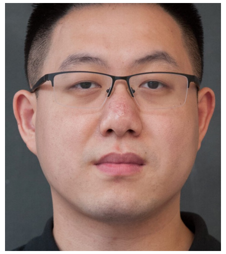
AI/ML for lattice

Dan Hackett
July 28, 2026
The 43rd International Symposium on Lattice Field Theory (Lattice 2026)

Fermilab

My collaborators (past and present; non-exhaustive; some affiliations at time of collab)


Phiala Shanahan


Yang Fu


Denis Boyda
→ **Meta**

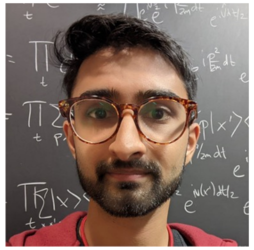

Julian Urban
→ **Uniper**


Sahil Pontula


Michael Albergo


Ryan Abbott


Fernando
Romero-López


Gurtej Kanwar


Kyle Cranmer


Chung-Chun Hsieh

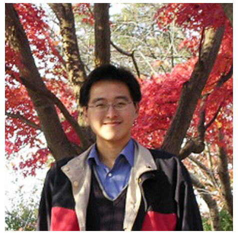

Sébastien Racanière

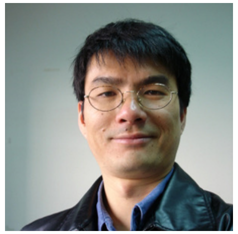

Alex
Matthews


Aleksandar Botev


Ali Razavi


Danilo
Rezende
→ **EllisonIT**


Kai-Feng Chen


Jiunn-Wei Chen

At Lattice 2026

MONDAY JULY 27

[204] S. Yasunaga (14:00) — Reducing the residual mass of domain-wall fermions using machine learning

[25] T. De Oliveira (14:20) — Non-perturbative tuning of relativistic heavy quarks using machine learning in lattice QCD

[328] A. Grebe (14:20) — Toward Compute-Bound Lattice-QCD Software *(Software & machines)*

[58] M. Colaço (14:40) — Mellin moments of proton unpolarized GPDs at nonzero skewness from lattice QCD with neural networks *(Hadron structure)*

[119] S. Pfahler (14:40) — Machine-Learning-Accelerated Multigrid Setup for Lattice QCD Dirac Solves

[129] L. Hostetler (15:00) — PerfAdvisor: An LLM-based Agent for Diagnosing Bottlenecks from GPU Systems Profiles *(Software & machines)*

[216] X. Jiang (16:10) — Use QUDA with Python *(Software & machines)*

[242] A. Singha (16:10) — Overcoming Topological Freezing with Multilevel Generative Plaquette Sampling

[253] R. Dionisio (16:30) — Optimal Paths through Distribution Space for Mitigating Topological Freezing

POSTERS

[304] V. Ayyar — Agentic AI for Accelerating Lattice Gauge Theory Research

[354] A. Hasenfratz — Renormalization-guided normalizing flows for lattice field generation

TUESDAY JULY 28

[280] D. Hackett (09:45) — AI/ML for Lattice ← You are here!

[199] T. Spriggs (11:00) — Neural Network Ground States for Lattice Gauge Theories

[31] D. Schuh (16:10) — Tackling the Sign Problem in the Doped Hubbard Model with Normalizing Flows

[18] F. Freche (16:30) — Machine-Learned Density of States for the Doped Hubbard Model

[301] N. Matsumoto (16:30) — AI/ML-inspired RMHMC kernels–CPN model study *(Software & machines)*

[30] J. Kreit (16:50) — Generative Models for the Low-Temperature Hubbard Model

WEDNESDAY JULY 29

[110] A. De Santis (09:00) — Operator Learning for spectral reconstruction in lattice QCD

[281] Y. Cahuana Medrano (09:00) — Inferring Parton Distributions with Gaussian Processes *(Hadron structure)*

[287] A. Niemiera (10:00) — Bridging Euclidean Methods for Parton Distributions with Machine Learning *(Hadron structure)*

[100] J. Karpie (10:20) — Imaging protons and partons with lattice QCD *(Hadron structure)*

[283] T. Bhattacharya (10:20) — Quantum data learning of phase transitions *(Quantum computing)*

[29] D. Ward (11:10) — Ground States of adjoint SU(2) from Neural Network assisted Variational Montecarlo *(Vacuum structure & confinement)*

[361] C. Minjae (11:50) — Learning Low Modes of the Wilson Dirac Operator with Gauge-Equivariant Networks

[90] H. Hsiao (12:10) — Updates on the Machine Learning Approach for Lattice Gauge Fixing

THURSDAY JULY 30

[147] R. Kapust (14:00) — Solving sign problems with physics-informed kernels

[64] F. Romero López (14:20) — Learning the generating functional for variance reduction in lattice QCD

[179] E. Cellini (14:40) — Neural-Enhanced Out-of-equilibrium sampling for Lattice Field Theory

[338] O. Vega (15:00) — Unbiased Diffusion-Based Generation via Scalable Non-Equilibrium Transport

[327] L. Parato (15:20) — Renormalization Group inspired inverse blocking with conditional normalizing flows

[3] S. Romiti (16:10) — Diagonalizing the Kogut-Susskind Hamiltonian with Physics-Informed Neural Networks

[187] S. Rajawat (16:30) — Neural Wavefunctions in Quantum Field Theory

[340] J. Lin (16:50) — 3+1d SU(3) Yang-Mills + theta-term with Neural Network Quantum States

FRIDAY JULY 31

[101] E. Lundstrum (14:00) — Agentic AI for QCD perturbation theory or: how I learned to stop worrying and love the bot

[214] T. Wada (14:20) — AI-Assisted Higher-Order Hopping-Parameter Expansion in Lattice QCD

[228] P. Butti (14:40) — A variational framework for variance reduction in lattice field theory

[346] U. Wenger (15:00) — Topology and thermodynamics from a machine-learned 4d SU(3) FP gauge action

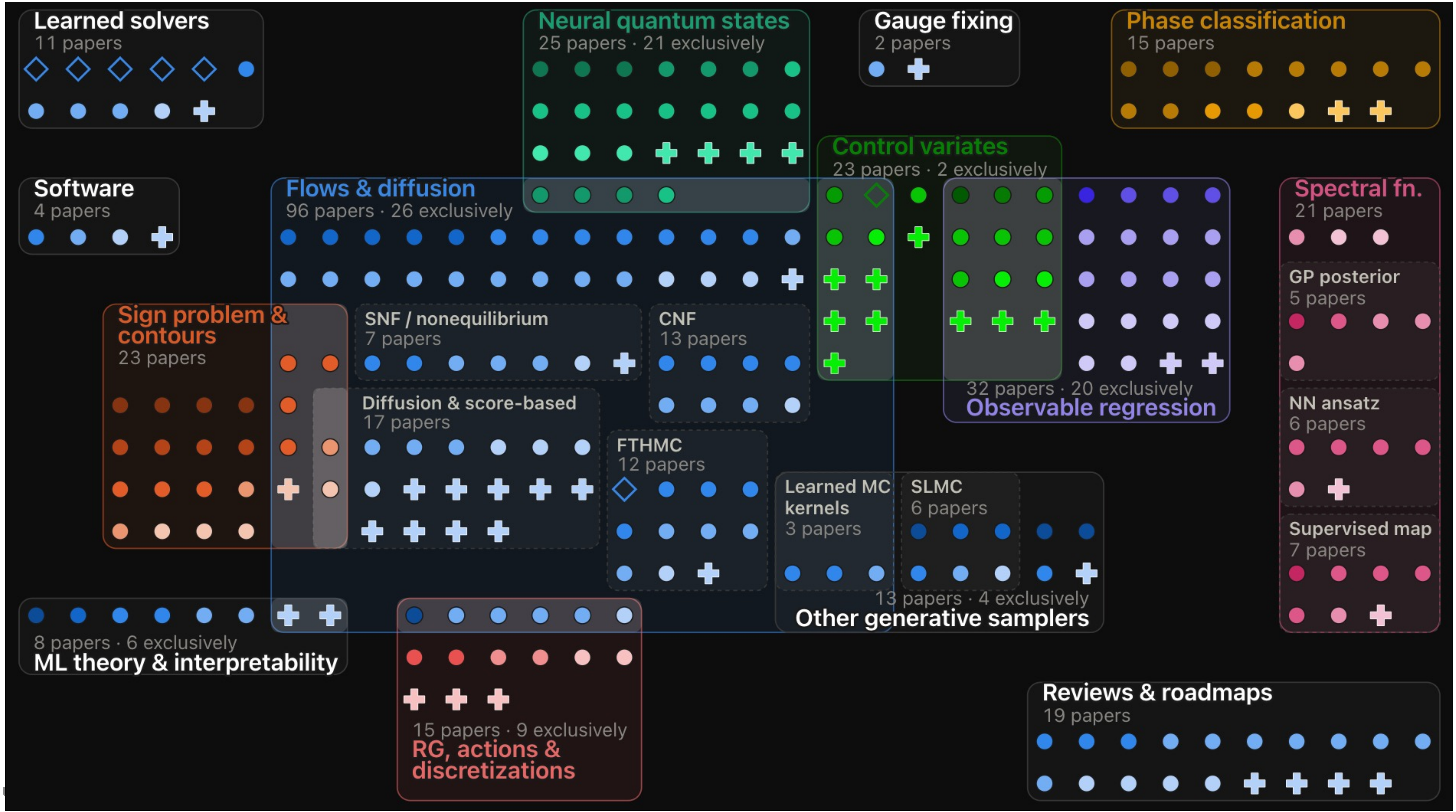
[364] S. Yamamoto (17:10) — Comparison of Smearing Kernels for Field-Transformation HMC via the Master-Field Technique

~ 38 presentations

Talks in Algorithms & AI unless *(noted otherwise)*

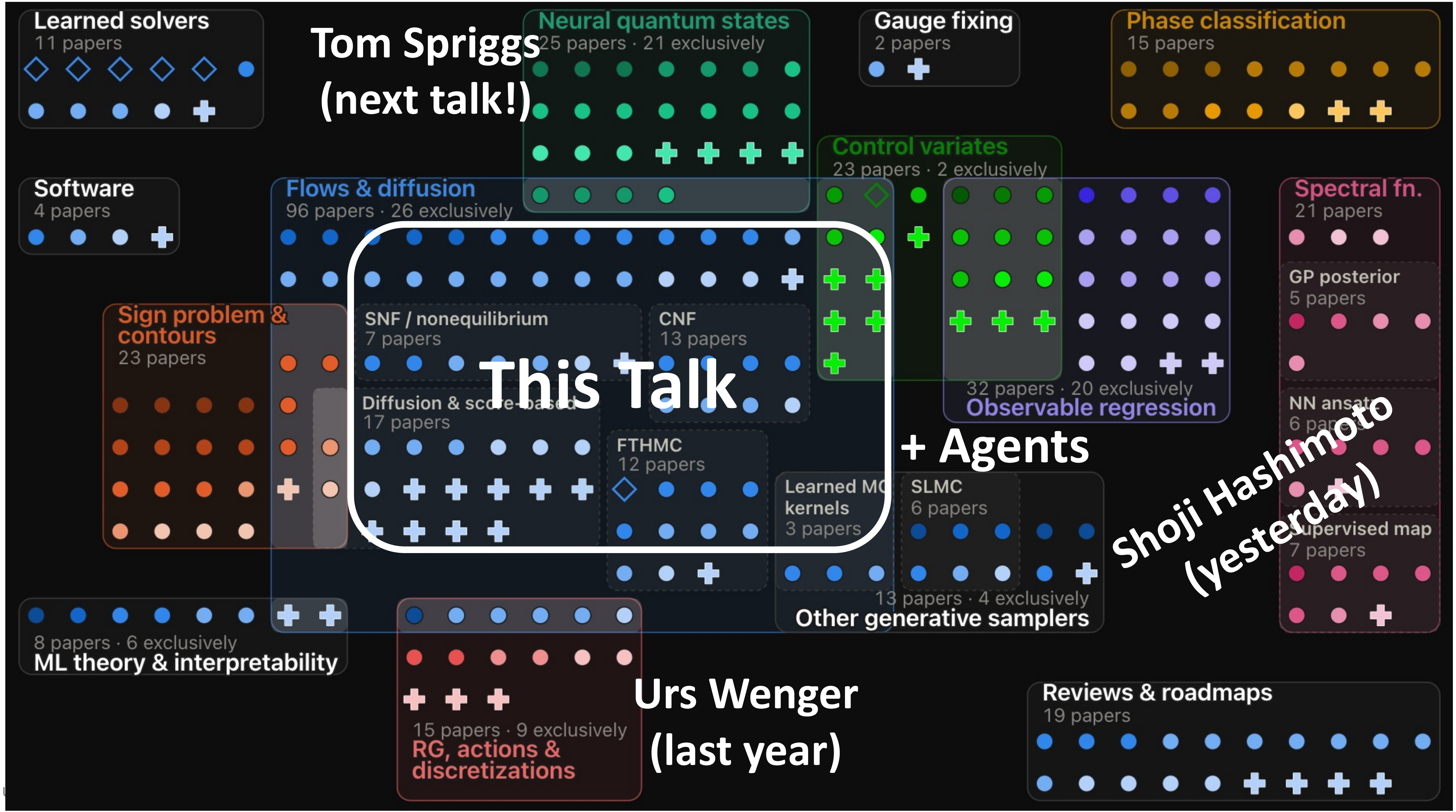
A Survey of the Field

~ 280 papers
+ : new since Lattice 2025



A Survey of the Field

~ 280 papers
+ : new since Lattice 2025



Outline

- Flows and exactness
  - …and challenges
- Applications of shorter flows
  - Hybrid methods (SNFs & T-REX)
  - OBC Defects
  - Flow-based variance reduction
- The quest for scalability
  - Diffusions, Transport, etc
- Agentic AI
- Outlook

Outline

Flows and exactness
- …and challenges

Applications of shorter flows
- Hybrid methods (SNFs & T-REX)
- OBC Defects
- Flow-based variance reduction

The quest for scalability
- Diffusions, Transport, etc

Agentic AI

Outlook

Emphasis: gauge theories, with an eye towards computational advantage in QCD calculations

In the end, cost is all that matters. Can we do rigorous QCD calculations cheaper with ML?

The State of the Art

Generate gauge fields with HMC

- One or few long MCMC chains
- Parallelize by forking streams (nontrivial decorrelation cost)
- Thermalization $\sim O(100 - 1000)$ trajectories
- Ensembles $\sim O(1000s)$ configs

Ensembles shared as community resources

The State of the Art

Generate gauge fields with HMC

- One or few long MCMC chains
- Parallelize by forking streams (nontrivial decorrelation cost)
- Thermalization $\sim O(100 - 1000)$ trajectories
- Ensembles $\sim O(1000s)$ configs

Ensembles shared as community resources

Pros:

- Well-established algo, lots of tricks / bells / whistles
- Highly optimized implementations
- Ensemble sharing/reuse amortizes generation cost

The State of the Art

Generate gauge fields with HMC

- One or few long MCMC chains
- Parallelize by forking streams (nontrivial decorrelation cost)
- Thermalization $\sim O(100-1000)$ trajectories
- Ensembles $\sim O(1000s)$ configs

Ensembles shared as community resources

Pros:

- Well-established algo, lots of tricks / bells / whistles
- Highly optimized implementations
- Ensemble sharing/reuse amortizes generation cost

Cons:

- **Big scale:** requires HPC, nontrivial parallelism
- **Serial:** long walltimes
- Difficult to tune hyperparameters
- Autocorrelations block fine a
  - Critical slowing down, topological freezing
- Poor signal-to-noise in many observables of interest

Disclaimer: throughout,
theory / action / distribution / density
used interchangeably

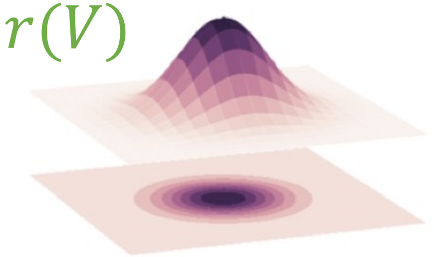
Recurring characters:
Target $p \propto e^{-S}$ (e.g. QCD)
Base/prior r
Model q

Flows and exactness

Normalizing* flows

* For QCD, should be "trivializing maps" or "Haarifying flows"

$r(V)$



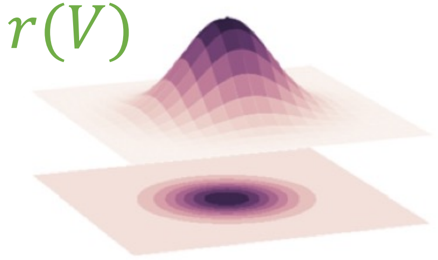
Base distribution

- Tractable density $r(V)$
- Exact sampling

e.g. Haar uniform ($\beta = 0$)

* For QCD, should be "trivializing maps" or "Haarifying flows"

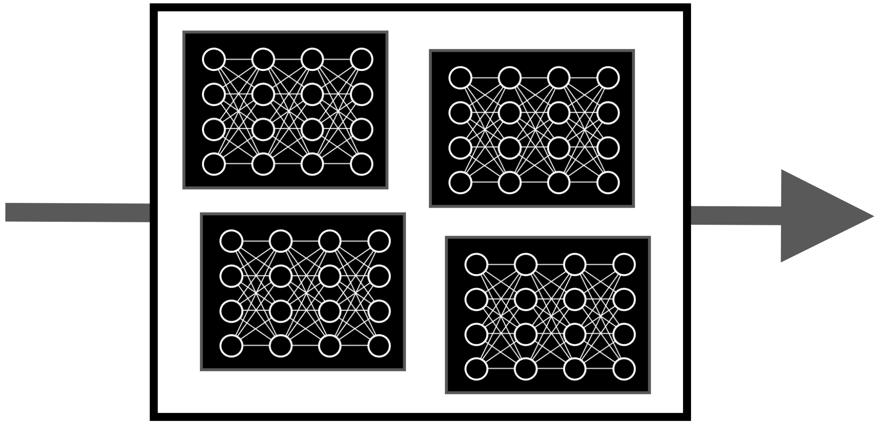
Normalizing* flows



Base distribution

- Tractable density $r(V)$
- Exact sampling

e.g. Haar uniform ($\beta = 0$)



"Flow" transformation

(a.k.a. a change of variables)

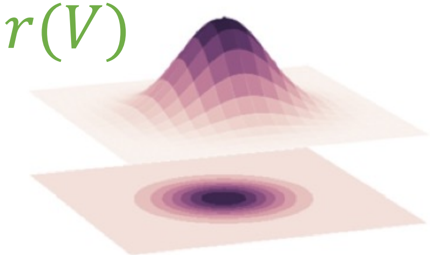
- *Parametrized* by NNs, but more structured
- Invertible (diffeomorphism!)
- Tractable Jacobian determinant

$$\det J = |\det \partial V / \partial U|$$

Normalizing* flows

* For QCD, should be "trivializing maps" or "Haarifying flows"

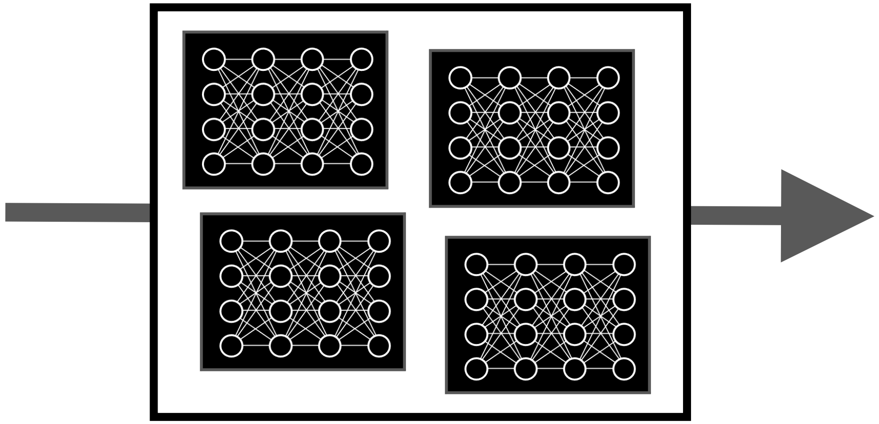
$r(V)$



Base distribution

- Tractable density $r(V)$
- Exact sampling

e.g. Haar uniform ($\beta = 0$)



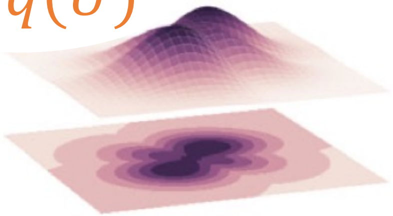
"Flow" transformation

(a.k.a. a change of variables)

- *Parametrized* by NNs, but more structured
- Invertible (diffeomorphism!)
- Tractable Jacobian determinant

$$\det J = |\det \partial V / \partial U|$$

$q(U)$



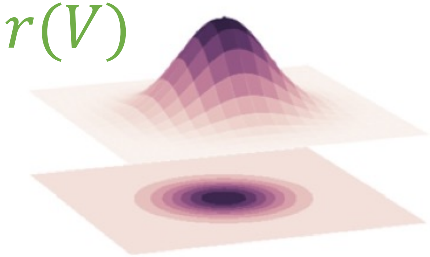
Learned **model distribution**

- Form is a function of flow params
- Exact sampling (inherit from base)
- Tractable density

$$q(U) = r(V) \det J$$

Normalizing* flows

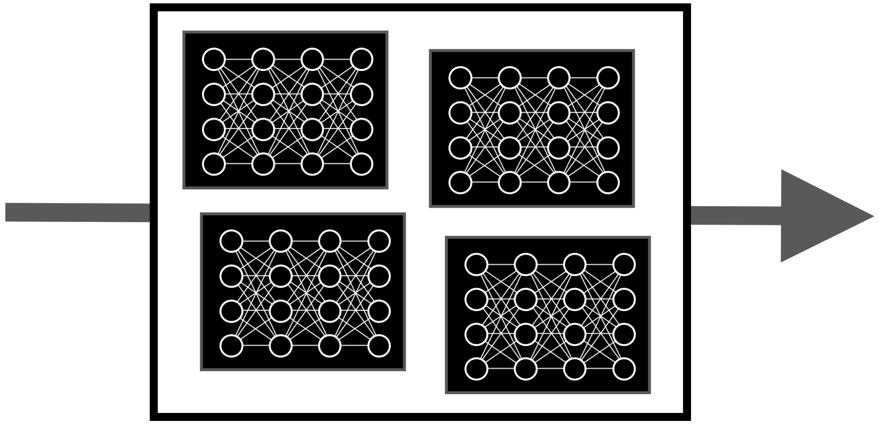
$r(V)$



Base distribution

- Tractable density $r(V)$
- Exact sampling

e.g. Haar uniform ($\beta = 0$)



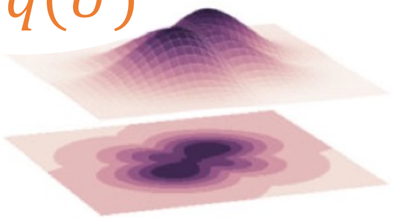
”Flow” transformation

(a.k.a. a change of variables)

- *Parametrized* by NNs, but more structured
- Invertible (diffeomorphism!)
- Tractable Jacobian determinant

$$\det J = |\det \partial V / \partial U|$$

$q(U)$



Learned model distribution

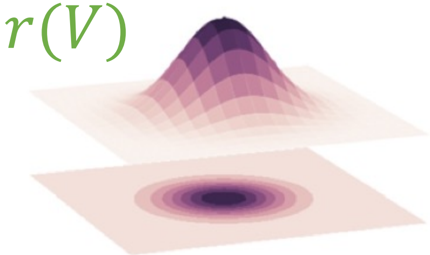
- Form is a function of flow params
- Exact sampling (inherit from base)
- Tractable density

$$\boxed{q(U) = r(V) \det J}$$

Enough info to reweight away biases from ML modeling!

Normalizing* flows

$r(V)$



Base distribution

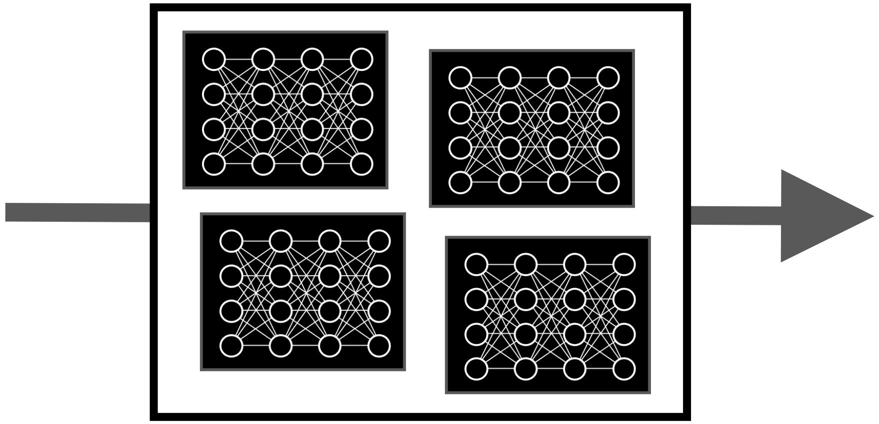
- Tractable density $r(V)$
- Exact sampling

e.g. Haar uniform ($\beta = 0$)

Modern variation:

More creative/opportunistic choices of base, e.g.

- Easier-to-sample theory
  e.g. low β, large m
- Unperturbed QCD in variance reduction setups



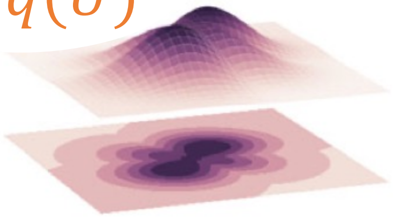
”Flow” transformation

(a.k.a. a change of variables)

- *Parametrized* by NNs, but more structured
- Invertible (diffeomorphism!)
- Tractable Jacobian determinant

$$\det J = |\det \partial V / \partial U|$$

$q(U)$



Learned **model distribution**

- Form is a function of flow params
- Exact sampling (inherit from base)
- Tractable density

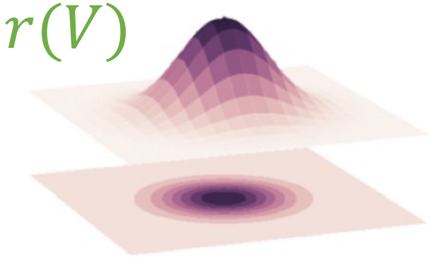
$$\boxed{q(U) = r(V) \det J}$$

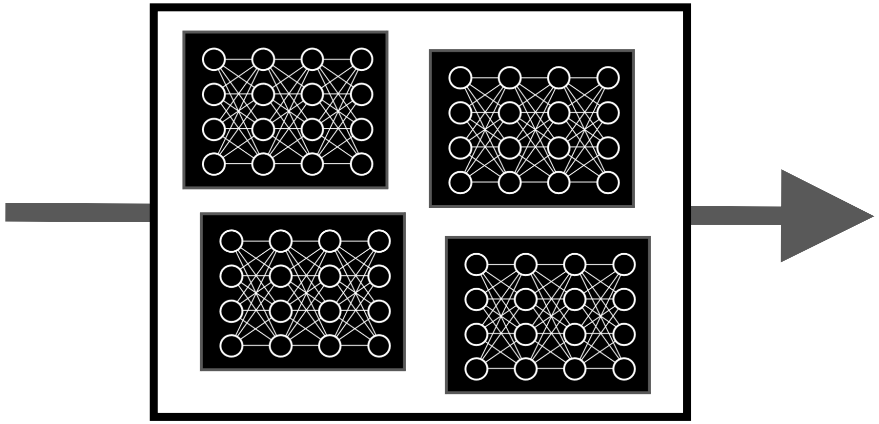
Enough info to reweight away biases from ML modeling!

* For QCD, should be “trivializing maps” or “Haarifying flows”

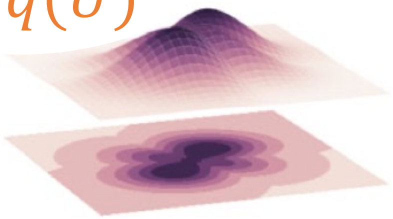
Normalizing* flows

$r(V)$





$q(U)$



Base distribution

- Tractable density $r(V)$
- Exact sampling

e.g. Haar uniform ($\beta = 0$)

Modern variation:

More creative/opportunistic choices of base, e.g.

- Easier-to-sample theory
  e.g. low β, large m
- Unperturbed QCD in variance reduction setups

”Flow” transformation

(a.k.a. a change of variables)

- *Parametrized* by NNs, but more structured
- Invertible (diffeomorphism!)
- Tractable Jacobian determinant

$$\det J = |\det \partial V / \partial U|$$

Modern variation: stochastic flows

q is intractable, but unbiased stochastic reweighting via Jarzynski and generalizations

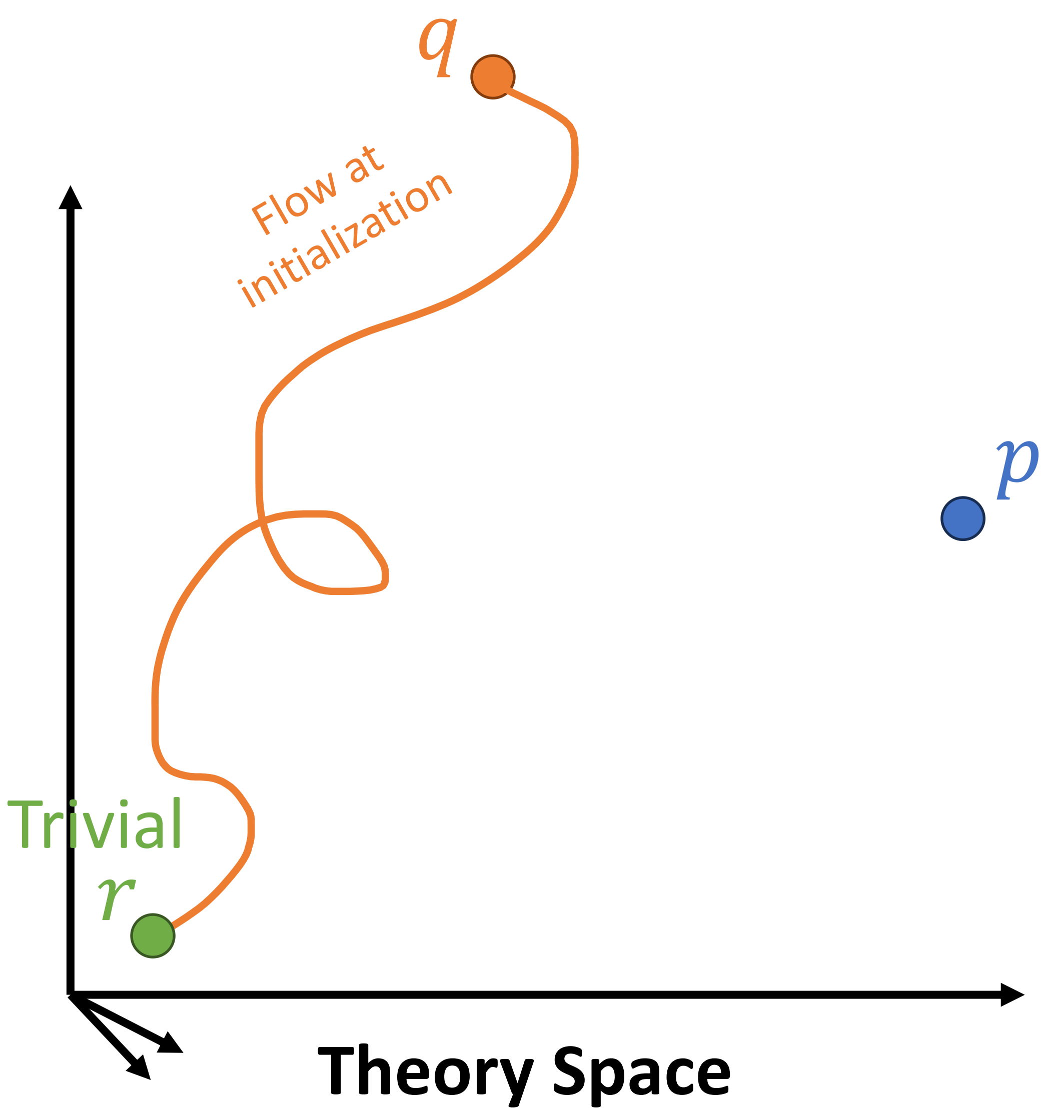
Learned model distribution

- Form is a function of flow params
- Exact sampling (inherit from base)
- Tractable density

$$\boxed{q(U) = r(V) \det J}$$

Enough info to reweight away biases from ML modeling!

