Diffusion-Based Generation with Non-Equilibrium Transport

In collaboration with **Gurtej Kanwar**

Octavio Vega

University of Illinois Urbana-Champaign

July 30, 2026 — 43rd International Symposium on LFT

UNIVERSITY OF MARYLAND, COLLEGE PARK

Motivation

Deep generative models (flows, diffusions, ...) arising in LFT

- **See:** Plenary talk by D. Hackett [Tue. 09:45]
- **See:** Talk by E. Cellini [Thu. 14:40]

ML samplers are imperfect → unbiased results via reweighting $w = p/q$

- But efficiency set by *effective sample size* (ESS)

$$\mathrm{ESS} = \frac{\mu_w^2}{\mu_w^2 + \sigma_w^2}$$

⚠️ Grows worse with volume: $\sigma_w \to \infty$, $\mathrm{ESS} \to 0$

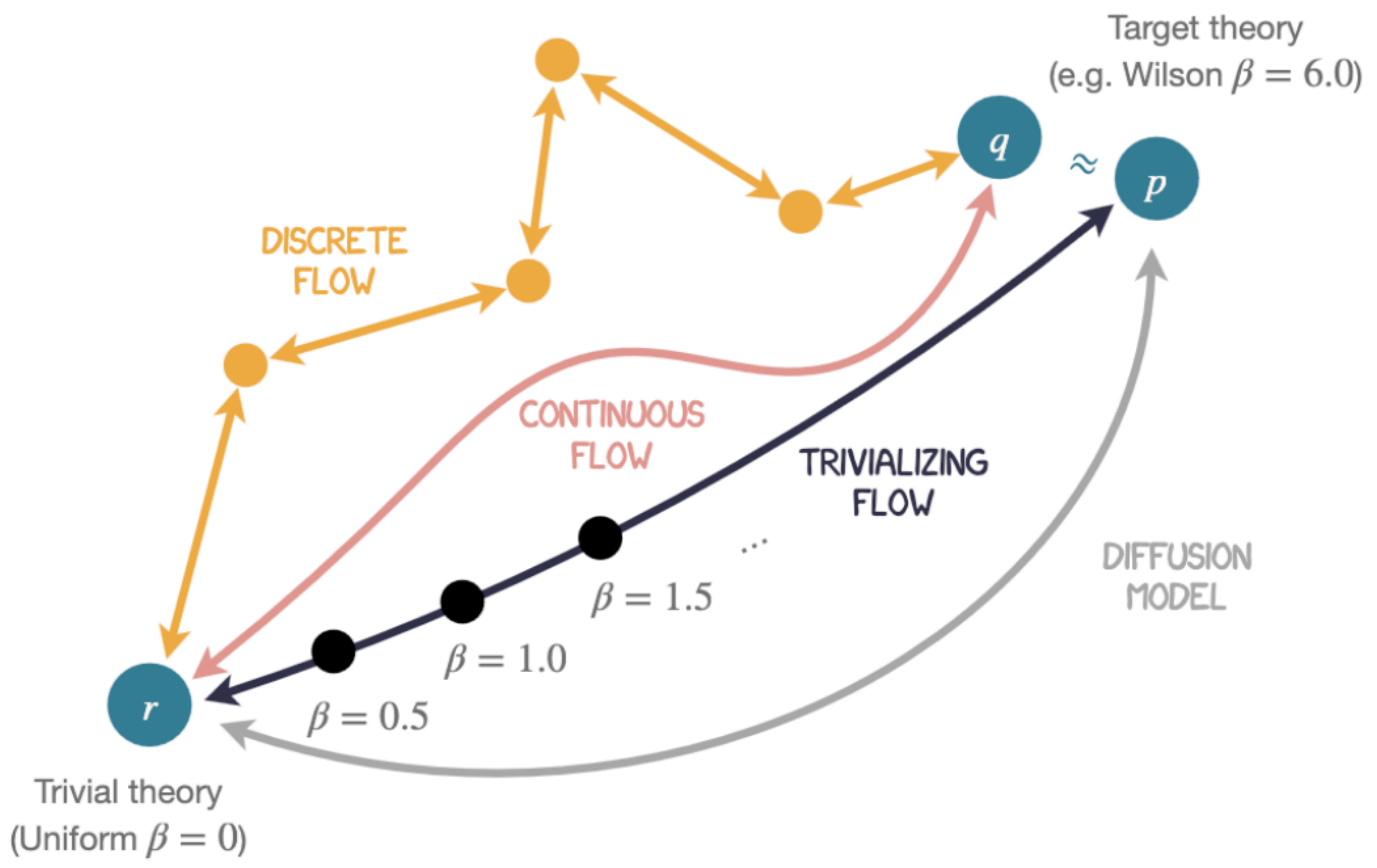


Image credit: G. Kanwar

Diffusion Sampling in LFT

Progress so far

- Exact equivariance in ϕ^4 and $\mathrm{U}(1)$ gauge theory: OV, J. Komijani, A. El-Khadra, M. Marinkovic, *JHEP* **07** (2026) 213
- Sampling on $\mathrm{SU}(N)$ with spectral heat kernel: G. Kanwar and OV, *PoS* **LATTICE2025** (2026) 040
- Schwinger Model with Fermion determinant: OV and A. El-Khadra, *PAI2026* `2606.27481`

Other works

- Diffusion as stochastic quantization: L. Wang et al, *JHEP* **05** (2024) 060
- Complex langevin: G. Aarts et al, `2510.01328`
- U(1) gauge theory + MALA: Q. Zhu et al, *JHEP* **03** (2026) 111
- U(2) gauge theory: G. Aarts et al, `2601.19552`
- SU(2) gauge theory: H. Alharazin et al, `2602.09045`
- SU(3) gauge theory: J. Komijani et al, `2605.06134`
- Noise scheduling on Lie groups: J. Komijani, `2605.17326`
- Scaling near criticality in ϕ^4: Y. Tan et al, `2607.08505`
- Self-learning in XY model: A. Tomiya, `2607.12587`

Background

Annealed Importance Sampling

R. Neal, *Stat. Comput.* **11** (2001)

Approach: break sampling process into smaller steps $t = 0,1,\ldots,T$

Design an annealing path π_t between densities $q(x)$ and $p(x)$

- e.g. $\pi_t(x) = p(x)^{\beta(t)}q(x)^{1-\beta(t)}$, $\beta(t) = \frac{t}{T}$

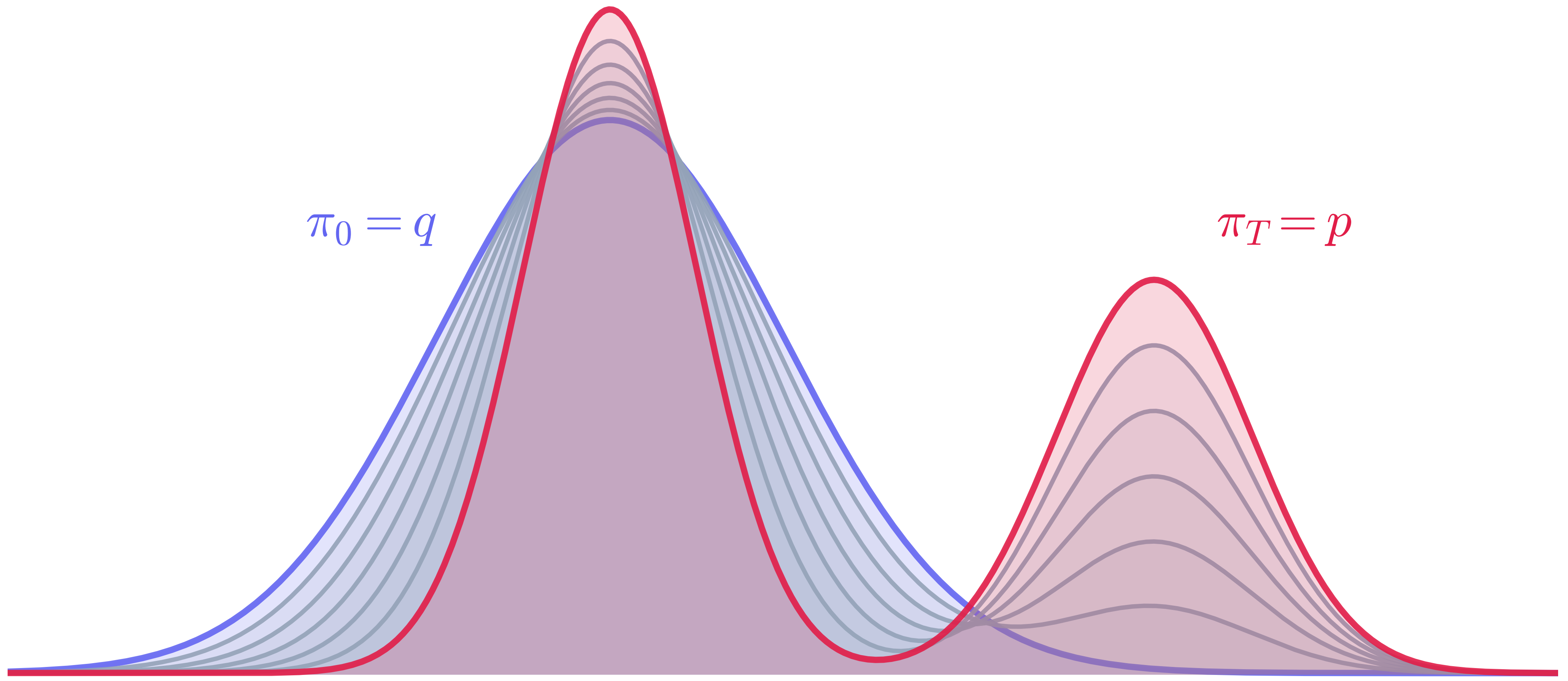
Two-steps per update:

non-equilibrium "work"

