MATCH THE RADIUS, LEARN THE ANGLES: RADIAL-ANGULAR FLOW MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

Flow Matching commonly relies on Gaussian sources and Euclidean conditional paths. For data with non-Gaussian amplitude statistics, this introduces a source-induced radial mismatch: the learned transport must reshape the distribution of norms while simultaneously modeling the directional structure of the data. We propose *Radial-Angular Flow Matching (RAFM)*, which matches the target radial law at the source and preserves it along spherical conditional paths, separating radial estimation from directional transport.

We further show that, although this construction keeps the radius constant along the path, conventional velocity regression still inherits the scale of the data. This motivates a scale-free angular formulation whose targets are uniformly bounded independently of the radial distribution, while still recovering the corresponding Flow Matching velocity field after radial rescaling. Our analysis characterizes the Gaussian radial mismatch removed by source matching and shows that, unlike velocity targets, angular targets require no finite radial moments. Across controlled heavy-tailed benchmarks and real vector, image, and audio data, ablations isolate the contributions of source matching, path geometry, and angular regression. The results show that these components provide complementary benefits while retaining the simulation-free training framework of Flow Matching.

1 INTRODUCTION

An isotropic Gaussian source distribution is a default design choice in many modern generative models. It underlies diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021), flow-based generative models (Rezende & Mohamed, 2015; Dinh et al., 2017; Chen et al., 2018), and more recently Flow Matching (FM) (Lipman et al., 2023). This choice is attractive for analytical and computational reasons, but it can impose a structural bias when the target distribution exhibits non-Gaussian radial behavior.

Heavy-tailed and non-Gaussian statistics arise in turbulent flows (Sosa-Correa et al., 2019), in financial returns with extreme fluctuations (Cont, 2001), and in meteorological extremes better described by heavy-tailed laws (Papalexiou et al., 2013). In high dimensions, norm statistics strongly shape the geometry of probability mass. In particular, for an isotropic Gaussian in ambient dimension d, the norm is typically of order $\sqrt{d}$, with $O(1)$ fluctuations around that scale (Vershynin, 2018). Consequently, when the source imposes Gaussian norm statistics while the target does not, the learned transport must correct a source-induced radial discrepancy while also learning the directional structure of the data. This initial source-level radial discrepancy is avoidable: it disappears when the source matches the target norm law.

This issue is especially relevant in Flow Matching (Lipman et al., 2023), where the generative design is specified directly through a source distribution and conditional probability paths, and the associated vector field is learned by regression. In this setting, source mismatch is not merely inherited from a forward noising process: it enters directly through the pair consisting of the source and the conditional transport geometry. This makes FM a particularly natural framework in which to study whether part of the modeling burden can already be removed at initialization, before learning the remaining transport. Recent work on multiplicative diffusion models (Gruhlke et al., 2026) similarly questions the suitability of Gaussian latent structure for heavy-tailed or anisotropic data. Our perspective is complementary. Rather than modifying the stochastic noising dynamics, we ask whether the same

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.

issue can be addressed directly within the standard simulation-free Flow Matching template through a coupled design of the source and the conditional path.

We propose *Radial-Angular Flow Matching (RAFM)*, which separates radial and angular roles in the transport. RAFM draws the source radius from the target radial law, which removes the radial part of the Gaussian mismatch at initialization: the data-to-Gaussian KL divergence decomposes exactly into this radial mismatch term and a remaining directional mismatch conditional on radius (Eq. (4)). Once radii are matched, the residual transport problem is primarily angular, so RAFM moves each source point along a spherical geodesic of the matched radius, which preserves the norm along the conditional path. Standard velocity regression on these paths gives RAFM-Vel. Because its targets still scale with the radius, RAFM-Ang regresses the velocity divided by the radius: its targets are bounded independently of the radius and of the radial moments, and the population velocity field is recovered after restoring the radial factor. Both variants keep the simulation-free Flow Matching pipeline and learn directional motion conditional on radius, so radius and direction may remain dependent; we also give guarantees for the empirical estimation of the radial law. Experiments on heavy-tailed synthetic data, particle image velocimetry (PIV) flow measurements, finance, weather, images and audio separate the effects of source choice, path geometry and regression objective, at substantially lower measured training cost than the evaluated implementations of Multiplicative Diffusion Models (MSGM) (Gruhlke et al., 2026).

2 Related Work

Flow Matching (Lipman et al., 2023) introduced a simulation-free framework for training continuous normalizing flows by regressing vector fields associated with prescribed conditional probability paths. Closely related contemporaneous formulations include Rectified Flow (Liu et al., 2023) and stochastic interpolants (Albergo & Vanden-Eijnden, 2023). Subsequent work such as simulation-free Schrödinger bridges (Tong et al., 2024), Functional Flow Matching (Kerrigan et al., 2024), and Optimal Flow Matching (Kornilov et al., 2024) has broadened the range of probability paths, couplings, and spaces considered within this framework. RAFM builds on this path-design perspective and extends it to the source: when the source is mismatched, part of the transport burden is source-induced. In RAFM, correcting the radial source mismatch leads to a matched-radius conditional transport problem whose natural geometry is angular and radius-preserving.

Manifold-aware generative models, including flows on tori and spheres (Rezende et al., 2020), Moser Flow (Rozen et al., 2021), flows and probability paths on manifolds (Ben-Hamu et al., 2022), Riemannian score-based models (De Bortoli et al., 2022), Flow Matching on general geometries (Chen & Lipman, 2024) and Metric Flow Matching (Kapuśniak et al., 2024), show that non-Euclidean interpolants can yield more meaningful transport than Euclidean paths. RAFM uses spherical geometry only conditionally: the data remain in Euclidean space, and each matched-radius pair is transported on the sphere determined by its radius.

Normalizing flows transform a simple base distribution (Rezende & Mohamed, 2015). This base can itself cause mismatch. Lipschitz triangular flows cannot map light-tailed bases to heavy-tailed targets (Jaini et al., 2020). Resampled (Stimper et al., 2022) and tail-adaptive bases (Laszkiewicz et al., 2022; Hickling & Prangle, 2025) address this support and tail mismatch. RAFM also adapts the source, but only its radial component: it corrects this component exactly at the population level (Eq. (4)) and leaves the remaining discrepancy to angular transport. Source-Guided Flow Matching (Wang et al., 2026) modifies the source distribution of a pre-trained flow for training-free guidance while keeping its vector field fixed, whereas RAFM designs the source jointly with the generative training path to reduce unconditional radial source mismatch.

Heavy-Tailed Diffusion Models (Pandey et al., 2025) and Multiplicative Diffusion Models (MSGM) (Gruhlke et al., 2026) directly question the suitability of Gaussian latent structure for heavy-tailed data. Pandey et al. (2025) also introduce t-Flow, which replaces the Gaussian source by a multivariate Student-t source and uses a linear Flow Matching path. This provides a parametric heavy-tailed source, whereas RAFM matches the data radial law directly and couples this source choice to radius-preserving spherical transport. Mirror Flow Matching (Guan et al., 2026) combines a Student-t prior with a regularized mirror map for generative modeling on convex domains; its main objective is constrained-domain generation rather than radial source matching in unconstrained Euclidean Flow Matching. MSGM also preserves radial structure, but does so through stochastic

multiplicative diffusion and score learning: its dynamics preserve sample norms while progressively randomizing directions, yielding a data-aware non-Gaussian latent distribution. RAFM builds the radial law directly into Flow Matching, through its source and deterministic matched-radius spherical paths, and can additionally regress scale-free angular targets.

3 METHOD

Standard Gaussian Conditional Flow Matching (CFM) must simultaneously correct the radial mismatch induced by the Gaussian source and learn the directional structure of the target distribution. For data with non-Gaussian norm statistics, this entangles two distinct tasks into a single transport.

RAFM separates these roles by modifying two components of CFM: the source distribution and the conditional path. Figure 1 illustrates this separation: Gaussian FM must progressively correct both radius and direction, whereas RAFM starts from the matched radial law and mainly transports mass angularly.

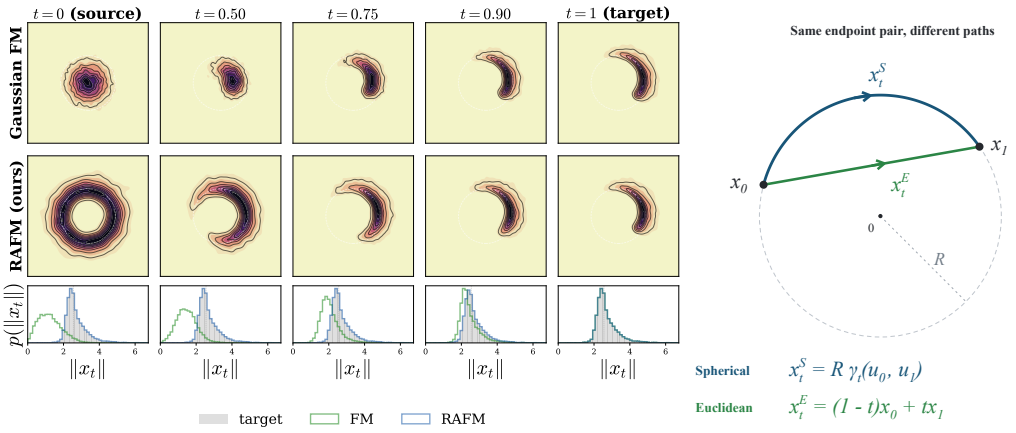


Figure 1: Source and path geometry in RAFM. Left: conditional-path marginals and radial diagnostics for a two-dimensional illustrative distribution. Gaussian FM must alter both radius and direction, whereas RAFM starts from the target radial law and reorganizes mass angularly. Right: between the same equal-radius endpoints, the spherical path (S) follows the sphere and preserves the radius, while the Euclidean path (E) crosses its interior. Here $u_i = x_i/R$ for $i = 0, 1$, and γ_t is the unit-sphere geodesic.

3.1 CONDITIONAL FLOW MATCHING SETUP

Let $(X_0, X_1) \sim \eta$ couple a source with p_{data} and let $X_t = \psi_t(X_0, X_1)$ be differentiable in time. Conditional Flow Matching (CFM) regresses its velocity (Lipman et al., 2023):

$$\mathcal{L}_{\text{CFM}}(v) = \mathbb{E}\Big[\|v(t, X_t) - \dot{X}_t\|^2\Big], \tag{1}$$

where $t \sim \text{Unif}[0, 1]$ is independent of the endpoints and $\dot{X}_t := \partial_t X_t$. When $\mathbb{E}\|\dot{X}_t\|^2 < \infty$, the unrestricted population minimizer is the conditional-mean field

$$v^\star(t, x) = \mathbb{E}[\dot{X}_t \mid X_t = x]. \tag{2}$$

Within this formulation, the two design choices relevant to RAFM are the source distribution and the conditional path. RAFM modifies these two objects in turn: first the source, to remove radial mismatch, then the path, to preserve that correction throughout transport.

3.2 MATCHING THE RADIAL LAW AT THE SOURCE

Let p_{data} be a distribution on $\mathbb{R}^d$, $d \geq 2$, with $p_{\text{data}}(\{0\}) = 0$. For $X_1 \sim p_{\text{data}}$, write $R = \|X_1\|$ and $U_1 = X_1/R \in S^{d-1}$, where $S^{d-1} := \{u \in \mathbb{R}^d : \|u\| = 1\}$, using the Euclidean norm and inner

product. We identify laws with densities when these exist. For each training endpoint $X_1 = RU_1$, draw $U_0 \sim \mathrm{Unif}(S^{d-1})$ independently of X_1 and set

$$X_0 := RU_0. \tag{3}$$

If μ_R is the law of R, the ideal source q_{rad} has radial law μ_R and uniform direction. Each coupled pair has the same radius, while the target direction can depend arbitrarily on that radius. At generation time, we sample radii from a one-dimensional empirical quantile-interpolation law $\widehat{\mu}_R$ fitted to training norms. This replaces the task of learning radial transport with radial estimation; angular transport remains conditional on the chosen radius.

Having matched the radius at the source, we can ask what part of the Gaussian source discrepancy has actually been removed. Let ϕ_d be the standard Gaussian density, p_{χ_d} its radial density, and p_R the target radial density. If p_{data} has a Lebesgue density and the KL divergences are finite, then

$$\mathrm{KL}(p_{\mathrm{data}} \| \phi_d) = \mathrm{KL}(p_{\mathrm{data}} \| q_{\mathrm{rad}}) + \mathrm{KL}(p_R \| p_{\chi_d}). \tag{4}$$

Thus radial matching removes exactly the radial component of the Gaussian source discrepancy; the remaining ideal-source discrepancy is directional, conditional on the matched radius (Theorem A.2). This identity concerns the population source; empirical radial estimation is treated separately in Appendix A.2.