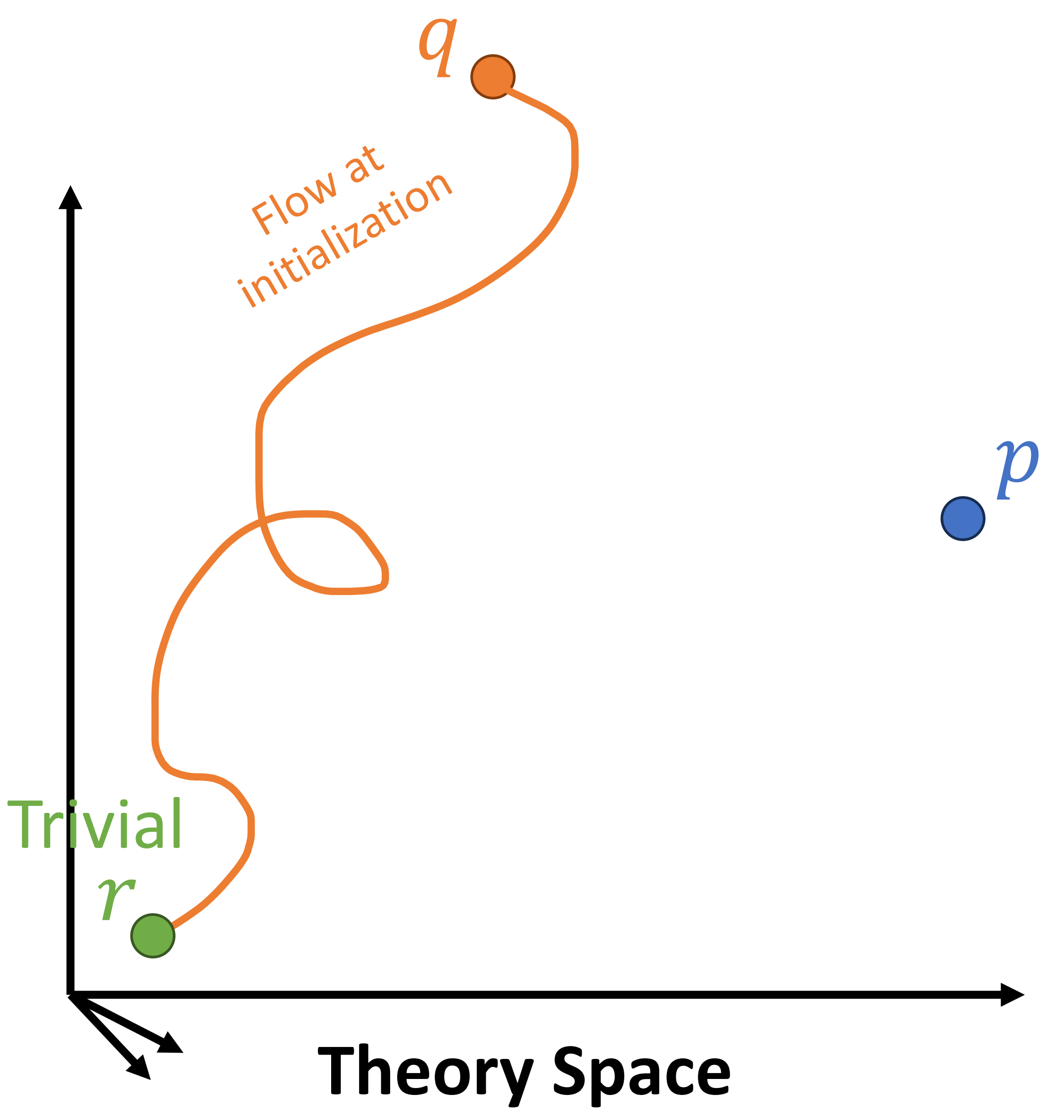
Flow model samplers



Step 1: write down some flow model ansatz q

- Flow traces a path through theory space
- q is a function of parameters in the flow

Flow model samplers

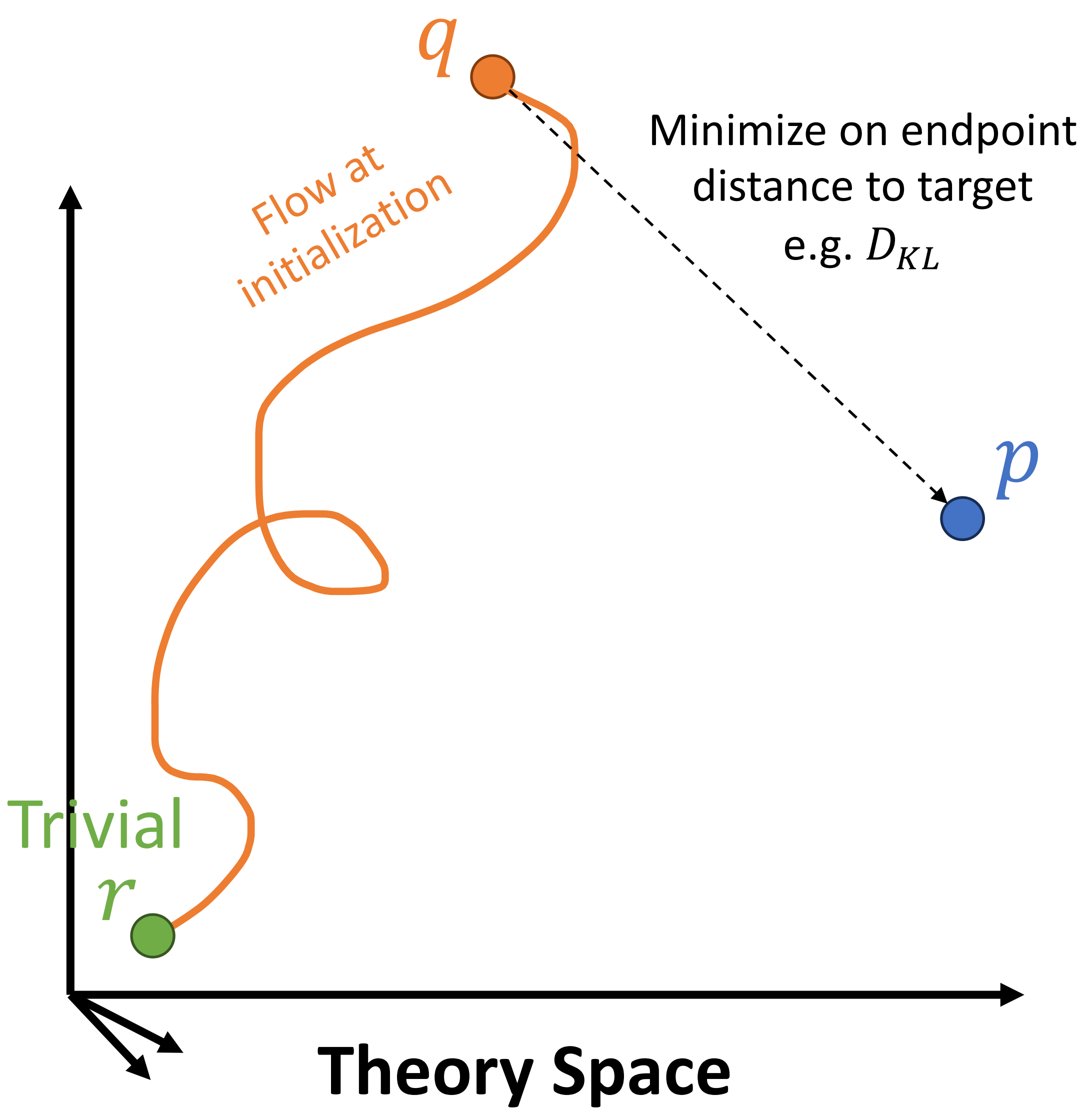


Step 1: write down some flow model ansatz q

- Flow traces a path through theory space
- q is a function of parameters in the flow

Step 2: train the flow so $q \approx p \propto e^{-S}$

Flow model samplers

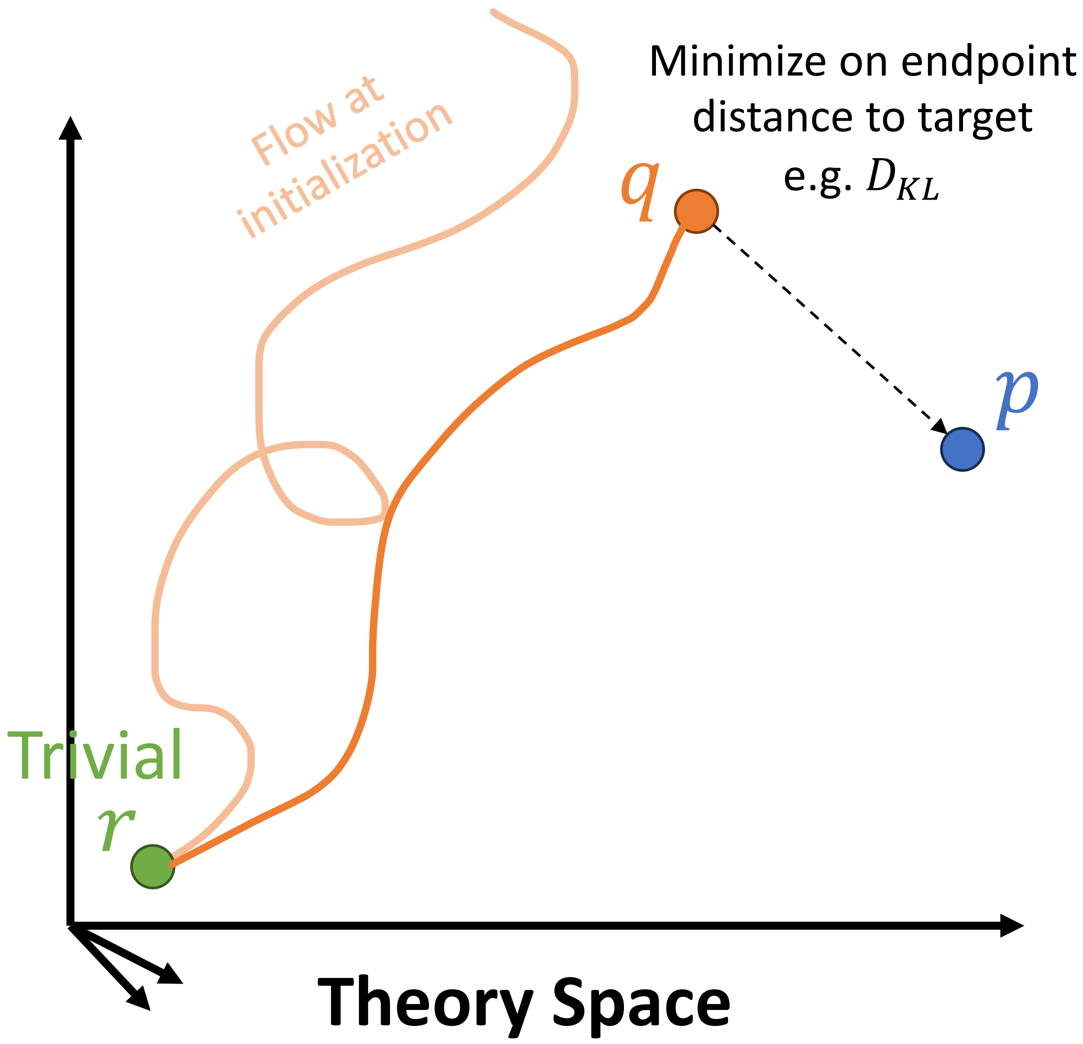


Step 1: write down some flow model ansatz q

- Flow traces a path through theory space
- q is a function of parameters in the flow

Step 2: train the flow so $q \approx p \propto e^{-S}$

Flow model samplers

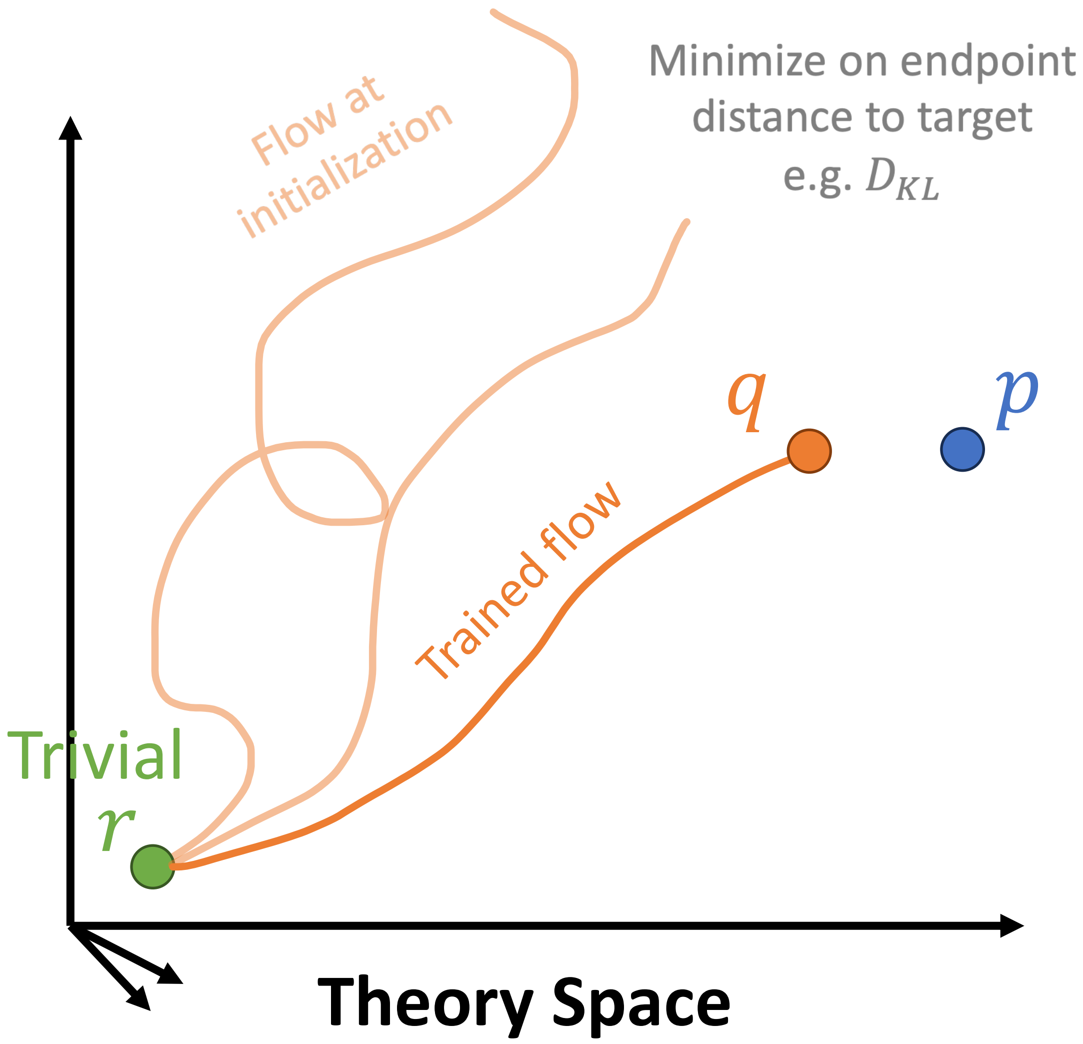


Step 1: write down some flow model ansatz q

- Flow traces a path through theory space
- q is a function of parameters in the flow

Step 2: train the flow so $q \approx p \propto e^{-S}$

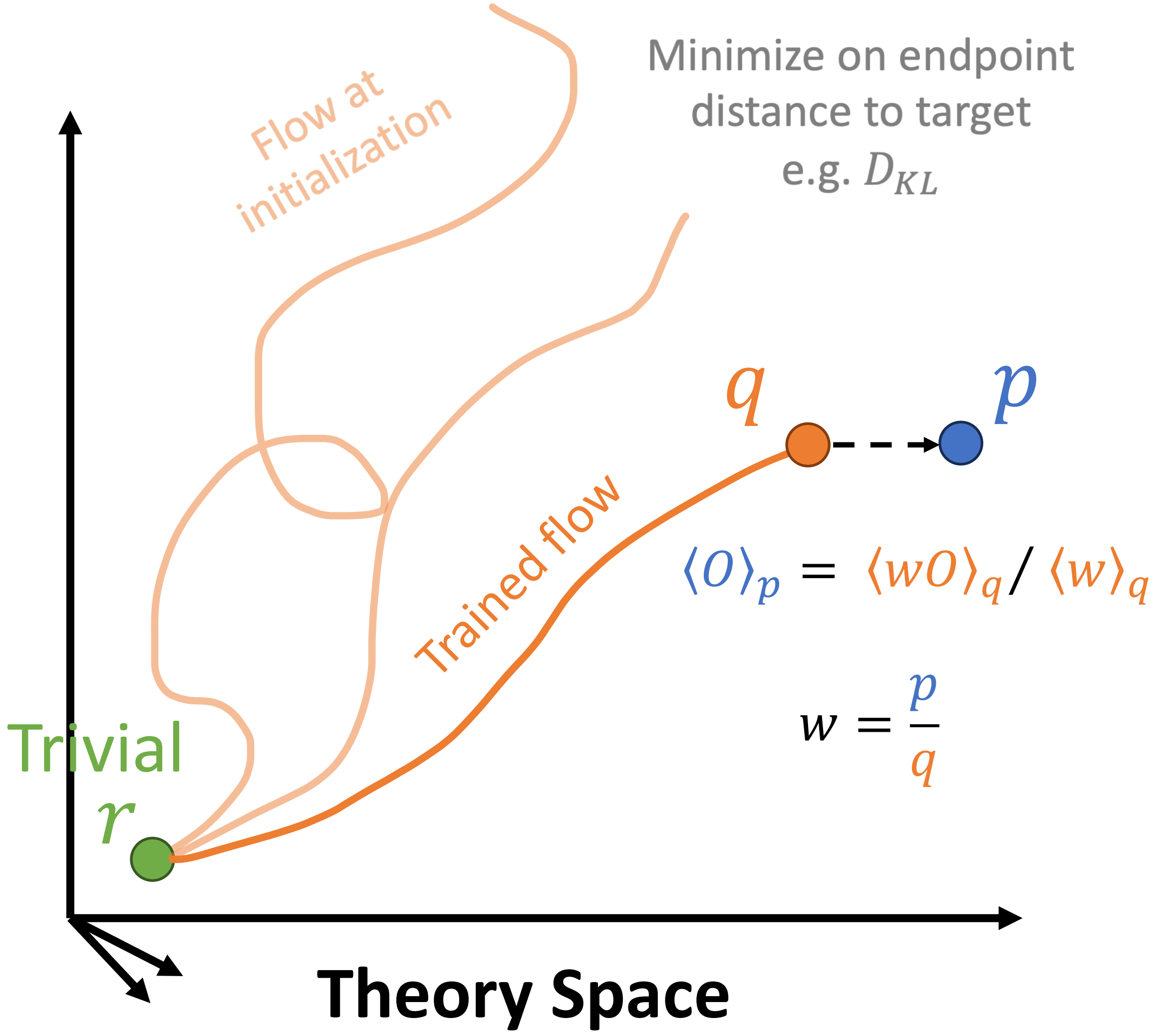
Flow model samplers



Step 1: write down some flow model ansatz q
- Flow traces a path through theory space
- q is a function of parameters in the flow

Step 2: train the flow so $q \approx p \propto e^{-S}$

Flow model samplers



Step 1: write down some flow model ansatz q

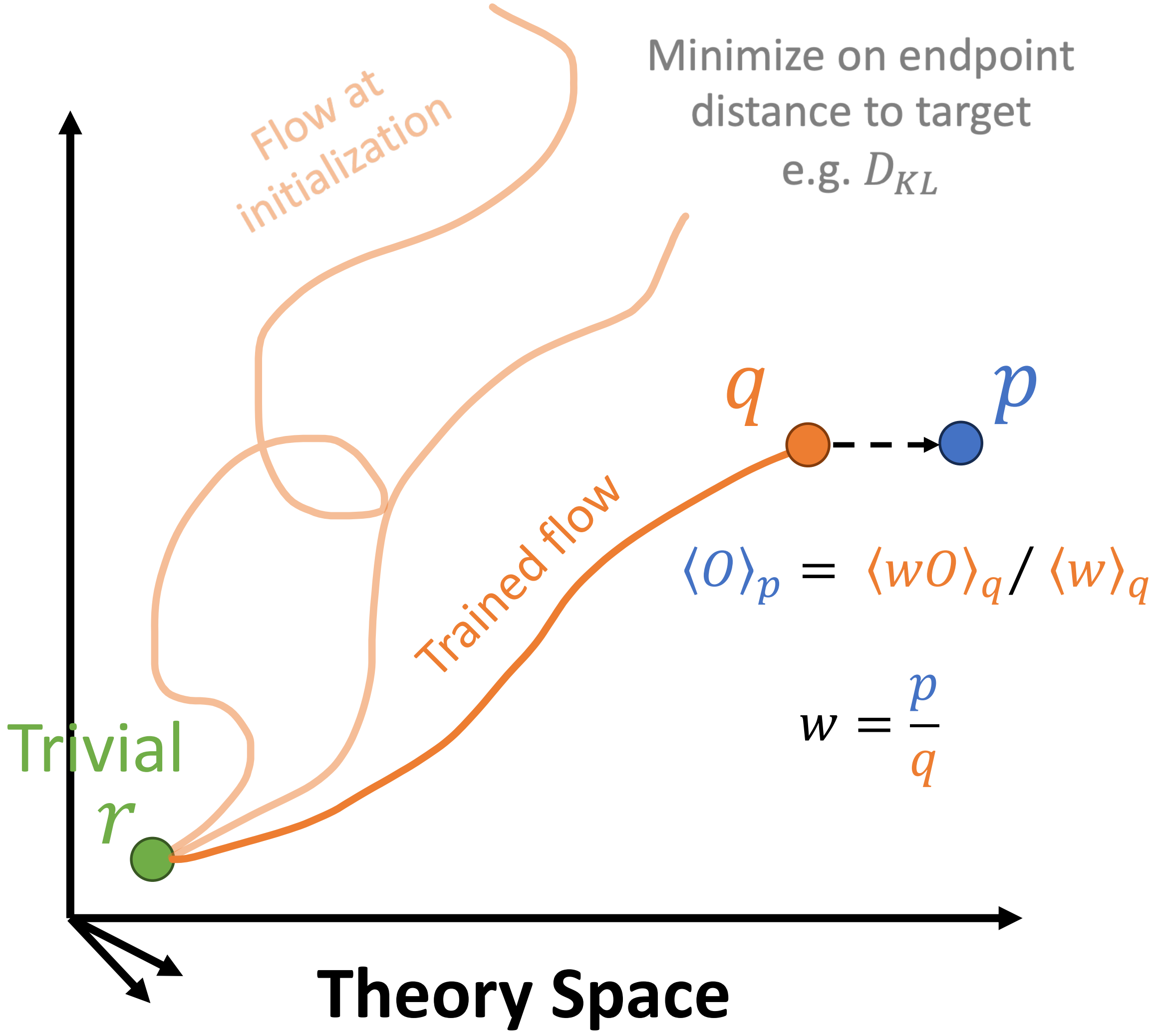
- Flow traces a path through theory space
- q is a function of parameters in the flow

Step 2: train the flow so $q \approx p \propto e^{-S}$

Step 3: compute **exact** expectations in target theory by reweighting*

* Can also do "Independence Metropolis" global proposals $\sim$ equivalent to reweighting.

Flow model samplers



Step 1: write down some flow model ansatz q

- Flow traces a path through theory space
- q is a function of parameters in the flow

Step 2: train the flow so $q \approx p \propto e^{-S}$

Step 3: compute **exact** expectations in target theory by reweighting*

Independent samples from q
→ No autocorrelations

Cost: extra variance from reweighting

$$\text{ESS} = \frac{1}{\langle w^2 \rangle_q} \sim \frac{1}{\text{Increase in variance due to reweighting}}$$

Note: Quality of training does not change exactness, just sampler efficiency

* Can also do "Independence Metropolis" global proposals ~ equivalent to reweighting.

The road to QCD

Flows for LQFT (scalar field theories)

[Albergo, Kanwar, Shanahan 1904.12072]

[**DH**, Hsieh, Pontula, Albergo, Boyda, JW Chen, KF Chen, Cranmer, Kanwar, Shanahan 2107.00734]

Gauge-equivariant flows

U(1) [Kanwar, Albergo, Boyda, Cranmer, **DH**, Racanière, Rezende, Shanahan 2003.06413]

SU(N) [Boyda Kanwar Racanière Rezende Albergo Cranmer DH Shanahan 2008.05456]

[Abbott Albergo Botev Boyda Cranmer **DH** Kanwar Matthews Racanière Razavi Rezende Romero-López Shanahan Urban 2305.02402]

Flows for fermions

[Albergo, Kanwar, Racanière, Rezende, Urban, Boyda, Cranmer, **DH**, Shanahan 2106.05934]

[Albergo, Boyda, Cranmer, **DH**, Kanwar, Racanière, Rezende, Romero-López, Shanahan, Urban 2202.11712]

[Abbott, Albergo, Boyda, Cranmer, **DH**, Kanwar, Racanière, Rezende, Romero-López, Shanahan, Tian, Urban 2207.08945]

Scaling properties

[Abbott, Albergo, Botev, Boyda, Cranmer, **DH**, Matthews, Racanière, Razavi, Rezende, Romero-López, Shanahan, Urban 2211.07541]

Applications to QCD (+ more this talk)

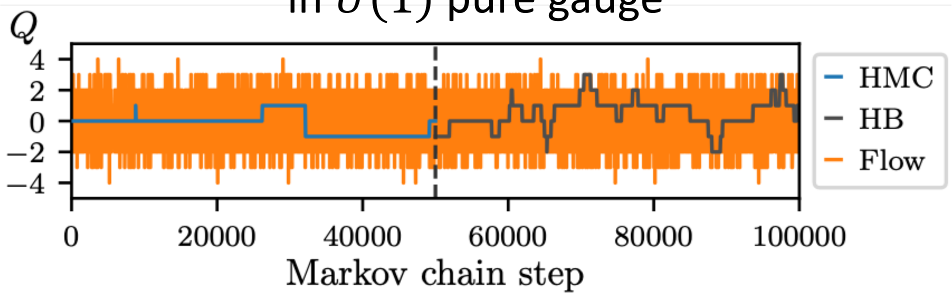
[Abbott, Albergo, Botev, Boyda, Cranmer, **DH**, Kanwar, Matthews, Racanière, Razavi, Rezende, Romero-López, Shanahan, Urban 2208.03832]

[Abbott Botev Boyda **DH** Kanwar Racanière Rezende Romero-López Shanahan Urban 2401.10874]

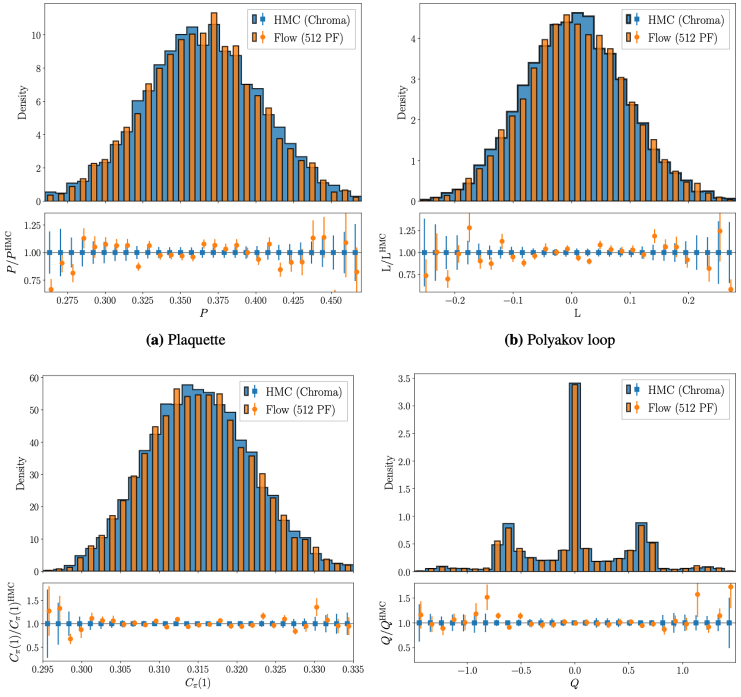
[Abbott, Albergo, Boyda, **DH**, Kanwar, Romero-López, Shanahan, Urban 2404.11674]

[Abbott, Boyda, Kanwar, Romero-López, **DH**, Shanahan, Urban 2502.00263]

Better topological sampling in $U(1)$ pure gauge



First demonstration of a QCD flow model



(a) Plaquette

(b) Polyakov loop

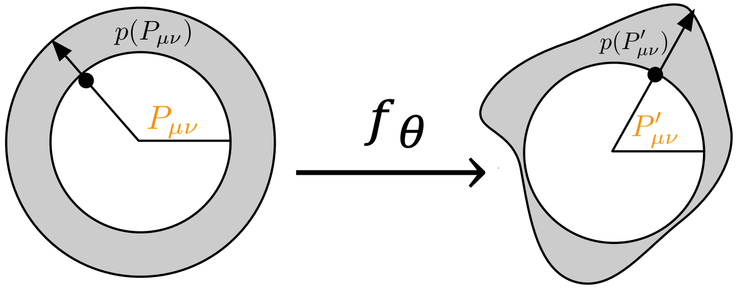
(c) Pion correlation function at $x_0 = 1$

(d) Topological charge at $t/a^2 = 4$

Along the road

Flows on compact variables

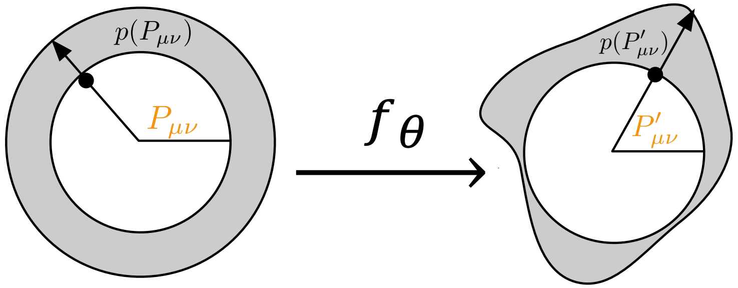
Structural necessity: must respect Lie group structure



Along the road

Flows on compact variables

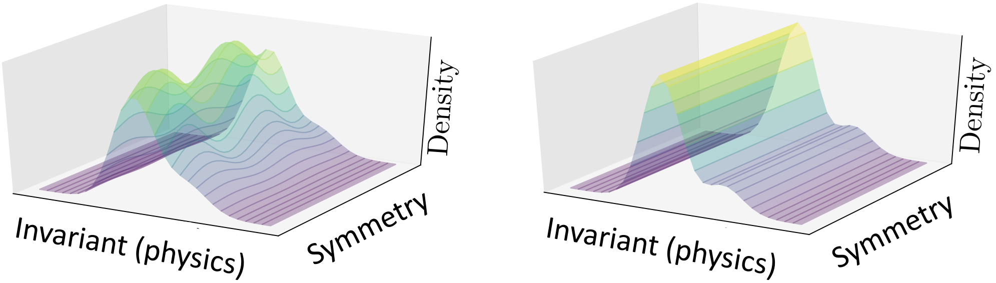
Structural necessity: must respect Lie group structure



Gauge-equivariant flows

Practical necessity: mismodeling high-dimensional symmetry kills ESS

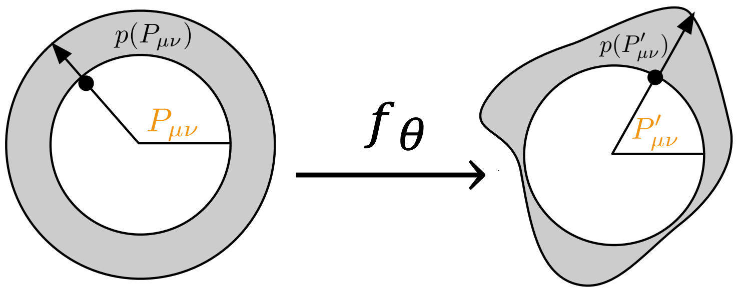
$$f_\theta\big(g(U)\big) = g(f_\theta(U))$$