- **Reweight:** $\log w_t = \log w_{t-1} + \log \pi_t(x_{t-1}) - \log \pi_{t-1}(x_{t-1})$
- **Transport:** $x_t \sim K_t(\,\cdot\,|\,x_{t-1})$

AIS transport kernel is the bottleneck…

- Can a learned transport map help?



Annealing with Learned Transport

Idea: augment MCMC transport with a learned, deterministic flow

- machine-learned drift field <u>approximates</u> perfect transport $\pi_t \to \pi_{t+1}$

Related: *stochastic normalizing flows* (SNF) [Wu et al *NeurIPS* **33** (2020), Caselle et al *PoS* **LATTICE2022** (2023) 005]

- Alternate learned transport with MCMC re-equilibration steps
- Re-weight by the accumulated work → unbiased
- Transport + correction at each step → weights stay low variance

Continuum limit: take infinitesimal annealing steps with $n_{\text{step}} \to \infty$

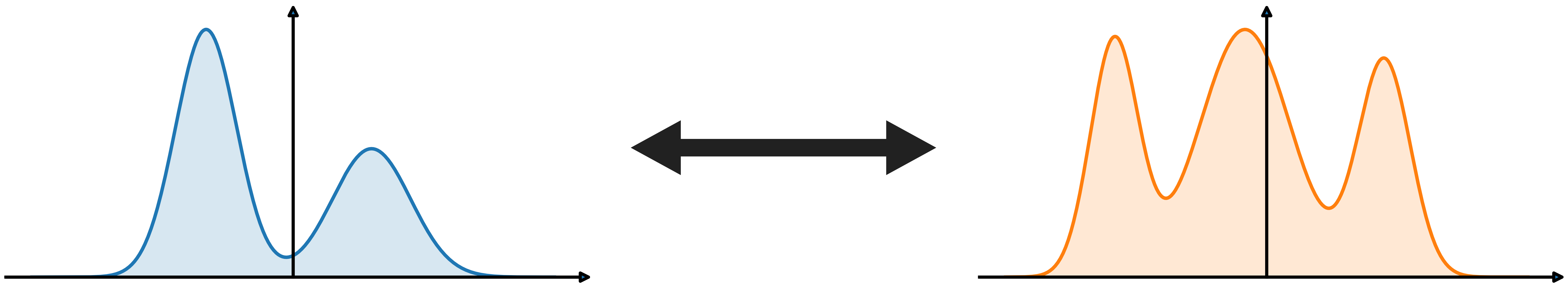
→ Probability density ρ_t evolves according to *Fokker-Planck Equation*

$$\partial_t \rho_t = -\nabla \cdot (\rho_t v_t) + \frac{g_t^2}{2} \nabla^2 \rho_t$$

Dynamical Measure Transport

Corresponding **SDE** gives continuous motion of samples $x_t \sim \rho_t(x)$

$$dx_t = v_t(x_t)dt + g_t dW_t$$



Theorem: Every forward process has a family of reverse processes,

$$dx_t = \left[v_t(x_t) - \frac{g_t^2 + \tilde{g}_t^2}{2} \boxed{\nabla \log \rho_t(x_t)} \right] dt + \tilde{g}_t dW_t$$

effective force

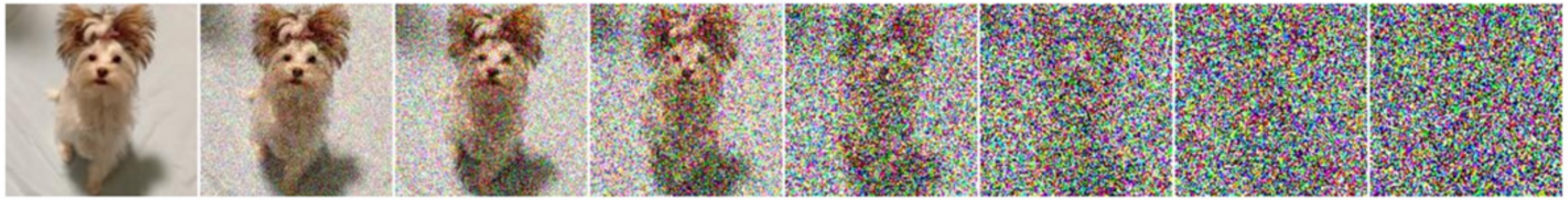
Diffusion as Annealing

Sohl-Dickstein et al, *PMLR* **37** (2015)
Ho et al, *NeurIPS* **33** (2020)

Diffusion process: defines the annealing path

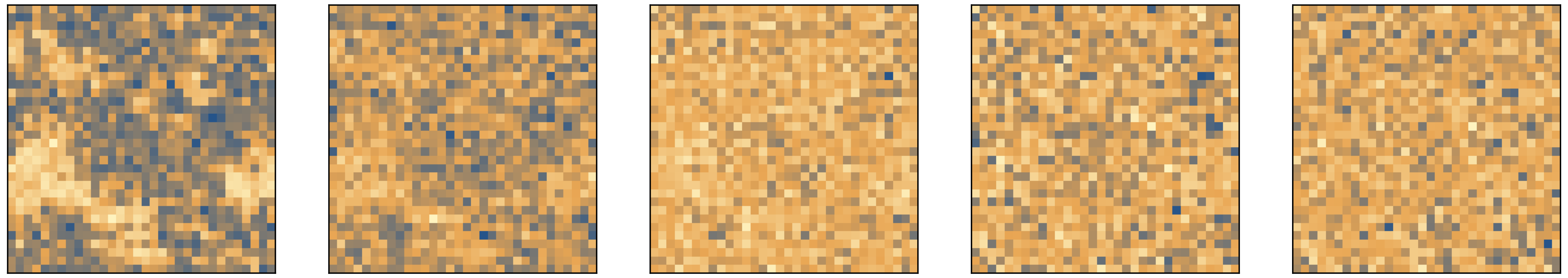
Image credit: Song et al, *ICLR* (2021)

data → noise



Denoising process: gives learned transport

noise → data



Forward diffusion supplies the path *and* transport known up to $\nabla \log \rho_t$

- Train neural net $\hat{b}_t$ to approximate reverse flow field   *e.g. score/flow matching*

Energy-Based Diffusion Models

SNF/AIS needs energy differences $\log \pi_t(x) - \log \pi_{t-1}(x)$, not just gradients

