Practical applications of machine-learned flows on gauge fields

Dan Hackett (Fermilab)

Machine-Learning-Based Sampling in Lattice Field Theory and Quantum Chemistry

Bonn, Germany

October 21, 2024

Flows: what are they for?

Flows are "bridges" between different distributions/theories/actions

$$r(U) = \frac{e^{-S_r(U)}}{Z_r} \quad \underset{f^{-1}}{\overset{f}{\Longleftrightarrow}} \quad q(U') = \frac{r(U)}{J_f(U)}$$

Exact bridge between r and q

Choose r, but flow induces q

For sampling applications: variationally optimize f so $q \approx p \propto e^{-S_p}$

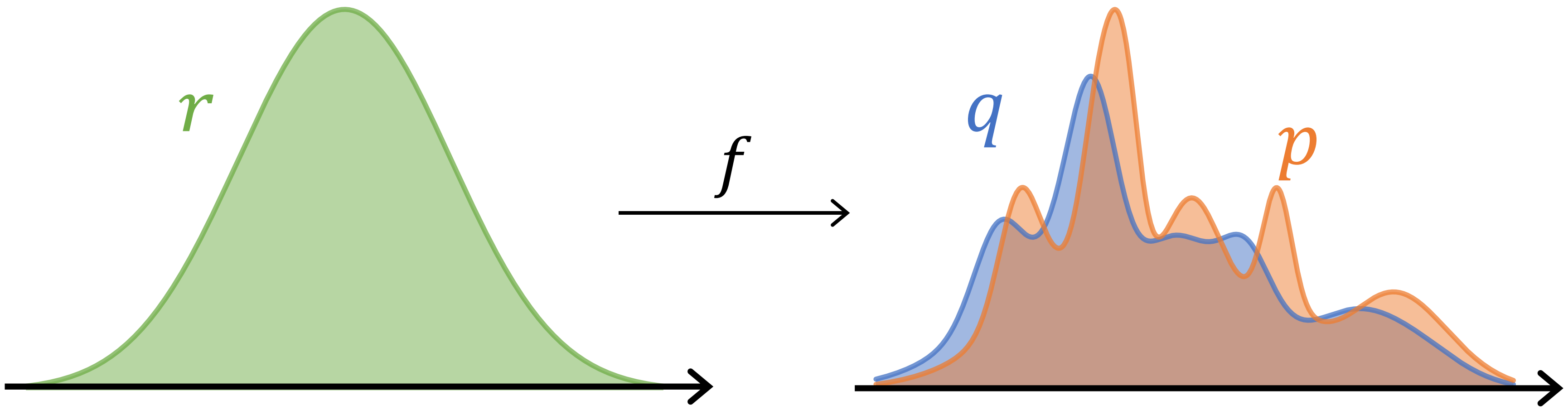
→ *Approximate* bridge between r and p

Requirements for exactness

Samples from model

$\phi_i \sim q$

$$\langle \mathcal{O} \rangle_p = \langle w\mathcal{O} \rangle_q = \int d\phi \, q(\phi) \, \frac{p(\phi)}{q(\phi)} \, \mathcal{O}(\phi) \approx \frac{1}{N} \sum_{i=1}^{N} w(\phi_i) \, \mathcal{O}(\phi_i)$$

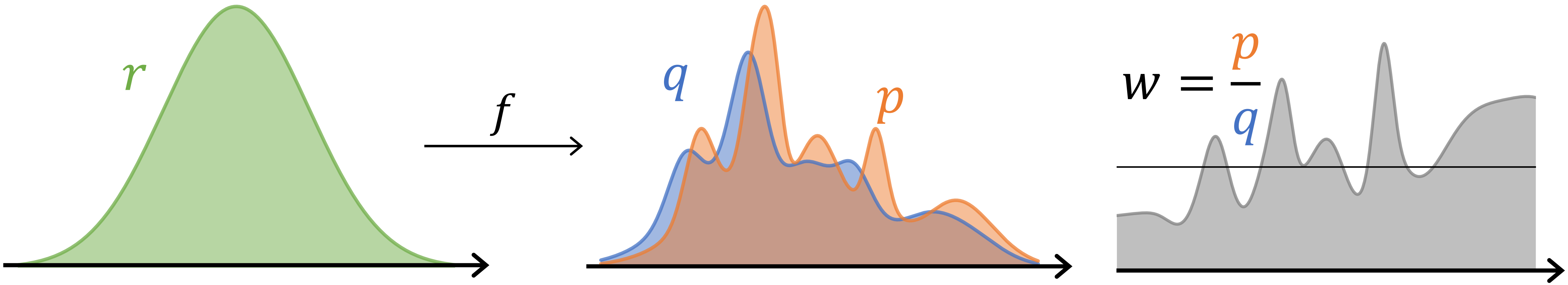


Requirements for exactness

Samples from model

$\phi_i \sim q$

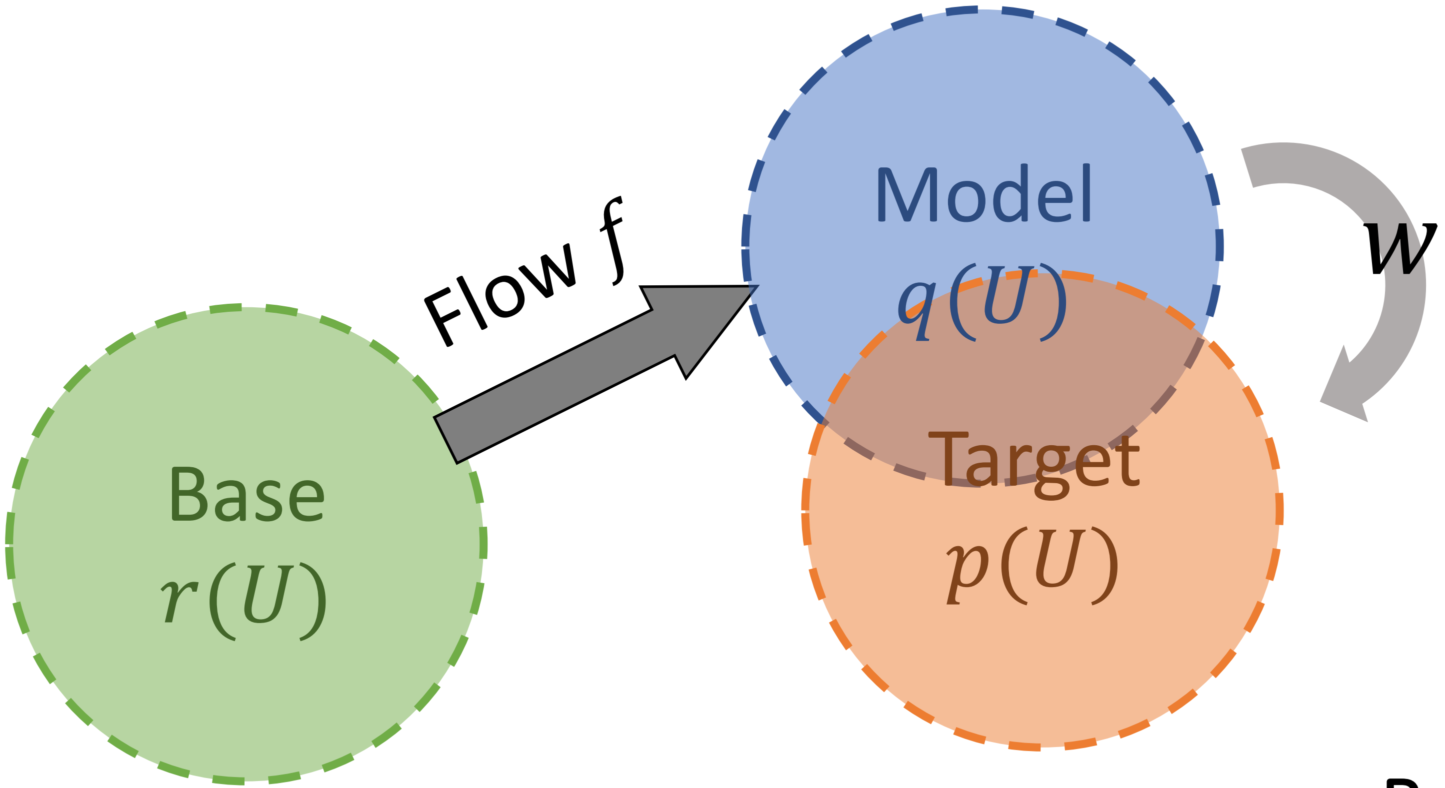
$$\langle \mathcal{O} \rangle_p = \langle w\mathcal{O} \rangle_q = \int d\phi \, q(\phi) \, \frac{p(\phi)}{q(\phi)} \, \mathcal{O}(\phi) \approx \frac{1}{N} \sum_{i=1}^{N} w(\phi_i) \, \mathcal{O}(\phi_i)$$



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

(Approximate) direct sampling with flows

Apply f to Haar uniform to get model q, tune f so $q \approx p \propto e^{-S_{\text{target}}}$



Reweight from $q \to p$

$w(U) = p(U) \,/\, q(U)$

$\langle O \rangle_p = \langle wO \rangle_q$

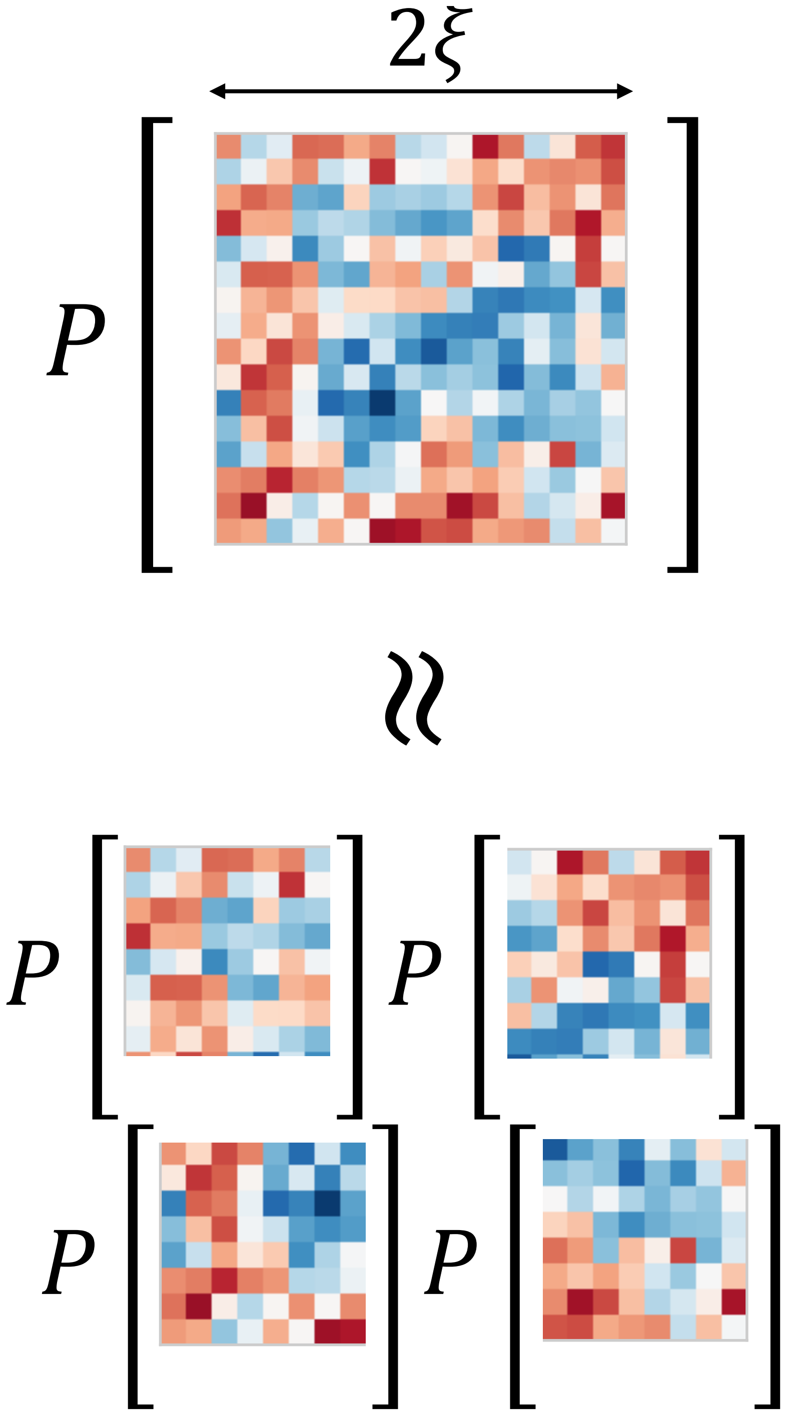
But: can choose whatever base distribution we want

