridge requires target-side terminal-velocity information that static datasets do not provide. We propose **V**ariational **T**erminal-**V**elocity **F**low **M**atching (**VTV-FM**), a second-order FM framework that derives the missing velocity by minimizing acceleration energy, yielding a closed-form closure for static data. The same minimum-acceleration variational construction also defines the OT pairing cost and the acceleration targets used for training. Experiments on low-dimensional datasets, PDE-governed physical fields, and CIFAR-10 show that VTV-FM improves transport geometry and generation quality over first-order and high-order FM baselines. Our code is available at `https://github.com/HaoyangJiang-WM/VTV-FM`.

1 Introduction

Flow Matching (FM) has emerged as a flexible generative model by learning continuous-time vector fields along prescribed probability paths [1, 2]. A common choice is the first-order linear bridge: given a paired source sample and target sample, the model learns a straight path from x_0 to x_1, corresponding to constant-velocity transport [1]. OT-based Flow Matching further improves this framework by replacing independent source–target pairing with an optimal coupling, thereby reducing transport cost and vector-field conflicts [3, 4]. However, even after the pairing is improved, the conditional bridge itself remains a first-order linear interpolant. For data distributions with curvature, branching structures, or separated modes, straight paths may pass through low-density regions and require the model to learn vector fields in poorly supported intermediate regions. Meanwhile, trajectory intersections can induce conflicting velocity supervision, forcing the learned vector field to average incompatible directions [5]. This motivates a second-order phase-space formulation, where transport can exhibit nonuniform speed and curved position paths with time-varying tangent directions, rather than being restricted to constant-velocity linear interpolation.

Constructing a second-order bridge is fundamentally more challenging than constructing a first-order path. A first-order bridge only requires endpoint positions, whereas a second-order phase-space bridge also requires endpoint velocities, since these boundary velocities determine the intermediate trajectory and the induced acceleration supervision. The initial velocity v_0 can be sampled from a source phase-space prior, but static datasets typically provide only the terminal position x_1, leaving

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

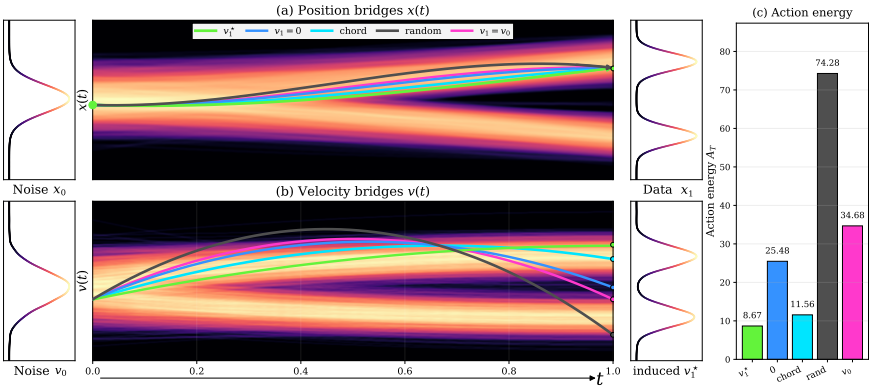


Figure 1: Effect of terminal-velocity choices on second-order bridge supervision. All curves in (a,b) use the same observable boundary information (x_0, v_0, x_1), but different terminal velocities v_1. In (a,b), the horizontal axis is normalized time t/T, and colors denote the same terminal-velocity choices: $v_1^\star$, $v_1 = 0$, chord velocity $(x_1 - x_0)/T$, random velocity, and $v_1 = v_0$. **(a)** The position bridges show the evolution of $x(t)$ under different terminal-velocity choices. **(b)** The velocity bridges show the corresponding evolution of $v(t)$. **(c)** The bars show the acceleration-energy cost for each terminal-velocity choice. Lower cost corresponds to smoother supervision, while externally specified terminal velocities may induce biased supervision.

the terminal velocity v_1 unspecified. Related high-order and acceleration-aware FM methods enrich path geometry through higher-order supervision, learned path constructions, boundary constraints, or acceleration-based refinement mechanisms [6–11].

VTV-FM instead formulates the static-data case by treating v_1 as a latent boundary variable and eliminating it variationally, tying the terminal-velocity completion to the bridge geometry and pairing cost. This issue is critical because v_1 controls the full phase-space bridge. Externally specified terminal velocities can therefore introduce unnecessary curvature or acceleration, biasing the regression targets. Figure 1 illustrates this sensitivity: under the same observable boundary information (x_0, v_0, x_1), different choices of v_1 induce different position trajectories, velocity profiles, and acceleration-energy costs. Since deterministic FM sampling relies on numerical integration of the learned dynamics at inference time, such trajectory distortions can make discretization errors accumulate more easily and increase sensitivity to solver choice and time discretization [5, 12].

To address these structural limitations, we propose **V**ariational **T**erminal-**V**elocity **F**low **M**atching (**VTV-FM**). Its core is a minimum-acceleration variational principle: the same variational problem simultaneously induces the second-order bridge, the terminal-velocity closure, and the OT pairing cost. For a complete boundary state (x_0, v_0, x_1, v_1), the minimum-acceleration problem yields a cubic Hermite bridge, together with the corresponding intermediate velocity and acceleration supervision. The optimal action value of this trajectory further defines a full-boundary second-order transport cost.

In static-data regimes where the terminal velocity is missing, VTV-FM does not introduce an auxiliary estimator or artificial constraint. Instead, it treats v_1 as a latent boundary variable and directly minimizes the same full-boundary cost with respect to v_1. This optimization has a closed-form solution, yielding the variational terminal-velocity closure $v_1^\star$. Substituting $v_1^\star$ back into the original cost gives a projected OT cost that depends only on the observable boundary variables (x_0, v_0, x_1). In this way, the missing v_1 in static data is determined intrinsically by the same minimum-acceleration principle, rather than being specified by an external mechanism. When the terminal velocity is observed or derivable, as in PDE data or temporal sequences, VTV-FM directly uses the full-boundary bridge without applying the closure. Thus, the terminal-velocity closure, the OT pairing cost, and the acceleration supervision induced by the second-order bridge are all determined by the same minimum-acceleration variational principle, rather than being specified separately by external rules.

Our contributions include: (1) a phase-space formulation of FM that identifies missing terminal velocity as the key obstacle for second-order bridges on static data; (2) together with the proposed VTV-FM framework, a closed-form terminal-velocity closure derived from a minimum-acceleration principle, which also defines the bridge and projected OT pairing cost; and (3) empirical validation on low-dimensional transport, PDE-governed fields, and CIFAR-10, where VTV-FM improves over first-order, OT-based, and high-order FM baselines.

2 Preliminaries

First-order FM and OT pairing. Let ρ_0 and ρ_1 denote the source and target distributions in position space. Given $x_0 \sim \rho_0$ and $x_1 \sim \rho_1$, standard FM constructs the linear bridge $x_\tau = (1-\tau)x_0 + \tau x_1$, $\tau \in [0,1]$, with conditional velocity $u_\tau = x_1 - x_0$. The velocity field is trained by $\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}\|v_\theta(x_\tau, \tau) - u_\tau\|_2^2$. OT-CFM [3] replaces the independent source–target pairing with an OT coupling. Given minibatch samples, it constructs a cost matrix using the quadratic cost $c_{\mathrm{OT}}(x_0, x_1) = \|x_1 - x_0\|_2^2$, and approximately solves the coupling by entropic OT / Sinkhorn iterations [13, 3]. From a variational view, this cost corresponds to the minimum value of the first-order kinetic action $\int_0^1 \|\dot{x}_t\|_2^2 dt$ between fixed endpoints, whose minimizer is a constant-velocity straight line. Thus, OT-CFM improves endpoint pairing, but the bridge remains linear and first-order.

Second-order phase-space dynamics. To construct second-order transport, we lift the state to phase space, $z_t = (x_t, v_t) \in \mathcal{Z} := \mathcal{X} \times \mathcal{V}$, where $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{V} = \mathbb{R}^d$ are the state and velocity space, respectively. We consider second-order dynamics $v_t = \dot{x}_t$, $a_t = \dot{v}_t = \ddot{x}_t$. Unlike a first-order bridge, which only requires endpoint positions (x_0, x_1), a second-order bridge typically requires the full phase-space boundary state (x_0, v_0, x_1, v_1). This phase-space formulation can be viewed as a second-order analogue of first-order OT: we use the acceleration action $\mathcal{A}_T[x] = \int_0^T \|\ddot{x}_t\|_2^2 dt$ to induce a second-order bridge and the corresponding endpoint cost. Given this cost, the pairing step remains a discrete OT problem and can be solved with minibatch Sinkhorn as in OT-CFM. The key difference is that static datasets usually provide only the terminal position x_1, but not the terminal velocity v_1. Therefore, determining the missing v_1 is the central obstacle in extending FM to second-order phase-space bridges. The next section shows how the acceleration action yields the bridge, terminal-velocity closure, and OT pairing cost.

3 Proposed Method: VTV-FM

We present VTV-FM, a second-order Flow Matching framework in phase space. Its central idea is to derive the transport bridge, terminal-velocity closure, and OT pairing cost from a single minimum-acceleration variational principle, rather than specifying them as separate design choices. This unified construction allows VTV-FM to use the full-boundary bridge when terminal velocities are available, and to close the missing terminal velocity variationally when they are not. We first derive the closed-form full-boundary bridge and cost for (x_0, v_0, x_1, v_1) (Section 3.1). We then project this cost onto (x_0, v_0, x_1), obtaining a closed-form terminal-velocity closure and projected OT cost for the latent-v_1 regime (Section 3.2). Finally, we use these quantities for minibatch pairing, trajectory supervision, acceleration matching, and sampling (Sections 3.3–3.4).

3.1 Full-Boundary Minimum-Acceleration Bridge

We consider the idealized full-boundary setting, where a complete boundary state (x_0, v_0, x_1, v_1) and a transport horizon $T > 0$ are given. Let the set of admissible trajectories be

$$\Gamma_T(x_0, v_0, x_1, v_1) = \left\{\gamma \in H^2([0,T]; \mathbb{R}^d) : \gamma(0) = x_0,\ \dot{\gamma}(0) = v_0,\ \gamma(T) = x_1,\ \dot{\gamma}(T) = v_1\right\},$$

which satisfy the full phase-space boundary conditions. Following the second-order OT viewpoint, each trajectory is measured by the squared-acceleration action $\mathcal{A}_T[\gamma] = \int_0^T \|\ddot{\gamma}(t)\|_2^2\, dt$. The full-boundary transport cost is the minimum action value, and the corresponding variational bridge is any action minimizer, which are defined as follows:

$$C_T(x_0, v_0, x_1, v_1) = \inf_{\gamma \in \Gamma_T(x_0, v_0, x_1, v_1)} \mathcal{A}_T[\gamma], \tag{1a}$$

$$\gamma_T^\star(x_0, v_0, x_1, v_1) \in \arg\min_{\gamma \in \Gamma_T(x_0, v_0, x_1, v_1)} \mathcal{A}_T[\gamma]. \tag{1b}$$

The bridge induces the phase-space trajectory and acceleration target:

$$x_t^\star = \gamma_T^\star(t), \qquad v_t^\star = \dot{\gamma}_T^\star(t), \qquad a_t^\star = \ddot{\gamma}_T^\star(t).$$

Equations (1a) and (1b) establish the basic variational objective used by VTV-FM. For a complete boundary state, the endpoint cost, intermediate trajectory, and acceleration target are all induced by the same minimum-acceleration principle. While these two equations specify the bridge and

cost implicitly, Proposition 3.1 provides their closed-form solutions as functions of the boundary states (x_0, v_0, x_1, v_1) and shows that the cubic Hermite bridge is not an additional path assumption, but the unique solution of the full-boundary minimum-acceleration problem. This cubic Hermite construction is classical in spline theory [14].