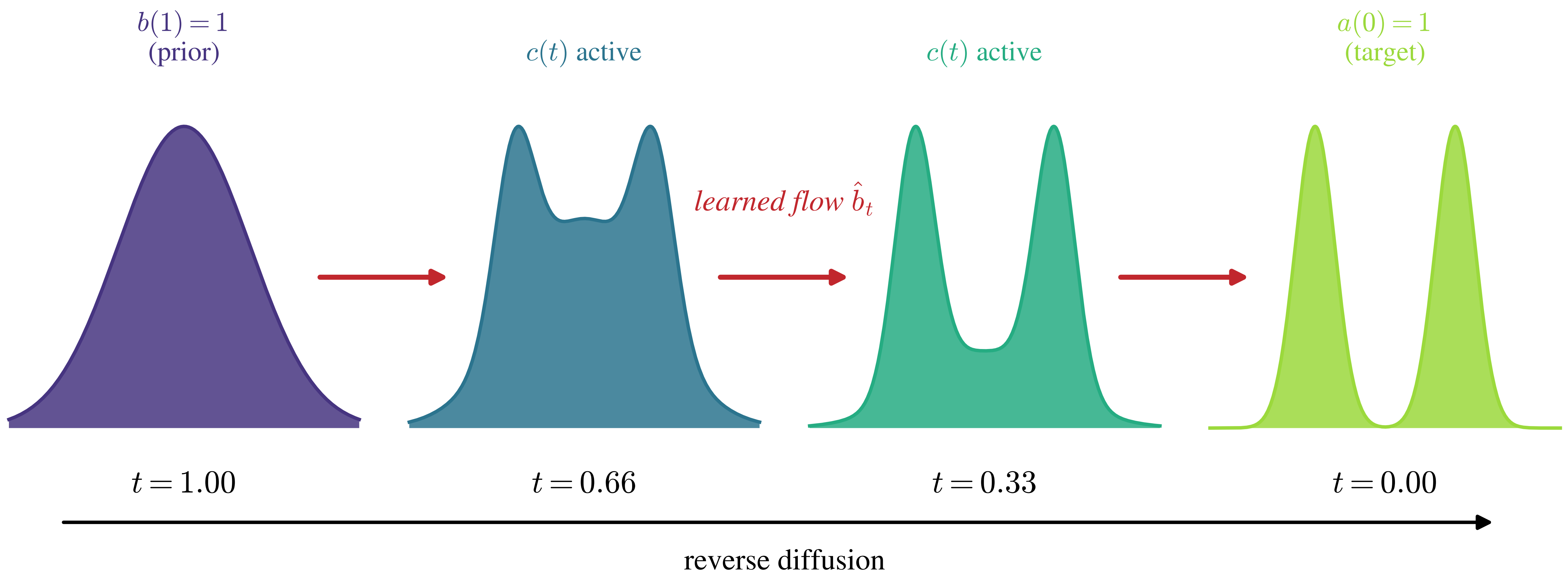
- Pure score model gives the flow, but not the weights

Parameterize energy function $E_t(x)$ through neural net

$$S_t^\theta(x) = a(t)E_0(x) + b(t)E_1(x) + c(t)E_t^\theta(x)$$

$a(0) = b(1) = 1$
$a(1) = b(0) = 0$
$c(0) = c(1) = 0$

Time-dependent $\rho_t \propto \exp(S_t)$ defines a continuous annealing path!



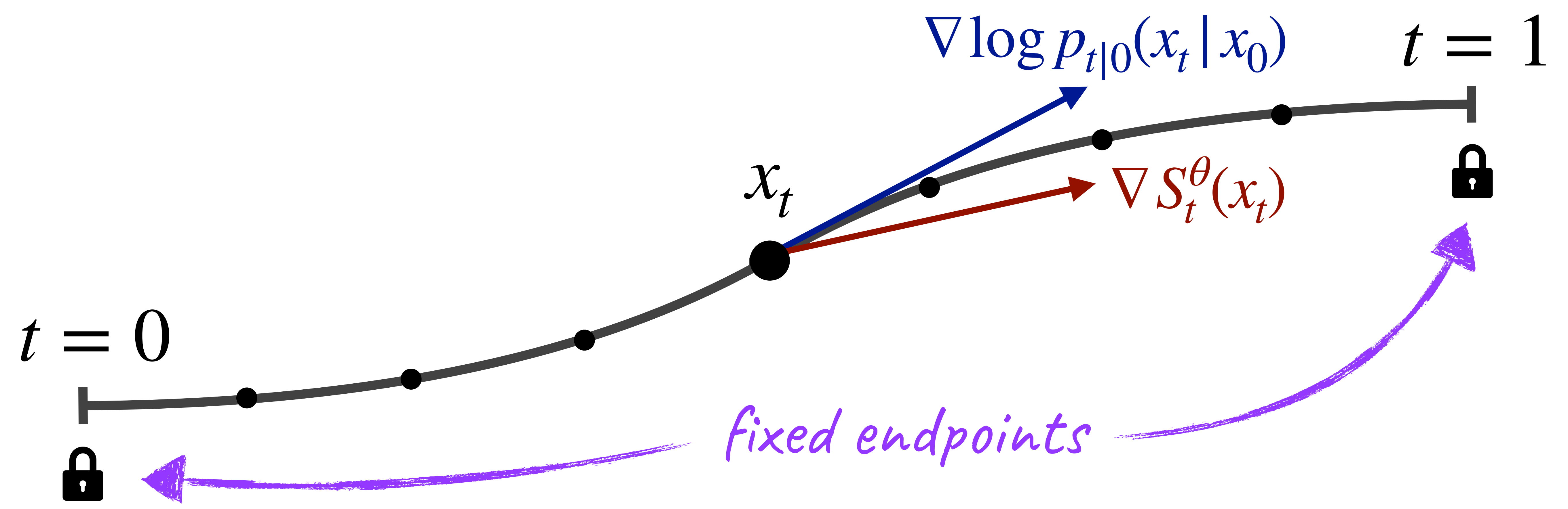
Training Diffusion EBMs

Neural energy trained through modified score-matching

$$\mathcal{L}(\theta) = \int_0^1 \mathbb{E}_{x_0 \sim p_0} \mathbb{E}_{x_t \sim p_{t|0}} \left[\left\| -\nabla S_t^\theta(x_t) - \nabla \log p_{t|0}(x_t \,|\, x_0) \right\|^2 \right] dt$$

Parameterization details:

- "Kinetic energy" through velocity field: $E_t^\theta(x) = \frac{1}{2} \left\| v_t^\theta(x) \right\|^2$
- constrained endpoints: $S_{t=0}^\theta \equiv S_0, S_{t=1}^\theta \equiv E_1$



Sampling: Non-Equilibrium Transport

integrate reverse ODE

Transport (flow)

$$x_t \to x_{t'}, \quad \log J = -\int \nabla \cdot b_t \, dt$$

accumulate dynamical weights w_t

Re-weight (correct)

$$\log w_{t'} = \log w_t + \log \pi_{t'} - \log \pi_t + \log J$$

SNF sampler!

π_t-invariant MCMC step to re-randomize samples

HMC (optional)

$$p_{\text{acc}} = \min\{1, e^{-\Delta H}\}$$

Results

Free Field: Exact Benchmark

Gaussian

Fourier transform $\phi(x) \xrightarrow{\mathcal{F}} \tilde{\phi}(k)$ reveals exact solvability:

$$S\left[\tilde{\phi}\right] = \int \frac{d^d k}{(2\pi)^d}(k^2 + m^2)\left|\tilde{\phi}(k)\right|^2 \Rightarrow p\left(\tilde{\phi}\right) \propto e^{-(k^2+m^2)\left|\tilde{\phi}(k)\right|^2}$$

Evolve samples through a forward diffusion process:

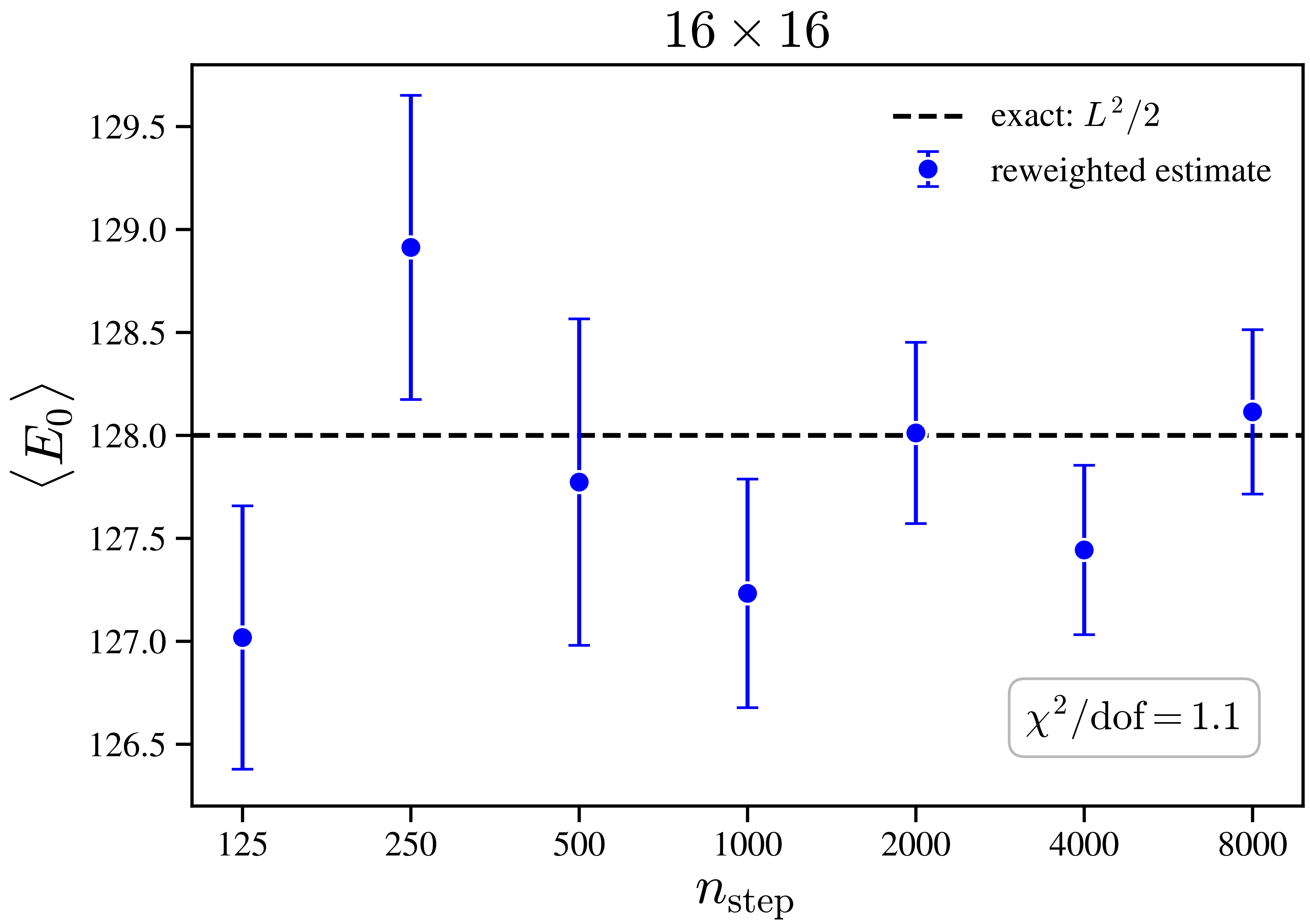
$$\tilde{\phi}_t(k) = \alpha_t \tilde{\phi}_0(k) + \sigma_t \eta, \quad \eta \sim \mathcal{N}(0, \mathbb{I})$$