Along the road

Flows on compact variables

Structural necessity: must respect Lie group structure



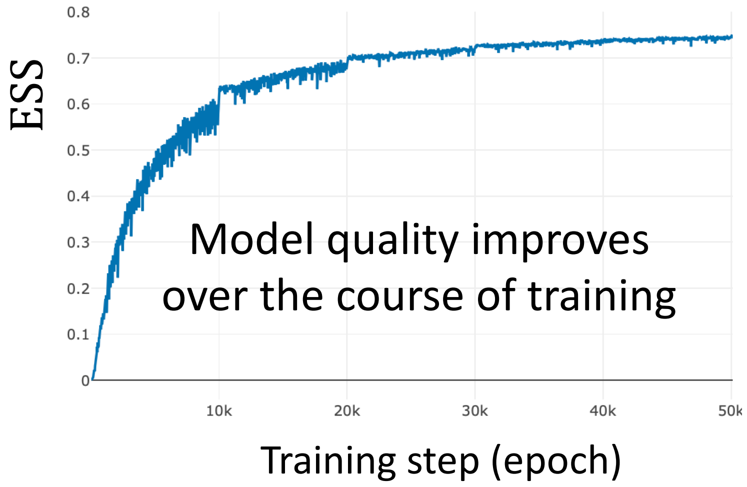
Reverse KL self-training

Minimize

$$D_{KL}(q||p) = \left\langle \log \frac{q}{p} \right\rangle_q$$

Only model samples to train

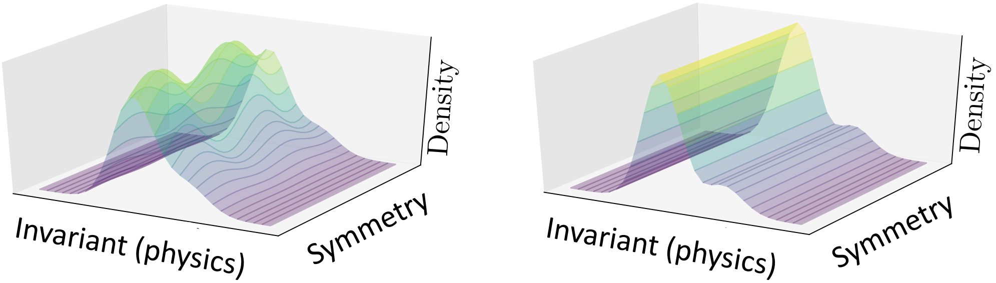
→ ~ infinite training data



Gauge-equivariant flows

Practical necessity: mismodeling high-dimensional symmetry kills ESS

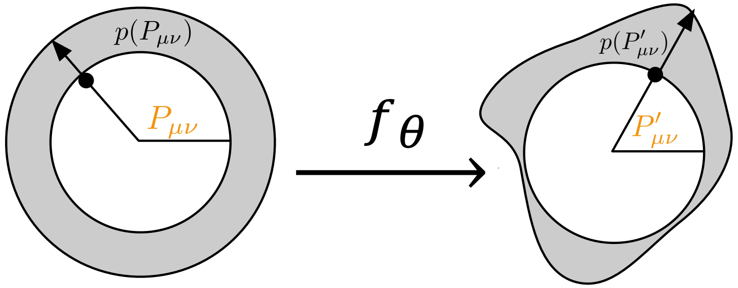
$$f_\theta\big(g(U)\big) = g\big(f_\theta(U)\big)$$



Along the road

Flows on compact variables

Structural necessity: must respect Lie group structure



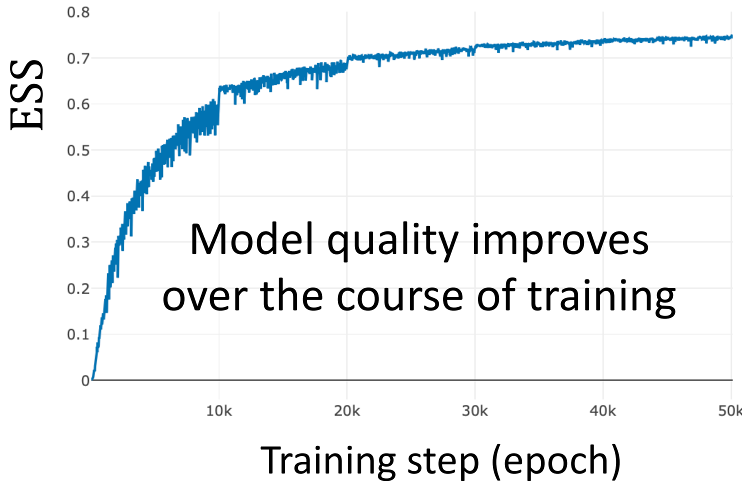
Reverse KL self-training

Minimize

$$D_{KL}(q||p) = \left\langle \log \frac{q}{p} \right\rangle_q$$

Only model samples to train

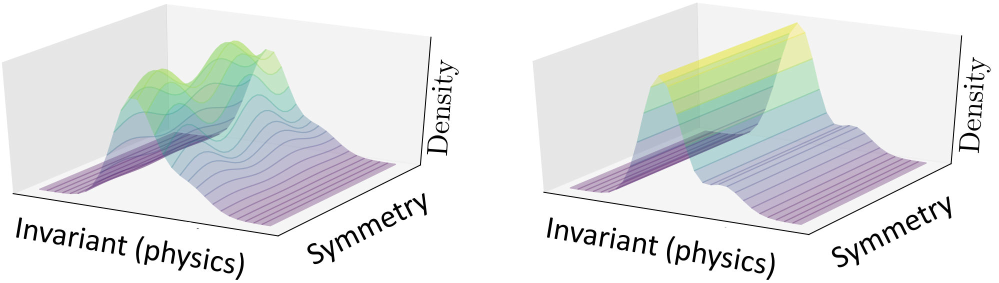
→ ~ infinite training data



Gauge-equivariant flows

Practical necessity: mismodeling high-dimensional symmetry kills ESS

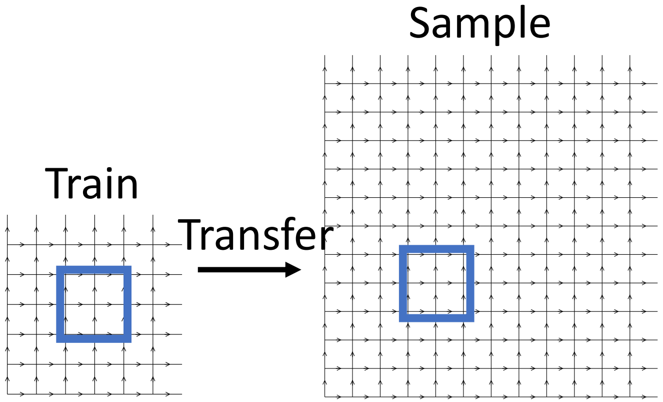
$$f_\theta\big(g(U)\big) = g\big(f_\theta(U)\big)$$



Convolutional models

Same flow works on different lattice geometries

→ **Volume transfer:** pay training costs $\sim L^4$ on smaller lattices than sampling target



$$-\log q = S_q \sim \sum_x \mathcal{L}_x(U)$$

Like any other lattice action, but w/ a built-in exact sampler

Flows on gauge theories

Spectral flows

[Kanwar Albergo Boyda Cranmer **DH** Racanière Rezende Shanahan 2003.06413]

[Boyda Kanwar Racanière Rezende Albergo Cranmer **DH** Shanahan 2008.05456]

[Komijani Marinkovic 2501.18288]

[Abbott Boyda Kanwar Romero-López **DH** Shanahan Urban 2502.00263]

Residual flows (learned gradient flow + masking)

[Abbott Albergo Botev Boyda Cranmer **DH** Kanwar Matthews Racanière Razavi Rezende Romero-López Shanahan Urban 2305.02402]

Gauge CNFs (learned gradient flow)

[Bacchio Kessel Schaefer Vaitl 2212.08469]

[Gerdes de Haan Bondesan Cheng 2410.13161]

Pseudofermion flows

[Abbott Albergo Boyda Cranmer **DH** Kanwar Racanière Rezende Romero-López Shanahan Tian Urban 2207.08945]

Hierarchical / multiscale flows

[Abbott Albergo Boyda **DH** Kanwar Romero-López Shanahan Urban 2404.10819]

Localized flows

[Finkenrath 2201.02216]

[Finkenrath 2402.12176]

[Abbott Albergo Boyda **DH** Kanwar Romero-López Shanahan Urban 2404.11674]

[Bonanno Bulgarelli Cellini Nada Panfalone Vadacchino Verzichelli 2510.25704]

Flows on gauge theories

Spectral flows

[Kanwar Albergo Boyda Cranmer **DH** Racanière Rezende Shanahan 2003.06413]
[Boyda Kanwar Racanière Rezende Albergo Cranmer **DH** Shanahan 2008.05456]
[Komijani Marinkovic 2501.18288]
[Abbott Boyda Kanwar Romero-López **DH** Shanahan Urban 2502.00263]

Residual flows (learned gradient flow + masking)

[Abbott Albergo Botev Boyda Cranmer **DH** Kanwar Matthews Racanière Razavi Rezende Romero-López Shanahan Urban 2305.02402]

Gauge CNFs (learned gradient flow)

[Bacchio Kessel Schaefer Vaitl 2212.08469]
[Gerdes de Haan Bondesan Cheng 2410.13161]

Pseudofermion flows

[Abbott Albergo Boyda Cranmer **DH** Kanwar Racanière Rezende Romero-López Shanahan Tian Urban 2207.08945]

Hierarchical / multiscale flows

[Abbott Albergo Boyda **DH** Kanwar Romero-López Shanahan Urban 2404.10819]

Localized flows

[Finkenrath 2201.02216]
[Finkenrath 2402.12176]
[Abbott Albergo Boyda **DH** Kanwar Romero-López Shanahan Urban 2404.11674]
[Bonanno Bulgarelli Cellini Nada Panfalone Vadacchino Verzichelli 2510.25704]

+ Gauge-equivariant architectures more generally

Parallel-transport CNNs

[Favoni Ipp Müller Schuh 2012.12901] (L-CNN)
[Lehner Wettig 2302.05419] (PTC/LPTC)
[Aronsson Müller Schuh 2303.11448] (L-CNN+)
[Lehner Wettig 2304.10438] (equiv. pooling)
[Pfahler Knüttel Lehner Wettig 2602.23840] (long-path PTC)
[Abbott Albergo Boyda Cranmer **DH** Kanwar Racanière Rezende Romero-López Shanahan Tian Urban 2207.08945] (PTCN)
[Komijani Marinkovic Turgut 2605.06134] (GaugeLinkConv)

Learned smearing

[Tomiya Nagai 2103.11965]
[Bellscheidt Brambilla Kronfeld Mayer-Steudte 2602.02436]
[Hsiao Choi Ohno Tomiya 2602.23731] (WLB layer)
[Jin 2201.01862] (NTHMC)
[Jin 2405.19692] (NTHMC, 4D SU(3))
[He Jin Osborn Zhao 2511.02018] (NTHMC CNN design)

Gauge-covariant transformers

[Nagai Ohno Tomiya 2501.16955] (CASK)

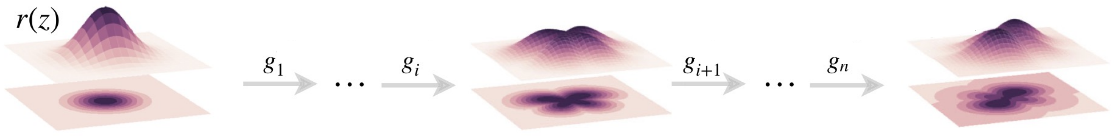
Gauge-equivariant GNNs

[Rayat Li Chern 2604.20797]

Flow architectures

Discrete flows: coupling layers

Build a flow by stacking many simple sub-flows



Variable partitioning:

Freeze some links

Update others (active links)

  …independently

  …conditioned on frozen links

$\phi =$  $+$ 

Active ϕ_A

Frozen ϕ_F

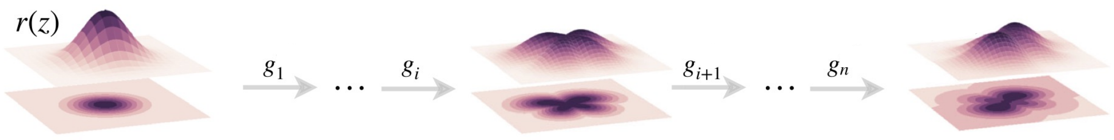
→ Triangular Jacobian J → easy to compute $\det J$

Alternate which links are masked per layer.

Flow architectures

Discrete flows: coupling layers

Build a flow by stacking many simple sub-flows



Variable partitioning:

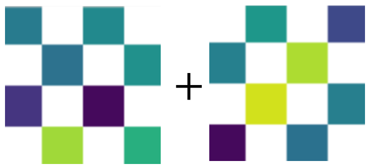
Freeze some links

Update others (active links)

…independently

…conditioned on frozen links

$\phi =$



Active ϕ_A

Frozen ϕ_F

→ Triangular Jacobian J → easy to compute det J

Alternate which links are masked per layer.

Spectral flows

Flow eigenvalue spectrum of open loop containing active link

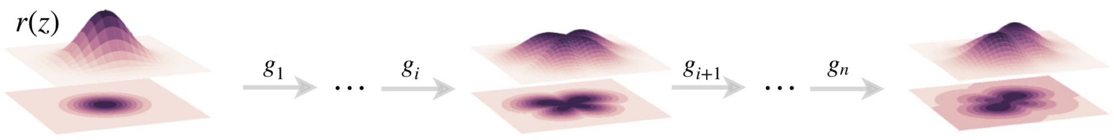
Residual flows

Masked hits of gradient* flow with a learned action

Flow architectures

Discrete flows: coupling layers

Build a flow by stacking many simple sub-flows



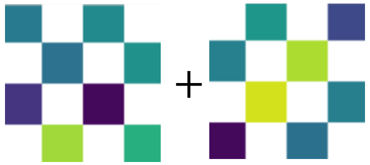
Variable partitioning:

Freeze some links

Update others (active links)

…independently

…conditioned on frozen links

$\phi =$ 

Active ϕ_A + Frozen ϕ_F

→ Triangular Jacobian J → easy to compute det J

Alternate which links are masked per layer.

Spectral flows

Flow eigenvalue spectrum of open loop containing active link

Residual flows

Masked hits of gradient* flow with a learned action

Continuous flows: learned gradient* flows

Flow by integrating ODE from $t = 0 \to 1$

$$\dot{U}_{x\mu}(t) = Z_{x\mu}(U(t), t)\, U_{x\mu}(t)$$

where Z is learned, $su(N)$ algebra-valued

Evaluate q by integrating conservation of probability

$$\partial_t q + \nabla \cdot (q\, Z) = 0$$

$$\Rightarrow \frac{d}{dt} \log q(t) = -\nabla \cdot Z(t)$$

Generalizes Luscher's gradient* flow for a learned, time-dependent action $\tilde{S}_\theta(t)$

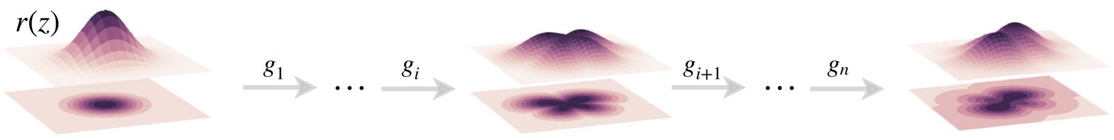
→ Automatically encode *all* symmetries of the theory by choosing $\tilde{S}_\theta(t)$ invariant

* Can use a more general $Z \neq \nabla \tilde{S}_\theta(t)$

Flow architectures

Discrete flows: coupling layers

Build a flow by stacking many simple sub-flows



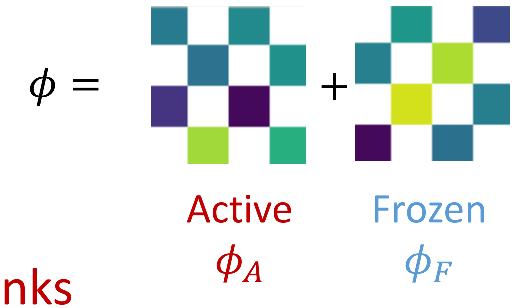
Variable partitioning:

Freeze some links

Update others (active links)

…independently

…conditioned on frozen links



→ Triangular Jacobian J → easy to compute $\det J$

Alternate which links are masked per layer.

Spectral flows

Flow eigenvalue spectrum of open loop containing active link

Residual flows

Masked hits of gradient* flow with a learned action

Continuous flows: learned gradient* flows

Flow by integrating ODE from $t = 0 \to 1$

$$\dot{U}_{x\mu}(t) = Z_{x\mu}(U(t), t)\, U_{x\mu}(t)$$

where Z is learned, $su(N)$ algebra-valued

Evaluate q by integrating conservation of probability

$$\partial_t\, q + \nabla \cdot (q\, Z) = 0$$

$$\Rightarrow \frac{d}{dt} \log q(t) = -\nabla \cdot Z(t)$$

Generalizes Luscher's gradient* flow for a learned, time-dependent action $\tilde{S}_\theta(t)$

→ Automatically encode *all* symmetries of the theory by choosing $\tilde{S}_\theta(t)$ invariant

* Can use a more general $Z \neq \nabla \tilde{S}_\theta(t)$

Arbitrary Z

Estimate $\nabla \cdot Z$ stochastically with Hutchinson

No expressivity limitations

Noise hurts ESS, induces bias

Restricted loop bases

$Z = \nabla[\text{sum over Wilson loops}]$

→ Can evaluate $\nabla \cdot Z$ analytically, cheaply

Limited expressivity

Fermions

Model learns distribution
including fermionic effects

$$q(U) \approx p(U) \propto e^{-S_g} \det D^\dagger D$$

[Albergo Kanwar Racanière Rezende Urban Boyda Cranmer DH Shanahan 2106.05934]

[Albergo Boyda Cranmer DH Kanwar Racanière Rezende Romero-López Shanahan Urban 2202.11712]

[Abbott Albergo Boyda Cranmer DH Kanwar Racanière Rezende Romero-López Shanahan Tian Urban 2207.08945]

Fermions

Model learns distribution
including fermionic effects

$$q(U) \approx p(U) \propto e^{-S_g} \det D^\dagger D$$

Training

- Smaller lattices: compute & backprop through exact determinant
- Larger lattices: stochastic gradient estimates

[Albergo Kanwar Racanière Rezende Urban Boyda Cranmer DH Shanahan 2106.05934]

[Albergo Boyda Cranmer DH Kanwar Racanière Rezende Romero-López Shanahan Urban 2202.11712]

[Abbott Albergo Boyda Cranmer DH Kanwar Racanière Rezende Romero-López Shanahan Tian Urban 2207.08945]

Fermions

Model learns distribution
including fermionic effects

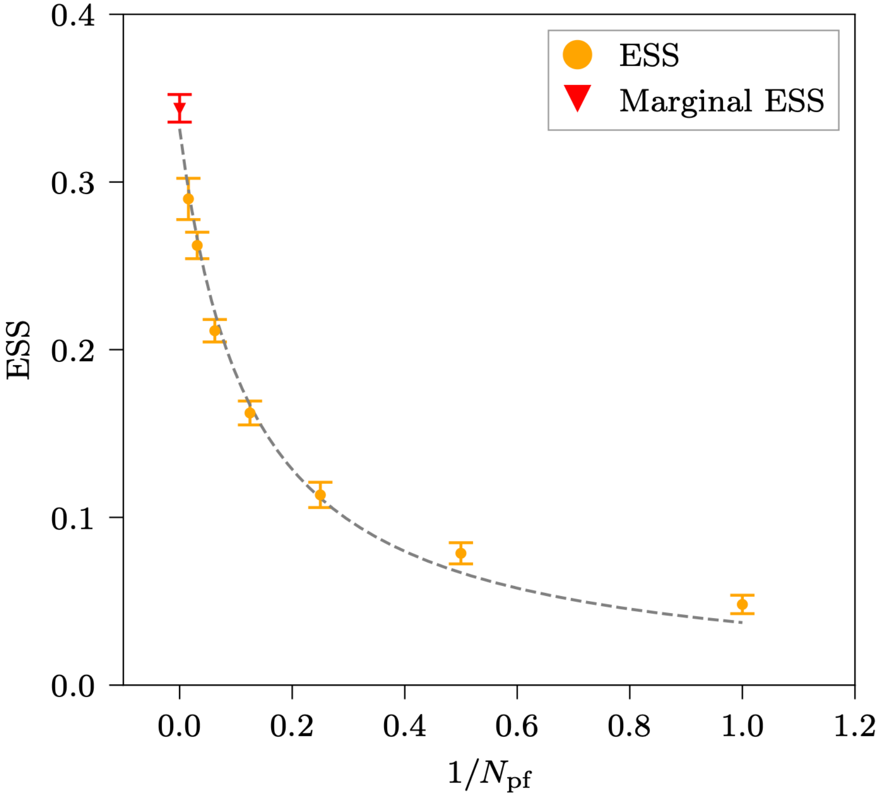
$$q(U) \approx p(U) \propto e^{-S_g} \det D^\dagger D$$

Training

- Smaller lattices: compute & backprop through exact determinant
- Larger lattices: stochastic gradient estimates

Sampling

- Just need unbiased estimate of p
- Use stochastic det estimators
- Improve variance with "pseudofermion flows"
  learned, easy-to-invert D^{-1}
- Similar cost to disconnected diagram calculation
- Pay once per configuration



[Albergo Kanwar Racanière Rezende Urban Boyda Cranmer DH Shanahan 2106.05934]

[Albergo Boyda Cranmer DH Kanwar Racanière Rezende Romero-López Shanahan Urban 2202.11712]

[Abbott Albergo Boyda Cranmer DH Kanwar Racanière Rezende Romero-López Shanahan Tian Urban 2207.08945]

Present Obstacles

Two compounding issues presently block construction of viable QCD flow samplers at scale

Present Obstacles

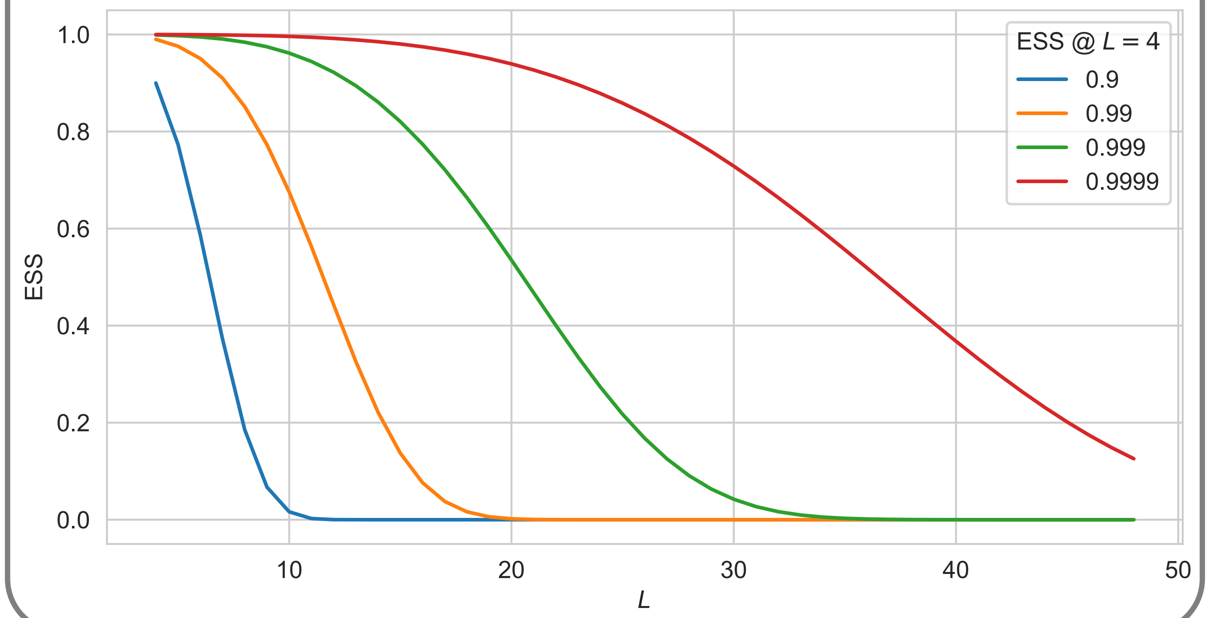
Two compounding issues presently block construction of viable QCD flow samplers at scale

1. Volume scaling of reweighting

Under direct volume transfer,

$$\mathrm{ESS}(V) \approx \mathrm{ESS}(V_0)^{V/V_0}$$

- Many caveats
- **Practical implication:** need excellent *local* modeling to do *global* reweighting



Present Obstacles

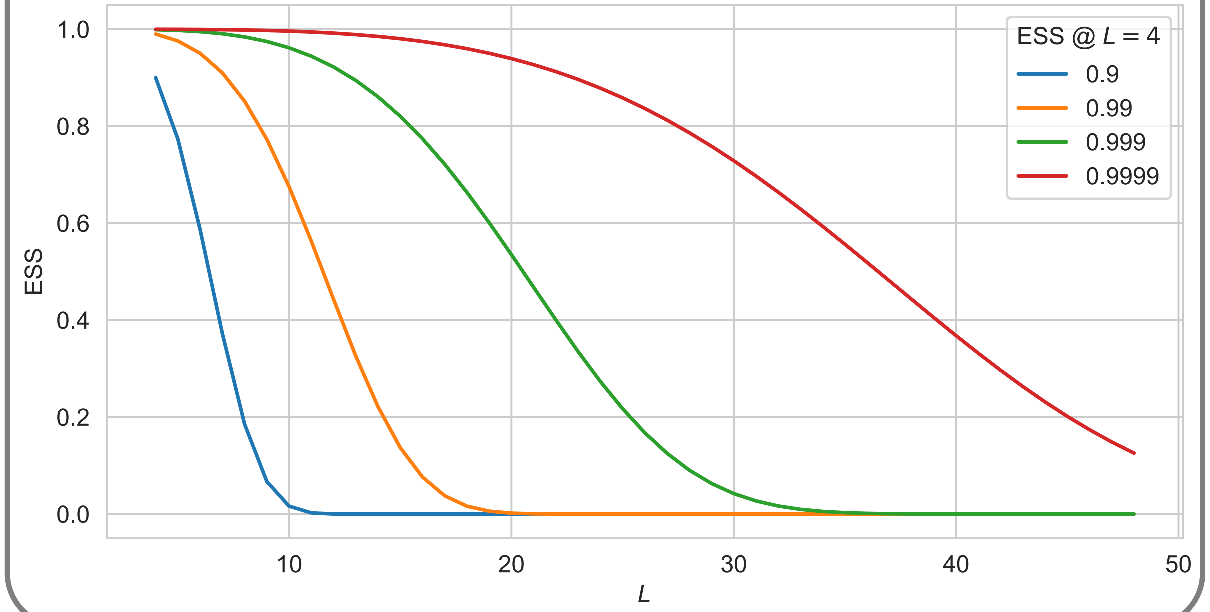
Two compounding issues presently block construction of viable QCD flow samplers at scale

1. Volume scaling of reweighting

Under direct volume transfer,

$$\mathrm{ESS}(V) \approx \mathrm{ESS}(V_0)^{V/V_0}$$

- Many caveats
- **Practical implication:** need excellent *local* modeling to do *global* reweighting



2. Model quality (non-)scalability

- Presently unable to construct high-quality flows between well-separated theories
- As yet, no known scalable approach: can't increase parameters/training to drive ESS arbitrarily high
- Combinatoric obstacles to increasing expressivity
  - Discrete: masking pattern sparsity
  - Continuous: loop basis size
- Unclear whether block is architecture/training/both

Present Obstacles

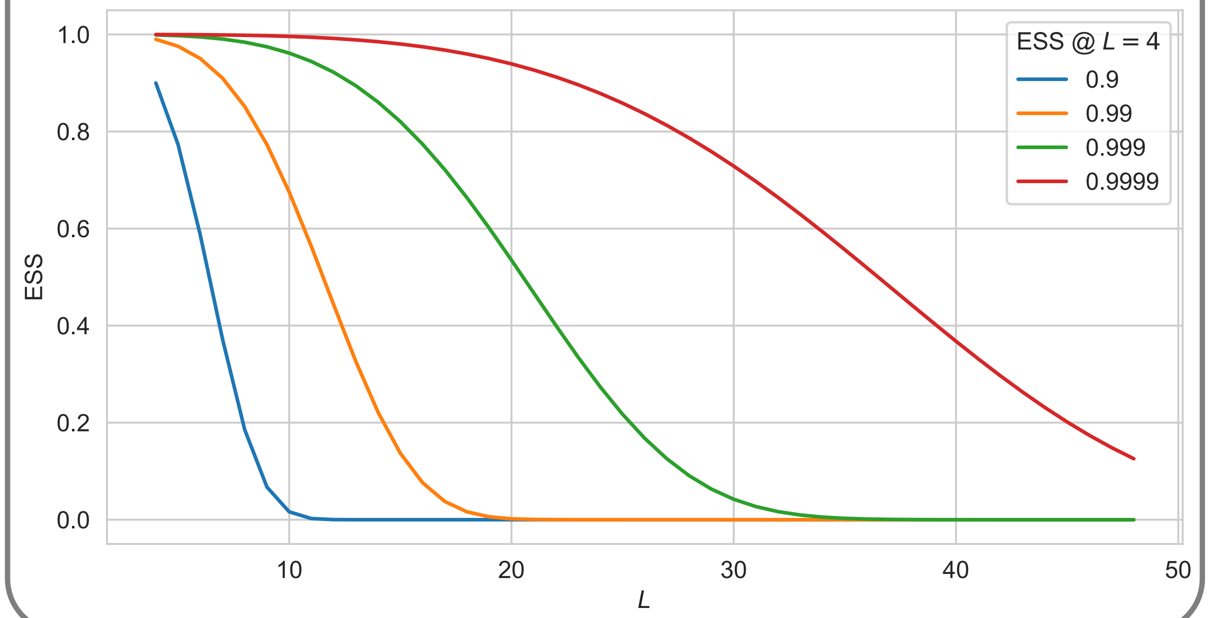
Two compounding issues presently block construction of viable QCD flow samplers at scale

1. Volume scaling of reweighting

Under direct volume transfer,

$$\mathrm{ESS}(V) \approx \mathrm{ESS}(V_0)^{V/V_0}$$

- Many caveats
- **Practical implication:** need excellent *local* modeling to do *global* reweighting



2. Model quality (non-)scalability

- Presently unable to construct high-quality flows between well-separated theories
- As yet, no known scalable approach: can't increase parameters/training to drive ESS arbitrarily high
- Combinatoric obstacles to increasing expressivity
  - Discrete: masking pattern sparsity
  - Continuous: loop basis size
- Unclear whether block is architecture/training/both

Present Efforts:

1. Applications of shorter flows
   What can we do *now* with flows of the quality we know how to construct?
2. The quest for scalable arch/training
   How can we construct better quality flows?

