MANIFOLD ATTRACTING FLOW MATCHING: PHYSICS-CONSTRAINED SAMPLING VIA CONTINUOUS GEOMETRIC DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Generating scientific data under previously unseen equality constraints requires satisfying the prescribed constraints without destroying the physical structure learned from data. Existing inference-time approaches often enforce physical constraints by repeatedly correcting predicted endpoints, but such discrete corrections can distort learned generative structure and introduce unstable feedback. In this work, we propose Manifold Attracting Flow Matching (MAFM), which continuously evolves a persistent anchor through an ODE whose dynamics drive it analytically toward the constraint manifold. Specifically, its geometric correction preserves model-guided tangent motion while prescribing the normal motion to contract constraint violations toward the manifold. This construction separates model tracking from constraint attraction and provides theoretical guarantees for residual contraction and local feedback stabilization. Experiments across geometric and partial differential equation benchmarks demonstrate improved generation quality over variant constraint-guided baselines while maintaining low constraint residuals. MAFM provides a flexible framework for coordinating learned physical structure and explicit constraints through controllable sampling dynamics, without retraining the underlying generative model. Our codes are available at https://anonymous.4open.science/r/MAFM-89EB

1 INTRODUCTION

Generative models have become an increasingly important tool for producing diverse and realistic samples across scientific and engineering domains. Yet in many applications, realism alone is not sufficient: we need to generate samples that satisfy what the task actually asks for. Crucially, many of these requirements are *hard* rather than preferential—a sample that violates them may be plausible under the learned data distribution, but still invalid for downstream use. For example, PDE reconstruction from sparse observations requires generated solution fields to agree with prescribed measurements or boundary values (Huang et al., 2024; Shysheya et al., 2024); controllable image generation may require specified regions or pixels to obey a prescribed spatial or color pattern (Endo, 2024; Shum et al., 2025); and safety-critical decision making requires states or actions to remain within prescribed safety limits rather than merely reducing violations on average (Ames et al., 2017; Shi et al., 2023; Hu et al., 2025). These examples highlight a common need: generating diverse samples while satisfying specific constraints exactly.

A particularly important setting arises when such constraints are specified only after training. A generative model may be trained once on a broad data distribution, but subsequently queried with new observations, boundary values, geometric specifications, or control targets (Wei et al., 2024; Feng et al., 2025). Collecting a dedicated dataset or retraining a conditional model for every new request would substantially limit the reuse of the learned generative prior. We therefore study *inference-time constrained generation*: an unconditional model learns the data distribution independently of downstream constraints, while an analytic equality constraint is supplied only to the sampler at inference time. The goal is to generate diverse samples from the portion of the learned distribution that satisfies the prescribed constraint.

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

Existing approaches incorporate physical information at different stages. Physics-informed training and constrained fine-tuning modify the learning objective to account for known physical requirements (Bastek et al., 2025; Tauberschmidt et al., 2026). Inference-time methods instead reuse a pretrained model: DiffusionPDE guides generation using residual gradients (Huang et al., 2024), while ECI alternates endpoint extrapolation, physical correction, and interpolation (Cheng et al., 2025). Physics-Constrained Flow Matching (PCFM) corrects predicted flow endpoints and re-interpolates with the original source to enforce sampling-time constraints (Utkarsh et al., 2026). Despite these advances, satisfying the prescribed constraint is only one requirement: constraint enforcement should not drive generated samples away from the manifold learned from data. We therefore ask whether a sampler can satisfy a prescribed hard constraint while preserving fidelity to the learned data manifold. Accordingly, we evaluate two complementary objectives: constraint satisfaction and data-manifold fidelity. When the probability distribution along the feasible manifold is also of interest, we further use distributional statistics to assess whether the generated samples reproduce the desired probability mass.

We introduce Manifold Attracting Flow Matching (MAFM), an inference-time method for imposing new equality constraints on a pretrained generative model without retraining. The key idea is *persistent-anchor dynamics*. Instead of injecting constraint guidance directly into the transient sampling state x_t, as in gradient-guidance or projection-based methods, MAFM guides constrained generation through the continuous geometric evolution of a persistent anchor, whose dynamics analytically converge to the constraint manifold. Learned endpoint predictions and constraint feedback are jointly accumulated in this anchor, which serves as the evolving candidate solution ultimately returned by the method. Its dynamics separate the assimilation of constraint-compatible model information from attraction toward the constraint manifold, allowing learned structure to be preserved while feasibility is controlled at an independently prescribed rate. We characterize the resulting constraint-residual dynamics under explicit assumptions and analyze how endpoint-prediction and adjoint-estimation errors propagate through the anchor trajectory.

We evaluate MAFM across five scientific-computing benchmarks involving Poisson, Burgers, reaction–diffusion (RD), Euler, and Navier–Stokes (NS) systems. Together, these tasks test constraints arising from partial observations, conservation laws, prescribed physical quantities, and recovery of feasible samples while retaining learned data structure. Across the benchmark suite, MAFM consistently reaches the prescribed numerical-feasibility tolerance while maintaining strong fidelity to the learned data support; among methods that achieve feasibility, it attains the lowest manifold-fidelity error across the evaluated tasks. So our contributions can be summarized as follows:

- We introduce a *persistent-anchor* sampling paradigm for imposing new equality constraints at inference time with a frozen flow model. The persistent anchor provides a shared endpoint representation that continuously integrates learned guidance from the pretrained flow with constraint feedback, directing both toward the eventual output.
- We characterize the constraint-residual dynamics induced by the persistent-anchor updates, establish conditions for convergence toward feasibility, and analyze how endpoint-prediction and adjoint-estimation errors affect the anchor trajectory.
- We evaluate MAFM across scientific-computing benchmarks involving partial observations, conservation laws, and prescribed physical quantities, demonstrating that hard constraints can be satisfied to numerical tolerance while preserving fidelity to the learned data support.

2 Related Work

Generative modeling for physical systems. Continuous normalizing flows and flow matching provide flexible generative models for high-dimensional physical states (Chen et al., 2018; Papamakarios et al., 2021; Lipman et al., 2022), with minibatch optimal transport often used to simplify conditional transport paths (Tong et al., 2023). When the underlying geometry is known, Riemannian score models and flow matching can construct dynamics directly on the prescribed manifold (De Bortoli et al., 2022; Chen & Lipman, 2024). In contrast, our setting learns the data support from ambient samples and introduces a separate analytic constraint only at inference time.

Physical information has been incorporated either during training or during sampling. Residual-based objectives and physics-constrained fine-tuning bias the learned model toward known physical

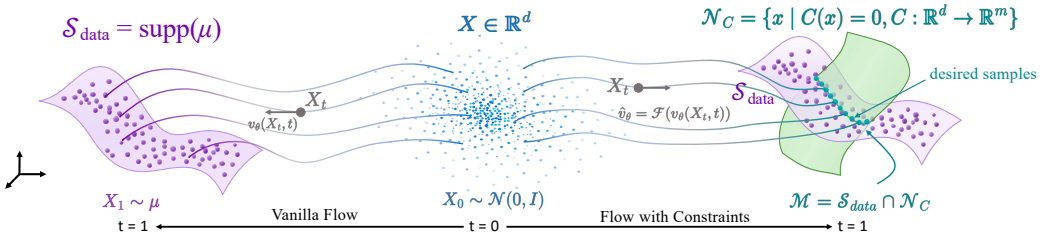

Figure 1: **Constrained sampling on a learned data support.** The broad data support (purple) meets the supplied constraint set (green) at the desired feasible intersection. The illustration denotes the data distribution by $\mu = p_{\text{data}}$, its support by $\mathcal{S}_{\text{data}} = \mathcal{S}$, and the constraint set by $\mathcal{N}_C = \mathcal{N}$.

requirements (Raissi et al., 2019; Bastek et al., 2025; Jacobsen et al., 2025; Baldan et al., 2026; Tauberschmidt et al., 2026). Inference-time approaches instead reuse a pretrained model: Diffusion-PDE applies residual-gradient guidance (Huang et al., 2024), ECI alternates endpoint extrapolation with hard correction and interpolation (Cheng et al., 2025), and PCFM repeatedly corrects predicted endpoints before re-interpolation (Utkarsh et al., 2026). However, constraint satisfaction alone does not ensure that these corrections preserve the physical relationships learned from data. While operating at inference time with a frozen flow model, MAFM maintains a *persistent endpoint anchor* across sampling steps, using learned endpoint guidance and constraint feedback to update a common candidate for the final output rather than attaching constraint control only to transient sampling states.

Geometric motivation and constraint stabilization. Stiff and fast–slow dynamical systems naturally separate rapid attraction toward a lower-dimensional set from slower evolution along it (Wanner & Hairer, 1996; Fenichel, 1979). Classical constrained integration similarly uses algebraic multipliers to enforce or stabilize equality constraints (Gear, 1988; Ascher & Petzold, 1993), with related ideas appearing in neural differential-algebraic models (Boesen et al., 2025; Diamond & Lill, 2025). These perspectives motivate the separation used in MAFM: learned endpoint information primarily guides motion compatible with the feasible geometry, while multiplier-based feedback independently controls contraction of the known constraint residual. We use this viewpoint as algorithmic motivation; our analysis establishes residual dynamics for the prescribed constraint and does not assume persistence of the unknown learned data support.

Neural operators for physical fields. Our PDE experiments use Fourier neural operator architectures to model high-dimensional physical fields (Li et al., 2020). Automatic differentiation and matrix-free linear algebra are used to apply constraint Jacobians when closed-form decompositions are unavailable (Baydin et al., 2018).

3 Manifold Attracting Flow Matching

In this section, we first formulate the problem setting and develop the persistent-anchor dynamics underlying MAFM. We then derive the corresponding sampling algorithm and its discrete implementation, followed by analyses of residual contraction, adjoint guidance, and finite-step error.

3.1 Problem Formulation

Let $\mathcal{X} = \mathbb{R}^d$ be the space of complete physical objects. We write the source and data points as $x_0 \sim p_0 = \mathcal{N}(0, I_d), x_1 \sim p_{\text{data}}$, and denote the data support by $\mathcal{S} = \text{supp}(p_{\text{data}})$. At generation time, an equality constraint specifies

$$\mathcal{C} : \mathcal{X} \to \mathbb{R}^m, \qquad \mathcal{N} = \{x \in \mathcal{X} : \mathcal{C}(x) = 0\}, \tag{1}$$

The feasible intersection is $\mathcal{M} = \mathcal{S} \cap \mathcal{N} \neq \varnothing$ as is shown in Figure 1. We use μ_q for a desired control distribution whose support is contained in $\mathcal{M}$, without prescribing probability weights on the intersection. The frozen unconditional model learns relationships in the broad data distribution; test-time constraints enter only the sampler. This assumption places the requested constraint within the physical range represented by the data. It allows the finite training set to contain no exactly

feasible sample. We assess the generated output through two complementary objectives,

$$\mathfrak{e}_{\mathrm{con}}(x) = \|\mathcal{C}(x)\|, \qquad \mathfrak{e}_{\mathrm{supp}}(x) = \mathrm{dist}(x, \mathcal{S}). \tag{2}$$

The unknown support makes the second objective difficult to compute directly. Geometry, PDE residuals, and independent rollouts test specified relationships in the data; distribution statistics provide auxiliary coverage diagnostics. An additional choice of probability weights and its corresponding evaluation appear in Appendix A.1.

Global and local constraints. For a differentiable constraint, define the row-oriented Jacobian $J_{\mathcal{C}}(a) = \partial \mathcal{C}(a)/\partial a \in \mathbb{R}^{m\times d}$. The product $J_{\mathcal{C}}(a)u$ gives the residual rate induced by a physical velocity u, while $J_{\mathcal{C}}(a)^\top \lambda$ gives a physical-space normal correction. A *global constraint* has a correction distributed broadly over the physical object, as for mass or energy. A *local observation constraint* has a correction supported at selected sensors, windows, or channels, as in $\mathcal{C}(x) = Hx - y$.

Frozen endpoint model. Let $(x_0, x_1) \sim \pi$, with marginals p_0 and p_{data}; the coupling π need not be independent. Vanilla flow matching uses the complete-space training interpolation

$$x_t^{\mathrm{train}} = (1-t)x_0 + t x_1, \qquad t \in [0,1], \tag{3}$$

whose target velocity is $x_1 - x_0$. The frozen model defines the Vanilla flow $\dot{x}_t^{\mathrm{FM}} = v_\theta(x_t^{\mathrm{FM}}, t)$ from $x_{t=0}^{\mathrm{FM}} = x_0$ (Lipman et al., 2022). We extract endpoint information through

$$D_\theta(x,t) = x + (1-t)v_\theta(x,t), \tag{4}$$

an endpoint-extrapolation form also used in PCFM (Utkarsh et al., 2026). Under finite second moments, for the population-optimal regression field $v^*(x,t) = \mathbb{E}[x_1 - x_0 \mid x_t^{\mathrm{train}} = x]$,

$$D^*(x,t) = x + (1-t)v^*(x,t) = \mathbb{E}[x_1 \mid x_t^{\mathrm{train}} = x] \tag{5}$$

on the training path almost everywhere. Thus D_θ is a learned data-endpoint estimate used repeatedly by the anchor dynamics. Its proof, parameterization, and query-region conditions are given in Appendix A.2.

3.2 Guided Manifold Attracting Dynamics

We construct the sampling dynamics from a persistent endpoint anchor, optional guidance through the frozen flow matching model, and a multiplier that controls the anchor's constraint residual. Intuitively, the anchor is the final pre-predicted endpoint state.

Query and persistent anchor. The new constraint concerns the final physical object which we noted as the anchor a_t. For each fixed source x_0, MAFM maintains a network query x_t and a persistent endpoint anchor a_t, both in $\mathcal{X}$, with

$$x_t = (1-t)x_0 + t a_t. \tag{6}$$

The query starts at $x_{t=0} = x_0$ and is rebuilt after every anchor update. At $t = 1$ the two states coincide, so the generated output is $x_1^{\mathrm{gen}} := x_{t=1} = a_1$. Here x_1 without a superscript remains the data endpoint in Eq. (3).