Challenge: volume scaling

Locality: system factorizes into patches $\sim \xi$

$\rightarrow$ For a fixed model under transfer $V_0 \rightarrow V$,

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



Challenge: volume scaling

Locality: system factorizes into patches $\sim \xi$

$\rightarrow$ For a fixed model under transfer $V_0 \rightarrow V$,

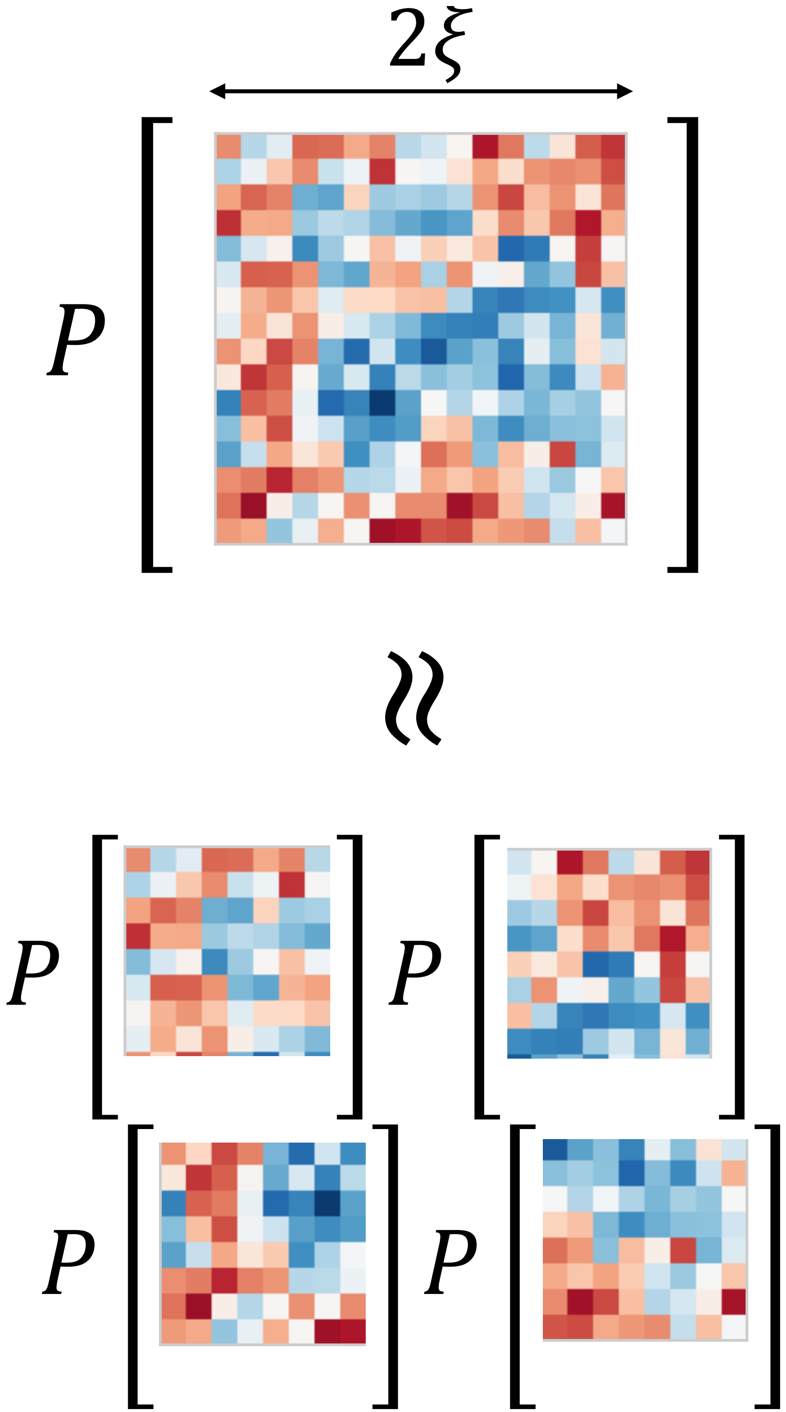
$$\mathrm{ESS}(V) \approx \mathrm{ESS}(V_0)^{V/V_0}$$

For QCD, $V \sim L^4$ so $L \rightarrow 2L$ means $V \rightarrow 16V$

$\rightarrow$ Need precise (at V_0) models!

Caveats:

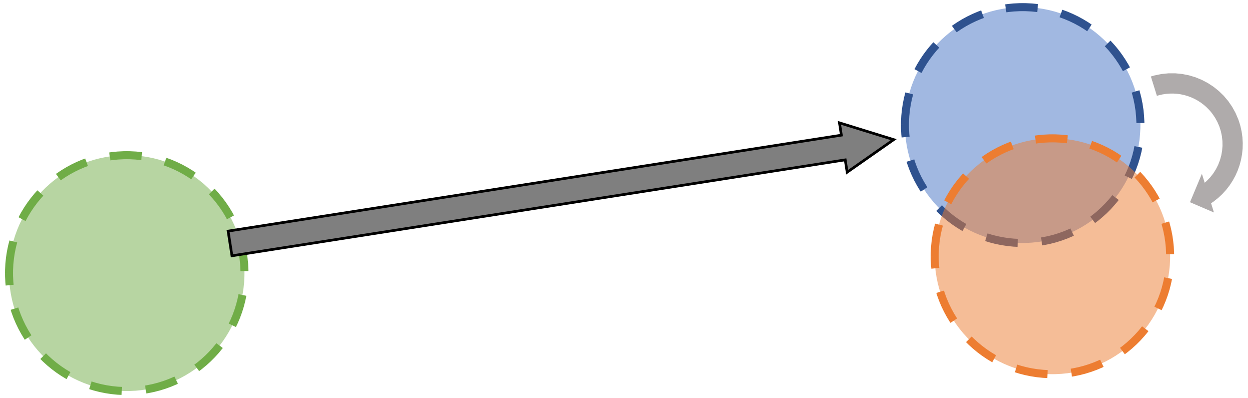
- Sometimes training directly at V beats transferring
- Obstacle to *thermodynamic limit* ($V \rightarrow \infty$, ξ fixed), but not to *continuum limit* ($\xi \rightarrow \infty$, V/ξ^d fixed)



Overview

~~Review/motivation~~

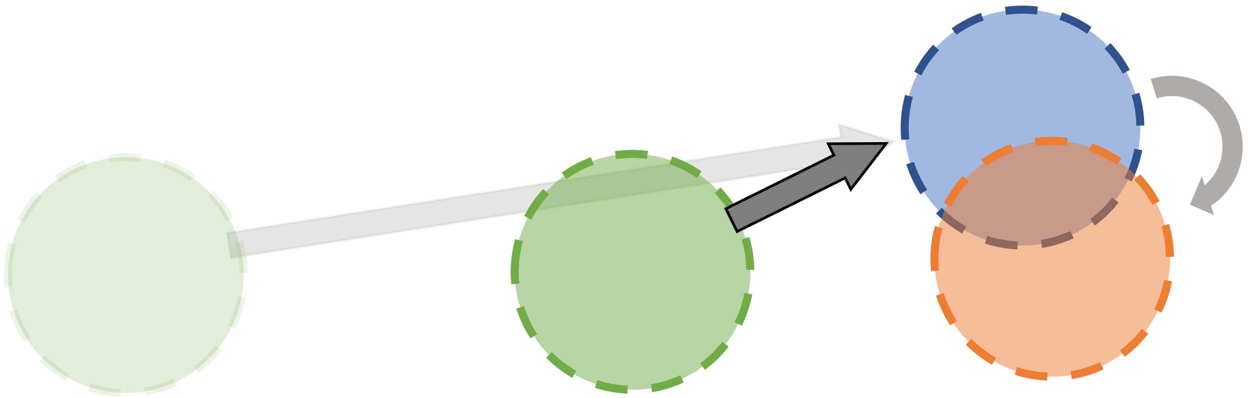
Basic idea:



Overview

~~Review/motivation~~

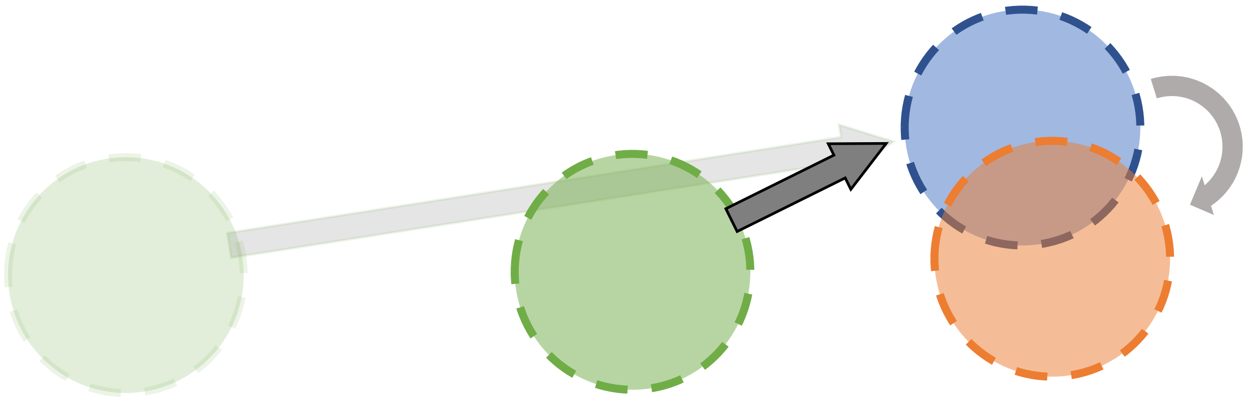
Basic idea: don't flow as far



Overview

~~Review/motivation~~

Basic idea: don't flow as far



Use cases:

1. Accelerating traditional sampling algorithms
2. Correlated ensembles (to improve derivative observables)

Collaborators (non-exhaustive)


Phiala Shanahan


Denis Boyda


Ryan Abbott


Julian Urban


Gurtej Kanwar


Michael Albergo

UNIVERSITÄT BERN


Fernando Romero-López


Kyle Cranmer

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

DeepMind


Sébastien Racanière


Danilo Rezende


Alex Matthews


Aleksandar Botev


Ali Razavi

Relevant Literature

[2401.10874]

Applications of flow models to the generation of correlated lattice QCD ensembles

Ryan Abbott,[1,2] Aleksandar Botev,[3] Denis Boyda,[1,2] Daniel C. Hackett,[4,1,2] Gurtej Kanwar,[5] Sébastien Racanière,[3] Danilo J. Rezende,[3] Fernando Romero-López,[1,2] Phiala E. Shanahan,[1,2] and Julian M. Urban[1,2]

[1] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[3] *Google DeepMind, London, UK*
[4] *Fermi National Accelerator Laboratory, Batavia, IL 60510, U.S.A.*
[5] *Albert Einstein Center, Institute for Theoretical Physics, University of Bern, 3012 Bern, Switzerland*

Machine-learned normalizing flows can be used in the context of lattice quantum field theory to generate statistically correlated ensembles of lattice gauge fields at different action parameters. This work demonstrates how these correlations can be exploited for variance reduction in the computation of observables. Three different proof-of-concept applications are demonstrated using a novel residual flow architecture: continuum limits of gauge theories, the mass dependence of QCD observables, and hadronic matrix elements based on the Feynman-Hellmann approach. In all three cases, it is shown that statistical uncertainties are significantly reduced when machine-learned flows are incorporated as compared with the same calculations performed with uncorrelated ensembles or direct reweighting.

[2404.11674]

Practical applications of machine-learned flows on gauge fields

Ryan Abbott,[b,c] Michael S. Albergo,[d] Denis Boyda,[b,c] Daniel C. Hackett,[a,b,c] Gurtej Kanwar,[e] Fernando Romero-López,[b,c] Phiala E. Shanahan[b,c] and Julian M. Urban[b,c]

[a] *Fermi National Accelerator Laboratory, Batavia, IL 60510, U.S.A.*
[b] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[c] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[d] *Center for Cosmology and Particle Physics, New York University, New York, NY 10003, USA*
[e] *Albert Einstein Center, Institute for Theoretical Physics, University of Bern, 3012 Bern, Switzerland*

E-mail: dhackett@fnal.gov

Normalizing flows are machine-learned maps between different lattice theories which can be used as components in exact sampling and inference schemes. Ongoing work yields increasingly expressive flows on gauge fields, but it remains an open question how flows can improve lattice QCD at state-of-the-art scales. We discuss and demonstrate two applications of flows in replica exchange (parallel tempering) sampling, aimed at improving topological mixing, which are viable with iterative improvements upon presently available flows.

