GEOMETRIC FLOW MATCHING FOR LEARNING QUANTUM MANY-BODY STATES WITH CLASSICAL SHADOWS

Anonymous authors
Paper under double-blind review

ABSTRACT

Classical shadows provide an efficient representation of quantum many-body states, enabling machine learning approaches to predict properties of unseen states. The distribution of shadows can be viewed as a tomographic representation of a quantum state, from which many observables can be estimated. Learning its Hamiltonian-conditional distribution therefore provides a generative surrogate for estimating multiple properties of quantum states unseen during training. Existing approaches, however, either generate qubits autoregressively or diffuse continuous token embeddings; neither explicitly exploits the discrete and physically meaningful geometry of shadow outcomes. We ask whether incorporating this structure improves conditional shadow generation. We introduce **ShadowFlow**, a non-autoregressive flow-matching framework that explicitly incorporates the geometry of shadow outcomes through two complementary constructions. We evaluate ShadowFlow on one-dimensional quantum systems up to 100 qubits and two-dimensional systems up to 6×6, comparing against direct regression, autoregressive generation, diffusion, and continuous and discrete flow baselines. Across these benchmarks, ShadowFlow improves observable estimation over existing generative approaches in several settings. We also further demonstrate its extension beyond Pauli measurements to other IC-POVMs.

1 INTRODUCTION

Flow Matching (FM) has recently emerged as a powerful generative modeling framework, achieving strong performance in both continuous (Lipman et al., 2022; Chen & Lipman, 2024; Lipman et al., 2024; Albergo et al., 2023) and discrete (Stärk et al., 2024; Gat et al., 2024; Davis et al., 2024; Cheng et al., 2024; 2025) domains. Extensions to Riemannian manifolds further enable generative modeling on non-Euclidean geometries (Chen & Lipman, 2024).

In parallel, increasingly large quantum many-body systems make it challenging to extract physical properties from their quantum states. A generic L-qubit density matrix has $4^L - 1$ real degrees of freedom, making full tomography intractable even at moderate scale. Classical shadow tomography (Huang et al., 2020) offers a different interface: randomized local measurements produce compact classical records that can be reused to estimate many observables without reconstructing the density matrix.

A promising direction is therefore to learn the Hamiltonian-conditional distribution of complete shadow records with a generative model. If such a model generalizes across a Hamiltonian family, it provides a reusable surrogate for states at unseen parameter values: generated shadows can be reused to estimate different observables without retraining an observable-specific predictor. Once trained, the model can generate additional shadows through classical inference without further quantum measurements.

Recent works have explored generative modeling of classical shadows using autoregressive models (Carrasquilla et al., 2019; Yao & You, 2024), energy-based models (Jayakumar et al., 2024), and diffusion models (Tang et al., 2025). Existing approaches, however, either generate outcomes sequentially or model them through continuous embeddings without explicitly incorporating the geometry of the measurement outcomes. For Pauli-6 measurements, for example, opposite spin

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

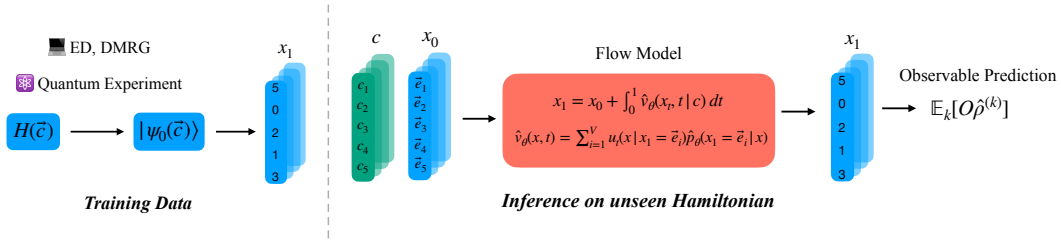

Figure 1: **Overview of ShadowFlow.** We train a flow model on shadow sequences x_1 obtained from randomized single-qubit measurements of ground states $|\psi_0(\mathbf{c})\rangle$ drawn from a Hamiltonian family $\{H(\mathbf{c})\}$. *Left:* Training states and their shadows are generated by exact diagonalization or DMRG (Schollwöck, 2005). *Right:* At inference, the learned Hamiltonian-conditional vector field $\hat{v}_\theta(x_t, t \mid \mathbf{c})$ transports prior samples x_0 toward shadow samples x_1 along the chosen geometry. A generated record gives the product snapshot $\hat{\rho}^{(k)} = \bigotimes_i \big(3U_i^{(k)\dagger}|b_i^{(k)}\rangle\langle b_i^{(k)}|U_i^{(k)} - \mathbb{I}\big)$, and observables are estimated by averaging $\mathrm{tr}(O\hat{\rho}^{(k)})$ over generated records.

outcomes along the same axis are antipodal on the Bloch sphere, whereas outcomes associated with different Pauli axes are orthogonal. This motivates our central question:

> *Can conditional shadow generation for predicting unseen quantum states be improved by explicitly incorporating the discreteness and geometry of shadows?*

Geometric flow matching provides a natural framework for addressing this question. Riemannian FM defines transport on curved manifolds (Chen & Lipman, 2024), while continuous-state discrete FM (CS-DFM) lifts categorical endpoints to continuous spaces and generates all tokens in parallel (Stärk et al., 2024; Davis et al., 2024; Cheng et al., 2024; 2025).

Building on these ideas, we develop **ShadowFlow**, a non-autoregressive geometric flow-matching framework for Hamiltonian-conditional generation of classical shadows. We first show analytically and empirically that confusing opposite spin outcomes is more damaging than changing the measurement basis at a matched corruption rate. We then introduce two complementary constructions: *Spherical Flow*, which transports outcomes along Bloch-sphere geodesics, and *Anisotropic Dirichlet Flow*, which incorporates target–antitarget structure into the categorical probability path. Both generate all sites jointly and in parallel while conditioning on the underlying Hamiltonian.

We summarize as follows:

- We motivate incorporating geometric structure into generative modeling of shadows, drawing inspiration from the Bloch sphere representation. We show experiments that demonstrate respecting this underlying geometry significantly improves the accuracy of observable estimation.
- We develop two geometric approaches to design flow matching specifically for learning shadows: (1) a *Riemannian* approach and (2) an *anisotropic probability path* based approach, which generalizes existing CS-DFM. These are motivated from the Bloch sphere geometry and the inherent (target, anti-target) pairing structure of shadows.
- We benchmark one-dimensional systems up to 100 qubits and two-dimensional systems up to 6×6 qubits and against direct regression, autoregressive, diffusion, and flow-based baselines. We also show the framework can be generalized to non-Pauli POVM-based shadows.

2 Background

2.1 Shadow Representation of Quantum States

A central challenge in machine learning quantum information is the exponential complexity of representing quantum states: a density matrix for L qubits is a $2^L \times 2^L$ Hermitian, trace-one matrix with $4^L - 1$ real degrees of freedom. Classical shadow tomography (Huang et al., 2020) avoids full reconstruction of this object for various computational tasks. Instead, each randomized measurement produces a compact snapshot whose expectation equals the state under a known inverse measurement

map. Intuitively, one shadow is a noisy view of the state: no single record is informative enough, but averaging many records enables accurately estimating important physical observables such as estimating correlations, subsystem purity. Crucially, the same records can be post-processed for a new observable.

The Pauli-6 POVM measures each qubit in an eigenbasis of X, Y, or Z. A single outcome is one of $\{|X^{\pm}\rangle, |Y^{\pm}\rangle, |Z^{\pm}\rangle\}$, and one shot on L qubits is a length-L categorical string $s = (s_1, \ldots, s_L)$. Its product snapshot:

$$\hat{\rho}(s) = \bigotimes_{i=1}^{L} (3|s_i\rangle\langle s_i| - \mathbb{I}), \tag{1}$$

serves as the estimator of the original quantum state and observables are estimated by averaging $\mathrm{tr}(O\hat{\rho})$ over shots. Further details about how to acquire shadow measurements for training data, to estimate interested observables from the inferred shadow and sample complexity details are in Section A.

2.2 Flow Matching

Linear Flow Matching. Flow matching (FM) (Lipman et al., 2022; Liu et al., 2022; Albergo et al., 2023) is a generative modeling framework that learns velocity fields transporting prior distributions to target data distributions, with simulation-free training. The linear flow matching framework trains a velocity field predictor v_θ with a conditional flow matching (CFM) loss:

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t,\, x_1 \sim q,\, x_t \sim p_t} \left\| v_\theta(x_t, t) - u_t(x_t \mid x_1) \right\|_2^2, \tag{2}$$

where q is target distribution, $p_t(\cdot|x_1)$ is the conditional probability path, and $u_t(x_t|x_1)$ is the conditional velocity field, usually defined with linear interpolation $(x_1 - x_t)/(1-t)$. Upon convergence, one can solve the ODE $\frac{dx_t}{dt} = \hat{v}_\theta(x_t, t)$, to generate samples from the target distribution.

Riemannian Flow Matching. When data lie on a manifold $\mathcal{M}$, straight Euclidean interpolation can leave the valid space. Riemannian FM (Chen & Lipman, 2024) replaces it by the geodesic

$$x_t = \exp_{x_0}\!\big(t \log_{x_0} x_1\big), \qquad u_t(x_t \mid x_1) = \frac{\log_{x_t}(x_1)}{1-t},$$

where $\exp$ and $\log$ move between the manifold and its tangent spaces. These are used as the regression target in the Riemannian Flow Matching (RFM) objective:

$$\mathcal{L}_{\mathrm{RFM}} = \mathbb{E}_{t,x_0,x_1} \left[\left\| v_\theta(x_t, t) - u_t(x_t|x_0, x_1) \right\|_g^2 \right]. \tag{3}$$

Discrete Flow Matching. One way to model discrete data with flow matching is to promote the discrete data to a continuous embedding space. We will refer to this as continuous-state discrete flow matching (CS-DFM) (Cheng et al., 2025). For a categorical endpoint v_k, CS-DFM lifts a hard token to a continuous manifold $\mathcal{M}$, learns the endpoint posterior, and forms

$$\hat{v}_\theta(x, t, c) = \sum_{k=1}^{K} u_t(x \mid v_k)\, \hat{p}_\theta(k \mid x, t, c). \tag{4}$$

Thus all sites move in parallel and the terminal state is decoded to a category. The lift and path encode the inductive bias: Statistical FM uses a Fisher–Rao spherical embedding, whereas Dirichlet FM prescribes probability paths on the simplex (Cheng et al., 2024; Davis et al., 2024; Stärk et al., 2024).

3 Related Work

Machine Learning Classical Shadows and Quantum Properties. Classical shadows (Huang et al., 2020; 2022; Hadfield et al., 2022; Koh & Grewal, 2022) estimate many observables from randomized local measurements. Autoregressive and energy-based models have been used to learn quantum measurement distributions (Carrasquilla et al., 2019; Jayakumar et al., 2024), while normalizing flows model quantum states through continuous POVM representations (Cranmer et al., 2019). More recently, ShadowGPT (Yao & You, 2024) and QuaDiM (Tang et al., 2025) learn

Hamiltonian-conditioned shadow distributions through autoregressive generation and Euclidean diffusion, respectively. Direct regression provides an alternative for predicting prespecified observables and can be highly competitive in measurement-limited settings (Zhao et al., 2025), but requires a separate predictor for each target quantity. In contrast, our goal is to learn a reusable shadow distribution from which multiple observables can be estimated, using non-autoregressive transport that explicitly incorporates measurement geometry.

Discrete Flow Matching and Riemannian Flow Matching. Statistical Flow Matching (Cheng et al., 2024; 2025; Davis et al., 2024) and Dirichlet Flow Matching (Stärk et al., 2024) extend flow matching to categorical data through Fisher–Rao and simplex probability paths, while Riemannian Flow Matching (Chen & Lipman, 2024) defines transport on non-Euclidean manifolds. Building on these frameworks, we design flow paths that explicitly incorporate the geometry of quantum measurement outcomes.