3.3 Preserving the radius along the conditional path

Matching the endpoint radii is not sufficient by itself. A standard Euclidean interpolation between two equal-radius points generally leaves their sphere, reintroducing radial motion along the conditional path. Writing $\Theta := \arccos(U_0^\top U_1) \in [0, \pi]$ gives

$$\|(1-t)X_0 + tX_1\|^2 = R^2[1 - 2t(1-t)(1-\cos\Theta)]\,. \tag{5}$$

For different directions, the path contracts before expanding to the target; for orthogonal endpoints, its midpoint radius is $R/\sqrt{2}$. Therefore source matching alone does not preserve the radial correction during transport: the conditional path introduces radial contraction and re-expansion that the velocity regression must represent. This suggests choosing a conditional path that remains on the matched-radius sphere; RAFM follows a shortest spherical geodesic (Ben-Hamu et al., 2022; Chen & Lipman, 2024):

$$\gamma_t(U_0, U_1) = \frac{\sin((1-t)\Theta)}{\sin\Theta} U_0 + \frac{\sin(t\Theta)}{\sin\Theta} U_1, \qquad X_t := R\gamma_t(U_0, U_1). \tag{6}$$

The formula extends continuously at $\Theta = 0$. Antipodal endpoints have probability zero for the independent uniform U_0; a deterministic choice of shortest geodesic there leaves the population construction unchanged (the spherical path is the shortest radius-preserving curve between its endpoints, traversed at constant speed; Appendix A.3). For a non-antipodal pair, at every $t \in [0, 1]$,

$$\|X_t\| = R, \qquad X_t^\top \dot{X}_t = 0, \qquad \|\dot{X}_t\| = R\Theta. \tag{7}$$

The first two identities show that the radial correction is preserved throughout transport. The third reveals a remaining issue: although radial motion has disappeared, the conventional CFM regression target still scales linearly with the radius.

3.4 From radius-preserving transport to scale-free regression

Applying standard CFM directly to this path yields RAFM-Vel: a network v_θ^{Vel} regresses $\dot{X}_t$ with the objective (1). However, Eq. (7) shows that its target magnitude is $R\Theta$. Thus the transport no longer changes the radius, yet the supervised target still inherits the radial scale of the data. This suggests factoring out the fixed radius from the regression target:

$$A_t := \frac{\dot{X}_t}{\|X_t\|} = \partial_t \gamma_t(U_0, U_1). \tag{8}$$

Because the radius is constant along each RAFM conditional path, this normalization removes scale without changing the underlying transport geometry.

Theorem 3.1 (Scale-free Angular Flow Matching). *Along the RAFM conditional path,*

$$X_t^\top A_t = 0, \qquad \|A_t\| = \Theta \le \pi, \qquad \mathbb{E}\|A_t\|^q \le \pi^q \quad \text{for every } q > 0. \tag{9}$$

Moreover, let p_t be the law of X_t and set $a^\star(t,x) := \mathbb{E}[A_t \mid X_t = x]$. Taking expectations under a regular conditional law at fixed t, we have, for p_t-almost every x,

$$v^\star(t,x) = \mathbb{E}[\dot{X}_t \mid X_t = x] = \|x\| a^\star(t,x). \tag{10}$$

The theorem provides the key consequence: RAFM can regress a uniformly bounded target while recovering the same population velocity field after restoring the factor $\|x\|$. The corresponding regularity and finite squared-loss conditions are given in Appendix A.4.

RAFM-Ang is the corresponding regression objective: we train $a_\theta : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ using

$$\mathcal{L}_{\text{RAFM-Ang}}(\theta) = \mathbb{E}\big[\|a_\theta(t, X_t) - A_t\|^2\big]. \tag{11}$$

For RAFM-Vel, the normalized representation $a_\theta^{\text{Vel}}(t,x) := v_\theta^{\text{Vel}}(t,x)/\|x\|$ for $x \neq 0$ makes the weighting explicit:

$$\mathcal{L}_{\text{RAFM-Vel}}(\theta) = \mathbb{E}\Big[\|v_\theta^{\text{Vel}}(t, X_t) - \dot{X}_t\|^2\Big] = \mathbb{E}\big[R^2 \|a_\theta^{\text{Vel}}(t, X_t) - A_t\|^2\big]. \tag{12}$$

The two objectives therefore target the same conditional direction field, since R is fixed at a given x, but velocity regression weights angular errors proportionally to R^2, whereas RAFM-Ang removes this radial weighting while retaining radius in the network input. An R^2-weighted angular control is used experimentally to isolate this weighting from the output parameterization.

Moment robustness. The boundedness above changes more than the numerical scale of the target: it changes which radial moments are required by squared-loss regression. For every $q \ge 1$, Gaussian linear FM, matched-source Euclidean FM and spherical velocity regression have finite qth target moments exactly when $\mathbb{E}[R^q] < \infty$; angular targets are bounded independently of those moments (Theorem A.8). For Student-t targets with $\nu = 3$, $\mathbb{E}[R^2] < \infty$ but $\mathbb{E}[R^4] = \infty$: the velocity-regression loss has a finite population mean, yet the squared target norm $\|\dot{X}_t\|^2 = R^2\Theta^2$ has infinite variance, so per-sample squared errors can have infinite variance and minibatch estimates can be dominated by rare large radii. RAFM-Ang instead regresses targets with $\|A_t\| \le \pi$, whose moments are all finite whatever the radial tail. Source matching and spherical geometry thus remove radial transport, but velocity regression still inherits radial moment requirements through its target magnitude $R\Theta$; angular normalization removes this dependence from the supervised target. This is a property of the regression target, not a guarantee of better optimization or generation, whose finite-budget effect is task-dependent (Section 4). Clipping and fixed-predictor concentration are treated in Appendix A.4.

3.5 Training and Sampling

The resulting procedure retains the standard simulation-free FM workflow: training samples endpoints, independent uniform source directions and times, then evaluates X_t and the target analytically ($\dot{X}_t$ for Vel, A_t for Ang), while generation draws $R \sim \widehat{\mu}_R$ and an independent uniform direction, initializes $x_0 = RU_0$, and integrates the corresponding tangent field. For $x \neq 0$, let $T_x = \{w : x^\top w = 0\}$. The projected sampling fields are:

$$\begin{aligned} \widetilde{v}_\theta^{\text{Vel}}(t,x) &= \Pi_{T_x} v_\theta^{\text{Vel}}(t,x), \quad v_\theta^{\text{Ang}}(t,x) = \|x\| \, \Pi_{T_x} a_\theta(t,x), \\ \Pi_{T_x} w &= w - \frac{x^\top w}{\|x\|^2} x. \end{aligned} \tag{13}$$

Both fields satisfy $x^\top v = 0$, so every exact nonzero trajectory obeys $\frac{d}{dt}\|x_t\|^2 = 0$ and keeps its initialized radius. Each flow therefore preserves its initialized radial law: the target law for the ideal source, or its estimated approximation for the empirical source. Algorithms 1 and 2 (Appendix A.6) give the Ang procedures; Vel replaces a_θ and A_t by v_θ^{Vel} and $\dot{X}_t$ in training, and uses $\widetilde{v}_\theta^{\text{Vel}}$ in sampling. Numerical safeguards are described in Appendix A.6; generation-stability bounds are given in Appendix A.5.

Conditional generation. Given a condition c, the same construction samples $R \sim \widehat{\mu}_R(\cdot \mid c)$ and an independent uniform direction, and uses $v_\theta^{\mathrm{Vel}}(t, x, c)$ or $a_\theta(t, x, c)$; the population identity becomes $v^\star(t, x, c) = \|x\|\mathbb{E}[A_t \mid X_t = x, c]$ under the corresponding conditional law. No independence between target radius and direction given c is assumed.

4 Experiments

We evaluate RAFM along the three choices developed in Section 3: whether matching the radial source removes a meaningful part of the transport burden, whether radius-preserving spherical paths improve upon source matching alone, and whether scale-free angular regression provides an additional benefit over conventional velocity regression. We also test whether these effects carry over to image and audio representations. We then use targeted controls to examine when radial and directional structure can be modeled separately.

4.1 Experimental setup

Datasets. Synthetic correlated Student-t vectors with $\nu = 3$ ($d = 16, 32$) give controlled heavy tails in the moment regime of Section 3. Particle image velocimetry (PIV) vorticity fields (Georgeault & Heitz, 2026) ($d = 64, 256$), Finance daily industry-portfolio returns ($d = 49$) and Weather ERA5 wind fields over Australia ($d = 96$) are structured real vectors. ImageNette images, encoded into latents by a frozen Deep Compression Autoencoder (DC-AE) (Chen et al., 2025), and FSD50K natural multi-label sound events (Fonseca et al., 2022) test higher-dimensional image and audio representations, with FSD50K clips generated conditionally on their sound-event tags. AudioMNIST spoken digits (Becker et al., 2024) provide a controlled setting: we represent each clip by its complex short-time Fourier transform (STFT) (Allen & Rabiner, 1977), normalize it to unit norm and multiply it by an independent random amplitude factor (gain) drawn from a lognormal mixture; by construction, the radius equals the gain (R), while the unit-norm STFT keeps its normalized time-frequency pattern (U).

Protocol. The main methods form an ablation ladder in which each step changes one design choice: Gaussian FM uses a standard Gaussian source, the Euclidean path and velocity regression; Matched-source FM changes the source, using the RAFM radial source but retaining the Euclidean path and velocity objective; RAFM-Vel changes the path to the radius-preserving spherical path; and RAFM-Ang changes the regression target to radius-normalized angular velocity. Fixed-spherical FM, which rescales source and target directions to a single radius, the training mean, and therefore removes radial variability from the transport, and RAFM-Ang-R^2, which keeps the RAFM-Ang predictor and sampling field and only weights each angular loss term by $R^2/m_{2,\mathrm{train}}$, where $m_{2,\mathrm{train}}$ is the mean of the squared training norms, reproducing the weighting implicit in velocity regression (Eq. (12); Section 4.4), are mechanism controls, not proposed methods. We abbreviate RAFM-Vel and RAFM-Ang as Vel and Ang when unambiguous. t-Flow (Pandey et al., 2025), Flow Matching from a Student-t source along a linear path, and MSGM (Gruhlke et al., 2026), a multiplicative-diffusion score model that preserves sample norms, are external non-Gaussian baselines, run as fixed-budget adaptations on the common backbones; we report one MSGM implementation, and Appendix C.1 compares both implementations and their training costs. Within each comparison, methods share data, splits, backbone, effective batch, and update and evaluation budgets; results summarize three training seeds. Radial 1-Wasserstein (W_1) distance (Peyré & Cuturi, 2019) and Kolmogorov–Smirnov (KS) statistic (Massey, 1951) measure how well the generated norm distribution matches the reference; Sliced W_1 (Bonneel et al., 2015) measures the joint fit of vector data; Fréchet Inception Distance (FID) (Heusel et al., 2017) and Fréchet Audio Distance (FAD) (Kilgour et al., 2019) measure image and audio feature-distribution quality, while precision (Kynkäänniemi et al., 2019) and coverage (Naeem et al., 2020) measure the fidelity and diversity of generated images in the same feature space; and waveform-energy KS compares generated and reference distributions of waveform energy (squared ℓ_2 norm of the signal). Full evaluation details are given in Appendix B.

4.2 Source, path, and regression on vector benchmarks

What source matching removes. On Student-t and PIV, Matched-source FM removes most of the gap between Gaussian FM and the best model, consistent with radial mismatch being a substantial

Dataset	Sliced $W_1 \downarrow$					
	Gaussian FM	Matched-source FM	RAFM (ours) Vel	RAFM (ours) Ang	t-Flow	MSGM
Student-t $d = 16$	$0.760_{\pm 0.232}$	$0.438_{\pm 0.064}$	$0.332_{\pm 0.030}$	$\mathbf{0.250}_{\pm 0.021}$	$4.810_{\pm 1.311}$	$0.385_{\pm 0.014}$
Student-t $d = 32$	$1.519_{\pm 0.026}$	$0.751_{\pm 0.204}$	$0.475_{\pm 0.060}$	$\mathbf{0.347}_{\pm 0.033}$	$7.955_{\pm 1.641}$	$0.722_{\pm 0.028}$
PIV ($d = 64$)	$0.0463_{\pm 0.0030}$	$0.0324_{\pm 0.0046}$	$\mathbf{0.0273}_{\pm 0.0029}$	$0.0279_{\pm 0.0025}$	$9.9476_{\pm 0.3396}$	$0.0629_{\pm 0.0008}$
PIV ($d = 256$)	$0.4498_{\pm 0.0024}$	$0.0382_{\pm 0.0019}$	$0.0242_{\pm 0.0030}$	$\mathbf{0.0227}_{\pm 0.0003}$	$72.6029_{\pm 0.0992}$	$0.0354_{\pm 0.0009}$
Finance ($d = 49$)	$0.194_{\pm 0.018}$	$0.208_{\pm 0.004}$	$0.174_{\pm 0.003}$	$\mathbf{0.172}_{\pm 0.004}$	$5.952_{\pm 0.306}$	$0.182_{\pm 0.002}$
Weather ($d = 96$)	$0.085_{\pm 0.007}$	$0.094_{\pm 0.007}$	$\mathbf{0.075}_{\pm 0.004}$	$0.077_{\pm 0.001}$	$196.345_{\pm 1.854}$	$0.126_{\pm 0.004}$

Table 1: Vector benchmarks (Student-t: $\nu = 3$), mean $\pm$ SD over three training seeds. Gray groups RAFM variants; bold marks the lowest mean. Full metrics and protocols: Appendices C.1, C.2, and B.

part of the transport burden. The exceptions are the chronological Finance and Weather splits, where matching the training radial law alone raises Sliced W_1 over Gaussian FM. We return to this degradation in the following path comparison; the chronological shift instead limits radial fidelity on Finance: the training-to-test radial W_1 is 1.386, and Gaussian FM, RAFM and MSGM remain close to it (Appendix C.2). Once the path changes, both RAFM variants still improve Finance and Weather joint fit over Gaussian FM and lower mean angular Sliced W_1 (Sliced W_1 between unit directions within radius quartiles) versus both Gaussian FM and Matched-source FM (Tables 6, 7). Source matching therefore removes a large part of the problem when the radial mismatch is large, but on real vectors its benefit depends on the path, whereas the spherical path improves the directional part of the fit even when radial fidelity is limited by distribution shift.