Technical details:

Target theory: $\mathrm{SU}(3)$ gluodynamics

Sample w/ heatbath (HB) + overrelaxation (OR)

"Residual flows" [2305.02402]

$\sim$ ODE flow + variable partitioning

$\rightarrow$ tractable/inexpensive exact Jacobian

Features:

Gauge equivariant flows

Translationally invariant $\leftrightarrow$ volume transferrable

Part 1: Accelerating sampling [2404.11674]

Generalize traditional algos:

- AIS → SNFs

  AIS = Annealed Importance Sampling SNF = Stochastic Normalizing Flow

- SMC → CRAFT

  SMC = Sequential Monte Carlo / Particle Filter CRAFT = Continuous Repeated Annealed Flow Transport

- REX → T-REX (L-REX)

  REX = Replica EXchange / Parallel Tempering T-REX = Transformed REX

Part 1: Accelerating sampling [2404.11674]

Generalize traditional algos:

- AIS → SNFs

  AIS = Annealed Importance Sampling

  SNF = Stochastic Normalizing Flow

- SMC → CRAFT

  SMC = Sequential Monte Carlo / Particle Filter

  CRAFT = Continuous Repeated Annealed Flow Transport

- **REX → T-REX (L-REX)**

  REX = Replica EXchange / Parallel Tempering

  T-REX = Transformed REX

How to "go shorter":

1. Bridge to nearby theory w/ faster sampling
2. Add small, unphysical "defect" designed to improve sampling

Sketch: Markov chain Monte Carlo sampling

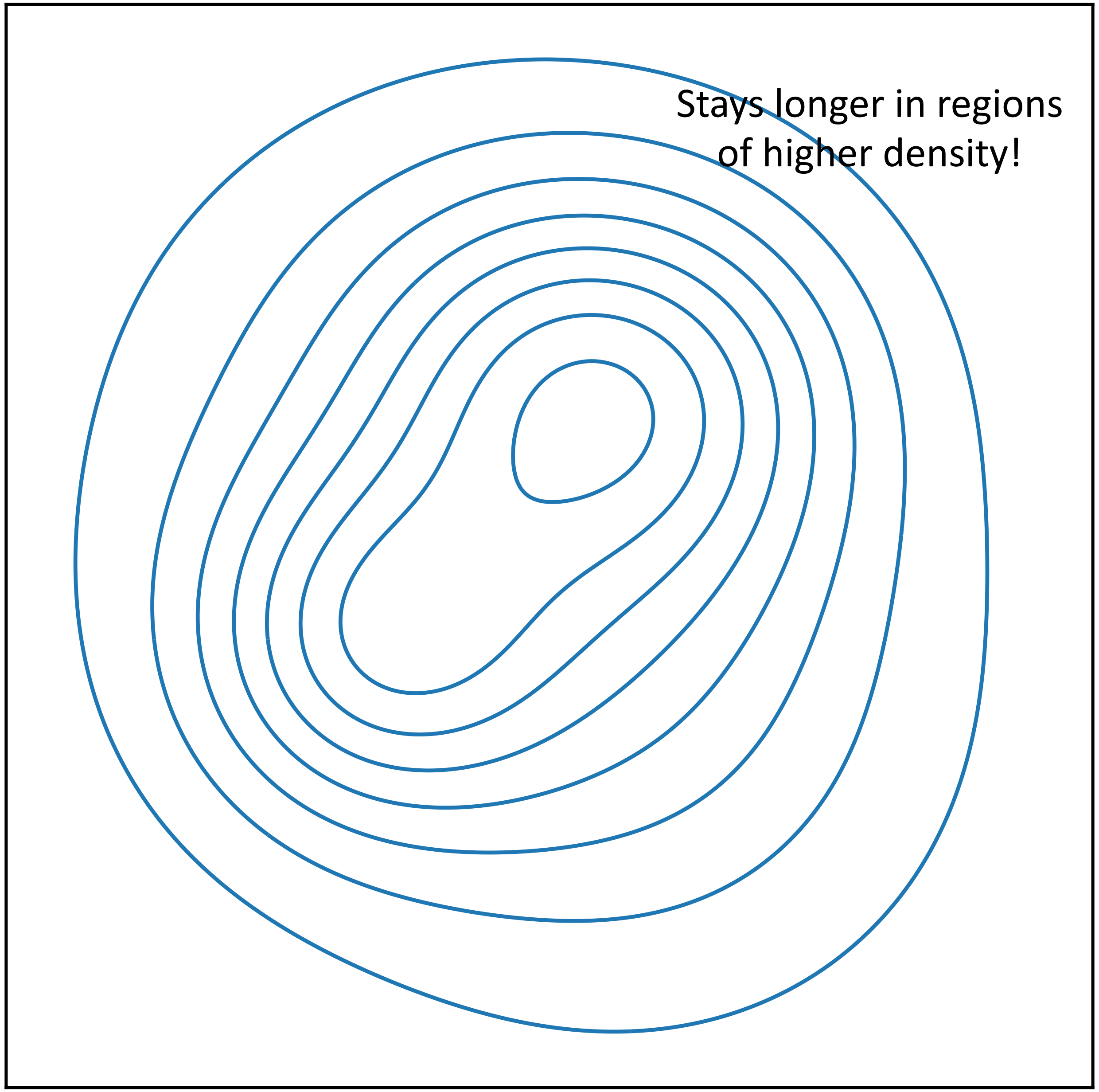
Sample by evolving the state of a Markov chain

Time average ⇔ Ensemble average

(If ergodic!)

Standard for QCD: HMC

Hybrid/Hamiltonian Monte Carlo



Sketch: Markov chain Monte Carlo sampling

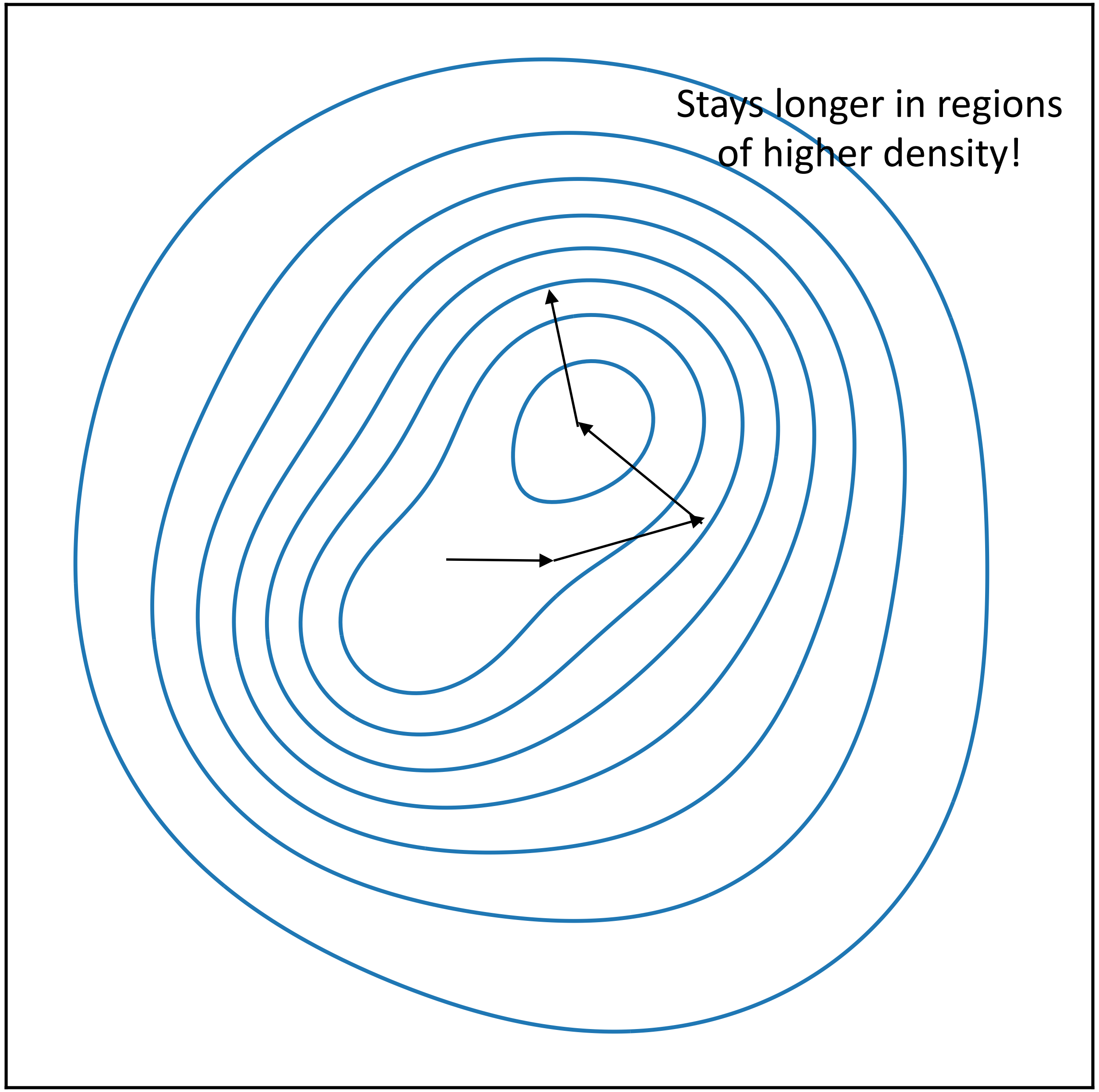
Sample by evolving the state of a Markov chain

Time average ⇔ Ensemble average

(If ergodic!)

Standard for QCD: HMC

Hybrid/Hamiltonian Monte Carlo



Sketch: Markov chain Monte Carlo sampling

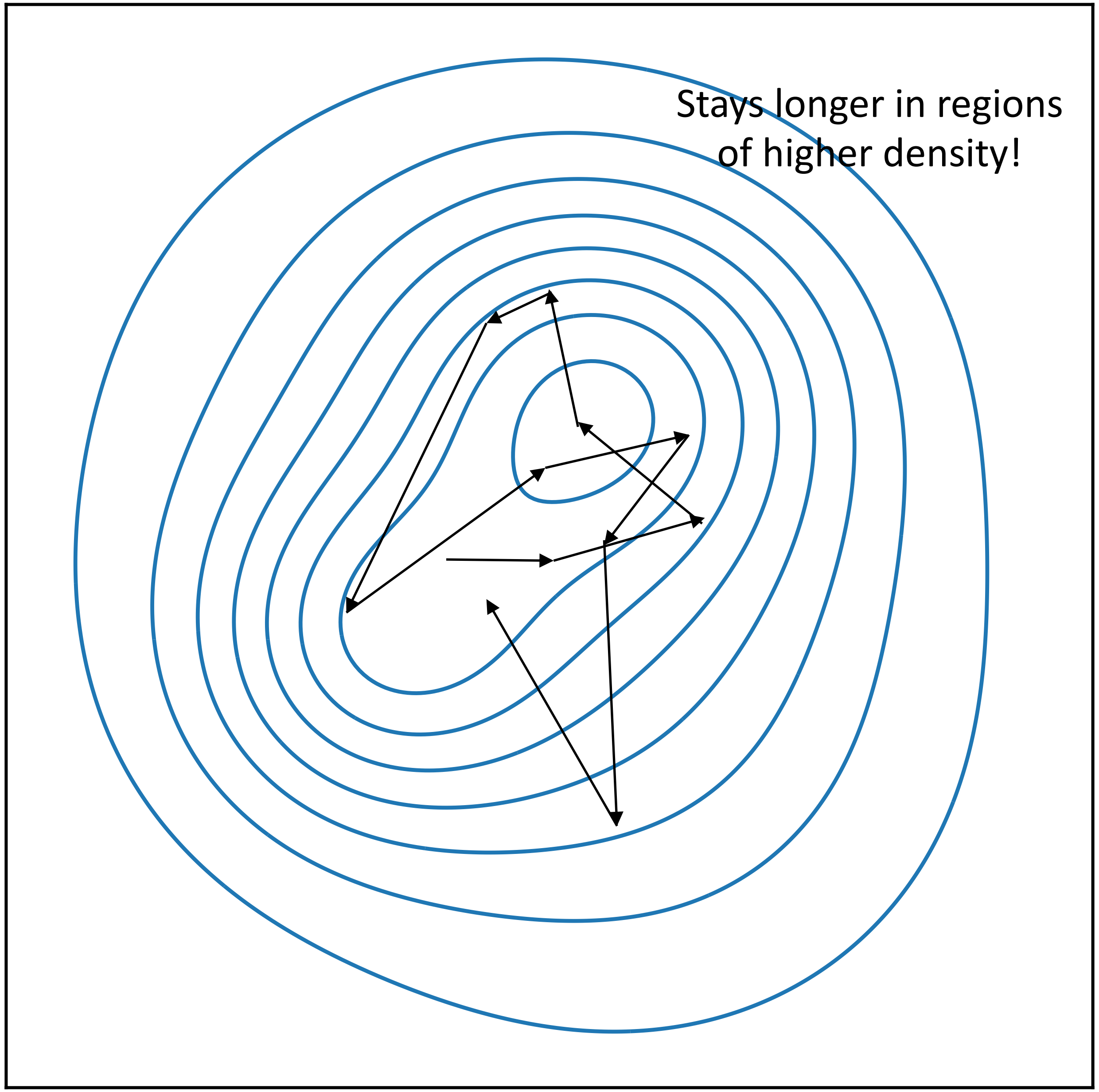
Sample by evolving the state of a Markov chain

Time average ⇔ Ensemble average

(If ergodic!)