Proposition 3.1 (Full-boundary minimum-acceleration bridge)**.** *For any $T > 0$ and complete boundary state (x_0, v_0, x_1, v_1), the variational bridge in Eq. (1b) exists and is unique. Let $\tau = t/T$. The unique minimizer is characterized by the cubic Hermite curve*

$$x_t^\star = h_{00}(\tau)x_0 + h_{10}(\tau)Tv_0 + h_{01}(\tau)x_1 + h_{11}(\tau)Tv_1, \tag{2}$$

where

$$\begin{aligned} h_{00}(\tau) &= 2\tau^3 - 3\tau^2 + 1, \qquad & h_{10}(\tau) &= \tau^3 - 2\tau^2 + \tau, \\ h_{01}(\tau) &= -2\tau^3 + 3\tau^2, \qquad & h_{11}(\tau) &= \tau^3 - \tau^2. \end{aligned}$$

The induced velocity and acceleration are

$$\begin{aligned} v_t^\star &= \frac{1}{T}\left[h'_{00}(\tau)x_0 + h'_{10}(\tau)Tv_0 + h'_{01}(\tau)x_1 + h'_{11}(\tau)Tv_1\right], \\ a_t^\star &= \frac{1}{T^2}\left[h''_{00}(\tau)x_0 + h''_{10}(\tau)Tv_0 + h''_{01}(\tau)x_1 + h''_{11}(\tau)Tv_1\right]. \end{aligned} \tag{3}$$

Moreover, the full-boundary cost in Eq. (1a) admits the closed form

$$C_T(x_0, v_0, x_1, v_1) = \frac{1}{T}\left[12\left\|\frac{x_1 - x_0}{T} - \frac{v_0 + v_1}{2}\right\|_2^2 + \|v_1 - v_0\|_2^2\right]. \tag{4}$$

The proof is given in Appendix C.1. Proposition 3.1 implies that, once the complete boundary state is specified, VTV-FM obtains both the second-order trajectory and the endpoint cost from a single variational principle. In the observed-v_1 regime, this full-boundary bridge can be used directly for pairing and acceleration supervision. In the latent-v_1 regime, the same cost C_T provides the starting point for eliminating the unobserved terminal velocity.

3.2 Terminal-Velocity Closure as Variational Projection

In settings where only the observable boundary tuple (x_0, v_0, x_1) is observed, the full-boundary bridge in Proposition 3.1 cannot be instantiated directly, since it depends on the terminal velocity v_1. We determine this unobserved terminal velocity by projecting the full-boundary cost onto the observable state space. Rather than relying on a heuristic, we minimize $C_T(x_0, v_0, x_1, v_1)$ with respect to v_1. The following theorem shows that minimizing C_T over the latent terminal velocity yields a closed-form terminal-velocity closure $v_1^\star$, and the associated projected cost $C_T^\star$.

Theorem 3.2 (Variational terminal-velocity projection)**.** *For any $T > 0$ and observable boundary tuple (x_0, v_0, x_1), the minimizer of the full-boundary cost over the latent terminal velocity is*

$$v_1^\star(x_0, v_0, x_1) = \arg\min_{v_1 \in \mathcal{V}} C_T(x_0, v_0, x_1, v_1) = \frac{3}{2T}(x_1 - x_0) - \frac{1}{2}v_0, \tag{5}$$

and the corresponding projected transport cost is

$$C_T^\star(x_0, v_0, x_1) = \inf_{v_1 \in \mathcal{V}} C_T(x_0, v_0, x_1, v_1) = \frac{3}{T}\left\|\frac{x_1 - x_0}{T} - v_0\right\|_2^2. \tag{6}$$

Moreover, this projection is exact: minimizing the projected cost over observable couplings is equivalent to minimizing the full-boundary cost over all lifted couplings with latent terminal velocities.

The proof is given in Appendix C.2. Theorem 3.2 provides the variational projection of the full-boundary cost onto the observable boundary variables. In this construction, the missing terminal velocity is not supplied by an auxiliary model or an external heuristic; instead, it is the analytic optimizer of the same minimum-acceleration principle that defines the bridge and the endpoint cost. In the latent-v_1 regime, VTV-FM therefore uses the completed boundary state $(x_0, v_0, x_1, v_1^\star(x_0, v_0, x_1))$ to instantiate the full-boundary bridge in Proposition 3.1, yielding the trajectory $x_t^\star$, velocity $v_t^\star$, and acceleration target $a_t^\star$.

Consequently, both regimes use the same full-boundary bridge construction. When v_1 is observed, VTV-FM uses the full-boundary cost C_T for pairing and the observed v_1 for bridge instantiation. When v_1 is latent, it uses the projected cost $C_T^\star$ for pairing and the variational closure $v_1^\star$ for bridge instantiation. The next section describes how C_T and $C_T^\star$ are used to construct the minibatch OT cost matrix and generate trajectory-level supervision.

3.3 Mini-batch Pairing and Trajectory Supervision

Theorem 3.2 justifies using $C_T^\star$ in Eq. (6) for latent-v_1 pairing: it gives the observable OT problem equivalent to the full-boundary formulation after optimizing out v_1. We now compute the variational costs on mini-batches to compute source–target pairings and trajectory-level supervision.

Given a source mini-batch $\mathcal{B}_0 = \{(x_0^{(i)}, v_0^{(i)})\}_{i=1}^m \sim \mu_0$ and a target mini-batch $\mathcal{B}_1 = \{x_1^{(j)}\}_{j=1}^n \sim \rho_1$, we can construct a pairing cost matrix $C \in \mathbb{R}^{m\times n}$ by

$$C_{ij} = \begin{cases} C_T(x_0^{(i)}, v_0^{(i)}, x_1^{(j)}, v_1^{(j)}), & \text{observed } v_1, \\ C_T^\star(x_0^{(i)}, v_0^{(i)}, x_1^{(j)}), & \text{latent } v_1. \end{cases} \tag{7}$$

Here $v_1^{(j)}$ denotes the terminal velocity estimated from temporal differences, simulator derivatives, or physical derivative information when available. We then solve the entropic OT problem with Sinkhorn iterations and use the resulting minibatch OT plan to select matched pairs (i, j). Details of the OT solver are provided in Appendix C.3.

To compute the bridge trajectory and acceleration target for each selected pair (i, j), we first complete its terminal velocity:

$$\widetilde{v}_1^{(i,j)} = \begin{cases} v_1^{(j)}, & \text{observed } v_1, \\ v_1^\star(x_0^{(i)}, v_0^{(i)}, x_1^{(j)}), & \text{latent } v_1. \end{cases} \tag{8}$$

Thus, both regimes produce a complete boundary tuple $\zeta^{(i,j)} = \big(x_0^{(i)}, v_0^{(i)}, x_1^{(j)}, \widetilde{v}_1^{(i,j)}\big)$. For $t \sim \mathrm{Unif}(0, T)$, we instantiate the full-boundary bridge in Proposition 3.1 as

$$x_t^\star = \gamma_T^\star(t; \zeta^{(i,j)}), \qquad v_t^\star = \dot\gamma_T^\star(t; \zeta^{(i,j)}), \qquad a_t^\star = \ddot\gamma_T^\star(t; \zeta^{(i,j)}).$$

The tuple $(x_t^\star, v_t^\star, t/T, a_t^\star)$ provides the trajectory supervision to train the acceleration field. This construction unifies pairing and supervision: both the matching cost and the acceleration target are induced by the same minimum-acceleration bridge, with the latent-v_1 closure given by Theorem 3.2.

3.4 Acceleration Matching and Sampling

Given the trajectory supervision $(x_t^\star, v_t^\star, \tau, a_t^\star)$, with $\tau = t/T$, we train a neural acceleration field $a_\theta : \mathcal{X} \times \mathcal{V} \times [0, 1] \to \mathcal{V}$. The primary objective is *acceleration matching*, which regresses the acceleration target induced by the variational bridge:

$$\mathcal{L}_{\mathrm{acc}}(\theta) = \mathbb{E}\left[\|a_\theta(x_t^\star, v_t^\star, \tau) - a_t^\star\|_2^2\right], \tag{9}$$

where the expectation is over source samples, mini-batch OT pairings, and sampled times.

We optionally add local phase-space consistency to encourage the learned dynamics to reproduce the short-time evolution of the same variational bridge. For a small step Δt, we use

$$\mathcal{L}_{\mathrm{va}} = \mathbb{E}\left[\left\|v_{t+\Delta t}^\star - v_t^\star - \Delta t\, a_\theta(x_t^\star, v_t^\star, \tau)\right\|_2^2\right], \tag{10a}$$

$$\mathcal{L}_{\mathrm{xv}} = \mathbb{E}\left[\left\|x_{t+\Delta t}^\star - x_t^\star - \Delta t\, v_t^\star - \frac{1}{2}\Delta t^2 a_\theta(x_t^\star, v_t^\star, \tau)\right\|_2^2\right]. \tag{10b}$$

Here $\mathcal{L}_{\mathrm{va}}$ enforces local velocity consistency, while $\mathcal{L}_{\mathrm{xv}}$ enforces the corresponding second-order position consistency.

Besides, inspired by flow-divergence matching [15], we introduce *acceleration-divergence matching* (ADM), a second-order phase-space analogue that regularizes the divergence of the learned

acceleration dynamics. We match the bridge-induced divergence target through

$$\mathcal{L}_{\mathrm{ADM}} = \mathbb{E}\left[\left(\frac{T-t}{d}\nabla_v \cdot a_\theta(x_t^\star, v_t^\star, \tau) + 3\right)^2\right]. \tag{11}$$

Here, $+3$ corresponds to the latent-v_1 regime; for fixed observed v_1, it becomes $+4$ (Appendix C.5). This term encourages the learned phase-space dynamics to match the local contraction or expansion structure of the variational bridge. Detailed derivations and implementation details are provided in Appendices C.4 and C.5. The total objective is

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{acc}} + \lambda_{\mathrm{va}}\mathcal{L}_{\mathrm{va}} + \lambda_{\mathrm{xv}}\mathcal{L}_{\mathrm{xv}} + \lambda_{\mathrm{ADM}}\mathcal{L}_{\mathrm{ADM}}.$$

Sampling. After training, samples are generated by integrating the learned second-order dynamics

$$\dot{x}_t = v_t, \qquad \dot{v}_t = a_\theta(x_t, v_t, t/T)$$

from an initial phase-space sample $(x_0, v_0) \sim \mu_0$. The generated sample is the terminal position x_T. In practice, rather than first-order Euler discretization, we use fixed-step solvers better aligned with second-order phase-space dynamics, including Heun, Taylor2, and Velocity Verlet [16–18]. For Velocity Verlet, with $\Delta t = T/N$ and $a_k = a_\theta(x_k, v_k, t_k/T)$, we first update the position as $x_{k+1} = x_k + \Delta t\, v_k + \frac{1}{2}\Delta t^2 a_k$. We then evaluate $a_{k+1} = a_\theta(x_{k+1}, v_k + \Delta t\, a_k, t_{k+1}/T)$ and correct the velocity by $v_{k+1} = v_k + \frac{\Delta t}{2}(a_k + a_{k+1})$. Complete update rules for Heun, Taylor2, and Velocity Verlet are provided in Appendix C.6. Algorithm 1 summarizes the training and sampling procedure:

Algorithm 1 VTV-FM: Variational Terminal-Velocity Flow Matching

Require: Source batch $B_0 = \{(x_0^i, v_0^i)\}_{i=1}^m$, and target batch $B_1 = \{x_1^j\}_{j=1}^n$, horizon T, acceleration model a_θ

for each training iteration **do**
- For the latent-v_1 regime, compute $v_1^\star$ by Eq. (5) and obtain $C_T^\star$ by Eq. (6).
- Construct the mini-batch OT cost matrix C_{ij} by Eq. (7).
- Solve mini-batch OT by Sinkhorn and obtain matched pairs (i, j).
- Complete the terminal velocity $\widetilde{v}_1^{(i,j)}$ by Eq. (8).
- Sample $t \sim \mathrm{Unif}(0, T)$ and set $\tau = t/T$.
- Compute trajectory supervision $(x_t^\star, v_t^\star, a_t^\star)$ using the full-boundary bridge in Eqs. (2), (3).
- Update a_θ by minimizing Eq. (9), optionally with the regularizers in Eqs. (10a), (10b), (11).

end for

Sampling: draw $(x_0, v_0) \sim \mu_0$, integrate the learned second-order dynamics, and return x_T.