4 METHOD

4.1 BLOCH-SPHERE GEOMETRY AND OBSERVABLE ERROR

The Bloch sphere (Bloch, 1946; Feynman et al., 1957; Nielsen & Chuang, 2010) is the natural geometric representation of pure single-qubit states. In this representation, the Pauli-6 outcomes are the six axis-aligned points $\{\pm e_x, \pm e_y, \pm e_z\} \subset S^2$: opposite outcomes in the same basis are antipodal, while outcomes from different bases are orthogonal. We give the projective-space and metric derivation in Section B.

Definition 4.1. Write a Pauli-6 token as $(a, s) \in \{X, Y, Z\} \times \{\pm 1\}$. Independently at each site and shot, $\mathcal{E}_{\mathrm{spin}}(p)$ maps (a, s) to $(a, -s)$ with probability p and otherwise leaves it unchanged. Under $\mathcal{E}_{\mathrm{basis}}(p)$, the token is unchanged with probability $1 - p$; with probability p, a is replaced uniformly by one of the other two axes and s is resampled uniformly from $\{\pm 1\}$.

Theorem 4.2. *Let $P = \bigotimes_{j \in S} \sigma_{t_j}$ be a Pauli string of weight $w = |S| \geq 1$, let $\mu = \mathrm{tr}(P\rho)$, and let $\hat{\mu}_{\mathcal{E}}$ be its M-shot Pauli-shadow estimator after applying one of the channels above. Then*

$$\mathbb{E}[\hat{\mu}_{\mathrm{spin}}] = (1 - 2p)^w \mu, \qquad \mathbb{E}[\hat{\mu}_{\mathrm{basis}}] = (1 - p)^w \mu. \tag{5}$$

For every $0 < p \leq \frac{1}{2}$ and $\mu \neq 0$, there is a finite M_0 such that, for all $M > M_0$,

$$\mathrm{MSE}(\hat{\mu}_{\mathrm{spin}}) > \mathrm{MSE}(\hat{\mu}_{\mathrm{basis}}). \tag{6}$$

One sufficient threshold is $M_0 = 3^{2w} / \{\mu^2([1 - (1 - 2p)^w]^2 - [1 - (1 - p)^w]^2)\}$.

The finite-M variance calculation and proof are in Section C. Intuitively, a basis error removes an informative contribution, whereas a spin error directly reverses its sign. This shows that spin error is more crucial than basis error in terms of observable estimation.

Figure 2 (left) confirms this distinction for adjacent XX, YY, and ZZ observables. At every nonzero corruption rate shown, the paired 95% confidence interval of $\mathrm{RMSE}_{\mathrm{spin}} - \mathrm{RMSE}_{\mathrm{basis}}$ lies above zero. This observation highlights that spin errors are more detrimental to accurately estimating physical observables. This result motivates us to design generative models for shadows such that measurement outcomes differing by a spin flip are placed further apart in the embedding space compared to those differing by a basis flip, as illustrated in Figure 2 (right).

4.2 GEOMETRIC FLOW MATCHING FOR SHADOWS

We introduce two complementary geometric approaches: (1) *Spherical Flow*, a Riemannian flow-matching model that respects the S^2 geometry induced by the Bloch map (Section 4.2.1), and (2) *Anisotropic Dirichlet Flow*, a probability-path construction that encodes the target–antitarget structure of shadows (Section 4.2.2).

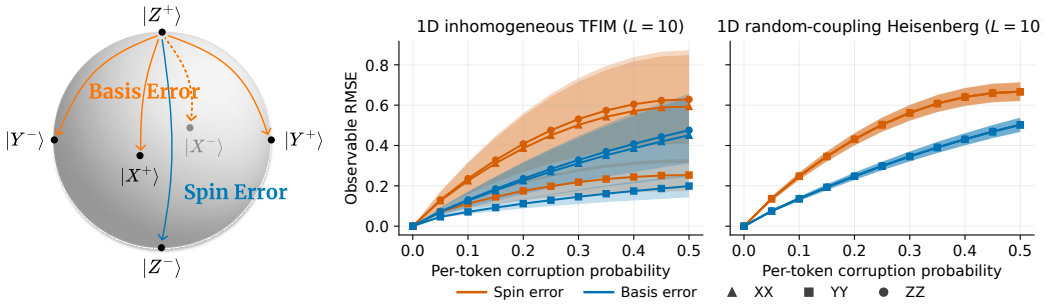

Figure 2: **Why measurement geometry matters. Left:** Illustration of bloch sphere and two error channels valid for shadows. **Right:** Effect of shadow errors on the prediction of observables. We estimate observables under errors in shadow for ground states of 1D TFIM and 1D Heisenberg, and then plot prediction error with respect to the true observable values, obtained under zero noise. The prediction errors are averaged over 1000 states. Notice that a spin error, which takes us across the Bloch sphere, is more detrimental across all error rate and leads to significantly higher error in reconstructing observable.

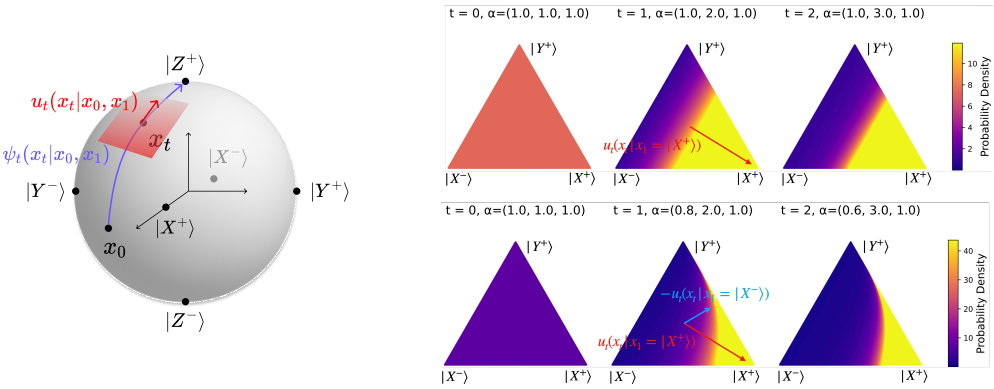

Figure 3: **Geometric flow matching for shadow generation. Left:** Spherical Flow transports continuous states along S^2 toward discrete POVM endpoints. **Right: Comparison between probability paths of different Anisotropic Dirichlet flows on Δ^2.** *Top:* $\gamma = 0$, which reduces to Dirichlet Flow. *Bottom:* $\gamma = 0.2$, which exhibits anisotropy in the probability path. This illustrates the evolution of the density over time, where densities are exponentiated for visual aid. Each vertex of the triangle represents distinct shadows. As time increases, the density becomes sharply concentrated near the $|X^+\rangle$ vertex, while simultaneously repelling from $|X^-\rangle$.

4.2.1 Riemannian Approach: Spherical Flow

Spherical Flow embeds the six Pauli outcomes as $\mathcal{V} = \{\pm e_1, \pm e_2, \pm e_3\} \subset S^2$. For $p, q \in S^2$ and $v \in T_p S^2$, the spherical maps are

$$\exp_p(v) = \cos(\|v\|)\, p + \sin(\|v\|)\, \frac{v}{\|v\|},$$
$$\log_p(q) = \frac{\cos^{-1}(\langle p, q\rangle)}{\sqrt{1 - \langle p, q\rangle^2}} \big(q - \langle p, q\rangle p\big). \tag{7}$$

The conditional path and velocity are $x_t = \exp_{x_0}(t \log_{x_0} v_k)$ and $u_t(x_t \mid v_k) = \log_{x_t}(v_k)/(1 - t)$. The base x_0 is Haar-uniform on S^2, i.e., rotation invariant and unbiased toward every axis and sign (Mardia & Jupp, 2000). We sample it by normalizing an isotropic Gaussian; stable coincident/antipodal limits and the basis-aligned coupling are in Definition B.1 and Section B.3.

We train the Hamiltonian-conditional endpoint classifier with

$$\mathcal{L}_{\mathrm{CE},S^2} = \mathbb{E}_{t,\, c,\, k \sim q(\cdot|c),\, x_0 \sim \kappa(\cdot|k)} \left[-\log \hat{p}_\theta(k \mid x_t, t, c) \right], \tag{8}$$

where $\kappa(\cdot \mid k)$ is a coupling with Haar-uniform marginal and c is the Hamiltonian condition. At inference, posterior-weighted geodesic velocities define an ODE on S^2; nearest-vertex decoding returns a discrete shadow token. Full maps, velocity, and decoding are in Section B.3.

4.2.2 PROBABILITY PATH APPROACH: ANISOTROPIC DIRICHLET FLOW

Dirichlet Flow. Dirichlet Flow (Stärk et al., 2024) prescribes the categorical path $p_t(x \mid e_i) = \mathrm{Dir}(x; \mathbf{1} + te_i)$ and trains an endpoint classifier with

$$\mathcal{L}_{\mathrm{CE,Dir}}(\theta) = \mathbb{E}_{t,x_1\sim q(\cdot),x\sim p_t(\cdot|x_1)}\left[-\log \hat{p}_\theta(x_1|x)\right]. \tag{9}$$

Anisotropic Dirichlet Flow. Our ADFM generalized Dirichlet Flow and modifies this path to encode the target–antitarget pairing. For target i, antipode $\bar{i}$, and hyperparameter $\gamma < 1$,

$$p_t^{\mathrm{AD}}(\mathbf{x} \mid x_1 = e_i) = \frac{1}{Z(t)} \prod_{j=1}^{K} x_j^{\alpha_j(t)-1}, \qquad \alpha_j(t) = \begin{cases} 1+t, & j = i, \\ 1-\gamma t, & j = \bar{i}, \\ 1, & j \in \{1,\dots,K\} \setminus \{i, \bar{i}\}. \end{cases} \tag{10}$$

so the path increases target concentration while decreasing its antipode. A continuity-equation solution has the form

$$u_t^{\mathrm{AD}}(x_t|x_1 = e_i) = \underbrace{C(x_i,t)(e_i - x_t)}_{\text{push toward } e_i} - \underbrace{\gamma D(x_{\bar{i}},t)(e_{\bar{i}} - x_t)}_{\text{pull away from } e_{\bar{i}}}, \tag{11}$$

The first term moves toward e_i and the second away from $e_{\bar{i}}$; $\gamma = 0$ recovers Dirichlet Flow. We use the endpoint loss in (9) with p_t^{AD}, combine conditional fields by the predicted endpoint posterior, solve the simplex ODE, and decode by $\arg\max$. The normalizer, C, D derivation, positivity constraint, and γ selection are in Sections D and F. This construction generalizes Dirichlet Flow by *pushing toward the target* and simultaneously *pulling away from the antitarget*.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP

For each Hamiltonian family, we use disjoint sets of 100/20/20 Hamiltonian instances for training, validation, and testing, with $M_{\mathrm{train}} = 1000$ measured shadows per training Hamiltonian. We denote by B the inference-time shadow budget, i.e., the number of generated shadows used to estimate downstream observables. Models are selected on the validation set at $B = 10$k and evaluated on the fixed test set at $B \in \{1\mathrm{k}, 10\mathrm{k}, 20\mathrm{k}\}$ with NFE$= 20$. Main results are averaged over three training seeds.

For an L-qubit system, our flow models generate all L site variables jointly and in parallel using a full-sequence Transformer with Hamiltonian conditioning. Self-attention models correlations across sites, while local Hamiltonian features encode the couplings incident on each site. Each site token is augmented with its position and incident couplings: (c_{i-1}, c_i, c_{i+1}) in 1D and the left, right, up, and down bonds in 2D, with missing boundary edges zero-padded. Full architecture, optimization, and evaluation details are provided in Section F.