Endpoint-guided candidate. For local observations, direct normal correction acts on the measured coordinates. The frozen endpoint model can also carry their mismatch to other coordinates through its input sensitivity. To obtain this additional guidance, define

$$\mathcal{E}_t(x) = \frac{1}{2}\|\mathcal{C}(D_\theta(x,t))\|^2, \tag{7}$$

This energy measures the constraint mismatch of the predicted endpoint; the normal correction below acts on the anchor residual $\mathcal{C}(a_t)$. When this option is active, the input adjoint at x_t is

$$g_t = \nabla_x \mathcal{E}_t(x_t) = J_{D_\theta}(x_t,t)^\top J_{\mathcal{C}}(D_\theta(x_t,t))^\top \mathcal{C}(D_\theta(x_t,t)). \tag{8}$$

The query relation in Eq. (6) has anchor derivative tI_d. Its chain rule therefore gives

$$\nabla_a \mathcal{E}_t((1-t)x_0 + ta)|_{a=a_t} = t\, g_t. \tag{9}$$

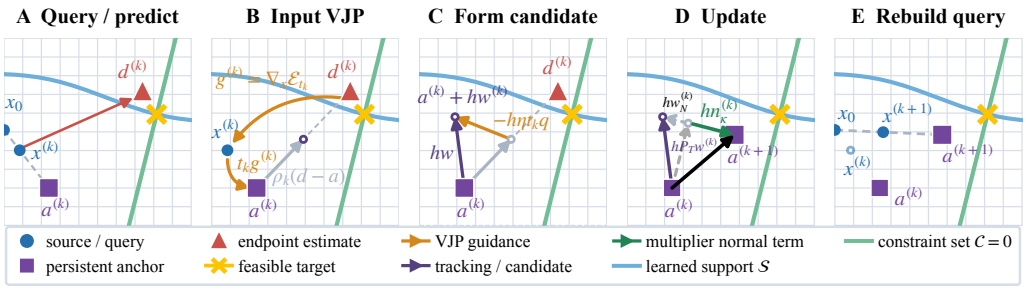

Figure 2: One step of persistent-anchor MAFM update.

The Jacobian transposes transfer the predicted endpoint's residual to the complete query; Eq. (9) then transfers it to the anchor. The model parameters remain frozen. Let $\gamma(t), \eta(t) \geq 0$ be the tracking and guidance rates. For $\gamma(t) > 0$, the optional adjoint can be expressed as an endpoint displacement. Set $\beta(t) = \eta(t)/\gamma(t)$ and define

$$\widetilde{D}_\theta(x_t, t) = D_\theta(x_t, t) - \beta(t) t g_t. \tag{10}$$

The factor t follows from Eq. (9). For all nonnegative rates, the candidate velocity is defined by

$$w_t = \underbrace{\gamma(t)[D_\theta(x_t, t) - a_t]}_{\text{endpoint tracking}} - \underbrace{\eta(t) t g_t}_{\text{endpoint-constraint guidance}}. \tag{11}$$

When $\gamma(t) > 0$, this equals $\gamma(t)[\widetilde{D}_\theta(x_t, t) - a_t]$; the virtual endpoint is an equivalent interpretation, not an additional state. At $\gamma(t) = 0$, Eq. (11) still gives $w_t = -\eta(t) t g_t$. The choice $\eta \equiv 0$ skips the input VJP, in which case the method reduces to PCFM(Utkarsh et al., 2026); its finite tracking rate accumulates successive endpoint predictions in the persistent anchor. Therefore, repeated predictions from the frozen endpoint model guide this anchor, while the correction below controls its constraint residual.

Constraint-compatible anchor dynamics. Let $G(a) = J_\mathcal{C}(a) J_\mathcal{C}(a)^\top$. An algebraic multiplier corrects the combined candidate velocity at the anchor, giving the unified dynamics

$$(G(a_t) + \delta I_m)\lambda_t = -\left[J_\mathcal{C}(a_t) w_t + \kappa(t)\mathcal{C}(a_t)\right], \tag{12}$$

$$\dot{a}_t = w_t + J_\mathcal{C}(a_t)^\top \lambda_t. \tag{13}$$

Here $\kappa(t) \geq 0$ specifies the target residual-contraction rate and $\delta \geq 0$ regularizes the Gram solve. For full-row-rank $J_\mathcal{C}$, define $P_T(a) = I_d - J_\mathcal{C}(a)^\top G(a)^{-1} J_\mathcal{C}(a)$. With an exact unregularized solve ($\delta = 0$), the dynamics become

$$\boxed{\dot{a}_t = P_T(a_t) w_t - J_\mathcal{C}(a_t)^\top G(a_t)^{-1} \kappa(t) \mathcal{C}(a_t)}. \tag{14}$$

The first term retains the candidate's constraint-neutral component:

$$P_T(a_t) w_t = \gamma(t) P_T(a_t)[D_\theta(x_t, t) - a_t] - \eta(t) t P_T(a_t) g_t.$$

Thus the adjoint changes tangential motion only after projection; $-\eta(t) t g_t$ need not itself be tangent. The second term replaces the candidate's normal component with the prescribed residual attraction. For $\delta > 0$, the multiplier still retains the tangential component but relaxes the exact normal replacement. The constrained-velocity interpretation and regularization effect are derived in Appendix A.3.

3.3 Properties and Sampling Algorithm

We first identify the exact continuous residual dynamics and the adjoint's instantaneous effect on constraint-neutral motion, then give the finite-step sampler and its residual bound.

Proposition 1 (Exact residual contraction). *Suppose $\mathcal{C} \in C^1$, $J_\mathcal{C}(a_t)$ has full row rank along an existing solution, κ is integrable, and Eq. (12) is solved exactly with $\delta = 0$. Then*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}(a_t) = -\kappa(t)\mathcal{C}(a_t), \qquad \mathcal{C}(a_t) = e^{-\int_0^t \kappa(r)\mathrm{d}r}\mathcal{C}(a_0). \tag{15}$$

Algorithm 1 Unified persistent-anchor MAFM sampling.

Require: source x_0, frozen D_θ, constraint $\mathcal{C}$, steps N, schedules γ, η, κ, ridge δ

$h \leftarrow 1/N;\ x^{(0)} \leftarrow x_0;\ a^{(0)} \leftarrow D_\theta(x_0, 0)$
for $k = 0, \dots, N-1$ **do**
  $t_k \leftarrow kh;\ d^{(k)} \leftarrow D_\theta(x^{(k)}, t_k);\ J^{(k)} \leftarrow J_{\mathcal{C}}(a^{(k)})$
  $q^{(k)} \leftarrow 0$
  if $\eta(t_k)t_k > 0$ **then**
    $q^{(k)} \leftarrow \mathrm{clip}_k\big(\nabla_x \mathcal{E}_{t_k}(x^{(k)})\big)$ ▷ one input VJP
  end if
  $w^{(k)} \leftarrow (\rho_k/h)(d^{(k)} - a^{(k)}) - \eta(t_k)t_k q^{(k)}$ ▷ tracking and guidance
  solve $\big(J^{(k)}(J^{(k)})^\top + \delta I_m\big)\lambda^{(k)} = -\big[J^{(k)}w^{(k)} + \kappa(t_k)\mathcal{C}(a^{(k)})\big]$ ▷ normal attraction
  $a^{(k+1)} \leftarrow a^{(k)} + h\big[w^{(k)} + (J^{(k)})^\top \lambda^{(k)}\big]$ ▷ advance persistent anchor
  $x^{(k+1)} \leftarrow (1 - t_{k+1})x_0 + t_{k+1}a^{(k+1)}$ ▷ rebuild model query
end for
return $x^{(N)} = a^{(N)}$

In this exact setting, the multiplier fixes the residual rate independently of the candidate velocity: κ controls normal attraction, while γ and η control tangent motion. Appendix A.6 proves this result and derives the additional algebraic terms caused by a ridge or an inexact multiplier solve.

Proposition 2 (Instantaneous projected-adjoint effect)**.** *Fix a source x_0, anchor a, and time t, and evaluate g_t at the common query $x = (1-t)x_0 + ta$. Suppose $J_{\mathcal{C}}(a)$ has full row rank and the multiplier is solved exactly with $\delta = 0$. Let $V_\eta(a, t; x_0)$ be the corrected anchor velocity from Eqs. (11)–(13), and let $V_0(a, t; x_0)$ use the same state and schedules with $\eta(t) = 0$. Then*

$$V_\eta - V_0 = -\eta(t)tP_T(a)g_t, \qquad J_{\mathcal{C}}(a)(V_\eta - V_0) = 0. \tag{16}$$

For $t > 0$ and $\eta(t) > 0$, the adjoint has no instantaneous effect if and only if $P_T(a)g_t = 0$, equivalently $g_t \in \mathrm{range}(J_{\mathcal{C}}(a)^\top)$.

Together, Propositions 1 and 2 separate the two roles of the correction: the multiplier prescribes residual contraction, while the input adjoint can steer constraint-neutral motion through its tangent projection. For sparse observations, the endpoint model's learned cross-coordinate sensitivity can carry information from measured to unmeasured coordinates through this correction. Appendix A.4 gives the sparse-observation specialization, and the Poisson ablation in Section 4.5 examines the adjoint's effect.

Under local common-query-region and Lipschitz assumptions, endpoint and input-adjoint errors yield a pathwise anchor perturbation bound, stated and proved in Appendix A.9.

Discrete sampler. Our algorithm is shown in Alg. 1, which discretizes the anchor dynamics with $h = 1/N$, $t_k = kh$, and relaxed tracking weight $\rho_k = 1 - e^{-\gamma(t_k)h}$. Its optional clip_k is the identity by default; Appendix B.1 gives the clipped variant. The query is rebuilt from the updated anchor as in Eq. (6).

Proposition 3 (Finite-step residual control)**.** *Consider Alg. 1 with $h > 0$, $\delta = 0$, full-row-rank $J^{(k)}$, and an exact multiplier solve. Set $u^{(k)} = w^{(k)} + (J^{(k)})^\top \lambda^{(k)}$, $c^{(k)} = \mathcal{C}(a^{(k)})$, and $\kappa_k = \kappa(t_k)$. Suppose $\mathcal{C} \in C^1$ and $J_{\mathcal{C}}$ is L_J-Lipschitz in operator norm along the step segment. Then*

$$c^{(k+1)} = (1 - h\kappa_k)c^{(k)} + R^{(k)}, \qquad \|R^{(k)}\| \le \tfrac{1}{2}L_J h^2 \|u^{(k)}\|^2. \tag{17}$$

The proof and the accumulated bound, including regularization and inexact linear solves, are in Appendix A.8. Under constant $\kappa > 0$, $0 < h\kappa \le 1$, bounded step velocities, and exact unregularized solves, the curvature contribution to the residual bound is $O(h)$.

4 Experiments

We evaluate MAFM on six physical-space generation problems spanning geometric constraints, partial observations, conservation laws, and prescribed physical quantities. The experiments test

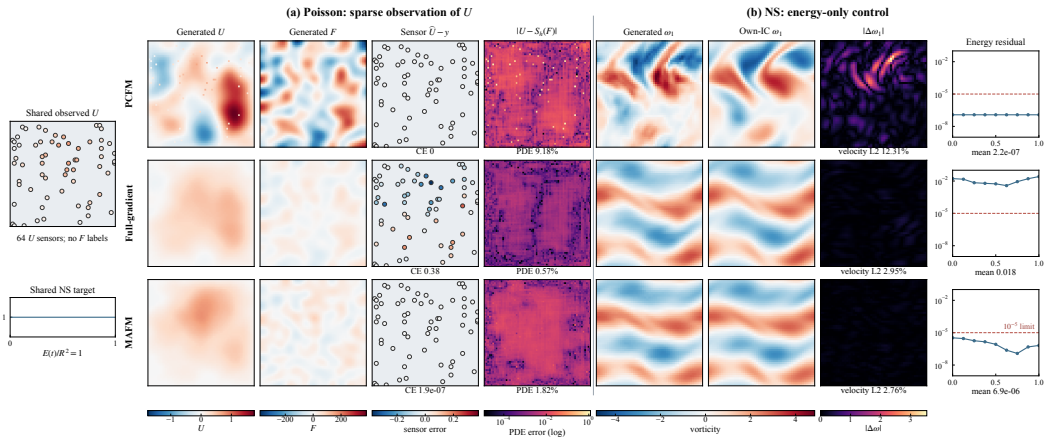

Figure 3: **Representative Poisson and fixed-energy controlled-NS outputs under shared conditions.**

whether hard constraints can be enforced at inference time while preserving the physical structure learned by the pretrained flow model. We first introduce the common experimental setup and evaluation protocol, and then present benchmark results and additional analyses.

4.1 Experimental setting and evaluation

Benchmarks. We evaluate MAFM on six physical-space generation tasks: sphere support recovery, partially observed Poisson fields, mass-constrained Burgers and reaction–diffusion trajectories, energy-conditioned Euler flows, and controlled Navier–Stokes (NS) flows. These benchmarks cover geometric constraints, partial observations, conservation laws, and prescribed physical quantities. All methods generate the complete physical state using the same pretrained flow models and inference-time conditions. Table 3 in Appendix C distinguishes the physical structure learned from data from the constraints supplied at inference time.

Baselines and metrics. We compare against Vanilla FM with and without projection, PCFM-Base/Relaxed (Utkarsh et al., 2026), ECI (Cheng et al., 2025), DiffusionPDE-style stop-gradient guidance (Huang et al., 2024), and full-gradient endpoint guidance. We evaluate two primary objectives: *constraint satisfaction* and *fidelity to the learned physical structure*. Constraint satisfaction is measured by the corresponding constraint error, while physical fidelity is evaluated using the PDE relative L2 for PDE systems. Distributional metrics are reported as auxiliary diagnostics. Complete metrics, feasibility thresholds, sample counts, and additional results are provided in Appendix D.

Implementation. All methods use identical frozen flow networks, source samples, conditions, and evaluation sets within each benchmark. Experiments are conducted on a workstation equipped with 8 NVIDIA GeForce RTX 4090 GPUs. Additional implementation details, hyperparameters, and compute settings are provided in the appendix.

4.2 Geometric and partially observed constraints

Learned support under a linear constraint. The sphere–plane task, a toy case, separates a known constraint from learned support: $z = 0$ permits a disk, whereas unit-sphere data occupy only its circular boundary. Terminal projection enforces the plane but cannot recover the circle. MAFM attains mean $|z| = 3.37 \times 10^{-6}$, radial $W_1 = 0.00493$, and the highest Joint coverage (73.0%). Full-gradient guidance has a slightly smaller radial W_1 but a worse full-field SWD, so neither metric alone determines the comparison. The full metric comparison and threshold sensitivity are in Appendix D. This illustrates the support–constraint intersection in Section 3.

Joint fields under partial observation. The Poisson condition observes sparse values of U, but the model must generate a coupled pair (U, F). Figure 3(a) shows why sensor agreement is insufficient: PCFM matches the observations yet leaves (U, F) inconsistency, while full-gradient

Table 1: **Physical consistency under supplied constraints.** (a) Poisson; (b) Burgers and reaction–diffusion; (c) Euler and controlled NS.