4 Experiments

We quantitatively evaluate VTV-FM on low-dimensional datasets, PDE-governed physical fields [19], and CIFAR-10 [20], and additionally provide qualitative proof-of-concept temporal examples on BAIR robot pushing [21] and KTH [22]. The low-dimensional datasets and CIFAR-10 evaluate the missing-v_1 regime, where terminal velocity is unobserved and must be inferred by the proposed closure. The low-dimensional setting further enables direct visualization of sample quality and transport trajectories. In contrast, the PDE and temporal examples evaluate the observed/derivable-v_1 regime, where terminal derivative information is available from governing dynamics, simulators, or adjacent frames. Additional qualitative temporal visualizations are provided in Appendix E.4.

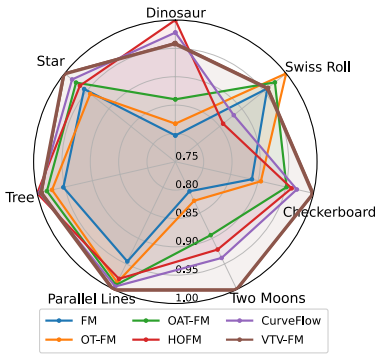
4.1 Low-Dimensional Case Studies

We evaluate VTV-FM on low-dimensional datasets, where generated samples and transport trajectories can be directly visualized. This setting compares sample quality and transport geometry on complex two-dimensional structures. Loss and solver analyses are provided in Appendix F.1.

Datasets. We use seven two-dimensional datasets: Dinosaur, Swiss Roll, Checkerboard, Two Moons, Parallel Lines, Tree, and Star. These datasets cover curved manifolds, multimodal structures, branching patterns, and sharp local direction changes, making them suitable for testing how different flow constructions model low-dimensional geometry.

Table 1: Low-dimensional results measured by sample-based W_2. Lower is better. The right radar plot visualizes the same results using the relative score $s_{d,m}$, where higher is better.

Dataset	CFM	OT-CFM	OAT-FM	HOFM	CurveFlow	VTV-FM
Dinosaur	0.1336	0.1302	0.1237	**0.1064**	0.1088	0.1109
Swiss Roll	0.1083	**0.1037**	0.1064	0.1209	0.1176	0.1081
Checkerboard	0.1229	0.1207	0.1148	0.1137	0.1126	**0.1092**
Two Moons	0.0978	0.0956	0.0884	0.0857	0.0842	**0.0790**
Parallel Lines	0.1441	0.1381	0.1374	0.1391	0.1370	**0.1362**
Tree	0.0760	0.0744	0.0737	**0.0724**	0.0727	0.0730
Star	0.0902	0.0915	0.0885	0.0893	0.0877	**0.0862**
Avg.	0.1104	0.1077	0.1047	0.1039	0.1029	**0.1004**



Baselines. We compare VTV-FM with CFM [1], OT-CFM [3], OAT-FM [11], HOFM [8], and CurveFlow [9]. CFM and OT-CFM are first-order baselines, where OT-CFM improves source–target pairing through minibatch OT while retaining a first-order linear conditional bridge. OAT-FM, HOFM, and CurveFlow are stronger acceleration-based or curve-based baselines, allowing us to compare whether high-order path modeling improves low-dimensional generation quality.

Metric. We evaluate generation quality using sample-based W_2. For each dataset, we denote the generated samples by $\{\hat{x}_i\}_{i=1}^n$ and the target samples by $\{y_i\}_{i=1}^n$. The corresponding empirical distributions are $\widehat{\rho}_{\text{gen}} = \sum_i \delta_{\hat{x}_i}/n$ and $\widehat{\rho}_{\text{data}} = \sum_i \delta_{y_i}/n$. We report $W_2(\widehat{\rho}_{\text{gen}}, \widehat{\rho}_{\text{data}}) = \left(\min_{\pi\in\Pi} \sum_{i,j} \pi_{ij} \|\hat{x}_i - y_j\|_2^2\right)^{1/2}$, where Π denotes the set of couplings between the two uniform empirical distributions. Lower values indicate better agreement between generated and target samples. For the radar plot, we convert the raw W_2 values into a per-dataset relative score. Given dataset d and method m, we define $s_{d,m} = \min_{m'} W_2(d, m')/W_2(d, m)$. Thus, the best method on each dataset has score 1, and higher radar scores indicate better relative performance.

Results and findings. Table 1 reports the main results on seven low-dimensional datasets. VTV-FM achieves the best average sample-based W_2, reaching 0.1004 and outperforming all baselines. Compared with CFM, OT-CFM, OAT-FM, HOFM, and CurveFlow, VTV-FM reduces the average W_2 by approximately 9.1%, 6.8%, 4.1%, 3.4%, and 2.4%, respectively. The radar relative score shows the same trend: VTV-FM obtains the highest average score of 0.987, compared with 0.963 for CurveFlow and 0.955 for HOFM, indicating more stable relative performance across datasets.

At the dataset level, VTV-FM achieves the best result on $4/7$ datasets: Checkerboard, Two Moons, Parallel Lines, and Star. Compared with the strongest competing baseline on these datasets, VTV-FM reduces W_2 by approximately 3.0%, 6.2%, 0.6%, and 1.7%, respectively. Checkerboard, Two Moons, and Star involve multimodality, separated modes, or sharp local direction changes, suggesting that second-order phase-space dynamics are particularly beneficial for complex low-dimensional structures. On another three datasets, VTV-FM is not the best method but remains competitive. On Tree, it is only about 0.8% worse than the best HOFM result. This suggests that high-order or OT-based path modeling can still be effective for certain geometries. However, in terms of average performance, VTV-FM shows the most robust behavior across low-dimensional structures. Additional controlled analyses of the source phase-space prior, terminal-velocity completion, pairing–bridge consistency, and computational efficiency on the low-dimensional benchmarks are provided in Appendices F.2–F.5.

4.2 PDE Case Studies in the Observed-v_1 Regime

We evaluate VTV-FM on PDE-governed generation tasks. Unlike two-dimensional datasets and CIFAR-10, PDE data provide explicit temporal evolution, so the terminal derivative can be obtained from the governing equation or numerical simulator. This corresponds to the observed-v_1 regime, where VTV-FM can use the full-boundary second-order bridge without terminal-velocity closure.

Datasets. We consider two two-dimensional PDE datasets: Navier–Stokes vorticity generation (NS2D) and Kuramoto–Sivashinsky generation (KS2D). For NS2D, the state variable is the vorticity field ω, governed by $\partial_t \omega + u \cdot \nabla \omega = \nu \Delta \omega + f$. For KS2D, the state variable is a scalar field u,

governed by $\partial_t u + \frac{1}{2}|\nabla u|^2 + \Delta u + \Delta^2 u = 0$. In both datasets, the terminal derivative can be obtained from the PDE right-hand side or simulator outputs.

Baselines. We compare VTV-FM with OT-CFM and HOFM. OT-CFM is the stronger first-order OT-based baseline, while HOFM is a competitive high-order baseline. This comparison evaluates whether terminal derivative information and the full-boundary bridge improve second-order generation.

Metrics. We evaluate PDE generation quality using MMD, PSD log-L_2, and a derivative-sensitive gap. MMD [23] measures global distributional discrepancy, PSD log-L_2 measures multiscale spectral fidelity, and the derivative-sensitive gap compares the distributions of per-sample L_2 norms of PDE-related quantities. For NS2D, we use $\nu\Delta\omega$; for KS2D, we use the PDE right-hand side.

Results and findings. Table 2 reports the PDE generation results on NS2D and KS2D. VTV-FM achieves the best result on both PDE datasets across all metrics, improving global distributional matching, spectral fidelity, and PDE-related derivative structure.

Table 2: PDE generation results.

Dataset	Method	MMD	PSD log-L_2	Deriv. gap
NS2D	OT-CFM	0.017	5.08	6.93×10^{-3}
	HOFM	0.015	4.22	1.06×10^{-3}
	VTV-FM	**0.014**	**3.48**	$\mathbf{2.80\times10^{-4}}$
KS2D	OT-CFM	0.035	9.00	5.50×10^{-1}
	HOFM	0.015	8.13	9.17×10^{-2}
	VTV-FM	**0.010**	**7.52**	$\mathbf{1.30\times10^{-2}}$

On NS2D, VTV-FM reduces MMD from 0.017 (OT-CFM) and 0.015 (HOFM) to 0.014, corresponding to relative reductions of approximately 17.6% and 6.7%, respectively. For spectral structure, it reduces PSD log-L_2 from 5.08 and 4.22 to 3.48, giving reductions of approximately 31.5% over OT-CFM and 17.5% over HOFM. The largest gain appears in the derivative-sensitive gap, VTV-FM reduces it from 6.93×10^{-3} and 1.06×10^{-3} to 2.80×10^{-4}, corresponding to reductions of approximately 96.0% and 73.6%. This indicates that VTV-FM better preserves diffusion-related derivative structure.

On KS2D, VTV-FM also consistently improves over both baselines. It reduces MMD from 0.035 (OT-CFM) and 0.015 (HOFM) to 0.010, yielding relative reductions of approximately 71.4% and 33.3%, respectively. For PSD log-L_2, VTV-FM reduces the value from 9.00 and 8.13 to 7.52, corresponding to reductions of approximately 16.4% and 7.5%. For the derivative-sensitive gap, it reduces the value from 5.50×10^{-1} and 9.17×10^{-2} to 1.30×10^{-2}, giving reductions of approximately 97.6% and 85.8%. This suggests that VTV-FM improves not only overall distribution matching, but also the PDE-related dynamical structure of the KS system.

Overall, the most pronounced improvements appear in the derivative-sensitive gap on both PDE systems. These results show that VTV-FM can directly exploit terminal derivative information and unify it with the full-boundary bridge geometry and trajectory supervision.

4.3 CIFAR-10 Case Study in the Missing-v_1 Regime

We next evaluate VTV-FM on CIFAR-10, a standard image-generation benchmark where target-side terminal derivatives are unavailable. This corresponds to the missing-v_1 regime: only target samples are observed, and the second-order bridge must be instantiated using the terminal-velocity closure derived in Section 3. This experiment tests whether the proposed construction remains effective in high-dimensional image generation.