5.2 LEARNING GROUND STATES OF ONE-DIMENSIONAL QUANTUM SYSTEMS

We first consider random-coupling 1D Heisenberg chains and inhomogeneous 1D Transverse Field Ising Models (TFIMs) with open boundary conditions and system sizes $L \in \{10, 30, 50, 100\}$. Refer to Section F for the definition of the Hamiltonians.

Random-coupling 1D Heisenberg chains. Table 1 shows that the geometric flows consistently improve over standard continuous and discrete flow baselines. Spherical Flow performs best at $L = 30$, while ADFM becomes increasingly effective as the system size grows, achieving the lowest prediction error at both $L = 50$ and $L = 100$. At $L = 100$ and $B = 10$k, ADFM reaches correlation/entropy RMSE of $0.086/0.141$, compared with $0.106/0.168$ for QuaDiM. This trend is consistent with Spherical Flow benefiting from denser training coverage at smaller system sizes, while the Dirichlet-based representation of ADFM becomes more competitive as the system size grows relative to the fixed training budget.

Figure 4 further visualizes the generative trajectory of correlation matrix on Heisenberg-30. ADFM recovers the spatial structure of the correlation matrix within the first few solver steps, whereas

Table 1: **Random-coupling 1D Heisenberg ground states.** Correlation and Entropy RMSE at $B = 1k, 10k, 20k$. Full means and standard deviations are in Table 8. Results directly from ground truth shadows is given only as a reference for the lower bound value.

METHOD	$L = 10$						$L = 30$						$L = 50$						$L = 100$					
	Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓		
	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$
Direct Shadows (lower bound)																								
CS	0.053	0.015	0.010	0.079	0.020	0.015	0.050	0.016	0.011	0.083	0.031	0.025	0.052	0.017	0.011	0.078	0.024	0.016	0.051	0.016	0.011	0.076	0.023	0.016
Classical ML																								
RBFK	0.087			0.151			0.140			0.222			0.167			0.249			0.202			0.316		
NTK	0.085			0.142			0.129			0.205			0.176			0.262			0.213			0.327		
Lasso	0.103			0.168			0.142			0.226			0.173			0.259			0.213			0.317		
Ridge	0.103			0.168			0.146			0.230			0.177			0.262			0.305			0.480		
Autoregressive																								
ShadowGPT	0.079	0.058	0.058	0.118	0.089	0.087	0.112	0.100	0.100	0.176	0.161	0.161	0.134	0.125	0.125	0.217	0.206	0.207	0.141	0.134	0.133	0.223	0.214	0.213
CFM / Diffusion																								
LinearFM	0.106	0.094	0.093	0.161	0.141	0.138	0.116	0.105	0.104	0.181	0.164	0.163	0.146	0.139	0.138	0.236	0.220	0.220	0.147	0.139	0.139	0.234	0.220	0.219
QuaDiM	0.101	0.083	0.083	0.144	0.114	0.113	0.110	0.100	0.098	0.168	0.150	0.148	0.122	0.113	0.113	0.197	0.181	0.180	0.117	0.106	0.105	0.187	0.168	0.167
CS-DFM																								
StatisticalFM	0.104	0.096	0.095	0.158	0.143	0.141	0.118	0.107	0.107	0.180	0.163	0.163	0.123	0.114	0.113	0.201	0.184	0.184	0.132	0.123	0.123	0.215	0.201	0.200
DirichletFM	0.076	0.066	0.063	0.117	0.101	0.096	0.109	0.095	0.095	0.170	0.150	0.150	0.179	0.173	0.172	0.265	0.254	0.252	0.102	0.090	0.089	0.163	0.145	0.145
Ours (AD)	0.072	0.055	0.054	0.109	0.083	0.081	0.108	0.095	0.096	0.167	0.150	0.149	**0.113**	**0.103**	**0.102**	**0.179**	**0.163**	**0.162**	**0.098**	**0.086**	**0.085**	**0.157**	**0.141**	**0.140**
Ours (Spherical)	**0.069**	**0.052**	**0.053**	**0.106**	**0.079**	**0.078**	**0.099**	**0.085**	**0.085**	**0.155**	**0.135**	**0.134**	0.134	0.125	0.124	0.214	0.199	0.198	0.123	0.112	0.111	0.193	0.177	0.176

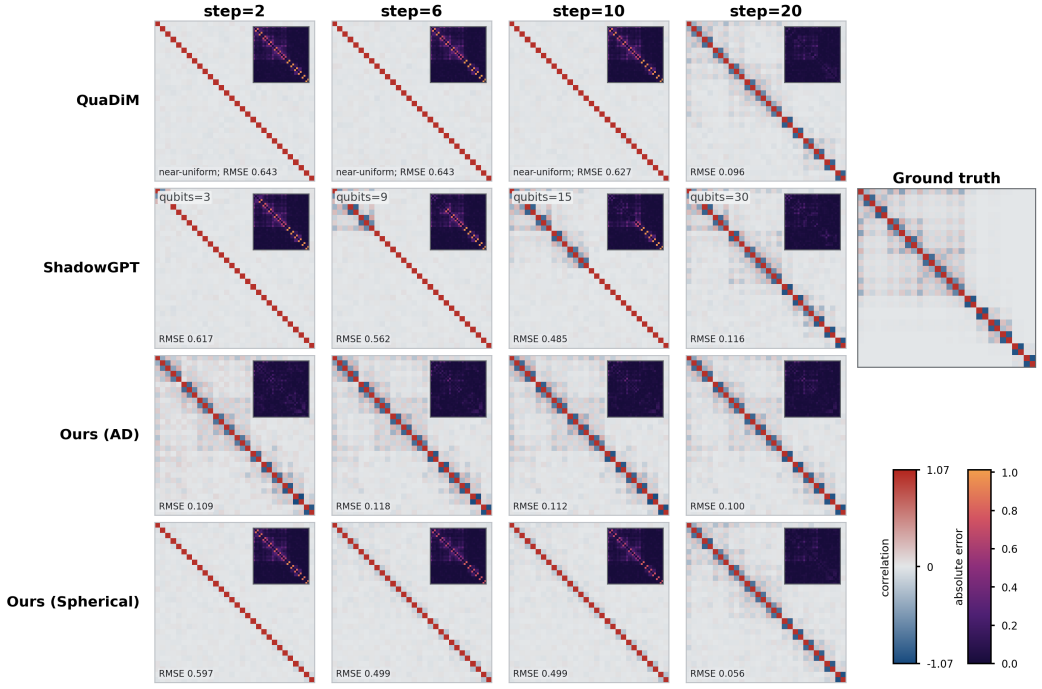

Figure 4: **Generative trajectories of correlation matrix of unseen quantum state. 1D Heisenberg-30.** Flow columns show solver steps $2, 6, 10, 20$, while ShadowGPT columns show prefixes of $3, 9, 15, 30$ generated qubits. Insets show absolute error with respect to the ground-truth correlation matrix. AD Flow is able to predict the correlation structure of unseen quantum state at early solver steps, while Spherical Flow achieves the lowest final prediction error.

Spherical Flow converges more gradually but ultimately yields the most accurate reconstruction. In contrast, autoregressive generation reveals the correlation structure sequentially as the shadow sequence is generated.

Inhomogeneous 1D TFIM. We next consider an inhomogeneous TFIM in which both bond couplings and local fields vary across sites. Table 2 shows a clear transition across system sizes: Spherical Flow performs best among the generative models at $L = 10$ and 30, while ADFM performs best at $L = 50$ and 100, outperforming the autoregressive ShadowGPT at the larger system sizes. Similarly, Spherical Flow is strong at small system while AD Flow becomes stronger at large system.

Table 2: **Inhomogeneous 1D TFIM ground states.** Correlation and Entropy RMSE at $B = 1k, 10k, 20k$; full means and standard deviations are in Table 7.

METHOD	$L = 10$						$L = 30$						$L = 50$						$L = 100$					
	Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓		
	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$
Direct Shadows (lower bound)																								
CS	0.052	0.015	0.011	0.071	0.020	0.014	0.053	0.017	0.012	0.076	0.025	0.017	0.054	0.018	0.012	0.075	0.024	0.016	0.055	0.017	0.012	0.076	0.022	0.016
Classical ML																								
RBFK	0.019			0.043			0.043			0.069			0.034			0.073			0.027			0.070		
NTK	0.018			0.038			0.016			0.077			0.021			0.075			0.013			0.074		
Lasso	0.011			0.062			0.018			0.170			0.009			0.122			0.009			0.120		
Ridge	0.023			0.066			0.045			0.194			0.065			0.245			0.081			0.274		
Autoregressive																								
ShadowGPT	0.056	0.021	0.018	0.075	0.038	0.034	0.057	0.023	0.020	0.102	0.063	0.061	0.056	0.021	0.018	0.108	0.067	0.064	0.063	0.036	0.032	0.166	0.142	0.143
CFM / Diffusion																								
LinearFM	0.058	0.029	0.027	0.100	0.056	0.055	0.061	0.032	0.029	0.157	0.130	0.127	0.057	0.026	0.023	0.107	0.075	0.073	0.062	0.036	0.034	0.179	0.159	0.158
QuaDiM	0.062	0.033	0.031	0.115	0.088	0.086	0.068	0.046	0.044	0.189	0.170	0.168	0.062	0.034	0.032	0.132	0.108	0.107	0.060	0.034	0.031	0.158	0.136	0.134
CS-DFM																								
StatisticalFM	0.057	0.025	0.022	0.096	0.055	0.052	0.065	0.038	0.035	0.147	0.122	0.122	0.062	0.035	0.032	0.149	0.120	0.118	0.063	0.038	0.036	0.160	0.135	0.133
DirichletFM	0.058	0.027	0.025	0.095	0.064	0.061	0.062	0.030	0.028	0.115	0.086	0.084	0.062	0.037	0.035	0.113	0.091	0.089	0.061	0.033	0.031	0.159	0.139	0.138
Ours (AD)	0.057	0.025	0.022	0.089	0.051	0.048	0.062	0.032	0.030	0.117	0.087	0.085	0.057	0.028	0.026	0.120	0.093	0.090	**0.059**	**0.027**	**0.024**	**0.118**	**0.090**	**0.088**
Ours (Spherical)	**0.053**	0.022	0.019	0.078	0.038	0.034	0.058	0.026	0.023	**0.092**	**0.054**	**0.051**	**0.055**	**0.024**	**0.020**	**0.095**	**0.055**	**0.051**	0.062	0.034	0.032	0.135	0.112	0.110

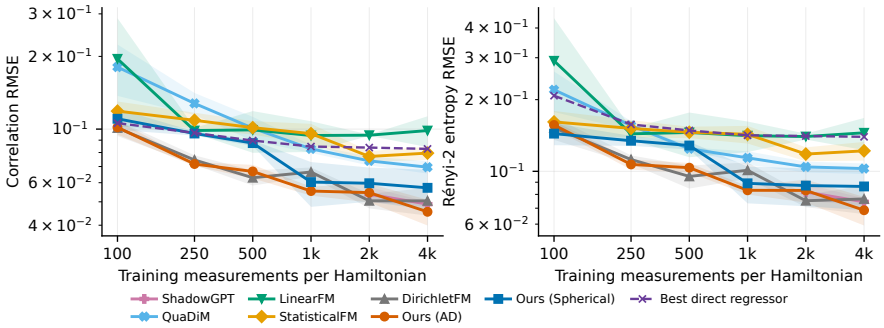

Figure 5: **Effect of training-set size on 1D Heisenberg-**10**.** Correlation and Rényi-2 entropy RMSE are reported as the number of training shadows per Hamiltonian, M_{train}, varies from 100 to 4000.

Effect of Training-Set Size. We study the effect of the number of training shadows per Hamiltonian, M_{train}, on 1D Heisenberg-10. As shown in Figure 5, prediction errors generally decrease as more training shadows are provided, while the relative performance of the generative models varies with training-set size. ADFM achieves the lowest errors at several intermediate and larger training-set sizes, whereas QuaDiM improves substantially as M_{train} increases but is less competitive in the low-data regime. Overall, the geometric flow variants maintain strong performance across a broad range of training-set sizes.