What remains after source matching. Matched-source FM and RAFM-Vel form the cleanest source/path ablation: they share the matched radial source, data, backbone and training budget, and differ only in the conditional path. RAFM-Vel has the lower mean Sliced W_1 in all six vector settings, and matched-source generation does not retain the matched radial law (Appendices C.1 and C.2). Source matching alone is therefore not sufficient: Euclidean interpolation between equal-radius endpoints leaves the sphere, contracting the radius to $R/\sqrt{2}$ at the midpoint for orthogonal endpoints (Eq. (5)), and the consistent improvement from Matched-source FM to RAFM-Vel is consistent with the benefit of preserving the matched radius throughout the conditional path. Vel versus Ang is more regime-dependent: Ang's clearest advantage is on the heavy-tailed Student-t targets, consistent with Section 3, whereas elsewhere the mean differences are smaller and change sign. Once the source is fixed, path geometry thus gives the most consistent improvement, and angular normalization an additional regime-dependent one; the Ang-R^2 control isolates the role of the implicit R^2 weighting (Section 4.4).

Non-Gaussian baselines. Under the evaluated fixed budgets, both RAFM variants have lower mean Sliced W_1 than MSGM in all six settings, although MSGM can have smaller radial errors, and RAFM-Ang has lower mean Sliced W_1 than a median-scaled Gaussian source in all six cases (Appendices C.1 and C.5). In our common-backbone adaptation, t-Flow has higher mean Sliced W_1 than all Flow Matching variants (Appendix B.5).

4.3 Beyond vector benchmarks: scaling to image and audio representations

We next run the same comparisons on image and audio representations.

Radial fidelity does not determine generative quality. On ImageNette (Table 2), Vel and Ang reach the same radial W_1, because both draw radii from the same source and preserve them along tangential dynamics, yet FID falls from 174.830 to 149.959: the FID improvement cannot be explained by a better match of the radial distribution, and neither Matched-source FM nor Vel improves FID over Gaussian FM. Radius is nevertheless not a nuisance variable: latent radius is associated with contrast and extreme-pixel prevalence, whereas the normalized direction carries most linearly decodable class information (0.616 from U versus 0.165 from R; Appendix C.4).

Method		FID ↓	Radial W_1 ↓	Precision ↑	Coverage ↑
Gaussian FM		161.917±0.933	4.747±0.170	0.102±0.007	0.055±0.003
Matched-source FM		171.894±1.238	2.616±0.150	0.101±0.004	0.047±0.004
Fixed-spherical FM		166.105±0.205	3.770±0.000	0.121±0.007	0.058±0.002
RAFM (ours)	Vel	174.830±0.918	**0.266**±0.000	0.110±0.008	0.049±0.004
	Ang	**149.959**±0.414	**0.266**±0.000	**0.145**±0.003	**0.079**±0.004
t-Flow		224.669±6.143	405.595±66.370	0.011±0.001	0.004±0.000
MSGM		280.835±2.214	5.110±0.001	0.023±0.005	0.002±0.001

Table 2: ImageNette: 3,000 generations, 2,680 common test-reference images, 100 network calls/sample; mean ± population SD, three seeds. Gray groups RAFM; bold marks best means. Full metrics/protocol: Table 8, Appendix B.4.

Method	FAD ↓	Waveform energy KS ↓
Gaussian FM	86.59±0.83	0.3424±0.0379
Matched-source FM	89.28±2.43	**0.1107**±0.0045
Fixed-spherical FM	78.98±0.79	0.6185±0.0023
Fixed-spherical FM + conditional gain	77.07±0.80	0.1159±0.0023
RAFM-Vel	90.28±1.01	0.1133±0.0000
RAFM-Ang	**75.81**±1.16	0.1146±0.0023
RAFM-Ang-R^2	93.32±1.55	0.1133±0.0000
t-Flow	97.88±4.31	0.7852±0.0170
MSGM	148.19±0.0000	0.1758±0.0000

Table 3: FSD50K: 10,000 updates, 160 network calls/sample, 256 generations and 256 common validation references per seed. Mean ± sample SD over three training seeds; FAD uses PANN CNN14 at 16 kHz. Bold marks the best mean. The gain control is a post-hoc gain rescaling of the fixed-spherical models.

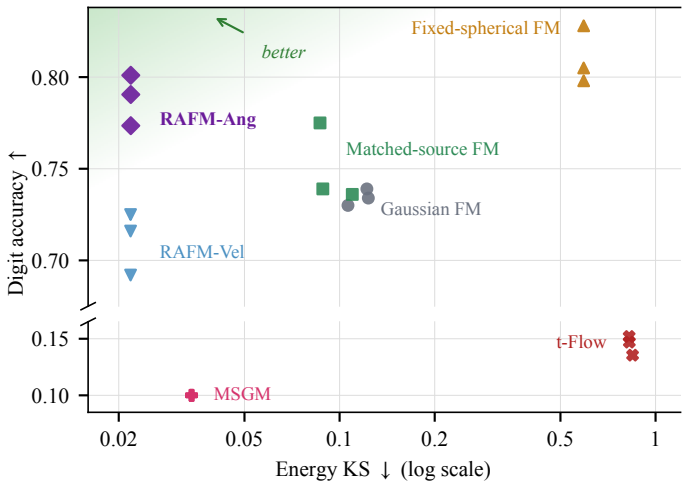


Figure 2: AudioMNIST digit accuracy versus energy KS (log scale); each marker is one training seed, three per method. Better results lie toward the upper left (high accuracy, low energy KS). The accuracy axis is broken to include t-Flow and MSGM. Mean ± SD values are in Table 9 (Appendix C.3).

Audio. On FSD50K (Table 3), matching the source improves the energy-distribution match (KS 0.3424 to 0.1107) without improving FAD (86.59 to 89.28), and Vel, Ang and Ang-R^2 have similar energy statistics but substantially different FAD, with the lowest mean for Ang: accurately matching the norm or energy distribution does not by itself determine generative quality. Post-hoc gain rescaling of fixed-spherical outputs, with gains drawn from the label-conditional radial law, is also competitive (77.07 versus 75.81), so FSD50K does not show that radius and content must be modeled jointly. On AudioMNIST, where the gain is independent of content by construction, modeling the gain

separately suffices: post-hoc independent gain rescaling of the fixed-spherical model recovers the energy distribution (KS 0.5923 to 0.0218, as for RAFM; Table 9) without changing any gain-invariant digit prediction, and therefore reaches slightly higher digit accuracy than RAFM-Ang (0.810 versus 0.7883); among trained generators, RAFM-Ang combines this energy fidelity with high digit accuracy (Figure 2).

4.4 Mechanism controls, cost and limitations

With the same angular parameterization and sampling field, restoring the implicit R^2 weighting (Ang-R^2) worsens the Student-t, ImageNette and FSD50K results, whereas on SEVIR weather frames (Veillette et al., 2020) Vel and Ang-R^2 improve over unweighted Ang, all three remaining below Gaussian FM and MSGM, so the preferred weighting can be task-dependent (Appendices C.8 and C.9). Separate radius resampling suffices when independence is imposed, as in AudioMNIST, but in a controlled synthetic example where direction depends on radius, independent output radii reach Sliced W_1 about 0.321 versus 0.027 for Ang, although all methods have the same radial error (radial W_1 0.01291) (Appendix C.6); RAFM does not impose $R \perp U$ and learns directional motion conditional on radius. Whether radius can be resampled separately is therefore a property of the data rather than of the method. Finally, RAFM retains simulation-free Flow Matching training and has substantially lower measured training cost than the evaluated MSGM implementations under matched update budgets, about $9.3\times$ on FSD50K and 52–$89\times$ on the Student-t/PIV comparisons (Appendix C.1).

Limitations. The empirical radial source cannot extrapolate beyond the radii represented in the training data and inherits train–test changes in the radial law, as Finance illustrates (Appendix C.2); more adaptive radial estimators are a natural extension. The scale-free objective removes radial-moment dependence from the target, but the best finite-budget weighting is task-dependent: Vel or R^2-weighted Ang can be preferable, as on SEVIR.

5 Conclusion

We introduced Radial-Angular Flow Matching, which replaces the default Gaussian source and Euclidean path with a matched radial source and radius-preserving spherical transport. This decomposition reveals that, even when the path keeps the radius constant, conventional velocity targets still scale with the sample radius. RAFM-Ang removes this scale dependence, yielding uniformly bounded targets independently of radial moments while recovering the corresponding Flow Matching velocity field after radial rescaling.

Experiments separate the effects of these design choices. Source matching removes a substantial part of the transport burden on heavy-tailed and PIV benchmarks, while spherical transport improves over matched-source Euclidean FM across all vector settings. On image and audio representations, matching the radial distribution alone does not determine generative quality, while the benefit of angular regression depends on how radius interacts with the remaining data structure. RAFM also preserves simulation-free Flow Matching training and achieves lower Sliced Wasserstein error than MSGM across the vector benchmarks at substantially lower measured training cost. Overall, our results show that for non-Gaussian data, source design, path geometry, and regression parameterization should be treated as coupled modeling choices rather than inherited Gaussian defaults.

AI USE STATEMENT

We used generative AI tools to assist with identifying related work; improving, optimizing, and debugging experimental code; and identifying candidate datasets based on research directions and criteria specified by the authors. These tools also supported the consistency checks of mathematical arguments, notation, and reported findings. Theoretical results and their proofs were developed by the authors. We additionally used these tools to rephrase and streamline selected passages, improve figures, and assist with manuscript formatting. All AI-assisted outputs were reviewed by the authors, who remain responsible for the final content of the paper, including its theoretical claims, experimental results, and accompanying code.

REPRODUCIBILITY STATEMENT

Appendix A states the assumptions and gives complete proofs of the theoretical results. Appendix B specifies data sources and preprocessing, splits, architectures, optimization hyperparameters, sampling budgets, training seeds and hardware, and Appendix C reports complete metrics for the comparisons summarized in the main text.

REFERENCES

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations (ICLR)*. OpenReview.net, 2023. URL https://openreview.net/forum?id=li7qeBbCR1t.

Jont B. Allen and Lawrence R. Rabiner. A unified approach to short-time Fourier analysis and synthesis. *Proceedings of the IEEE*, 65(11):1558-1564, 1977. doi: 10.1109/PROC.1977.10770.

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, 2 edition, 2008.

Sören Becker, Johanna Vielhaben, Marcel Ackermann, Klaus-Robert Müller, Sebastian Lapuschkin, and Wojciech Samek. AudioMNIST: Exploring explainable artificial intelligence for audio analysis on a simple benchmark. *Journal of the Franklin Institute*, 361(1):418–428, 2024.

Heli Ben-Hamu, Samuel Cohen, Joey Bose, Brandon Amos, Aditya Grover, Maximilian Nickel, Ricky T. Q. Chen, and Yaron Lipman. Matching normalizing flows and probability paths on manifolds. In *International Conference on Machine Learning (ICML)*, 2022.

Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.

Junyu Chen, Han Cai, Junsong Chen, Enze Xie, Shang Yang, Haotian Tang, Muyang Li, Yao Lu, and Song Han. Deep compression autoencoder for efficient high-resolution diffusion models. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

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

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 31, 2018.

Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223, 2001.

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.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real NVP. In *International Conference on Learning Representations (ICLR)*, 2017.

Eduardo Fonseca, Xavier Favory, Jordi Pons, Frederic Font, and Xavier Serra. FSD50K: An open dataset of human-labeled sound events. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:829-852, 2022.

Philippe Georgeault and Dominique Heitz. Non-time-resolved PIV Dataset of Flow over a Circular Cylinder at Reynolds Number 3900. Recherche Data Gouv, 2026. URL https://doi.org/10.57745/DHJXM6.

Robert Gruhlke, Valentin Resseguier, and Merveille Talla. Multiplicative diffusion models: Beyond Gaussian latents. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026.

Yunrui Guan, Krishna Balasubramanian, and Shiqian Ma. Mirror flow matching with heavy-tailed priors for generative modeling on convex domains. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/d33a2d61ff49d6156e0abfd906b23cf1-Abstract-Conference.html.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Tennessee Hickling and Dennis Prangle. Flexible tails for normalizing flows. In *International Conference on Machine Learning (ICML)*, 2025.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Priyank Jaini, Ivan Kobyzev, Yaoliang Yu, and Marcus Brubaker. Tails of Lipschitz triangular flows. In *International Conference on Machine Learning (ICML)*, pp. 4673–4681. PMLR, 2020.

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

Gavin Kerrigan, Giosue Migliorini, and Padhraic Smyth. Functional flow matching. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2024.

Kevin Kilgour, Mauricio Zuluaga, Dominik Roblek, and Matthew Sharifi. Fréchet audio distance: A reference-free metric for evaluating music enhancement algorithms. In *Interspeech*, pp. 2350-2354, 2019. doi: 10.21437/Interspeech.2019-2219.

Nikita Kornilov, Petr Mokrov, Alexander Gasnikov, and Alexander Korotin. Optimal flow matching: Learning straight trajectories in just one step. *Advances in Neural Information Processing Systems*, 37:104180–104204, 2024.

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Mike Laszkiewicz, Johannes Lederer, and Asja Fischer. Marginal tail-adaptive normalizing flows. In *International Conference on Machine Learning (ICML)*, pp. 12020–12048. PMLR, 2022.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations (ICLR)*. OpenReview.net, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations (ICLR)*. OpenReview.net, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

Frank J. Massey, Jr. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, and Jaejun Yoo. Reliable fidelity and diversity metrics for generative models. In *International Conference on Machine Learning (ICML)*, 2020.