Present Obstacles

(deterministic)

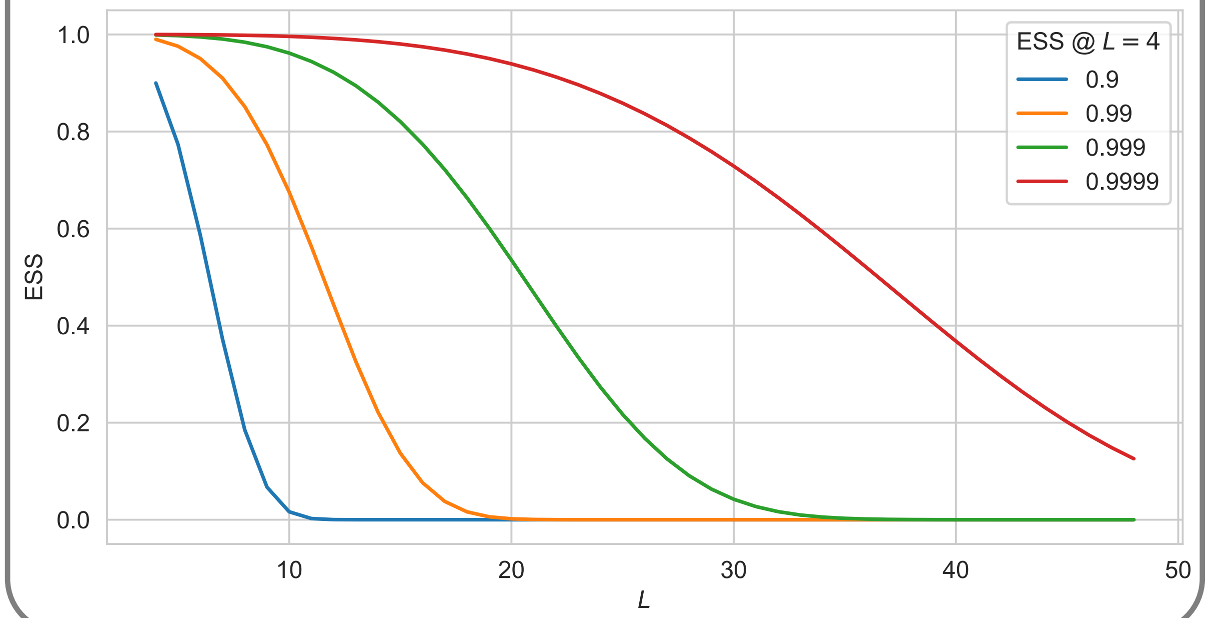
Two compounding issues presently block construction of viable QCD flow samplers at scale

1. Volume scaling of reweighting

Under direct volume transfer,

$$\mathrm{ESS}(V) \approx \mathrm{ESS}(V_0)^{V/V_0}$$

- Many caveats
- **Practical implication:** need excellent *local* modeling to do *global* reweighting



2. Model quality (non-)scalability

- Presently unable to construct high-quality flows between well-separated theories
- As yet, no known scalable approach: can't increase parameters/training to drive ESS arbitrarily high
- Combinatoric obstacles to increasing expressivity
  Discrete: masking pattern sparsity
  Continuous: loop basis size
- Unclear whether block is architecture/training/both

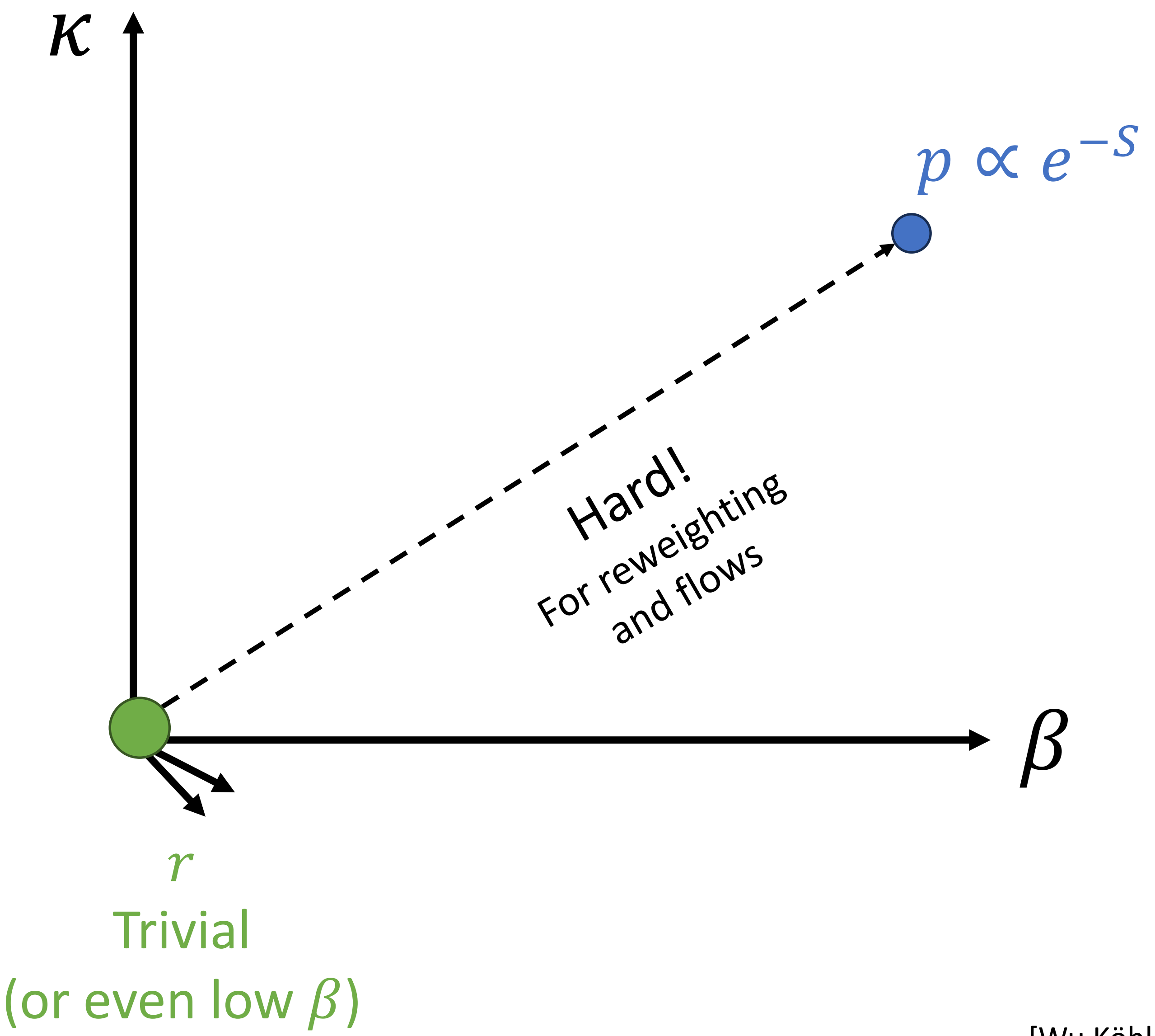
Present Efforts:

1. Applications of shorter flows
   What can we do *now* with flows of the quality we know how to construct?
2. The quest for scalable arch/training
   How can we construct better quality flows?

Relevant talks at Lattice 2026:
[253] R. Dionisio (M16:30) — Optimal Paths through Distribution Space for Mitigating Topological Freezing
[63] G. Fuwa (M16:50) — Parallel tempered Metadynamics in full QCD
[171] V. Granados-Pinto (M17:10) — Exploration of Parallel Tempering on Boundary Conditions
[179] E. Cellini (Th14:40) — Neural-Enhanced Out-of-equilibrium sampling for Lattice Field Theory
[338] O. Vega (Th15:00) — Unbiased Diffusion-Based Generation via Scalable Non-Equilibrium Transport

Applications of shorter flows:
Hybrid methods

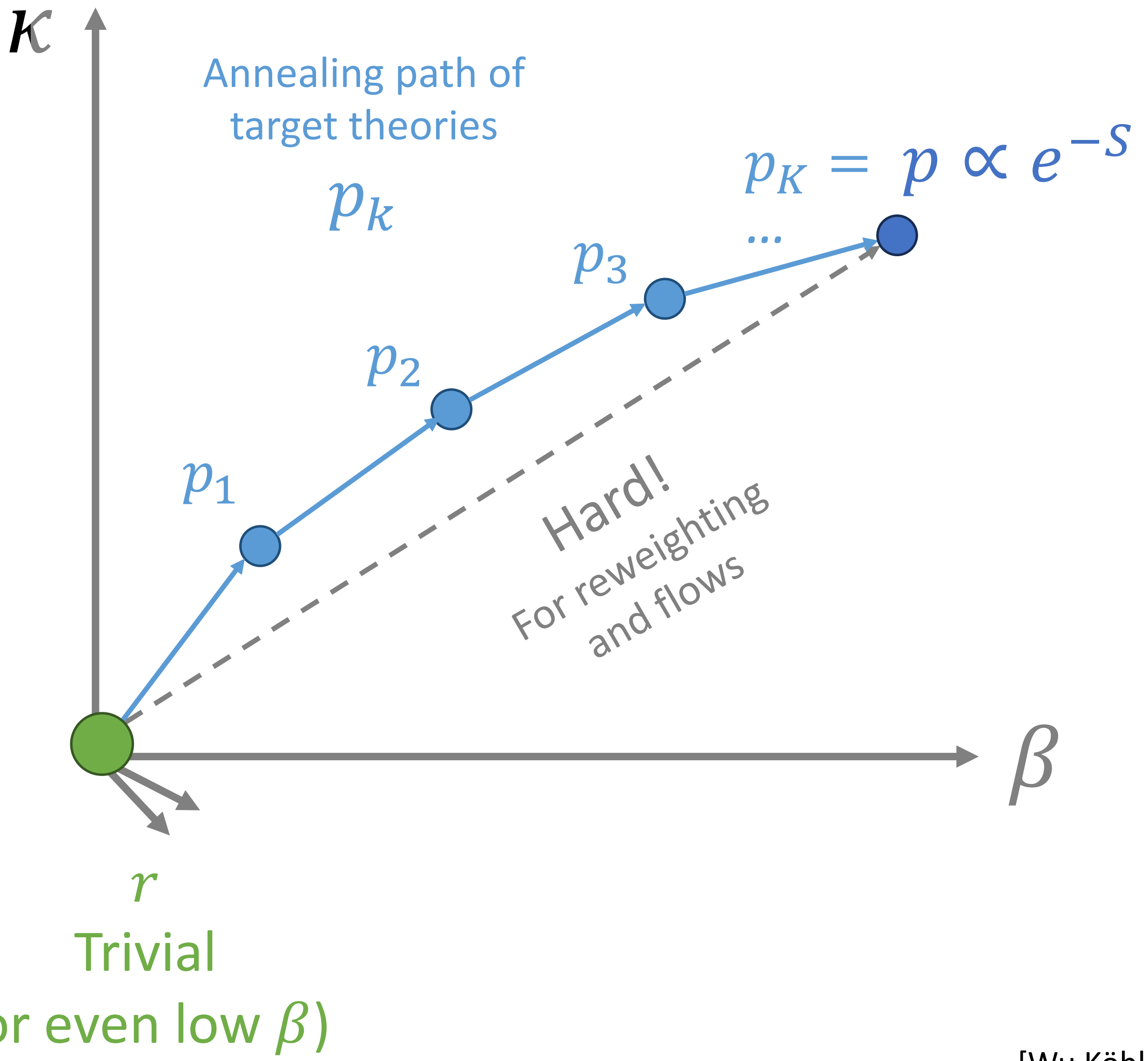
Stochastic Normalizing Flows (SNFs)



[Wu Köhler Noé 2002.06707]
[Caselle Cellini Nada Panero 2210.03139]

[Jarzynski cond-mat/9610209, cond-mat/9707325]
[Crooks cond-mat/9901352] [Neal physics/9803008]

Stochastic Normalizing Flows (SNFs)



AIS / Jarzcynski / Non-Equilibrium MCMC

At each p_k, apply an MCMC hit and track the work

$$\Delta W_k = S_{k+1}(U_k) - S_k(U_k)$$

Unbiased reweighting:

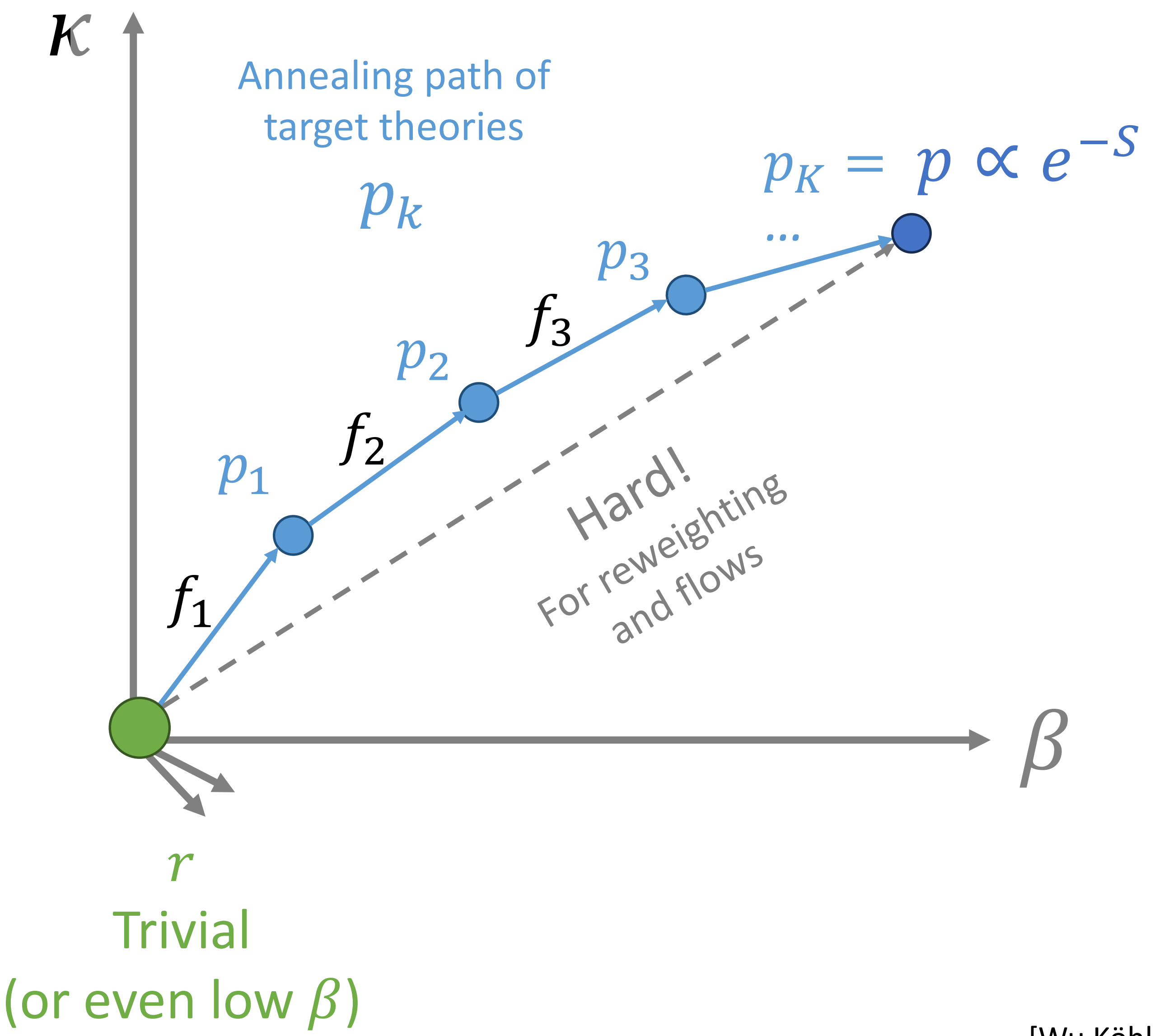
$$\hat{w} = e^{-\sum_k \Delta W_k}$$

$$\langle O \rangle_p = \langle \hat{w}\, O(U_K) \rangle \,/\, \langle \hat{w} \rangle$$

[Wu Köhler Noé 2002.06707]
[Caselle Cellini Nada Panero 2210.03139]

[Jarzynski cond-mat/9610209, cond-mat/9707325]
[Crooks cond-mat/9901352] [Neal physics/9803008]

Stochastic Normalizing Flows (SNFs)



AIS / Jarzcynski / Non-Equilibrium MCMC

At each p_k, apply an MCMC hit and track the work

$$\Delta W_k = S_{k+1}(U_k) - S_k(U_k)$$

Unbiased reweighting:

$$\hat{w} = e^{-\sum_k \Delta W_k}$$

$$\langle O \rangle_p = \langle \hat{w}\, O(U_K) \rangle \,/\, \langle \hat{w} \rangle$$

Stochastic normalizing flows

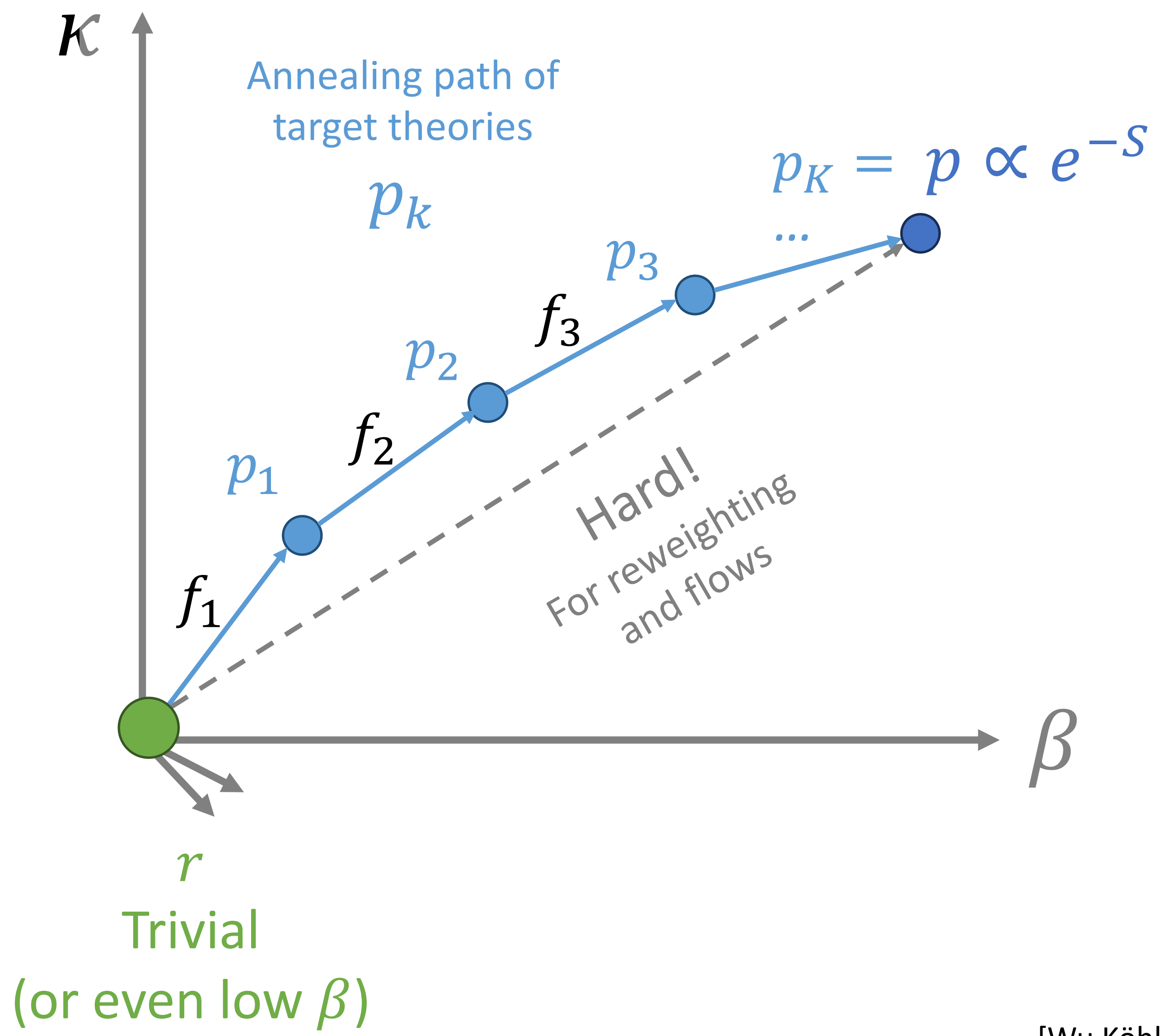
Add deterministic flows bridging between theories

Generalized work:

$$\Delta W_k = S_{k+1}\big(f_k(U_k)\big) - S_k(U_k) - \log J_k(U_k)$$

[Wu Köhler Noé 2002.06707]
[Caselle Cellini Nada Panero 2210.03139]
[Jarzynski cond-mat/9610209, cond-mat/9707325]
[Crooks cond-mat/9901352] [Neal physics/9803008]

Stochastic Normalizing Flows (SNFs)



AIS / Jarzcynski / Non-Equilibrium MCMC

At each p_k, apply an MCMC hit and track the work

$$\Delta W_k = S_{k+1}(U_k) - S_k(U_k)$$

Unbiased reweighting:

$$\hat{w} = e^{-\sum_k \Delta W_k}$$

$$\langle O \rangle_p = \langle \hat{w}\, O(U_K) \rangle \,/\, \langle \hat{w} \rangle$$

Stochastic normalizing flows

Add deterministic flows bridging between theories

Generalized work:

$$\Delta W_k = S_{k+1}\big(f_k(U_k)\big) - S_k(U_k) - \log J_k(U_k)$$

Zero variance limits:

- Perfect flows (obvious, wouldn't need MCMC hits)
- **Adiabatic limit (independent of flow quality!)**

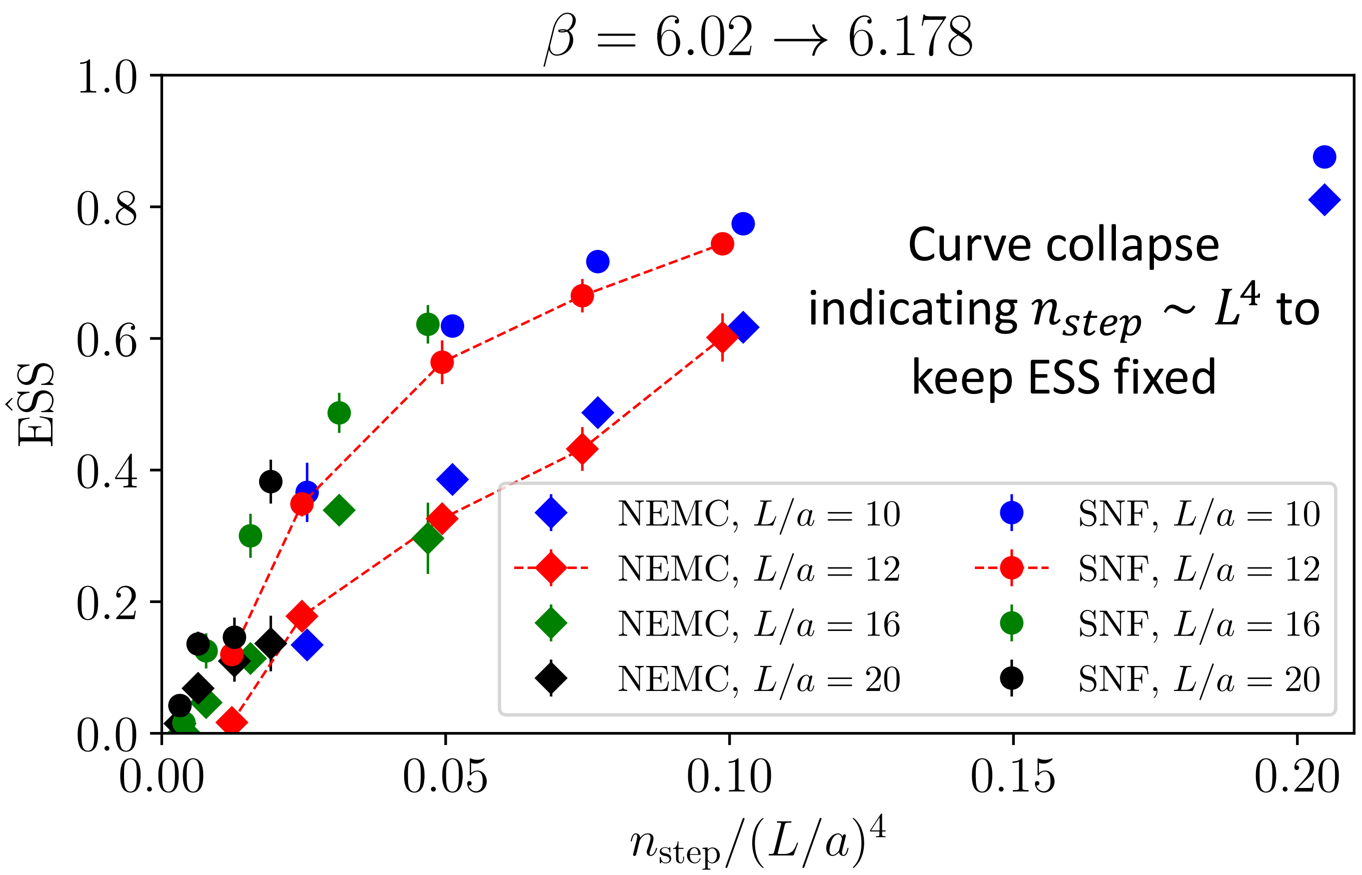
[Wu Köhler Noé 2002.06707]
[Caselle Cellini Nada Panero 2210.03139]
[Jarzynski cond-mat/9610209, cond-mat/9707325]
[Crooks cond-mat/9901352] [Neal physics/9803008]

SNF volume scaling

In pure gauge, follows from

$$\mathrm{ESS}(\beta \to \beta + \Delta\beta) \sim e^{-c(\beta)\, V\, \Delta\beta^2}$$

$$n_{\text{steps}} \sim V \text{ for fixed ESS}$$

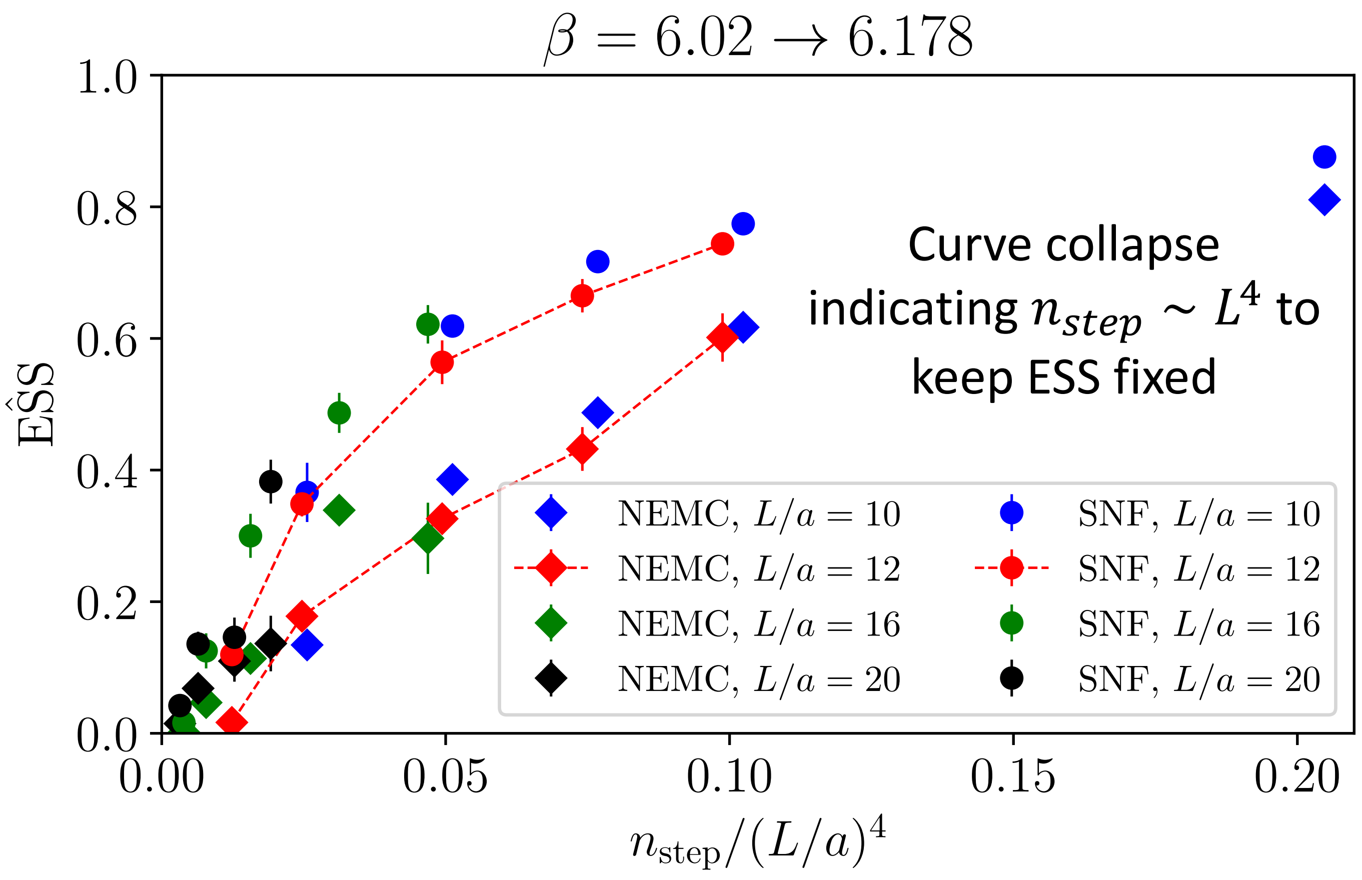


From [Bulgarelli, Cellini, Nada 2412.00200]

SNF volume scaling

$n_{\text{steps}} \sim V$ for fixed ESS

In pure gauge, follows from
$\mathrm{ESS}(\beta \to \beta + \Delta\beta) \sim e^{-c(\beta)\, V\, \Delta\beta^2}$



From [Bulgarelli, Cellini, Nada 2412.00200]

Note: applies also for NE-MCMC, with no flows

e.g. [Caselle Nada Panero 1801.03110]

Flows reduce *prefactor*

i.e. overall number of steps required

Can construct viable SNF samplers from presently obtainable flows!