Standard for QCD: HMC

Hybrid/Hamiltonian Monte Carlo



Sketch: Markov chain Monte Carlo sampling

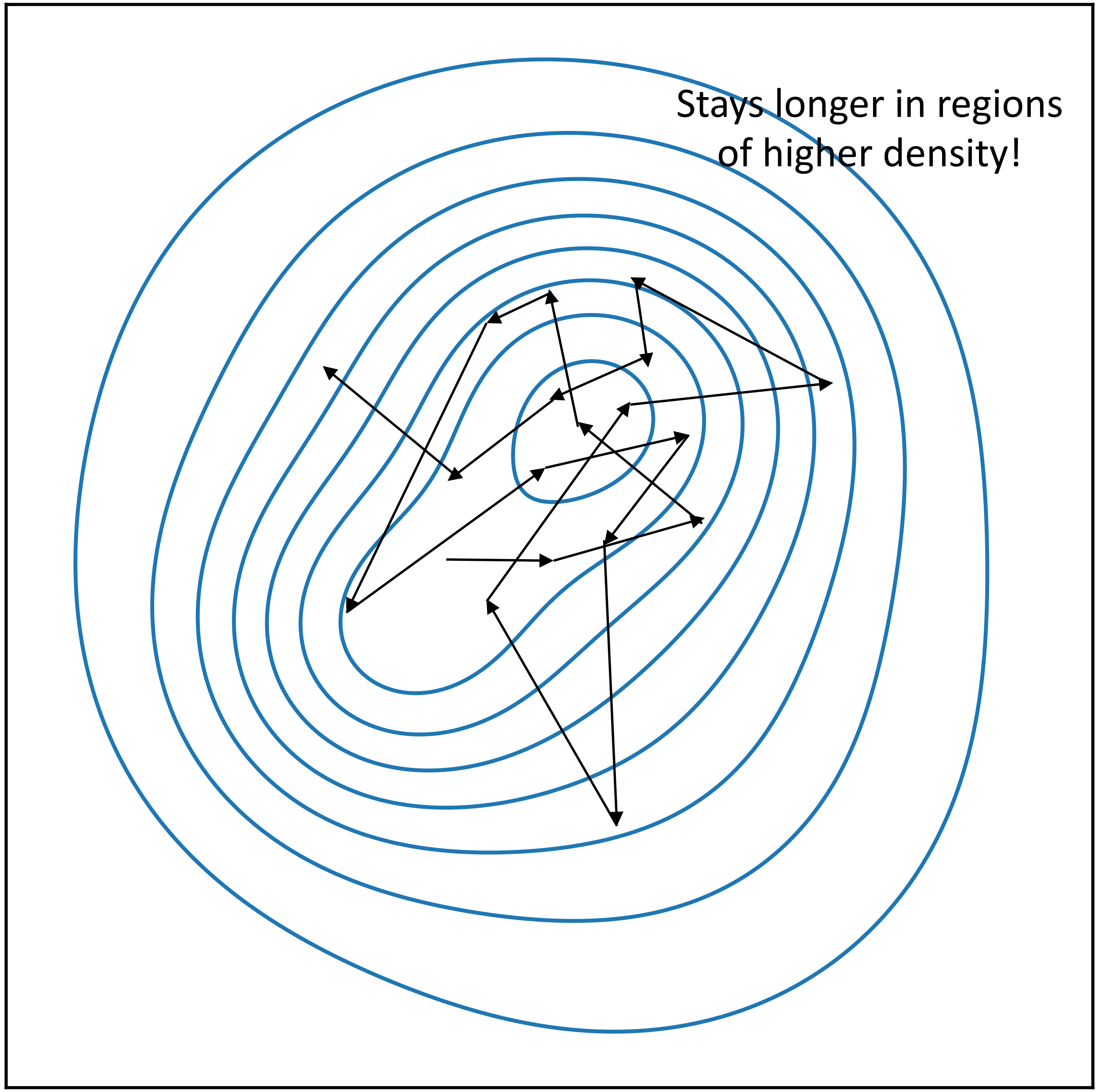
Sample by evolving the state of a Markov chain

Time average ⇔ Ensemble average

(If ergodic!)

Standard for QCD: HMC

Hybrid/Hamiltonian Monte Carlo



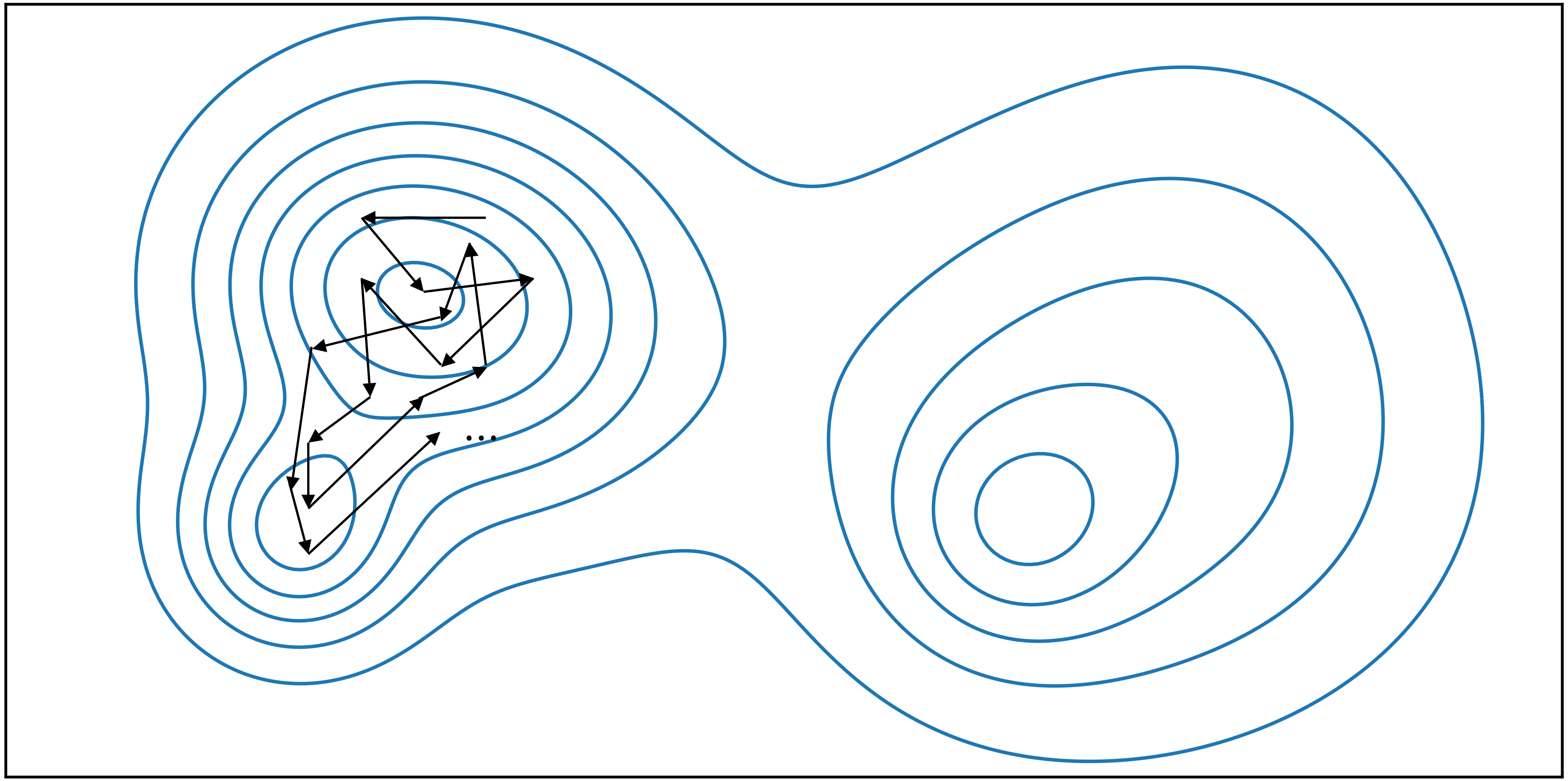
Topological freezing

Topological charge

$$Q \sim \int_x F\tilde{F}$$

Problem: gauge field distribution is multimodal

- Must sample different **topological sectors**
- Exponentially slow tunneling as $a \to 0$



Topological freezing

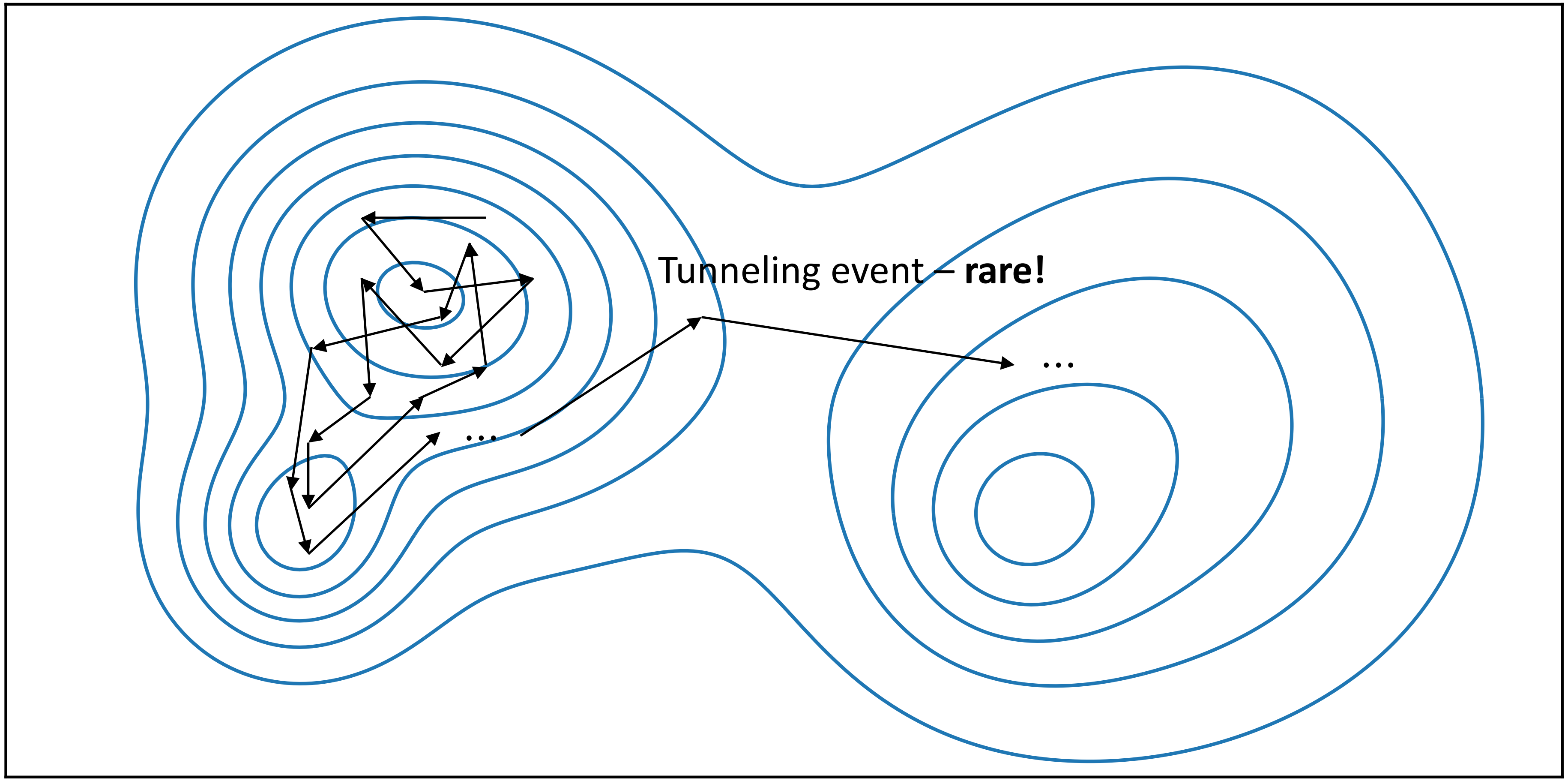
Topological charge

$$Q \sim \int_x F\tilde{F}$$

Problem: gauge field distribution is multimodal

Must sample different **topological sectors**

Exponentially slow tunneling as $a \to 0$



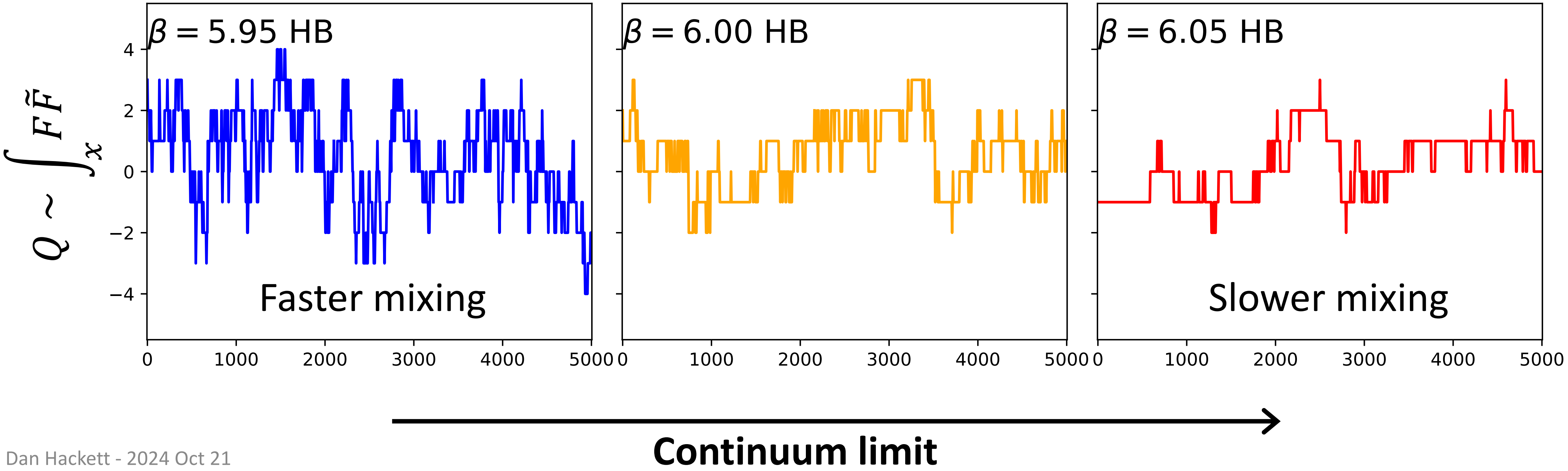
Topological freezing

Problem: gauge field distribution is multimodal

- Must sample different **topological sectors**
- Exponentially slow tunneling as $a \to 0$