Baselines and evaluation. We compare VTV-FM with representative Flow Matching baselines, including FM, I-CFM, OT-CFM, OAT-FM, HOFM, and CurveFlow. FM and I-CFM use independent source–target coupling, while OT-CFM uses minibatch OT pairing. OAT-FM further refines first-order couplings through acceleration transport, whereas HOFM and CurveFlow provide high-order or curve-based path-modeling baselines. Because these methods use different dynamical orders and native sampling procedures, we report both solver-controlled comparisons and method-specific inference protocols. For the controlled cross-method comparison, VTV-FM uses acceleration matching only, without the local consistency or ADM regularizers.

Solver-controlled comparison. Table 3 separates method design from solver choice. Under the shared Heun-100 protocol with 200 NFE, FM, OT-CFM, and VTV-FM obtain FIDs of 3.62, 3.48, and 3.17, respectively. Under the same adaptive Dopri5 tolerance, VTV-FM reaches an FID of 3.23 with 81 NFE, compared with 3.73 with 147 NFE for FM and 3.60 with 134 NFE for OT-CFM. These

Table 3: CIFAR-10 FID under controlled inference protocols. For adaptive Dopri5, the measured number of function evaluations (NFE) is shown in parentheses.

Method	Euler-100	Dopri5	Heun-100	Taylor2-100	Verlet-100
FM	4.58	3.73 (147)	3.62	–	–
I-CFM	4.48	3.68 (146)	3.59	–	–
OT-CFM	4.46	3.60 (134)	3.48	–	–
FM + OAT-FM	3.91	3.56 (135)	3.51	–	–
I-CFM + OAT-FM	3.77	3.50 (138)	3.45	–	–
OT-CFM + OAT-FM	3.72	3.47 (126)	3.42	–	–
CurveFlow	3.47	3.41 (130)	3.40	–	–
HOFM	–	–	–	3.45	–
VTV-FM (Acc. only)	**3.28**	**3.23 (81)**	**3.17**	**3.25**	**3.10**

results show that the improvement persists under shared general-purpose solvers and therefore cannot be attributed solely to Velocity Verlet, auxiliary regularization, or a larger sampling budget.

Role of the auxiliary objectives. We next evaluate the effect of local consistency and acceleration-divergence matching (ADM) within VTV-FM. Table 4 reports the results under three fixed-step phase-space solvers. Acceleration matching alone already achieves FIDs of 3.10, 3.17, and 3.25 with Velocity Verlet, Heun, and Taylor2, respectively. Adding local consistency improves these values to 3.06, 3.16, and 3.23. Adding ADM alone gives 3.05, 3.14, and 3.19, showing a larger individual improvement than local consistency under all three solvers. Combining both auxiliary terms gives the best overall results, including an FID of 2.99 with 100-step Velocity Verlet sampling.

Here, $\mathrm{Cons.} = \mathcal{L}_{\mathrm{va}} + \mathcal{L}_{\mathrm{xv}}$. The ablation indicates that the main gain is already present with bridge-induced acceleration supervision, while local consistency and ADM provide complementary refinements. ADM yields the larger individual gain, and combining both auxiliary terms gives the strongest overall performance. Accordingly, we treat acceleration matching as the core learning objective and the consistency and ADM terms as optional refinements.

Table 4: VTV-FM objective ablation.

Objective	Verlet	Heun	Taylor2
Acc.	3.10	3.17	3.25
Acc. + Cons.	3.06	3.16	3.23
Acc. + ADM	3.05	3.14	3.19
Acc. + Cons. + ADM	**2.99**	**3.13**	**3.18**

Results and findings. The controlled comparisons show that VTV-FM improves over first-order FM and OT-based FM even when solver choice is held fixed. In particular, under Heun-100, VTV-FM reduces FID from 3.62 for FM and 3.48 for OT-CFM to 3.17. Under adaptive Dopri5, VTV-FM also achieves a lower FID with fewer function evaluations. Within VTV-FM, the full objective further improves the best fixed-step result from 3.10 with acceleration matching alone to 2.99 when both consistency and ADM are included.

These results support two complementary conclusions. First, the improvement of VTV-FM is not a solver-specific effect: it persists under shared Heun and Dopri5 inference protocols. Second, the auxiliary regularizers refine an already strong acceleration-matching model rather than accounting for the primary gain. Together, the results show that the variationally induced phase-space construction remains effective in the high-dimensional missing-v_1 regime. Qualitative CIFAR-10 samples are provided in Appendix E.3. Additional ImageNet-256 scaling results are reported in Appendix F.5.

5 Related Work

Flow Matching and OT-based coupling. Flow Matching learns generative transport by regressing time-dependent vector fields along prescribed probability paths, with stochastic interpolants and rectified flows providing related formulations [1, 2, 5]. Minibatch optimal transport improves source–target coupling and simplifies transport paths and regression targets [3, 4]. In OT-CFM with linear conditional bridges, optimal transport improves sample pairing, but each conditional trajectory remains a straight line between the paired endpoints.

Higher-order dynamics and path design. Beyond improving sample coupling, existing methods enrich generative dynamics through higher-order supervision and alternative dynamical construc-

tions. HOMO introduces acceleration and higher-order derivative supervision into one-step shortcut models [6], and NRFlow jointly learns velocity and acceleration fields to improve noise robustness and trajectory stability [7]. HOFM develops a unified framework incorporating arbitrary-order trajectory derivatives, with conditional regression formulations, derivative consistency constraints, and statistical convergence rates [8]. Related theory also studies estimation errors in higher-order trajectory refinement [24]. In terms of dynamical structure and path geometry, HRF models velocity, acceleration, and higher-order variables through hierarchically coupled rectified flows [25], and CurveFlow explores curvature-guided path objectives [9]. For few-step generation, Terminal Velocity Matching matches the derivative of a two-time displacement model with respect to its terminal time [26]; this derivative target is distinct from the data-side endpoint velocity considered here.

Minimum-acceleration transport and boundary conditions. A closely related variational approach uses squared-acceleration action to construct trajectories and transport costs. Classical spline theory shows that minimizing this action with fixed endpoint positions and velocities yields a cubic Hermite curve [14]. At the distribution level, second-order optimal transport studies acceleration-minimizing paths and cubic Wasserstein splines [27]. OAT-FM applies optimal acceleration transport to flow matching [11], and OTIKIN studies second-order discrepancies between distributions induced by acceleration costs [28]. Endpoint information also plays an important role: boundary-constrained flow matching imposes prescribed endpoint behavior on learned first-order velocity fields [10]. Our setting concerns second-order phase-space bridges whose terminal velocity is not provided by static data. VTV-FM treats this unspecified velocity as an optimization variable in the minimum-acceleration problem and eliminates it analytically, jointly determining the terminal-velocity closure, bridge, corresponding OT pairing cost, and acceleration supervision.

6 Conclusion

We introduced **VTV-FM**, a second-order FM framework for generative transport in phase space. The central challenge is that second-order bridges require a target-side terminal velocity, which is often unavailable for static data. VTV-FM addresses this issue with a minimum-acceleration variational principle: when the terminal velocity is missing, the principle yields an analytic terminal-velocity closure and a projected OT cost; when the terminal velocity is observed or derivable, the same full-boundary bridge directly incorporates it. This construction unifies terminal-velocity closure, source–target pairing, and acceleration supervision under a single variational framework. Experiments on low-dimensional datasets, PDE-governed fields, and CIFAR-10 show that VTV-FM improves generation quality over first-order, OT-based, and higher-order FM baselines, while producing more reliable second-order transport trajectories. Future work includes developing richer kinetic priors, extending the framework to conditional and controlled generation, and scaling second-order transport constructions to larger generative modeling tasks.

Acknowledgements and Disclosures

Funding. The work done by H. Jiang and Y. He was supported in part by the National Science Foundation (NSF) under Grant Nos. IIS-2544353 (CAREER), IIS-2505719, IIS-2441449, and IOS-2446522, and in part by the Commonwealth Cyber Initiative (CCI). The work done by Y. Li and H. Chen was supported by the NSF under IIS-2348405, the Coastal Virginia Center for Cyber Innovation (COVA CCI), CCI, the William & Mary Faculty Research Grant, and generous gifts from NVIDIA and Amazon. And the work done by D. Yang and Y. Xiong was supported by the COVA CCI and the CCI.

Competing interests. The authors declare no financial relationships or competing interests outside the submitted work that could reasonably be perceived as influencing this work.

References

[1] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow Matching for Generative Modeling. In *ICLR*, 2023.

[2] 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.

[3] Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and Generalizing Flow-Based Generative Models with Minibatch Optimal Transport. *Transactions on Machine Learning Research*, March 2024.

[4] Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample Flow Matching: Straightening Flows with Minibatch Couplings. In *ICML*, volume 202, pages 28100–28127, 2023.

[5] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. In *ICLR*, 2023.

[6] Bo Chen, Chengyue Gong, Xiaoyu Li, Yingyu Liang, Zhizhou Sha, Zhenmei Shi, Zhao Song, and Mingda Wan. High-Order Matching for One-Step Shortcut Diffusion Models. In *ICLRW*, 2025. DeLTa Workshop.

[7] Bo Chen, Chengyue Gong, Xiaoyu Li, Yingyu Liang, Zhizhou Sha, Zhenmei Shi, Zhao Song, Mingda Wan, and Xugang Ye. NRFlow: Towards noise-robust generative modeling via high-order mechanism. In *UAI*, volume 286, pages 673–704, 2025.

[8] Maojiang Su, Jerry Yao-Chieh Hu, Yi-Chen Lee, Ning Zhu, Jui-Hui Chung, Shang Wu, Zhao Song, Minshuo Chen, and Han Liu. High-Order Flow Matching: Unified Framework and Sharp Statistical Rates. In *NeurIPS*, volume 38, pages 51192–51304, 2025.

[9] Yan Luo, Drake Du, Hao Huang, Yi Fang, and Mengyu Wang. CurveFlow: Curvature-guided flow matching for image generation. In *CVPRW*, pages 9020–9029, 2026.

[10] Xixi Hu, Runlong Liao, Keyang Xu, Bo Liu, Yeqing Li, Eugene Ie, Hongliang Fei, and Qiang Liu. Improving Rectified Flow with Boundary Conditions. In *ICCV*, pages 18177–18186, 2025.

[11] Angxiao Yue, Anqi Dong, and Hongteng Xu. OAT-FM: Optimal acceleration transport for improved flow matching. *arXiv preprint arXiv:2509.24936*, 2025.

[12] Neta Shaul, Juan C. Pérez, Ricky T. Q. Chen, Ali K. Thabet, Albert Pumarola, and Yaron Lipman. Bespoke Solvers for Generative Flow Models. In *ICLR*, 2024.

[13] Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *NeurIPS*, volume 26, pages 2292–2300, 2013.

[14] Carl de Boor. *A Practical Guide to Splines*, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1978.

[15] Yuhao Huang, Taos Transue, Shih-Hsin Wang, William M. Feldman, Hong Zhang, and Bao Wang. Improving Flow Matching by Aligning Flow Divergence. In *ICML*, volume 267, pages 25813–25834, 2025.

[16] Ernst Hairer, Syvert P. Nørsett, and Gerhard Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*, volume 8 of *Springer Series in Computational Mathematics*. Springer, Berlin, second edition, 1993.

[17] John Charles Butcher. *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons, third edition, 2016.

[18] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *The Journal of Chemical Physics*, 76(1):637–649, 1982.