SNFs vs T-REX (= Transformed Replica EXchange)

T-REX: flows enable higher swap rates in parallel tempering / replica exchange

Requires *same flows as SNF* over same path

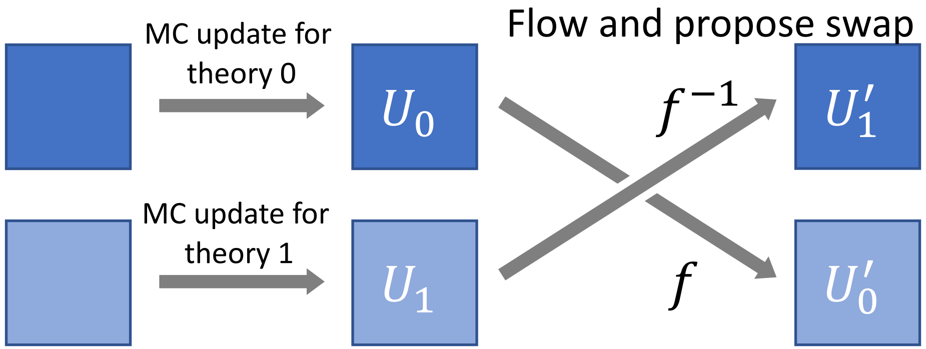
→ Can train SNF flows in REX setup

See **[179] E. Cellini (Th 14:40)**

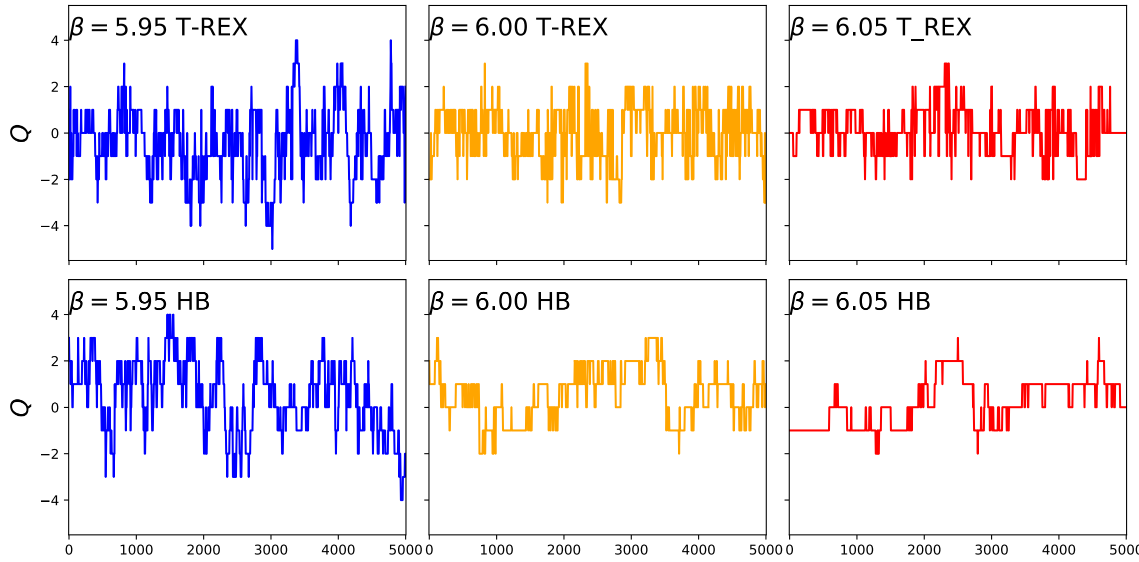
Similar volume scaling as SNFs

Performance may differ in detail.

e.g. T-REX more parallelizable, at cost of coordination problems (different integrator walltimes)



$$p_{\text{acc}} = \min\left[1, \frac{p_0(U_1')\, p_1(U_0')}{p_0(U_0)\, p_1(U_1)}\, J_f(U_0)\, J_{f^{-1}}(U_1)\right]$$



REX: [Swendsen Wang 1986] [Marinari Parisi hep-lat/9205018] [Hukushima Nemoto cond-mat/9512035]

L-REX: [Invernizzi Krämer Clementi Noé 2210.14104]

T-REX: [Abbott, Albergo, Boyda, **DH**, Kanwar, Romero-López, Shanahan, Urban 2404.11674]

Defects and topological freezing

Localized OBC defect $\sim L_d^3$ accelerates mixing

PTBC: Anneal defect away with REX

[Hasenbusch 1706.04443] [Bonnano Bonati D'Elia 2012.14000]

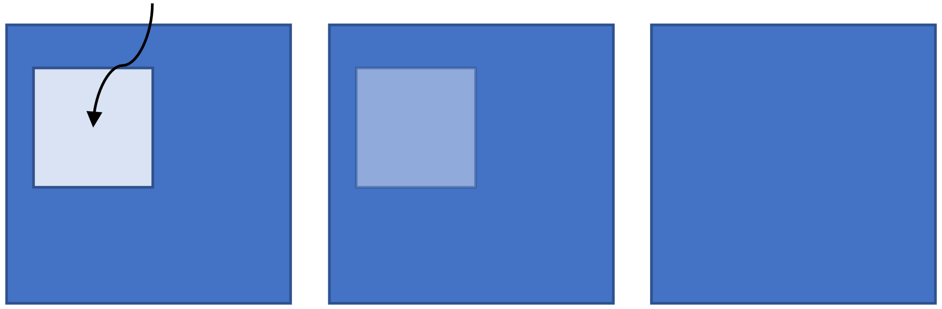
[Bonanno Clemente D'Elia Maio Parente 2404.14151]

Alternately, use NE-MCMC / Jarzynski

[Bonanno Nada Vadacchino 2310.11979, 2402.06561, 2411.00620]

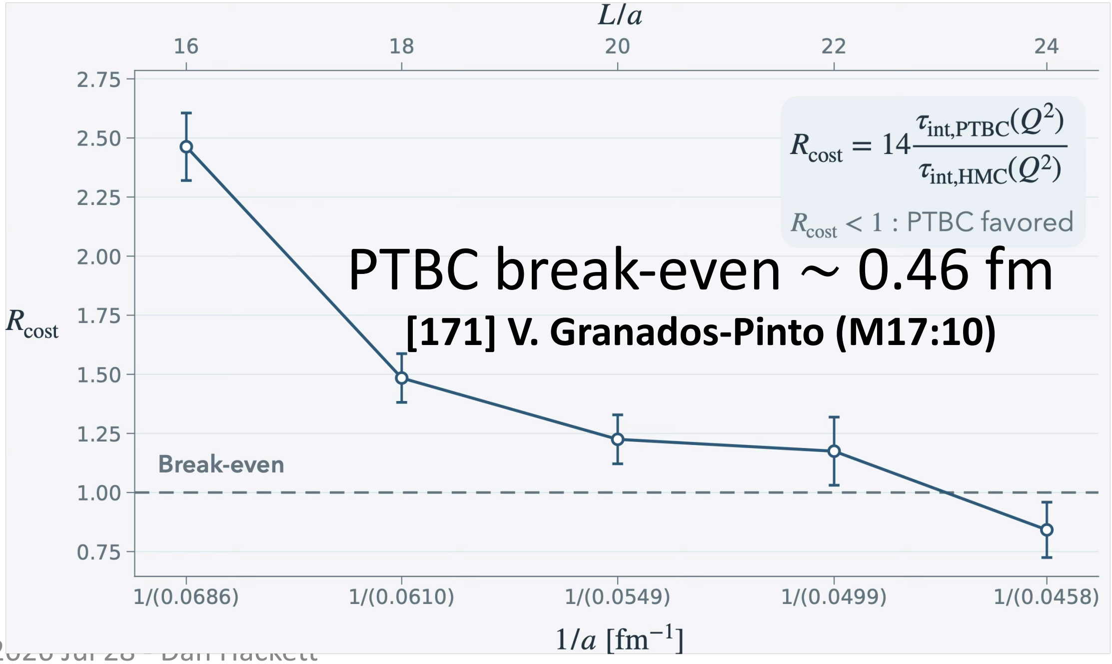
Volume scaling $\sim e^{-L_d^3}$ *independent of target volume!*

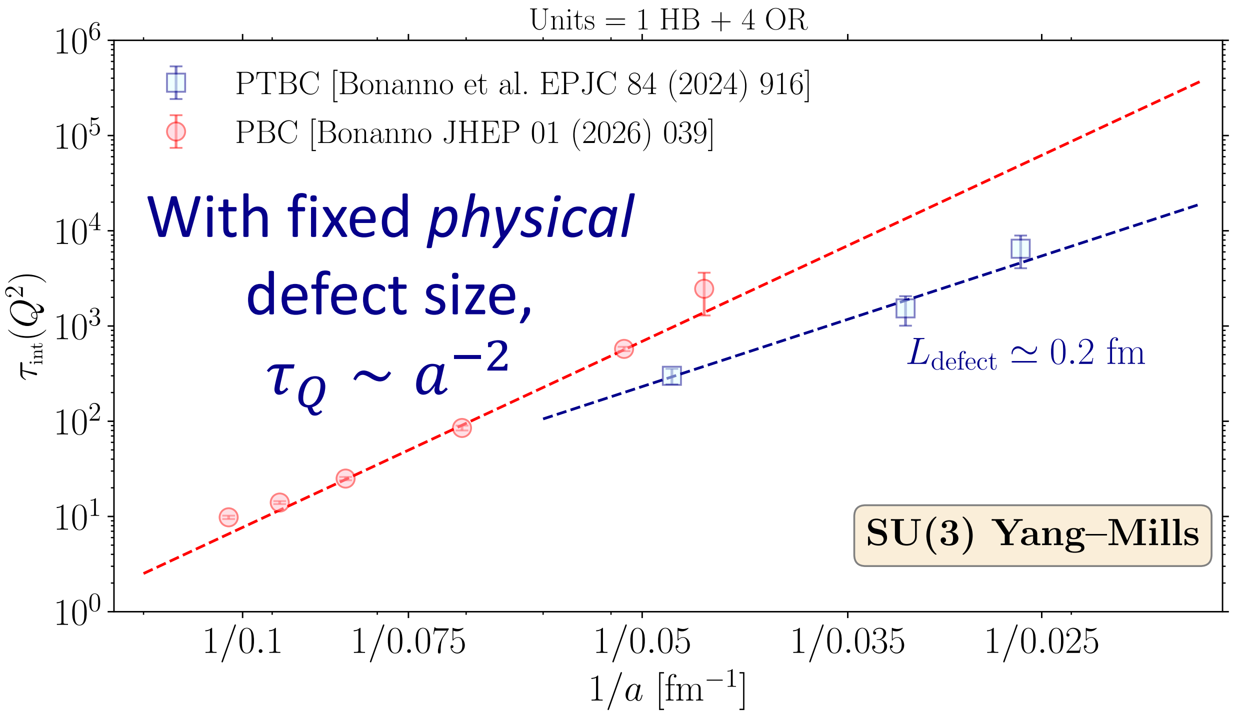
Defect: patch of OBCs on one timeslice



Defect theory ⟷ Target: PBCs

Anneal to full PBCs



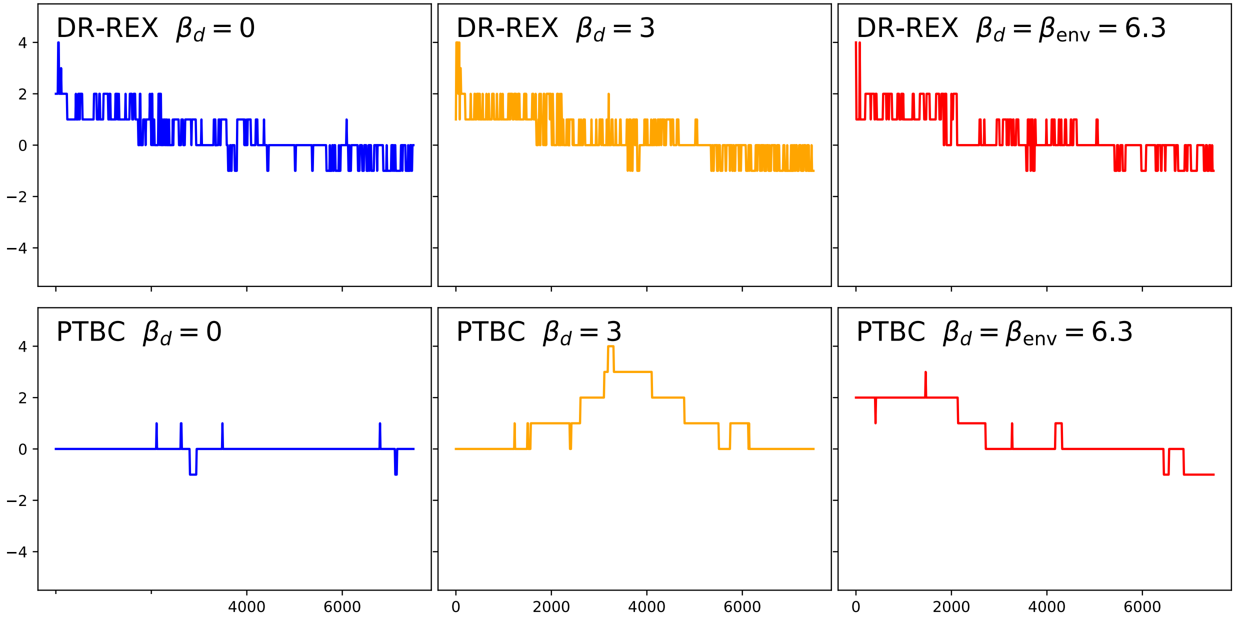


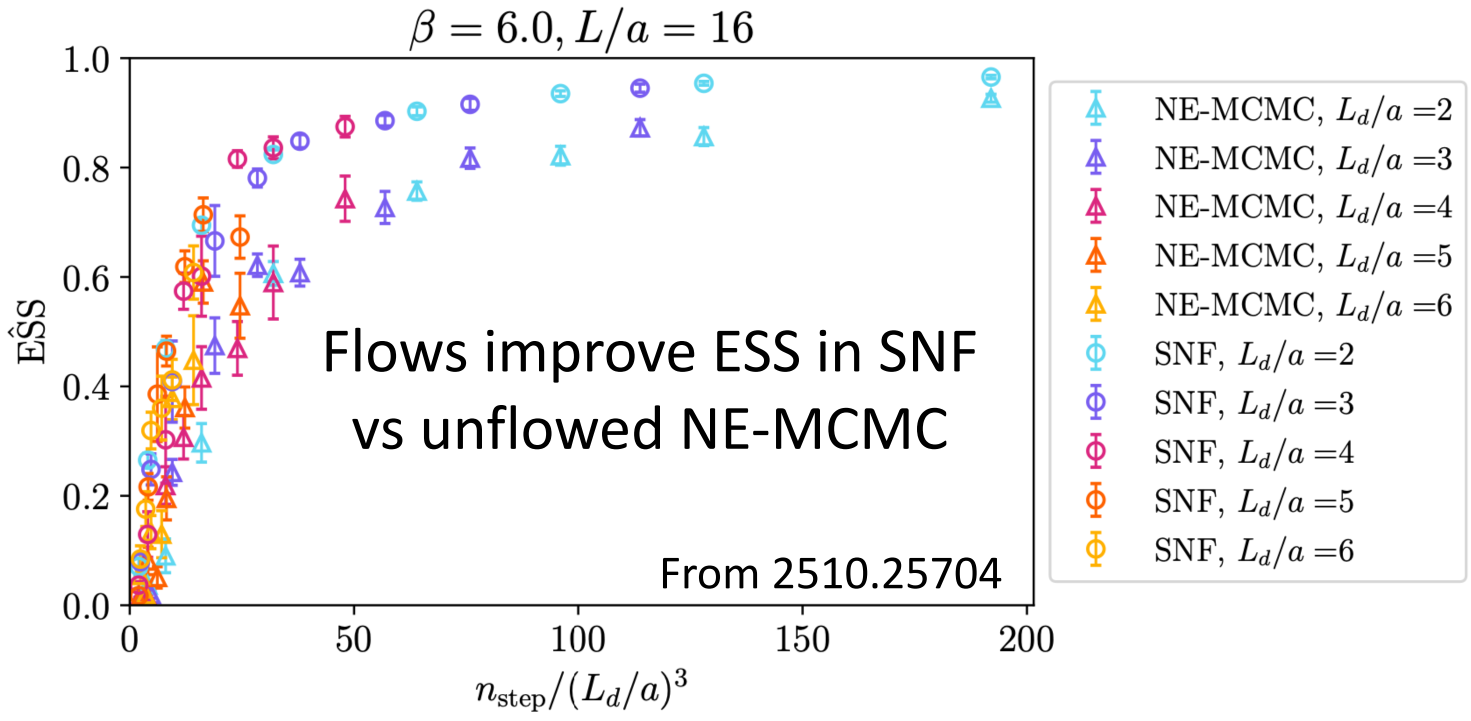
Defect Repair Flows

Flows can reduce overall scale required, allow fewer replicas/steps

DR-REX: Flows accelerate topological mixing by enabling swaps out-of-range for PTBC

[Abbott, Albergo, Boyda, **DH**, Kanwar, Romero-López, Shanahan, Urban 2404.11674]





SNF version:

Project advantage in pure gauge below $a \sim 0.03$ fm!

Better flows: advantage at coarser lattices

[Bonanno Bulgarelli Cellini Nada Panfalone Vadacchino Verzichelli 2510.25704, 2601.20708]

SNF & T-REX: Outstanding Challenges

- Needs demonstration with fermions
  - SNFs and T-REX may admit different treatments.
- Many intermediate flows to train
  - Interpolation can help
  - e.g. [Bulgarelli, Cellini, Nada 2412.00200]
- Many intermediate MCMC updates
  - HMC integrator setups laborious
  - Expensive!

SNF & T-REX: Outstanding Challenges

Needs demonstration with fermions

- SNFs and T-REX may admit different treatments.

Many intermediate flows to train

- Interpolation can help
- e.g. [Bulgarelli, Cellini, Nada 2412.00200]

Many intermediate MCMC updates

- HMC integrator setups laborious
- Expensive!

For general-purpose sampling:

Overall $O(V^2)$ scaling vs HMC's $O\left(V^{5/4}\right)$

Need $n_{step} \sim 10$s for advantage

- cf. thermalization for $O(100 - 1000)$
- Better flows will help

SNF & T-REX: Outstanding Challenges

Needs demonstration with fermions

SNFs and T-REX may admit different treatments.

Many intermediate flows to train

Interpolation can help

e.g. [Bulgarelli, Cellini, Nada 2412.00200]

Many intermediate MCMC updates

HMC integrator setups laborious

Expensive!

For general-purpose sampling:

Overall $O(V^2)$ scaling vs HMC's $O\left(V^{5/4}\right)$

Need $n_{step} \sim 10$s for advantage

cf. thermalization for $O(100 - 1000)$

Better flows will help

For resolving topological freezing:

OBC defects solve freezing *without flows*, but require large overall scale

Many intermediate theories

Flows can push this scale down to more practical/useful levels

Localized defects: easier volume scaling

Another option: metadynamics potentials

See [63] G. Fuwa (M16:50)

Flows could help here too!

Relevant talks at Lattice 2026:

[253] R. Dionisio (M16:30) — Optimal Paths through Distribution Space for Mitigating Topological Freezing

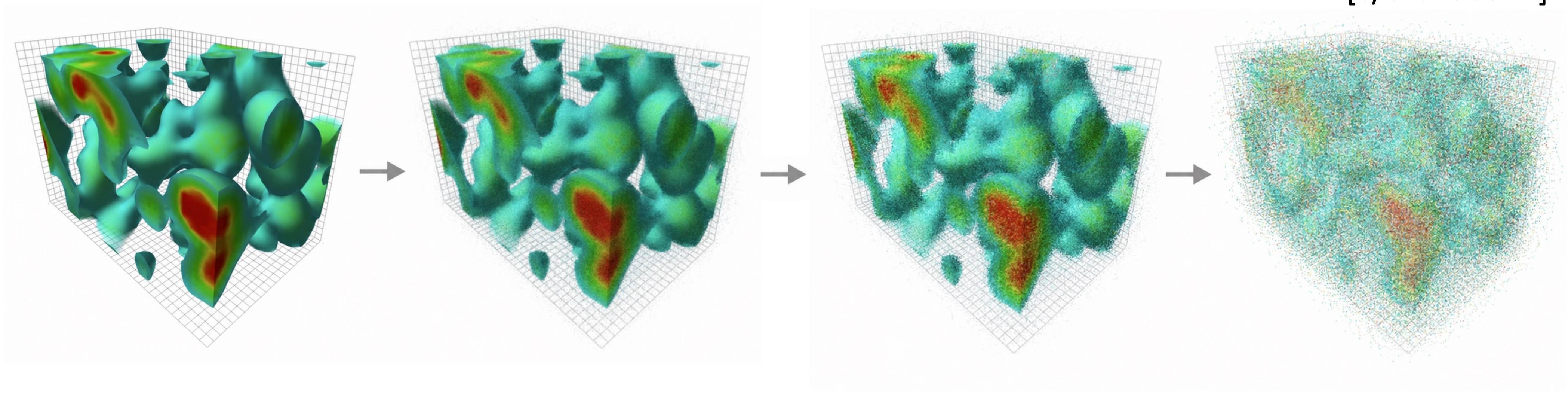
[338] O. Vega (Th15:00) — Unbiased Diffusion-Based Generation via Scalable Non-Equilibrium Transport

The quest for scalable approaches:

Diffusions, Transport, etc

What are diffusions, transport, etc?

[c/o ChatGPT]

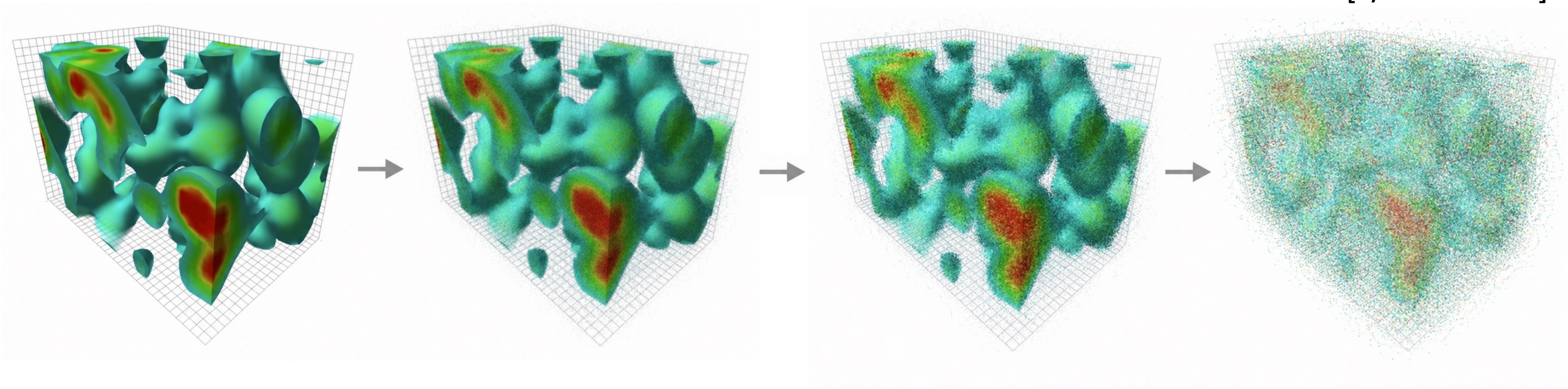


Very rich mathematics/statistics for:

- Stochastic Differential Equations (SDEs)
- Transport (through theory space)

What are diffusions, transport, etc?

[c/o ChatGPT]



Very rich mathematics/statistics for:

- Stochastic Differential Equations (SDEs)
- Transport (through theory space)

Think very hard about them, get:

1. **Improved training methods**

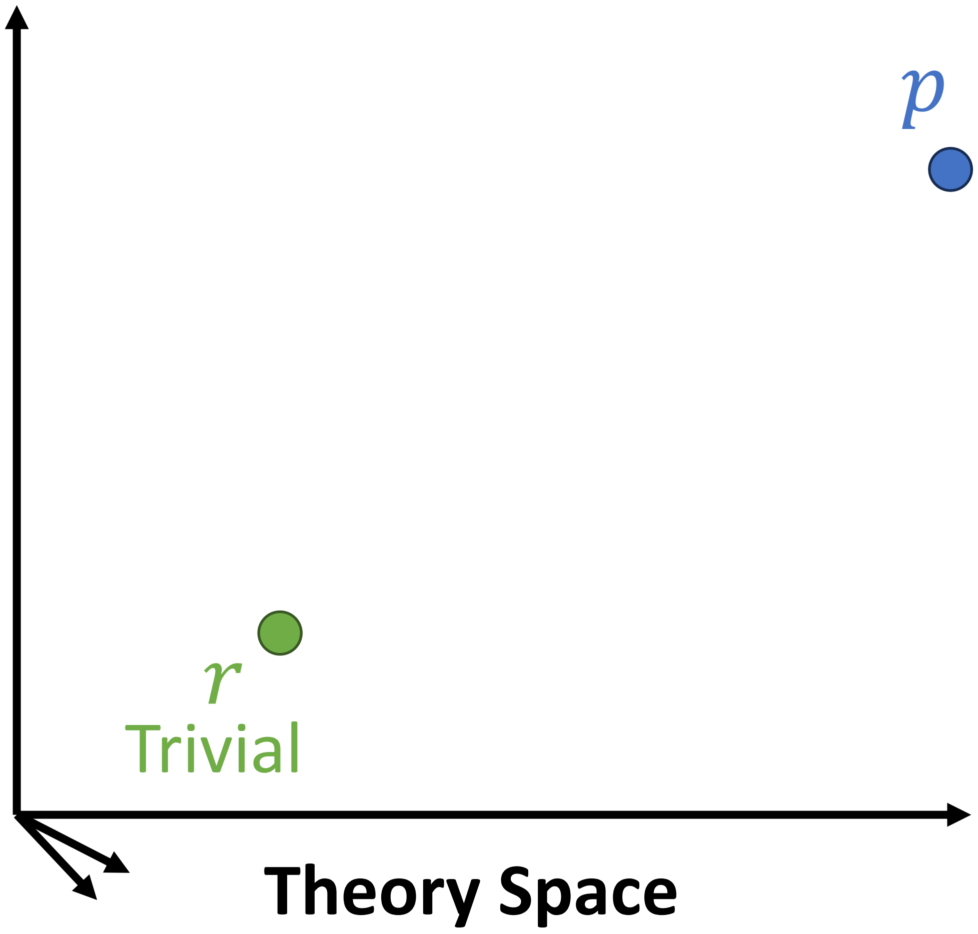
   Denoising score matching

   Flow matching / stochastic interpolants

   …

2. **New exact sampling schemes**

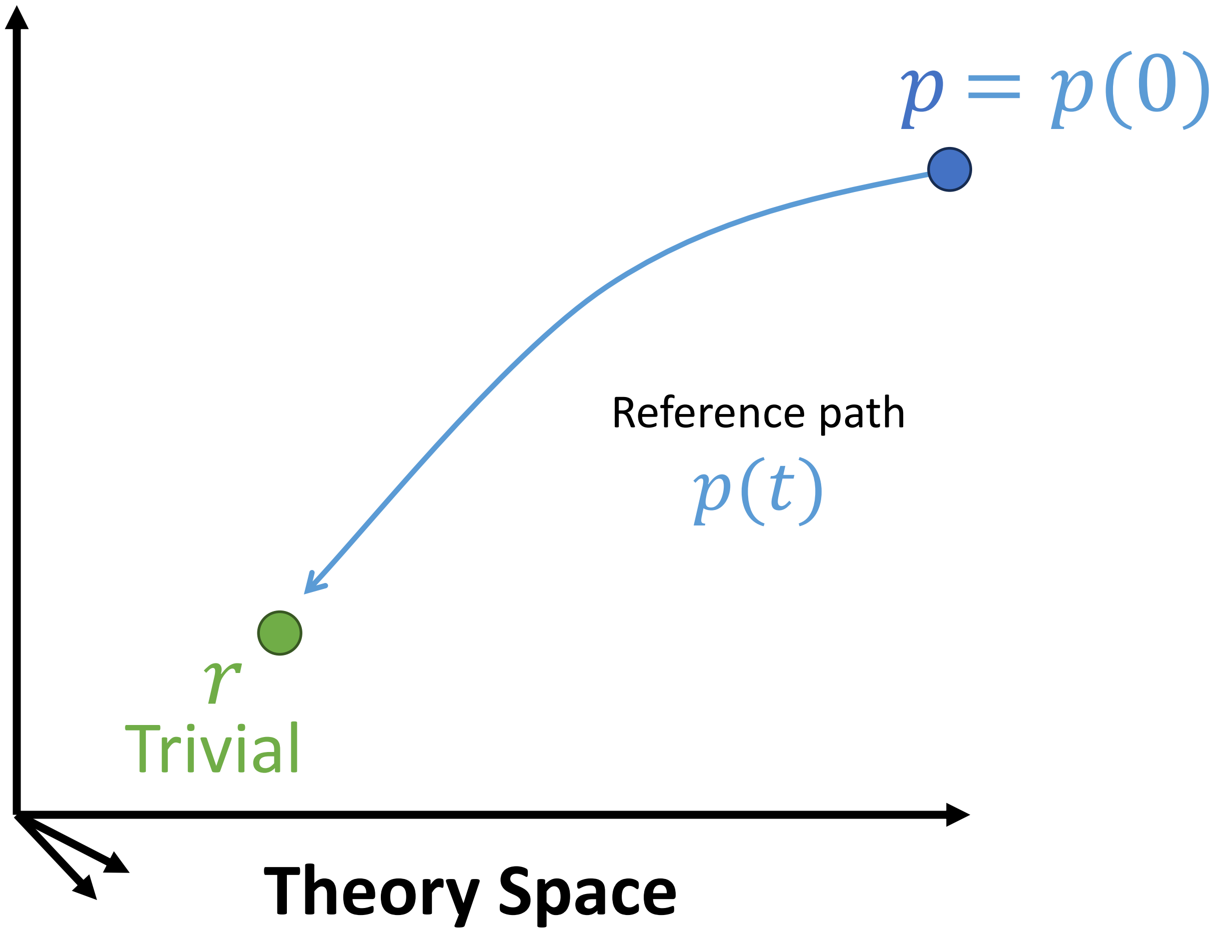
Diffusions: training



Diffusions: training

Step 1: Diffusion process out of p defines a **reference path**

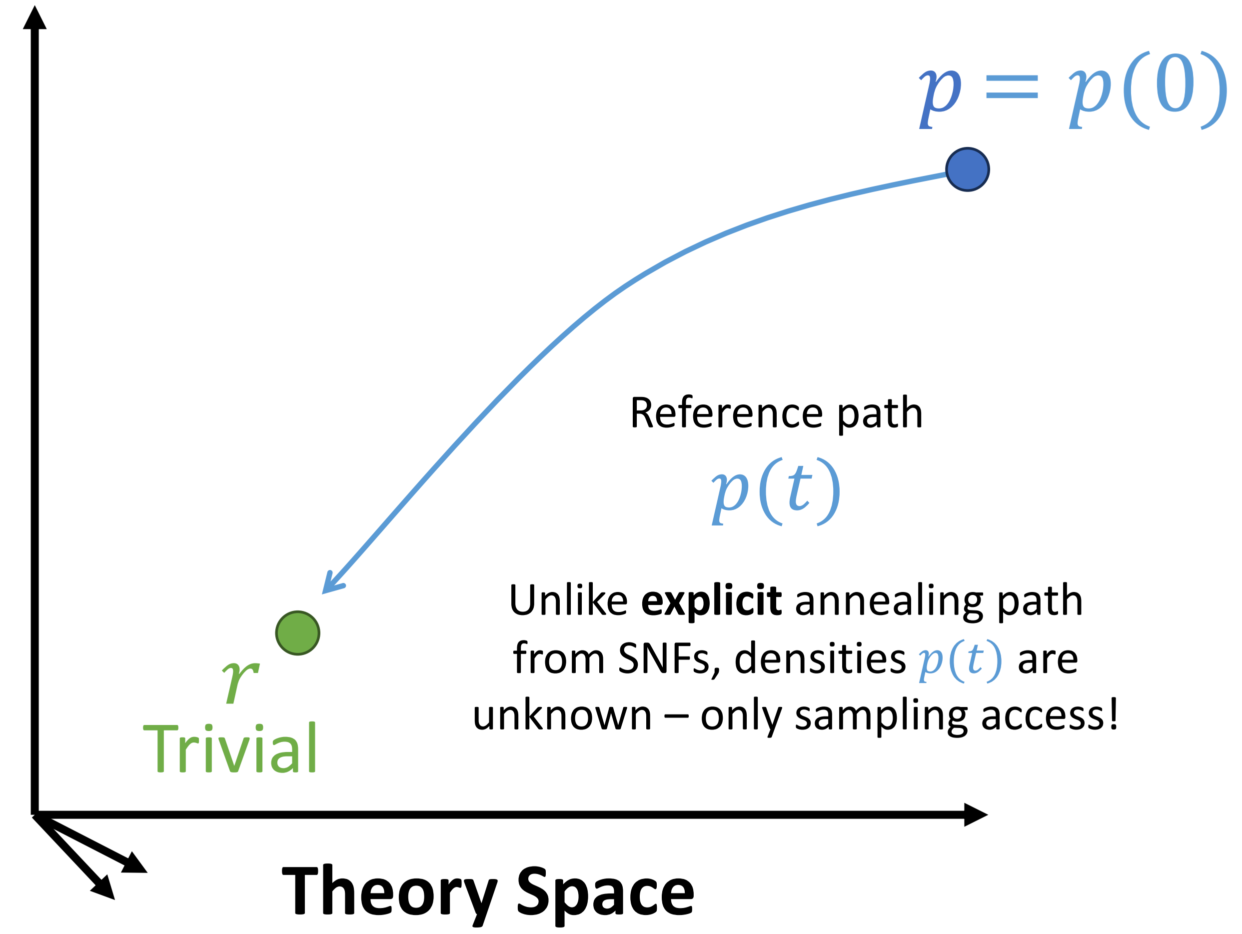
- Iteratively adding noise to samples from p trivializes $p \to r$ as t increases
- **Note:** requires target samples $\sim p$



Diffusions: training

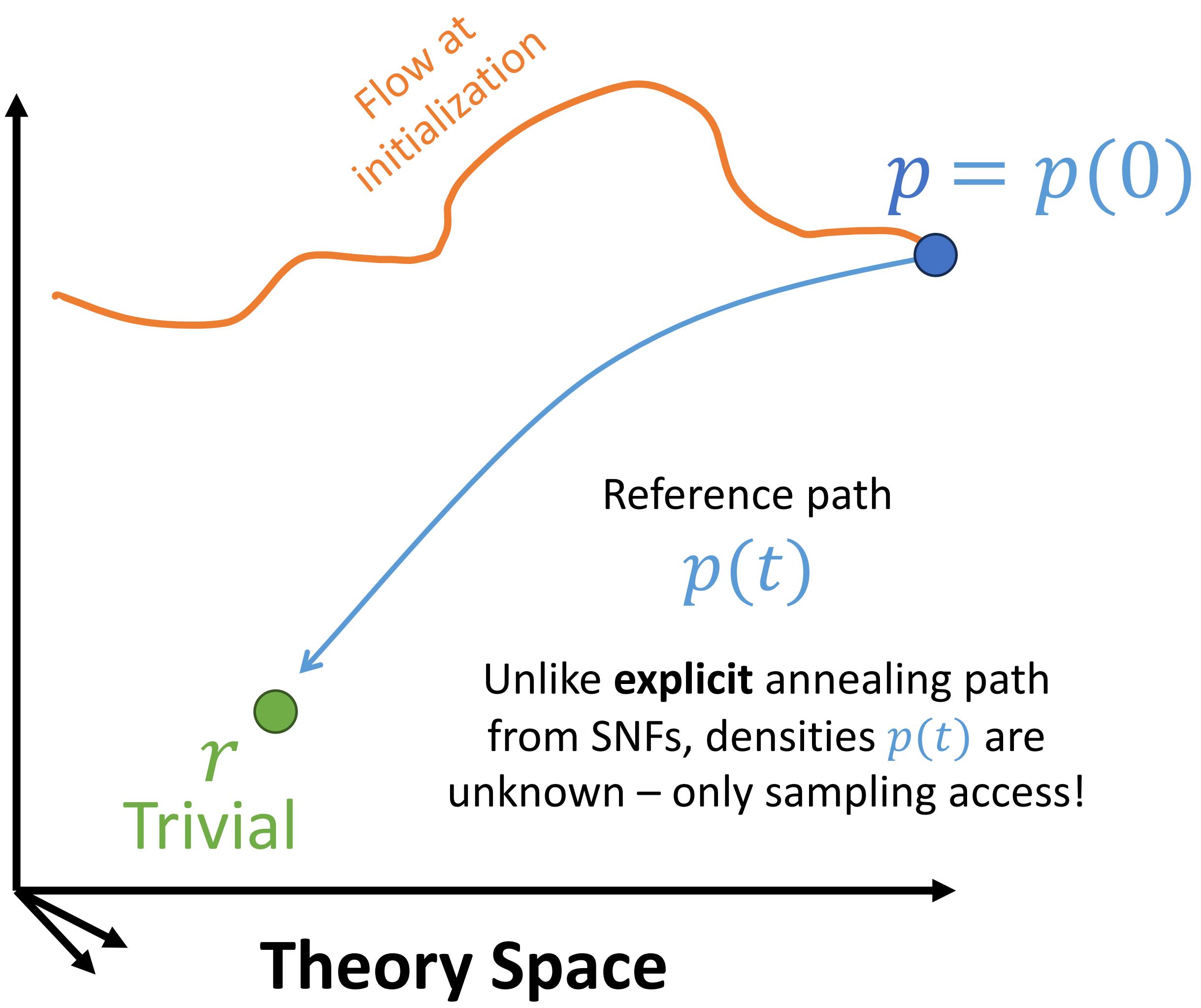
Step 1: Diffusion process out of p defines a **reference path**

- Iteratively adding noise to samples from p trivializes $p \to r$ as t increases
- **Note:** requires target samples $\sim p$



Unlike **explicit** annealing path from SNFs, densities $p(t)$ are unknown – only sampling access!