Kushagra Pandey, Jaideep Pathak, Yilun Xu, Stephan Mandt, Michael Pritchard, Arash Vahdat, and Morteza Mardani. Heavy-tailed diffusion models. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025. URL https://openreview.net/forum?id=tozlOEN4qp.

S. M. Papalexiou, D. Koutsoyiannis, and C. Makropoulos. How extreme is extreme? an assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, 17(2):851–862, 2013.

Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning (ICML)*, pp. 1530–1538. PMLR, 2015.

Danilo Jimenez Rezende, George Papamakarios, Sébastien Racaniere, Michael Albergo, Gurtej Kanwar, Phiala Shanahan, and Kyle Cranmer. Normalizing flows on tori and spheres. In *International Conference on Machine Learning (ICML)*, pp. 8083–8092. PMLR, 2020.

Noam Rozen, Aditya Grover, Maximilian Nickel, and Yaron Lipman. Moser flow: Divergence-based generative modeling on manifolds. *Advances in Neural Information Processing Systems*, 34: 17669–17680, 2021.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning (ICML)*, pp. 2256–2265. PMLR, 2015.

Yang Song, Jascha Sohl-Dickstein, Durk P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://iclr.cc/virtual/2021/oral/3402.

W Sosa-Correa, RM Pereira, AMS Macêdo, EP Raposo, DSP Salazar, and GL Vasconcelos. Emergence of skewed non-Gaussian distributions of velocity increments in isotropic turbulence. *Physical Review Fluids*, 4(6):064602, 2019.

Vincent Stimper, Bernhard Schölkopf, and José Miguel Hernández-Lobato. Resampling base distributions of normalizing flows. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 4915–4936. PMLR, 2022.

Alexander Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free Schrödinger bridges via score and flow matching. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2024.

Mark Veillette, Siddharth Samsi, and Chris Mattioli. SEVIR: A storm event imagery dataset for deep learning applications in radar and satellite meteorology. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://papers.neurips.cc/paper/2020/hash/fa78a16157fed00d7a80515818432169-Abstract.html.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.

Zifan Wang, Alice Harting, Matthieu Barreau, Michael M. Zavlanos, and Karl Henrik Johansson. Source-guided flow matching. In *The Fourteenth International Conference on Learning Representations (ICLR)*. OpenReview.net, 2026. URL https://openreview.net/forum?id=p56ZAQUCUr.

A ADDITIONAL THEORY FOR RAFM

A.1 RADIAL SOURCE: CONSTRUCTION AND COMPARISON WITH A GAUSSIAN SOURCE

This appendix complements Section 3.2. We first formalize the ideal radial source q_{rad}, then compare it to the standard Gaussian source. The empirical radial extension is deferred to Appendix A.2.

Polar notation. Let $X \sim p_{\mathrm{data}}$ be a random vector in $\mathbb{R}^d$, with $d \geq 2$, and assume $p_{\mathrm{data}}(\{0\}) = 0$. We write

$$R := \|X\| \in \mathbb{R}_+, \qquad U := \frac{X}{\|X\|} \in S^{d-1}.$$

We denote by p_R the density of R on $\mathbb{R}_+$, by σ the surface measure on S^{d-1}, and by $p_{U|R}(\cdot \mid r)$ the conditional angular density with respect to σ. Under polar coordinates $x = ru$, the Lebesgue measure decomposes as

$$dx = r^{d-1}\, dr\, d\sigma(u).$$

Proposition A.1 (Density of the radial source). *Let*

$$X_0 = R\,U_0, \qquad R \sim p_R, \qquad U_0 \sim \mathrm{Unif}(S^{d-1}), \qquad R \perp U_0.$$

Then the law q_{rad} of X_0 is absolutely continuous on $\mathbb{R}^d \setminus \{0\}$, with density

$$q_{\mathrm{rad}}(x) = \frac{p_R(\|x\|)}{|S^{d-1}|\,\|x\|^{d-1}}, \qquad x \neq 0.$$

Proof. Let $f : \mathbb{R}^d \to \mathbb{R}$ be bounded measurable. Using polar coordinates and the independence of R and U_0,

$$\mathbb{E}[f(X_0)] = \int_0^\infty \int_{S^{d-1}} f(ru)\, \frac{1}{|S^{d-1}|}\, d\sigma(u)\, p_R(r)\, dr.$$

Since $dx = r^{d-1}\, dr\, d\sigma(u)$, this can be rewritten as

$$\mathbb{E}[f(X_0)] = \int_{\mathbb{R}^d \setminus \{0\}} f(x)\, \frac{p_R(\|x\|)}{|S^{d-1}|\,\|x\|^{d-1}}\, dx.$$

Hence the density of q_{rad} is exactly the claimed expression. □

Exact radial preservation. By construction, if $X_0 \sim q_{\mathrm{rad}}$, then

$$X_0 = R\,U_0, \qquad \|U_0\| = 1 \quad \text{a.s.}, \qquad R \overset{d}{=} \|X\| \text{ for } X \sim p_{\mathrm{data}}.$$

Hence

$$\|X_0\| \overset{d}{=} \|X\|, \qquad \mathbb{P}(\|X_0\| > t) = \mathbb{P}(\|X\| > t) \quad \forall t \geq 0.$$

Theorem A.2 (KL decomposition for the ideal radial source). *Assume that the relevant conditional densities exist and that the KL divergences below are finite. Then*

$$\mathrm{KL}(p_{\mathrm{data}} \| q_{\mathrm{rad}}) = \mathbb{E}_{r \sim p_R}\Big[\mathrm{KL}\big(p_{U|R}(\cdot \mid r) \,\|\, \mathrm{Unif}(S^{d-1})\big)\Big],$$

whereas

$$\mathrm{KL}(p_{\mathrm{data}} \| \phi_d) = \mathrm{KL}(p_R \| p_{\chi_d}) + \mathbb{E}_{r \sim p_R}\Big[\mathrm{KL}\big(p_{U|R}(\cdot \mid r) \,\|\, \mathrm{Unif}(S^{d-1})\big)\Big].$$

Consequently,

$$\mathrm{KL}(p_{\mathrm{data}} \| q_{\mathrm{rad}}) \leq \mathrm{KL}(p_{\mathrm{data}} \| \phi_d),$$

with equality if and only if $p_R = p_{\chi_d}$ almost everywhere.

Proof. Under polar coordinates $x = ru$, the data density may be written as

$$p_{\text{data}}(x) = \frac{p_R(r)\, p_{U|R}(u \mid r)}{r^{d-1}}, \qquad r > 0,\ u \in S^{d-1},$$

where $p_{U|R}(\cdot \mid r)$ is a density with respect to σ. By Proposition A.1,

$$q_{\text{rad}}(x) = \frac{p_R(r)}{|S^{d-1}|\, r^{d-1}}.$$

Therefore

$$\log \frac{p_{\text{data}}(x)}{q_{\text{rad}}(x)} = \log p_{U|R}(u \mid r) + \log |S^{d-1}|.$$

Integrating against $p_{\text{data}}(x)\, dx = p_R(r)\, p_{U|R}(u \mid r)\, dr\, d\sigma(u)$ gives

$$\mathrm{KL}(p_{\text{data}} \| q_{\text{rad}}) = \int_0^\infty p_R(r) \int_{S^{d-1}} p_{U|R}(u \mid r) \log\Big(p_{U|R}(u \mid r)\, |S^{d-1}|\Big)\, d\sigma(u)\, dr.$$

Since the uniform density on the sphere is $|S^{d-1}|^{-1}$ with respect to σ, this is exactly

$$\mathbb{E}_{r \sim p_R}\Big[\mathrm{KL}\big(p_{U|R}(\cdot \mid r) \,\|\, \mathrm{Unif}(S^{d-1})\big)\Big].$$

For the standard Gaussian,

$$\phi_d(x) = \frac{p_{\chi_d}(r)}{|S^{d-1}|\, r^{d-1}},$$

since the Gaussian is conditionally uniform in direction at fixed radius. Hence

$$\log \frac{p_{\text{data}}(x)}{\phi_d(x)} = \log \frac{p_R(r)}{p_{\chi_d}(r)} + \log p_{U|R}(u \mid r) + \log |S^{d-1}|.$$

Integrating again yields

$$\mathrm{KL}(p_{\text{data}} \| \phi_d) = \int_0^\infty p_R(r) \log \frac{p_R(r)}{p_{\chi_d}(r)}\, dr + \mathbb{E}_{r \sim p_R}\Big[\mathrm{KL}\big(p_{U|R}(\cdot \mid r) \,\|\, \mathrm{Unif}(S^{d-1})\big)\Big].$$

The first term is exactly $\mathrm{KL}(p_R \| p_{\chi_d})$, which proves the decomposition. The inequality and equality condition follow immediately. □

A.2 Empirical radial estimation and statistical guarantees

We now turn to the empirical extension of the radial source. At the density level, replacing the ideal radial law p_R by a positive approximation $\widehat{p}_R$ preserves the same radial–angular decomposition up to a radial KL term. This provides the direct analogue of the ideal KL comparison.

In practice, however, RAFM is typically initialized from the empirical radial distribution itself, either through the empirical CDF or by direct resampling of the observed training radii, rather than from a strictly positive density estimator. For this reason, the practical guarantees are more naturally stated in terms of CDF and Wasserstein control, which we study next.

Theorem A.3 (KL decomposition for an empirical radial source)**.** *Let $\widehat{p}_R$ be a strictly positive density on the support of p_R, and define*

$$\widehat{q}_{\text{rad}}(x) = \frac{\widehat{p}_R(\|x\|)}{|S^{d-1}|\, \|x\|^{d-1}}, \qquad x \neq 0.$$

Assume that all KL divergences below are finite. Then

$$\mathrm{KL}(p_{\text{data}} \| \widehat{q}_{\text{rad}}) = \mathrm{KL}(p_R \| \widehat{p}_R) + \mathbb{E}_{r \sim p_R}\Big[\mathrm{KL}\big(p_{U|R}(\cdot \mid r) \,\|\, \mathrm{Unif}(S^{d-1})\big)\Big].$$

In particular,

$$\mathrm{KL}(p_{\text{data}} \| \widehat{q}_{\text{rad}}) - \mathrm{KL}(p_{\text{data}} \| \phi_d) = \mathrm{KL}(p_R \| \widehat{p}_R) - \mathrm{KL}(p_R \| p_{\chi_d}).$$

Proof. The proof is identical to that of Theorem A.2, replacing p_R by $\widehat{p}_R$ in the source density:

$$\log \frac{p_{\mathrm{data}}(x)}{\widehat{q}_{\mathrm{rad}}(x)} = \log \frac{p_R(r)}{\widehat{p}_R(r)} + \log p_{U|R}(u \mid r) + \log |S^{d-1}|.$$

Integrating against the polar factorization of p_{data} yields the result. □

We reserve $\widehat{\mu}_{R,n}$ for the discrete empirical measure and write $\widehat{\mu}_R$ for the quantile-interpolation law used by the sampler. The CDF and Wasserstein guarantees below concern $\widehat{\mu}_{R,n}$; applying them to $\widehat{\mu}_R$ additionally requires controlling the interpolation error.

A.2.1 Statistical properties of the empirical radial source

This appendix quantifies the approximation error introduced when the radial law is estimated empirically from training data.

Empirical radial law. Let $X_1, \dots, X_n \sim p_{\mathrm{data}}$ be i.i.d. training samples, and define the corresponding radii

$$R_i := \|X_i\|, \qquad i = 1, \dots, n.$$

Let μ_R denote the law of $R = \|X\|$ for $X \sim p_{\mathrm{data}}$, and let

$$F_R(r) := \mathbb{P}(R \le r), \qquad r \ge 0,$$

be its cumulative distribution function. The empirical radial measure is

$$\widehat{\mu}_{R,n} := \frac{1}{n} \sum_{i=1}^{n} \delta_{R_i},$$

with empirical CDF

$$\widehat{F}_{R,n}(r) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{R_i \le r\}}.$$

We define the empirical radial source by

$$\widehat{X}_0 := \widehat{R}\, U_0, \qquad \widehat{R} \sim \widehat{\mu}_{R,n}, \qquad U_0 \sim \mathrm{Unif}(S^{d-1}), \qquad \widehat{R} \perp U_0,$$

and denote its law by $\widehat{q}_{\mathrm{rad},n}$.

Theorem A.4 (Dvoretzky–Kiefer–Wolfowitz bound)**.** *For every $\varepsilon > 0$,*

$$\mathbb{P}\left(\sup_{r \ge 0} \left| \widehat{F}_{R,n}(r) - F_R(r) \right| > \varepsilon \right) \le 2e^{-2n\varepsilon^2}.$$

Equivalently, for every $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\sup_{r \ge 0} \left| \widehat{F}_{R,n}(r) - F_R(r) \right| \le \sqrt{\frac{1}{2n} \log \frac{2}{\delta}}.$$

Proof. This is the classical Dvoretzky–Kiefer–Wolfowitz inequality applied to the i.i.d. sample $(R_i)_{i=1}^n$. □

Proposition A.5 (Transfer from radial estimation error to source estimation error)**.** *Let $p \ge 1$. Then*

$$W_p(\widehat{q}_{\mathrm{rad},n}, q_{\mathrm{rad}}) \le W_p(\widehat{\mu}_{R,n}, \mu_R).$$

In particular, any convergence of the empirical radial law in Wasserstein distance induces the same convergence for the corresponding radial source.