	(a) Partial observation			**(b) Conservation**							
	Poisson, 64 sensors ($n = 1{,}024$)			**Burgers** ($n = 768$)				**Reaction–diffusion** ($n = 384$)			
Method	Feas.	CE↓	PDE RelL2↓	Feas.	CE↓	IC RelL2↓	SWD↓	Feas.	CE↓	IC RelL2↓	SWD↓
Vanilla FM	No	1.40	1.13%	No	3.57e−1	1.13%	0.270	No	5.50e−2	1.31%	0.0965
FM + projection	**Yes**	0	22.7%	**Yes**	5.49e−8	27.0%	**0.0778**	**Yes**	5.84e−8	10.4%	0.0853
PCFM-Base	**Yes**	0	11.5%	**Yes**	5.39e−8	7.61%	0.178	**Yes**	5.89e−8	3.03%	0.130
PCFM-Relaxed	**Yes**	0	10.3%	No	3.17e−6	7.61%	0.178	No	3.72e−4	2.46%	0.116
ECI*	**Yes**	0	11.5%	**Yes**	5.40e−8	7.54%	0.177	**Yes**	5.86e−8	2.96%	0.125
Stop-grad.	**Yes**	3.29e−9	5.55%	**Yes**	5.38e−8	1.03%	0.288	**Yes**	5.92e−8	6.04%	0.122
Full-grad. (DPDE)	No	0.610	1.05%	No	2.63e−2	17.2%	0.290	No	1.70e−3	1.56%	0.0637
MAFM	**Yes**	1.72e−7	**1.71%**	**Yes**	4.22e−8	**0.965%**	0.157	**Yes**	1.13e−7	**2.74%**	**0.0666**
+ projection	**Yes**	0	**1.71%**	**Yes**	5.32e−8	**0.965%**	0.157	**Yes**	6.77e−8	**2.74%**	**0.0666**

(c) Energy-only held-out tests (512 samples per system, NFE=100)

	Euler						**Controlled NS**					
Method	Feas.	E↓	IC RelL2↓	Div.↓	Mean↓	SWD↓	Feas.	E↓	IC RelL2↓	Div.↓	Mean↓	SWD↓
Vanilla FM	No	1.85e−2	3.68%	0.0635	0.0131	0.0662	No	1.80e−2	3.46%	0.0624	0.0111	0.0495
Terminal projection	**Yes**	2.29e−7	3.62%	0.0636	0.0131	**0.0662**	**Yes**	2.32e−7	3.40%	0.0624	0.0113	**0.0495**
PCFM-Base	**Yes**	2.12e−7	13.6%	0.232	0.0276	0.114	**Yes**	2.21e−7	12.3%	0.213	0.0301	0.0945
PCFM-Relaxed	No	1.84e−3	13.6%	0.229	0.0297	0.115	No	1.81e−3	12.3%	0.212	0.0301	0.0948
ECI†	**Yes**	2.12e−7	13.6%	0.232	0.0276	0.114	**Yes**	2.21e−7	12.3%	0.213	0.0301	0.0945
Stop-grad. guidance	No	2.13e−2	8.64%	0.112	0.0214	0.118	No	1.97e−2	3.47%	0.0494	0.00893	0.0792
Full-grad. (DPDE)	No	1.60e−2	3.42%	0.0604	0.0106	0.0667	No	1.78e−2	2.95%	0.0515	0.00729	0.0630
MAFM	**Yes**	6.84e−6	**2.94%**	**0.0483**	**0.0108**	0.0712	**Yes**	6.89e−6	**2.76%**	**0.0474**	**0.00881**	0.0533
+ projection	**Yes**	2.37e−7	**2.94%**	**0.0483**	**0.0108**	0.0712	**Yes**	2.43e−7	**2.76%**	**0.0474**	**0.00881**	0.0533

guidance reduces the PDE discrepancy but misses the observation condition. Across 1,024 samples with 64 sensors(1.5% observation points), MAFM is feasible and lowers the best feasible PDE RelL2 from 5.55% to 1.71% (Table 1a). Terminal projection reaches zero observation error but has 22.7% PDE RelL2. This pattern is consistent with endpoint sensitivity guiding unobserved coordinates while normal correction enforces the observed ones. With 256 sensors, MAFM again gives the lowest feasible PDE error (2.25%); visualization details and complete results are in Appendix D.

4.3 Conservation with dynamical consistency

Burgers fixes the spatial mean throughout 65 frames, and reaction–diffusion fixes total two-species mass. Neither condition determines how the remaining field evolves. We therefore compare each generated trajectory with the independent own-IC rollout defined above, after checking mass feasibility. In Table 1b, MAFM improves the best feasible Burgers rollout error from 1.03% to 0.965%; on reaction–diffusion it lowers the best feasible error from 2.96% to 2.74% and obtains the lowest feasible Field SWD, 0.0666. The ECI* row is operator-equivalent to PCFM-Base on Poisson; it uses the fixed ECI-2 configuration on Burgers and the registered ECI configuration on reaction–diffusion. All evaluated variants appear in the appendix.

The projection baseline clarifies the distinction between the two objectives. It satisfies mass in both systems, yet its own-IC error rises to 27.0% on Burgers and 10.4% on reaction–diffusion. Its low Burgers SWD does not make those generated trajectories dynamically consistent. MAFM's joint result is consistent with the separation in Eq. (14): normal correction regulates the supplied invariant, while the learned candidate continues to guide motion along the constraint set.

4.4 Energy-conditioned prior recovery

The Euler and controlled-NS samplers receive only a kinetic-energy target; divergence, spatial mean, and time evolution are not supplied as conditions. We assess them separately to ask whether the samples still resemble the physical flows learned from data. Guidance strengths are selected on a disjoint 64-sample validation set, and all reported test metrics use 512 held-out samples from fixed-energy checkpoints. Figure 3(b) shows one controlled-NS sample per method: generated and own-IC vorticity, their absolute difference, and the energy residual over time. The velocity RelL2 and

Table 2: **Core-mechanism ablations.**

	(a) Poisson ($n = 1{,}024$)	(b) Conservation					
		Burgers ($n = 768$)			Reaction–diffusion ($n = 384$)		
Variant	PDE RelL2↓	Feas.	CE↓	IC RelL2↓	Feas.	CE↓	IC RelL2↓
Full MAFM	**1.70%**	**Yes**	4.22e−8	**0.965%**	**Yes**	1.13e−7	**2.74%**
w/o adjoint	38.0%		—			—	
Reset anchor	—	No	9.00e−3	9.44%	No	4.41e−2	6.14%
w/o support	—	**Yes**	1.18e−7	2.70%	**Yes**	1.15e−7	10.6%
w/o normal	—	No	0.270	1.29%	No	3.19e−2	2.19%

mean energy numbers printed below the panels summarize the full held-out set, rather than just the pictured sample.

The figure exposes two distinct failure modes. PCFM-Base nearly eliminates energy error but its sampled flow departs from its own-IC rollout; full-gradient guidance produces a closer rollout but misses the energy band. MAFM keeps the energy residual within that band while preserving a closer rollout. In the full comparison (Table 1c), MAFM has the lowest own-IC, divergence, and mean errors among distinct energy-feasible methods on both systems. Own-IC error falls from 3.62% to 2.94% on Euler and from 3.40% to 2.76% on controlled NS relative to the strongest feasible alternatives. Terminal projection has the lowest Field SWD; SWD alone therefore gives a different ranking from the joint feasibility and physical-consistency test. The ECI† row is tensor-identical to PCFM-Base in this setting and is retained as an implementation audit, not an independent improvement.

The optional MAFM projection rows apply a matched terminal correction to the same frozen samples and rerun the own-IC evaluation. Projection reduces the energy residual while leaving physical-consistency metrics unchanged at the reported precision (exact deltas are in the appendix). MAFM's favorable own-IC rollout therefore reflects the guided persistent-anchor trajectory, consistent with the method's separate control of energy residual and learned motion.

4.5 Core-mechanism ablations

Table 2 compares targeted changes to frozen models with matched sources, conditions, NFE $= 100$, and the main feasibility criteria; no variant is retrained or tuned. Dashes mark task–variant pairs that were not evaluated. Both Poisson variants remain feasible with the same observation CE (1.72×10^{-7}), but removing the adjoint raises PDE RelL2 from 1.70% to 38.0%, indicating a substantial role for adjoint guidance in joint U–F consistency, consistent with the pathway in Corollary 4. Removing support tracking remains feasible yet increases own-IC error from 0.965% to 2.70% on Burgers and from 2.74% to 10.6% on reaction–diffusion. Resetting the anchor degrades both feasibility and dynamics, while removing normal correction makes both systems infeasible. Thus adjoint guidance governs joint consistency, persistent support tracking preserves dynamics, and normal correction supplies feasibility.

5 Conclusion

We study inference-time sampling from the intersection of a learned data support and an equality constraint supplied only at generation time. MAFM addresses this problem through the continuous evolution of a persistent physical anchor driven by a frozen endpoint model. Successive endpoint predictions are incorporated through finite-rate tracking, while a geometric correction separates model-guided tangent motion from normal constraint attraction. Under full-rank Jacobians and exact unregularized multiplier solves, the resulting dynamics yield a prescribed residual-contraction law, while the persistent-anchor relaxation admits a local feedback-stabilization analysis.

Across six benchmarks, MAFM achieves numerical feasibility while maintaining strong consistency with the physical structure learned by the frozen model. The ablations further support the complementary roles of persistent tracking, tangent guidance, and normal correction. While MAFM provides explicit control over constraint-residual dynamics, specifying and matching a particular probability law on the feasible intersection remains a separate problem. Extending the framework toward distributionally controlled constrained sampling and broader physical regimes is an important direction for future work.

Reproducibility Statement

We make efforts to facilitate reproducibility of this work. The supplementary material includes the complete code and resources required to reproduce our experiments, together with training and evaluation scripts, configurations, environment information, and relevant execution records. The main paper provides the target definition, assumptions, sampler equations, pseudocode, and key experimental results, while Appendix A presents the theoretical derivations and proofs, and Appendix D provides detailed experimental settings, including physical tensors, constraints, evaluation metrics, complete result tables, and projection audits. We also record code and input hashes, logs, checkpoints, generated samples, and evaluation outputs, and clearly distinguish the single training seed from the number of generated samples. Negative or mixed outcomes are reported alongside improvements. Together, these materials provide the information and implementation needed for independent verification and reproduction of our reported results.

Ethics Statement

The work uses procedurally generated geometric and physical data and involves no personal data or human participants. A relevant downstream risk is treating a small declared residual as certification that a sample follows the correct physical distribution or governing dynamics. We therefore evaluate these properties separately and recommend independent verification for scientific or safety-critical use.

AI Use Statement

Generative AI tools assisted with language editing, manuscript organization, and LaTeX assembly. The authors selected the scientific claims and verified the equations, citations, and numerical values against the theory, source code, and saved experiment results. The authors take responsibility for the manuscript.

Data, Conflicts, and Funding

All experiments use synthetic data described in the paper and appendix. The anonymous authors declare no competing interests and no external funding that creates a conflict of interest.

References

Aaron D. Ames, Xiangru Xu, Jessy W. Grizzle, and Paulo Tabuada. Control barrier function based quadratic programs for safety critical systems. *IEEE Transactions on Automatic Control*, 62 (8):3861–3876, August 2017. ISSN 1558-2523. doi: 10.1109/tac.2016.2638961. URL http://dx.doi.org/10.1109/TAC.2016.2638961.

Uri M Ascher and Linda R Petzold. Stability of computational methods for constrained dynamics systems. *SIAM Journal on Scientific Computing*, 14(1):95–120, 1993.

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

Jan-Hendrik Bastek, WaiChing Sun, and Dennis Kochmann. Physics-informed diffusion models. In *International Conference on Learning Representations*, volume 2025, pp. 3360–3385, 2025.

Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, 18 (153):1–43, 2018.

Tue Boesen, Eldad Haber, and Uri M Ascher. Neural daes: Constrained neural networks. *SIAM Journal on Scientific Computing*, 47(2):C291–C312, 2025.

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

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Chaoran Cheng, Boran Han, Danielle Maddix, Abdul Fatir Ansari, Andrew Stuart, Michael W Mahoney, and Bernie Wang. Gradient-free generation for hard-constrained systems. In *International Conference on Learning Representations*, volume 2025, pp. 100510–100539, 2025.

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

Justin S Diamond and Markus A Lill. Neural shake: geometric constraints in neural differential equations: J. diamond, m. lill. *Journal of Cheminformatics*, 17(1):115, 2025.

Yuki Endo. Masked-attention diffusion guidance for spatially controlling text-to-image generation: Y. endo. *The visual computer*, 40(9):6033–6045, 2024.

Ruiqi Feng, Chenglei Yu, Wenhao Deng, Peiyan Hu, and Tailin Wu. On the guidance of flow matching. *arXiv preprint arXiv:2502.02150*, 2025.

Neil Fenichel. Geometric singular perturbation theory for ordinary differential equations. *Journal of Differential Equations*, 31(1):53–98, 1979. ISSN 0022-0396. doi: https://doi.org/10.1016/0022-0396(79)90152-9. URL https://www.sciencedirect.com/science/article/pii/0022039679901529.

Charles William Gear. Differential-algebraic equation index transformations. *SIAM Journal on Scientific and Statistical Computing*, 9(1):39–47, 1988.

Peiyan Hu, Xiaowei Qian, Wenhao Deng, Rui Wang, Haodong Feng, Ruiqi Feng, Tao Zhang, Long Wei, Yue Wang, Zhi-Ming Ma, et al. From uncertain to safe: Conformal adaptation of diffusion models for safe pde control. *arXiv preprint arXiv:2502.02205*, 2025.

Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. Diffusionpde: Generative pde-solving under partial observation. *Advances in Neural Information Processing Systems*, 37: 130291–130323, 2024.

Christian Jacobsen, Yilin Zhuang, and Karthik Duraisamy. Cocogen: Physically consistent and conditioned score-based generative models for forward and inverse problems. *SIAM Journal on Scientific Computing*, 47(2):C399–C425, 2025.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

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

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Ming Shi, Yingbin Liang, and Ness Shroff. A near-optimal algorithm for safe reinforcement learning under instantaneous hard constraints. In *International Conference on Machine Learning*, pp. 31243–31268. PMLR, 2023.

Ka Chun Shum, Binh-Son Hua, Duc Thanh Nguyen, and Sai-Kit Yeung. Color alignment in diffusion. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28446–28455. IEEE, 2025.

Aliaksandra Shysheya, Cristiana Diaconu, Federico Bergamin, Paris Perdikaris, José M Hernández-Lobato, Richard E Turner, and Emile Mathieu. On conditional diffusion models for pde simulations. *Advances in Neural Information Processing Systems*, 37:23246–23300, 2024.

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

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

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

Gerhard Wanner and Ernst Hairer. *Solving ordinary differential equations II*, volume 375. Springer Berlin Heidelberg New York, 1996.

Long Wei, Peiyan Hu, Ruiqi Feng, Haodong Feng, Yixuan Du, Tao Zhang, Rui Wang, Yue Wang, Zhi-Ming Ma, and Tailin Wu. Diffphycon: A generative approach to control complex physical systems. *Advances in Neural Information Processing Systems*, 37:4090–4147, 2024.