Diffusions: training



Step 1: Diffusion process out of p defines a **reference path**

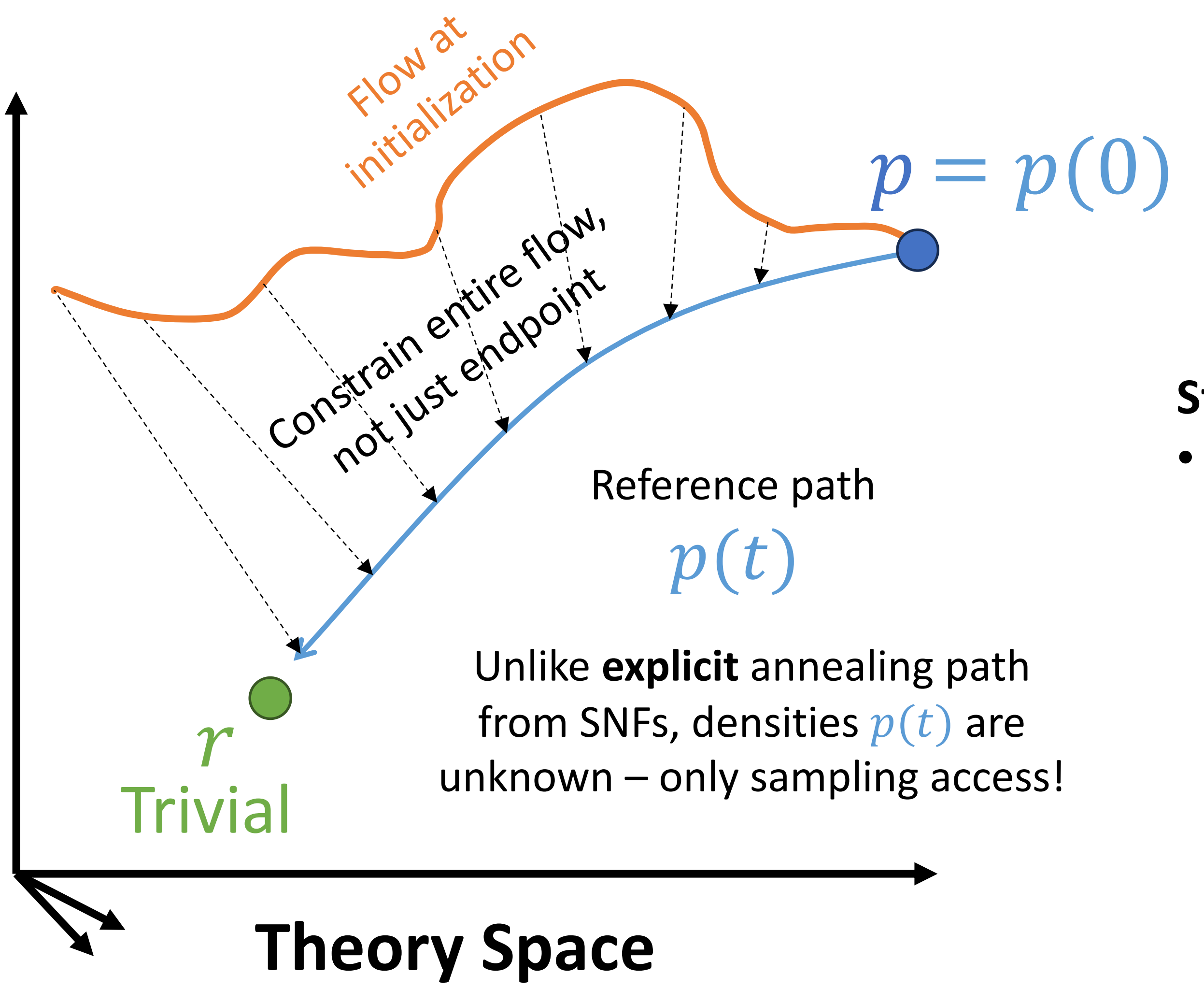
- Iteratively adding noise to samples from p trivializes $p \to r$ as t increases
- **Note:** requires target samples $\sim p$

Step 2: Set up a trainable flow out of p

- Continuous, stochastic: noise + learned drift $s_\theta(t)$ cf. Langevin / stochastic quantization

[Wang Aarts Zhou 2309.17082]

Diffusions: training



Step 1: Diffusion process out of p defines a **reference path**

- Iteratively adding noise to samples from p trivializes $p \to r$ as t increases
- **Note:** requires target samples $\sim p$

Step 2: Set up a trainable flow out of p

- Continuous, stochastic: noise + learned drift $s_\theta(t)$ cf. Langevin / stochastic quantization

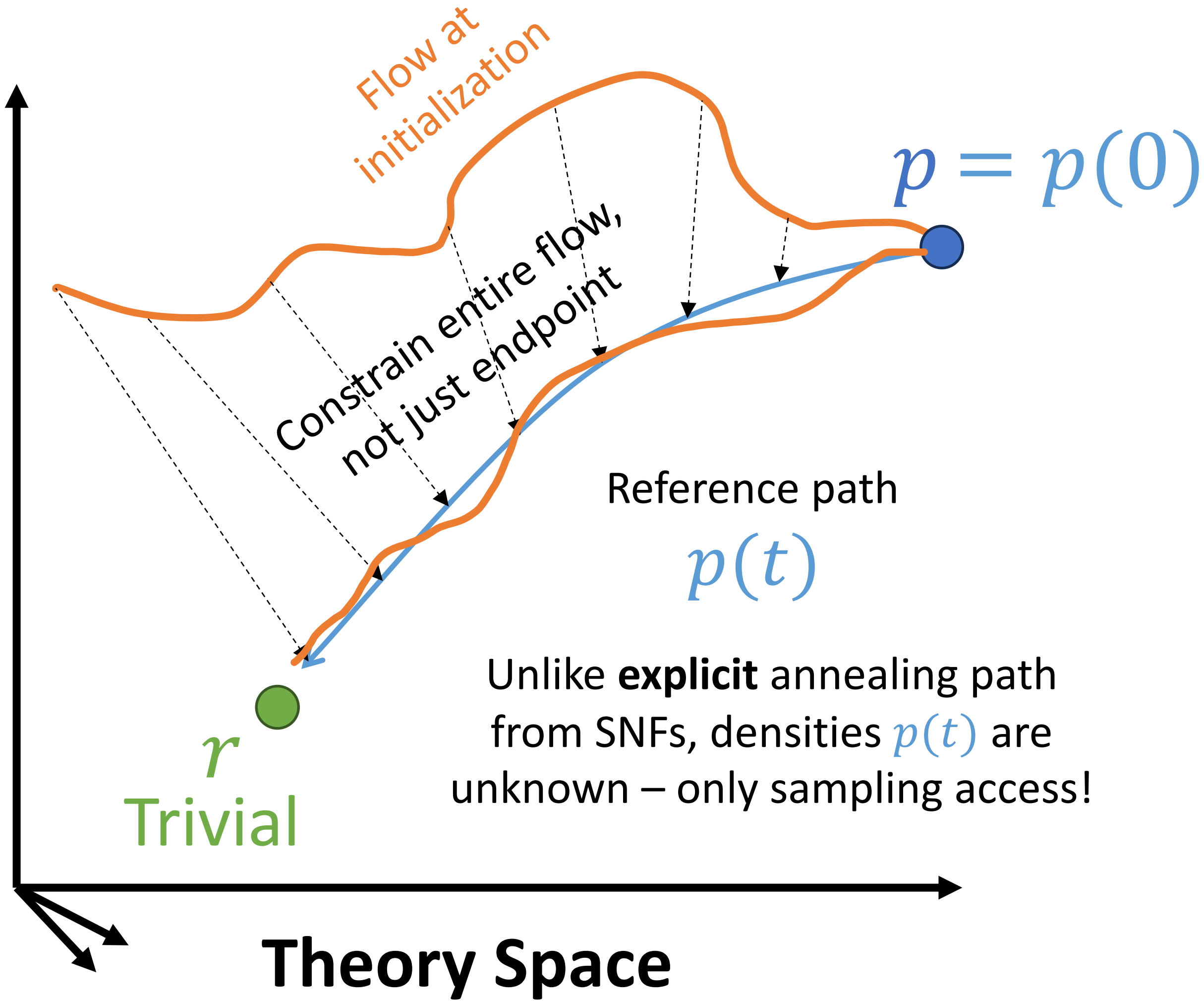
[Wang Aarts Zhou 2309.17082]

Step 3: Minimize flow to approximate the reference path

- Train on a **regression** target

$$L(\theta) \sim \int_0^1 dt \left\langle \left[s_\theta(t) - \begin{pmatrix} \text{known,} \\ \text{simple fn} \end{pmatrix}(t) \right]^2 \right\rangle$$

Diffusions: training



Unlike **explicit** annealing path from SNFs, densities $p(t)$ are unknown – only sampling access!

Step 1: Diffusion process out of p defines a **reference path**

- Iteratively adding noise to samples from p trivializes $p \to r$ as t increases
- **Note:** requires target samples $\sim p$

Step 2: Set up a trainable flow out of p

- Continuous, stochastic: noise + learned drift $s_\theta(t)$
  cf. Langevin / stochastic quantization

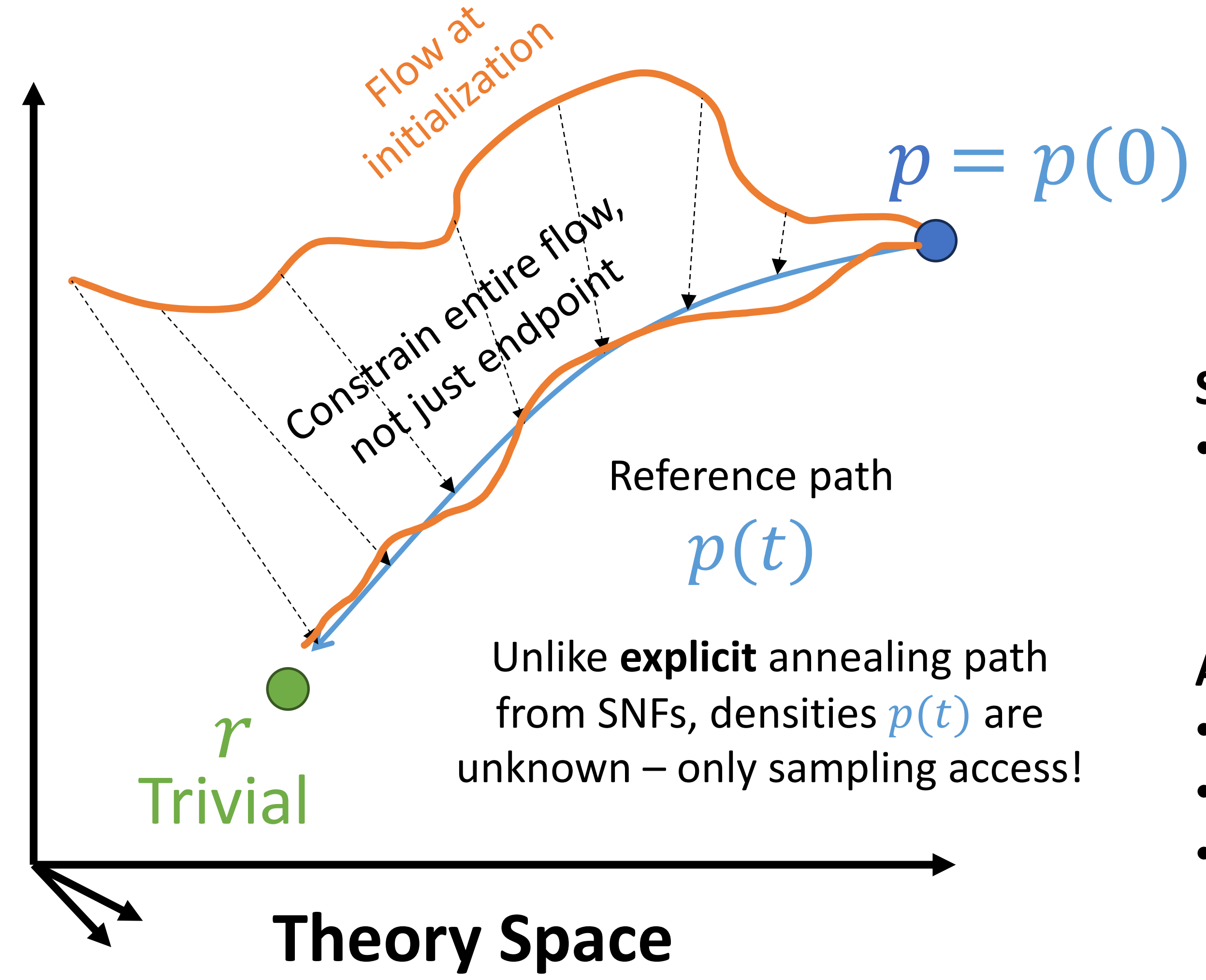
[Wang Aarts Zhou 2309.17082]

Step 3: Minimize flow to approximate the reference path

- Train on a **regression** target

$$L(\theta) \sim \int_0^1 dt \left\langle \left[s_\theta(t) - \left(\begin{matrix} \text{known,} \\ \text{simple fn} \end{matrix} \right)(t) \right]^2 \right\rangle$$

Diffusions: training



Unlike **explicit** annealing path from SNFs, densities $p(t)$ are unknown – only sampling access!

Step 1: Diffusion process out of p defines a **reference path**

- Iteratively adding noise to samples from p trivializes $p \to r$ as t increases
- **Note:** requires target samples $\sim p$

Step 2: Set up a trainable flow out of p

- Continuous, stochastic: noise + learned drift $s_\theta(t)$
  cf. Langevin / stochastic quantization

[Wang Aarts Zhou 2309.17082]

Step 3: Minimize flow to approximate the reference path

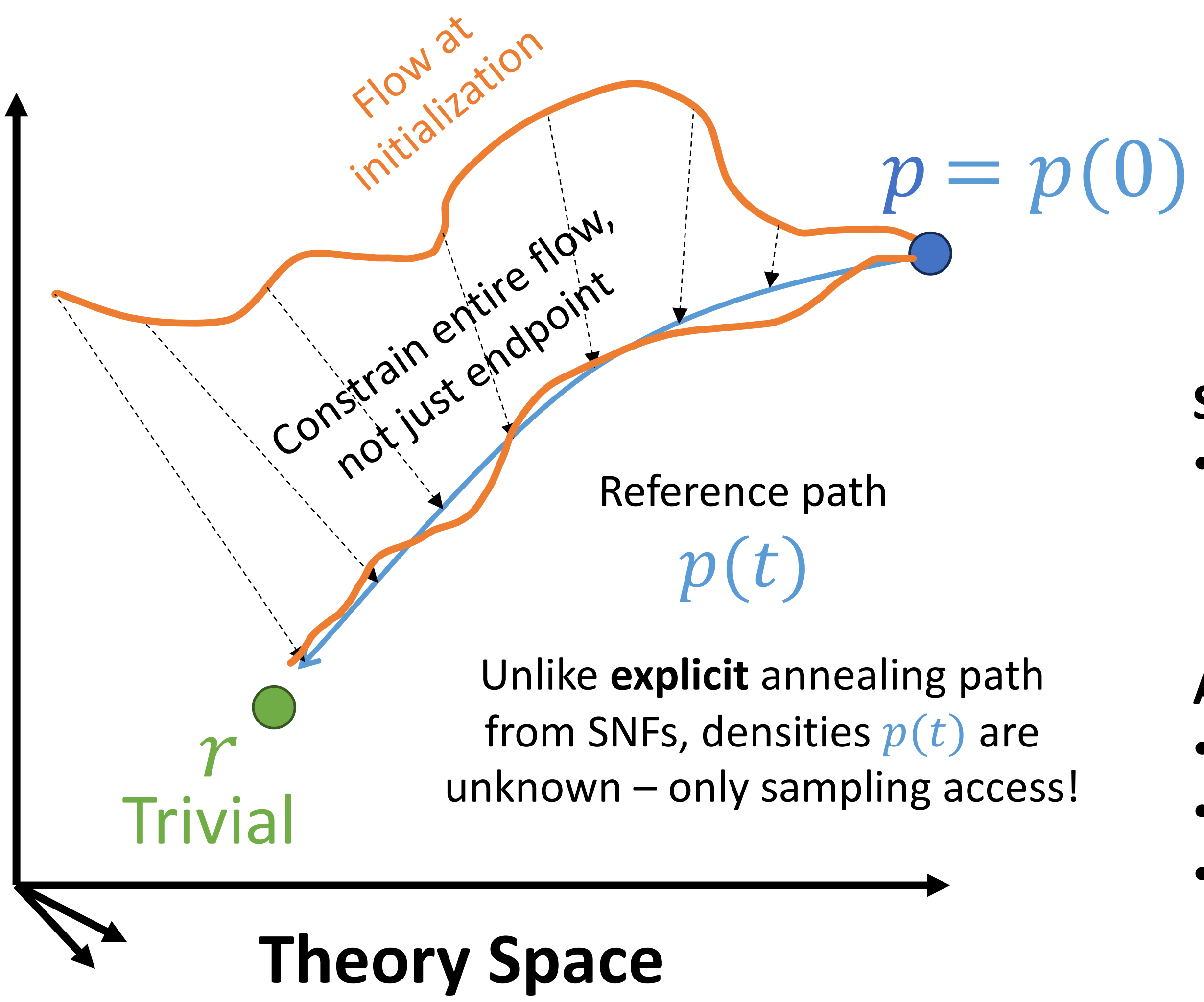
- Train on a **regression** target

$$L(\theta) \sim \int_0^1 dt \left\langle \left[s_\theta(t) - \begin{pmatrix} \text{known,} \\ \text{simple fn} \end{pmatrix}(t) \right]^2 \right\rangle$$

Advantages:

- No evaluation of p, q – don't need to run the flow
- Sampling $p(t)$ usually trivial for arbitrary t
- Parallelize training across entire path

Diffusions: training



Unlike **explicit** annealing path from SNFs, densities $p(t)$ are unknown – only sampling access!

Step 1: Diffusion process out of p defines a **reference path**

- Iteratively adding noise to samples from p trivializes $p \to r$ as t increases
- **Note:** requires target samples $\sim p$

Step 2: Set up a trainable flow out of p

- Continuous, stochastic: noise + learned drift $s_\theta(t)$ cf. Langevin / stochastic quantization

[Wang Aarts Zhou 2309.17082]

Step 3: Minimize flow to approximate the reference path

- Train on a **regression** target

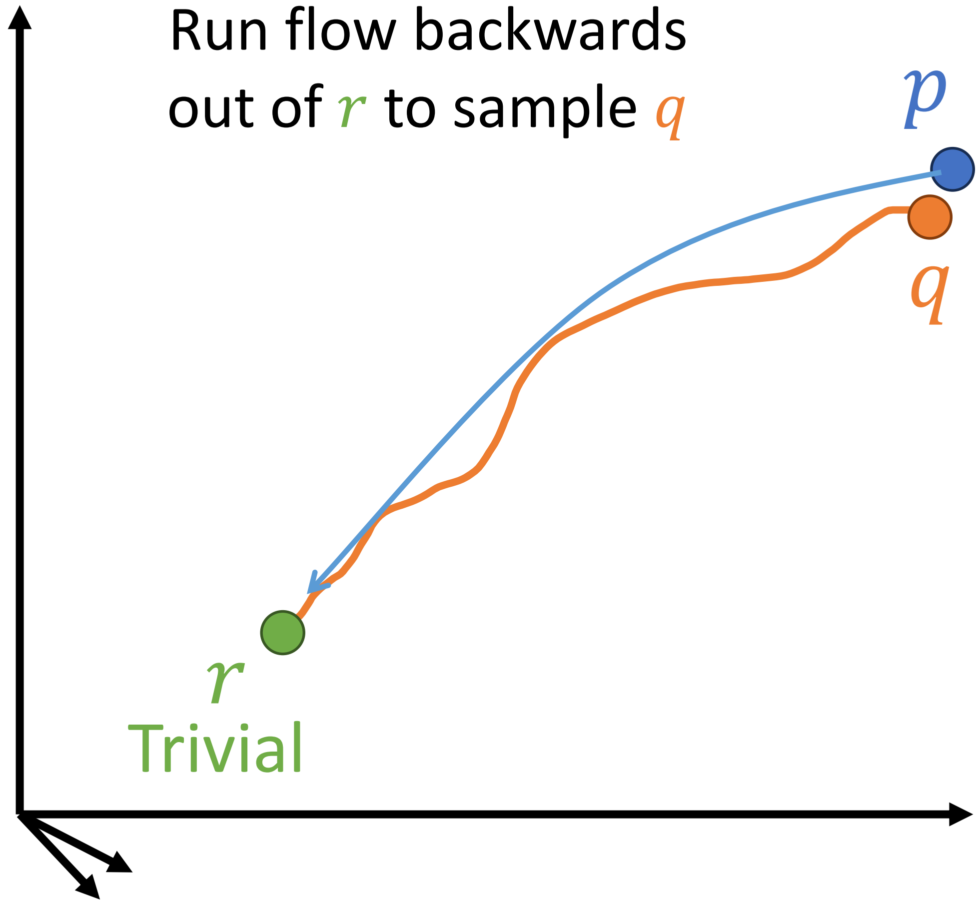
$$L(\theta) \sim \int_0^1 dt \left\langle \left[s_\theta(t) - \begin{pmatrix} \text{known,} \\ \text{simple fn} \end{pmatrix} (t) \right]^2 \right\rangle$$

Advantages:

- No evaluation of p, q – don't need to run the flow
- Sampling $p(t)$ usually trivial for arbitrary t
- Parallelize training across entire path

* **Stochastic interpolants / flow matching:** same picture, but define a reference path by interpolating between r, p samples, and constrain a deterministic flow.

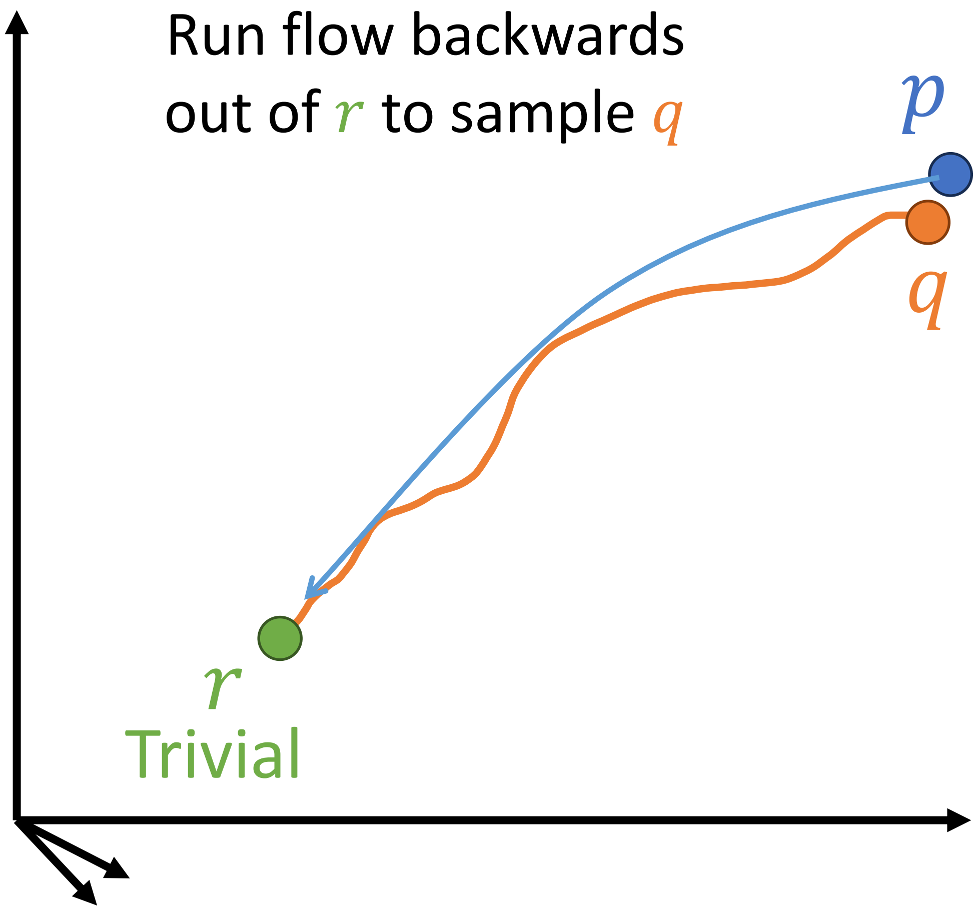
Diffusions: sampling



Diffusions: sampling

Not fundamentally different from flows!

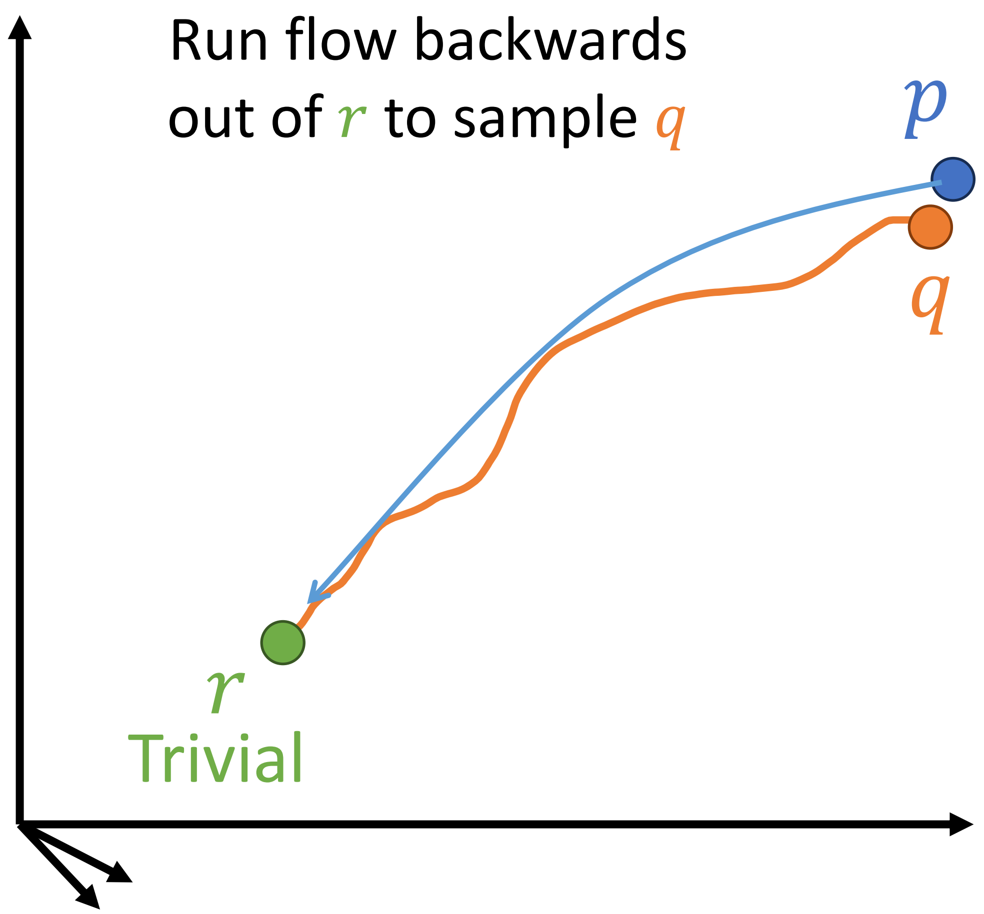
Model q needs correction → target p



Diffusions: sampling

Not fundamentally different from flows!