Can result in *effective* loss of ergodicity [2202.1172]

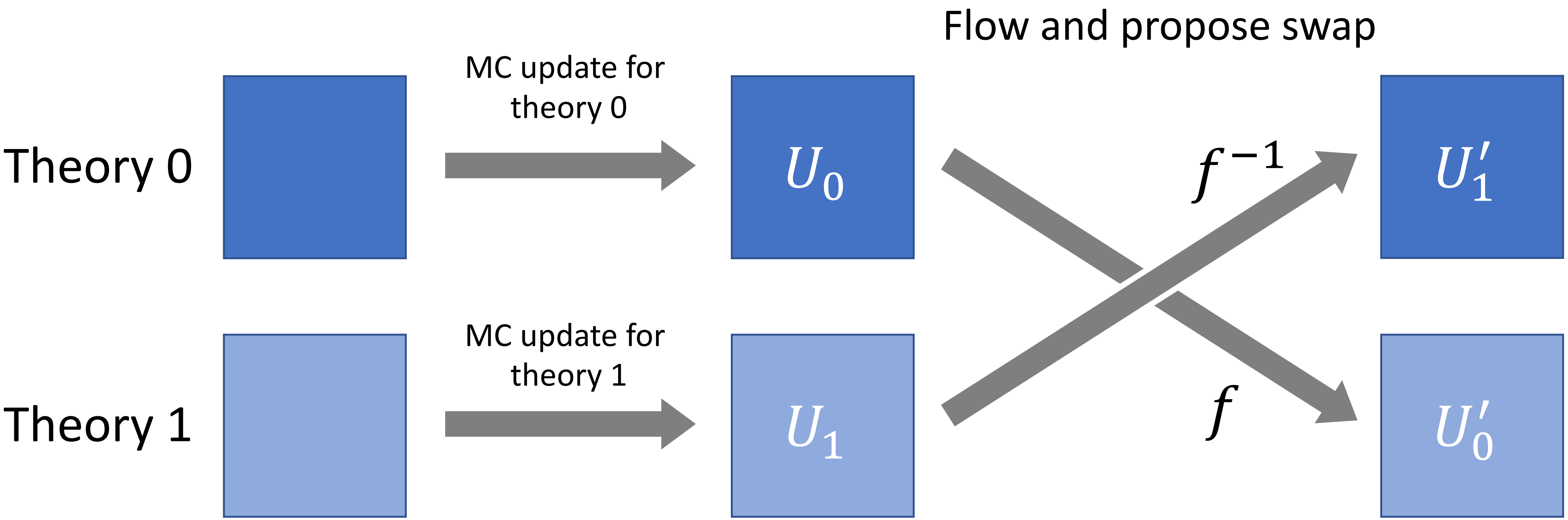
- → Apparent convergence to wrong answers at achievable sample sizes



Application 1: Transformed Replica EXchange (T-REX)

(REX a.k.a. parallel tempering) [Invernizzi Krämer Clemente Noé 2210.14104]

Simultaneously sample chains for different targets

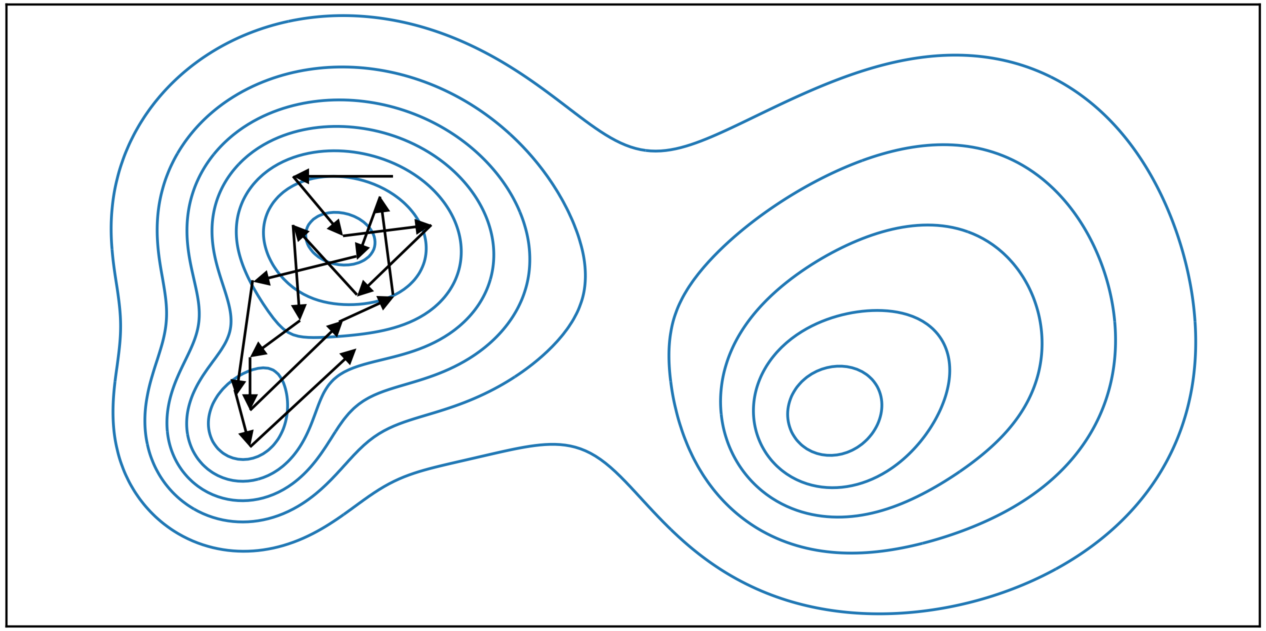


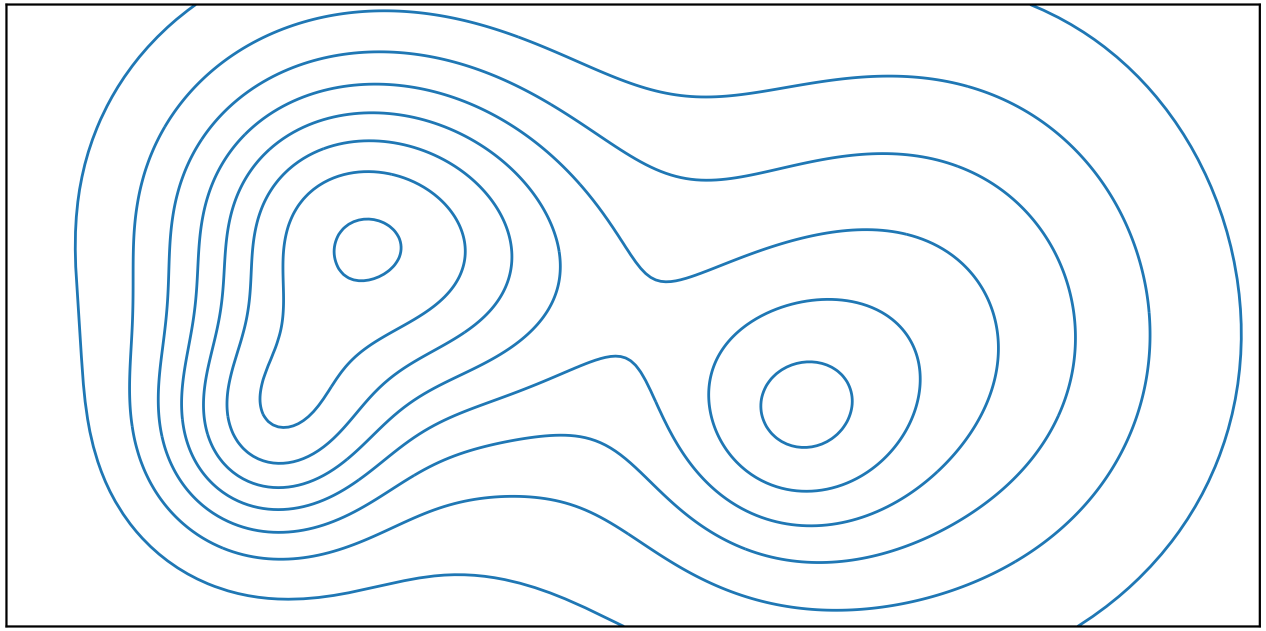
$$p_{\rm acc} = \min\left[1, \frac{p_0(U_1')\, p_1(U_0')}{p_0(U_0)\, p_1(U_1)}\, J_f(U_0)\, J_{f^{-1}}(U_1)\right]$$

How REX can accelerate topological mixing

Swapping allows system to temporarily evolve in “easier” theory where modes are less separated

∼ Temporary bridge between modes in ”harder” theory

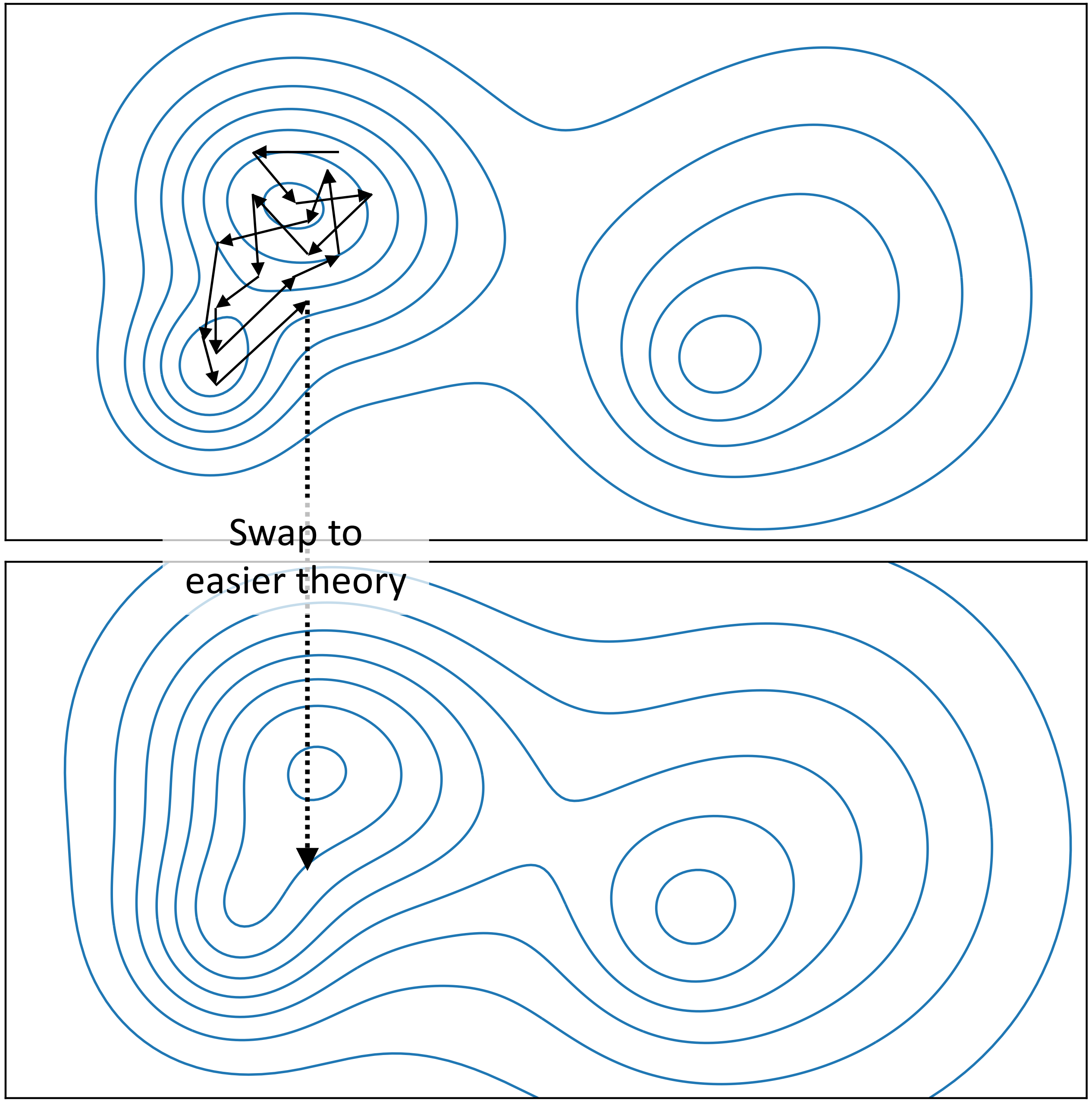




How REX can accelerate topological mixing

Swapping allows system to temporarily evolve in “easier” theory where modes are less separated