APPENDIX CONTENTS

A. Proofs and derivations 13
B. Additional algorithmic details 18
C. Experimental protocols 18
D. Supplementary results and diagnostics 21
E. Numerical and applicability boundaries 31

A PROOFS AND DERIVATIONS

A.1 TARGET-MEASURE GEOMETRY

The main text specifies the support of μ_q through $\mathcal{M} = \mathcal{S} \cap \mathcal{N}$. A distributional target requires an additional choice of weights on that intersection. For a measurable set B, one such choice is $p_{\mathrm{data}}(B \mid \mathcal{N})$ when $p_{\mathrm{data}}(\mathcal{N}) > 0$. For a zero-probability constraint event, fix the residual norm and normalization and, when the denominator is positive and the weak limit exists, define the tubular conditional law

$$\nu_\varepsilon(B) = \frac{p_{\mathrm{data}}(B \cap \{x : \|\mathcal{C}(x)\| \le \varepsilon\})}{p_{\mathrm{data}}(\{x : \|\mathcal{C}(x)\| \le \varepsilon\})}, \qquad \nu = \lim_{\varepsilon \downarrow 0} \nu_\varepsilon. \tag{18}$$

Suppose $\mathcal{S}$ is a smooth $d_{\mathcal{S}}$-dimensional manifold with $d_{\mathcal{S}} \ge m$, $\mathcal{C}|_{\mathcal{S}}$ has full row rank near $\mathcal{M}$, and p_{data} has density $p_{\mathcal{S}}$ with respect to $\mathcal{H}^{d_{\mathcal{S}}}$. Subject to the regularity and integrability conditions for the normalized limit, the coarea formula gives

$$\mathrm{d}\nu(x) \propto \frac{p_{\mathcal{S}}(x)}{\sqrt{\det(J_{\mathcal{S}}(x) J_{\mathcal{S}}(x)^\top)}} \, \mathrm{d}\mathcal{H}^{d_{\mathcal{S}} - m}(x), \qquad x \in \mathcal{M}, \tag{19}$$

where $J_{\mathcal{S}}$ is the Jacobian of $\mathcal{C}$ in an orthonormal tangent basis of $T_x\mathcal{S}$. The tube law depends on the chosen residual units and norm; reparameterizing the same zero set may change its limiting weights.

Let $\nu^*_{\mathrm{MAFM}} = \mathrm{Law}(a_1^*)$ be the terminal law of the ideal anchor dynamics and $\widehat{a}^{(N)}$ a numerical output. When these laws and ν have finite second moments, the Wasserstein triangle inequality gives

$$\begin{aligned} W_2(\mathrm{Law}(\widehat{a}^{(N)}), \nu) \le\ & W_2(\mathrm{Law}(\widehat{a}^{(N)}), \nu^*_{\mathrm{MAFM}}) \\ & + W_2(\nu^*_{\mathrm{MAFM}}, \nu). \end{aligned} \tag{20}$$

The first term can reflect model and numerical errors; separating it further requires assumptions linking their coupled effects. The second term is the structural difference between ideal anchor dynamics and the additionally specified target law.

A.2 ENDPOINT ESTIMATION AND QUERY-REGION CONDITIONS

Conditional mean and regression error. For the linear bridge in Eq. (3), the population flow field is $v^*(x,t) = \mathbb{E}[x_1 - x_0 \mid x_t^{\mathrm{train}} = x]$. The pointwise identity $x_t^{\mathrm{train}} + (1-t)(x_1 - x_0) = x_1$ implies, after conditioning on $x_t^{\mathrm{train}} = x$,

$$x + (1-t) v^*(x,t) = \mathbb{E}[x_1 \mid x_t^{\mathrm{train}} = x]. \tag{21}$$

The conditional expectation is also the L^2-optimal endpoint predictor. For finite second moments, its orthogonality decomposition is

$$\begin{aligned} \mathbb{E}_\pi \|D_\theta(x_t^{\mathrm{train}}, t) - x_1\|^2 =\ & \mathbb{E}_\pi \|D_\theta(x_t^{\mathrm{train}}, t) - D^*(x_t^{\mathrm{train}}, t)\|^2 \\ & + \mathbb{E}_\pi \|x_1 - D^*(x_t^{\mathrm{train}}, t)\|^2. \end{aligned} \tag{22}$$

The second term quantifies uncertainty about a paired endpoint even for an optimal predictor. Under an independent coupling at $t = 0$, $D^*(x_0, 0) = \mathbb{E}[x_1]$, which can fall between data modes.

Endpoint extrapolation and ODE integration. Let χ_r solve the frozen Vanilla ODE from $\chi_t = x$. Its terminal state and the endpoint extrapolation are

$$\chi_1 = x + \int_t^1 v_\theta(\chi_r, r) \mathrm{d}r, \qquad D_\theta(x,t) = x + (1-t) v_\theta(x,t). \tag{23}$$

If the velocity along this trajectory is differentiable with $\|\mathrm{d}v_\theta(\chi_r, r)/\mathrm{d}r\| \le A$ for $r \in [t, 1]$, then

$$\|\chi_1 - D_\theta(x, t)\| \le \tfrac{1}{2}A(1-t)^2. \tag{24}$$

This local extrapolation bound depends on the stated trajectory regularity. PCFM also uses endpoint extrapolation (Utkarsh et al., 2026); the persistent MAFM anchor incorporates successive predictions across changing queries.

Parameterization and source law. For $t < 1$, a directly trained endpoint predictor induces the velocity $v_\theta(x, t) = [D_\theta(x, t) - x]/(1-t)$. The two parameterizations have corresponding population optima under the linear path, while their cross-time finite-model losses carry different $(1-t)^2$ weights. The conditional-mean identity requires finite moments and the linear path, rather than a standard-Gaussian source. More general flow matching permits other source laws (Tong et al., 2023). MAFM draws x_0 from the source distribution used to train the frozen Vanilla model; task-specific Gaussian covariances are recorded with their training and sampling protocols.

Query region. For a fixed x_0, let $\mathcal{A}_t$ contain the compared anchor paths and define $\Omega_t = \{(1-t)x_0 + ta : a \in \mathcal{A}_t\}$. The stability result uses endpoint and adjoint error bounds on this common query region. These are local model assumptions: endpoint regression on the training path alone does not supply uniform error or Jacobian bounds over every later query.

A.3 Lagrange-multiplier correction

At a fixed anchor and time, write $J = J_{\mathcal{C}}(a)$ and $c = \mathcal{C}(a)$, and let $w = w_t$ be the candidate in Eq. (11). The base specialization sets $\eta = 0$. Under full row rank, the exact unregularized velocity is the unique solution of

$$\min_{u \in \mathbb{R}^d} \tfrac{1}{2}\|u - w\|^2 \quad \text{subject to} \quad Ju = -\kappa c. \tag{25}$$

For the Lagrangian $\mathcal{L}(u, \lambda) = \frac{1}{2}\|u - w\|^2 - \lambda^\top(Ju + \kappa c)$, stationarity gives $u = w + J^\top\lambda$. Substituting into the constraint gives $JJ^\top\lambda = -(Jw + \kappa c)$, which yields the main-text multiplier system in Eq. (12). The correction lies in $\mathrm{range}(J^\top)$ and retains the tangent component of the candidate velocity.

For $\delta > 0$, the regularized system instead minimizes

$$\tfrac{1}{2}\|u - w\|^2 + \frac{1}{2\delta}\|Ju + \kappa c\|^2. \tag{26}$$

Stationarity is $u - w + \delta^{-1}J^\top(Ju + \kappa c) = 0$. Setting $\lambda = -\delta^{-1}(Ju + \kappa c)$ gives $(JJ^\top + \delta I_m)\lambda = -(Jw + \kappa c)$ and $u = w + J^\top\lambda$. Regularization therefore implements a quadratic penalty on the prescribed residual rate.

Base-specialization endpoint filtering. Let $V_D(a, t; x_0)$ denote the exact unregularized base-MAFM velocity using endpoint field D. At the same anchor, source, and time,

$$V_{D+e}(a, t; x_0) - V_D(a, t; x_0) = \gamma(t)P_T(a)e(x, t), \qquad x = (1-t)x_0 + ta. \tag{27}$$

For the pointwise perturbation in Eq. (27), the two velocities share a, t, x_0, and hence J, P_T, and $\mathcal{C}(a)$. Subtracting their base-specialization forms in Eq. (14) leaves $\gamma(t)P_T(a)e(x, t)$. In particular, errors in $\mathrm{range}(J_{\mathcal{C}}(a)^\top)$ have no instantaneous effect on the velocity. With regularization $\delta > 0$, P_T is replaced by $A_\delta = I_d - J^\top(G + \delta I_m)^{-1}J$: normal errors are attenuated rather than eliminated. Its tangent eigenvalues are one and its normal eigenvalues are $\delta/(\sigma_i(J)^2 + \delta)$, so $\|A_\delta\| \le 1$.

A.4 Projected-adjoint effect and sparse observations

Proof of Proposition 2. At the fixed anchor and time in the proposition, write $J = J_{\mathcal{C}}(a)$, $G = JJ^\top$, and $P_T = I_d - J^\top G^{-1}J$. The candidate velocities with and without the adjoint satisfy

$$w_t^{(\eta)} = \gamma(t)[D_\theta(x, t) - a] - \eta(t)tg_t, \qquad w_t^{(0)} = \gamma(t)[D_\theta(x, t) - a]. \tag{28}$$

Equation (14) gives $V_\eta = P_T w_t^{(\eta)} - J^\top G^{-1}\kappa(t)\mathcal{C}(a)$ and the same expression with $w_t^{(0)}$ for V_0. Subtracting cancels the normal attraction term and yields $V_\eta - V_0 = -\eta(t)tP_T g_t$. Since $JP_T = 0$,

this difference induces no first-order residual change. Moreover, P_T is the orthogonal projector onto $\ker J$, with $\ker P_T = \mathrm{range}(J^\top)$. For $t > 0$ and $\eta(t) > 0$, the velocity difference therefore vanishes exactly when $g_t \in \mathrm{range}(J^\top)$. □

Corollary 4 (Sparse-observation pullback)**.** *After reordering coordinates, let $\mathcal{C}(z) = Hz - y$ with $H = [I_m\ 0]$, partition the physical query as $x = (x_O, x_U)$. Write $D_O(x,t) = HD_\theta(x,t)$. Then*

$$P_T g_t = \begin{bmatrix} 0 \\ J_{D_O,x_U}(x_t,t)^\top [D_O(x_t,t) - y] \end{bmatrix}. \tag{29}$$

Thus the adjoint can affect unobserved coordinates through the sensitivity of the predicted observed endpoint to those coordinates.

Proof. Here $HH^\top = I_m$ and $P_T = I_d - H^\top H = \mathrm{diag}(0, I_{d-m})$. The input adjoint is $g_t = J_{D_O}(x_t,t)^\top [D_O(x_t,t) - y]$. Partitioning this gradient into its x_O and x_U blocks and applying P_T gives Eq. (29). □

For the Poisson state $x = (U, F)$, x_U in this corollary contains both unobserved U coordinates and the forcing field F. A nonzero $J^\top_{D_O,x_F}[D_O - y]$ provides a model-mediated route from an observed-U residual to a tangent update of F. The ablation in Section 4.5 is consistent with this mechanism; the identity alone does not guarantee improved joint consistency.

Scope and diagnostics. At a common anchor, let $J = J_{\mathcal{C}}(a)$, $G = JJ^\top$, and $A_\delta = I_d - J^\top (G + \delta I_m)^{-1} J$. For an exact regularized multiplier solve with $\delta > 0$, the same fixed-state comparison gives

$$V_{\eta,\delta} - V_{0,\delta} = -\eta(t) t A_\delta g_t, \qquad J(V_{\eta,\delta} - V_{0,\delta}) = -\eta(t) t \delta (G + \delta I_m)^{-1} J g_t. \tag{30}$$

Thus regularization does not fully remove the adjoint's normal component. For a finite step with nonlinear $\mathcal{C}$, curvature can also make the next-step residuals differ, as in Eq. (17).

The per-unit-η instantaneous speed is $A_t = t\|P_T(a_t) g_t\|$. A relative diagnostic, when the base tracking speed is appreciable, is

$$R_t = \frac{\eta(t) A_t}{\gamma(t) \|P_T(a_t)[D_\theta(x_t,t) - a_t]\| + \varepsilon}, \qquad \varepsilon > 0. \tag{31}$$

The denominator uses base tracking rather than the guided $P_T w_t$. Neither score establishes that guidance is beneficial. In the discrete sampler, the realized tangent displacement also includes h; this quantity should be logged when assessing adjoint activity.

A.5 History weights for affine constraints

For $\mathcal{C}(a) = Ha - y$ with full-row-rank H, the constant tangent projector $P_T = I_d - H^\top (HH^\top)^{-1} H$ satisfies $P_T H^\top = 0$. Applying it to Alg. 1 removes the multiplier correction. For the base specialization $\eta = 0$, an exact unregularized update gives

$$\begin{aligned} P_T a^{(k+1)} &= (1 - \rho_k) P_T a^{(k)} + \rho_k P_T d^{(k)}, \\ c^{(k+1)} &= (1 - h\kappa_k) c^{(k)}, \end{aligned} \qquad c^{(k)} = \mathcal{C}(a^{(k)}), \quad \kappa_k = \kappa(t_k). \tag{32}$$

The tangential identity holds for any multiplier values, including regularized or inexact solves. For the normal identity, the exact unregularized solve gives $H(w^{(k)} + H^\top \lambda^{(k)}) = -\kappa_k c^{(k)}$; affinity then gives the stated recurrence. Writing $u^{(k)} = P_T a^{(k)}$ and $d_T^{(k)} = P_T d^{(k)}$, repeated substitution gives

$$u^{(N)} = \left[\prod_{j=0}^{N-1} (1 - \rho_j) \right] u^{(0)} + \sum_{i=0}^{N-1} \left[\rho_i \prod_{j=i+1}^{N-1} (1 - \rho_j) \right] d_T^{(i)}. \tag{33}$$

Empty products equal one. For $\gamma(t_k) \geq 0$, all weights are nonnegative and sum to one. The initial-state weight is $\exp[-h \sum_{j=0}^{N-1} \gamma(t_j)]$. This makes the role of the attraction schedule explicit: it determines how much past information remains in the tangential anchor. The endpoint estimates themselves depend on evolving queries. Active adjoint guidance adds $-h\eta(t_k) t_k P_T q^{(k)}$ to the tangent recurrence, with $q^{(k)}$ as in Alg. 1; the normal recurrence remains the one in Eq. (32).

A.6 Proof of Proposition 1

By the chain rule and Eq. (13),

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}(a_t) = J_{\mathcal{C}}(a_t)\dot{a}_t \tag{34}$$

$$= J_{\mathcal{C}}(a_t)w_t + G(a_t)\lambda_t. \tag{35}$$

With $\delta = 0$ and the exact multiplier solution, $G(a_t)\lambda_t = -[J_{\mathcal{C}}(a_t)w_t + \kappa(t)\mathcal{C}(a_t)]$. Substitution yields $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}(a_t) = -\kappa(t)\mathcal{C}(a_t)$. Solving this linear vector ODE gives Eq. (15).