Model q needs correction → target p



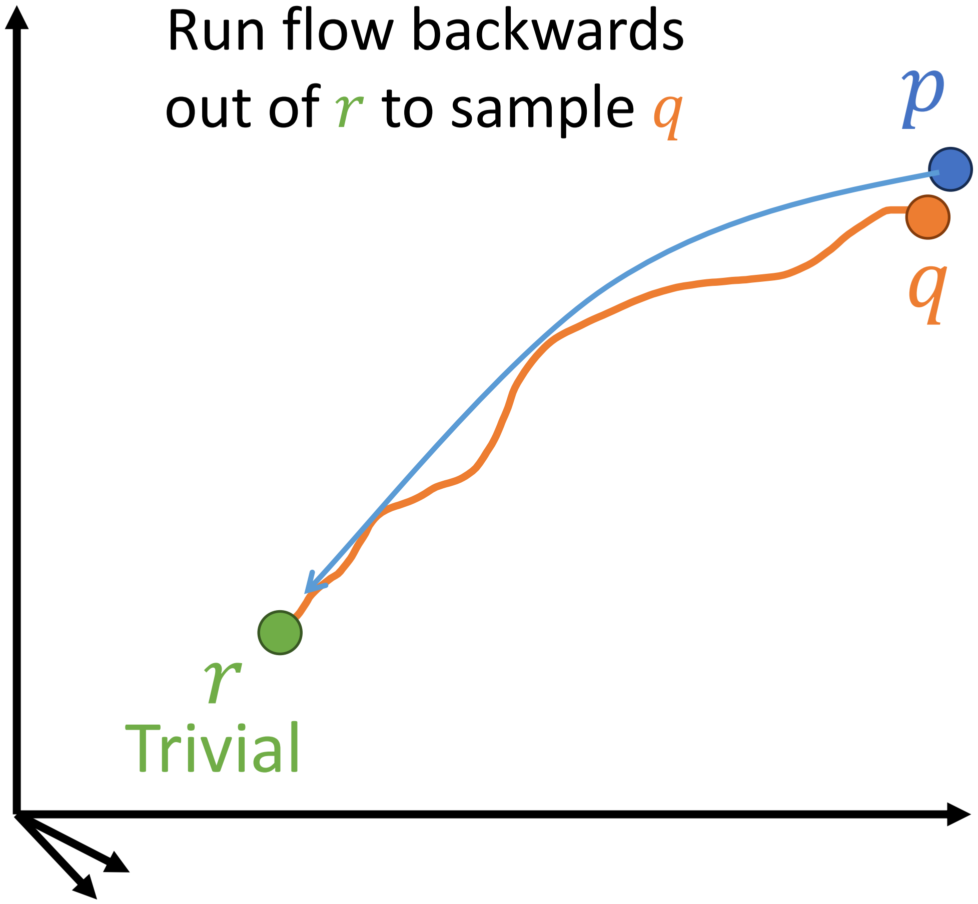
Mode 1: Deterministic Flows

Every SDE has a corresponding "probability flow ODE" which samples ≈ same q

→ Evaluate as an ODE flow

→ Diffusions are just better CNF training.

Diffusions: sampling



Not fundamentally different from flows!

Model q needs correction → target p

Mode 1: Deterministic Flows

Every SDE has a corresponding "probability flow ODE" which samples ≈ same q

→ Evaluate as an ODE flow

→ Diffusions are just better CNF training.

...but new kinds of sampling schemes are possible

Mode 2: Stochastic Flows

Integrate (discretized) SDE

Con: irreducible path noise

Pro: may allow less-constrained architectures!

Some schemes under exploration, but **underexplored**

Diffusions/etc: WIP

Rapid recent progress in this area!

[Cotler Rezchikov 2308.12355]

[Wang Aarts Zhou 2309.17082]

[Wang Aarts Zhou 2311.03578]

[Albergo Vanden-Eijnden 2410.02711]

[Aarts Habibi Wang Zhou 2410.21212]

[Fukushima Kamata Hirono 2411.11297]

[Habibi Aarts Wang Zhou 2412.01919]

[Chen Aarts Lucini 2502.02127]

[Aarts Habibi Wang Zhou 2510.01328]

[Ihssen Kapust Pawlowski 2510.26678]

[Chen Qian Aarts Lucini Zhou 2606.13790]

[Tan Aarts Habibi Lucini Wang 2607.08505]

[Tomiya 2607.12587]

Bold = new since Lattice 2025!

Diffusions/etc: WIP

Rapid recent progress in this area!

[Cotler Rezchikov 2308.12355]

[Wang Aarts Zhou 2309.17082]

[Wang Aarts Zhou 2311.03578]

[Albergo Vanden-Eijnden 2410.02711]

[Aarts Habibi Wang Zhou 2410.21212]

[Fukushima Kamata Hirono 2411.11297]

[Habibi Aarts Wang Zhou 2412.01919]

[Chen Aarts Lucini 2502.02127]

[Aarts Habibi Wang Zhou 2510.01328]

[Ihssen Kapust Pawlowski 2510.26678]

[Chen Qian Aarts Lucini Zhou 2606.13790]

[Tan Aarts Habibi Lucini Wang 2607.08505]

[Tomiya 2607.12587]

Emerging: gauge theory applications

[Zhu Aarts Wang Zhou Wang 2502.05504]

[Vega Komijani El-Khadra Marinkovic 2510.26081]

[Kanwar Vega 2512.19877]

[Aarts Habibi Ipp Müller Ranner Wang Wang Zhu 2601.19552]

[Alharazin Panteleeva Sun 2602.09045]

[Komijani Marinkovic Turgut 2605.06134]

[Komijani 2605.17326]

[Vega El-Khadra 2606.27481]

Bold = new since Lattice 2025!

Diffusions/etc: WIP

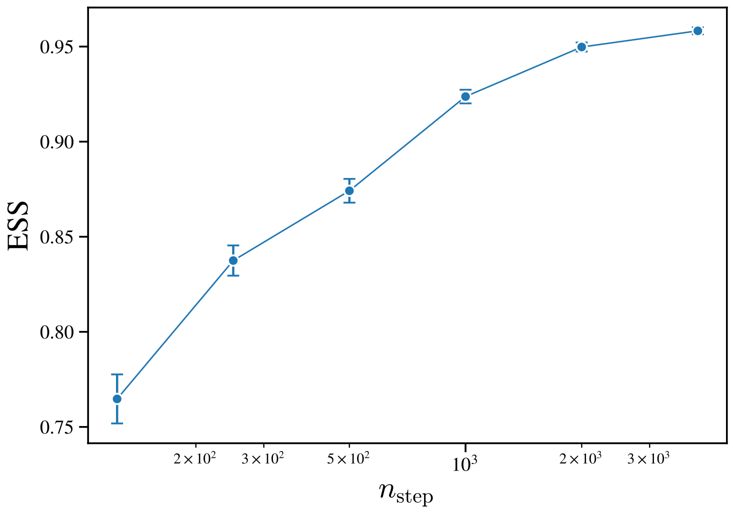
Rapid recent progress in this area!

[Cotler Rezchikov 2308.12355]
[Wang Aarts Zhou 2309.17082]
[Wang Aarts Zhou 2311.03578]
[Albergo Vanden-Eijnden 2410.02711]
[Aarts Habibi Wang Zhou 2410.21212]
[Fukushima Kamata Hirono 2411.11297]
[Habibi Aarts Wang Zhou 2412.01919]
[Chen Aarts Lucini 2502.02127]
[Aarts Habibi Wang Zhou 2510.01328]
[Ihssen Kapust Pawlowski 2510.26678]
[Chen Qian Aarts Lucini Zhou 2606.13790]
[Tan Aarts Habibi Lucini Wang 2607.08505]
[Tomiya 2607.12587]

Bold = new since Lattice 2025!

[338] O. Vega (Th15:00) — Unbiased Diffusion-Based Generation via Scalable Non-Equilibrium Transport

Diffusions + SNF-like MCMC insertions drive ESS → 1



Emerging: gauge theory applications

[Zhu Aarts Wang Zhou Wang 2502.05504]
[Vega Komijani El-Khadra Marinkovic 2510.26081]
[Kanwar Vega 2512.19877]
[Aarts Habibi Ipp Müller Ranner Wang Wang Zhu 2601.19552]
[Alharazin Panteleeva Sun 2602.09045]
[Komijani Marinkovic Turgut 2605.06134]
[Komijani 2605.17326]
[Vega El-Khadra 2606.27481]

Relevant talks at Lattice 2026:
[64] F. Romero López (Th14:20) — Learning the generating functional for variance reduction in lattice QCD
[228] P. Butti (F14:40) — A variational framework for variance reduction in lattice field theory
[31] D. Schuh (Tu16:10) — Tackling the Sign Problem in the Doped Hubbard Model with Normalizing Flows
[147] R. Kapust (Th14:00) — Solving sign problems with physics-informed kernels
[275] E. Ortiz Pacheco (Th15:20) — Alleviating the sign problem with contour deformations in heavy-dense QCD

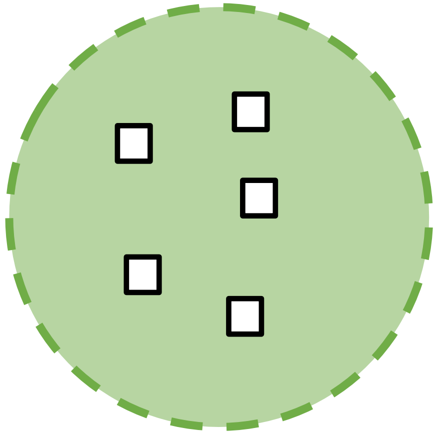
Applications of shorter flows:
Flow-based variance reduction & control variates

See also:
[217] L. Jin (M15:00) — Noise reduction for disconnected diagrams
[175] B. Luke (W10:20) — Optimized solver algorithms for computing Tr(A^-1)
[44] S. Lahrtz (W11:30) — Improved calculational setup for isospin-breaking corrections to the HVP contribution to g-2
[42] A. Evangelista (F15:00) — LMA performance on twisted-mass fermions ensembles at physical pion masses
[182] V. Moningi (F16:10) — Leveraging Existing HISQ Eigenvectors for Improved Statistical Precision in the HVP Contribution to muon g-2
[117] F. Margari (F16:50) — The smeared R-ratio in isospin symmetric QCD from first-principles lattice simulations

Correlated ensembles & the derivative trick

A flowed ensemble is correlated w/ the original

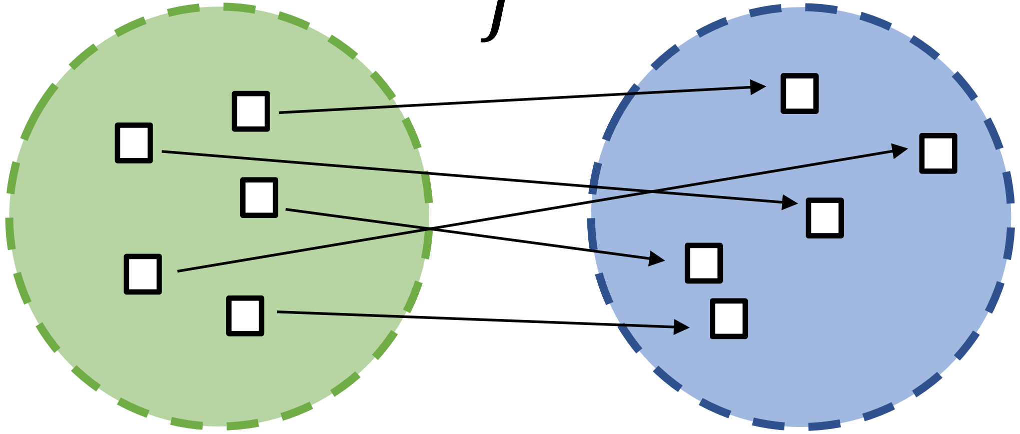
$\{U\} \sim r$



Correlated ensembles & the derivative trick

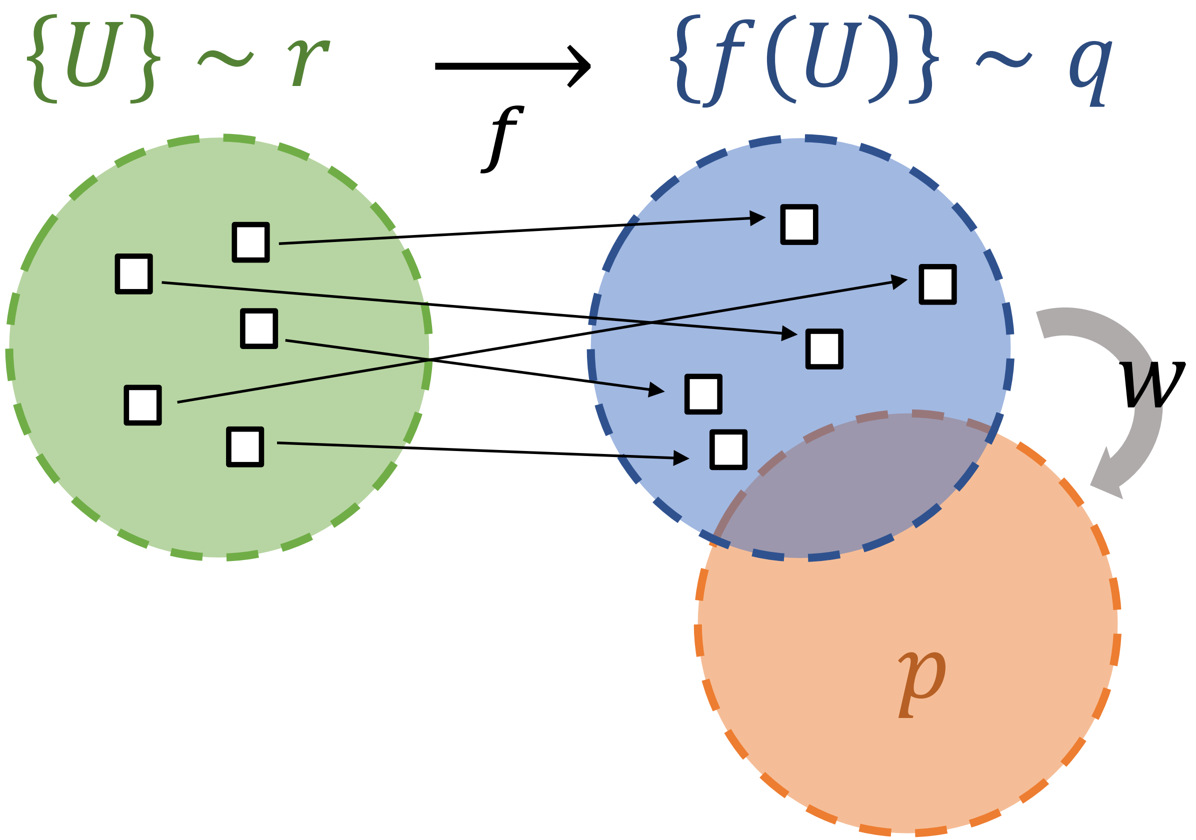
A flowed ensemble is correlated w/ the original

$$\{U\} \sim r \xrightarrow{f} \{f(U)\} \sim q$$



Correlated ensembles & the derivative trick

A flowed ensemble is correlated w/ the original



Correlated ensembles & the derivative trick

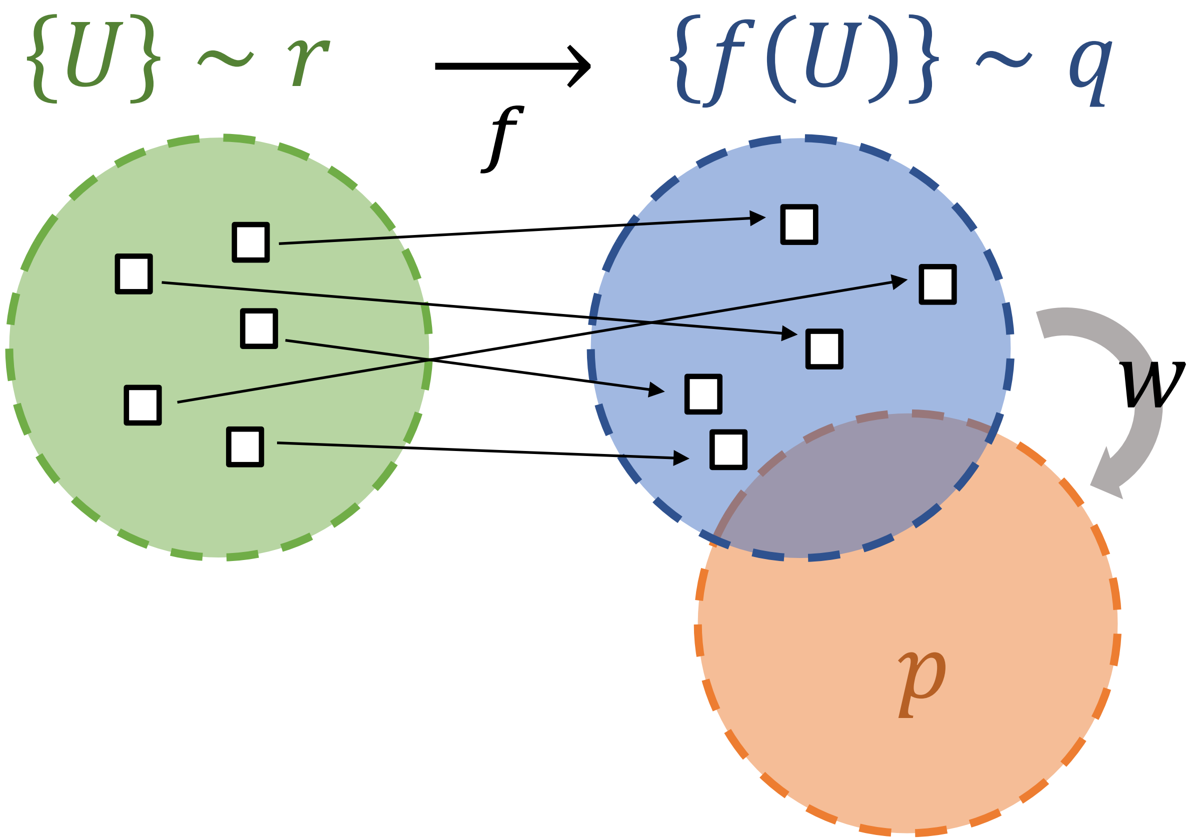
Perturb theory w/ some operator $\mathcal{Q}$

$$S \to S_\lambda = S - \lambda \mathcal{Q}$$

Derivative inserts $\mathcal{Q}$ into n-pt fn $\langle \mathcal{O} \rangle$

$$\left.\frac{\partial \langle \mathcal{O} \rangle_\lambda}{\partial \lambda}\right|_{\lambda=0} = \langle \mathcal{O}\mathcal{Q} \rangle_0 - \langle \mathcal{O} \rangle_0 \langle \mathcal{Q} \rangle_0$$

A flowed ensemble is correlated w/ the original



Correlated ensembles & the derivative trick

Perturb theory w/ some operator $\mathcal{Q}$

$$S \to S_\lambda = S - \lambda \mathcal{Q}$$

Derivative inserts $\mathcal{Q}$ into n-pt fn $\langle \mathcal{O} \rangle$

$$\left.\frac{\partial \langle \mathcal{O} \rangle_\lambda}{\partial \lambda}\right|_{\lambda=0} = \langle \mathcal{O}\mathcal{Q} \rangle_0 - \langle \mathcal{O} \rangle_0 \langle \mathcal{Q} \rangle_0$$

The trick: Finite-difference approx.:

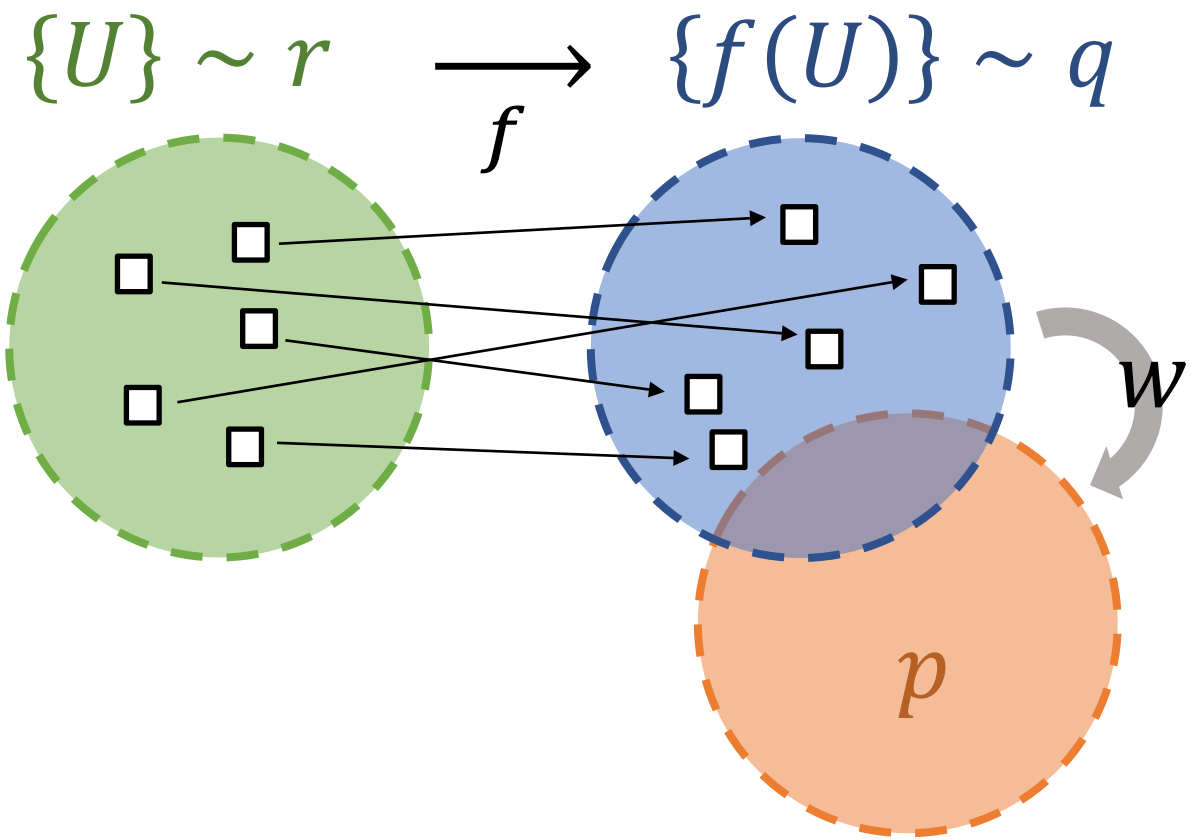
$$\left.\frac{\partial \langle \mathcal{O} \rangle_\lambda}{\partial \lambda}\right|_0 \approx \frac{1}{\lambda}\left[\langle \mathcal{O} \rangle_\lambda - \langle \mathcal{O} \rangle_0\right]$$

Train a flow f to map $S_0 \to S_\lambda$, use to evaluate both expectations on same ensemble

$$= \frac{1}{\lambda}\Big\langle w\big(f(U)\big)\mathcal{O}\big(f(U)\big) - \mathcal{O}(U)\Big\rangle_0$$

→ **Correlated noise cancellation!**

A flowed ensemble is correlated w/ the original



Correlated ensembles & the derivative trick

Perturb theory w/ some operator $\mathcal{Q}$

$$S \to S_\lambda = S - \lambda \mathcal{Q}$$

Derivative inserts $\mathcal{Q}$ into n-pt fn $\langle \mathcal{O} \rangle$

$$\left.\frac{\partial \langle \mathcal{O} \rangle_\lambda}{\partial \lambda}\right|_{\lambda=0} = \langle \mathcal{O}\mathcal{Q} \rangle_0 - \langle \mathcal{O} \rangle_0 \langle \mathcal{Q} \rangle_0$$

The trick: Finite-difference approx.:

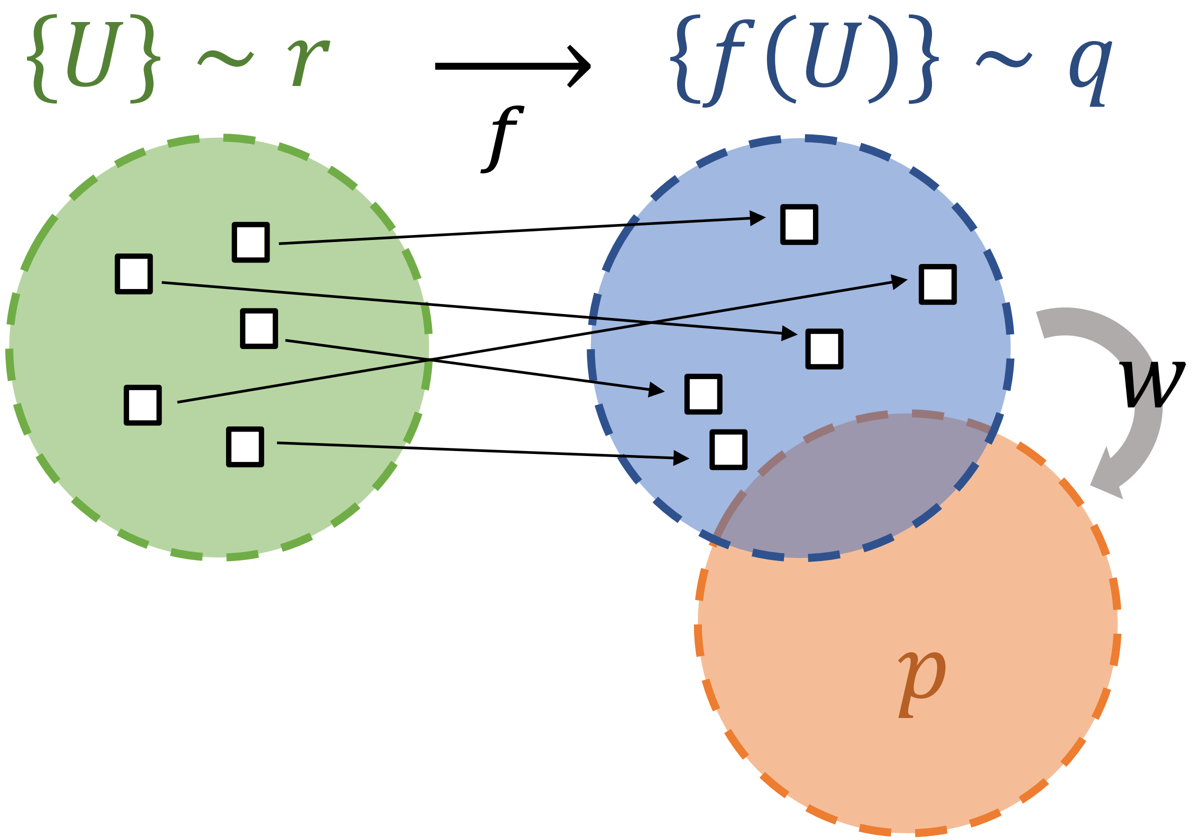
$$\left.\frac{\partial \langle \mathcal{O} \rangle_\lambda}{\partial \lambda}\right|_0 \approx \frac{1}{\lambda}\left[\langle \mathcal{O} \rangle_\lambda - \langle \mathcal{O} \rangle_0\right]$$

Train a flow f to map $S_0 \to S_\lambda$, use to evaluate both expectations on same ensemble

$$= \frac{1}{\lambda}\left\langle w\big(f(U)\big)\mathcal{O}\big(f(U)\big) - \mathcal{O}(U)\right\rangle_0$$

→ **Correlated noise cancellation!**

A flowed ensemble is correlated w/ the original



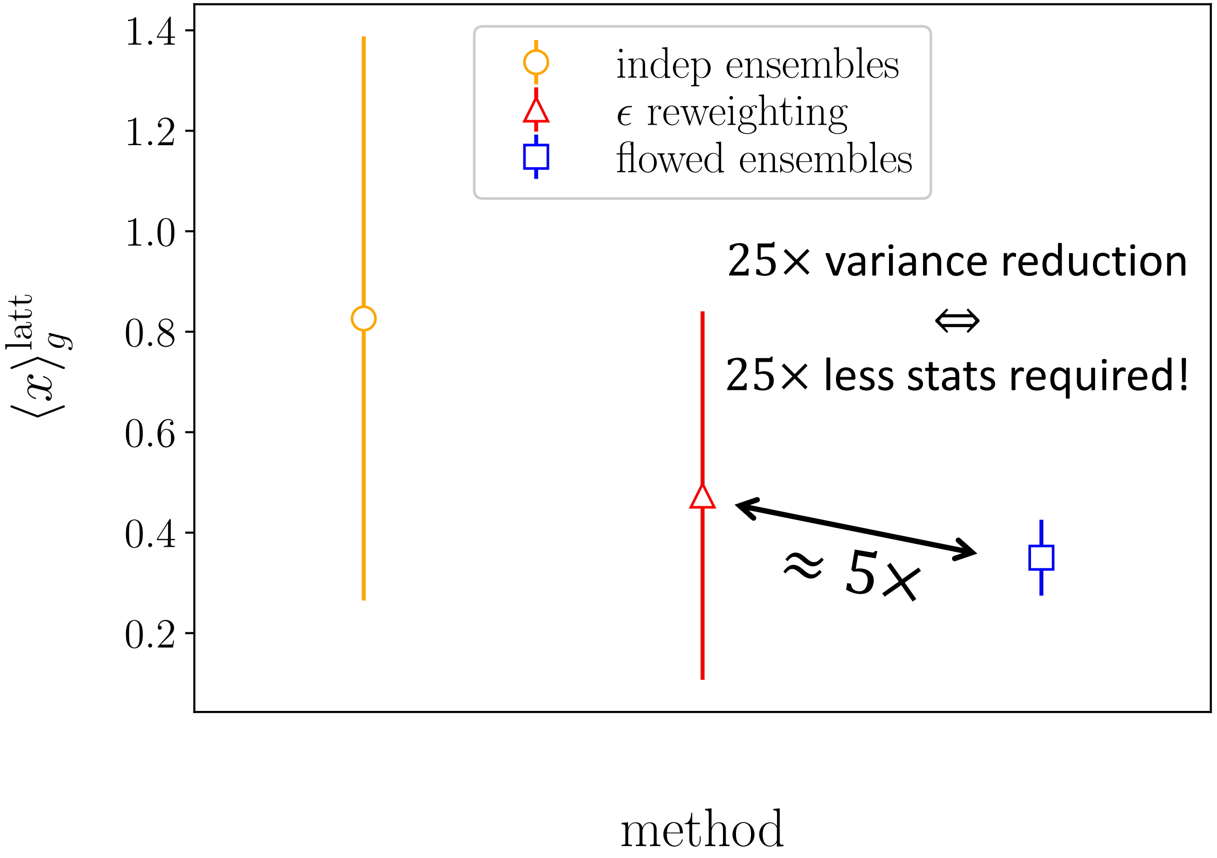
Idea also generalizes straightforwardly to Feynman-Hellmann method

Demos: pion $\langle x \rangle_g$ w/ flowed Feynman-Hellmann

First demo: Quenched QCD

$8^3{\times}16 \quad \beta = 6 \quad \kappa = 0.132$

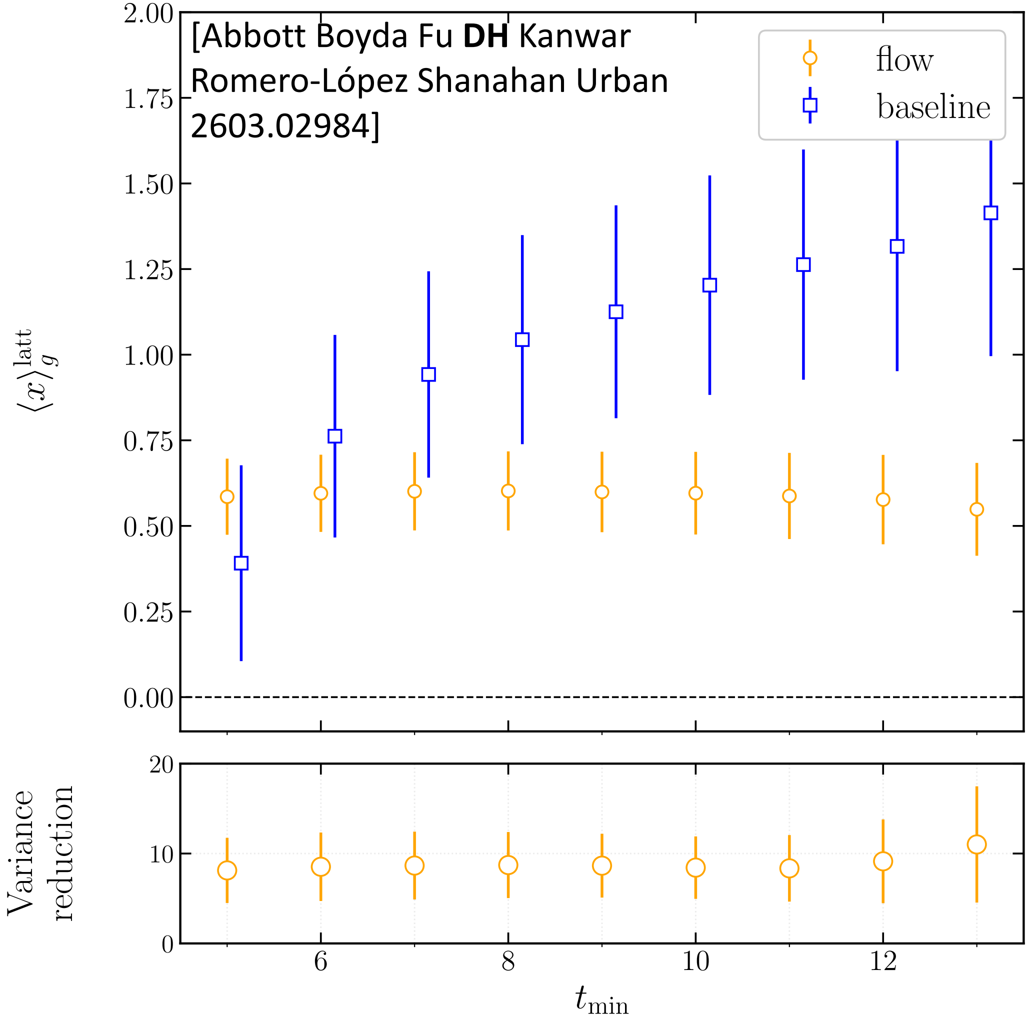
[Abbott Botev Boyda **DH** Kanwar Racanière Rezende Romero-López Shanahan Urban 2401.10874]