5.3 Learning Ground States of Two-Dimensional Systems

We next extend the benchmark to two-dimensional systems, where ground-state preparation becomes substantially more challenging with increasing lattice size. We consider OBC Heisenberg models on 4×4, 5×5, and 6×6 lattices. Table 3 reports observable prediction from generated shadows across the three lattice sizes. At 6×6 and $B = 10$k, Spherical Flow achieves the lowest correlation RMSE (0.076), while ADFM achieves the lowest entropy RMSE (0.108), improving over QuaDiM (0.083/0.141) and ShadowGPT (0.092/0.132). These results show that the gains of geometry-aware flow matching extend from 1D chains to larger 2D lattices. This trend is again consistent with Spherical Flow benefiting from denser training coverage at smaller system sizes (4×4), while the Dirichlet-based flows, particularly ADFM, become more competitive as the system size grows (6×6).

5.4 Extension to Non-Pauli POVMs

We further evaluate ShadowFlow beyond Pauli-6 using the tetrahedral SIC-POVM with four equiangular outcomes and the Cube-8 IC-POVM with eight outcomes forming four antipodal pairs. ADFM uses uniform repulsion over the three non-target outcomes for the tetrahedral POVM, while Cube-8

Table 3: **2D OBC Heisenberg.** Correlation and Entropy RMSE at $B = 1k, 10k, 20k$. Standard deviation is given in Appendix Table 9.

METHOD	4×4						5×5						6×6					
	Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓		
	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$
Direct Shadows (lower bound)																		
CS	0.097	0.030	0.021	0.080	0.025	0.017	0.094	0.029	0.020	0.082	0.025	0.017	0.094	0.030	0.021	0.085	0.025	0.017
Classical ML																		
RBFK	0.087			0.138			0.109			0.151			0.124			0.186		
NTK	0.087			0.136			0.108			0.149			0.124			0.184		
Lasso	0.086			0.132			0.109			0.145			0.137			0.188		
Ridge	0.087			0.132			0.114			0.148			0.152			0.197		
Autoregressive																		
ShadowGPT	0.125	0.088	0.084	0.150	0.125	0.124	0.135	0.100	0.098	0.174	0.154	0.152	0.129	0.092	0.090	0.158	0.132	0.131
CFM / Diffusion																		
LinearFM	0.130	0.094	0.094	0.187	0.167	0.166	0.127	0.087	0.084	0.182	0.164	0.163	0.129	0.093	0.090	0.178	0.157	0.156
QuaDiM	0.124	0.085	0.081	0.164	0.141	0.140	0.124	0.085	0.083	0.170	0.149	0.146	0.126	0.083	0.081	0.165	0.141	0.139
CS-DFM																		
StatisticalFM	0.122	0.080	0.079	0.168	0.147	0.147	0.126	0.092	0.089	0.169	0.152	0.151	0.125	0.085	0.082	0.168	0.145	0.142
DirichletFM	0.117	0.080	0.077	0.140	0.115	0.114	0.121	0.081	0.078	**0.139**	**0.111**	**0.108**	0.116	0.078	0.075	0.136	0.109	0.107
Ours (AD)	0.116	0.076	0.073	0.137	0.108	0.107	**0.120**	**0.078**	**0.075**	0.141	0.112	0.109	0.117	0.077	**0.073**	0.136	**0.108**	**0.106**
Ours (Spherical)	**0.111**	**0.070**	**0.067**	**0.127**	**0.100**	**0.099**	0.121	0.080	0.078	0.147	0.121	0.119	**0.116**	**0.076**	0.073	**0.134**	0.109	0.108

Table 4: **Cube-8, 1D Heisenberg-**10**.**

METHOD	Corr. ↓			Ent. ↓		
	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$
Direct Shadows (lower bound)						
CS	0.051	0.016	0.011	0.073	0.023	0.016
Classical ML						
RBFK	0.088			0.151		
NTK	0.083			0.141		
Lasso	0.101			0.167		
Ridge	0.101			0.167		
Autoregressive						
ShadowGPT	0.085	0.070	0.068	0.128	0.108	0.104
CFM / Diffusion						
LinearFM	0.119	0.110	0.112	0.182	0.165	0.168
QuaDiM	0.108	0.095	0.095	0.161	0.139	0.139
CS-DFM						
StatisticalFM	0.110	0.099	0.100	0.164	0.144	0.146
DirichletFM	0.076	0.061	0.061	0.116	0.092	0.092
Ours (AD)	0.080	0.060	0.060	0.121	0.092	0.093
Ours (Spherical)	**0.074**	**0.054**	**0.050**	**0.110**	**0.080**	**0.075**

Table 5: **Tetrahedral, 1D Heisenberg-**10**.**

METHOD	Corr. ↓			Ent. ↓		
	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$
Direct Shadows (lower bound)						
CS	0.039	0.012	0.009	0.051	0.016	0.011
Classical ML						
RBFK	0.094			0.159		
NTK	0.097			0.159		
Lasso	0.124			0.199		
Ridge	0.124			0.198		
Autoregressive						
ShadowGPT	0.064	0.053	0.052	0.097	0.081	0.078
CFM / Diffusion						
LinearFM	0.096	0.091	0.089	0.142	0.131	0.130
QuaDiM	0.088	0.077	0.077	0.119	0.100	0.100
CS-DFM						
StatisticalFM	0.087	0.077	0.076	0.129	0.111	0.110
DirichletFM	0.066	0.055	0.054	0.095	0.082	0.080
Ours (AD)	**0.064**	**0.053**	**0.051**	**0.092**	**0.078**	**0.076**
Ours (Spherical)	0.072	0.062	0.061	0.105	0.088	0.087

admits direct target–antitarget pairing. Tables 4 and 5 report results on 1D Heisenberg-10. At $B = 10$k, Spherical Flow performs best on Cube-8 ($0.054/0.080$ correlation/entropy RMSE), while ADFM performs best overall on the tetrahedral POVM ($0.053/0.078$). This complementary behavior is consistent with Spherical Flow benefiting from the denser angular geometry of Cube-8 and ADFM performing well on the sparse four-outcome tetrahedral representation.

6 LIMITATION AND OUTLOOK

ShadowFlow learns a reusable distribution of classical shadows, but this is a more demanding task than directly regressing a single prespecified observable, for which specialized regressors can be more accurate and computationally efficient. Its performance also depends on the number of available measurements and on the learnability of the underlying family of quantum states. Extending the framework to more challenging and physically intriguing state families remains an important direction for future work.

AI USE STATEMENT

Generative AI tools were used to assist with proof verification, experimental implementation and analysis, and manuscript writing and editing. The authors reviewed AI-assisted work and take responsibility for the content of this work, including its text, claims, code, and artifacts.

REPRODUCIBILITY STATEMENT

The appendices provide details on data generation, model architectures, hyperparameters, checkpoint selection, and evaluation protocols. Source code, experiment configurations, and the data required to reproduce the reported results will be released upon acceptance.

REFERENCES

Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*, 2023.

F. Bloch. Nuclear induction. *Phys. Rev.*, 70:460–474, Oct 1946. doi: 10.1103/PhysRev.70.460. URL https://link.aps.org/doi/10.1103/PhysRev.70.460.

Juan Carrasquilla, Giacomo Torlai, Roger G Melko, and Leandro Aolita. Reconstructing quantum states with generative models. *Nature Machine Intelligence*, 1(3):155–161, 2019.

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

Chaoran Cheng, Jiahan Li, Jian Peng, and Ge Liu. Categorical flow matching on statistical manifolds. *arXiv preprint arXiv:2405.16441*, 2024.

Chaoran Cheng, Jiahan Li, Jiajun Fan, and Ge Liu. α-flow: A unified framework for continuous-state discrete flow matching models. *arXiv preprint arXiv:2504.10283*, 2025.

Kyle Cranmer, Siavash Golkar, and Duccio Pappadopulo. Inferring the quantum density matrix with machine learning. *arXiv preprint arXiv:1904.05903*, 2019.

Oscar Davis, Samuel Kessler, Mircea Petrache, Ismail Ceylan, Michael Bronstein, and Joey Bose. Fisher flow matching for generative modeling over discrete data. *Advances in Neural Information Processing Systems*, 37:139054–139084, 2024.

Richard P. Feynman, Jr. Vernon, Frank L., and Robert W. Hellwarth. Geometrical representation of the schrödinger equation for solving maser problems. *Journal of Applied Physics*, 28(1):49–52, 01 1957. ISSN 0021-8979. doi: 10.1063/1.1722572. URL https://doi.org/10.1063/1.1722572.

Itai Gat, Tal Remez, Neta Shaul, Felix Kreuk, Ricky TQ Chen, Gabriel Synnaeve, Yossi Adi, and Yaron Lipman. Discrete flow matching. *Advances in Neural Information Processing Systems*, 37: 133345–133385, 2024.

Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. Measurements of quantum hamiltonians with locally-biased classical shadows. *Communications in Mathematical Physics*, 391(3):951–981, 2022.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V Albert, and John Preskill. Provably efficient machine learning for quantum many-body problems. *Science*, 377(6613):eabk3333, 2022.

Abhijith Jayakumar, Marc Vuffray, and Andrey Y Lokhov. Learning energy-based representations of quantum many-body states. *Physical Review Research*, 6(3):033201, 2024.

Dax Enshan Koh and Sabee Grewal. Classical shadows with noise. *Quantum*, 6:776, 2022.

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

Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky TQ Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code. *arXiv preprint arXiv:2412.06264*, 2024.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

Kanti V. Mardia and Peter E. Jupp. *Directional Statistics*. Wiley, Chichester, 2000.

Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

Ulrich Schollwöck. The density-matrix renormalization group. *Reviews of modern physics*, 77(1): 259–315, 2005.

Hannes Stärk, Bowen Jing, Chenyu Wang, Gabriele Corso, Bonnie Berger, Regina Barzilay, and Tommi Jaakkola. Dirichlet flow matching with applications to DNA sequence design. *arXiv preprint arXiv:2402.05841*, 2024.

Yehui Tang, Mabiao Long, and Junchi Yan. Quadim: A conditional diffusion model for quantum state property estimation. In *The Thirteenth International Conference on Learning Representations*, 2025.

Jian Yao and Yi-Zhuang You. Shadowgpt: Learning to solve quantum many-body problems from randomized measurements. *arXiv preprint arXiv:2411.03285*, 2024.

Yusheng Zhao, Chi Zhang, and Yuxuan Du. Rethink the role of deep learning towards large-scale quantum systems. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, 2025.