For $\delta > 0$, use $\lambda = -(G + \delta I_m)^{-1}(J_{\mathcal{C}}w + \kappa\mathcal{C})$ and $I_m - G(G + \delta I_m)^{-1} = \delta(G + \delta I_m)^{-1}$ to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}(a_t) = \delta(G + \delta I_m)^{-1}J_{\mathcal{C}}w_t - \kappa(t)G(G + \delta I_m)^{-1}\mathcal{C}(a_t). \tag{36}$$

If the numerical multiplier leaves algebraic residual $r_\lambda = (G + \delta I_m)\widehat{\lambda} + J_{\mathcal{C}}w + \kappa\mathcal{C}$, then $G(G + \delta I_m)^{-1}r_\lambda$ adds a further forcing term to the residual dynamics.

A.7 Base-specialization local feedback stability

To isolate anchor memory, fix a source x_0, time $\bar{t}$, and affine constraint $Ha = y$. Write $a = p + a_\perp$, with $p \in T = \ker H$ and $a_\perp = H^\top(HH^\top)^{-1}y$, and define $F(p) = P_T D_\theta((1-\bar{t})x_0 + \bar{t}(p + a_\perp), \bar{t})$. Direct replacement uses $p^+ = F(p)$, while the relaxed persistent anchor uses $p^+ = (1-\rho)p + \rho F(p)$.

Proposition 5 (Local feedback stabilization)**.** *Suppose $F : T \to T$ is continuously differentiable near a fixed point p^*, and let $K = \mathrm{D}F(p^*)$. For $0 < \rho \leq 1$, the relaxed update has linearization $M_\rho = (1-\rho)I_T + \rho K$, with eigenvalues*

$$\Lambda_\rho(\zeta) = 1 - \rho + \rho\zeta, \qquad \zeta \in \mathrm{spec}(K). \tag{37}$$

The fixed point is locally asymptotically stable if $\max_{\zeta\in\mathrm{spec}(K)} |\Lambda_\rho(\zeta)| < 1$. In particular, a real mode $\zeta < -1$ is damped when $0 < \rho < 2/(1-\zeta)$.

Proof. Fix $\bar{t}$, x_0, and the feasible affine slice as in the proposition. For $a = p + a_\perp$, orthogonal projection of an endpoint onto $Ha = y$ is $a_\perp + P_T D_\theta((1-\bar{t})x_0 + \bar{t}a, \bar{t})$. Thus its tangential direct-replacement map is F, and the frozen-time MAFM map is $H_\rho(p) = (1-\rho)p + \rho F(p)$. For $\rho > 0$, $H_\rho(p) = p$ if and only if $F(p) = p$. Differentiating at p^* gives $\mathrm{D}H_\rho(p^*) = (1-\rho)I_T + \rho K$. When the endpoint estimator is differentiable at the associated query x^*, $K = \bar{t}\,P_T J_{D_\theta}(x^*, \bar{t})|_T$. The spectral mapping identity for this affine polynomial in K gives Eq. (37).

If every eigenvalue of M_ρ has modulus less than one, there is an equivalent norm on T whose induced matrix norm satisfies $\|M_\rho\|_* < 1$. Continuity of $\mathrm{D}H_\rho$ then makes H_ρ a contraction in a sufficiently small ball around p^* in this norm. Its iterates converge to p^* from this ball, proving local asymptotic stability. The spectral condition does not require a symmetric K; for nonnormal matrices it need not imply contraction at every step in the Euclidean norm.

For a possibly complex eigenvalue ζ,

$$|1 - \rho + \rho\zeta|^2 = 1 + 2\rho(\mathrm{Re}\,\zeta - 1) + \rho^2|1-\zeta|^2. \tag{38}$$

Hence a mode is strictly damped exactly when $\mathrm{Re}\,\zeta < 1$ and $0 < \rho < 2(1 - \mathrm{Re}\,\zeta)/|1-\zeta|^2$, subject also to $\rho \leq 1$ for the relaxed update considered here. This must hold for every mode to obtain the proposition's stability result. For real $\zeta < -1$ the condition reduces to $0 < \rho < 2/(1-\zeta)$, whereas direct replacement has amplification factor ζ and is unstable in that direction. For $0 < \zeta < 1$, the relaxed factor lies between ζ and one, so relaxation slows convergence in this mode. Real $\zeta > 1$ remains unstable for every $\rho > 0$. □

The scalar example $\zeta = -2$, $\rho = 1/2$ illustrates this spectral mechanism. The proposition concerns a fixed-time local map; the sampler uses a changing time and query, for which a uniform stability analysis would require additional bounds on the sequence of linearizations.

A.8 FINITE-STEP RESIDUAL CONTROL AND ACCUMULATED ERROR

Proof of Proposition 3. For $a^{(k+1)} = a^{(k)} + hu^{(k)}$, the fundamental theorem of calculus gives $c^{(k+1)} = c^{(k)} + hJ^{(k)}u^{(k)} + R^{(k)}$, where

$$R^{(k)} = h\int_0^1 [J_{\mathcal{C}}(a^{(k)} + rhu^{(k)}) - J^{(k)}]u^{(k)}\,\mathrm{d}r, \quad \|R^{(k)}\| \le \tfrac{1}{2}L_J h^2\|u^{(k)}\|^2. \tag{39}$$

The exact unregularized solve gives $J^{(k)}u^{(k)} = -\kappa_k c^{(k)}$, proving Eq. (17). For an affine constraint the Jacobian is constant, so $R^{(k)} = 0$. The factor $1 - h\kappa_k$ has magnitude less than one precisely when $0 < h\kappa_k < 2$, and is nonnegative when $0 \le h\kappa_k \le 1$. □

For a full-row-rank affine constraint $\mathcal{C}(a) = Ha - y$, exact unregularized normal correction can also be written as a split step:

$$\begin{aligned} a^{(k+1)} = a^{(k)} &+ \rho_k P_T(d^{(k)} - a^{(k)}) \\ &- h\eta(t_k)t_k P_T q^{(k)} \\ &- \alpha_k H^\top(HH^\top)^{-1}\mathcal{C}(a^{(k)}). \end{aligned} \tag{40}$$

Since $HP_T = 0$, this step has the exact affine residual law

$$Ha^{(k+1)} - y = (1 - \alpha_k)(Ha^{(k)} - y). \tag{41}$$

The explicit Euler update in Alg. 1 uses $\alpha_k = h\kappa_k$. With frozen κ_k, the alternatives $\alpha_k = 1 - e^{-h\kappa_k}$ and $\alpha_k = 1$ give exponential residual decay and one-step affine feasibility, respectively. The split form also applies when $\rho_k = 0$.

For a regularized or inexact solve, let $G^{(k)} = J^{(k)}(J^{(k)})^\top$ and define the algebraic residual $r_\lambda^{(k)} = (G^{(k)} + \delta I_m)\widehat{\lambda}^{(k)} + J^{(k)}w^{(k)} + \kappa_k c^{(k)}$. With $u^{(k)} = w^{(k)} + (J^{(k)})^\top\widehat{\lambda}^{(k)}$, direct substitution gives

$$c^{(k+1)} = (1 - h\kappa_k)c^{(k)} + he^{(k)} + R^{(k)}, \tag{42}$$

$$e^{(k)} = \delta(G^{(k)} + \delta I_m)^{-1}(J^{(k)}w^{(k)} + \kappa_k c^{(k)}) \tag{43}$$

$$+ G^{(k)}(G^{(k)} + \delta I_m)^{-1}r_\lambda^{(k)}. \tag{44}$$

Here the inverse exists for $\delta > 0$, or for $\delta = 0$ with full-row-rank $J^{(k)}$. The two terms in $e^{(k)}$ separate regularization from linear-solve error. If the smallest eigenvalue of $G^{(k)}$ is $\ell_k > 0$, then $\|e^{(k)}\| \le [\delta/(\ell_k + \delta)]\|J^{(k)}w^{(k)} + \kappa_k c^{(k)}\| + \|r_\lambda^{(k)}\|$.

Set $q_k = |1 - h\kappa_k|$ and $f_k = h\|e^{(k)}\| + \frac{1}{2}L_J h^2\|u^{(k)}\|^2$, using a common Lipschitz bound over the step segments. Iteration yields

$$\|c^{(N)}\| \le \left(\prod_{k=0}^{N-1} q_k\right)\|c^{(0)}\| + \sum_{i=0}^{N-1} f_i \prod_{j=i+1}^{N-1} q_j. \tag{45}$$

In particular, for constant $\kappa > 0$, $0 < h\kappa \le 1$, $\|u^{(k)}\| \le U$, and $\|e^{(k)}\| \le \varepsilon_e$, this becomes

$$\|c^{(N)}\| \le (1 - h\kappa)^N\|c^{(0)}\| + \frac{\varepsilon_e + \frac{1}{2}L_J hU^2}{\kappa}[1 - (1 - h\kappa)^N]. \tag{46}$$

Under these uniform bounds, the estimate quantifies both the contraction of the initial residual and the accumulated effects of curvature and numerical corrections for the explicit anchor step in Alg. 1.

A.9 ENDPOINT STABILITY ON A COMMON QUERY REGION

Proposition 6 (Local model-perturbation bound)**.** *Let a_t^* use a reference endpoint field D^* and its input adjoint g^*, and let a_t^θ use D_θ and g^θ. Couple the paths through the same source, constraint, and schedules, with $\delta = 0$, exact multiplier solves, and full-row-rank $J_{\mathcal{C}}$ along a common trajectory neighborhood. Suppose the anchor vector field has an integrable $L(t)$-Lipschitz bound there and, at a common state in Ω_t, $\|D_\theta - D^*\| \le \varepsilon_D(t)$ and $\|g^\theta - g^*\| \le \varepsilon_g(t)$, with finite error integrals. Then*

$$\|a_t^\theta - a_t^*\| \le e^{\int_0^t L(q)\mathrm{d}q}\left[\|a_0^\theta - a_0^*\| + \int_0^t (\gamma(r)\varepsilon_D(r) + \eta(r)r\varepsilon_g(r))\,\mathrm{d}r\right]. \tag{47}$$

Proof. Let $V_{D,g}(a,t;x_0)$ denote the exact unregularized anchor vector field after inserting Eq. (6). At a common anchor,

$$\|V_{D_\theta,g^\theta}(a,t;x_0) - V_{D^*,g^*}(a,t;x_0)\| \le \gamma(t)\varepsilon_D(t) + \eta(t)t\varepsilon_g(t), \tag{48}$$

because the orthogonal projector $P_T(a)$ has norm at most one. Comparing the two paths and using the anchor-Lipschitz bound gives

$$\|V_{D_\theta,g^\theta}(a_t^\theta,t;x_0) - V_{D^*,g^*}(a_t^*,t;x_0)\| \le L(t)\|a_t^\theta - a_t^*\| + \gamma(t)\varepsilon_D(t) + \eta(t)t\varepsilon_g(t). \tag{49}$$

Grönwall's inequality gives Eq. (47). □

For exact unregularized solves, Eq. (27) also permits replacing $\varepsilon_D(t)$ in the trajectory bound by a uniform bound on $\|P_T(a)[D_\theta(x,t) - D^*(x,t)]\|$ over the same queried region. For $\eta \equiv 0$, the adjoint-error term vanishes. If the query region extends beyond the training-path support, the reference field D^* requires a specified extension with the assumed regularity. The optional target-law comparison in Eq. (20) addresses a separate question from this pathwise model-perturbation bound.

B Additional algorithmic details

B.1 Optional adjoint-step clipping

The default clip_k in Alg. 1 is the identity. For full-row-rank $J_{\mathcal{C}}(a^{(k)})$, an optional radius τ_k and norm $\|\cdot\|_\star$ define

$$\mathrm{clip}_k(v) = \min\left\{1, \frac{\tau_k}{h\eta(t_k)t_k\|P_T(a^{(k)})v\|_\star + \varepsilon}\right\} v, \qquad \varepsilon > 0. \tag{50}$$

Taking $\tau_k = \infty$ recovers the default. A finite radius bounds the projected adjoint displacement in $\|\cdot\|_\star$; the VJP is skipped when $\eta(t_k)t_k = 0$.

B.2 Poisson adjoint scaling

The main text uses the ordinary residual norm and input gradient. The reported partially observed Poisson sampler uses fixed channel scales $s_U, s_F > 0$ from the training data. With m observed U entries and $\mathcal{C}(z) = Hz - y$, its endpoint energy and physical guidance direction are

$$\mathcal{E}_t^{\mathrm{P}}(x) = \frac{\|HD_\theta(x,t) - y\|^2}{2ms_U^2}, \qquad g_t^{\mathrm{P}} = \big(s_U^2[\nabla_x \mathcal{E}_t^{\mathrm{P}}]_U, s_F^2[\nabla_x \mathcal{E}_t^{\mathrm{P}}]_F\big). \tag{51}$$

Its discrete adjoint step uses $q^{(k)} = \mathrm{clip}_k(g_{t_k}^{\mathrm{P}})$, with the reported per-step guidance coefficient corresponding to $h\eta(t_k)$ in the split update of Eq. (40). The channel factors adjust the relative U and F motion; they are part of this numerical specialization, not needed for the core dynamics. The multiplier still enforces the unscaled observation equation $Ha = y$. The projected-adjoint identity remains valid with g_t^{P} substituted for g_t. The Poisson clipping norm is the root mean square of componentwise v/s, where s equals s_U or s_F by channel; its radius is the configured trust factor times the norm of the tangent tracking displacement $P_T(a^{(k)})(d^{(k)} - a^{(k)})$.

C Experimental protocols

This section specifies the physical data, equality conditions, and evaluation for the six examples in Section 4. In every case, the unconditional model and each constrained sampler act on the same complete physical object. The numerical PDE solvers generate reference data and measure physical consistency; they do not supply an unobserved field to the sampler. Table 3 separates structure learned from training data from information explicitly supplied at inference. Here τ denotes physical time, while t denotes the flow or sampling time. The notation $\langle f \rangle_\Omega$ means the spatial average, evaluated as the corresponding grid average, and ξ denotes the Burgers spatial coordinate. The Euler and controlled-NS entries describe their fixed-energy, energy-only tests.

Table 3: Equations, learned structure, and supplied conditions for the six benchmarks.