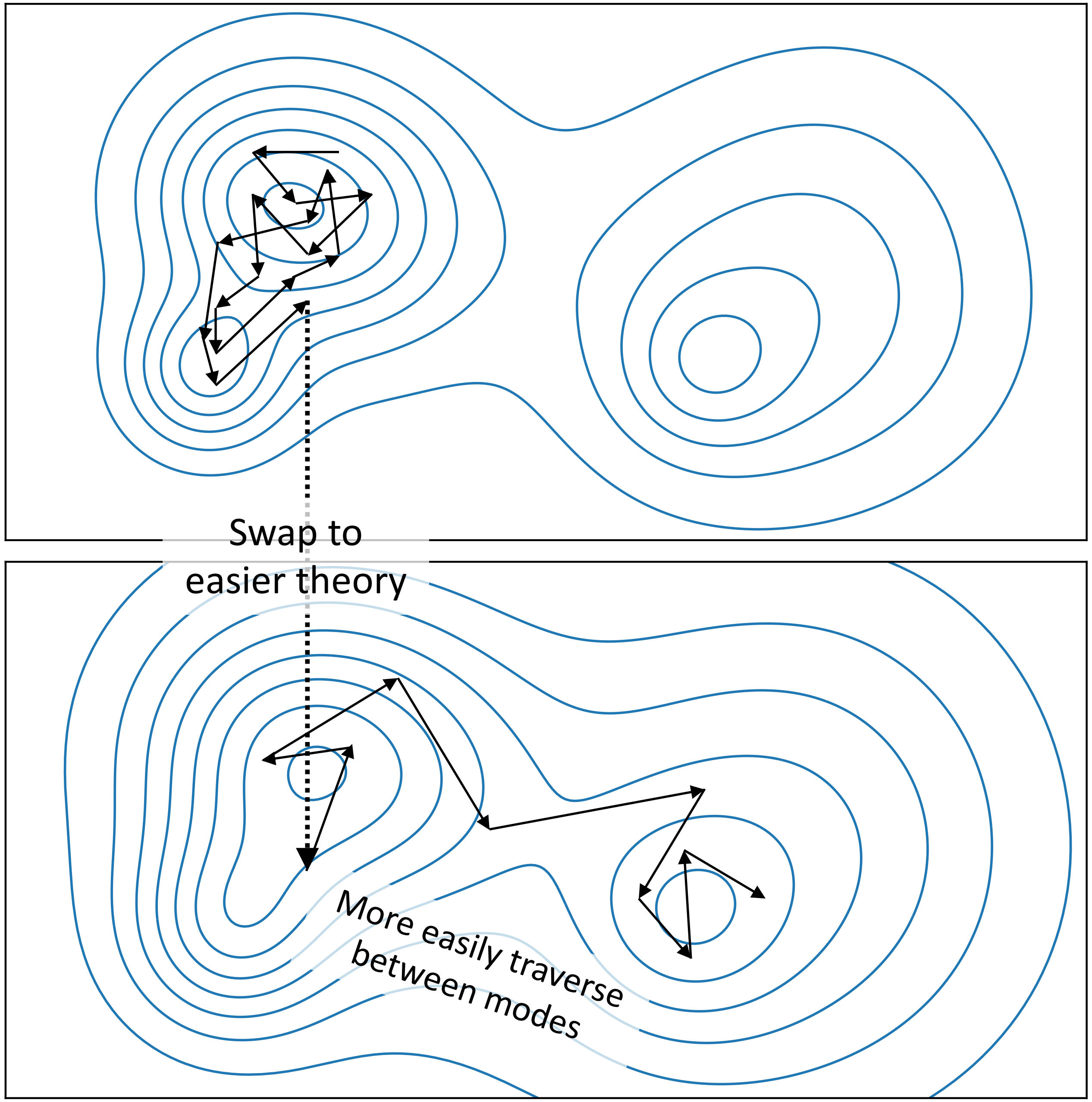
~ Temporary bridge between modes in ”harder” theory



How REX can accelerate topological mixing

Swapping allows system to temporarily evolve in “easier” theory where modes are less separated

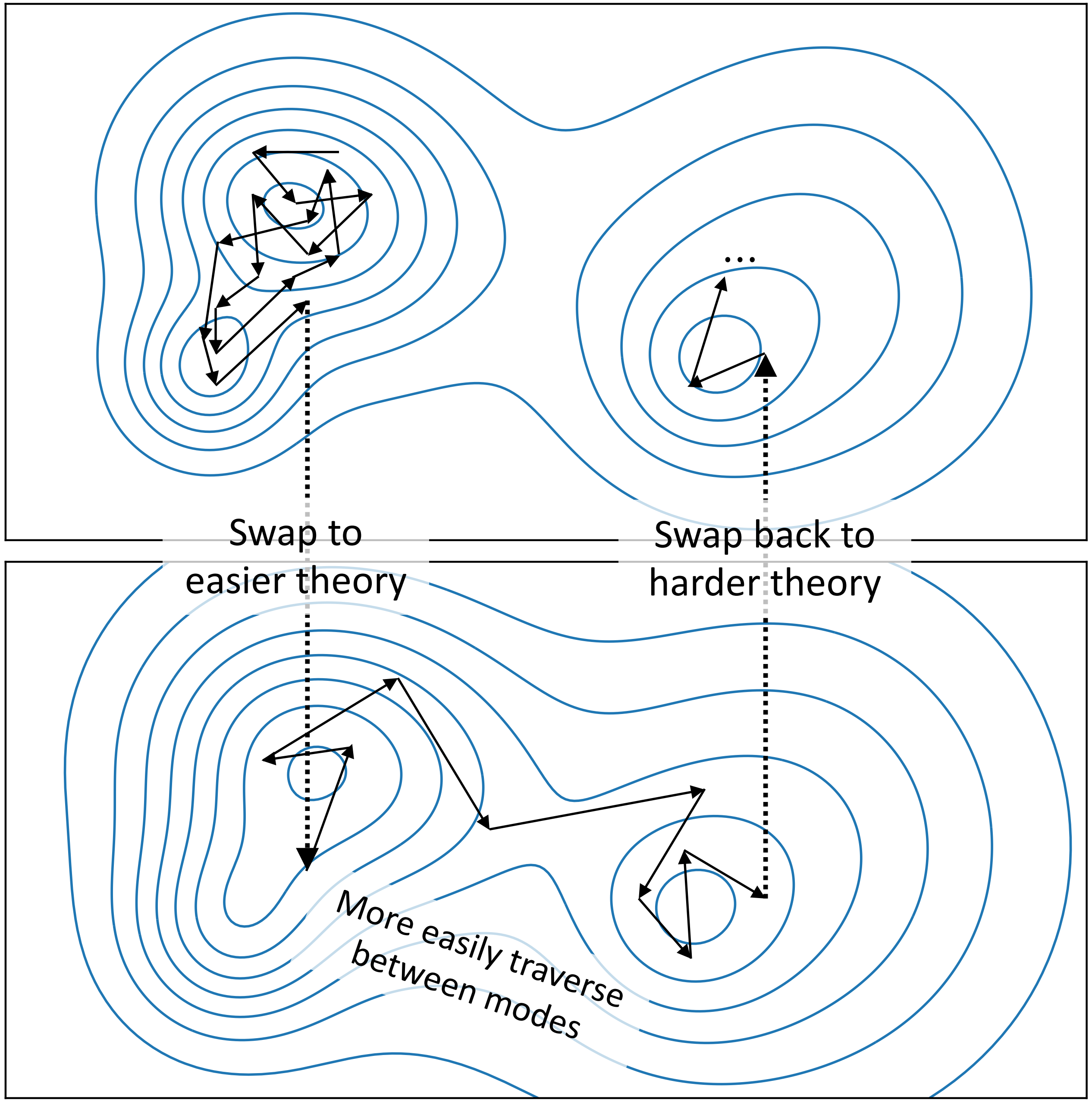
∼ Temporary bridge between modes in ”harder” theory



How REX can accelerate topological mixing

Swapping allows system to temporarily evolve in "easier" theory where modes are less separated