A CLASSICAL SHADOW REPRESENTATION

Algorithm 1 Measuring Shadows

Input: B copies of unknown n-qubit state ρ, $\mathcal{U} = \{X, Y, Z\}$
Output: List of shadows $\mathcal{S}$

```
Initialize empty list S
for i = 1 to B do
    for k = 1 to n do
        Sample l_i^(k) ~ Cat(1/3, 1/3, 1/3)
        U_i^(k) ← U[l_i^(k)]
    end for
    U_i ← U_i^(1) ⊗ ··· ⊗ U_i^(n)
    l_i ← (l_i^(1), ..., l_i^(n))
    Measure U_i ρ U_i^†, obtain b_i ∈ {0,1}^n
    Append (l_i, b_i) to S
end forS
```

Algorithm 2 Observable Estimation

Input: List of shadows $\mathcal{S} = \{(l_i, b_i)\}_{i=1}^{B}$ of a n-qubit state, observable $O, \mathcal{U} = \{X, Y, Z\}$
Output: Estimation of the expectation of the observable $\langle \hat{O} \rangle$

```
for i = 1 to B do
    for k = 1 to n do
        U_i^(k) ← U[l_i^(k)]
    end for
    U_i ← U_i^(1) ⊗ ··· ⊗ U_i^(n)
    ρ̂_i ← 3 U_i^† |b_i⟩⟨b_i| U_i − I
end for
ρ̃ ← (1/B) Σ_{i=1}^{B} ρ̂_i
⟨Ô⟩ ← Tr(O ρ̃) ⟨Ô⟩
```

Protocol. Let $b_i^{(k)} \in \{0, 1\}$ denote the measurement outcome on the ith qubit, and $U_i^{(k)}$ the unitary mapping that maps the chosen Pauli basis to the computational basis. Then, we can define the local estimator for the ith qubit: $\hat{\rho}_i^{(k)} = 3\, U_i^{(k)\dagger} |b_i^{(k)}\rangle\langle b_i^{(k)}| U_i^{(k)} - \mathbb{I}$. The global estimator can then be constructed by taking the tensor product of the local estimators: $\hat{\rho}^{(k)} = \bigotimes_{i=1}^{N} \hat{\rho}_i^{(k)}$. This shadow representation significantly reduces the computational overhead while still providing unbiased estimators for local observables. With M measurements of the shadow, one can reconstruct the density matrix as $\hat{\rho} = \frac{1}{M} \sum_{k=1}^{M} \hat{\rho}^{(k)}$, and estimate the expectation value of observable as $\langle \hat{O} \rangle = \frac{1}{M} \sum_{k=1}^{M} \mathrm{Tr}(O \hat{\rho}^{(k)})$.

Efficiency of Shadows. A key advantage of classical shadows is their favorably efficient sample complexity. Suppose O is an observable with Pauli rank r (i.e., the number of non-zero coefficients when decomposed into Pauli strings). Then, to estimate $\langle O \rangle$ within additive error ϵ and failure probability δ, it suffices to use $M = \mathcal{O}(r \log(1/\delta)/\epsilon^2)$ shadow measurements Huang et al. (2020). For example, pairwise correlations that we are using as an evaluation metric in this work have low Pauli rank and are thus efficiently able to be estimated.

Example of Estimating Local Observables. As an example, consider estimating the expectation of two-qubit observable $O = Z_1 Z_2$ from classical shadows. For each k-th shot, we measure all qubits independently in random Pauli bases. We only use shadows where qubits 1 and 2 are both measured in the Z basis, i.e., $U_1^{(k)} = U_2^{(k)} = \mathbb{I}$. Given the measurement outcomes $b_1^{(k)}, b_2^{(k)} \in \{0, 1\}$ on those qubits, we compute:

$$\hat{O}^{(k)} \equiv \mathrm{Tr}(O \hat{\rho}^{(k)}) = \mathrm{Tr}(Z_1 Z_2 \hat{\rho}_1^{(k)} \otimes \hat{\rho}_2^{(k)}) = 3(-1)^{b_1^{(k)}} \cdot 3(-1)^{b_1^{(k)}} = 9(-1)^{b_1^{(k)} + b_2^{(k)}} \tag{12}$$

as an unbiased single-shot estimator for $\langle Z_1 Z_2 \rangle$. This factor of 9 arises from the definition of single-qubit inverse map used in classical shadows: $\hat{\rho}_i^{(k)} = 3|b_i^{(k)}\rangle\langle b_i^{(k)}| - \mathbb{I}$, which yields $\mathrm{Tr}(Z_i \hat{\rho}_i^{(k)}) = 3(-1)^{b_i^{(k)}}$.

The final estimate is given by averaging the single shot estimator $\hat{O}^{(k)}$ over the M' shots that satisfy the basis condition: $\langle O \rangle = \mathbb{E}[\mathrm{Tr}(O \hat{\rho}^{(k')})] \approx \frac{1}{M'} \sum_{k'=1}^{M'} \hat{O}^{(k')}$, k' denotes the indices of snapshots where qubits 1 and 2 are both measured in the Z basis.

B Additional Spherical Flow Background

B.1 Bloch-Sphere Geometry of Single-Qubit Outcomes

A pure qubit state is a ray in $\mathbb{C}^2$: $|\psi\rangle$ and $e^{i\gamma}|\psi\rangle$ represent the same physical state. The resulting projective space is $\mathbb{CP}^1$. With the parametrization

$$|\psi(\theta,\phi)\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle,$$

the Fubini–Study line element is

$$ds^2_{\mathrm{FS}} = \langle d\psi|d\psi\rangle - |\langle\psi|d\psi\rangle|^2 = \tfrac{1}{4}\big(d\theta^2 + \sin^2\theta\, d\phi^2\big).$$

The Bloch map $[\psi] \mapsto \langle\psi|\boldsymbol{\sigma}|\psi\rangle$ is a smooth bijection from $\mathbb{CP}^1$ to S^2. Since $ds^2_{S^2} = 4ds^2_{\mathrm{FS}}$, the two metrics agree up to a constant factor. This justifies representing rank-one qubit POVM outcomes as points on the sphere and using spherical geodesics in our transport path.

B.2 Haar-Uniform Spherical Prior

Definition B.1 (Haar-uniform spherical prior)**.** Let μ_{H} denote the normalized rotation-invariant measure on S^2,

$$d\mu_{\mathrm{H}}(x) = \frac{dA(x)}{4\pi}, \qquad \mu_{\mathrm{H}}(RA) = \mu_{\mathrm{H}}(A) \tag{13}$$

for every measurable $A \subseteq S^2$ and $R \in \mathrm{SO}(3)$. Equivalently, μ_{H} is the probability measure on $S^2 \simeq \mathrm{SO}(3)/\mathrm{SO}(2)$ induced by Haar measure on $\mathrm{SO}(3)$ (Mardia & Jupp, 2000).

We sample this prior by drawing $g \sim \mathcal{N}(0, I_3)$ and setting

$$x_0 = \frac{g}{\|g\|_2}. \tag{14}$$

Equivalently, $(x_{0,1}^2, x_{0,2}^2, x_{0,3}^2) \sim \mathrm{Dirichlet}(\frac{1}{2},\frac{1}{2},\frac{1}{2})$, with independent uniform coordinate signs. For an L-site shadow sequence, the base distribution is $\mu_{\mathrm{H}}^{\otimes L}$.

Our earlier construction instead sampled $w \sim \mathrm{Dir}(1,1,1)$ and set $x_0 = s(\sqrt{w_1}, \sqrt{w_2}, \sqrt{w_3})$ with a shared $s \in \{\pm 1\}$. Its support is restricted to two opposite octants and is therefore not rotation invariant. We replace it with μ_{H} to remove this coordinate preference.

For the optional basis-aligned coupling, we permute the coordinates of a Haar-uniform draw so that a prescribed absolute-coordinate rank is assigned to a uniformly sampled measurement axis. Averaging over the uniformly sampled axis preserves the marginal μ_{H} by signed-permutation symmetry. The measurement outcome sign is not used in this coupling.

B.3 Spherical-Flow Maps and Decoding

For $p, q \in S^{d-1}$ and $v \in T_pS^{d-1}$, the sphere exponential and logarithmic maps are

$$\exp_p(v) = \cos(\|v\|)p + \sin(\|v\|)\frac{v}{\|v\|}, \tag{15}$$

$$\log_p(q) = \frac{\cos^{-1}(\langle p,q\rangle)}{\sqrt{1-\langle p,q\rangle^2}}\big(q - \langle p,q\rangle p\big). \tag{16}$$

We use numerically stable limiting branches near coincident and antipodal points. For endpoint v_k, the conditional velocity is $u_t(x_t \mid v_k) = \log_{x_t}(v_k)/(1-t)$ and the learned marginal field is $\hat{v}_\theta(x_t, t \mid c) = \sum_{k=1}^K u_t(x_t \mid v_k)\hat{p}_\theta(k \mid x_t, t, c)$. We integrate this field independently for every generated record and decode the terminal point by $\arg\max_k\langle x_1, v_k\rangle$. Independent base draws followed by deterministic integration preserve conditional independence across generated records, as required by the U-statistic entropy estimator.

C Proof of Theorem C.2

Definition C.1 (Spin and basis error channels)**.** A Pauli-6 token is a pair (a, s), where $a \in \{X, Y, Z\}$ denotes the measurement axis and $s \in \{\pm 1\}$ the measured eigenvalue. Under $\mathcal{E}_{\text{spin}}(p)$, the sign is flipped with probability p while the axis is unchanged. Under $\mathcal{E}_{\text{basis}}(p)$, with probability p the axis is replaced uniformly by one of the other two axes and the sign is resampled uniformly; otherwise the token is unchanged. The channels act independently across sites and measurement shots.

Theorem C.2. *Let*

$$P = \bigotimes_{j \in S} \sigma_{q_j}, \qquad w = |S|, \qquad \mu = \operatorname{tr}(P\rho),$$

be a Pauli observable of weight w. Let $\hat{\mu}_{\text{spin}}$ and $\hat{\mu}_{\text{basis}}$ denote the standard Pauli classical-shadow estimators constructed from M shots corrupted independently by $\mathcal{E}_{\text{spin}}(p)$ and $\mathcal{E}_{\text{basis}}(p)$, respectively. Then

$$\mathbb{E}[\hat{\mu}_{\text{spin}}] = (1-2p)^w \mu, \qquad \mathbb{E}[\hat{\mu}_{\text{basis}}] = (1-p)^w \mu. \tag{17}$$

Consequently, for $0 \le p \le \frac{1}{2}$,

$$|\mathbb{E}[\hat{\mu}_{\text{spin}}] - \mu| \ge |\mathbb{E}[\hat{\mu}_{\text{basis}}] - \mu|\,. \tag{18}$$

Moreover, for $0 < p \le \frac{1}{2}$ and $\mu \ne 0$, define

$$\Delta_{p,w,\mu} = \mu^2 \left([1-(1-2p)^w]^2 - [1-(1-p)^w]^2 \right) > 0. \tag{19}$$

A sufficient condition for $\text{MSE}(\hat{\mu}_{\text{spin}}) > \text{MSE}(\hat{\mu}_{\text{basis}})$ is

$$M > \frac{3^{2w}}{\Delta_{p,w,\mu}}. \tag{20}$$

Proof. For a single site and target Pauli σ_q, a Pauli-shadow snapshot satisfies

$$\operatorname{tr}(\sigma_q \hat{\rho}) = 3s\, \mathbf{1}\{a = q\}, \qquad \hat{\rho} = 3U^\dagger |b\rangle\langle b| U - I, \tag{21}$$

where (a, s) is the observed Pauli-6 token. Thus the single-shot estimator of P is

$$X = \prod_{j \in S} 3 s_j\, \mathbf{1}\{a_j = q_j\}, \qquad |X| \le 3^w, \qquad \mathbb{E}[X] = \mu. \tag{22}$$

Under $\mathcal{E}_{\text{spin}}(p)$, each sign is multiplied by an independent random variable $\eta_j \in \{\pm 1\}$ satisfying $\mathbb{E}[\eta_j] = 1 - 2p$. Therefore,

$$\mathbb{E}[X_{\text{spin}}] = \mathbb{E}\left[X \prod_{j \in S} \eta_j \right] = (1-2p)^w \mu. \tag{23}$$

Under $\mathcal{E}_{\text{basis}}(p)$, the original local contribution is retained with probability $1 - p$. If the axis is replaced, the new sign is uniform and hence the corrupted local contribution has conditional mean zero. Independence across sites therefore gives

$$\mathbb{E}[X_{\text{basis}}] = (1-p)^w \mu. \tag{24}$$

Averaging over M independent shots proves Equation (17).

For $0 \le p \le \frac{1}{2}$,

$$0 \le (1-2p)^w \le (1-p)^w \le 1.$$

Hence

$$1 - (1-2p)^w \ge 1 - (1-p)^w \ge 0,$$

which immediately yields Equation (18).