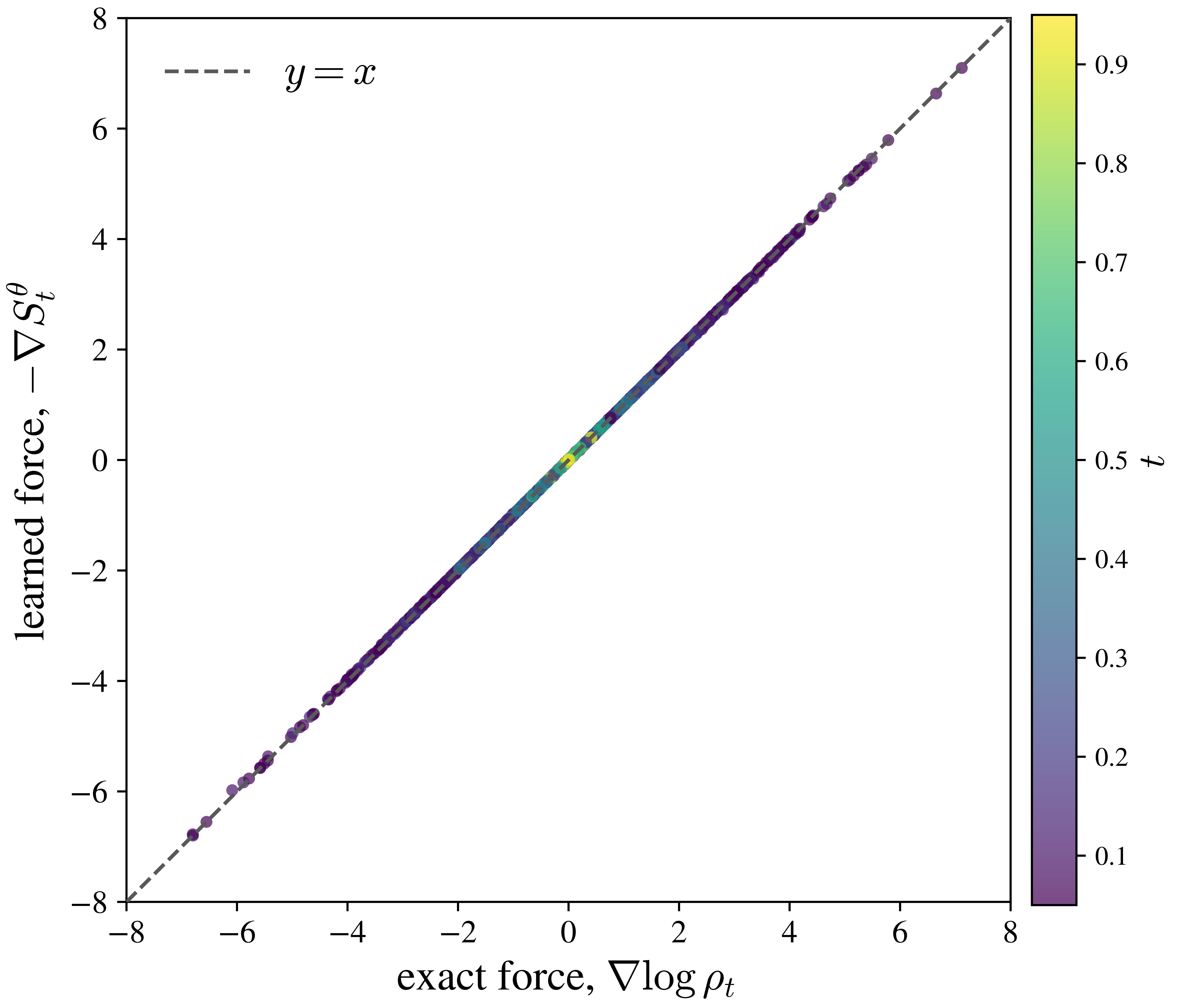
Exact energy and force easily computed

$$E_t\left(\tilde{\phi}_t(k)\right) = \frac{1}{2}\frac{\omega(k)^2}{\alpha_t^2 + \sigma_t^2 \omega(k)^2}\left|\tilde{\phi}_t(k)\right|^2, -\nabla E\left(\tilde{\phi}_t(k)\right) = -\frac{\omega(k)^2}{\alpha_t^2 + \sigma_t^2 \omega(k)^2}\tilde{\phi}_t(k)$$