Proof. Let π be any coupling between $\widehat{\mu}_{R,n}$ and μ_R, and let $U \sim \mathrm{Unif}(S^{d-1})$ be independent of $(\widehat{R}, R) \sim \pi$. Then $(\widehat{R}U, RU)$ is a coupling between $\widehat{q}_{\mathrm{rad},n}$ and q_{rad}. Hence

$$W_p^p(\widehat{q}_{\mathrm{rad},n}, q_{\mathrm{rad}}) \le \mathbb{E}\big[\|\widehat{R}U - RU\|^p\big] = \mathbb{E}\big[|\widehat{R} - R|^p \|U\|^p\big] = \mathbb{E}\big[|\widehat{R} - R|^p\big].$$

Taking the infimum over all couplings π yields the result. □

A.3 SPHERICAL PATH

This appendix complements Section 3.3. We define the matched-radius spherical interpolation, record its basic geometric properties, and derive its explicit time derivative.

Matched-radius spherical interpolation. Let $x_0 = Ru_0$ and $x_1 = Ru_1$ with $R > 0$ and $u_0, u_1 \in S^{d-1}$. For non-antipodal pairs $u_0 \neq -u_1$, define

$$\Theta := \arccos(\langle u_0, u_1 \rangle) \in [0, \pi),$$

$$\gamma_t(u_0, u_1) = \frac{\sin((1-t)\Theta)}{\sin\Theta}\, u_0 + \frac{\sin(t\Theta)}{\sin\Theta}\, u_1, \qquad t \in [0,1],$$

and

$$\psi_t(x_0, x_1) := R\,\gamma_t(u_0, u_1).$$

All displayed quotients involving Θ are interpreted by continuous extension at $\Theta = 0$; in particular, $\gamma_t(u, u) = u$ and $\Theta/\sin\Theta = 1$ at zero.

When $u_0 = -u_1$, the minimizing geodesic is not unique. Under the RAFM sampling scheme, this event has conditional probability zero, so any fixed deterministic completion may be used on the antipodal set without affecting the construction almost surely.

Proposition A.6 (Geometry of the spherical path)**.** *For any non-antipodal pair $x_0 = Ru_0$, $x_1 = Ru_1$ with $R > 0$, the path $X_t = R\gamma_t(u_0, u_1)$ satisfies, for every $t \in [0,1]$,*

$$X_0 = x_0, \qquad X_1 = x_1, \qquad \|X_t\| = R, \qquad X_t^\top \dot{X}_t = 0, \qquad \|\dot{X}_t\| = R\Theta.$$

Among all absolutely continuous curves $\xi : [0,1] \to RS^{d-1}$ with the same endpoints, it also minimizes $\int_0^1 \|\dot{\xi}_t\|^2\, dt$, with minimum $R^2\Theta^2$.

Proof. For $0 < \Theta < \pi$, set

$$e := \frac{u_1 - \cos(\Theta)u_0}{\sin(\Theta)}.$$

Then (u_0, e) is orthonormal, and the interpolation and its derivative are

$$\gamma_t = \cos(t\Theta)u_0 + \sin(t\Theta)e,$$
$$\dot{\gamma}_t = \Theta[-\sin(t\Theta)u_0 + \cos(t\Theta)e].$$

The endpoints follow directly, while orthonormality gives $\|\gamma_t\| = 1$, $\gamma_t^\top \dot{\gamma}_t = 0$, and $\|\dot{\gamma}_t\| = \Theta$. Since $X_t = R\gamma_t$, these imply all the claimed identities. For $\Theta = 0$, the continuous extension is the constant path and the same identities hold.

Every admissible curve ξ has length at least the intrinsic spherical distance $R\Theta$ between the endpoints. Cauchy-Schwarz therefore gives

$$\int_0^1 \|\dot{\xi}_t\|^2\, dt \geq \left(\int_0^1 \|\dot{\xi}_t\|\, dt\right)^2 \geq R^2\Theta^2.$$

The constant speed $\|\dot{X}_t\| = R\Theta$ attains equality. □

Velocity of the spherical path. The next result gives the explicit time derivative of the interpolation.

Proposition A.7 (Derivative of the spherical interpolation)**.** *For any non-antipodal pair (u_0, u_1),*

$$\partial_t \gamma_t(u_0, u_1) = \frac{\Theta}{\sin\Theta}\left[-\cos((1-t)\Theta)\, u_0 + \cos(t\Theta)\, u_1\right].$$

Consequently,

$$\partial_t \psi_t(x_0, x_1) = R\frac{\Theta}{\sin\Theta}\left[-\cos((1-t)\Theta)\, u_0 + \cos(t\Theta)\, u_1\right].$$

Proof. Differentiate the coefficients of $\gamma_t(u_0, u_1)$ with respect to t:

$$\frac{d}{dt}\frac{\sin((1-t)\Theta)}{\sin\Theta} = -\frac{\Theta\cos((1-t)\Theta)}{\sin\Theta}, \qquad \frac{d}{dt}\frac{\sin(t\Theta)}{\sin\Theta} = \frac{\Theta\cos(t\Theta)}{\sin\Theta}.$$

Substituting into the definition of γ_t yields the first identity, and multiplying by R gives the second. □

A.4 Angular target and regression statistics

Recall that $X_t = R\gamma_t(U_0, U_1)$ and define $A_t := \dot{X}_t/\|X_t\| = \partial_t\gamma_t(U_0, U_1)$. Proposition A.6 gives

$$X_t^\top A_t = 0, \qquad \|A_t\| = \Theta \le \pi. \tag{14}$$

In particular, A_t has moments of every positive order independently of the radial law.

Conditional means and regression. Recall that p_t is the law of X_t. All conditional expectations below are taken under a regular conditional law given $X_t = x$ at fixed t. For p_t-almost every x, this law satisfies $R = \|x\|$ and $\|\dot{X}_t\| \le \pi\|x\|$. Thus both conditional means are finite without assuming a finite first or second radial moment, and

$$v^\star(t,x) := \mathbb{E}[\dot{X}_t \mid X_t = x] = \mathbb{E}[RA_t \mid X_t = x] \tag{15}$$
$$= \|x\|\, a^\star(t,x). \tag{16}$$

When $\mathbb{E}[R] = \infty$, the velocity expectation here is understood through this conditional law; it is locally integrable with respect to p_t because $\|v^\star(t,x)\| \le \pi\|x\|$. For any prediction $b \in \mathbb{R}^d$, expanding the conditional squared loss around its mean gives

$$\mathbb{E}[\|b - A_t\|^2 \mid X_t = x] = \|b - a^\star(t,x)\|^2 + \mathbb{E}[\|A_t - a^\star(t,x)\|^2 \mid X_t = x].$$

The velocity loss for prediction $\|x\|b$ is $\|x\|^2$ times this conditional loss. Both therefore have the same normalized conditional minimizer, even when the integrated velocity MSE is infinite.

Path marginals. For a continuously differentiable, compactly supported test function $f : \mathbb{R}^d \to \mathbb{R}$, differentiation along the paths yields

$$\frac{d}{dt}\int f(x)\, p_t(dx) = \mathbb{E}[\nabla f(X_t)^\top \dot{X}_t] = \int \nabla f(x)^\top v^\star(t,x)\, p_t(dx).$$

If the support of ∇f is contained in a ball of radius M, the integrand along each path is bounded by $\pi M \sup_x \|\nabla f(x)\|$, which justifies differentiation without a radial moment assumption. This is the weak continuity equation for p_t, with endpoints q_{rad} and p_{data}. To identify these marginals with a deterministic ODE flow, a sufficient additional assumption is that $v^\star$ admits an everywhere tangent representative that is continuous in (t,x) and locally Lipschitz in x, uniformly in t, on $\mathbb{R}^d \setminus \{0\}$. Its trajectories remain on their initial spheres, and uniqueness of characteristics then gives the marginals p_t: tangency preserves the radial law and localizes the continuity equation to bounded annuli, where the field is bounded and Lipschitz, so the uniqueness results of Ambrosio et al. (2008, Sec. 8.1) apply. This establishes the population-field statement in Theorem 3.1 and specifies the additional regularity required for its ODE interpretation.

Theorem A.8 (Moment transition for the regression targets). *Let $q \geq 1$, and let $t \sim \mathrm{Unif}[0,1]$ be independent of the endpoint variables. Let $Z \sim \mathcal{N}(0, I_d)$ be independent of X_1, and define the standard Gaussian linear-CFM target $Y_{\mathrm{G}} := X_1 - Z$. For the matched-radius coupling, define the Euclidean target $Y_{\mathrm{E}} := R(U_1 - U_0)$, the spherical velocity target $Y_{\mathrm{V}} := R\partial_t\gamma_t(U_0, U_1)$, and the angular target $Y_{\mathrm{A}} := \partial_t\gamma_t(U_0, U_1)$. With $e_1 = (1, 0, \ldots, 0)^\top$ and $\widetilde{U} \sim \mathrm{Unif}(S^{d-1})$, define*

$$c^{\mathrm{E}}_{d,q} := \mathbb{E}\|e_1 - \widetilde{U}\|^q, \qquad c^{\mathrm{V}}_{d,q} := \mathbb{E}\Big[\arccos(e_1^\top \widetilde{U})^q\Big]. \tag{17}$$

Then

$$\mathbb{E}\|Y_{\mathrm{G}}\|^q < \infty \iff \mathbb{E}[R^q] < \infty, \tag{18}$$

$$\mathbb{E}\|Y_{\mathrm{E}}\|^q = c^{\mathrm{E}}_{d,q}\mathbb{E}[R^q], \tag{19}$$

$$\mathbb{E}\|Y_{\mathrm{V}}\|^q = c^{\mathrm{V}}_{d,q}\mathbb{E}[R^q], \tag{20}$$

$$\mathbb{E}\|Y_{\mathrm{A}}\|^q = c^{\mathrm{V}}_{d,q} \leq \pi^q. \tag{21}$$

Proof. For the Gaussian target, $Y_{\mathrm{G}} = X_1 - Z$ and $X_1 = Y_{\mathrm{G}} + Z$. Since Gaussian moments are finite, Minkowski's inequality proves both directions of the first equivalence. For the matched-radius targets, condition on (R, U_1). Rotational invariance of the independent uniform U_0 makes both $\mathbb{E}[\|U_1 - U_0\|^q \mid R, U_1]$ and $\mathbb{E}[\Theta^q \mid R, U_1]$ constants depending only on (d, q). Using $\|Y_{\mathrm{V}}\| = R\Theta$ and $\|Y_{\mathrm{A}}\| = \Theta \leq \pi$ proves the result. □

Proposition A.9 (Loss-non-increasing tangential clipping). *For any measurable predictor $a : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ and $x \neq 0$, let $\mathcal{C}_x = \{w : x^\top w = 0,\ \|w\| \leq \pi\}$ and let $P_{\mathcal{C}_x}$ be Euclidean projection onto this set. Then, almost surely,*

$$\ell := \|P_{\mathcal{C}_{X_t}} a(t, X_t) - A_t\|^2 \leq \|a(t, X_t) - A_t\|^2, \qquad 0 \leq \ell \leq 4\pi^2. \tag{22}$$

Proof. By (14), $A_t \in \mathcal{C}_{X_t}$. The metric projection onto a nonempty closed convex set cannot increase distance to any point in that set, which gives the first inequality in (22). Both the projected prediction and target have norm at most π, giving the upper bound $4\pi^2$. □

For a fixed predictor and n independent RAFM-Ang training triples $(t_i, X_{t_i}, A_{t_i})_{i=1}^n$, let $\ell_1, \ldots, \ell_n$ be the corresponding squared losses, and define $\mathcal{R} := \mathbb{E}[\ell]$ and $\widehat{\mathcal{R}}_n := n^{-1}\sum_{i=1}^n \ell_i$. Hoeffding's inequality yields, for every $\varepsilon > 0$,

$$\Pr\Big(|\widehat{\mathcal{R}}_n - \mathcal{R}| > \varepsilon\Big) \leq 2\exp\left(-\frac{n\varepsilon^2}{8\pi^4}\right). \tag{23}$$

The bound contains no radial moment. It applies to the constrained or post-composed predictor; the boundedness of the target itself requires no constraint.

A.5 Tangential dynamics and generation stability

This appendix complements Section 3.5. We study the learned flow induced by v_θ and relate target approximation to generation accuracy.

Learned dynamics. Let $v_\theta : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ be a learned time-dependent vector field and consider the ODE

$$\dot{Y}_t = v_\theta(t, Y_t), \qquad Y_0 \sim q_{\mathrm{rad}}.$$

Whenever the ODE is well posed, we denote by Φ_t^θ the associated flow map, so that

$$Y_t = \Phi_t^\theta(Y_0), \qquad \mathrm{Law}(Y_t) = (\Phi_t^\theta)_{\#} q_{\mathrm{rad}}.$$

Standard sufficient regularity conditions. A standard sufficient set of conditions for the learned field v_θ is the following: it is measurable in t, globally Lipschitz in space uniformly in time with constant L_θ, and has at most linear growth. That is, there exist constants $L_\theta, M_\theta \geq 0$ such that, for all $t \in [0,1]$ and all $x, y \in \mathbb{R}^d$,

$$\|v_\theta(t,x) - v_\theta(t,y)\| \leq L_\theta \|x - y\|,$$

and

$$\|v_\theta(t,x)\| \leq M_\theta(1 + \|x\|).$$

These conditions ensure that the learned dynamics induce a well-defined flow map and justify the Grönwall argument used below.