Finally, for either error channel $\mathcal{E}$,

$$\text{MSE}(\hat{\mu}_{\mathcal{E}}) = \left(\mathbb{E}[\hat{\mu}_{\mathcal{E}}] - \mu \right)^2 + \frac{\text{Var}(X_{\mathcal{E}})}{M}. \tag{25}$$

Since $|X_{\mathcal{E}}| \leq 3^w$,

$$0 \leq \mathrm{Var}(X_{\mathcal{E}}) \leq 3^{2w}. \tag{26}$$

The squared-bias difference between the two channels is $\Delta_{p,w,\mu}$. Therefore,

$$\begin{aligned} &\mathrm{MSE}(\hat{\mu}_{\mathrm{spin}}) - \mathrm{MSE}(\hat{\mu}_{\mathrm{basis}}) \\ &= \Delta_{p,w,\mu} + \frac{\mathrm{Var}(X_{\mathrm{spin}}) - \mathrm{Var}(X_{\mathrm{basis}})}{M} \\ &\geq \Delta_{p,w,\mu} - \frac{3^{2w}}{M}. \end{aligned} \tag{27}$$

The right-hand side is positive whenever $M > 3^{2w}/\Delta_{p,w,\mu}$, proving the finite-shot claim. □

D Derivation of the Conditional Velocity Field of the Anisotropic Dirichlet Flow

We derive the time-dependent conditional velocity field $u_t : \Delta^{K-1} \to T\Delta^{K-1}$ satisfying $\partial_t p_t + \nabla \cdot (p_t u_t) = 0$ for the anisotropic probability path below. Here $t \in [0, T]$ is concentration time, $T = \alpha_{\max} - 1$, and $\gamma T < 1$ ensures positive concentrations:

$$p_t(x|x_1 = e_i) = \frac{1}{Z(t)} \prod_{j=1}^{K} x_j^{\alpha_j(t)-1}, \quad \alpha_i(t) = 1+t,\ \alpha_{\bar{i}}(t) = 1-\gamma t,\ \alpha_j(t) = 1\ (j \notin \{i, \bar{i}\}), \tag{28}$$

where,

$$Z(t) = \frac{\prod_{j=1}^{K} \Gamma(\alpha_j(t))}{\Gamma(\alpha_\Sigma(t))}, \quad \alpha_\Sigma(t) = \sum_{j=1}^{K} \alpha_j(t) = K + (1-\gamma)t. \tag{29}$$

Note that $\bar{i}$ denotes the conjugate partner of i. This indicates the conjugate relation of the shadow measurement (i.e. $\sigma^+ \leftrightarrow \sigma^-$, $\sigma \in \{X, Y, Z\}$). In our implementation, we set $\{(i, \bar{i})\} \in \{(0,1), (2,3), (4,5)\}$, which represents 3 conjugate pairs corresponding to each Pauli matrix X, Y, Z.

Eq. 28 is a generalized Dirichlet probability path which we call *cross polytope probability path*. We incorporate anisotropy to reflect the geometry motivated from the Bloch sphere representation of shadows. Solving the continuity equation for this conditional probability path will give us the conditional velocity field u_t such that generates the given p_t.

Solving the Continuity Equation. Only $\alpha_i, \alpha_{\bar{i}}$ depend on t, so

$$\partial_t \log p_t = \dot{\alpha}_i \log x_i + \dot{\alpha}_{\bar{i}} \log x_{\bar{i}} - \partial_t \log Z(t) = \log x_i - \gamma \log x_{\bar{i}} - \big[\mathbb{E}_{p_t}[\log X_i] - \gamma \mathbb{E}_{p_t}[\log X_{\bar{i}}]\big], \tag{30}$$

hence

$$\partial_t p_t = p_t \big[(\log x_i - \mathbb{E}_{p_t}[\log X_i]) - \gamma(\log x_{\bar{i}} - \mathbb{E}_{p_t}[\log X_{\bar{i}}])\big]. \tag{31}$$

We seek a velocity field that solves the continuity equation Eq. 31, that acts only on coordinates $i, \bar{i}$, tangent to the simplex, and vanishing on the boundary. Motivated from Stärk et al. (2024), we posit the following ansatz for the velocity field:

$$u_t(x|x_1 = e_i) = C(x_i, t)\,(e_i - x) - \gamma D(x_{\bar{i}}, t)\,(e_{\bar{i}} - x). \tag{32}$$

If we set $F_j(x) = f_j(x_j)\,(e_j - x)$, one finds

$$\nabla \cdot (p_t F_j) = p_t \Big[-(\alpha_\Sigma - 1)\, f_j(x_j) + \frac{\alpha_j - 1}{x_j}\, f_j(x_j) - (x_j - 1)\, f_j'(x_j)\Big], \quad \alpha_\Sigma = \sum_{k=1}^{K} \alpha_k(t). \tag{33}$$

Dividing by p_t gives

$$\frac{\nabla \cdot (p_t F_j)}{p_t} = -(\alpha_\Sigma - 1)\, f_j(x_j) + \frac{\alpha_j - 1}{x_j}\, f_j(x_j) - (x_j - 1)\, f_j'(x_j). \tag{34}$$

Matching this to $-\partial_t \log p_t$ for $j = i, \bar{i}$ yields

$$(x_j - 1)\, f_j'(x_j) + \Big[\alpha_\Sigma - 1 - \frac{\alpha_j - 1}{x_j}\Big] f_j(x_j) = \log x_j - \mathbb{E}_{p_t}[\log X_j]. \tag{35}$$

This is an ODE problem for solving $f_j(x_j)$.

Solving the ODE problem. Write the ODE as $f' + P(x_j)f = Q(x_j)$ by dividing by $x_j - 1$. Then,

$$P(x_j) = \frac{\alpha_\Sigma - 1 - \frac{\alpha_j - 1}{x_j}}{x_j - 1} = -\frac{\alpha_\Sigma - 1}{1 - x_j} + \frac{\alpha_j - 1}{x_j(1 - x_j)}. \tag{36}$$

Defining the integrating factor as,

$$\mu_j(x_j) \equiv \exp\Big(\int P(x_j)\, dx_j\Big) = x_j^{\alpha_j - 1}\, (1 - x_j)^{\,\alpha_\Sigma - \alpha_j}, \tag{37}$$

we can derive the unique solution vanishing at $x_j = 0$:

$$f_j(x_j) = -\frac{1}{\mu_j(x_j)} \int_0^{x_j} \mu_j(s)\, \frac{\log s - \mathbb{E}_{p_t}[\log X_j]}{1 - s}\, ds. \tag{38}$$

Identify $C(x_i, t) = f_i(x_i),\ D(x_{\bar{i}}, t) = f_{\bar{i}}(x_{\bar{i}})$. Note that the expectation value of $\log X_j$ with respect to p_t can be calculated as $\mathbb{E}_{p_t}[\log X_j] = \psi(\alpha_j) - \psi(\alpha_\Sigma)$, where $\psi(z) = \frac{d}{dz} \log \Gamma(z)$ is the digamma function.

Alternative form via incomplete Beta function. Here, we show the connection of Eq. 38 to the formulation of conditional velocity field in Stärk et al. (2024). Define the incomplete Beta function as $I_x(a, b) = B_x(a, b)/B(a, b) = \int_0^x s^{a-1}(1 - s)^{b-1} / \int_0^1 s^{a-1}(1 - s)^{b-1}$. Then, one can show

$$C(x_i, t) = -\frac{B(t + 1, K - 1 - \gamma t)}{x_i^t (1 - x_i)^{K - 1 - \gamma t}}\, \partial_a I_{x_i}(a, K - 1 - \gamma t)\big|_{a = t+1}, \tag{39}$$

$$D(x_{\bar{i}}, t) = -\frac{B(1 - \gamma t, K - 1 + t)}{x_{\bar{i}}^{-\gamma t} (1 - x_{\bar{i}})^{K - 1 + t}}\, \partial_a I_{x_{\bar{i}}}(a, K - 1 + t)\big|_{a = 1 - \gamma t}, \tag{40}$$

where the derivative of the incomplete Beta function with respect to the first argument is

$$\partial_a I_x(a, b) = \frac{1}{B(a, b)} \int_0^x s^{a-1}(1 - s)^{b-1} \big[\log s - (\psi(a) - \psi(a + b))\big]\, ds. \tag{41}$$

In other words,

$$C(x_i, t) = -x_i^{-\alpha_i(t) + 1}\, (1 - x_i)^{-\alpha_\Sigma(t) + \alpha_i(t)} \int_0^{x_i} s^{\alpha_i(t) - 1}\, (1 - s)^{\alpha_\Sigma(t) - \alpha_i(t)}\, \frac{\log s - \langle \log X_i \rangle_t}{1 - s}\, \mathrm{d}s, \tag{42}$$

$$D(x_{\bar{i}}, t) = -x_{\bar{i}}^{-\alpha_{\bar{i}}(t) + 1}\, (1 - x_{\bar{i}})^{-\alpha_\Sigma(t) + \alpha_{\bar{i}}(t)} \int_0^{x_{\bar{i}}} s^{\alpha_{\bar{i}}(t) - 1}\, (1 - s)^{\alpha_\Sigma(t) - \alpha_{\bar{i}}(t)}\, \frac{\log s - \langle \log X_{\bar{i}} \rangle_t}{1 - s}\, \mathrm{d}s, \tag{43}$$

where $\langle \log X_i \rangle_t \equiv \mathbb{E}_{p_t}[\log X_i] = \psi(\alpha_i) - \psi(\alpha_\Sigma)$, $\psi(z) = \frac{d}{dz} \log(\Gamma(z))$ is the digamma function, $I_x(a, b)$ is the regularized incomplete Beta function, and $B(a, b)$ is the Beta function.

If $\gamma = 0$, then $C(x_i, t) = -\frac{B(t+1,K-1)}{(1-x_i)^{K-1}x_i^t}\partial_a I_{x_i}(a, K-1)\big|_{a=t+1}$ and Eq. 32 becomes $u_t(x|x_1 = e_i) = C(x_i, t)(e_i - x)$. Hence this is a generalized Dirichlet flow that is consistent with the result from Stärk et al. (2024).

The factors $(e_j - x)$ make the field tangent to the simplex, and the construction above satisfies the continuity equation for the prescribed path. Hence, $u_t(x|x_1 = e_i)$ generates $p_t(x|x_1 = e_i)$.

Unpaired POVMs. For a target i, let $w_{ij} \geq 0$, $w_{ii} = 0$, and $\sum_{j\neq i} w_{ij} = 1$. We use $\alpha_i(t) = 1 + t$ and $\alpha_j(t) = 1 - \gamma w_{ij} t$ for $j \neq i$. The same coordinatewise derivation gives $u_t(x \mid e_i) = C_i(x_i, t)(e_i - x) - \gamma \sum_{j\neq i} w_{ij} D_{ij}(x_j, t)(e_j - x)$, where each D_{ij} uses the Beta marginal induced by $\alpha_j(t)$. The Pauli-6 construction is recovered by placing unit weight on the antipodal outcome. For the tetrahedral SIC-POVM, whose distinct Bloch vectors have equal inner product $-1/3$, we use uniform weights $w_{ij} = 1/(K-1)$. Positivity requires $\gamma T \max_{j\neq i} w_{ij} < 1$.

E Evaluation Metrics

We evaluate the isotropic two-site correlation $C_{ij} = \mathrm{tr}(O_{ij}\rho)$ with $O_{ij} = (X_iX_j + Y_iY_j + Z_iZ_j)/3$. For local shadow snapshots $\hat{\rho}_i^{(m)} = 3|s_i^{(m)}\rangle\langle s_i^{(m)}| - \mathbb{I}$, its estimator is

$$\hat{C}_{ij} = \frac{1}{3\,M_{\mathrm{infer}}} \sum_{o\in\{x,y,z\}} \sum_{m=1}^{M_{\mathrm{infer}}} \mathrm{tr}\big(\sigma_i^o \otimes \sigma_j^o \hat{\rho}_i^{(m)} \otimes \hat{\rho}_j^{(m)}\big), \tag{44}$$

The 2D tables report ZZ; SU(2) symmetry gives $\langle XX\rangle = \langle YY\rangle = \langle ZZ\rangle$ for the exact Heisenberg target.