Still Unbiased

Reweighted $\langle E_0 \rangle$ consistent with exact value

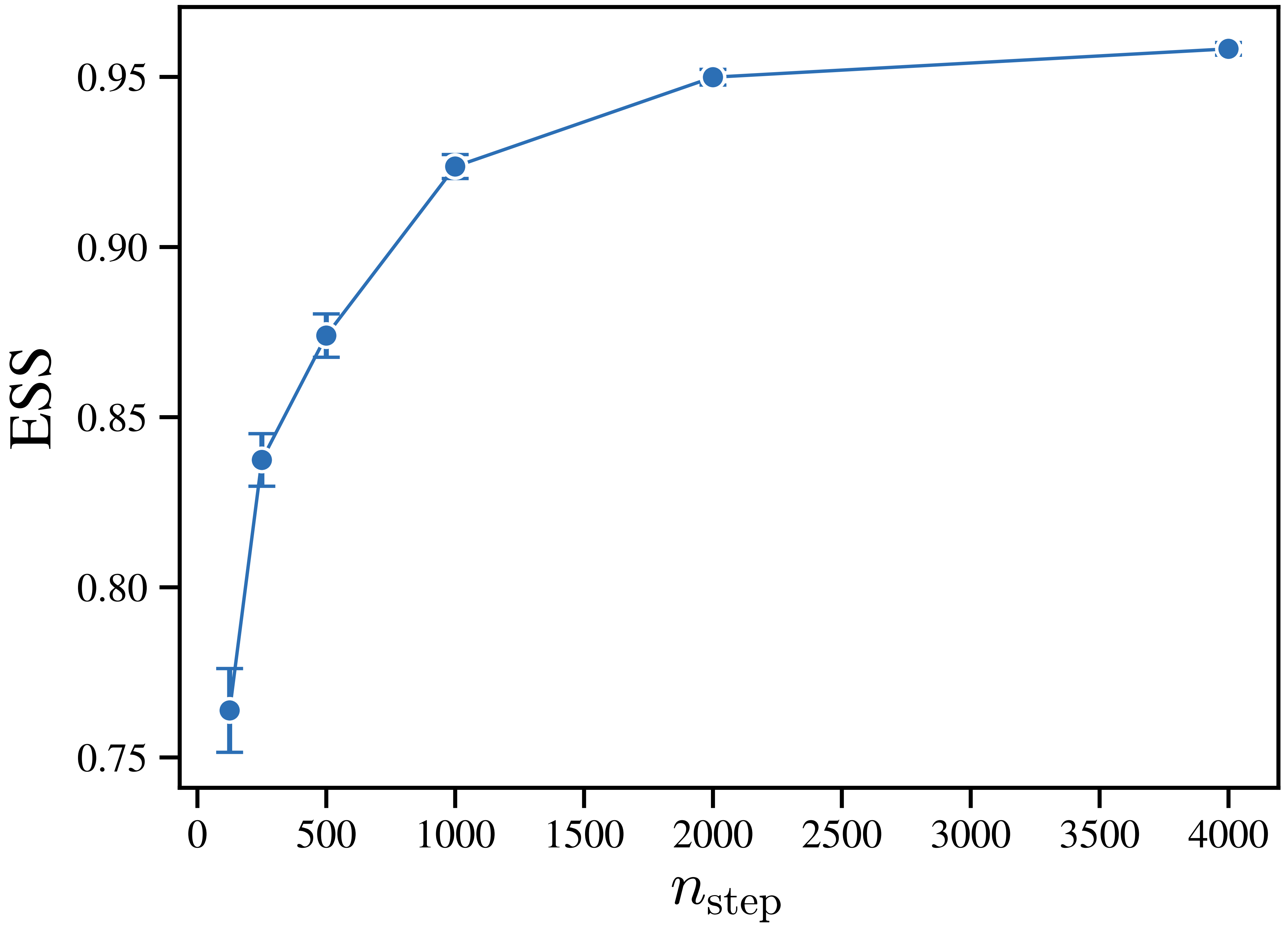


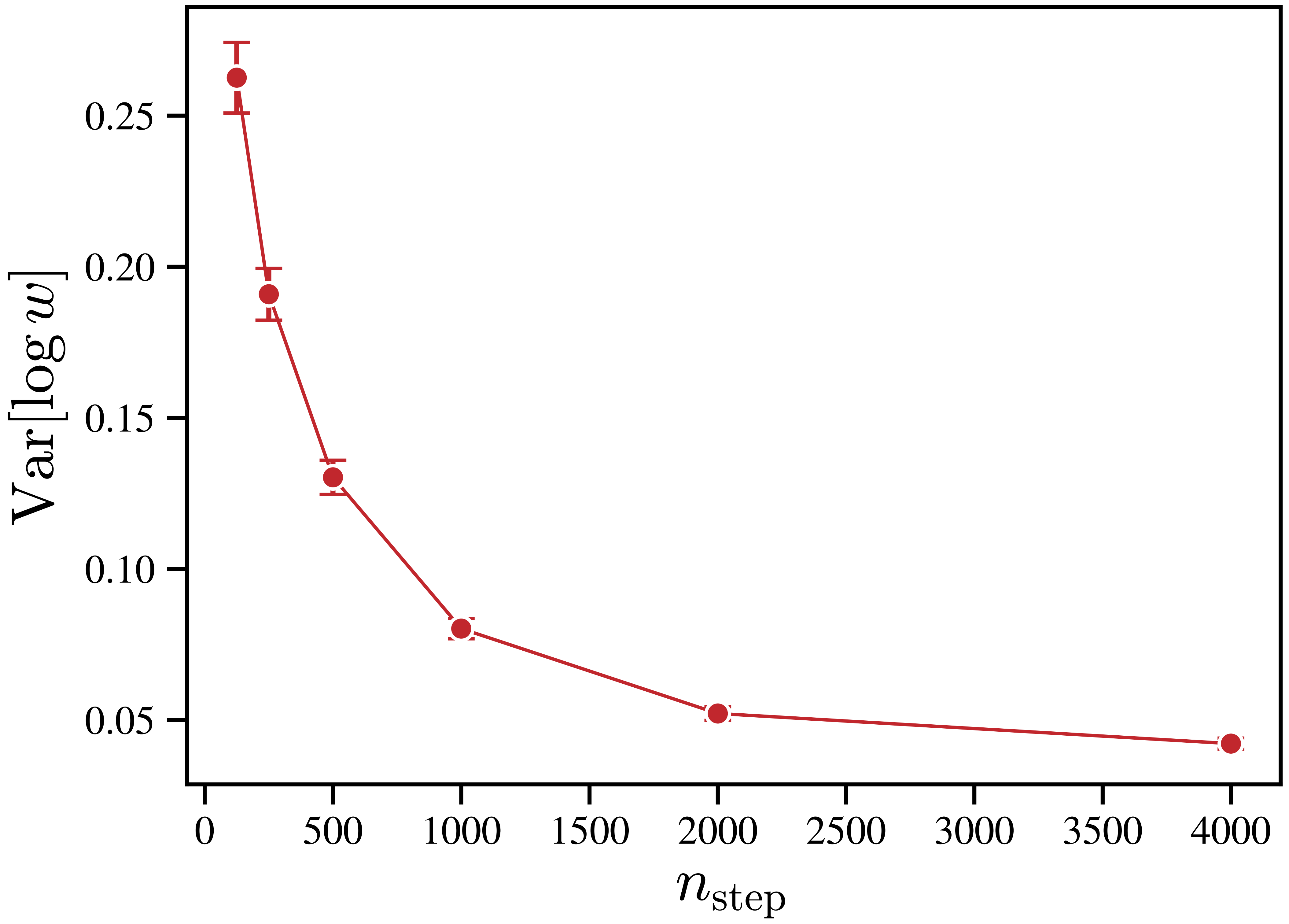


Controlled Scaling

Statistical efficiency is a tunable knob

- No retraining needed!



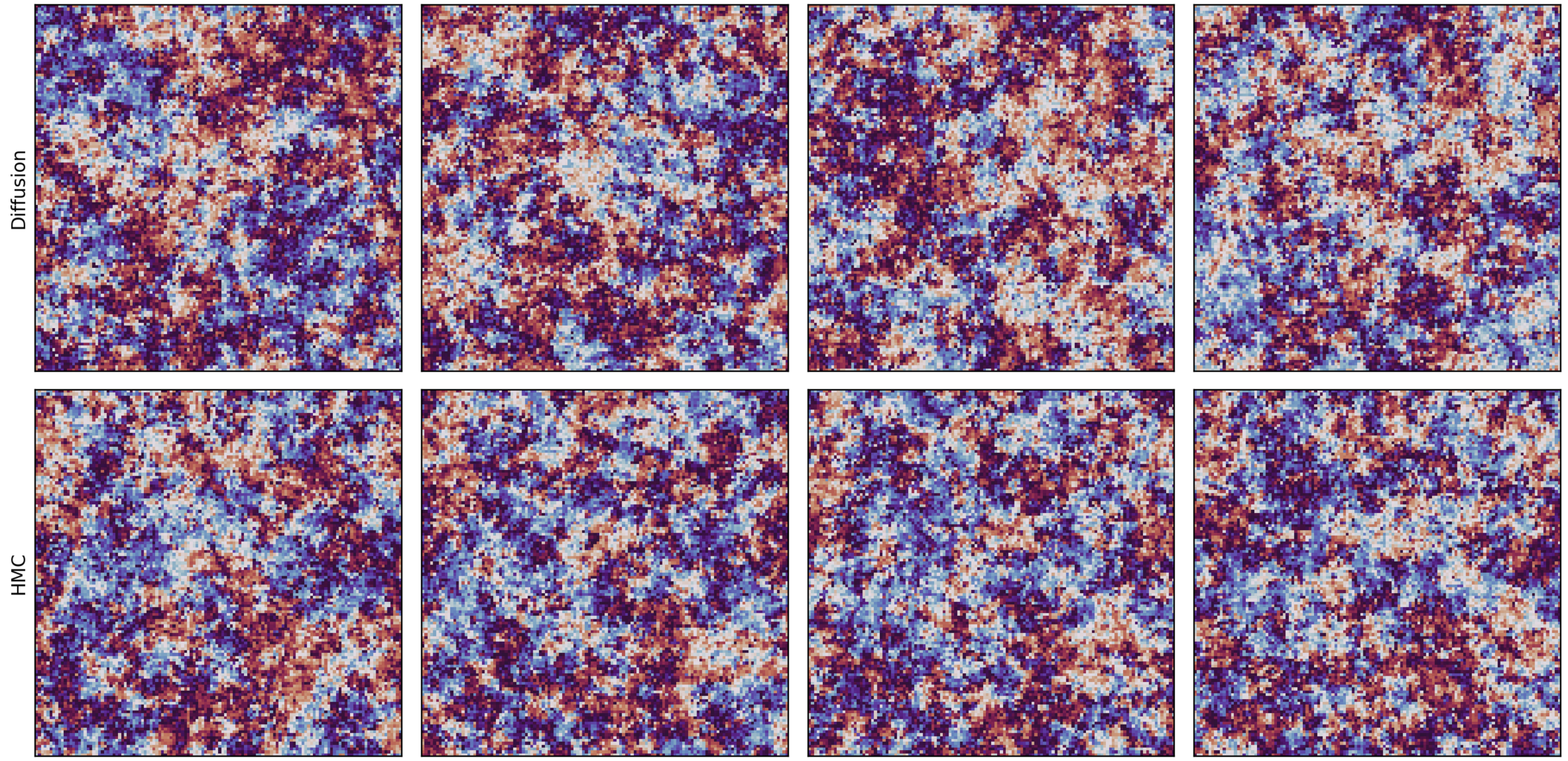


$\mathrm{O}(3)$ Sigma Model in 2D

$$S[\phi] = -\beta \sum_x \sum_\mu \vec{\phi}(x) \cdot \vec{\phi}(x+\hat{\mu}), \quad \vec{\phi}(x) \in \mathbb{S}^2$$