Radial–tangential decomposition. For $x \neq 0$, any vector field can be decomposed uniquely as

$$v_\theta(t,x) = \alpha_\theta(t,x)\, x + w_\theta(t,x), \qquad x^\top w_\theta(t,x) = 0,$$

where

$$\alpha_\theta(t,x) := \frac{x^\top v_\theta(t,x)}{\|x\|^2}, \qquad w_\theta(t,x) := v_\theta(t,x) - \alpha_\theta(t,x)x.$$

The scalar α_θ is the radial component, while w_θ is the tangential component.

Proposition A.10 (Purely tangential dynamics preserve the norm)**.** *Assume that*

$$x^\top v_\theta(t,x) = 0 \qquad \forall (t,x) \in [0,1] \times (\mathbb{R}^d \setminus \{0\}).$$

Let Y_t solve

$$\dot{Y}_t = v_\theta(t, Y_t)$$

with $Y_0 \neq 0$. Then

$$\|Y_t\| = \|Y_0\| \qquad \forall t \in [0,1].$$

In particular, if $Y_0 \sim q_{\mathrm{rad}}$, then

$$\|Y_t\| \overset{d}{=} \|Y_0\| \overset{d}{=} \|X\|, \qquad X \sim p_{\mathrm{data}}.$$

Proof. Differentiate $\frac{1}{2}\|Y_t\|^2$:

$$\frac{d}{dt}\frac{1}{2}\|Y_t\|^2 = Y_t^\top \dot{Y}_t = Y_t^\top v_\theta(t, Y_t) = 0.$$

Hence $\|Y_t\|^2 = \|Y_0\|^2$ for all t. □

Reference path and population regression error. Recall the matched-radius coupling of RAFM:

$$X_1 \sim p_{\mathrm{data}}, \qquad X_0 = \|X_1\| U_0, \qquad U_0 \sim \mathrm{Unif}(S^{d-1}), \qquad X_t = \psi_t(X_0, X_1).$$

We define the population RAFM regression error by

$$\mathcal{E}_{\mathrm{RAFM}}(\theta) := \mathbb{E} \int_0^1 \left\| v_\theta(t, X_t) - \partial_t \psi_t(X_0, X_1) \right\|_2^2 dt.$$

Theorem A.11 (Generation stability from target approximation)**.** *Assume that $\mathbb{E}[\|X_1\|^2] < \infty$. Under the standard sufficient regularity conditions above, let Y_t be the learned trajectory driven by v_θ from the same initial condition X_0:*

$$\dot{Y}_t = v_\theta(t, Y_t), \qquad Y_0 = X_0.$$

Then

$$\mathbb{E}\|Y_1 - X_1\|^2 \leq e^{2L_\theta}\, \mathcal{E}_{\mathrm{RAFM}}(\theta).$$

Consequently,

$$W_2^2\big((\Phi_1^\theta)_{\#} q_{\mathrm{rad}},\, p_{\mathrm{data}}\big) \leq e^{2L_\theta}\, \mathcal{E}_{\mathrm{RAFM}}(\theta),$$

and therefore

$$W_2\big((\Phi_1^\theta)_{\#} q_{\mathrm{rad}},\, p_{\mathrm{data}}\big) \leq e^{L_\theta}\, \mathcal{E}_{\mathrm{RAFM}}(\theta)^{1/2}.$$

Proof. Let

$$\Delta_t := Y_t - X_t.$$

Since $\dot{X}_t = \partial_t \psi_t(X_0, X_1)$, we have

$$\dot{\Delta}_t = v_\theta(t, Y_t) - v_\theta(t, X_t) + \big(v_\theta(t, X_t) - \dot{X}_t\big).$$

Taking norms and using the Lipschitz property of v_θ in space,

$$\frac{d}{dt}\|\Delta_t\| \le L_\theta \|\Delta_t\| + \|v_\theta(t, X_t) - \dot{X}_t\|$$

for almost every $t \in [0,1]$. Since $\Delta_0 = 0$, Grönwall's inequality yields

$$\|\Delta_1\| \le \int_0^1 e^{L_\theta(1-s)} \|v_\theta(s, X_s) - \dot{X}_s\|\, ds \le e^{L_\theta} \int_0^1 \|v_\theta(s, X_s) - \dot{X}_s\|\, ds.$$

Squaring and using Jensen's inequality,

$$\|\Delta_1\|^2 \le e^{2L_\theta} \left(\int_0^1 \|v_\theta(s, X_s) - \dot{X}_s\|\, ds \right)^2 \le e^{2L_\theta} \int_0^1 \|v_\theta(s, X_s) - \dot{X}_s\|^2\, ds.$$

Taking expectations proves

$$\mathbb{E}\|Y_1 - X_1\|^2 \le e^{2L_\theta}\, \mathcal{E}_{\mathrm{RAFM}}(\theta).$$

Since $Y_0 = X_0 \sim q_{\mathrm{rad}}$ and $Y_1 = \Phi_1^\theta(Y_0)$,

$$\mathrm{Law}(Y_1) = (\Phi_1^\theta)_{\#} q_{\mathrm{rad}}.$$

Moreover, $X_1 \sim p_{\mathrm{data}}$. Therefore the joint law of (Y_1, X_1) is a coupling between $(\Phi_1^\theta)_{\#} q_{\mathrm{rad}}$ and p_{data}, so by the definition of Wasserstein distance,

$$W_2^2\big((\Phi_1^\theta)_{\#} q_{\mathrm{rad}},\, p_{\mathrm{data}}\big) \le \mathbb{E}\|Y_1 - X_1\|^2.$$

Combining both inequalities concludes the proof. □

The bound controls the error of the velocity field that is integrated. For RAFM-Ang, $v_\theta(t,x) = \|x\| \Pi_{T_x} a_\theta(t,x)$ and $\dot{X}_t = R A_t$, so $\mathcal{E}_{\mathrm{RAFM}}(\theta) = \mathbb{E}\int_0^1 R^2 \|\Pi_{T_{X_t}} a_\theta(t, X_t) - A_t\|^2 dt \le \mathbb{E}\int_0^1 R^2 \|a_\theta(t, X_t) - A_t\|^2 dt$, the R^2-weighted angular error. With the empirical source $\widehat{q}_{\mathrm{rad},n}$, the e^{L_θ}-Lipschitz flow map and Proposition A.5 add $e^{L_\theta} W_2(\widehat{\mu}_{R,n}, \mu_R)$ to the W_2 bound.

A.6 Practical training and sampling details

Algorithm 1 RAFM-Ang training

Require: Training set $\mathcal{D}$, batch size B, angular predictor a_θ

1: **repeat**
2: Sample $\{x_1^{(b)}\}_{b=1}^B$ from $\mathcal{D}$
3: For each b, independently draw $u_0^{(b)} \sim \mathrm{Unif}(S^{d-1})$ and $t_b \sim \mathrm{Unif}[0,1]$
4: $R_b \leftarrow \|x_1^{(b)}\|$, $u_1^{(b)} \leftarrow x_1^{(b)} / R_b$
5: $x_{t_b}^{(b)} \leftarrow R_b \gamma_{t_b}(u_0^{(b)}, u_1^{(b)})$
6: $A_b \leftarrow \dot{\gamma}_{t_b}(u_0^{(b)}, u_1^{(b)})$
7: Take an optimizer step on $\frac{1}{B}\sum_{b=1}^B \big\| a_\theta(t_b, x_{t_b}^{(b)}) - A_b \big\|^2$
8: **until** convergence

Algorithm 2 RAFM-Ang sampling

Require: Trained a_θ, fitted radial law $\widehat{\mu}_R$, numerical ODE solver

1: Independently draw $\widehat{R} \sim \widehat{\mu}_R$ and $u_0 \sim \mathrm{Unif}(S^{d-1})$
2: $x_0 \leftarrow \widehat{R}\, u_0$
3: Integrate $\dot{x}_t = v_\theta^{\mathrm{Ang}}(t, x_t)$ from $t = 0$ to $t = 1$, starting at x_0, with the field in (13)
4: **return** $x_{t=1}$

Algorithms 1 and 2 give the RAFM-Ang procedures; Section 3.5 specifies the RAFM-Vel substitutions. Both use the following numerical safeguards.

The implementation divides by a norm clamped away from zero during training and skips tangent projection below a small numerical threshold at sampling. It applies no π-norm clipping to the network output. Where applied, tangential projection removes the radial velocity component; the fixed-step solver does not renormalize states; exact norm preservation is therefore a continuous-time statement and small discretization drift may remain.

B EXPERIMENTAL DETAILS

Code and configuration files reproduce every run; this section gives the essential settings.

B.1 DATA AND PREPROCESSING

Vectors. Student-t targets contain 50,000 samples $X = M_{\mathrm{mix}} z$ with $z_i \overset{\text{i.i.d.}}{\sim} \text{Student-}t(3)$ and a Gaussian mixing matrix $(M_{\mathrm{mix}})_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, for $d \in \{16, 32\}$, without further normalization. PIV uses the public cylinder-wake dataset at Reynolds number 3900 (DOI `10.57745/DHJXM6`); vorticity is computed on the full 740×545 grid, subsampled to 8×8 ($d = 64$) or 16×16 ($d = 256$), scaled by $1/2.5$ and centered (998 snapshots). Finance uses the 49 value-weighted daily industry portfolios of the Kenneth French Data Library (14,349 days, July 1969–May 2026); Weather uses ERA5 10-metre wind components over Australia from WeatherBench2 (5,847 vectors, $d = 96$). Both are standardized per coordinate with training statistics. Splits are $60\%/20\%/20\%$, random for Student-t and PIV and chronological for Finance and Weather:

Dataset	Training	Validation	Test
Student-t	30,000	10,000	10,000
PIV	598	199	201
Finance	8,609	2,869	2,871
Weather	3,508	1,169	1,170

Images and audio. ImageNette images (13,394) are encoded by a frozen DC-AE autoencoder into $32 \times 8 \times 8$ latents ($d = 2{,}048$), scaled by 0.41407 and centered with the training mean; the split is $8{,}036/2{,}678/2{,}680$. AudioMNIST clips are resampled to 8 kHz and represented by a complex STFT of shape $2 \times 129 \times 63$; each unit-norm STFT receives an independent gain from a lognormal mixture (weights 0.6, 0.4, log-means $\log 1, \log 4$, log-SD 0.25, clipped at 0.1), with 10,200 generator-training and 3,000 external test clips. FSD50K (Fonseca et al., 2022) clips are resampled to 16 kHz, cropped or padded to four seconds and represented by a complex STFT of shape $2 \times 257 \times 251$, conditioned on 200-dimensional multi-hot labels (36,796 training and 4,170 validation clips). SEVIR (Veillette et al., 2020) uses one central 64×64 VIL frame per event (403 training and 121 validation events), generated unconditionally.

B.2 ARCHITECTURES AND TRAINING

All vector methods share a three-hidden-layer MLP of width 128 with Swish activations and input $[x; t]$ ($N(d) = 33{,}280 + 257d$ parameters). AudioMNIST and FSD50K use U-Nets of base width 96 (27.9M and 50.5M parameters), ImageNette a class-conditional SiT (width 384, depth 12, 32.5M parameters), and SEVIR an MLP with three hidden layers of width 4,096. Within each benchmark, all methods share the backbone, data, splits, effective batch and update budget:

Benchmark	Updates	Batch	Optimizer / rate	EMA
Student-t, PIV	10,000	256	Adam / 10^{-3}	none
Weather	10,000	2,048	Adam / 10^{-3}	none
Finance	10,000	4,096	Adam / 10^{-3}	none
AudioMNIST	24,000	32	AdamW / 2×10^{-4}	0.999
ImageNette	40,000	64	AdamW / 10^{-4}	0.9999
FSD50K	10,000	32	AdamW / 2×10^{-4}	0.999
SEVIR	10,000	256	AdamW / 2×10^{-4}	0.999

Learning rates are constant, weight decay is zero, and no gradient clipping or augmentation is used. Matched-source FM and RAFM fit the empirical radial law on training norms only. RAFM-Ang-R^2 multiplies each angular loss term by $R^2/m_{2,\mathrm{train}}$, where $m_{2,\mathrm{train}}$ is the mean of the squared training norms.

B.3 SAMPLING, SEEDS AND HARDWARE

FM and RAFM variants integrate their learned field with RK4 at a common network-call budget (NFE): 512 for vectors, 160 for AudioMNIST, FSD50K and SEVIR, and 100 for ImageNette. RAFM applies tangential projection $v - \langle x, v\rangle x/\|x\|^2$ at inference and samples radii by quantile interpolation of training norms. Generated samples per seed: 10,000 (Student-t, PIV), the test-set size (Finance, Weather), 2,000 (AudioMNIST), 3,000 (ImageNette), 256 (FSD50K) and 121 (SEVIR). Models are evaluated at the final update (final EMA weights for audio, images and SEVIR), without checkpoint selection. All results summarize three training seeds: $\{8925, 77395, 65457\}$ for vectors, $\{8925, 1234, 7\}$ for AudioMNIST and ImageNette, $\{8925, 8926, 8927\}$ for FSD50K and $\{8925, 7, 1234\}$ for SEVIR. Vector experiments and their timings use an NVIDIA RTX 2000 Ada Generation GPU (16 GB); ImageNette training uses AMD Instinct MI210 GPUs (64 GB); FSD50K and SEVIR training and timings use one NVIDIA H200 per run.

B.4 DOWNSTREAM PROTOCOLS

AudioMNIST. Energy metrics compare generated radii with the 3,000 external test gains. The digit classifier (95.5% accuracy on held-out real clips) standardizes log magnitudes of unit-normalized STFT directions and is therefore invariant to positive global gain. The post-hoc gain control rescales each fixed-spherical output Y to $GY/\|Y\|$, with G drawn by quantile interpolation of training norms, without retraining.