[19] Phillip Lippe, Bastiaan S. Veeling, Paris Perdikaris, Richard E. Turner, and Johannes Brandstetter. PDE-Refiner: Achieving accurate long rollouts with neural PDE solvers. In *NeurIPS*, volume 36, pages 67398–67433, 2023.

[20] Alex Krizhevsky. Learning Multiple Layers of Features from Tiny Images. Technical report, University of Toronto, 2009.

[21] Frederik Ebert, Chelsea Finn, Alex X. Lee, and Sergey Levine. Self-Supervised Visual Planning with Temporal Skip Connections. In *CoRL*, volume 78, pages 344–356, 2017.

[22] Christian Schuldt, Ivan Laptev, and Barbara Caputo. Recognizing Human Actions: A Local SVM Approach. In *ICPR*, volume 3, pages 32–36, 2004.

[23] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.

[24] Chengyue Gong, Xiaoyu Li, Yingyu Liang, Jiangxuan Long, Zhenmei Shi, Zhao Song, and Yu Tian. Theoretical Guarantees for High Order Trajectory Refinement in Generative Flows. *arXiv preprint arXiv:2503.09069*, 2025.

[25] Yichi Zhang, Yici Yan, Alex Schwing, and Zhizhen Zhao. Towards Hierarchical Rectified Flow. In *ICLR*, 2025.

[26] Linqi Zhou, Mathias Parger, Ayaan Haque, and Jiaming Song. Terminal velocity matching. In *ICLR*, 2026.

[27] Jean-David Benamou, Thomas O. Gallouët, and François-Xavier Vialard. Second-order models for optimal transport and cubic splines on the Wasserstein space. *Foundations of Computational Mathematics*, 19(5):1113–1143, 2019.

[28] Giovanni Brigati, Jan Maas, and Filippo Quattrocchi. Kinetic Optimal Transport (OTIKIN) – Part 1: Second-Order Discrepancies Between Probability Measures. *arXiv preprint arXiv:2502.15665*, 2025.

[29] 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 *ECCV*, pages 23–40, 2024.

[30] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

Technical Appendices and Supplementary Material

A Limitations. **14**

B Broader impacts. **14**

C Core Derivations and Algorithm Details **14**

C.1 Proof of Proposition 3.1 14

C.2 Proof of Theorem 3.2: Variational Terminal-Velocity Projection 14

C.3 Mini-batch Pairing and Supervision Details 15

C.4 Local Integral Consistency 15

C.5 Acceleration Divergence Matching 16

C.6 Numerical Solvers 17

D Implementation Details **17**

D.1 Low-Dimensional Experiments 17

D.2 Shared Source Phase-Space and Training Setup 17

D.3 PDE-Governed Field Experiments 18

D.4 CIFAR-10 Experiments 18

E Additional Qualitative Visualizations **18**

E.1 Low-Dimensional Trajectory Visualizations 18

E.2 Additional PDE Visualization 20

E.3 Additional CIFAR-10 Samples 20

E.4 Additional Video Visualization 20

F Additional Controlled Experiments **20**

F.1 Loss and Solver Analysis on Low-Dimensional Datasets 20

F.2 Source Phase-Space Prior and Sensitivity Analysis 21

F.3 Terminal-Velocity Ablation Study 21

F.4 Pairing–Bridge Cross-Ablation Study 22

F.5 Computational Efficiency and ImageNet-256 Scaling 22

G VTV-FM as an Extension of OT-CFM **23**

G.1 OT-CFM as First-Order Variational Flow Matching 23

G.2 Second-Order Lift to Phase Space 23

G.3 Projection under Missing Terminal Velocity 23

G.4 Bridge-Shape Ablation: Minimum Acceleration versus Minimum Jerk 24

A Limitations.

VTV-FM relies on a minimum-acceleration principle to close the missing terminal velocity in static-data regimes. Although this gives an analytic and internally consistent bridge, richer kinetic structures or multimodal terminal velocities may require more flexible closures. Our experiments include low-dimensional datasets, PDE-governed fields, CIFAR-10, an ImageNet-256 scaling study, and qualitative video examples. Broader evaluation on higher-resolution image generation, quantitative video generation, and conditional generation remains future work.

B Broader impacts.

This work is primarily methodological and may benefit scientific generation, physical simulation, and data-driven modeling by improving transport geometry and derivative-aware generation. As with other generative modeling methods, improved sample quality may also be misused for synthetic image or video generation. The paper does not propose a deployed system or use private personal data, and our experiments are limited to standard benchmarks and simulated physical fields.

C Core Derivations and Algorithm Details

We prove the full-boundary bridge and terminal-velocity projection (subsections C.1 and C.2), then provide minibatch pairing, auxiliary-loss, and sampling details (subsections C.3, C.4, C.5, and C.6) for Section 3.

C.1 Proof of Proposition 3.1

Proof. Let $\Delta x = x_1 - x_0$ and define

$$\gamma^\star(t) = x_0 + v_0 t + \beta t^2 + \alpha t^3, \qquad \alpha = \frac{-2\Delta x + T(v_0 + v_1)}{T^3}, \qquad \beta = \frac{3\Delta x - T(2v_0 + v_1)}{T^2}.$$

This polynomial satisfies all four boundary conditions and is equivalent to the Hermite representation in Eq. (2). Its derivatives give Eq. (3).

For any other admissible $\gamma \in \Gamma_T(x_0, v_0, x_1, v_1)$, write $r = \gamma - \gamma^\star$. Then $r(0) = r(T) = \dot{r}(0) = \dot{r}(T) = 0$. Since $(\gamma^\star)^{(4)} = 0$, integration by parts gives

$$\int_0^T \langle \ddot{\gamma}^\star, \ddot{r} \rangle \, dt = 0.$$

Consequently,

$$\mathcal{A}_T[\gamma] = \mathcal{A}_T[\gamma^\star] + \int_0^T \|\ddot{r}(t)\|_2^2 \, dt \geq \mathcal{A}_T[\gamma^\star].$$

Equality requires $\ddot{r} = 0$ almost everywhere; the homogeneous boundary conditions then imply $r = 0$. Thus $\gamma^\star$ is the unique global minimizer in the admissible Sobolev class.

Finally, $\ddot{\gamma}^\star(t) = 6\alpha t + 2\beta$ yields

$$\mathcal{A}_T[\gamma^\star] = 12T^3\|\alpha\|_2^2 + 12T^2\langle \alpha, \beta \rangle + 4T\|\beta\|_2^2.$$

Substituting α, β and completing the square gives Eq. (4). □

C.2 Proof of Theorem 3.2: Variational Terminal-Velocity Projection

Proof. Completing the square in Eq. (4) gives

$$C_T(x_0, v_0, x_1, v_1) = \frac{3}{T^3}\|x_1 - x_0 - Tv_0\|_2^2 + \frac{4}{T}\left\| v_1 - \left(\frac{3}{2T}(x_1 - x_0) - \frac{1}{2}v_0 \right) \right\|_2^2.$$

Since $T > 0$, the second term has a unique minimizer in v_1. This proves the closure and projected cost in Eqs. (5)–(6).

To establish the coupling-level statement, write $s = (x_0, v_0, x_1)$ and let $\eta \in \Pi(\mu_0, \rho_1)$ be an observable coupling. Define its full-boundary lifts by

$$\mathcal{L}(\eta) = \{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{V} \times \mathcal{X} \times \mathcal{V}) : p_{\#}\pi = \eta\}, \qquad p(s, v_1) = s.$$

For any $\pi \in \mathcal{L}(\eta)$, the completed-square identity yields

$$\int C_T(s, v_1)\, d\pi = \int C_T^\star(s)\, d\eta + \frac{4}{T} \int \|v_1 - v_1^\star(s)\|_2^2\, d\pi \geq \int C_T^\star(s)\, d\eta.$$

Equality is attained by the deterministic lift $\pi^\star = (\mathrm{id}, v_1^\star)_{\#}\eta$, which assigns $v_1 = v_1^\star(s)$ to each observable tuple. Hence

$$\inf_{\pi \in \mathcal{L}(\eta)} \int C_T\, d\pi = \int C_T^\star\, d\eta.$$

Taking the infimum over observable couplings gives

$$\inf_{\eta \in \Pi(\mu_0, \rho_1)} \int C_T^\star\, d\eta = \inf_{\eta \in \Pi(\mu_0, \rho_1)} \inf_{\pi \in \mathcal{L}(\eta)} \int C_T\, d\pi,$$

which proves the exact variational projection. □

C.3 Mini-batch Pairing and Supervision Details

For the cost matrix C in Eq. (7), we use uniform minibatch weights $a_i = 1/m$ and $b_j = 1/n$. The entropic coupling solves

$$\min_{\substack{P \geq 0 \\ P\mathbf{1} = a,\ P^\top \mathbf{1} = b}} \sum_{i,j} P_{ij} C_{ij} + \varepsilon \sum_{i,j} P_{ij}(\log P_{ij} - 1).$$

With $K_{ij} = \exp(-C_{ij}/\varepsilon)$, Sinkhorn iterations give

$$u \leftarrow \frac{a}{Kw}, \qquad w \leftarrow \frac{b}{K^\top u}, \qquad P^\star = \mathrm{diag}(u) K\, \mathrm{diag}(w),$$

where divisions are elementwise. These updates may equivalently be performed in the log domain for numerical stability.

Matched pairs are drawn jointly from $P^\star$, or by sampling $i \sim a$ followed by

$$j \sim q(\cdot \mid i), \qquad q(j \mid i) = \frac{P_{ij}^\star}{a_i} = m P_{ij}^\star.$$

For each pair, Eq. (8) supplies the terminal velocity. Sampling $t \sim \mathrm{Unif}(0, T)$ and evaluating Eqs. (2) and (3) then gives $(x_t^\star, v_t^\star, t/T, a_t^\star)$ for acceleration matching. The complete training procedure is given in Algorithm 1.

C.4 Local Integral Consistency

For a variational bridge and $0 \leq t \leq T - \Delta t$, its exact increments satisfy

$$v_{t+\Delta t}^\star - v_t^\star = \int_t^{t+\Delta t} a_s^\star\, ds,$$

$$x_{t+\Delta t}^\star - x_t^\star - \Delta t\, v_t^\star = \int_t^{t+\Delta t} (t + \Delta t - s) a_s^\star\, ds.$$

Approximating each integral using the learned acceleration at its left endpoint gives

$$v_{t+\Delta t}^\star - v_t^\star \approx \Delta t\, a_\theta(x_t^\star, v_t^\star, t/T),$$

$$x_{t+\Delta t}^\star - x_t^\star - \Delta t\, v_t^\star \approx \frac{1}{2} \Delta t^2 a_\theta(x_t^\star, v_t^\star, t/T).$$

The expected squared residuals are precisely $\mathcal{L}_{\mathrm{va}}$ and $\mathcal{L}_{\mathrm{xv}}$ in Eqs. (10a)–(10b). They supplement instantaneous acceleration matching with finite-step increment supervision. Because the acceleration varies along a cubic bridge, these are local approximations rather than exact integration constraints. We use $\Delta t = T/N$, as specified in Appendix D.

C.5 Acceleration Divergence Matching

The phase-space drift is $b_\theta(x, v, t) = (v, a_\theta(x, v, t/T))$. Since v is an independent phase-space coordinate, $\nabla_x \cdot v = 0$, and therefore

$$\nabla_{(x,v)} \cdot b_\theta = \nabla_v \cdot a_\theta.$$

This identifies the velocity divergence of acceleration as the compressibility term for the learned dynamics.