Physical features:

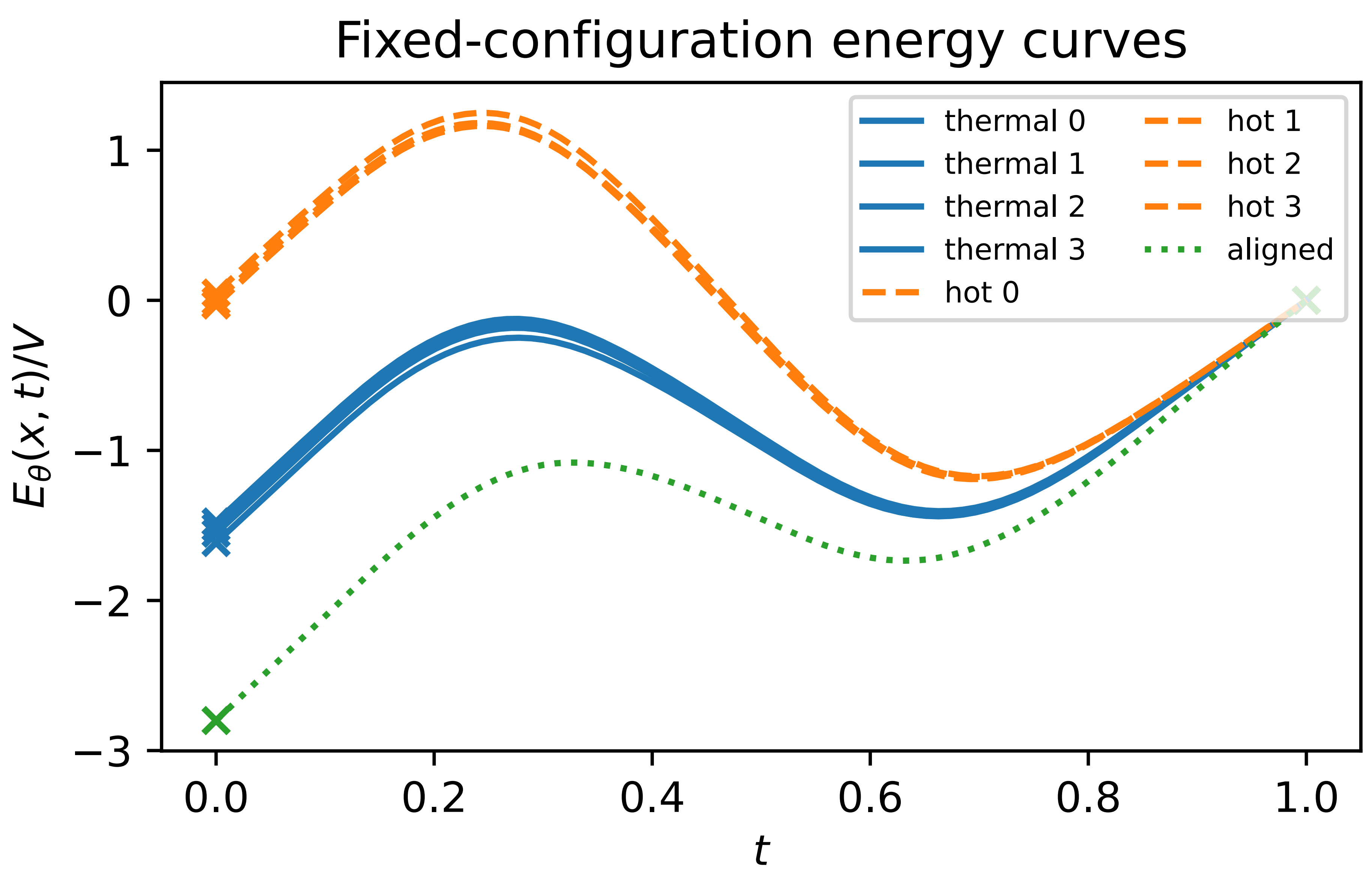
- Fields live on sphere → curved target, geometry-aware transport
- Asymptotic freedom + topological sectors → HMC hits topological freezing