For two-site Rényi-2 entropy $E_{ij} = -\log \mathrm{tr}(\rho_{ij}^2)$, independent generated snapshots give the U-statistic

$$\hat{E}_{ij} = -\log\left[\max\left\{\epsilon, \frac{1}{M_{\mathrm{infer}}(M_{\mathrm{infer}}-1)} \sum_{m\neq n} \mathrm{tr}\Big(\mathbf{S}_{ij}\hat{\rho}_{ij}^{(m)} \otimes \hat{\rho}_{ij}^{(n)}\Big)\right\}\right], \tag{45}$$

where $\mathbf{S}_{ij}$ swaps the two subsystem copies and $\epsilon = 10^{-12}$ is a numerical floor.

F Experimental Settings

Topology-aligned conditioning. Let $x_i(t)$ be the noisy shadow variable at site i, p_i its positional embedding, $\ell_i(h)$ the embedding of Hamiltonian coefficients incident on that site, and $g(h)$ a global summary of the complete Hamiltonian. The initial token is

$$z_i^{(0)} = E_x(x_i(t)) + p_i + E_{\mathrm{loc}}(\ell_i(h)). \tag{46}$$

At Transformer block r, the global condition and flow time modulate the normalized hidden state through

$$\tilde{z}_i^{(r)} = \big(1 + s_r(g(h), t)\big) \odot \mathrm{LN}(z_i^{(r)}) + b_r(g(h), t), \tag{47}$$

followed by global self-attention and an MLP. Thus local features identify which interaction touches each token, while attention still permits every site to use the full Hamiltonian context. A geometry-specific output head parameterizes the velocity or denoising field and decodes all sites in parallel.

We train on 100 Hamiltonians with $M_{\mathrm{train}} = 1000$ shadows per Hamiltonian and select on 20 disjoint validation Hamiltonians. Although the stored test pool contains 100 states, the reporting protocol freezes its first 20 before evaluation, giving an effective 100/20/20 train/validation/test split. The remaining 80 test-pool states are not used in table metrics or model selection. The primary TFIM benchmark uses the inhomogeneous OBC Hamiltonian in Equation (49), with condition vector $[J_1, \ldots, J_{L-1}, c_1, \ldots, c_L]$. The Heisenberg condition contains one coupling per physical OBC bond.

Table 6: **Default implementation and model-selection protocol.**

Component	Setting
Backbone	6 Transformer blocks, 8 heads, width 512, AdaLN; batch size 1024.
Optimization	Adam, learning rate 2×10^{-4}; at most 1000 epochs.
Validation	Every 20 epochs; checkpoint and hyperparameters selected at $B = 10$k on 20 validation states.
Test	Frozen 20-state split; $B \in \{1\text{k}, 10\text{k}, 20\text{k}\}$; NFE$= 20$ unless stated otherwise.
ADFM γ	Selected from $\{0.01, 0.03, 0.1, 0.2\}$ using seed 0 and then fixed across seeds.

Task definitions. For an open 1D chain, the random-coupling Heisenberg and inhomogeneous TFIM Hamiltonians are

$$H_{\text{Heis}}(\mathbf{c}) = \sum_{i=1}^{L-1} c_i \left(\sigma_i^x \sigma_{i+1}^x + \sigma_i^y \sigma_{i+1}^y + \sigma_i^z \sigma_{i+1}^z\right), \tag{48}$$

$$H_{\text{TFIM}}(\mathbf{J}, \mathbf{c}) = -\sum_{i=1}^{L-1} J_i \sigma_i^z \sigma_{i+1}^z - \sum_{i=1}^{L} c_i \sigma_i^x. \tag{49}$$

We use positive random Heisenberg couplings and smooth quenched TFIM disorder whose mean transverse field is stratified over $[0.25, 1.75]$. In the inhomogeneous TFIM benchmark, both J_i and c_i vary.

For the uniform phase-transition control, we use

$$H_{\text{uTFIM}}(c) = -\sum_{i=1}^{L-1} \sigma_i^z \sigma_{i+1}^z - c \sum_{i=1}^{L} \sigma_i^x, \tag{50}$$

with the Ising coupling fixed to $J = 1$. The sweep coordinate $h \in (0, 1)$ is parameterized as $c = h/(1-h)$, which preserves the ground states up to a positive rescaling.

For an open 2D square lattice, the random-coupling Heisenberg Hamiltonian is

$$H_{\text{2D-Heisenberg}}(\mathbf{c}) = \sum_{\langle i,j \rangle} c_{ij} \left(\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y + \sigma_i^z \sigma_j^z\right), \tag{51}$$

where $\langle i, j \rangle$ denotes nearest-neighbor pairs on the physical edges of the open square lattice.

Baselines. Fixed-record classical shadows (CS) serve as a reference rather than a learned baseline. RBFK, NTK, Lasso, and Ridge directly regress individual observables; ShadowGPT is autoregressive; and QuaDiM, LinearFM, and StatisticalFM serve as diffusion, Euclidean-flow, and categorical-flow baselines, respectively. For the 6×6 system, CS estimates use exact two-site marginals, which are sufficient for the reported observables. For the uniform-TFIM visualization, we restrict to the global-X even-parity sector to avoid arbitrary mixing of the near-degenerate ordered-phase doublet.

Checkpoints are selected by minimizing the scale-normalized validation score

$$\max\left(\frac{\text{RMSE}_{\text{corr}}}{s_{\text{corr}}}, \frac{\text{RMSE}_{\text{ent}}}{s_{\text{ent}}}\right) + 10^{-3} \sqrt{\frac{\text{RMSE}_{\text{corr}} \text{RMSE}_{\text{ent}}}{s_{\text{corr}} s_{\text{ent}}}}, \tag{52}$$

where s_{corr} and s_{ent} are the RMS magnitudes of the corresponding validation labels. The test set is not used for model selection. The selected ADFM values are $\gamma = 0.01$ for Heisenberg-10/Cube-8, 0.03 for Heisenberg-100, 0.1 for Heisenberg-$30/50$, tetrahedral, 4×4, and TFIM-100, and 0.2 for the remaining TFIM and $5 \times 5/6 \times 6$ experiments.

For Hamiltonian conditioning, each 1D token receives its adjacent couplings and local field, while each 2D token receives its four incident couplings; missing OBC edges are zero-padded. Global self-attention is unrestricted. ShadowGPT uses its published 4-layer, 4-head, width-128 architecture. For tetrahedral Spherical Flow, the prior is a validation-selected, target-independent symmetric tangent-Gaussian mixture with full support on S^2.

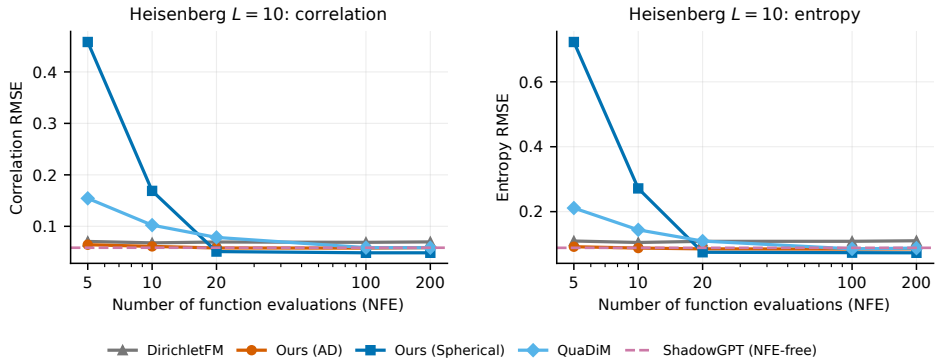

Figure 6: **Effect of the number of function evaluations (NFE) on 1D Heisenberg-10.** NFE denotes the number of iterative model evaluations during sampling. We vary NFE while keeping the validation-selected checkpoint fixed. ShadowGPT is autoregressive and therefore does not use NFE.

Sampling cost. For $B = 1000$ on a single H100, generation takes 0.60 s for ShadowGPT, 6.02 s for QuaDiM, and 12.08–15.62 s for the flow-based models. Within a fixed flow model, NFE$= 10$ and NFE$= 100$ require approximately $0.52\times$ and $4.9\times$ the runtime of NFE$= 20$, respectively.

G Additional Ground-State Results

G.1 NFE Analysis

Figure 6 studies the effect of inference-time NFE while keeping the validation-selected checkpoint fixed. Spherical Flow reaches nearly converged performance by NFE $= 20$, whereas QuaDiM continues to improve as the number of denoising steps increases up to NFE $= 100$. ADFM and DirichletFM show smaller changes across the evaluated solver budgets. These results indicate that Spherical Flow achieves strong performance with relatively few sampling steps on Heisenberg-10. This analysis uses fixed checkpoints and is separate from the primary three-seed evaluation.

H Full Uncertainty Tables

The following tables report the three training-seed mean and standard deviation for the main 1D and 2D budget sweeps. Model selection is unchanged: validation at $B = 10k$ precedes evaluation on the frozen test-20.

Table 7: **Full inhomogeneous-TFIM results.** Entries report mean (sample SD) over three runs where available.

Method	$L = 10$						$L = 30$					
	Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓		
	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$
Direct Shadows (lower bound)												
CS Huang et al. (2020)	0.052	0.015	0.011	0.071	0.020	0.014	0.053	0.017	0.012	0.076	0.025	0.017
Classical ML												
RBFK Huang et al. (2022)		0.019			0.043			0.043			0.069	
NTK Huang et al. (2022)		0.018			0.038			0.016			0.077	
Lasso Zhao et al. (2025)		0.011			0.062			0.018			0.170	
Ridge Zhao et al. (2025)		0.023			0.066			0.045			0.194	
Autoregressive												
ShadowGPT Yao & You (2024)	0.052	0.023	0.018	0.087	0.046	0.034	0.056	0.024	0.020	0.098	0.054	0.061
CFM / Diffusion												
LinearFM Lipman et al. (2022)	0.058 (0.002)	0.029 (0.005)	0.027 (0.006)	0.1 (0.007)	0.056 (0.009)	0.055 (0.009)	0.061 (0.004)	0.032 (0.006)	0.029 (0.006)	0.157 (0.019)	0.13 (0.02)	0.127 (0.019)
QuaDiM Tang et al. (2025)	0.062 (0.003)	0.033 (0.005)	0.03 (0.006)	0.115 (0.002)	0.088 (0.001)	0.086 (0.003)	0.068 (0.007)	0.046 (0.012)	0.044 (0.012)	0.189 (0.03)	0.17 (0.04)	0.168 (0.038)
CS-DFM												
StatisticalFM Cheng et al. (2024)	0.057 (0.002)	0.025 (0)	0.022 (0.002)	0.096 (0.01)	0.055 (0.01)	0.052 (0.008)	0.065 (0.007)	0.037 (0.005)	0.035 (0.005)	0.147 (0.031)	0.122 (0.033)	0.122 (0.032)
DirichletFM	0.058 (0.005)	0.027 (0.005)	0.025 (0.004)	0.095 (0.006)	0.064 (0.007)	0.061 (0.006)	0.062 (0.005)	0.03 (0.006)	0.028 (0.005)	0.115 (0.009)	0.086 (0.011)	0.084 (0.011)
Ours (AD)	0.057 (0.004)	0.025 (0.002)	0.022 (0.002)	0.089 (0.003)	0.051 (0.007)	0.048 (0.008)	0.062 (0.004)	0.032 (0.006)	0.03 (0.007)	0.117 (0.017)	0.087 (0.016)	0.085 (0.015)
Ours (Spherical)	0.053 (0.003)	0.021 (0.003)	0.019 (0.003)	0.078 (0.003)	0.037 (0.005)	0.034 (0.004)	0.058 (0.004)	0.026 (0.004)	0.023 (0.005)	0.092 (0.003)	0.054 (0.004)	0.051 (0.003)