ImageNette. FID, precision, recall, density and coverage use Inception features of 3,000 generations against 2,680 held-out test images; radial metrics compare generated and test latents.

FSD50K. RAFM draws radii from the training law conditional on the requested labels (exact label set, else maximal Jaccard overlap). FAD uses PANN CNN14 embeddings at 16 kHz (Kilgour et al., 2019), with 256 generations and 256 common validation references per seed; the fixed-spherical gain control draws its radius from the same conditional law.

SEVIR. Unconditional generations are scored by the decoded-field energy score against the 121 validation events. The fixed-spherical gain control rescales each output to a radius drawn independently from the training radial law.

B.5 BASELINES

MSGM. We use the authors' public implementation (Gruhlke et al., 2026) with empirical-radius initialization, $T = 1$, $\beta_{\min} = 0.1$, $\beta_{\max} = 20$ and sliced-score matching. The sparse implementation stores $2d$ nonzeros per multiplicative-noise tensor, whereas the dense formulation stores the full generator tensor. The main text reports the sparse implementation; FSD50K and SEVIR use a sparse adaptation with eight fixed Hamiltonian-cycle generators, norm-projected forward and reverse steps, and 160 network calls.

t-Flow. Following Pandey et al. (2025, Appendix B), t-Flow uses the source $N = sZ/\sqrt{K}$ with $Z \sim \mathcal{N}(0, I_d)$ and $K \sim \chi^2_\nu/\nu$, the path $X_t = tX_1 + (1-t)N$, the noise-prediction loss $\mathbb{E}\|\epsilon_\theta(t, X_t) - N\|^2$ and the field $v_\theta = (x - \epsilon_\theta)/t$, integrated with Heun steps from $t = 10^{-3}$. The source parameters (ν, s) are selected on validation data among $\nu \in \{3, 5, 7\}$ and three scales around the median-matching scale, then frozen for the three final trainings.

B.6 EVALUATION METRICS

Radial W_1 and KS compare generated and test norm distributions. Sliced W_1 averages one-dimensional W_1 distances over random projections; angular Sliced W_1 applies the same metric to unit directions within test-radius quartiles. RBF MMD^2 uses a Gaussian kernel. FID and FAD are Fréchet distances between Inception and PANN embeddings, respectively.

Dataset	Method	Radial $W_1 \downarrow$	KS $\downarrow$	Sliced $W_1 \downarrow$	Training time
Student-t ($\nu = 3, d = 16$)	Gaussian FM	3.3415±0.9074	0.1745±0.0627	0.7595±0.2315	**18.3**±0.2 s
	Matched-source FM	0.3986±0.0444	0.0207±0.0055	0.4379±0.0644	19.0±0.3 s
	RAFM (ours)				
	Vel	0.2264±0.0476	**0.0119**±0.0029	0.3316±0.0298	35.2±0.1 s
	Ang	0.467±0.059	0.0130±0.0013	**0.250**±0.021	33.87±0.18 s
	MSGM (dense)	0.3760±0.0650	0.0142±0.0036	0.4864±0.0279	49.3±0.3 min
	MSGM (sparse)	**0.2101**±0.0365	**0.0119**±0.0014	0.3852±0.0144	50.3±1.0 min
Student-t ($\nu = 3, d = 32$)	Gaussian FM	8.3012±0.5585	0.2960±0.0208	1.5186±0.0261	**18.4**±0.2 s
	Matched-source FM	1.4295±0.2491	0.0369±0.0090	0.7513±0.2040	19.2±0.0 s
	RAFM (ours)				
	Vel	0.6162±0.0076	0.0120±0.0008	0.4749±0.0601	35.4±0.2 s
	Ang	0.401±0.043	**0.0107**±0.0020	**0.347**±0.033	34.66±0.14 s
	MSGM (dense)	**0.3747**±0.0811	0.0112±0.0026	0.6812±0.0210	51.6±0.5 min
	MSGM (sparse)	0.8604±0.0046	0.0204±0.0003	0.7223±0.0281	49.5±0.5 min
PIV ($d = 64$)	Gaussian FM	0.3043±0.0218	0.2272±0.0032	0.0463±0.0030	**18.3**±0.0 s
	Matched-source FM	0.1017±0.0032	0.1068±0.0056	0.0324±0.0046	19.3±0.0 s
	RAFM (ours)				
	Vel	0.0482±0.0019	0.0469±0.0026	**0.0273**±0.0029	35.8±0.2 s
	Ang	0.0482±0.0019	0.0469±0.0026	0.0279±0.0025	43.02±0.61 s
	MSGM (dense)	**0.0459**±0.0037	0.0474±0.0039	0.0539±0.0003	46.9±0.4 min
	MSGM (sparse)	0.0461±0.0034	**0.0448**±0.0044	0.0629±0.0008	49.2±0.2 min
PIV ($d = 256$)	Gaussian FM	8.9522±0.0243	1.0000±0.0000	0.4498±0.0024	**18.5**±0.0 s
	Matched-source FM	0.5696±0.0164	0.3971±0.0123	0.0382±0.0019	19.4±0.1 s
	RAFM (ours)				
	Vel	**0.0371**±0.0013	**0.0579**±0.0014	0.0242±0.0030	35.4±0.1 s
	Ang	**0.0371**±0.0013	**0.0579**±0.0014	**0.0227**±0.0003	65.5±1.6 s
	MSGM (dense)	0.0429±0.0030	0.0665±0.0023	0.0498±0.0015	5.39±0.02 h
	MSGM (sparse)	0.0502±0.0041	0.0718±0.0030	0.0354±0.0009	56.3±0.7 min

Table 4: Full Student-t and PIV results corresponding to Table 1. Distributional results are mean and standard deviation over three training seeds. All methods use batch 256, 10,000 training updates and 10,000 generated samples. Bold marks the lowest displayed mean within each dataset and metric. Training times and protocols are specified in Appendices C.1 and B.5.

C Additional results

C.1 MSGM implementations and training cost

Table 4 gives the complete Student-t and PIV comparison. MSGM has two implementations. The dense formulation stores the full multiplicative-noise generator tensor; the sparse implementation of the authors' public code keeps $2d$ nonzeros per noise tensor (Appendix B.5). In the evaluated settings, the dense formulation is at best comparable in cost and is substantially more expensive at higher dimension: on Student-t and PIV $d = 64$ the two have similar training times (46.9–51.6 min dense, 49.2–50.3 min sparse), but on PIV $d = 256$ dense training takes 5.39 h versus 56.3 min, and the dense generator tensor (d^3 float32 entries) would require 34.4 GB for ImageNette and 17.2 TB for AudioMNIST (Appendix C.1). Both implementations remain far slower than RAFM, which trains in at most 65.5 s. Both RAFM variants have a lower mean Sliced W_1 than both MSGM implementations on all six vector benchmarks (Tables 4, 6 and 7), although either MSGM implementation can reach a smaller radial error. The main text therefore reports the sparse implementation as its single MSGM baseline, labeled MSGM; this is a presentation choice, and both implementations are reported in these tables. FSD50K and SEVIR use the sparse-union8 adaptation of Appendix B.5.

Table 5 reports measured end-to-end training times for 10,000 updates, averaged over three seeds and excluding compilation and warm-up.

FSD50K and SEVIR. Both comparisons use an NVIDIA H200 and shared backbones and effective batches within each dataset. FSD50K nominal training times for 10,000 updates average 2,883.35 s (Ang), 2,888.90 s (Vel), 26,973.69 s (sparse-union8 MSGM), and 2,833.38 s (t-Flow) over three seeds. They include data preparation, objectives, optimizer/EMA updates and MSGM forward simulation, but exclude initialization, checkpoints, evaluation and lost or replayed computation.

MSGM accumulates two batches of 16 to the common effective batch 32. Its ratios to Ang/Vel are $9.355\times$ /$9.337\times$, at equal update budgets rather than equal quality.

For SEVIR, comparable windows give 8.536 ms/update (Ang), 8.499 ms/update (Vel), and 225.044 ms/update (MSGM), ratios $26.36\times$ /$26.48\times$. Each seed contributes 75 windows of 50 updates between updates 5,000 and 10,000, excluding evaluation/checkpoint intervals and including forward simulation. These are update-time ratios, not full training-job totals. Source-selection costs are separate (Appendix B.5); neither comparison establishes a generation-time speedup or a speedup over t-Flow.

Dense MSGM: Finance batch $4{,}096$**, Weather batch** $2{,}048$

Dataset	RAFM-Ang (s)	MSGM (dense, s)	Speedup
Finance	43.02	18,657.99	$433.7\times$
Weather	43.90	20,085.23	$457.5\times$

Sparse MSGM: batch 256 **for both methods**

Setting	Batch	Sparse MSGM	RAFM-Ang	Speedup
Student-t ($d = 16$)	256	50.3 min	33.87 s	$89.2\times$
Student-t ($d = 32$)	256	49.5 min	34.66 s	$85.7\times$
PIV ($d = 64$)	256	49.2 min	43.02 s	$68.6\times$
PIV ($d = 256$)	256	56.3 min	65.5 s	$51.6\times$
Finance MLP	256	49.7 min	51 s	$58.4\times$
Weather MLP	256	46.0 min	44 s	$62.7\times$

Table 5: Measured end-to-end training times for $10{,}000$ updates, averaged over three seeds. Dense and sparse panels use the stated, distinct batch sizes; ratios use unrounded timings.

C.2 Structured-data metrics

Tables 6 and 7 report radial, joint and angular metrics for Finance and Weather.

Finance radial shift. The training-to-test radial W_1 is 1.386. The model radial errors in Table 6 remain near this shift, illustrating the limitation of preserving a training-derived radial law under chronological distribution shift.

C.3 ImageNette and AudioMNIST

Tables 8 and 9 report the complete ImageNette and AudioMNIST metrics.

C.4 What the DC-AE radius encodes

For a nonzero centered, officially scaled DC-AE latent X, let $R := \|X\|$ and $U := X/R$. For $\tau \in (0, 1)$, let $\mathcal{Q}_R(\tau)$ denote the empirical τ-quantile of the training-radius distribution. This radius is distinct from pixel-space image energy. The diagnostics in Table 10 associate global radius with contrast and extreme-pixel prevalence, defined as the fraction of grayscale pixels below 0.1 or above 0.9, while normalized direction retains nearly all linear class information. Correlations use a fixed random subset of $3{,}000$ real images. The multinomial probes use a deterministic $60\%/20\%/20\%$ split (8,036 training and 2,680 test latents), and the decoded intervention (Figure 3) averages 32 held-out directions. These are separate real-latent representation diagnostics with training-only centering; they do not use the FID reference or generated samples of Table 8.

C.5 Gaussian source-scale control

This control rescales the Gaussian source to $\mathcal{N}(0, \sigma^2 I_d)$ with $\sigma = \text{median}\{\|x_i\|\}/F^{-1}_{\chi_d}(1/2)$, fitted once on training norms, keeping the straight-line velocity objective. Scale adjustment substantially reduces several gaps, especially on Student-t and PIV, while RAFM-Ang has lower mean Sliced W_1 in all six conditions (Table 11); the smallest relative margin is on Finance (0.1721 versus 0.1768). The gains are therefore not explained solely by a Gaussian scale mismatch.

Metric	Gaussian FM	Matched-source FM	RAFM (ours) Vel	RAFM (ours) Ang	MSGM (dense)	MSGM (sparse)
Radial W_1	$1.4179_{\pm 0.1132}$	$1.7068_{\pm 0.0398}$	$1.4229_{\pm 0.0401}$	$1.4229_{\pm 0.0401}$	$\mathbf{1.4168}_{\pm 0.0335}$	$1.4831_{\pm 0.0446}$
Radial KS	$0.2184_{\pm 0.0090}$	$0.2460_{\pm 0.0009}$	$0.2003_{\pm 0.0028}$	$0.2003_{\pm 0.0028}$	$\mathbf{0.1975}_{\pm 0.0059}$	$0.2053_{\pm 0.0058}$
Sliced W_1	$0.1941_{\pm 0.0180}$	$0.2076_{\pm 0.0045}$	$0.1741_{\pm 0.0026}$	$\mathbf{0.1721}_{\pm 0.0037}$	$0.1820_{\pm 0.0029}$	$0.1816_{\pm 0.0021}$
RBF MMD^2	$0.0117_{\pm 0.0019}$	$0.0131_{\pm 0.0004}$	$0.0098_{\pm 0.0006}$	$\mathbf{0.0095}_{\pm 0.0005}$	$0.0124_{\pm 0.0005}$	$0.0118_{\pm 0.0009}$
Angular Sliced W_1 (mean)	$0.0236_{\pm 0.0025}$	$0.0205_{\pm 0.0006}$	$0.0173_{\pm 0.0003}$	$\mathbf{0.0172}_{\pm 0.0006}$	$0.0195_{\pm 0.0004}$	$0.0180_{\pm 0.0005}$

Table 6: Complete Finance results under the chronological split. Entries are mean and standard deviation over three training seeds. All metrics are lower-is-better; bold marks the smallest mean at the displayed precision. Sparse protocol: Appendix B.5.