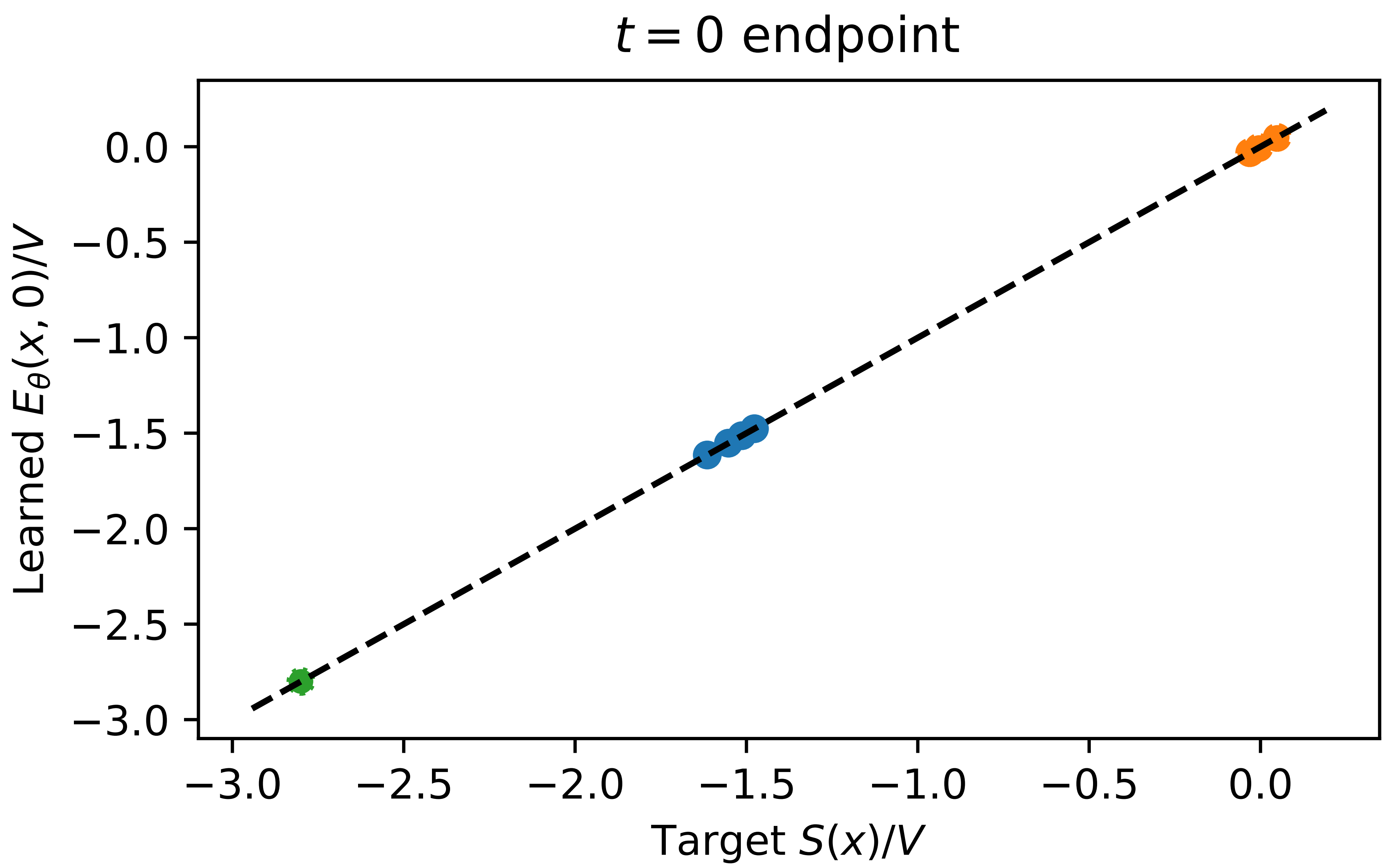


A Calibrated Path

Learned energy interpolates prior ↔ target across time

- Endpoints pinned by construction



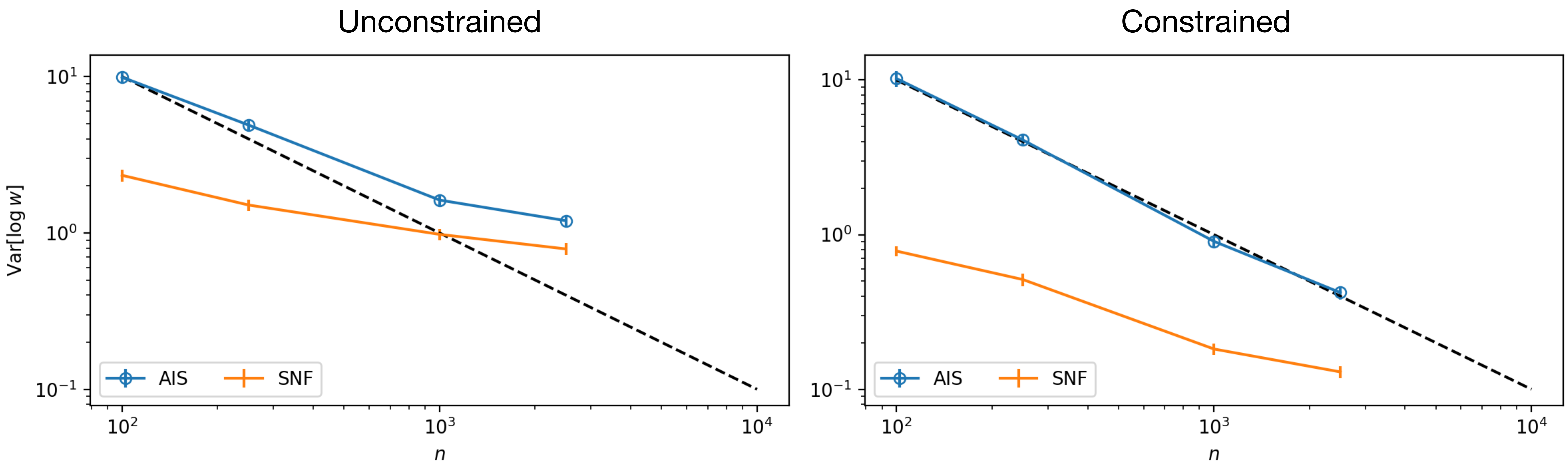


$$16 \times 16,\ \beta = 1.4$$

SNF beats AIS

Ideal scaling: $1/n$

- Endpoint constraint restores better scaling ✔
- At fixed steps: SNF < AIS ✔



Summary

Recap:

- Replaced the MCMC transport bottleneck of AIS with a learned continuous flow
- Diffusion-based sampler with in-dynamics bias correction
- Endpoint-constrained energy guarantees calibrated path from prior to target
- ESS ↑ and $\mathrm{Var}[\log w]$ ↓ with refined annealing schedule → tunable efficiency

Key takeaways:

- Truly unbiased diffusion sampling is possible
- Systematic improvement of sample quality also achievable

Outlook:

- Large design space of EBMs to be explored
- Test on non-Abelian gauge theories, larger volumes

References

R. M. Neal, *Annealed Importance Sampling*, Stat. Comput. **11** (2001) 125 [physics/9803008].

H. Wu, J. Köhler, and F. Noe, *Stochastic Normalizing Flows*, Adv. Adv. Neural Inf. Process. Syst. **33** (2020) 5933 [2002.06707].

M. Caselle, E. Cellini, A. Nada, and M. Panero, *Stochastic normalizing flows as non-equilibrium transformations*, JHEP **07** (2022) 015 [2201.08862].

M. Caselle, E. Cellini, A. Nada, and M. Panero, *Stochastic normalizing flows for lattice field theory*, PoS **LATTICE2022** (2023) 005 [2210.03139].

C. Bonanno, A. Nada, and D. Vadacchino, *Out-of-equilibrium simulations to fight topological freezing*, PoS **LATTICE2023** (2024) 005 [2310.11979].

J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*, PMLR **37** (2015) 2256 [1503.03585].

Y. Song and S. Ermon, *Generative modeling by estimating gradients of the data distribution*, Adv. Neural Inf. Process. Syst. **32** (2019) 11918 [1907.05600].

J. Ho, A. Jain, and P. Abbeel, *Denoising diffusion probabilistic models*, Adv. Neural Inf. Process. Syst. **33** (2020) 6840 [2006.11239].

Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, *Score-Based Generative Modeling through Stochastic Differential Equations*, *ICLR* (2021) [2011.13456].

M. Albergo and E. Vanden-Eijnden, *NETS: A Non-Equilibrium Transport Sampler*, PMLR **267** (2025) 1026 [2410.02711].

X. Peng and A. Gao, *Flow perturbation to accelerate Boltzmann sampling*, Nat. Commun. **16** (2025) 6604 [2407.10666].

Backup Slides

Importance Sampling

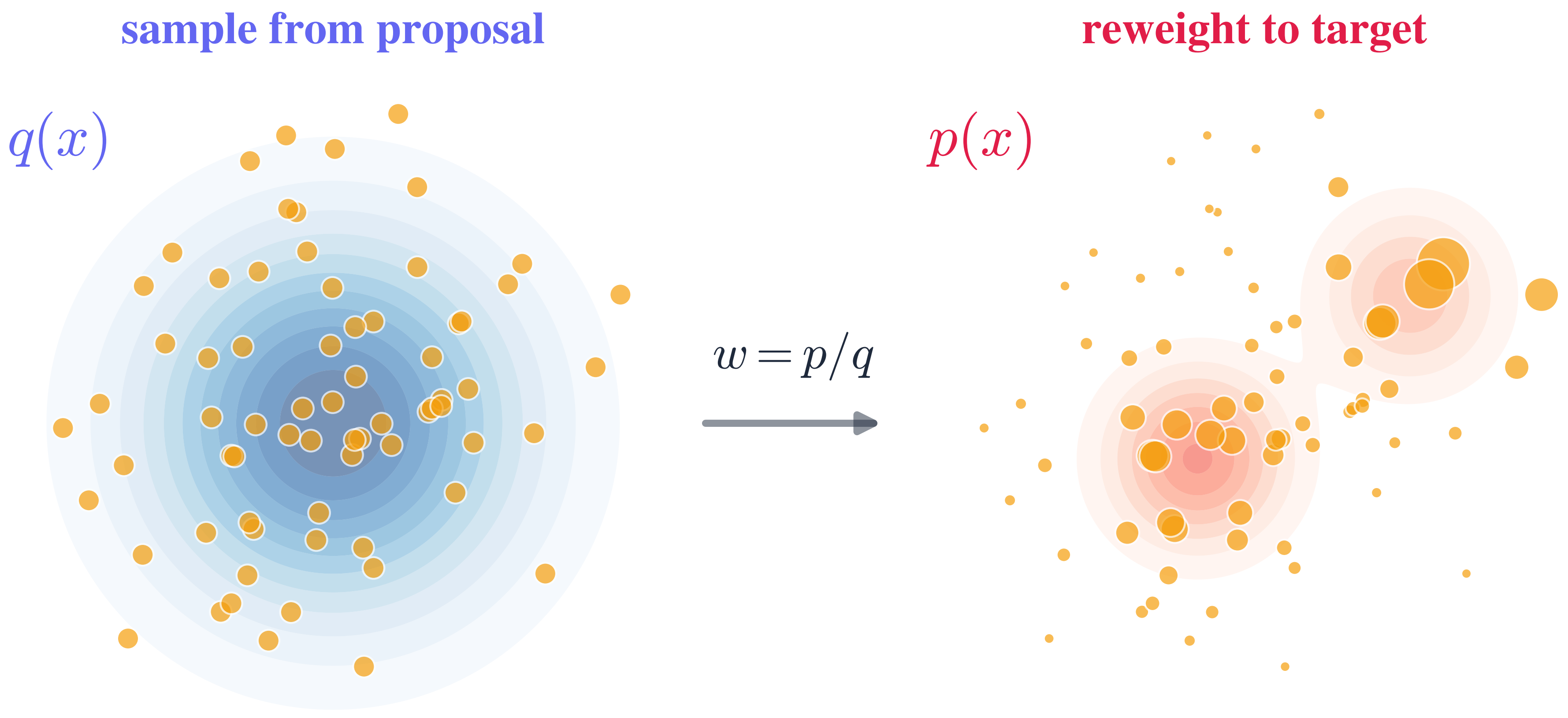
unknown normalization;
hard to sample

Goal: sample a target density $p(x)$, compute estimates

$$\mathbb{E}_p[f] \approx \frac{1}{N}\sum_i f(x_i), \quad p(x) = \frac{1}{\mathcal{Z}} e^{-S(x)}$$