Path-error motivation. Let $F_t^\star(x, v) = \mathbb{E}[a_t^\star \mid x_t^\star = x, v_t^\star = v]$ be the marginal reference acceleration field. Under sufficient regularity, the reference and learned densities $\mu_t^\star, \mu_t^\theta$ satisfy their phase-space continuity equations. For $\eta_t = \mu_t^\theta - \mu_t^\star$, subtraction gives

$$\partial_t \eta_t + \nabla_{(x,v)} \cdot (b_\theta \eta_t) = \nabla_v \cdot \big[(F_t^\star - a_\theta)\mu_t^\star\big].$$

The forcing term expands as

$$(F_t^\star - a_\theta) \cdot \nabla_v \mu_t^\star + \mu_t^\star \big(\nabla_v \cdot F_t^\star - \nabla_v \cdot a_\theta\big).$$

Thus, both acceleration and divergence mismatch contribute to the path error, motivating divergence regularization alongside acceleration matching [15].

Conditional continuation targets. Let $\Delta = T - t > 0$. Applying Proposition 3.1 and Theorem 3.2 to the remaining interval gives the current acceleration for a bridge from (x, v) to position y:

$$F_{\text{free}}^\star(x, v, t; y) = \frac{3}{\Delta^2}(y - x - \Delta v),$$

when the terminal velocity is free. If the terminal velocity is fixed at w, the corresponding acceleration is

$$F_{\text{fix}}^\star(x, v, t; y, w) = \frac{6}{\Delta^2}(y - x - \Delta v) - \frac{2}{\Delta}(w - v).$$

Taking the velocity divergence with the terminal data held fixed gives

$$\nabla_v \cdot F_{\text{free}}^\star = -\frac{3d}{\Delta}, \qquad \nabla_v \cdot F_{\text{fix}}^\star = -\frac{4d}{\Delta}.$$

Normalizing by Δ/d yields the auxiliary objective

$$\mathcal{L}_{\text{ADM}} = \mathbb{E}\left[\left(\frac{T-t}{d}\nabla_v \cdot a_\theta(x_t^\star, v_t^\star, t/T) + c\right)^2\right], \qquad c = \begin{cases} 3, & \text{latent } v_1, \\ 4, & \text{observed } v_1. \end{cases}$$

The latent-v_1 case is Eq. (11).

Conditional and marginal divergence. The constants above are exact for conditional continuation fields. The marginal reference field averages over terminal data conditioned on the current state. For example, in the latent-v_1 case, let $m_t(x, v) = \mathbb{E}[x_1 \mid x_t^\star = x, v_t^\star = v]$. Then

$$F_t^\star(x, v) = \frac{3}{\Delta^2}\big(m_t(x, v) - x - \Delta v\big),$$

and, wherever differentiable,

$$\nabla_v \cdot F_t^\star = -\frac{3d}{\Delta} + \frac{3}{\Delta^2}\nabla_v \cdot m_t(x, v).$$

Accordingly, ADM regularizes the learned field toward the contraction of the conditional bridge. Exact marginal-divergence matching would also require the state dependence of the conditional terminal mean. The same conditional–marginal distinction applies when terminal velocities are observed.

Trace estimation. For low-dimensional data, we compute $\nabla_v \cdot a_\theta = \operatorname{Tr}(\partial a_\theta / \partial v)$ directly by automatic differentiation. For high-dimensional data, we use a Hutchinson estimator

$$\nabla_v \cdot a_\theta \approx \epsilon^\top \frac{\partial a_\theta}{\partial v}\epsilon, \qquad \mathbb{E}[\epsilon\epsilon^\top] = I,$$

with a standard Gaussian or Rademacher vector ϵ. The quadratic form is evaluated using a vector–Jacobian product without materializing the full Jacobian.

C.6 Numerical Solvers

Sampling integrates

$$\dot{z}(t) = F_\theta(z(t), t), \qquad z = (x, v), \qquad F_\theta(z, t) = \begin{pmatrix} v \\ a_\theta(x, v, t/T) \end{pmatrix}.$$

Let $h = T/N$, $t_k = kh$, and $a_k = a_\theta(x_k, v_k, t_k/T)$.

Heun solver. The predictor and corrected update are

$$\widetilde{z}_{k+1} = z_k + hF_\theta(z_k, t_k), \qquad z_{k+1} = z_k + \frac{h}{2}\left[F_\theta(z_k, t_k) + F_\theta(\widetilde{z}_{k+1}, t_{k+1})\right].$$

Taylor2 update. The update used in our comparisons is

$$x_{k+1} = x_k + hv_k + \frac{1}{2}h^2 a_k, \qquad v_{k+1} = v_k + ha_k.$$

Here *Taylor2* denotes the second-order expansion of the position update combined with an Euler velocity update. For a general learned acceleration field, this scheme is not second-order accurate for the full phase-space state.

Velocity Verlet update. We first predict the position and velocity:

$$x_{k+1} = x_k + hv_k + \frac{1}{2}h^2 a_k, \qquad \overline{v}_{k+1} = v_k + ha_k.$$

We then evaluate the acceleration at the predicted state and correct the velocity:

$$\widehat{a}_{k+1} = a_\theta(x_{k+1}, \overline{v}_{k+1}, t_{k+1}/T), \qquad v_{k+1} = v_k + \frac{h}{2}(a_k + \widehat{a}_{k+1}).$$

For velocity-dependent acceleration, this is an explicit predictor–corrector variant of the Verlet update. At the next step, a_{k+1} is evaluated at the corrected state (x_{k+1}, v_{k+1}).

D Implementation Details

This appendix summarizes the implementation and evaluation settings for the low-dimensional, PDE, and CIFAR-10 experiments. We report both method-specific and solver-controlled inference protocols, with solver choices and sampling budgets specified for each experiment.

D.1 Low-Dimensional Experiments

For the low-dimensional datasets, all methods are evaluated with the same sampling budget of 100 steps and EMA weights. For each dataset, we use $n = 2000$ target samples and $n = 2000$ generated samples to compute the sample-based W_2 metric. CFM, OT-CFM, OAT-FM, and CurveFlow are evaluated with Euler sampling, following their standard first-order sampling protocol. HOFM is evaluated with the Taylor2 solver, following its original high-order setting. For VTV-FM, we evaluate Heun, Taylor2, and Velocity Verlet, and report the best validation result in the main table. The solver-specific results are reported separately in the ablation table.

D.2 Shared Source Phase-Space and Training Setup

Source phase-space prior. For the low-dimensional and CIFAR-10 experiments, we sample

$$v_0 \sim \mathcal{N}(0, \sigma^2 I),$$

independently of the source position x_0. Unless otherwise stated, we use $\sigma = 1$. For static data, v_0 is a modeling prior used to provide a non-degenerate lifting from position space to phase space; it is not interpreted as an observed or physical velocity. The variational terminal-velocity closure itself does not depend on the Gaussian assumption.

The resulting projected cost is

$$C_T^\star(x_0, v_0, x_1) = \frac{3}{T^3} \|x_1 - x_0 - Tv_0\|_2^2 .$$

When $v_0 = 0$, this cost is proportional to the standard quadratic OT cost, so their unregularized optimal couplings coincide. The induced position trajectory follows a cubic time parameterization rather than a constant-velocity linear bridge. The exact constant-velocity trajectory is recovered only when $v_0 = (x_1 - x_0)/T$, for which the residual $x_1 - x_0 - Tv_0$ and the required acceleration vanish.

Loss coefficients and local consistency window. The full training objective is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{acc}} + \lambda_{\text{va}}\mathcal{L}_{\text{va}} + \lambda_{\text{xv}}\mathcal{L}_{\text{xv}} + \lambda_{\text{ADM}}\mathcal{L}_{\text{ADM}}.$$

Unless otherwise stated, we use

$$\lambda_{\text{va}} = \lambda_{\text{xv}} = \lambda_{\text{ADM}} = 0.1,$$

with coefficient 1 on $\mathcal{L}_{\text{acc}}$. For the two local consistency losses, the local time window is matched to the actual fixed sampling step,

$$\Delta t = \frac{T}{N},$$

and is not tuned as an additional hyperparameter.

D.3 PDE-Governed Field Experiments

For PDE-governed field generation, all methods are evaluated with 100 sampling steps. Unlike static image data, PDE data provide terminal derivative information through the PDE right-hand side or simulator outputs. Therefore, VTV-FM directly uses the observed or derivable terminal velocity and does not apply terminal-velocity closure in this setting. This corresponds to the observed/derivable-v_1 regime described in the main text.

The baseline solver choices follow the same principle as in the low-dimensional experiments. First-order FM variants use their standard first-order samplers, HOFM uses the Taylor2 solver, and VTV-FM is evaluated with phase-space solvers such as Heun, Taylor2, and Velocity Verlet. Other evaluation settings, including sample size and metric computation, follow the low-dimensional experiment protocol unless otherwise specified.

D.4 CIFAR-10 Experiments

CIFAR-10 represents the high-dimensional missing-v_1 regime, where only terminal images are observed and target-side terminal derivatives are unavailable. Therefore, VTV-FM uses the variational terminal-velocity closure derived in Section 3 to instantiate the second-order bridge.

For the controlled cross-method comparison in Table 3, we evaluate FM, I-CFM, OT-CFM, OAT-FM, CurveFlow, and VTV-FM under the applicable shared inference protocols. VTV-FM uses acceleration matching only in this comparison, without the local consistency or ADM regularizers. Heun-100 provides the primary fixed-step solver-controlled comparison, while Dopri5 provides a shared adaptive-solver comparison with measured NFE.

Table 4 separately evaluates the effect of the auxiliary objectives within VTV-FM under Velocity Verlet, Heun, and Taylor2. This separates the cross-method comparison from the contribution of local consistency and acceleration-divergence matching.

E Additional Qualitative Visualizations

This appendix complements the quantitative results in Section 4 with transport trajectories and generated samples.

E.1 Low-Dimensional Trajectory Visualizations

Figure 2 compares transport geometry on six representative datasets, complementing the sample-based W_2 results in Section 4.1.