Metric	Gaussian FM	Matched-source FM	RAFM (ours) Vel	RAFM (ours) Ang	MSGM (dense)	MSGM (sparse)
Radial W_1	$0.1786_{\pm 0.0311}$	$0.3797_{\pm 0.0371}$	$\mathbf{0.0873}_{\pm 0.0048}$	$\mathbf{0.0873}_{\pm 0.0048}$	$0.0987_{\pm 0.0282}$	$0.1179_{\pm 0.0096}$
Radial KS	$0.0442_{\pm 0.0102}$	$0.0838_{\pm 0.0178}$	$\mathbf{0.0330}_{\pm 0.0018}$	$\mathbf{0.0330}_{\pm 0.0018}$	$\mathbf{0.0330}_{\pm 0.0093}$	$0.0385_{\pm 0.0014}$
Sliced W_1	$0.0854_{\pm 0.0075}$	$0.0944_{\pm 0.0070}$	$\mathbf{0.0746}_{\pm 0.0045}$	$0.0767_{\pm 0.0007}$	$0.1199_{\pm 0.0029}$	$0.1264_{\pm 0.0042}$
RBF MMD^2	$0.0027_{\pm 0.0007}$	$0.0030_{\pm 0.0005}$	$\mathbf{0.0019}_{\pm 0.0004}$	$0.0020_{\pm 0.0001}$	$0.0042_{\pm 0.0002}$	$0.0053_{\pm 0.0002}$
Angular Sliced W_1 (mean)	$0.0150_{\pm 0.0001}$	$0.0150_{\pm 0.0006}$	$\mathbf{0.0134}_{\pm 0.0005}$	$0.0144_{\pm 0.0005}$	$0.0187_{\pm 0.0003}$	$0.0198_{\pm 0.0004}$

Table 7: Complete Weather results under the chronological split, using the main three-hidden-layer multilayer perceptron. Entries are mean and standard deviation over three training seeds. All metrics are lower-is-better; bold marks the smallest mean at the displayed precision. Sparse protocol: Appendix B.5.

C.6 A CONTROLLED RADIUS-DIRECTION DEPENDENCE TEST

Let $\log R \sim \mathcal{N}(0, 0.7^2)$ and $\varepsilon \sim \mathrm{vonMises}(0, 8)$ be independent, and set $X = R(\cos\Theta, \sin\Theta)$ with $\Theta = \pi \tanh(\log S/0.7) + \varepsilon$. The dependent case uses $S = R$; the independent case draws S independently from the same law, so both cases share the same radial and angular marginals. RAFM-Ang, RAFM-Vel and a fixed-spherical model followed by independent training-law radii use the common $d = 2$ MLP and vector protocol (50,000 samples, 30,000/10,000/10,000 split). All methods have the same radial errors (radial W_1 0.01291); independent output radii miss the joint structure only in the dependent case (Figure 4).

C.7 FSD50K COMPLEMENTARY METRICS

Table 12 complements Table 3. At similar radial errors, RAFM-Ang has higher mean embedding recall than RAFM-Vel and Ang-R^2 (0.124 versus 0.031 and 0.013).

C.8 ANGULAR-LOSS WEIGHTING CONTROL

RAFM-Ang-R^2 keeps the angular predictor, input and sampling field of RAFM-Ang and only multiplies each loss term by $R^2/m_{2,\mathrm{train}}$. Table 13 shows that unweighted angular regression has lower mean Sliced W_1 on Student-t and lower FID on ImageNette; Table 3 gives the corresponding FSD50K comparison and Table 14 the SEVIR one, where the weighted objective is preferred.

C.9 SEVIR

Table 14 evaluates unconditional generation of SEVIR VIL frames. Here Ang-R^2 and RAFM-Vel improve over unweighted Ang, indicating task-dependent radial weighting; RAFM-Vel is close to the TRAIN resampling reference.

Image-space metrics

Method	FID ↓	Precision ↑	Recall ↑	Density ↑	Coverage ↑
Gaussian FM	161.917±0.933	0.1023±0.0066	0.6840±0.0096	0.0315±0.0014	0.0547±0.0029
Matched-source FM	171.894±1.238	0.1007±0.0044	**0.7007**±0.0048	0.0287±0.0019	0.0470±0.0037
Fixed-spherical FM	166.105±0.205	0.1208±0.0075	0.6960±0.0092	0.0378±0.0011	0.0585±0.0017
RAFM (ours)					
Vel	174.830±0.918	0.1096±0.0081	0.6757±0.0197	0.0342±0.0019	0.0494±0.0044
Ang	**149.959**±0.414	**0.1453**±0.0026	0.6831±0.0155	**0.0501**±0.0013	**0.0794**±0.0044
t-Flow	224.669±6.143	0.0113±0.0009	0.1787±0.0540	0.0024±0.0002	0.0044±0.0004
MSGM sparse	280.835±2.214	0.0230±0.0049	0.0002±0.0004	0.0047±0.0011	0.0021±0.0006

Centered-latent metrics

Method	Radial W_1 ↓	Radial KS ↓	Sliced W_1 ↓
Gaussian FM	4.747±0.170	0.429±0.017	0.09483±0.00290
Matched-source FM	2.616±0.150	0.238±0.014	0.06635±0.00173
Fixed-spherical FM	3.770±0.000	0.535±0.000	0.04752±0.00056
RAFM (ours)			
Vel	0.266±0.000	0.040±0.000	**0.04505**±0.00064
Ang	0.266±0.000	0.040±0.000	0.04707±0.00082
t-Flow	405.595±66.370	0.674±0.090	7.24313±1.16149
MSGM sparse	5.110±0.001	0.468±0.000	0.11513±0.00024

Table 8: Final class-conditional ImageNette comparison: mean ± population SD over three training seeds. Each model generates 3,000 images at 100 network calls per image; all image-space metrics use the same 2,680-image generator-test reference. Latent metrics use the corresponding centered test latents. Gray groups the RAFM variants. Bold marks the best mean, except that the practically identical Vel/Ang radial scores are left unranked. Zero displayed SD can reflect rounding. Protocol: Appendices B.4 and B.5.

Method	Digit accuracy ↑	Energy KS ↓	Mass above target $\mathcal{Q}_g(0.95)$	Mass above target $\mathcal{Q}_g(0.99)$
Gaussian FM	0.734±0.004	0.117±0.008	0.0587±0.0041	0.0195±0.0032
Matched-source FM	0.750±0.018	0.095±0.010	0.0397±0.0013	0.0118±0.0002
Fixed-spherical FM	**0.810**±0.013	0.592±0.000	0.0000±0.0000	0.0000±0.0000
Fixed-spherical FM + independent gain*	**0.810**±0.013	**0.0218**±0.0000	0.0500±0.0000	0.0145±0.0000
RAFM (ours)				
Vel	0.711±0.014	**0.0218**±0.0000	0.0500±0.0000	0.0145±0.0000
Ang	0.7883±0.0113	**0.0218**±0.0000	0.0500±0.0000	0.0145±0.0000
t-Flow	0.1448±0.0069	0.8323±0.0096	0.8755±0.0110	0.7640±0.0178
MSGM (sparse)	0.100±0.000	0.0340±0.0000	0.0465±0.0000	0.0105±0.0000

Table 9: AudioMNIST digit accuracy, energy-distribution match, and tail coverage at 24,000 training steps. Values are mean ± standard deviation over three training seeds, using 2,000 generated clips per seed. $\mathcal{Q}_g$ denotes the external-test gain quantiles. *Post-hoc independent gains on the fixed-spherical models, without retraining; the gain-invariant classifier leaves every prediction, and hence the fixed-spherical accuracy, unchanged (Appendix B.4). Protocols are in Appendices B.4 and B.5.

Diagnostic	Value	Interpretation
Pearson correlation: radius vs. extreme-pixel fraction	0.700	Associated with extreme pixels
Pearson correlation: radius vs. contrast	0.435	Associated with contrast
Class accuracy from R	0.165	Weak linear class signal
Class accuracy from U	0.616	Higher linear class signal
Class accuracy from full latent	0.611	Similar probe accuracy
Fixed U, contrast at $\mathcal{Q}_R(0.05) \to \mathcal{Q}_R(0.95)$	$0.254 \to 0.329$	+29% over 32 directions

Table 10: Representation-level diagnostics for the ImageNette DC-AE latent. The class results use the same multinomial linear-probe protocol. The fixed-direction intervention averages 32 directions and changes only the global radius.



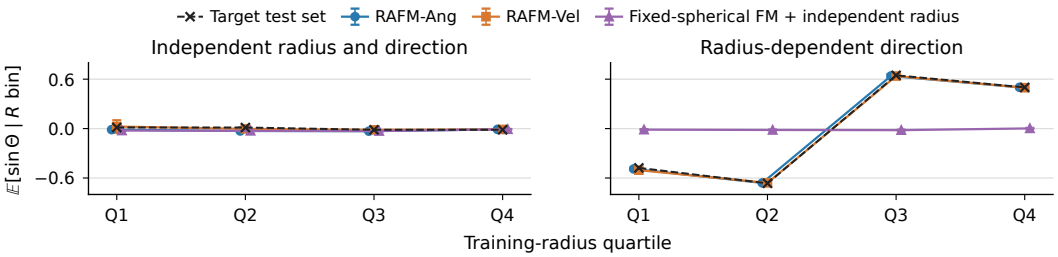
Figure 3: Fixed-direction interventions in the ImageNette DC-AE latent. Each row starts from a held-out real-image latent and shows its original-radius reconstruction followed by reconstructions at five training-set radius quantiles. The centered unit direction is fixed: object identity and layout remain stable while contrast changes with radius. Table 10 reports the aggregate diagnostic over 32 separately sampled held-out directions.

	Gaussian FM	Scaled Gaussian FM	RAFM-Ang
Student-t $d = 16$	.7595$\pm$.2315	.3765$\pm$.0527	**.250**$\pm$.021
Student-t $d = 32$	1.5186$\pm$.0261	.7467$\pm$.1440	**.347**$\pm$.033
PIV $d = 64$	.0463$\pm$.0030	.0323$\pm$.0016	**.0279**$\pm$.0025
PIV $d = 256$	.4498$\pm$.0024	.0376$\pm$.0033	**.0227**$\pm$.0003
Finance	.1941$\pm$.0180	.1768$\pm$.0103	**.1721**$\pm$.0037
Weather	.0854$\pm$.0075	.0914$\pm$.0022	**.0767**$\pm$.0007

Table 11: Gaussian scale control: Sliced $W_1 \downarrow$, mean $\pm$ population SD over three seeds. Bold marks the lowest mean. Gaussian FM and RAFM-Ang values are those of Table 1.

Table 12: FSD50K complementary metrics: mean $\pm$ sample SD over three training seeds, using the same 256 generated and 256 validation clips as Table 3. W_1 compares the indicated radii; recall and coverage use PANN embeddings. A dagger marks positive SD below 0.001. Gain is conditional on labels and requires no new training.

Method	STFT $W_1 \downarrow$	Waveform $W_1 \downarrow$	Recall $\uparrow$	Coverage $\uparrow$
Gaussian FM	73.144$\pm$3.822	10.590$\pm$0.545	0.035$\pm$0.010	0.461$\pm$0.022
Matched-source FM	24.788$\pm$1.102	3.618$\pm$0.174	0.021$\pm$0.019	0.409$\pm$0.018
Fixed-spherical FM	93.330$\pm$0.001	13.263$\pm$0.057	0.096$\pm$0.009	0.542$\pm$0.011
Fixed-spherical FM + conditional gain	23.234†	3.435$\pm$0.048	0.085$\pm$0.030	0.569$\pm$0.012
RAFM-Vel	23.231$\pm$0.001	3.392$\pm$0.001	0.031$\pm$0.008	0.371$\pm$0.010
RAFM-Ang	23.231†	3.438$\pm$0.041	0.124$\pm$0.018	0.577$\pm$0.010
RAFM-Ang-R^2	23.234†	3.465$\pm$0.017	0.013$\pm$0.006	0.365$\pm$0.033
t-Flow	476.154$\pm$47.274	58.709$\pm$6.665	0.003$\pm$0.002	0.250$\pm$0.038
MSGM sparse-union8	23.234†	6.722†	0.000$\pm$0.000	0.078$\pm$0.000



	Independent	Dependent
RAFM-Ang	.02835±.00249	**.02723**±.00447
RAFM-Vel	.03556±.00862	.02755±.00438
Fixed-spherical FM + independent radius	**.02553**±.00236	.32051±.00288

Figure 4: Conditional angular summaries from the radius-direction control: test-set and generated mean $\sin\Theta$ within training-radius quartiles. Below: joint Sliced $W_1 \downarrow$. Means, SDs, and error bars use three seeds (population SD), from the 18 runs. Independent output radii miss the dependent angular pattern.

Condition / metric ↓	Angular, unweighted	Angular, normalized R^2
Student-t, $d = 16$ / Sliced W_1	**0.250**±0.021	0.3886±0.0371
Student-t, $d = 32$ / Sliced W_1	**0.347**±0.033	0.5418±0.0309
ImageNette / FID	**149.959**±0.414	158.132±0.960

Table 13: Identical angular predictor and sampling field; only training-example weights change, from 1 to $R^2/m_{2,\text{train}}$. Mean ± population SD over three seeds; Student-t has $\nu = 3$.

Table 14: SEVIR validation energy score after 10,000 updates and 160 network calls, with 121 generated and 121 validation frames per seed. Entries are means ± sample SD over three training seeds. The separate TRAIN resampling reference is one shared deterministic control and has no training-seed SD. Lower is better.

Method	Decoded-field energy score ↓
Gaussian FM	118.906±0.273
Matched-source FM	116.665±0.386
Fixed-spherical FM + independent gain	112.080±0.356
RAFM-Vel	110.079±0.030
RAFM-Ang	111.725±0.237
RAFM-Ang-R^2	110.360±0.166
t-Flow	13987.968±24.698
MSGM sparse-union8	118.137±0.001
TRAIN resampling reference	110.253