Benchmark	Equation or data distribution	Structure learned from data	Explicit inference-time condition	Constraint type
Sphere–plane	$X \sim \mathrm{Unif}(\mathbb{S}^2)$	Uniform unit-sphere distribution, $\|X\|_2 = 1$.	Plane $X_3 = 0$; the desired support is the intersection circle.	Affine, single-coordinate algebraic
Poisson	$-\Delta U = F$, $U\|_{\partial\Omega} = 0$	Joint (U, F) samples satisfy the discrete Poisson equation and homogeneous boundary conditions.	$H_U U = y_{\mathrm{obs}}$; only U is observed, with no direct observations of F.	Sparse, local linear observations
Burgers	$\partial_\tau u + \partial_\xi(u^2/2) = 0.01\partial_{\xi\xi} u$	Periodic viscous dynamics and their trajectory statistics.	Prescribed spatial mean $\langle u(\tau)\rangle_\Omega = \mu_\star$ at every frame.	Framewise global linear integral
Reaction–diffusion	$\partial_\tau U = 0.1\Delta U + \mathcal{R}$, $\partial_\tau V = 10\Delta V - \mathcal{R}$	No-flux nonlinear dynamics, positive concentration support, and trajectory statistics.	Prescribed total mass $\langle U(\tau) + V(\tau)\rangle_\Omega = \mu_\star$ at every frame.	Framewise global linear integral
Euler	$\partial_\tau \boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} + \nabla p = 0$	Incompressible, zero-mean velocity trajectories generated by Euler dynamics.	Prescribed kinetic energy at every frame; incompressibility and zero mean are left to the learned prior.	Global nonlinear quadratic
Controlled NS	$\partial_\tau \boldsymbol{u} = -\mathsf{P}[(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}] + \nu\Delta\boldsymbol{u} + \beta_{\mathrm{NS}}(\tau)\boldsymbol{u}$	Incompressible, zero-mean velocity trajectories generated by controlled NS dynamics.	Prescribed kinetic energy at every frame; incompressibility and zero mean are left to the learned prior.	Global nonlinear quadratic

Shared comparison and metrics. Methods within a case use the same frozen unconditional model, paired Gaussian sources, prescribed conditions, and nominal number of function evaluations (NFE). A constraint error (CE) measures the residual of the condition supplied at sampling. It is distinct from a physical-consistency error: for PDE trajectories, we evolve each generated initial state with an independent numerical solver and compare its later frames with the generated trajectory (own-IC relative L_2); for Poisson, we solve from the generated F and compare with the generated U. Distributional and paired-truth diagnostics are reported separately. The Poisson protocol requires both mean and q99 observation CE below 10^{-6}. For concise comparisons, we mark mass CE below 10^{-6} and fixed-energy error below 10^{-5} as numerically feasible; these two bands are reporting conventions. The sphere's 10^{-3} plane tolerance belongs only to its joint-support diagnostic, not to a universal feasibility rule. Once a method reaches the applicable numerical band, the physical and distributional metrics distinguish its samples.

C.1 Sphere–plane intersection

The training law and the inference-time plane condition are

$$X \sim \mathrm{Unif}(\mathbb{S}^2) \subset \mathbb{R}^3, \qquad X_3 = 0. \tag{52}$$

We draw a standard Gaussian vector and normalize it to unit length, then train an unconditional velocity MLP on full Cartesian three-vectors with a Gaussian source and linear flow-matching bridge. The width-192, depth-three MLP uses velocity MSE, batch size 512, and 100,000 optimizer updates for its frozen checkpoint. It receives no sphere equation or plane condition. At inference the only supplied condition is the plane equation above; its intersection with the learned support is the unit circle in that plane. There is no physical-time evolution or boundary condition in this algebraic example. The paired 10,000-sample comparison uses 100 NFE and independent uniform-circle references. We report plane residual, radial and angular distribution diagnostics, sliced Wasserstein distance (SWD), and joint support coverage over several tolerances. The radius is an evaluation diagnostic, not an additional constraint given to the sampler.

C.2 Partially observed Poisson equation

On $\Omega = (0,1)^2$, the data satisfy

$$\begin{aligned} -\Delta U &= F \quad \text{in } \Omega, \\ U &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{53}$$

Each physical state is the joint pair (U, F) on the 64×64 interior grid ($h = 1/65$); boundary nodes are fixed rather than generated. To draw F, independent grid Gaussian noise is smoothed by a reflected Gaussian kernel of width three grid cells, scaled to spatial RMS about 20, and given an independent Gaussian constant offset with standard deviation 5. A float64 five-point Dirichlet solve supplies U. The joint physical-field network predicts both channels and minimizes a channel-scaled full-field endpoint MSE. It uses 8,192 training and 1,024 validation pairs; the width-32, depth-three joint U-Net trains with batch size 32. The final 40,000-update AdamW run selects its checkpoint by validation loss. Fixed channel scales balance the two physical units and do not change the generated state space.

At inference, the supplied observation equation is

$$H_U U = y_{\text{obs}}, \tag{54}$$

which selects either 64 or 256 distinct interior U values; F is never observed. Two fixed sensor layouts are evaluated for each count, with 64 independently generated conditions and eight paired draws per condition, at 100 NFE. The held-out pair is used only to form observations and evaluation targets. Observation CE is normalized by the truth-U RMS; PDE consistency is $\|\widehat{U} - S_h(\widehat{F})\|_2 / \|S_h(\widehat{F})\|_2$, where S_h is the same discrete boundary-value solver. Errors against the unobserved U and F are supplementary paired-truth diagnostics, not conditional-distribution distances. The Poisson-specific adjoint scaling and tangent-step radius are defined in Appendix B.2.

C.3 Periodic viscous Burgers equation

We solve

$$\partial_\tau u + \partial_\xi \left(\frac{u^2}{2} \right) = 0.01 \partial_{\xi\xi} u \tag{55}$$

on $\xi \in [0, 1)$ with periodic boundary conditions, recording 65 frames over $\tau \in [0, 1]$ on 512 grid points. Each initial field is $u_0 = m + Ag$: $m \sim \text{Unif}[0.3, 1.3]$, $A \sim \text{Unif}[0.3, 0.8]$, and g is a zero-mean, unit-RMS smooth periodic Gaussian field built from sine and cosine frequencies 1–8 with amplitude proportional to $(1 + k^2)^{-1}$. A float64 Cole–Hopf solve generates the physical trajectories. Its Fourier calculations are numerical solver operations; the jointly trained endpoint network sees complete $[65, 1, 512]$ trajectories. Training uses 4,096 trajectories and 512 validation trajectories. A joint circular-convolution U-Net mixes all frames as channels and uses a full-trajectory endpoint MSE with batch size 32. The reported model is the selected 52,000-update checkpoint after the registered time-allocation continuation.

The inference condition prescribes the spatial mean $m_\star$ of every generated frame. Held-out tests use $m_\star \in \{0.55, 0.90, 1.05\}$, with 256 independent sources at each value and 100 NFE. For each raw generated initial field, an independent Burgers solve produces the own-IC future reference. We compare the remaining 64 frames by relative L_2; mean-subtracted fluctuation error and distributional statistics check whether a low error merely reflects correct means. Mean CE is the framewise RMS deviation from $m_\star$, normalized by $|m_\star|$.

C.4 Mass-conserving reaction–diffusion system

On $\Omega = [0, 10]^2$ with no-flux boundaries, the two concentrations obey

$$\begin{aligned} \partial_\tau U &= 0.1\Delta U + \mathcal{R}(U, V), \qquad \partial_\tau V = 10\Delta V - \mathcal{R}(U, V), \\ \mathcal{R}(U, V) &= \left(0.067 + \frac{U^2}{1 + U^2} \right) V - U. \end{aligned} \tag{56}$$

The opposite reaction signs and no-flux boundary conserve the spatial mean of $U + V$. Initial states are positive: a smooth random total concentration is normalized to a mass drawn uniformly from $[2, 2.5]$, then split into the two species by a smooth random fraction with one to three reflected Gaussian

bumps. A conservative cell-centered Neumann solver uses diffusion substeps and Runge–Kutta reaction steps to produce 33 frames over $\tau \in [0, 40]$ on a native 64×64 grid. The joint endpoint U-Net learns complete $[33, 2, 64, 64]$ trajectories from 4,096 training examples, with 512 validation examples and 60,000 optimizer updates. Its width is 48, its batch size is eight, and full-field endpoint MSE trains both species and all frames together; validation loss selects the frozen checkpoint.

The sampler is told only the desired total mass at every frame. The held-out values are 2.15, 2.25, and 2.35, each with 128 paired sources at 100 NFE. CE is the framewise residual of the mean of $U + V$; an independent rollout from each generated first frame gives own-IC relative L_2 over the remaining 32 frames. Species-wise errors, distributional distances, and the fraction of negative generated concentrations are supplementary checks. Negative values are counted on the raw output rather than clipped.

C.5 Two-dimensional Euler flow

The data solve

$$\begin{aligned} \partial_\tau \boldsymbol{u} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} + \nabla p = 0, \\ \nabla \cdot \boldsymbol{u} = 0 \end{aligned} \tag{57}$$

on the periodic torus $[0, 2\pi)^2$. Smooth random velocity fields combine randomly phased low-frequency patterns and smoothed noise; they are projected to zero-mean, divergence-free fields and normalized to the common kinetic-energy level before numerical rollout. The resulting training objects are nine complete two-component velocity frames on a 49×49 grid over $\tau \in [0, 1]$. A physical-field endpoint network learns their joint distribution using full-field MSE and an auxiliary full-field spectral error. It has width 48, four spectral blocks, and 12 modes per spatial direction; these features are internal computations, not generated states. Training has 4,096 histories, 1,024 validation histories, batch size 16, and 48,000 updates.

The energy-only inference test supplies the target kinetic energy at every frame; divergence-free and zero-mean structure must still come from the learned law. On 512 held-out paired sources at 100 NFE, we measure energy error, divergence, spatial mean, field and feature distribution distances, and own-IC rollout error from each generated first frame. The separate joint-oracle comparison supplies divergence and zero mean as additional conditions and is reported in Table 9; it does not enter the energy-only ranking.

C.6 Energy-controlled Navier–Stokes flow

The controlled system uses the same periodic domain, grid, nine-frame horizon, and initial-field family as Euler, with

$$\partial_\tau \boldsymbol{u} = -\mathsf{P}[(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}] + 0.02\Delta\boldsymbol{u} + \beta_{\mathrm{NS}}(\tau)\boldsymbol{u}, \qquad \beta_{\mathrm{NS}}(\tau) = -\frac{\langle \boldsymbol{u}, -\mathsf{P}[(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}] + 0.02\Delta\boldsymbol{u}\rangle}{\langle \boldsymbol{u}, \boldsymbol{u}\rangle}. \tag{58}$$

Here P is the divergence-free, zero-mean Leray projection; the feedback offsets instantaneous energy change. The solver constructs complete Cartesian velocity histories, and the same physical-field architecture and training sizes as Euler learn them jointly. At inference, only framewise energy is imposed. The fixed-energy 512-sample held-out comparison uses 100 NFE and the same constraint, distributional, and own-IC diagnostics as Euler. These fixed-energy checkpoints do not test generalization across unseen energy levels.

D Supplementary results and diagnostics

The tables below give the extended metrics behind the compact main-text comparison. They include the 256-sensor Poisson setting, matched-model ablations, and the separate three-constraint flow comparison. The subsequent diagnostics and sample panels clarify particular behaviors without changing the evaluation conditions in Appendix C.

Main-figure visualization protocol. In Figure 3(a), circles mark the shared observed sites; the columns show generated fields, sensor mismatch, and $|\widehat{U} - S_h(\widehat{F})|$. The numerical solver S_h is used only for evaluation, and no hidden paired field is supplied as a visual target. The illustrated

PCFM row uses PCFM-Relaxed in the Poisson panel and PCFM-Base in the controlled-NS panel, the registered variants for those tasks. Fields and residual maps show individual samples; the tables report full evaluation sets.

D.1 EXTENDED METRIC TABLES

Table 4's plane indicator uses the 10^{-3} tolerance inside Joint coverage. In Table 5, CE is the mean observation residual and PDE is the relative discrepancy between the generated U and the numerical solution from its generated F; two fixed sensor layouts contribute to each density. In the Burgers table CE is the mean-trajectory mass residual. Burgers and reaction–diffusion relative L_2 entries are percentages, including their fluctuation and q95 columns.

Table 4: Complete sphere–plane results (10,000 samples).

Method	mean $\lvert z\rvert$	q99 $\lvert z\rvert$	Radial W_1	SWD	MMD^2	Joint
Vanilla FM	0.5009	0.2877	0.2158	0.1281	0.05516	0.06%
Terminal projection	0	0	0.2158	0.1078	0.03237	6.47%
PCFM-Base	0	0	0.006273	0.009357	0.000559	56.57%
PCFM-Relaxed	2.103e−5	1.358e−5	0.006328	0.009551	0.000562	56.08%
ECI (operator-equivalent)	0	0	0.006273	0.009357	0.000559	56.57%
Stop-grad. guidance	0	0	0.006310	0.01885	0.000768	57.35%
Full-grad. guidance	2.521e−6	2.122e−6	0.003888	0.02417	0.000930	70.52%
MAFM	3.372e−6	2.866e−6	0.004929	0.009568	0.000534	72.97%
MAFM + projection	0	0	0.004929	0.009568	0.000534	72.97%

Table 5: Complete Poisson results at two sensor densities.

	64 sensors		256 sensors	
Method	CE	PDE RelL2	CE	PDE RelL2
Vanilla FM	1.395	1.128%	1.425	1.128%
Terminal projection	0	22.73%	0	46.31%
PCFM-Base	0	11.51%	0	18.05%
PCFM-Relaxed	0	10.33%	0	16.07%
ECI	0	11.51%	0	18.05%
ECI-2	0	11.50%	0	18.14%
ECI-4	0	11.50%	0	18.19%
Stop-grad. guidance	3.291e−9	5.554%	1.350e−9	5.636%
Full-grad. guidance	0.610	1.050%	0.371	0.974%
MAFM	1.721e−7	1.705%	1.751e−7	2.252%
MAFM + projection	0	1.705%	0	2.252%

Table 6: **Complete periodic Burgers results** (768 independent samples). CE is the mean-trajectory mass error. RelL2 and fluctuation RelL2 are percentages.

Method	RelL2	CE	Fluct. RelL2	Field SWD	MMD^2
Vanilla FM	1.134%	3.565e−1	3.163%	0.2695	0.2323
Terminal projection	26.97%	5.490e−8	74.48%	0.07782	0.0883
PCFM-Base	7.610%	5.392e−8	13.48%	0.1776	0.5160
PCFM-Relaxed	7.610%	3.169e−6	13.47%	0.1777	0.5163
ECI (operator-equivalent)	7.610%	5.392e−8	13.48%	0.1776	0.5160
ECI-2	7.537%	5.397e−8	13.38%	0.1766	0.5125
ECI-4	7.890%	5.396e−8	13.93%	0.1834	0.5333
ECI-5 (resampled)	2.708%	5.370e−8	6.965%	0.06150	0.1347
Stop-grad. guidance	1.030%	5.382e−8	16.80%	0.2876	0.8484
Full-grad. guidance	17.18%	2.634e−2	23.78%	0.2901	0.8569
MAFM	0.9648%	4.216e−8	4.828%	0.1574	0.4556
MAFM + projection	0.9648%	5.320e−8	4.828%	0.1574	0.4556

Table 7: **Complete mass-conserving reaction–diffusion results** (384 independent samples). RelL2 and q95 RelL2 are percentages.