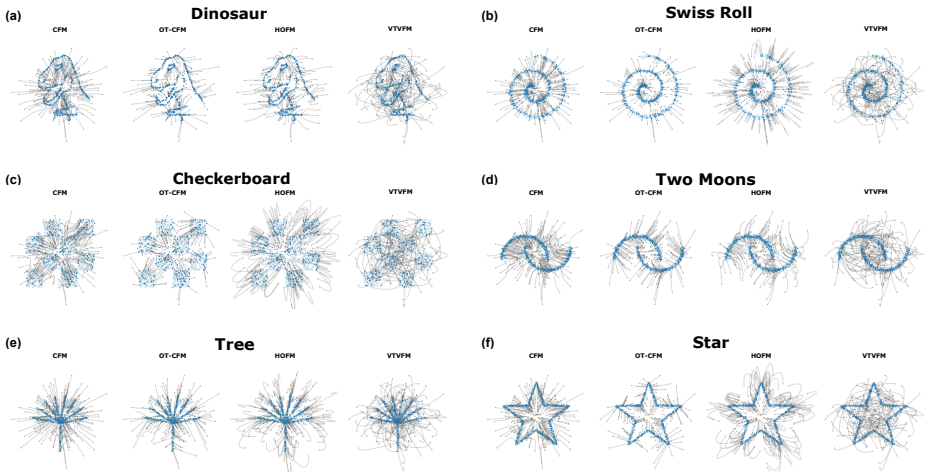
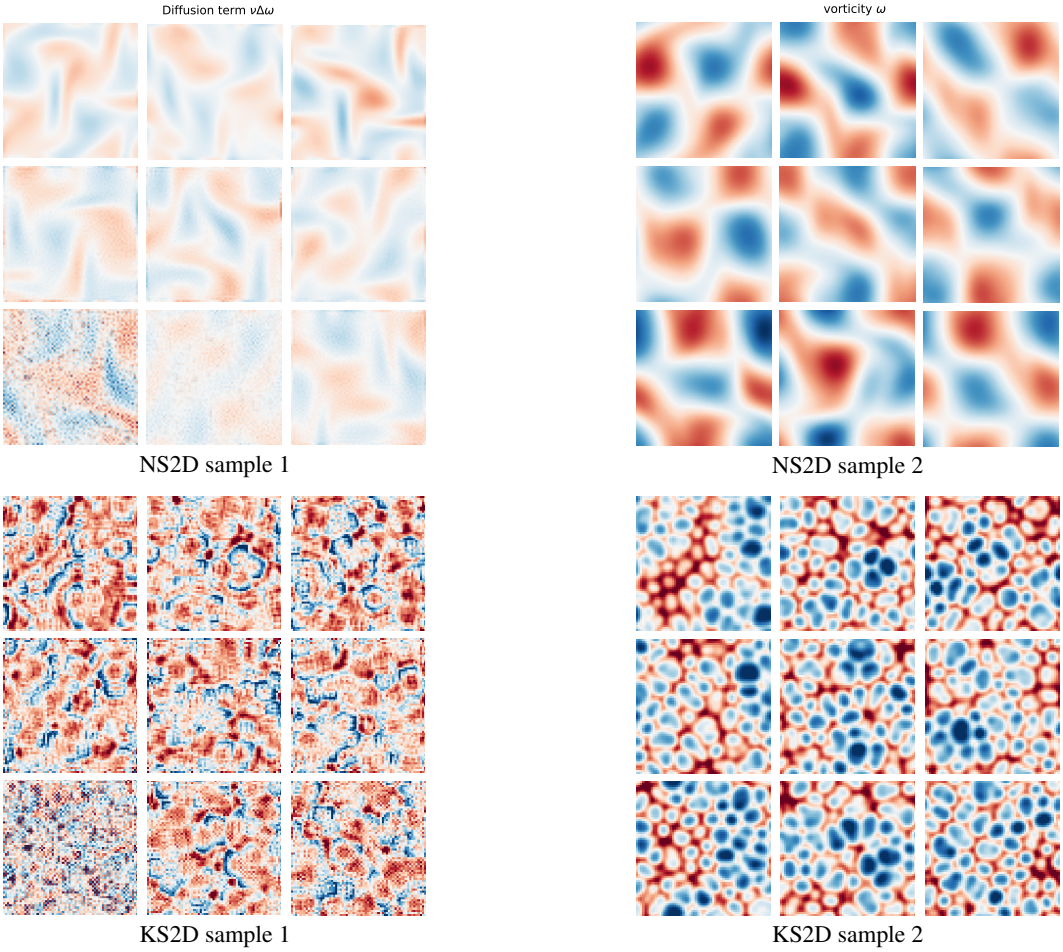


Figure 2: Transport trajectories on six low-dimensional datasets. Each panel compares CFM, OT-CFM, HOFM, and VTV-FM.



NS2D sample 1 NS2D sample 2

KS2D sample 1 KS2D sample 2

Figure 3: NS2D and KS2D generation examples. Within each example, the top, middle, and bottom rows show real, VTV-FM, and OT-CFM samples, respectively.

E.2 Additional PDE Visualization

Figure 3 compares real and generated fields for the NS2D and KS2D experiments in Section 4.2.

E.3 Additional CIFAR-10 Samples

Figure 4 uses the best VTV-FM configuration in Table 4: the full objective, variational terminal-velocity closure, and 100-step Velocity Verlet sampling.

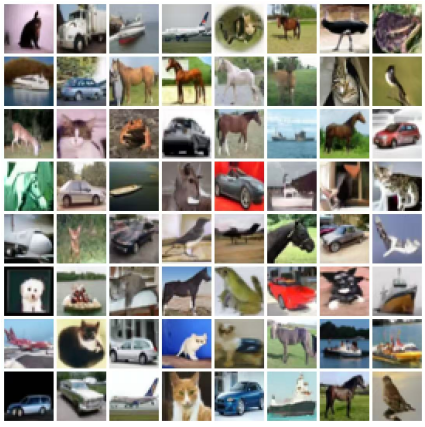


Figure 4: Qualitative CIFAR-10 samples generated by VTV-FM.

E.4 Additional Video Visualization

Figure 5 provides qualitative examples on KTH, BAIR, and Navier–Stokes temporal sequences. Terminal velocities are approximated by finite differences between adjacent frames and used in the full-boundary bridge. These examples illustrate the observed/derivable-v_1 construction on temporal data; they do not constitute a quantitative video-generation benchmark.

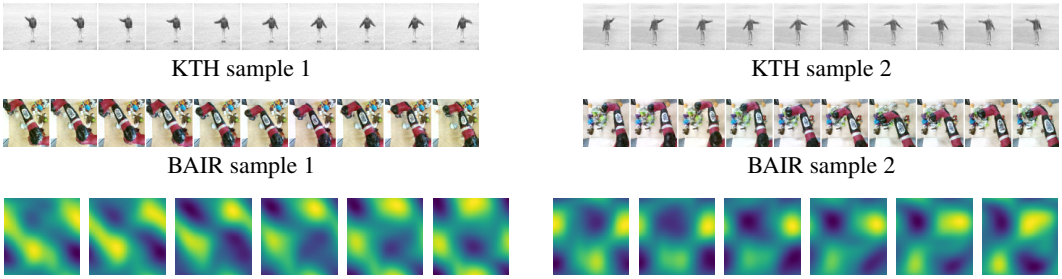


Figure 5: Qualitative temporal examples from KTH, BAIR, and Navier–Stokes sequences.

F Additional Controlled Experiments

This appendix provides additional controlled experiments that complement the main results. We examine loss and solver analysis on low-dimensional datasets, the source phase-space prior, terminal-velocity completion, the consistency between endpoint pairing and bridge construction, and computational efficiency and scalability. We do not repeat the solver-controlled CIFAR-10 comparisons already reported in Section 4.3.

F.1 Loss and Solver Analysis on Low-Dimensional Datasets

We provide additional ablation and solver analysis for VTV-FM on the low-dimensional datasets. Table 5 summarizes two studies. First, we compare different training objectives: acceleration-only

training, acceleration with local consistency, acceleration with acceleration-divergence matching (ADM), and a velocity–acceleration joint variant. Second, we compare three fixed-step solvers: Heun, Taylor2, and Velocity Verlet, using the best-performing loss setting for each dataset.

The acceleration-only objective is already competitive, achieving an average sample-based W_2 of 0.1018. Adding local consistency gives the best average among the loss variants, reducing the mean W_2 to 0.1012, while ADM gives a similar average of 0.1013. This suggests that the main gain comes from the bridge-induced acceleration supervision, with auxiliary terms providing modest refinements. For solver choice, all three fixed-step solvers perform similarly under the same sampling budget. Heun obtains the best average W_2 of 0.1007, while Taylor2 and Velocity Verlet remain close at 0.1011 and 0.1013, respectively. These results indicate that VTV-FM is not tied to a particular solver; its gains mainly come from the second-order phase-space transport construction.

Table 5: Ablation and fixed-step solver analysis for VTV-FM on low-dimensional datasets. Left: loss variants. Right: solver comparison under the best-performing loss setting for each dataset.

	Loss ablation				Solver comparison		
Dataset	Acc.	+Cons.	+ADM	Vel.+Acc.	Heun	Taylor2	Verlet
Dinosaur	0.1119	0.1114	0.1112	**0.1109**	0.1113	0.1121	**0.1109**
Swiss	0.1112	**0.1081**	0.1105	0.1127	**0.1081**	0.1092	0.1105
Checker	**0.1092**	0.1094	0.1098	0.1092	**0.1092**	0.1096	0.1099
Moons	0.0835	0.0826	0.0807	**0.0790**	**0.0790**	0.0803	0.0808
Lines	**0.1362**	0.1366	0.1380	0.1387	0.1370	**0.1362**	0.1376
Tree	0.0741	0.0739	**0.0730**	0.0752	0.0736	0.0734	**0.0730**
Star	0.0868	0.0866	**0.0862**	0.0877	0.0869	0.0867	**0.0862**
Avg.	0.1018	**0.1012**	0.1013	0.1019	**0.1007**	0.1011	0.1013

F.2 Source Phase-Space Prior and Sensitivity Analysis

For static datasets, we introduce a source velocity

$$v_0 \sim \mathcal{N}(0, \sigma^2 I),$$

sampled independently of the source position x_0. Unless otherwise stated, we use $\sigma = 1$.

The role of v_0 is to define the source phase-space state required by the second-order formulation. For static data, it is a modeling variable rather than an observed physical velocity, and its scale is not intended to be a finely tuned performance parameter.

To evaluate sensitivity to this choice, we vary σ while keeping the remaining training and evaluation settings fixed. Table 6 reports the resulting average W_2 over the seven low-dimensional datasets.

Table 6: Sensitivity to the source-velocity scale, averaged over the seven low-dimensional datasets.

σ	0	0.5	1.0	1.5	2.0
Avg. W_2	0.1283	**0.1002**	0.1004	0.1011	0.1056

As shown in Table 6, performance remains similar over the moderate nonzero range $\sigma \in [0.5, 1.5]$, indicating limited sensitivity to the precise source-velocity scale. In contrast, $\sigma = 0$ collapses the velocity variation in the source phase-space distribution and gives substantially worse performance. Thus, v_0 primarily provides the source phase-space representation, while the precise choice of a reasonable nonzero scale has only a modest effect on performance.

F.3 Terminal-Velocity Ablation Study

We directly compare the proposed variational terminal-velocity closure

$$v_1^\star = \frac{3}{2T}(x_1 - x_0) - \frac{1}{2}v_0$$

with five alternative completion rules: *Zero*, $v_1 = 0$; *Chord*, $v_1 = (x_1 - x_0)/T$; *Random*, which uses a random direction with the same norm as $v_1^\star$; *Copy-*v_0, $v_1 = v_0$; and *Tangent*, which sets v_1 along

the local target-data tangent estimated by one-dimensional PCA around x_1, with its sign aligned to $v_1^\star$ and its norm matched to $\|v_1^\star\|_2$.

We consider two settings. In the *fixed-pairing* setting, the projected-2OT source–target pairs are held fixed and only the terminal-velocity rule is changed. This isolates the effect of terminal-velocity completion and the resulting bridge supervision. In the *end-to-end* setting, each alternative terminal-velocity rule is substituted into the full-boundary cost, after which pairing and training are recomputed; VTV-FM uses its projected cost and the corresponding variational bridge. Table 7 summarizes both settings, averaged over the seven low-dimensional datasets. As shown in Table 7, with pairing fixed,

Table 7: Terminal-velocity ablation averaged over the seven low-dimensional datasets.

Setting	$v_1^\star$	Zero	Chord	Random	Copy-v_0	Tangent
Fixed pairing	**0.1004**	0.1050	0.1020	0.1094	0.1062	0.1013
End-to-end	**0.1004**	0.1041	0.1008	0.1087	0.1054	0.1005

the proposed closure gives the lowest average error among all terminal-velocity rules. The advantage remains in the end-to-end setting, where pairing and training are recomputed for each completion rule. These results show that the variational terminal-velocity completion provides a robust bridge construction beyond simply changing the endpoint pairing.