New: extension to dynamical QCD

$N_f = 2$ twisted mass

$16^3{\times}32 \quad a \approx 0.10$ fm $\quad M_\pi \approx 540$ MeV



Volume scalability

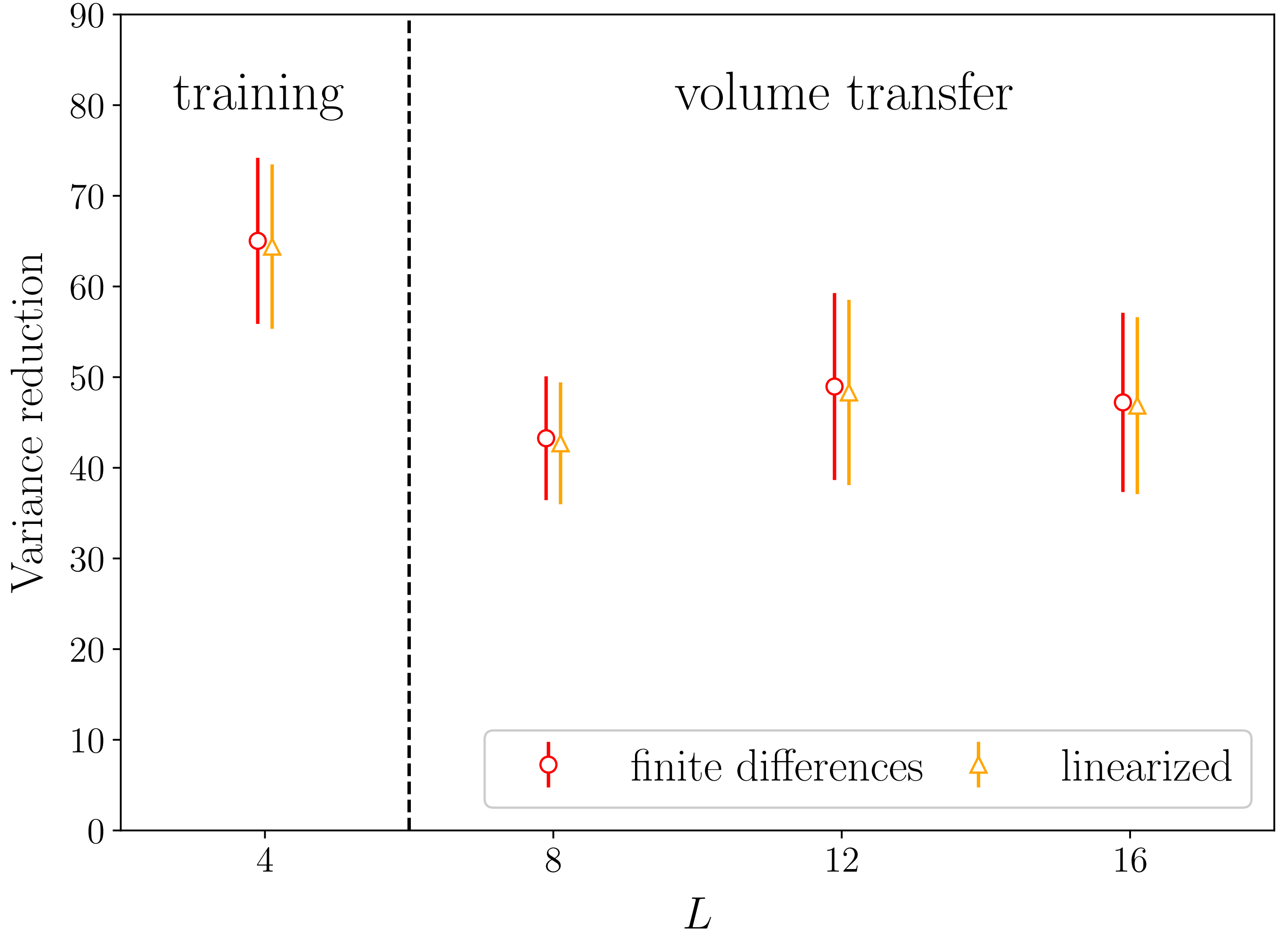
[Abbott Boyda Fu **DH** Kanwar
Romero-López Shanahan Urban
2603.02984]

No degradation of performance with increasing volume!

→ No obstacle to at-scale applications

Works by different mechanism than sampling; ESS not the correct metric

Closely related: [Catumba Ramos 2502.15570]



Example: glueball two-point $C(t)$ at $t = 2a$ in Yang-Mills on $L^3 \times 32$

Linearized formalism

[Abbott Boyda Fu **DH** Kanwar Romero-López Shanahan Urban 2603.02984]

Linearize flow for small λ

Algebra-valued "flow field"

$$f_\lambda(U) = [1 + i\lambda \boxed{F(U)} + \cdots]U$$

Expand finite difference

$$\left.\frac{d\langle \mathcal{O} \rangle}{d\lambda}\right|_{\lambda=0} = \lim_{\lambda \to 0} \frac{1}{\lambda} \left\langle w_\lambda(U) \mathcal{O}\big(f_\lambda(U)\big) - \mathcal{O}(U) \right\rangle_0$$

Linearized formalism

[Abbott Boyda Fu **DH** Kanwar Romero-López Shanahan Urban 2603.02984]

Linearize flow for small λ

Algebra-valued "flow field"

$$f_\lambda(U) = [1 + i\lambda F(U) + \cdots]U$$

Expand finite difference

$$\left.\frac{d\langle \mathcal{O}\rangle}{d\lambda}\right|_{\lambda=0} = \lim_{\lambda\to 0}\frac{1}{\lambda}\left\langle w_\lambda(U)\mathcal{O}\big(f_\lambda(U)\big) - \mathcal{O}(U)\right\rangle_0$$

$$= \langle \mathcal{O}\mathcal{Q}\rangle_0 - \langle \mathcal{O}\rangle_0\langle \mathcal{Q}\rangle_0 - \langle \mathcal{L}[F\mathcal{O}]\rangle_0$$

Original observable

Control variate:

- $\langle \cdots \rangle = 0 \quad \Rightarrow$ no bias
- Correlated with $\langle \mathcal{O}\mathcal{Q}\rangle_0 - \langle \mathcal{O}\rangle_0\langle \mathcal{Q}\rangle_0$
  $\Rightarrow$ Reduced variance estimator

Linearized formalism

[Abbott Boyda Fu **DH** Kanwar Romero-López Shanahan Urban 2603.02984]

Linearize flow for small λ

Algebra-valued "flow field"

$$f_\lambda(U) = [1 + i\lambda \boxed{F(U)} + \cdots]U$$

Expand finite difference

$$\left.\frac{d\langle\mathcal{O}\rangle}{d\lambda}\right|_{\lambda=0} = \lim_{\lambda\to 0}\frac{1}{\lambda}\left\langle w_\lambda(U)\mathcal{O}\big(f_\lambda(U)\big) - \mathcal{O}(U)\right\rangle_0$$

$$= \boxed{\langle\mathcal{O}\mathcal{Q}\rangle_0 - \langle\mathcal{O}\rangle_0\langle\mathcal{Q}\rangle_0} - \boxed{\langle\mathcal{L}[F\mathcal{O}]\rangle_0}$$

Original observable

Control variate:

- $\langle\cdots\rangle = 0 \quad\Rightarrow\quad$ no bias
- Correlated with $\langle\mathcal{O}\mathcal{Q}\rangle_0 - \langle\mathcal{O}\rangle_0\langle\mathcal{Q}\rangle_0$
  $\Rightarrow$ Reduced variance estimator

Stein operator:

$$\mathcal{L}[g] = \nabla\cdot g - g\cdot\nabla S$$

$\langle\mathcal{L}[g]\rangle = 0$ for any g

Linearized flows construct Stein / Schwinger-Dyson control variates!

Learning the generating functional

[Abbott Fu **DH** Kanwar Romero-López Shanahan 2606.15986]

$$S_{\boldsymbol{\lambda}} = S_0 - \lambda_1 O_1 - \lambda_2 O_2 - \cdots$$

For unnormalized weight $\widehat{w}_{\boldsymbol{\lambda}} = e^{-[S_\lambda - S_0]} \det J$,

$$\langle \widehat{w}_{\boldsymbol{\lambda}} \rangle = \frac{Z_{\boldsymbol{\lambda}}}{Z_0}$$

is the *generating functional!*

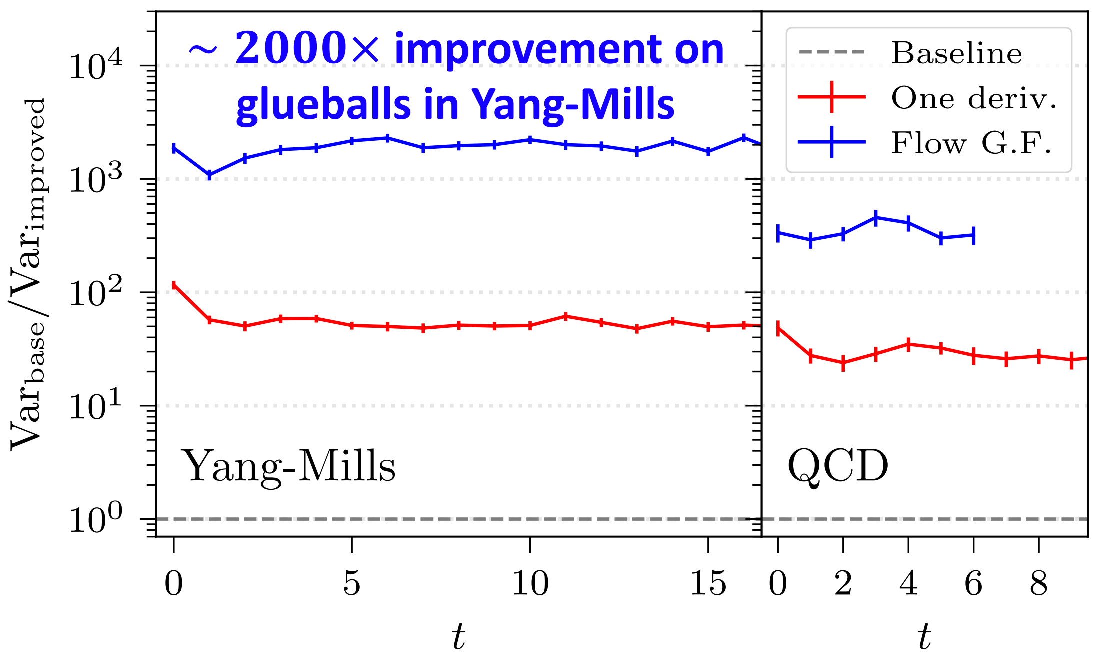
$$\Rightarrow \langle O_1 O_2 \dots O_n \rangle = \frac{\partial^n}{\partial\lambda_1 \partial\lambda_2 \cdots \partial\lambda_n} \langle \widehat{w}_{\boldsymbol{\lambda}} \rangle$$

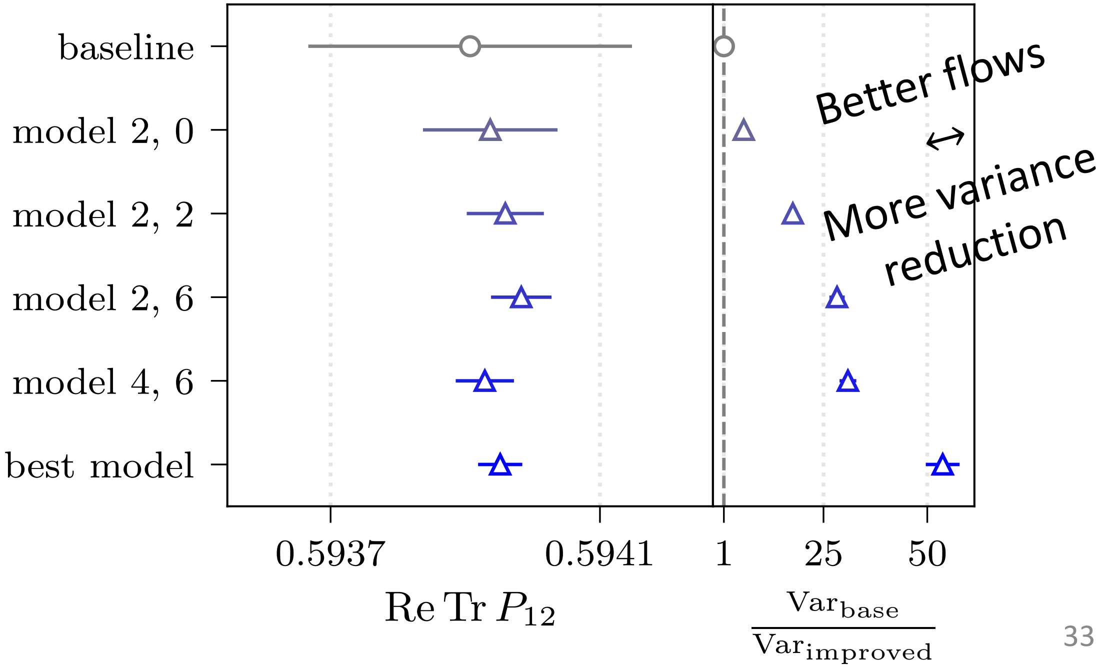
Train flows to insert each operator, compose to improve multiple derivatives

Different timeslices for free

When linearized, also admits interpretation as a Stein/SD control variate

Zero-variance estimators in limit of perfect flows!





Learning the generating functional

[Abbott Fu **DH** Kanwar Romero-López Shanahan 2606.15986]

$$S_\lambda = S_0 - \lambda_1 O_1 - \lambda_2 O_2 - \cdots$$

For unnormalized weight $\widehat{w}_\lambda = e^{-[S_\lambda - S_0]} \det J$,

$$\langle \widehat{w}_\lambda \rangle = \frac{Z_\lambda}{Z_0}$$

is the *generating functional!*

$$\Rightarrow \langle O_1 O_2 \dots O_n \rangle = \frac{\partial^n}{\partial\lambda_1 \partial\lambda_2 \cdots \partial\lambda_n} \langle \widehat{w}_\lambda \rangle$$

Train flows to insert each operator, compose to improve multiple derivatives

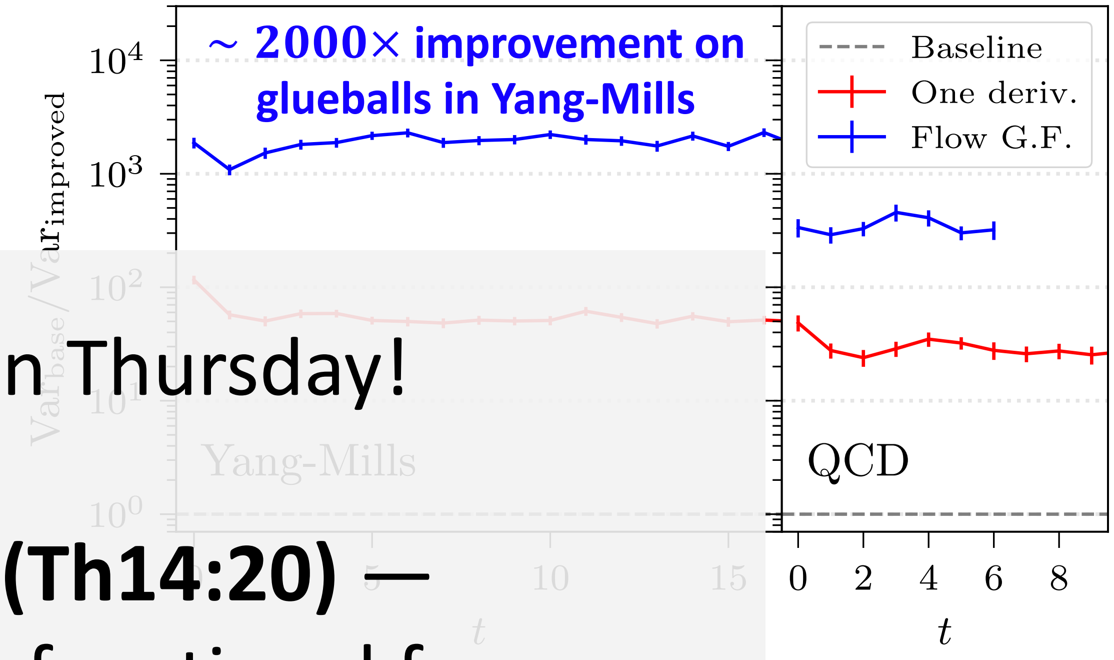
Different timeslices for free

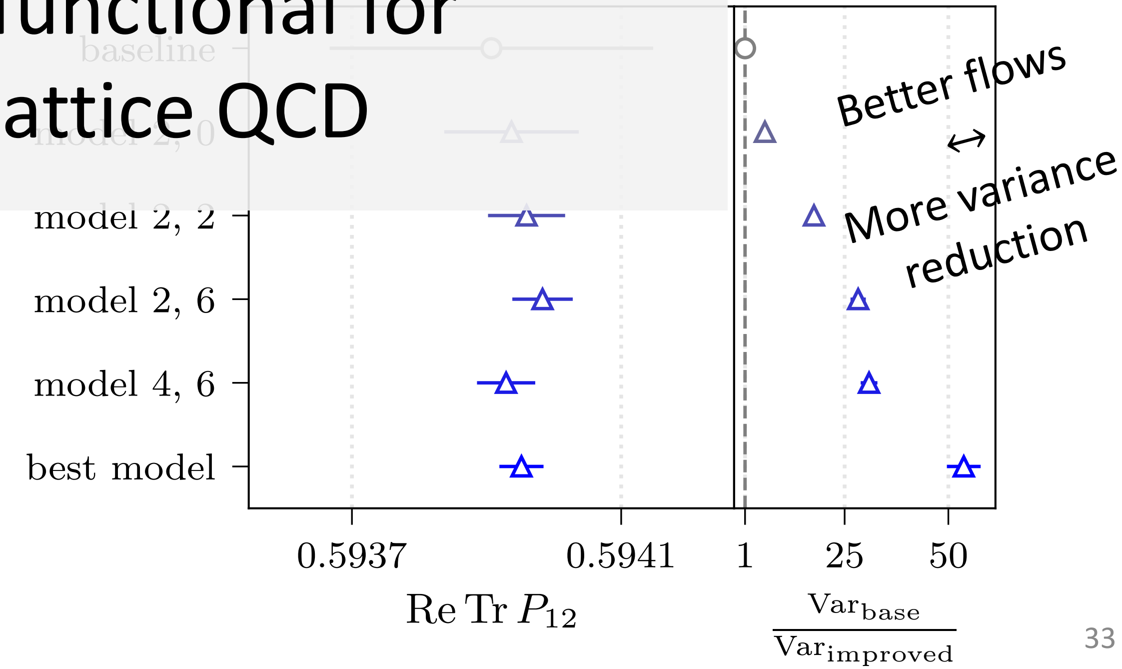
When linearized, also admits interpretation as a Stein/SD control variate

Zero-variance estimators in limit of perfect flows!

See Fernando's talk on Thursday!

[64] F. Romero López (Th14:20) —
Learning the generating functional for variance reduction in lattice QCD





Flows for variance reduction

Framework to construct families of useful Stein / Schwinger-Dyson control variates

Linearized formalism:

- No recomputation of observables on many flowed ensembles
- Instead, compute extra "flow-field contractions" in hadronic correlators

New, independent scaling direction to improve QCD calculations

- Use ML to reduce noise at *fixed ensemble size,* just additional meas. cost

WIP: fermionic operators (see Fernando's talk!)

WIP: demonstration in an at-scale hadron structure calculation

[304] V. Ayyar (poster) — Agentic AI for Accelerating Lattice Gauge Theory Research
[328] A. Grebe (M14:20) — Toward Compute-Bound Lattice-QCD Software
[129] L. Hostetler (M15:00) — PerfAdvisor: An LLM-based Agent for Diagnosing Bottlenecks from GPU Systems Profiles
[101] E. Lundstrum (F14:00) — Agentic AI for QCD perturbation theory or: how I learned to stop worrying and love the bot
[214] T. Wada (F14:20) — AI-Assisted Higher-Order Hopping-Parameter Expansion in Lattice QCD

Agentic AI

Agentic AI

Late 2025: ~~coding~~ computational science agents suddenly got very good, including at lattice-related math and tasks

Emerging Applications

Agentic AI

Late 2025: ~~coding~~ computational science agents suddenly got very good, including at lattice-related math and tasks

Emerging Applications

Code optimization

Perfect verification signal! No danger.

Hardware/task-specific optimizations now feasible

See [328] A. Grebe (M14:20)

[129] L. Hostetler (M15:00)

Agentic AI

Late 2025: ~~coding~~ computational science agents suddenly got very good, including at lattice-related math and tasks

Emerging Applications

Code optimization

Perfect verification signal! No danger.
Hardware/task-specific optimizations now feasible
See [328] A. Grebe (M14:20)
[129] L. Hostetler (M15:00)

Hyperparameter Optimization

Tuning multigrid, HMC, etc (with some coaching)
See [304] V. Ayyar (poster)

NOTE: more labor-intensive approaches now feasible! e.g. many SNF/T-REX integrator setups

Agentic AI

Late 2025: ~~coding~~ computational science agents suddenly got very good, including at lattice-related math and tasks

FemtoMind

Emerging Applications

Code optimization

Perfect verification signal! No danger.
Hardware/task-specific optimizations now feasible
See [328] A. Grebe (M14:20)
[129] L. Hostetler (M15:00)

Hyperparameter Optimization

Tuning multigrid, HMC, etc (with some coaching)
See [304] V. Ayyar (poster)

NOTE: more labor-intensive approaches now feasible! e.g. many SNF/T-REX integrator setups

Workflow Orchestration

Running and debugging jobs on clusters, managing outputs, analyzing data
High reliability a requirement at expensive scales!

Agentic AI

Late 2025: ~~coding~~ computational science agents suddenly got very good, including at lattice-related math and tasks

Emerging Applications

Code optimization
Perfect verification signal! No danger.
Hardware/task-specific optimizations now feasible
See [328] A. Grebe (M14:20)
[129] L. Hostetler (M15:00)

Hyperparameter Optimization
Tuning multigrid, HMC, etc (with some coaching)
See [304] V. Ayyar (poster)

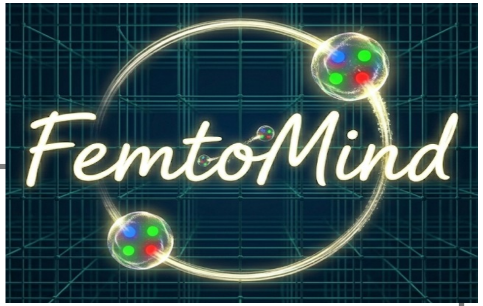
NOTE: more labor-intensive approaches now feasible! e.g. many SNF/T-REX integrator setups

Workflow Orchestration
Running and debugging jobs on clusters, managing outputs, analyzing data
High reliability a requirement at expensive scales!

Rapid prototyping
Coding up and testing new ideas at laptop/few-GPU scale
Off-the-shelf great for this with minimal coaching, checking

Agentic AI

Late 2025: ~~coding~~ computational science agents suddenly got very good, including at lattice-related math and tasks



Emerging Applications

Code optimization

Perfect verification signal! No danger.
Hardware/task-specific optimizations now feasible
See [328] A. Grebe (M14:20)
[129] L. Hostetler (M15:00)

Hyperparameter Optimization

Tuning multigrid, HMC, etc (with some coaching)
See [304] V. Ayyar (poster)

NOTE: more labor-intensive approaches now feasible! e.g. many SNF/T-REX integrator setups

Workflow Orchestration

Running and debugging jobs on clusters, managing outputs, analyzing data
High reliability a requirement at expensive scales!

Rapid prototyping

Coding up and testing new ideas at laptop/few-GPU scale
Off-the-shelf great for this with minimal coaching, checking

Literature Analysis

AI can digest & analyze hundreds of papers
(e.g. to help prepare a plenary talk)

Agentic AI

Late 2025: ~~coding~~ computational science agents suddenly got very good, including at lattice-related math and tasks

Emerging Applications

Code optimization
Perfect verification signal! No danger.
Hardware/task-specific optimizations now feasible
See [328] A. Grebe (M14:20)
[129] L. Hostetler (M15:00)

Hyperparameter Optimization
Tuning multigrid, HMC, etc (with some coaching)
See [304] V. Ayyar (poster)

NOTE: more labor-intensive approaches now feasible! e.g. many SNF/T-REX integrator setups

Workflow Orchestration
Running and debugging jobs on clusters, managing outputs, analyzing data
High reliability a requirement at expensive scales!

Rapid prototyping
Coding up and testing new ideas at laptop/few-GPU scale
Off-the-shelf great for this with minimal coaching, checking

Literature Analysis
AI can digest & analyze hundreds of papers
(e.g. to help prepare a plenary talk)

Completely new class of capabilities
As yet unclear what else is now possible.

Thermodynamics, entanglement, etc

Many calculations require expensive bespoke ensemble generation e.g.

- Equation of state (integral method)
- Replica-trick entanglement
- α_s with Schrodinger functional

Previous promising explorations of Jarczynski/AIS for these problems

(S)NFs can improve on this!

[Nicoli Anders Funcke Hartung Jansen Kessel Nakajima Stornati 2007.07115]
[Caselle Cellini Nada Panero 2210.03139]
[Bulgarelli Cellini Jansen Kühn Nada Nakajima Nicoli Panero 2410.14466]

Field-Transform HMC (FTHMC)

Luscher's original idea – seems to work when fermions dominate and using (very) simple flows

[Boyle 2401.16620] [Yamamoto Boyle Izubuchi Jin Lehner Matsumoto 2502.05452]

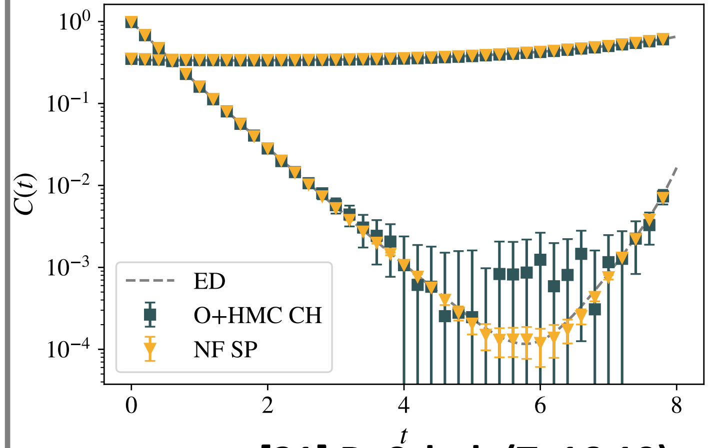
- **[282] P. Boyle (M15:20)** — Algorithms for domain wall fermions
- **[342] X.-Y. Jin (Tu15:00)** — Crafting the change of variables for HMC
- **[364] S. Yamamoto (F17:10)** — Comparison of Smearing Kernels for Field-Transformation HMC via the Master-Field Technique

Other improvements on HMC:

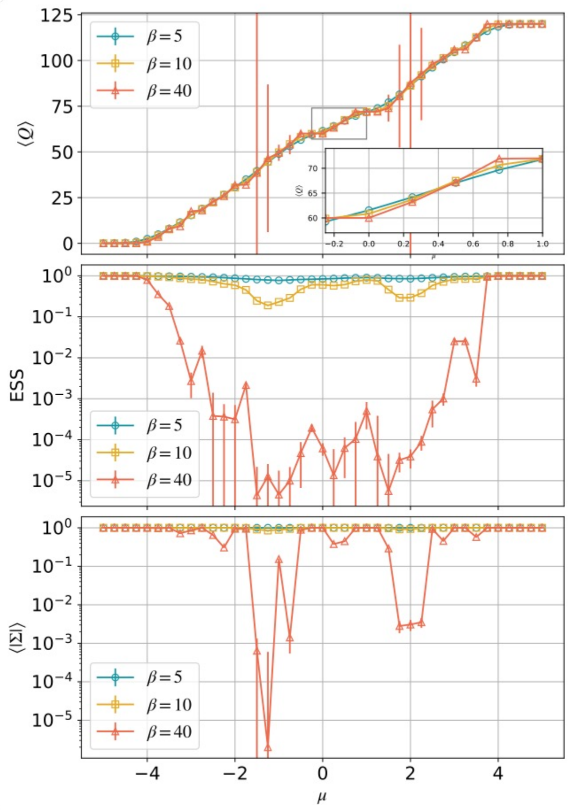
- **[301] N. Matsumoto (Tu16:30)** — AI/ML-inspired RMHMC kernels — CP^N model study
- **[254] B. Izett (F16:50)** — Simulating the string formulation of SU(3) lattice gauge theory

Condensed Matter applications

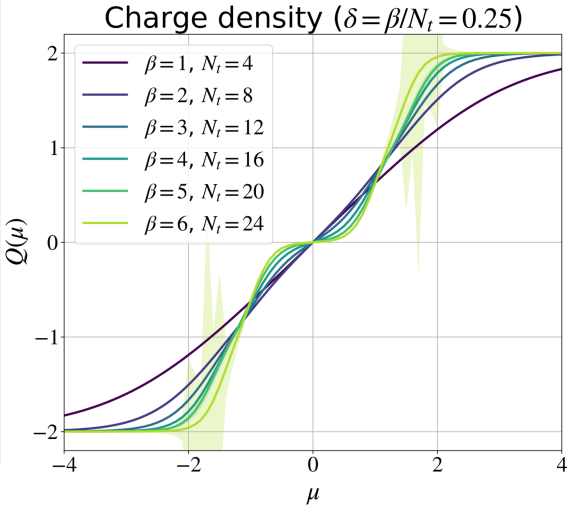
Fewer dof → easier to model than QCD
Much progress inside/outside hep-lat



[31] D. Schuh (Tu16:10) — Tackling the Sign Problem in the Doped Hubbard Model with Normalizing Flows



[30] J. Kreit (Tu16:50) — Generative Models for the Low-Temperature Hubbard Model



[18] F. Freche (Tu16:30) — Machine-Learned Density of States for the Doped Hubbard Model

Conclusions / Outlook

Comments on ML practice

Weak generalization to gauge theories
Overused

ϕ^4 theory

Trivial dynamics
Exact solutions exist
Overused

Pure gauge $U(1)$ $d=2$

Pure gauge $\mathrm{SU}(N)$ $d=2$

Simple fermionic theories
Nontrivial dynamics

Abelian

Schwinger model

No topology

QCD $d=2$

Seems to capture essential sampling challenges

Pure gauge $\mathrm{SU}(N)$ $d=4$

QCD $d=4$

Increasing testbed quality

Agentic AI changes the playing field

- AI coding → easy to implement more-complicated gauge demos
- AI optimization → true cost comparisons more feasible

For samplers, still not worth worrying about training costs

- …until we’ve found a viable, advantageous sampling scheme

Outlook

Flows for variance reduction beginning to bear fruit

- Early demonstrations of computational advantage*
- No volume scaling obstacles, can be deployed at SOTA scales

Promising path to solving topological freezing: hybrid approaches

- OBC defects with PTBC/Jarzynski is a solution, but requires large scale
- SNFs, T-REX: flows can push this scale down to more practical(?) levels

More general sampling: the quest for a scalable approach is on

- Diffusions/transport provide rich framework of opportunities

We can try all of our ideas!

- AI reduces obstacles to experimentation
- ML & other algorithms, new lattice formulations, …
- Whole new branches of fruit made low-hanging