Use a simple proposal density $q(x)$

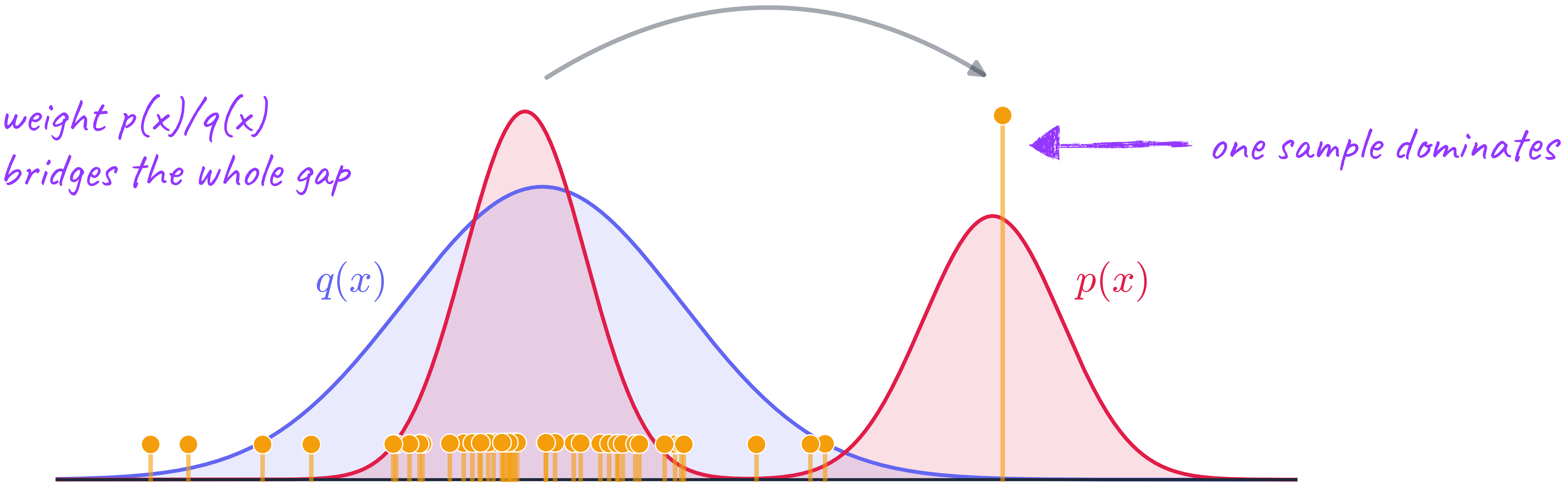
$$\int p(x) f(x) dx = \int \frac{p(x)}{q(x)} q(x) f(x) dx = \mathbb{E}_q[wf], \quad w(x) = \frac{p(x)}{q(x)}$$



The Failure of a Single Leap

Simple proposals are rarely mode-covering

- when q and p don't overlap well, IS collapses



A *single* jump cannot bridge the gap...

Model Likelihoods

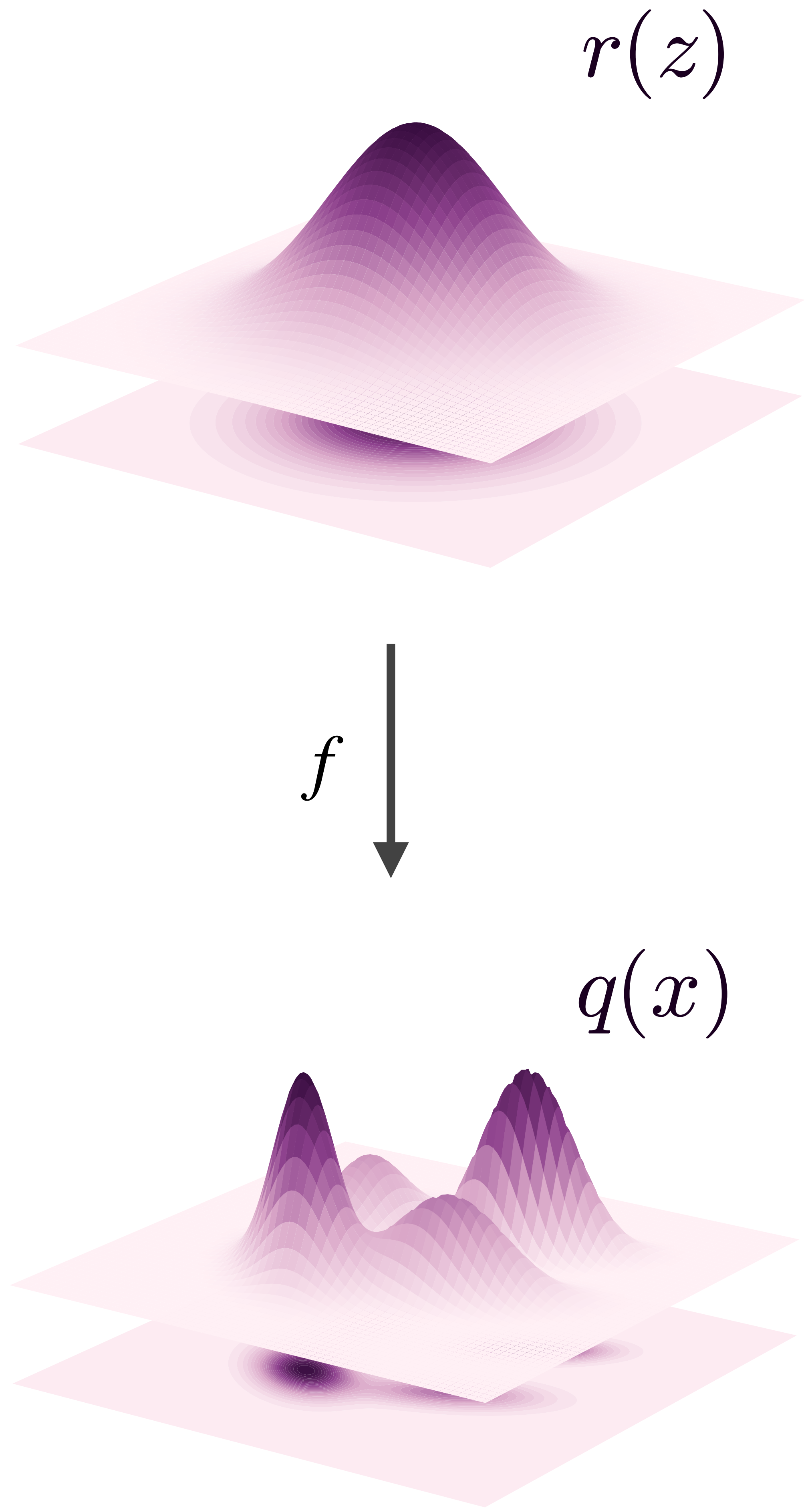
Change of variables: diffeomorphism $x = f(z)$, $z \sim r(z)$

$$q(x) = r(z) \left| \det_{i,j} \frac{\partial z_i}{\partial x_j} \right|$$

Continuous time: solution to the Fokker-Plank equation

$$q_t(x_t) = \underbrace{\exp\left[-\int_{t'}^{t} \nabla \cdot \hat{b}_\tau(x_\tau)\, d\tau \right]}_{\text{Jacobian}} q_{t'}(x_{t'})$$

Challenge: efficiently compute divergence $\nabla \cdot \hat{b}$



Divergence Computation

Problem: log-likelihood requires divergence $\nabla \cdot \hat{b} = \operatorname{Tr} \dfrac{\partial \hat{b}}{\partial x}$

- Exact trace costs $O(d)$ backward passes

Hutchinson: stochastic trace estimator with random vectors ξ

$$\operatorname{Tr}(A) = \mathbb{E}\left[\xi^\top A \xi\right]$$

Flow-Perturbation: unbiased $\det J$ with sphere-averaging [Peng and Gao, *Nat. Commun.* **16** (2025)]

$$|\det A|^{-1} = \mathbb{E}_{\epsilon \sim \mathbb{S}^{d-1}}\left[\|A\epsilon\|^{-d}\right]$$

Related Works

Non-equilibrium sampling approaches

- Caselle et al, *JHEP* **07** (2022)
- Bonanno et al, *PoS* **LATTICE2023** (2024)
- Albergo and Vanden-Eijnden, *PMLR* **267** (2025) 1026