Method	Field SWD	RelL2	CE	q95 RelL2	Negative fraction
Vanilla FM	0.09650	1.312%	5.501e−2	5.201%	4.48e−6
Terminal projection	0.08529	10.39%	5.837e−8	41.51%	8.75e−5
PCFM-Base	0.1303	3.030%	5.894e−8	10.26%	1.79e−3
PCFM-Relaxed	0.1157	2.458%	3.720e−4	8.482%	1.37e−3
ECI	0.1246	2.962%	5.855e−8	10.18%	1.78e−3
Stop-grad. guidance	0.1220	6.040%	5.922e−8	23.45%	1.53e−6
Full-grad. guidance	0.06375	1.563%	1.695e−3	5.878%	4.88e−6
MAFM	0.06659	2.740%	1.134e−7	11.07%	2.33e−4
MAFM + projection	0.06659	2.740%	6.773e−8	11.07%	2.33e−4

Task	Variant	CE	Physical RelL2	Field SWD	Δ RelL2	Paired 95% CI
Poisson	Full MAFM	1.721e−7	0.01705	–	–	–
	w/o adjoint	1.721e−7	0.3803	–	0.3632	[0.3521, 0.3751]
Burgers	Full MAFM	4.216e−8	0.009648	0.1574	–	–
	Reset anchor	0.008997	0.09437	0.1767	0.08473	[0.07948, 0.09028]
	w/o support	1.181e−7	0.02703	0.3135	0.01738	[0.01652, 0.01828]
	w/o normal	0.2698	0.01286	0.2750	0.003207	[0.002873, 0.003559]
Reaction–diffusion	Full MAFM	1.134e−7	0.02740	0.06659	–	–
	Reset anchor	0.04413	0.06145	0.2110	0.03405	[0.02654, 0.04189]
	w/o support	1.147e−7	0.1058	0.2102	0.07839	[0.07624, 0.08044]
	w/o normal	0.03186	0.02190	0.09998	−0.005500	[−0.007283, −0.003750]

Table 8: Core-ablation aggregates with paired confidence intervals.

The tables report values to at most four significant digits. Table 5 adds the 256-sensor comparison. Table 8 reports 10,000-resample paired intervals; Poisson resamples conditions while retaining layouts and draws, and the dynamical tasks resample sample IDs within each prescribed-mass stratum. Its Δ column is ablation minus Full MAFM: Poisson uses PDE-consistency relative L_2, while the dynamical cases use own-IC relative L_2. Field SWD was not part of this Poisson ablation comparison.

D.2 Three-constraint oracle comparison

Table 1c is the complete 512-sample energy-only test: no omitted energy-only method has been used to determine its ranking. The corresponding joint-oracle comparison below supplies energy, divergence, and zero mean simultaneously. It records projection behavior and the effect of additional analytic information, but it is not part of the energy-only ranking.

Table 9: Three-constraint oracle comparison for Euler and controlled NS (64 samples each).

Case	Method	Feature SWD	Field SWD	RelL2	Div.	Mean	Energy
Euler	Vanilla FM	.1521	.06619	.03680	.06346	.01314	.01863
	Terminal projection	.1521	.06630	.03317	1.45e-6	2.86e-8	4.78e-7
	PCFM-Base	.4237	.1148	.1248	1.48e-6	1.65e-8	4.77e-7
	PCFM-Relaxed	.4265	.1161	.1243	.007660	.000965	.001822
	ECI (operator-equivalent)	.4237	.1148	.1248	1.48e-6	1.65e-8	4.77e-7
	Stop-grad. guidance	.4267	.09845	.04287	.05980	.01609	.02623
	Full-grad. guidance	.1844	.06669	.03411	.06015	.01057	.01597
	MAFM	.1572	.06763	.03196	1.62e-6	1.86e-7	7.15e-6
	MAFM + projection	.1572	.06763	.03196	1.44e-6	2.22e-8	4.78e-7
NS	Vanilla FM	.06904	.04946	.03462	.06236	.01109	.01811
	Terminal projection	.06335	.04949	.03087	1.45e-6	2.22e-8	4.77e-7
	PCFM-Base	.3747	.09432	.1123	1.47e-6	1.79e-8	4.73e-7
	PCFM-Relaxed	.3869	.09466	.1122	.007071	.000961	.001789
	ECI (operator-equivalent)	.3747	.09432	.1123	1.47e-6	1.79e-8	4.73e-7
	Stop-grad. guidance	.4051	.07969	.03545	.04859	.008778	.02011
	Full-grad. guidance	.2518	.06222	.02887	.05039	.006922	.01725
	MAFM	.09609	.05057	.02938	1.62e-6	1.89e-7	7.18e-6
	MAFM + projection	.09609	.05057	.02938	1.44e-6	2.14e-8	4.66e-7

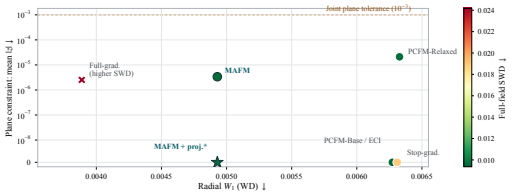

Figure 4: Sphere–plane radial distance, plane residual, and distributional distance.

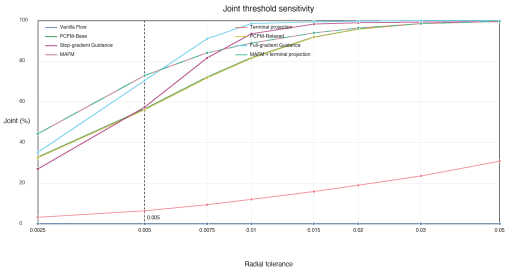

Figure 5: Sphere–plane joint coverage at varying tolerances.

D.3 Additional diagnostics

Matched ablations. The variants in Table 8 share each case's frozen checkpoint, paired sources, conditions, and 100 NFE. Each own-IC reference is solved from the initial condition produced by that variant. This isolates the inference mechanism from changes in model training or reference selection.

Terminal projection. An optional final projection sharpens numerical feasibility without explaining the consistency advantage: sphere SWD and the Burgers and reaction–diffusion own-IC errors barely change. In the 512-sample flow tests, mean energy error falls from about 6.9×10^{-6} to about 2.4×10^{-7} for both Euler and controlled NS, while own-IC error and field SWD are effectively unchanged.

Sphere threshold sensitivity. Figure 4 places radial distance and plane residual on its axes, with color for full-field SWD. The full-gradient point attains a slightly smaller radial W_1 but a poorer distributional match; the projected MAFM point reaches zero plane residual with effectively unchanged distances. This radial tail also worsens angular and kernel distances. MAFM has the highest Joint coverage at the strict 0.0025 and 0.005 support tolerances; conclusions at looser thresholds are reported as sensitivity analysis rather than universal support claims.

Burgers uncertainty. The independent-sample bootstrap preserves MAFM's advantage in own-IC rollout error. The analysis also exposes a remaining failure mode: generated fluctuation amplitude is slightly contracted even when the mean trajectory is exact.

Reaction–diffusion nonnegativity. Nonnegativity is not supplied to any sampler. MAFM's negative-value fraction is 2.33×10^{-4}: better than PCFM variants but worse than Vanilla and stop-gradient guidance. This auxiliary diagnostic does not reverse the primary ranking and remains a target for future work.

PCFM normalization. Changing the relaxed-PCFM normalization can improve selected distribution distances, but the same changes can worsen feasibility and own-IC rollout error. The paper therefore reports the official Eq. 6 arm and labels operator-equivalent duplicates instead of counting them as independent evidence.

D.4 Inference-time computational cost

All samplers in a given case reuse the same frozen physical-field model, so the comparisons below concern inference rather than repeated model training. Table 10 transcribes recorded remote sampling times for two representative tasks: one fixed 64-sensor Poisson layout with 64 conditions and eight draws per condition, and the 384-sample reaction–diffusion test over three prescribed masses. The timers cover sample generation and transfer to host memory, but exclude independent PDE-reference solves and metric aggregation. Time units are stated per row because the two tasks use different physical states and batch protocols; their absolute times should not be compared across tasks.

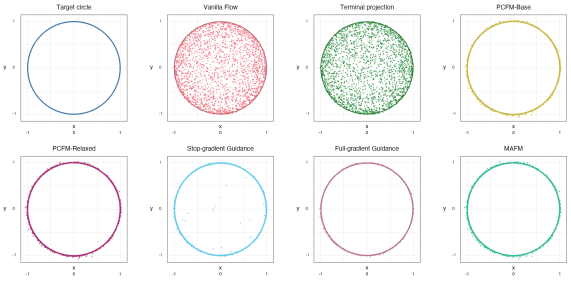

Figure 6: Sphere–plane samples.

Case	Sampler	NFE	Input VJPs	Sampling time	CE	Physical RelL2
Poisson (64 sensors)	Vanilla FM	100	0	6.615 s/512	1.366	1.128%
	PCFM-Relaxed	100	0	15.77 s/512	0	10.26%
	MAFM	100	100	17.88 s/512	1.696e−7	1.710%
Reaction–diffusion	PCFM-Base	100	0	0.04249 s/sample	5.894e−8	3.030%
	ECI	100	0	0.04361 s/sample	5.855e−8	2.962%
	MAFM	100	0	0.04375 s/sample	1.134e−7	2.740%
	Full-grad. guidance	100	100	0.08584 s/sample	1.695e−3	1.563%

Table 10: Recorded inference cost and physical consistency for two frozen-model evaluations. Poisson values use one fixed 64-sensor layout (512 samples); reaction–diffusion values cover 384 samples. CE is the supplied-constraint error. Lower physical RelL2 is preferable among numerically feasible methods.

On Poisson, MAFM's active adjoint requires an input VJP at each of 100 steps and takes 17.88 seconds for the 512 outputs, compared with 15.77 seconds for PCFM-Relaxed. Both satisfy the observation tolerance, while MAFM reduces PDE RelL2 from 10.26% to 1.710%. The unguided Vanilla sampler is faster, but its observation CE is far outside the feasible band.

On reaction–diffusion, base MAFM needs no input VJP and takes 0.04375 seconds per sample, close to ECI's 0.04361 and PCFM-Base's 0.04249, while reducing feasible own-IC RelL2 to 2.740% from 2.962% and 3.030%, respectively.

These two evaluations show that MAFM can improve feasible physical consistency with limited additional sampling time relative to constrained baselines. The gain is especially pronounced on Poisson, where the adjoint also increases cost. The records do not establish a uniform runtime or memory advantage across all six tasks.

D.5 Qualitative samples

These panels illustrate individual outputs from the six cases. The sphere panel checks recovery of the circular intersection from a supplied plane and a learned sphere prior; the radius was never imposed during sampling. The Poisson U and F panels show corresponding joint-field samples. The trajectory and vorticity panels show selected outputs rather than the full evaluation set, so quantitative conclusions follow from Figure 4, Table 4, and the threshold sensitivity in Figure 5.

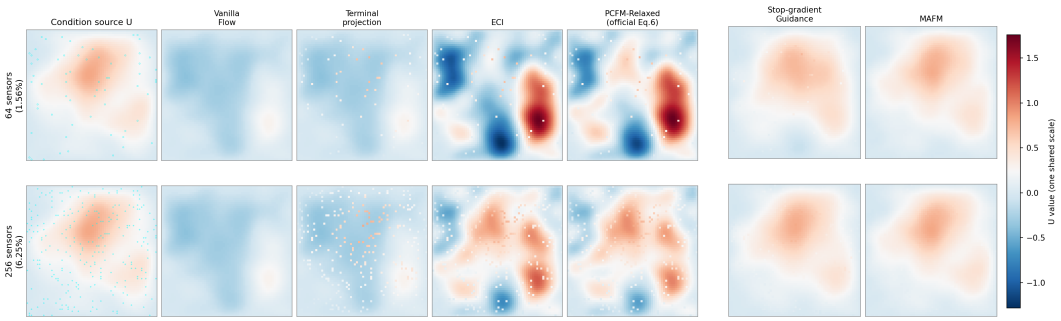

Figure 7: Poisson generated solution fields U under partial observations. One representative ECI column is shown; the ECI-2 and ECI-4 variants remain in Table 5.

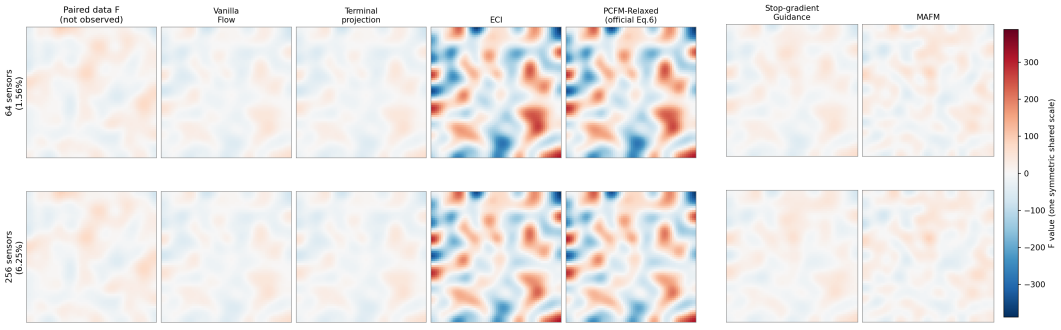

Figure 8: Poisson source fields F paired with Figure 7, with the same representative ECI column.