F.4 Pairing–Bridge Cross-Ablation Study

We further separate the effects of endpoint pairing and bridge construction by crossing standard and projected OT pairing with first- and second-order bridges. Table 8 reports the four resulting combinations.

Table 8: Pairing–bridge cross-ablation on the low-dimensional datasets.

Method	Bridge	Pairing	Avg. W_2
OT-CFM	Linear	Standard OT	0.1077
2OT-CFM	Linear	Projected 2OT	0.1109
OT-VTV-FM	Second-order	Standard OT	0.1283
2OT-VTV-FM	Second-order	Projected 2OT	**0.1004**

As shown in Table 8, neither mismatched cross-combination improves performance. Projected 2OT does not improve the linear bridge, while combining the second-order bridge with standard OT also degrades performance. The strongest result is obtained when the projected pairing cost and second-order bridge are used together. This supports the interpretation that the improvement is associated with the variational consistency among terminal-velocity closure, pairing cost, and bridge construction, rather than with replacing any single component in isolation.

F.5 Computational Efficiency and ImageNet-256 Scaling

Computational efficiency on low-dimensional benchmarks. We measure wall-clock inference time and peak allocated GPU memory using the same hardware, evaluation batch size, and solver configuration. VTV-FM uses acceleration matching only in this comparison. Table 9 reports the average W_2, number of function evaluations, wall-clock inference time, and peak GPU memory on the seven low-dimensional benchmarks.
As shown in Table 9, under Heun-100, VTV-FM improves average W_2 with essentially unchanged wall-clock time and peak memory relative to the first-order baselines. Under Dopri5, it also achieves the lowest error with fewer function evaluations and lower inference time. Although VTV-FM augments the dynamical state from x to (x, v), the measured increase in peak memory remains small.

ImageNet-256 scaling. We additionally evaluate VTV-FM on ImageNet-256 to test scalability beyond CIFAR-10. We fine-tune the official SiT checkpoint [29] and otherwise follow the OAT-FM evaluation setting [11], including CFG $= 1.5$, SDE sampling, and the official SiT FID evaluation protocol. VTV-FM obtains an FID of 1.97, compared with 2.06 for SiT and 2.00 for OAT-FM. We primarily interpret this experiment as evidence that the proposed construction scales beyond CIFAR-10, rather than as evidence that minimum acceleration is universally superior to other high-order objectives.

Table 9: Compute-matched comparison on the low-dimensional benchmarks.

Solver	Method	Avg. W_2	NFE	Time (ms)	Peak memory (MB)
Heun-100	CFM	0.1085	200	858	62.30
	OT-CFM	0.1079	200	861	62.30
	VTV-FM	**0.1020**	200	859	62.37
Dopri5	CFM	0.1069	72.8	486	62.58
	OT-CFM	0.1066	42.2	306	62.59
	VTV-FM	**0.1018**	**38.0**	**242**	63.66
Verlet-100	VTV-FM	**0.1009**	200	687	62.96

G VTV-FM as an Extension of OT-CFM

We relate the construction in Section 3 to first-order and kinetic optimal transport, then compare the minimum-acceleration bridge with a minimum-jerk alternative.

G.1 OT-CFM as First-Order Variational Flow Matching

For fixed endpoints, the least kinetic action is

$$\inf_{\substack{\gamma(0)=x_0\\ \gamma(T)=x_1}} \int_0^T \|\dot\gamma(t)\|_2^2\,dt = \frac{1}{T}\|x_1-x_0\|_2^2,$$

attained by the linear bridge in Section 2. At the distribution level, the Benamou–Brenier formulation [30] gives

$$\frac{1}{T}W_2^2(\rho_0,\rho_1) = \inf_{\rho_t,u_t} \int_0^T \int \|u_t(x)\|_2^2\,d\rho_t(x)\,dt,$$

subject to

$$\partial_t\rho_t + \nabla_x\cdot(\rho_t u_t) = 0, \qquad \rho_{t=0}=\rho_0, \quad \rho_{t=T}=\rho_1.$$

Thus, the quadratic pairing cost and constant conditional velocity target arise from the same first-order action.

G.2 Second-Order Lift to Phase Space

For phase-space laws $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^d)$, the analogous kinetic transport problem minimizes acceleration energy:

$$\mathsf{d}_T^2(\mu,\nu) = T\inf_{\mu_t,F_t} \int_0^T \int \|F_t(x,v)\|_2^2\,d\mu_t(x,v)\,dt,$$

subject to

$$\partial_t\mu_t + v\cdot\nabla_x\mu_t + \nabla_v\cdot(F_t\mu_t) = 0, \qquad \mu_{t=0}=\mu, \quad \mu_{t=T}=\nu.$$

Its static formulation [28] is

$$\mathsf{d}_T^2(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int c_T(x,v,y,w)\,d\pi, \qquad c_T = T\,C_T.$$

Here C_T is the full-boundary cost in Eq. (4); the factor T accounts for the normalization used above. The minimizing bridge and its derivatives are given in Eqs. (2)–(3); Appendix C.1 proves Proposition 3.1.

G.3 Projection under Missing Terminal Velocity

When only the target position law ρ_1 is prescribed, the target velocity is optimized out. Theorem 3.2 and Appendix C.2 show that the resulting observable transport problem is

$$\inf_{\eta\in\Pi(\mu_0,\rho_1)} \int C_T^\star(x_0,v_0,x_1)\,d\eta.$$

The minimizing lift assigns $v_1 = v_1^\star(x_0, v_0, x_1)$, with the closure and projected cost given in Eqs. (5)–(6). Their use in pairing and supervision is described in Section 3.3 and Appendix C.3. This gives the following correspondence:

Method	Action	Endpoint cost	Conditional bridge
OT-CFM	$\int_0^T \|\dot\gamma\|_2^2\, dt$	$\|x_1 - x_0\|_2^2/T$	Linear
VTV-FM	$\int_0^T \|\ddot\gamma\|_2^2\, dt$	$C_T^\star(x_0, v_0, x_1)$	Projected cubic

Both constructions derive the pairing cost and supervision from one action; the second-order construction additionally determines the unobserved terminal velocity.

G.4 Bridge-Shape Ablation: Minimum Acceleration versus Minimum Jerk

To assess the role of the cubic bridge, we replace the acceleration action by

$$\mathcal{J}_T[\gamma] = \int_0^T \|\gamma^{(3)}(t)\|_2^2\, dt.$$

The alternative fixes $\gamma(0) = x_0$, $\dot\gamma(0) = v_0$, and $\gamma(T) = x_1$, imposes $\ddot\gamma(0) = \ddot\gamma(T) = 0$, and leaves the terminal velocity free. We call it VTV-FM-JERK-ZA. The Euler–Lagrange equation $\gamma^{(6)} = 0$, together with the natural condition $\gamma^{(4)}(T) = 0$, gives the projected quintic bridge below.

Let $\tau = t/T$ and $\delta = x_1 - x_0 - Tv_0$. Both bridges have the form

$$x_q(\tau) = x_0 + T\tau v_0 + q(\tau)\delta, \qquad v_q(\tau) = v_0 + \frac{q'(\tau)}{T}\delta, \qquad a_q(\tau) = \frac{q''(\tau)}{T^2}\delta,$$

with

$$q_{\mathrm{acc}}(\tau) = \frac{3}{2}\tau^2 - \frac{1}{2}\tau^3, \qquad q_{\mathrm{jerk}}(\tau) = \frac{5}{2}\tau^3 - \frac{15}{8}\tau^4 + \frac{3}{8}\tau^5.$$

Here q_{acc} follows by substituting Eq. (5) into Eq. (2). The jerk bridge has terminal velocity and projected cost

$$v_{1,\mathrm{jerk}}^\star = \frac{15}{8T}(x_1 - x_0) - \frac{7}{8}v_0, \qquad J_T^\star = \frac{45}{T^5}\|\delta\|_2^2 = \frac{15}{T^2}C_T^\star.$$

For fixed T, these proportional costs have the same unregularized optimal couplings. This equivalence extends to entropic OT if the entropy coefficient is scaled with the cost: $\varepsilon_{\mathrm{jerk}} = (15/T^2)\varepsilon_{\mathrm{acc}}$. Proportional costs alone do not ensure identical Sinkhorn couplings at fixed ε.

Table 10 compares the two variants using acceleration matching only and the same acceleration-network architecture. Their average W_2 values are nearly identical, with no consistent advantage across datasets. We retain the minimum-acceleration bridge because it matches the second-order variational formulation without imposing zero endpoint accelerations.

Table 10: Bridge-shape ablation using acceleration matching only. VTV-FM-JERK-ZA uses a minimum-jerk bridge with zero endpoint accelerations.

Dataset	VTV-FM	VTV-FM-JERK-ZA
Dinosaur	0.1119	**0.1113**
Swiss	**0.1112**	0.1120
Checker	**0.1092**	0.1096
Moons	0.0835	**0.0829**
Lines	0.1362	**0.1356**
Tree	**0.0741**	0.0745
Star	0.0868	**0.0863**
Average	0.1018	**0.1017**

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes].

   Justification: The abstract and introduction accurately describe the variational terminal-velocity closure, bridge construction, and pairing cost developed in Section 3, with empirical claims supported by Section 4 and Appendix H.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes].

   Justification: Appendix A discusses the assumptions of the minimum-acceleration closure, the scope of the experiments, and directions requiring further evaluation.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes].

Justification: Section 3 states the variational setting and assumptions for Proposition 3.1 and Theorem 3.2. Complete proofs are provided in Appendices E.1 and E.2.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes].

Justification: Sections 3–4 and Appendices D–E describe the method and evaluation protocols. Detailed implementations, experimental configurations, and reproduction instructions are provided in the accompanying open-source repository.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes].

   Justification: Code, data-access instructions, and reproduction scripts are publicly available at `https://github.com/HaoyangJiang-WM/VTV-FM`.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes].

   Justification: Section 4 and Appendix D describe the experimental settings and evaluation protocols. Complete training and test configurations, including optimizer settings and hyperparameters, are provided with the open-source code.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No].

   Justification: The reported results do not include error bars, confidence intervals, or statistical significance tests across independent training runs.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No].

   Justification: Appendix H.4 reports inference time, function evaluations, and peak GPU memory for the low-dimensional comparisons. Complete hardware specifications and training compute budgets are not reported for every experiment.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes].

   Justification: The research conforms to the NeurIPS Code of Ethics and uses synthetic data, simulated physical fields, and established benchmarks without human-subject studies or private personal data collection.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes].

Justification: Appendix B discusses potential benefits for scientific generation and physical simulation, together with potential misuse of improved image and video generation.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A].

Justification: The work does not release high-risk pretrained models or newly scraped datasets requiring dedicated release safeguards.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes].

Justification: Existing assets are credited in the paper and accompanying repository, which documents their applicable licenses and terms of use. These terms are respected in the experiments and code release.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes].

    Justification: The released VTV-FM implementation is accompanied by documentation covering dependencies, experiment configurations, training and evaluation instructions, and licensing.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A].

    Justification: The research does not involve crowdsourcing, user studies, or human-subject experiments.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A].

Justification: The research does not involve human-subject studies requiring IRB approval or equivalent review.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A].

Justification: LLMs are not an important, original, or non-standard component of the proposed method or experimental evaluation.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.