Method	$L = 50$						$L = 100$					
	Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓		
	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$
Direct Shadows (lower bound)												
CS	0.054	0.018	0.012	0.075	0.024	0.016	0.055	0.017	0.012	0.076	0.022	0.016
Classical ML												
RBFK		0.034			0.073			0.027			0.070	
NTK		0.021			0.075			0.013			0.074	
Lasso		0.009			0.122			0.009			0.120	
Ridge		0.065			0.245			0.081			0.274	
Autoregressive												
ShadowGPT	0.056	0.021	0.018	0.108	0.067	0.064	0.063	0.036	0.032	0.166	0.142	0.143
CFM / Diffusion												
QuaDiM	0.062 (0.002)	0.034 (0.003)	0.032 (0.003)	0.132 (0.01)	0.108 (0.016)	0.107 (0.016)	0.06 (0.002)	0.034 (0.004)	0.031 (0.005)	0.158 (0.007)	0.136 (0.01)	0.134 (0.009)
LinearFM	0.057 (0.002)	0.026 (0.005)	0.023 (0.005)	0.107 (0.01)	0.075 (0.017)	0.073 (0.017)	0.062 (0.003)	0.036 (0.005)	0.034 (0.005)	0.179 (0.027)	0.159 (0.029)	0.158 (0.029)
CS-DFM												
StatisticalFM	0.062 (0.002)	0.035 (0.003)	0.032 (0.004)	0.149 (0.033)	0.12 (0.037)	0.118 (0.038)	0.063 (0.003)	0.038 (0.006)	0.036 (0.007)	0.16 (0.044)	0.135 (0.048)	0.133 (0.048)
DirichletFM	0.062 (0.005)	0.037 (0.005)	0.035 (0.006)	0.113 (0.011)	0.091 (0.011)	0.089 (0.011)	0.061 (0.003)	0.033 (0.005)	0.031 (0.006)	0.159 (0.009)	0.139 (0.012)	0.138 (0.013)
Ours (AD)	0.057 (0.003)	0.028 (0.003)	0.026 (0.003)	0.12 (0.011)	0.093 (0.013)	0.09 (0.014)	0.059 (0.002)	0.027 (0.002)	0.024 (0.002)	0.118 (0.013)	0.09 (0.016)	0.088 (0.016)
Ours (Spherical)	0.055 (0)	0.024 (0.002)	0.02 (0.003)	0.095 (0.016)	0.055 (0.018)	0.051 (0.019)	0.062 (0.003)	0.034 (0.005)	0.032 (0.005)	0.135 (0.032)	0.112 (0.035)	0.11 (0.035)

Table 8: **Full 1D Heisenberg results.** Entries report mean (sample SD) over three runs where available.

Method	$L = 10$						$L = 30$					
	Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓		
	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$
Direct Shadows (lower bound)												
CS Huang et al. (2020)	0.053	0.015	0.010	0.079	0.020	0.015	0.050	0.016	0.011	0.083	0.031	0.025
Classical ML												
RBFK Huang et al. (2022)		0.087			0.151			0.140			0.222	
NTK Huang et al. (2022)		0.085			0.142			0.129			0.205	
Lasso Zhao et al. (2025)		0.103			0.168			0.142			0.226	
Ridge Zhao et al. (2025)		0.103			0.168			0.146			0.230	
Autoregressive												
ShadowGPT Yao & You (2024)	0.079	0.058	0.058	0.118	0.089	0.087	0.112 (0.003)	0.1 (0.004)	0.1 (0.003)	0.176 (0.008)	0.161 (0.009)	0.161 (0.007)
CFM / Diffusion												
LinearFM Lipman et al. (2022)	0.106 (0.011)	0.094 (0.014)	0.093 (0.014)	0.161 (0.015)	0.141 (0.021)	0.138 (0.02)	0.116 (0.009)	0.105 (0.007)	0.104 (0.008)	0.181 (0.015)	0.164 (0.011)	0.163 (0.011)
QuaDiM Tang et al. (2025)	0.101 (0.005)	0.083 (0.003)	0.083 (0.003)	0.144 (0.01)	0.114 (0.007)	0.113 (0.008)	0.11 (0.002)	0.1 (0.003)	0.098 (0.003)	0.168 (0.003)	0.15 (0.006)	0.148 (0.006)
CS-DFM												
StatisticalFM Cheng et al. (2024)	0.104 (0.009)	0.096 (0.009)	0.095 (0.009)	0.158 (0.012)	0.143 (0.012)	0.141 (0.012)	0.118 (0.01)	0.107 (0.01)	0.107 (0.01)	0.18 (0.013)	0.163 (0.012)	0.163 (0.012)
DirichletFM	0.076 (0.002)	0.066 (0.003)	0.063 (0.001)	0.117 (0.007)	0.101 (0.008)	0.096 (0.006)	0.109 (0.006)	0.095 (0.006)	0.095 (0.006)	0.17 (0.01)	0.15 (0.008)	0.15 (0.009)
Ours (AD)	0.071 (0.006)	0.055 (0.002)	0.054 (0.004)	0.109 (0.009)	0.083 (0.003)	0.081 (0.006)	0.108 (0.007)	0.095 (0.004)	0.096 (0.004)	0.167 (0.01)	0.15 (0.008)	0.149 (0.009)
Ours (Spherical)	0.069 (0.004)	0.052 (0.002)	0.053 (0.003)	0.106 (0.007)	0.079 (0.005)	0.078 (0.006)	0.099 (0.004)	0.085 (0.004)	0.085 (0.003)	0.155 (0.005)	0.135 (0.006)	0.134 (0.004)

Method	$L = 50$						$L = 100$					
	Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓		
	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$
Direct Shadows (lower bound)												
CS	0.052	0.017	0.011	0.078	0.024	0.016	0.051	0.016	0.011	0.076	0.023	0.016
Classical ML												
RBFK		0.167			0.249			0.202			0.316	
NTK		0.176			0.262			0.213			0.327	
Lasso		0.173			0.259			0.213			0.317	
Ridge		0.177			0.262			0.305			0.480	
Autoregressive												
ShadowGPT	0.134	0.125	0.125	0.217	0.206	0.207	0.141	0.134	0.133	0.223	0.214	0.213
CFM / Diffusion												
QuaDiM	0.122 (0.008)	0.113 (0.008)	0.113 (0.008)	0.197 (0.012)	0.181 (0.011)	0.18 (0.012)	0.117 (0.001)	0.106 (0.002)	0.105 (0.002)	0.187 (0.002)	0.168 (0.001)	0.167 (0)
LinearFM	0.146 (0.016)	0.139 (0.017)	0.138 (0.017)	0.236 (0.024)	0.22 (0.025)	0.22 (0.025)	0.147 (0.01)	0.139 (0.011)	0.139 (0.011)	0.234 (0.016)	0.22 (0.016)	0.219 (0.017)
CS-DFM												
StatisticalFM	0.123 (0.008)	0.114 (0.008)	0.113 (0.008)	0.201 (0.01)	0.184 (0.01)	0.184 (0.01)	0.132 (0.001)	0.123 (0.001)	0.123 (0.001)	0.215 (0.003)	0.201 (0.003)	0.2 (0.003)
DirichletFM	0.179 (0.001)	0.173 (0.002)	0.172 (0.002)	0.265 (0.002)	0.254 (0.003)	0.252 (0.003)	0.102 (0.007)	0.09 (0.007)	0.089 (0.007)	0.163 (0.01)	0.145 (0.01)	0.145 (0.01)
Ours (AD)	0.113 (0.002)	0.103 (0.001)	0.102 (0.002)	0.179 (0.004)	0.163 (0.002)	0.162 (0.002)	0.098 (0.003)	0.086 (0.003)	0.085 (0.003)	0.157 (0.005)	0.141 (0.004)	0.14 (0.004)
Ours (Spherical)	0.134 (0.013)	0.125 (0.014)	0.124 (0.014)	0.214 (0.014)	0.199 (0.014)	0.198 (0.014)	0.123 (0.005)	0.112 (0.007)	0.111 (0.007)	0.193 (0.005)	0.177 (0.01)	0.176 (0.01)

Table 9: **Two-dimensional OBC Heisenberg results with uncertainty.** Learned entries report mean (sample SD) over three runs where available.

Method	4×4						5×5						6×6					
	Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓			Corr. ↓			Ent. ↓		
	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$	$1k$	$10k$	$20k$
Direct Shadows (lower bound)																		
CS	0.097	0.030	0.021	0.080	0.025	0.017	0.094	0.029	0.020	0.082	0.025	0.017	0.094	0.030	0.021	0.085	0.025	0.017
Classical ML																		
RBFK		0.087			0.138			0.109			0.151			0.124			0.186	
NTK		0.087			0.136			0.108			0.149			0.124			0.184	
Lasso		0.086			0.132			0.109			0.145			0.137			0.188	
Ridge		0.087			0.132			0.114			0.148			0.152			0.197	
Autoregressive																		
ShadowGPT	0.125	0.088	0.084	0.150	0.125	0.124	0.135	0.100	0.098	0.174	0.154	0.152	0.129	0.092	0.090	0.158	0.132	0.131
CFM / Diffusion																		
LinearFM	0.13(0.015)	0.094(0.017)	0.094(0.017)	0.187(0.032)	0.167(0.033)	0.166(0.033)	0.127(0.006)	0.087(0.001)	0.084(0.001)	0.182(0.003)	0.164(0.003)	0.163(0.003)	0.129(0.013)	0.093(0.017)	0.09(0.017)	0.178(0.022)	0.157(0.021)	0.156(0.02)
QuaDiM	0.124(0.008)	0.085(0.008)	0.081(0.009)	0.164(0.006)	0.141(0.002)	0.14(0.002)	0.124(0.004)	0.085(0.003)	0.083(0.003)	0.17(0.006)	0.149(0.003)	0.146(0.005)	0.126(0.002)	0.083(0.005)	0.081(0.005)	0.165(0.009)	0.141(0.01)	0.139(0.01)
CS-DFM																		
StatisticalFM	0.122(0.009)	0.08(0.008)	0.079(0.006)	0.168(0.008)	0.147(0.011)	0.147(0.01)	0.126(0.009)	0.092(0.007)	0.089(0.008)	0.169(0.016)	0.152(0.018)	0.151(0.018)	0.125(0.006)	0.085(0.002)	0.082(0.003)	0.168(0.011)	0.145(0.011)	0.142(0.011)
DirichletFM	0.117(0.006)	0.08(0.004)	0.077(0.003)	0.14(0.003)	0.115(0.001)	0.114(0.001)	0.121(0.001)	0.081(0.002)	0.078(0.002)	0.139(0.002)	0.111(0.001)	0.108(0.002)	0.116(0.001)	0.078(0.001)	0.075(0.002)	0.136(0.003)	0.109(0.002)	0.107(0.002)
Ours (AD)	0.116(0.003)	0.076(0.006)	0.073(0.005)	0.137(0.011)	0.108(0.01)	0.107(0.011)	0.12(0.002)	0.078(0.002)	0.075(0.002)	0.141(0.012)	0.112(0.007)	0.109(0.008)	0.117(0.002)	0.077(0.001)	0.073(0.001)	0.136(0.001)	0.108(0.002)	0.106(0.001)
Ours (Spherical)	0.111(0.004)	0.07(0.003)	0.067(0.002)	0.127(0.007)	0.1(0.006)	0.099(0.007)	0.121(0.004)	0.08(0.001)	0.078(0.001)	0.147(0.001)	0.121(0.001)	0.119(0.001)	0.116(0.003)	0.076(0.002)	0.073(0.002)	0.134(0.006)	0.109(0.004)	0.108(0.003)