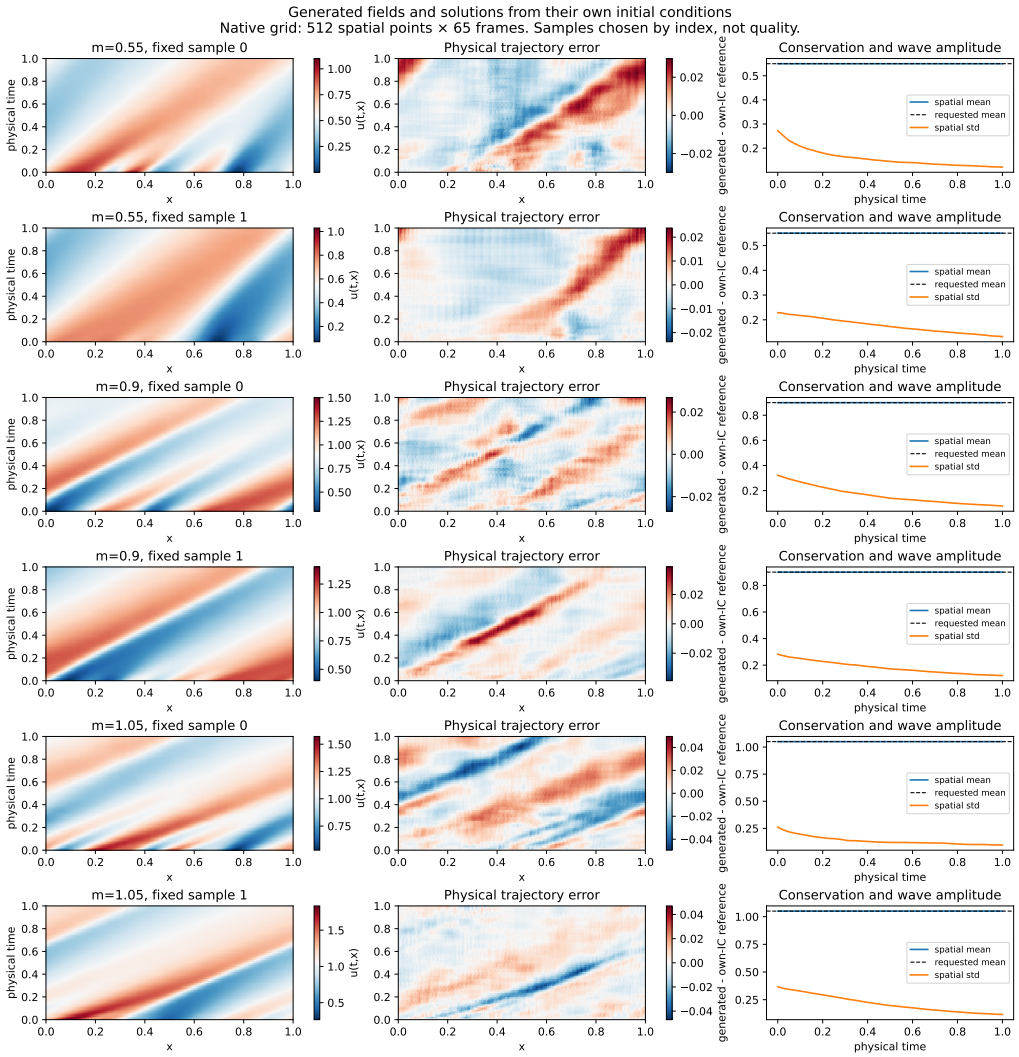

Figure 9: Periodic Burgers samples.

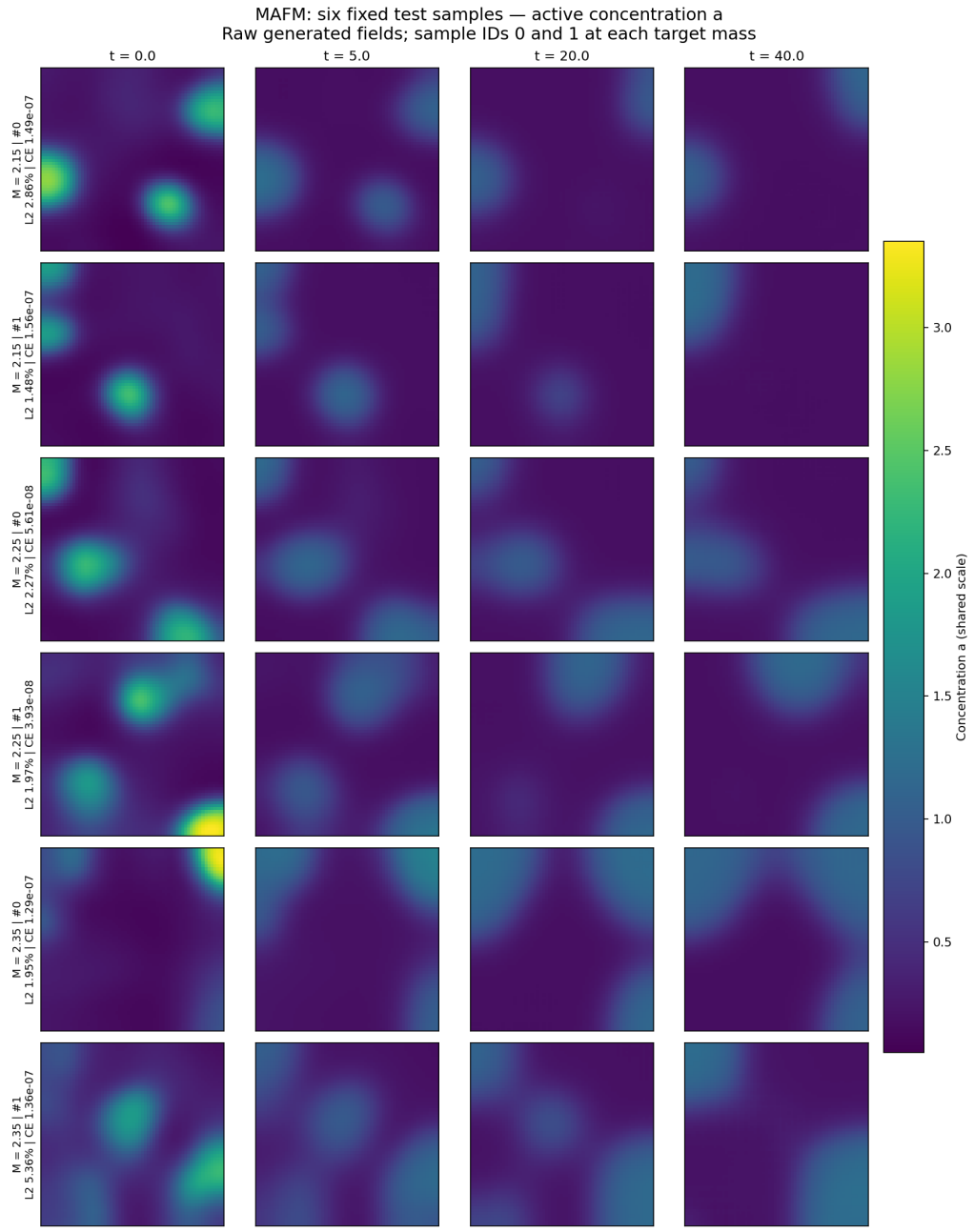

Figure 10: Reaction–diffusion samples.

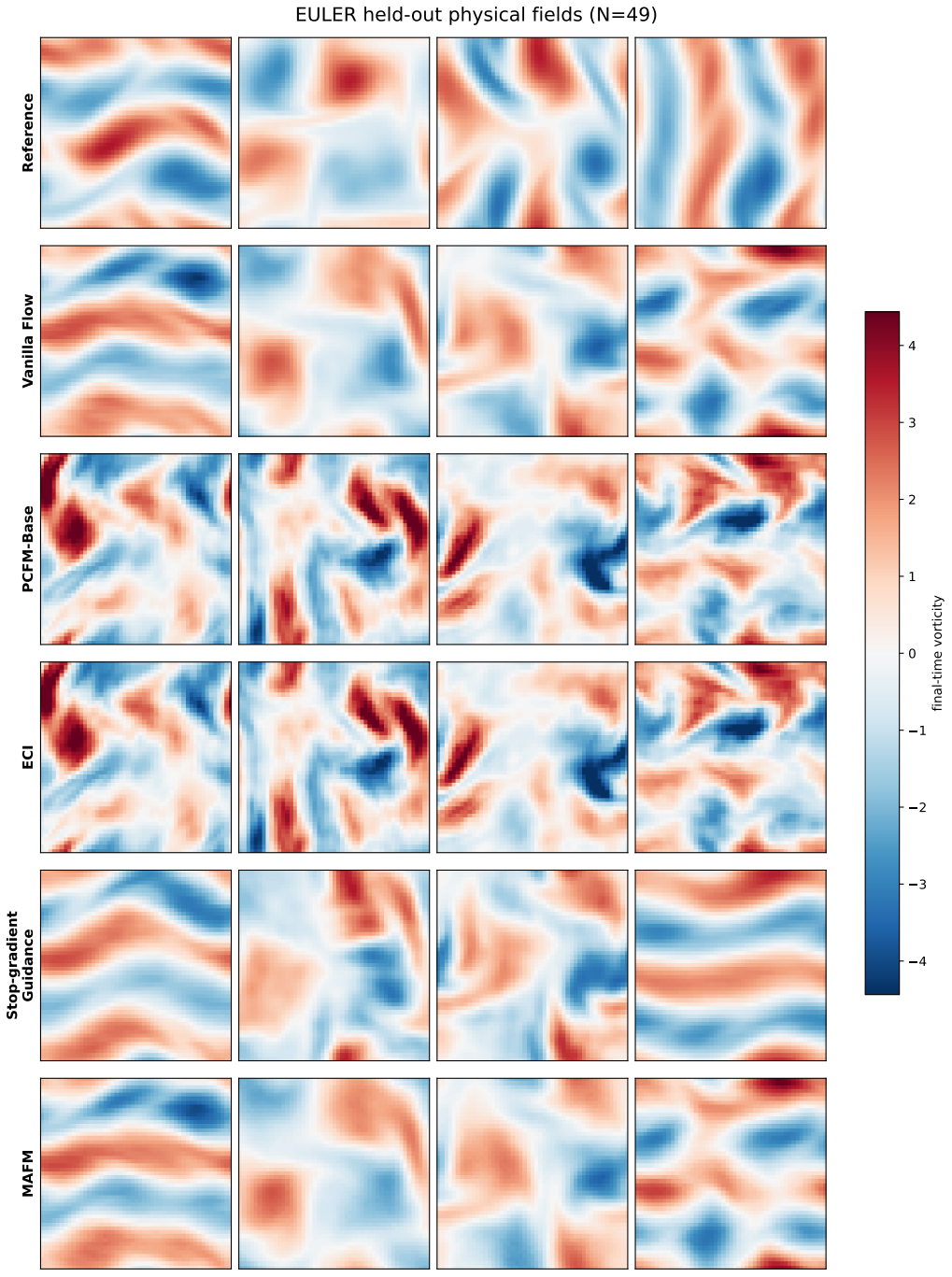

Figure 11: Euler vorticity samples at fixed energy.

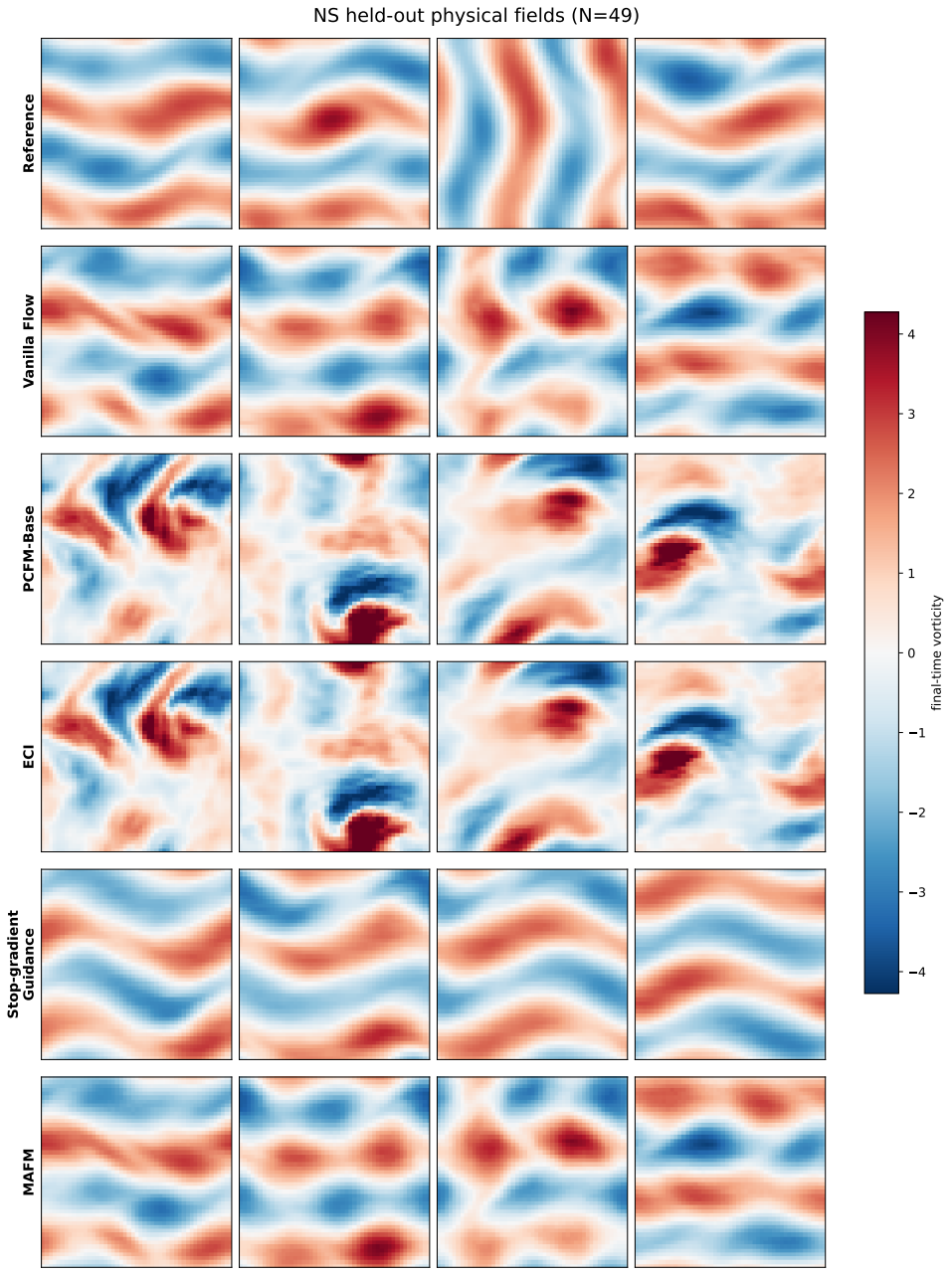

Figure 12: Controlled-NS vorticity samples at fixed energy.

E SCOPE AND PRACTICAL LIMITATIONS

The residual-contraction law describes ideal continuous dynamics with a full-rank constraint Jacobian and exact multiplier solves. In a numerical sampler, Gram regularization, solve error, and time discretization can alter that law. A finite attraction budget $\int_0^1 \kappa(t)\mathrm{d}t$ also gives a finite reduction unless the sampler starts feasible or applies an exact normal step. Thus a small measured constraint error supports numerical feasibility, while matching a separately prescribed law on the feasible intersection remains a distinct question (Eq. (20)).

The quality of the guidance depends on the frozen endpoint model. Its conditional-mean prediction may average modes or become less reliable away from the training distribution; the $\eta > 0$ adjoint is useful when the predicted-endpoint residual and input sensitivity convey a stable correction (Appendix A.4). Each active adjoint step adds one input VJP, so its benefit must be read together with the sampling-time comparison in Table 10. The model parameters remain frozen.

The empirical evidence is limited to one training seed per task. The Euler and controlled-NS tests use fixed energy and do not establish behavior across unseen energy levels. The Burgers samples show mild fluctuation-amplitude contraction, and reaction–diffusion samples retain a small negative-value tail. These observations define the scope of the present results and motivate broader seed and condition studies.