~ Temporary bridge between modes in "harder" theory



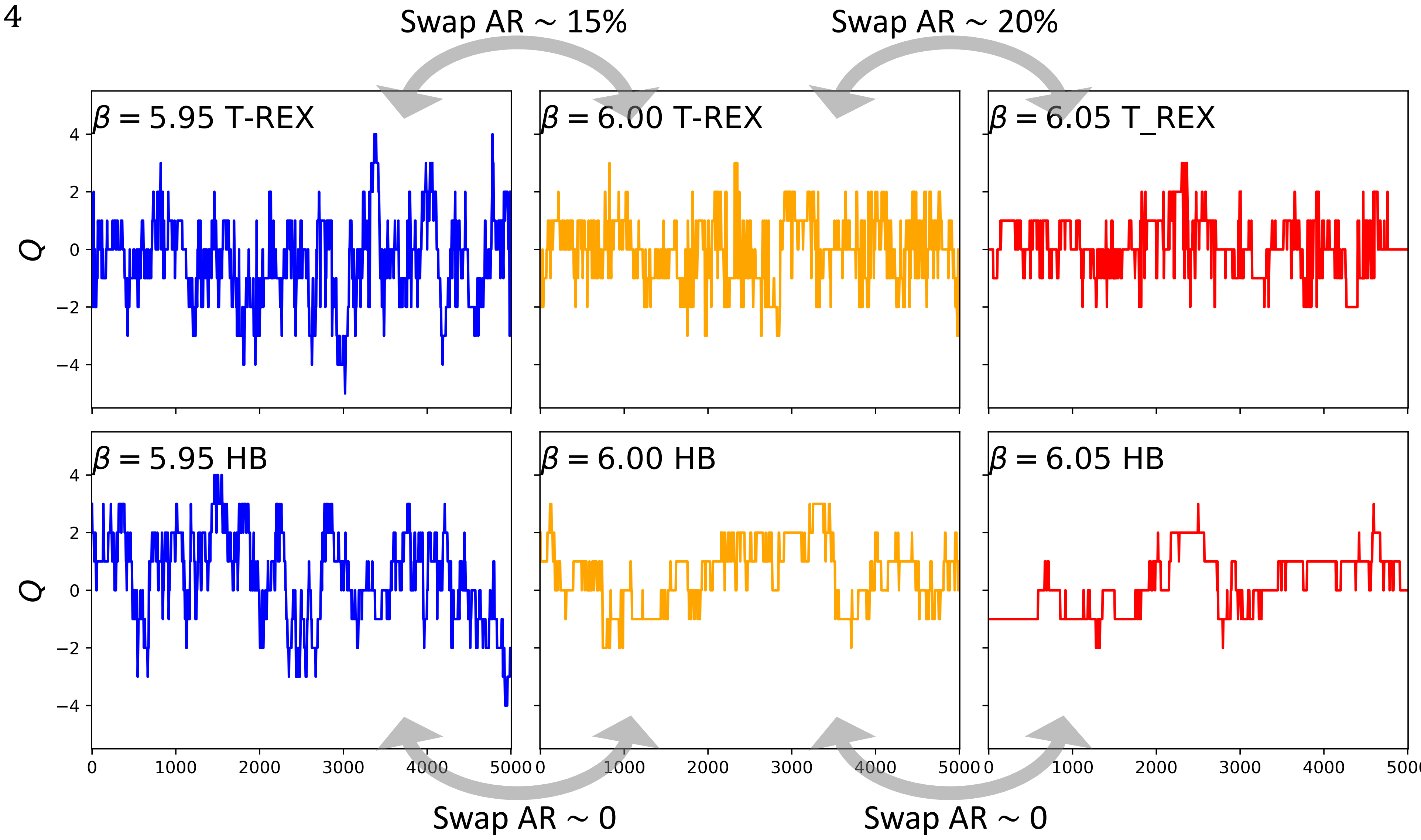
Demo: T-REX

Three target βs on 12^4

Two different flows

$$5.95 \leftrightarrow 6$$

$$6 \leftrightarrow 6.05$$



1 step = 5 HB + 2 OR, propose swaps every 5 steps

Demo: T-REX

Three target βs on 12^4

Two different flows

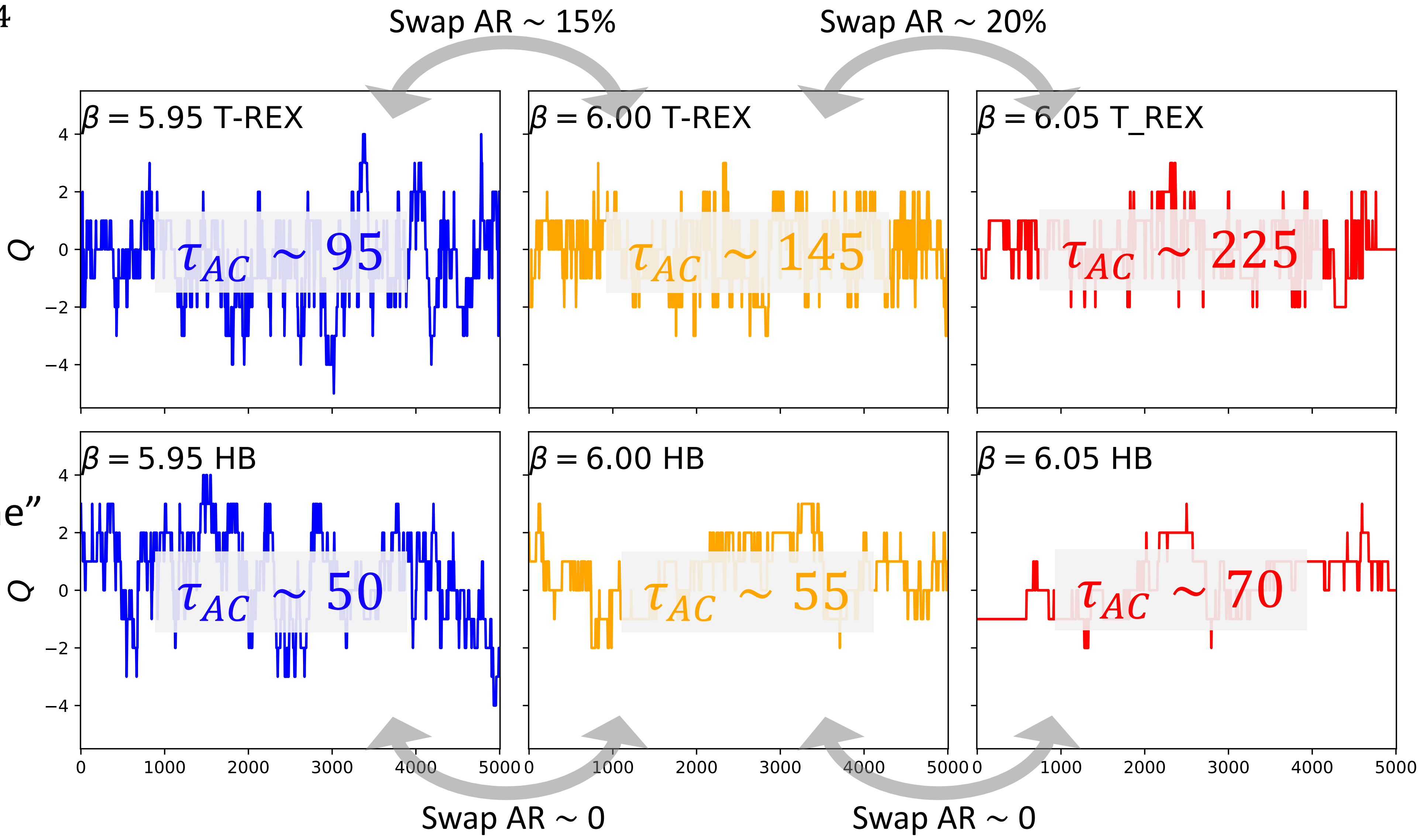
$5.95 \leftrightarrow 6$

$6 \leftrightarrow 6.05$

τ_{AC}: "autocorrelation time"

Sampling cost $\propto \tau_{AC}$

$$\text{ESS} \sim \frac{1}{1+2\,\tau_{AC}}$$



1 step = 5 HB + 2 OR, propose swaps every 5 steps

Demo: T-REX

Three target βs on 12^4

Two different flows

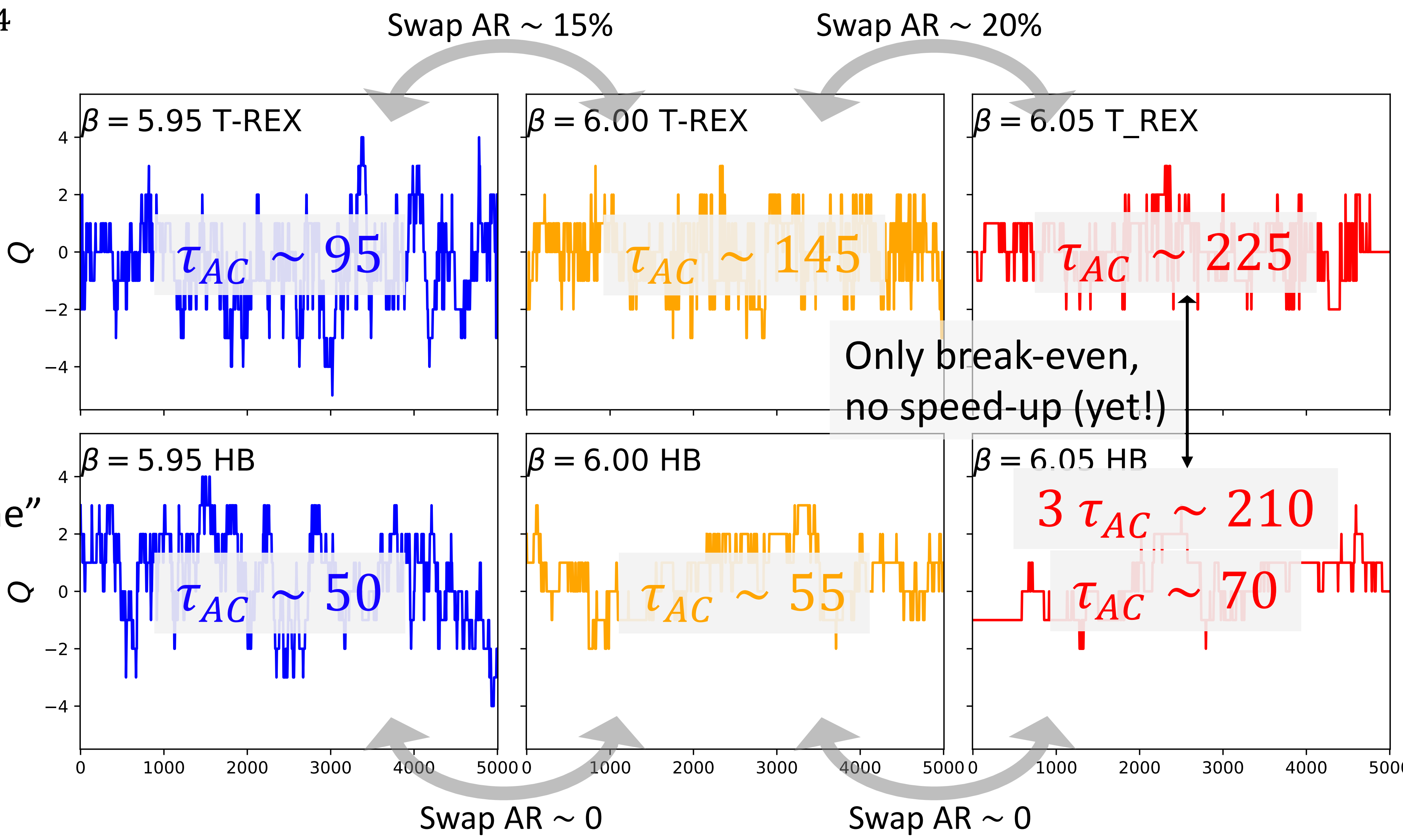
$5.95 \leftrightarrow 6$

$6 \leftrightarrow 6.05$

τ_{AC}: "autocorrelation time"

Sampling cost $\propto \tau_{AC}$

$$\text{ESS} \sim \frac{1}{1+2\,\tau_{AC}}$$



1 step = 5 HB + 2 OR, propose swaps every 5 steps

Parallel Tempering on Boundary Conditions (PTBC)

[Hasenbusch 1706.04443]

[Bonnano Bonati D'Elia 2012.14000]

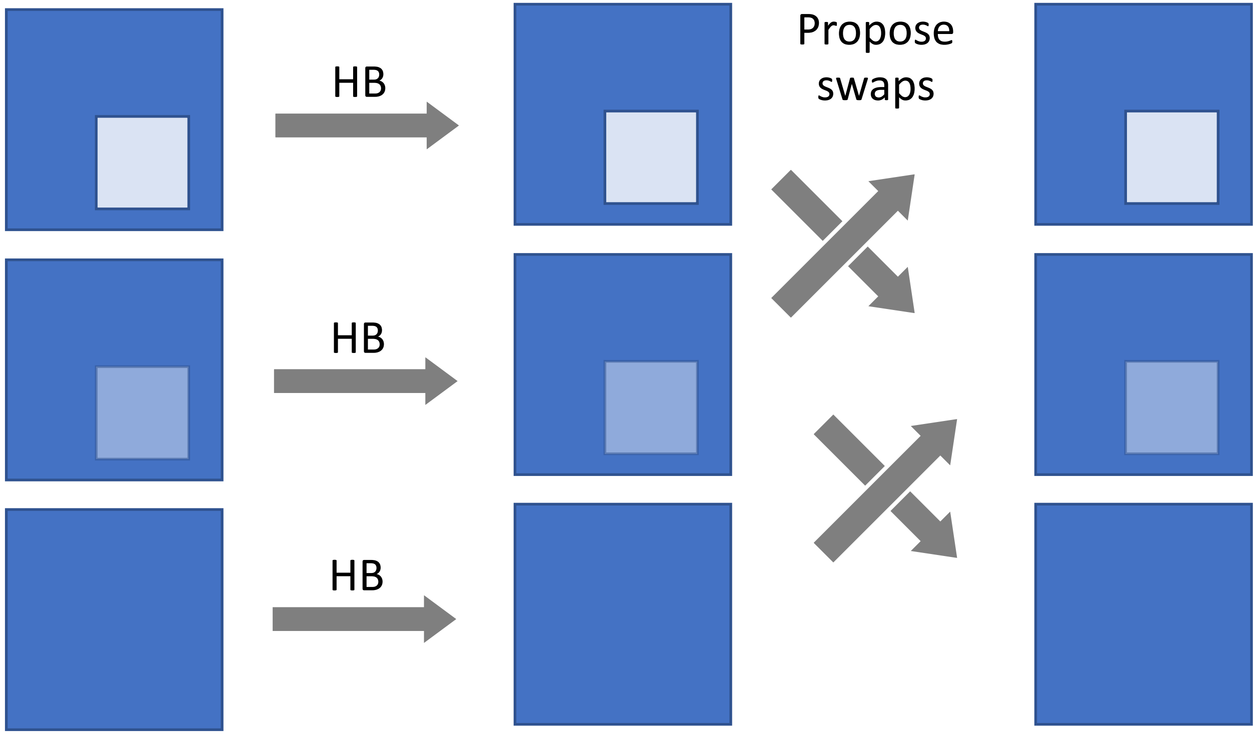
Open Boundary Conditions (OBC) are known to accelerate topology, but harder to do physics with

Idea: introduce localized defect

$\beta_{\mathbf{defect}} < \beta_{\mathbf{target}}$

"Poke a hole in the boundary"

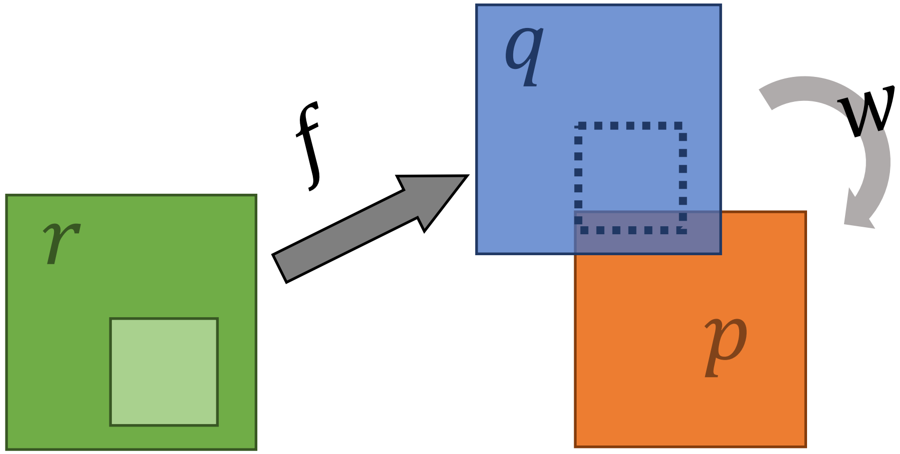
Remove defect w/ REX



Defect Repair Replica EXchange (DR-REX)

Train flow to repair defect

(Or, multiple flows for several steps of partial repair)



Defect has localized physical effects

→ Flow can act only on a subvolume

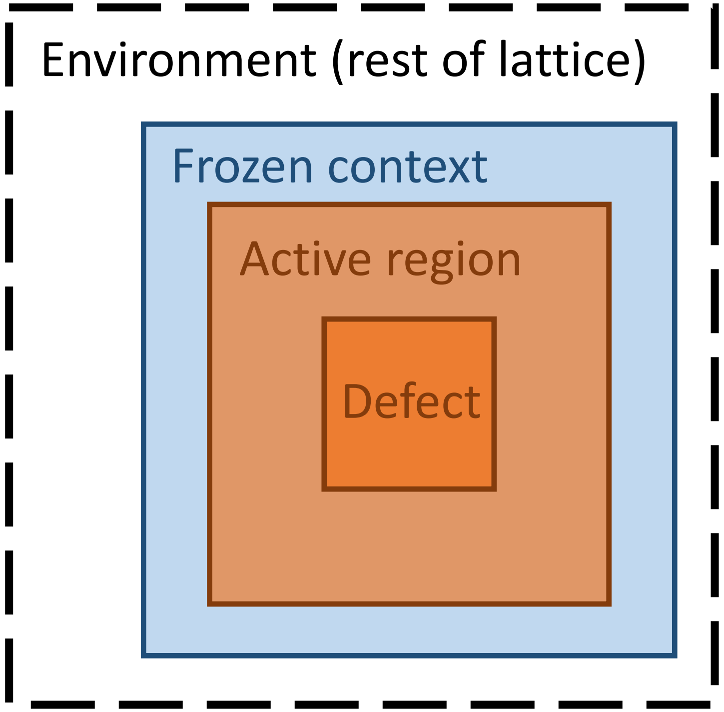
Pros:

Volume-independent computational cost

Volume-independent model quality

Con(?):

Requires conditional flow



Demo: DR-REX

Target: $\beta = 6.3$ on 16^4

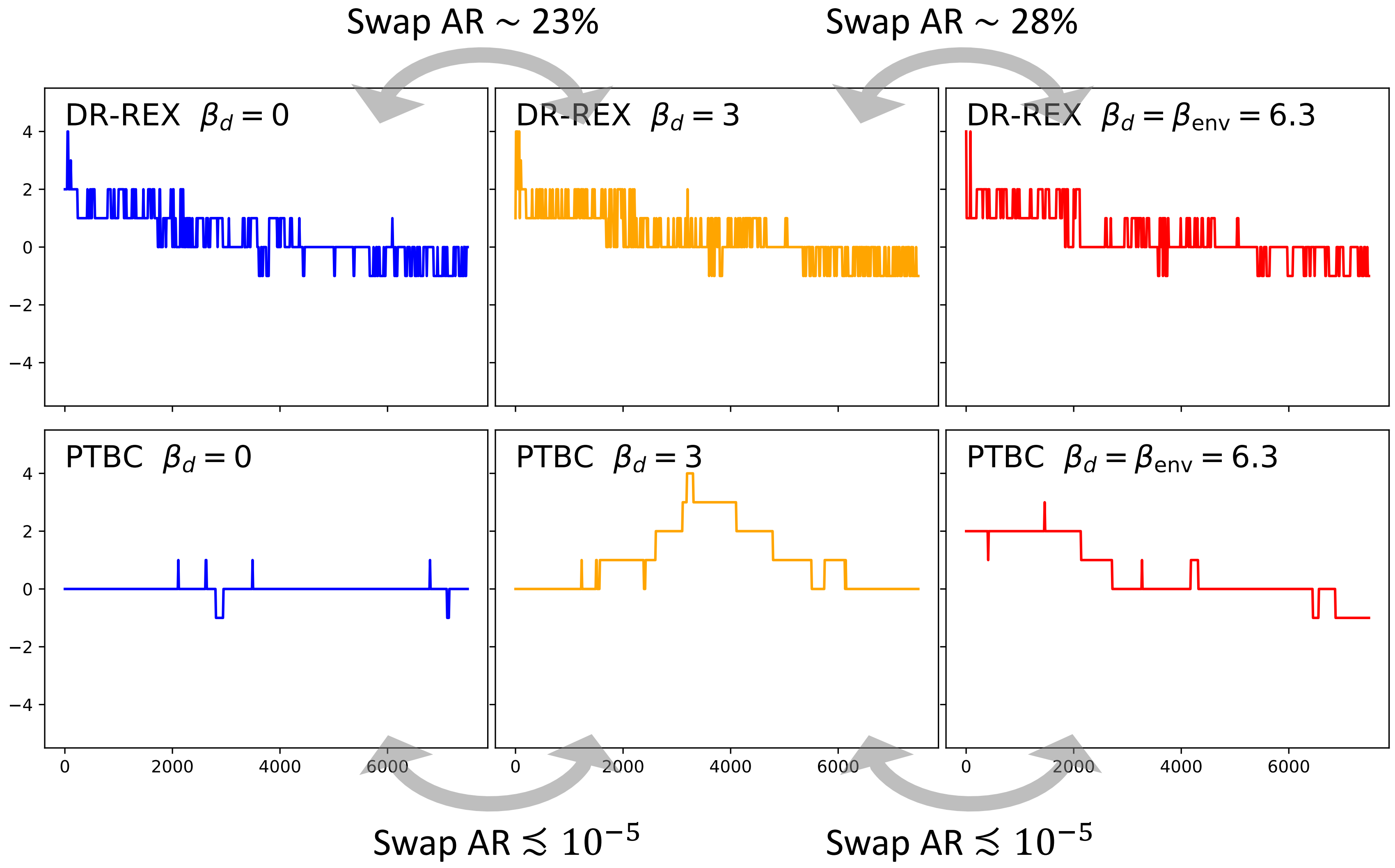
2^3 OBC defect

Two flows to repair

$$\beta_d = 0 \to 3 \to 6.3$$

Flows act on 8^4 subvolume

Similar swap AR w/o flows requires 7-8 chains



1 step = 1 HB + 5 OR, propose swaps every 10 steps

Demo: DR-REX

Target: $\beta = 6.3$ on 16^4

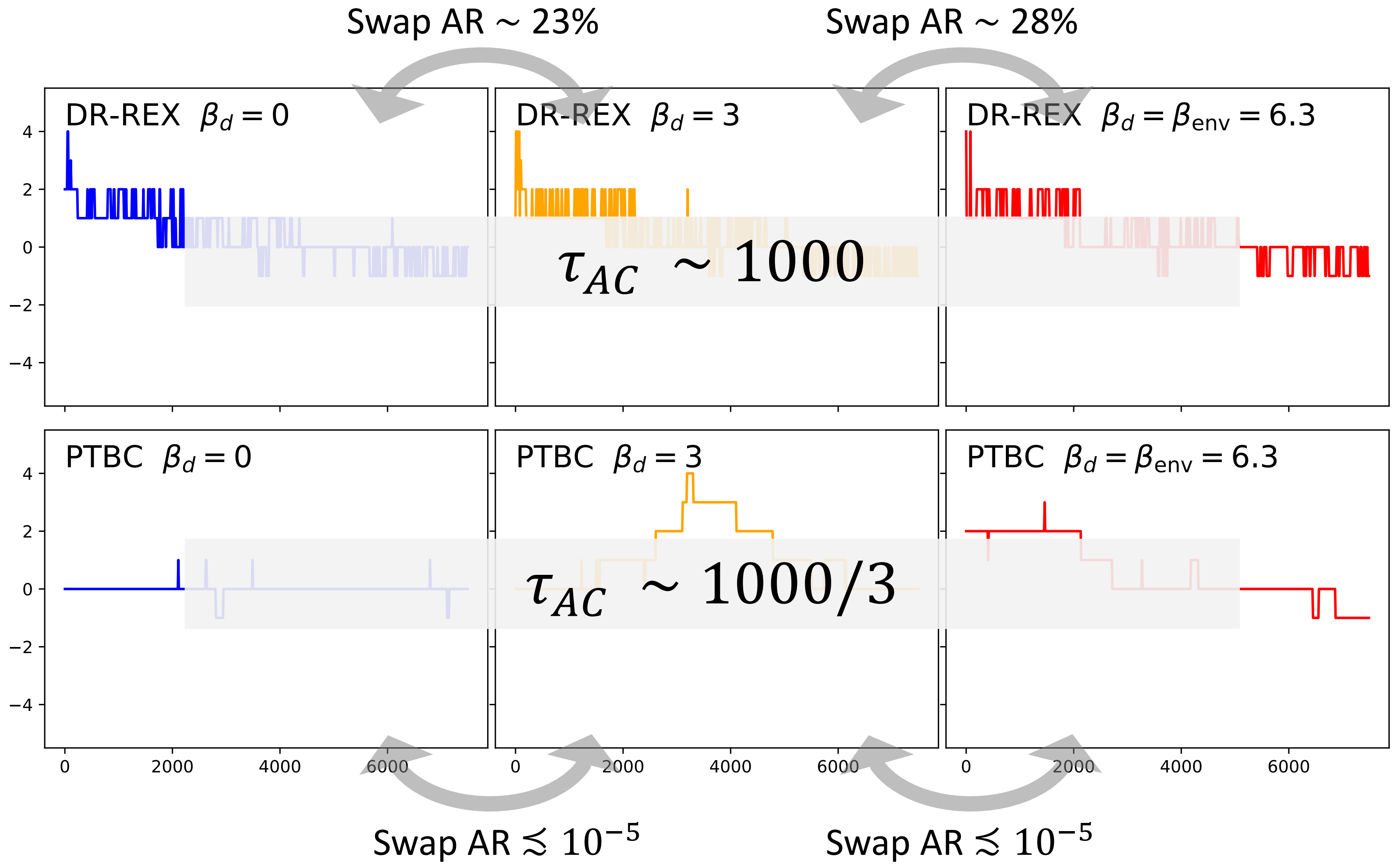
2^3 OBC defect

Two flows to repair

$$\beta_d = 0 \to 3 \to 6.3$$

Flows act on 8^4 subvolume

Similar swap AR w/o flows requires 7-8 chains



1 step = 1 HB + 5 OR, propose swaps every 10 steps

Demo: DR-REX

Target: $\beta = 6.3$ on 16^4

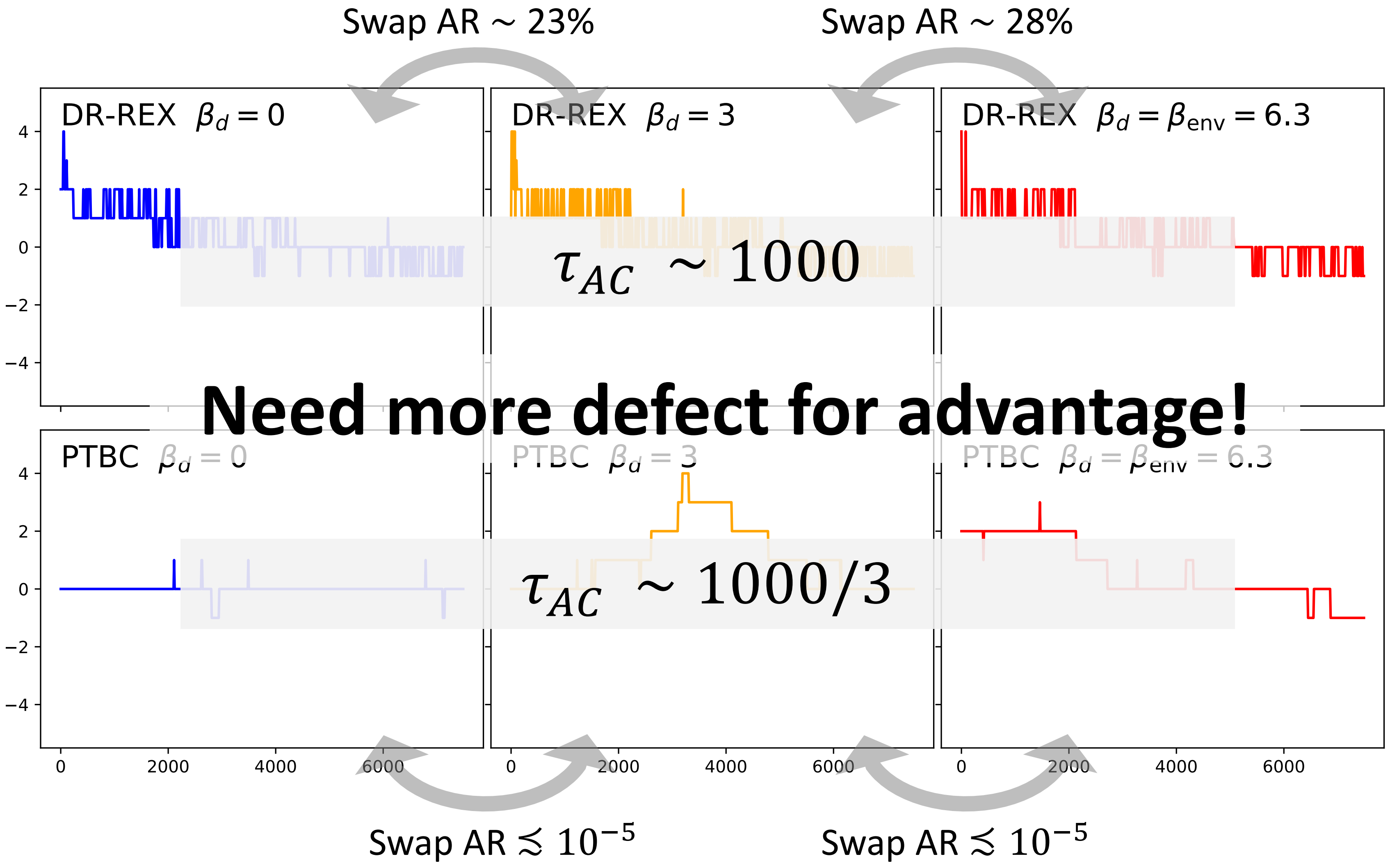
2^3 OBC defect

Two flows to repair

$$\beta_d = 0 \to 3 \to 6.3$$

Flows act on 8^4 subvolume

Similar swap AR w/o flows requires 7-8 chains



1 step = 1 HB + 5 OR, propose swaps every 10 steps

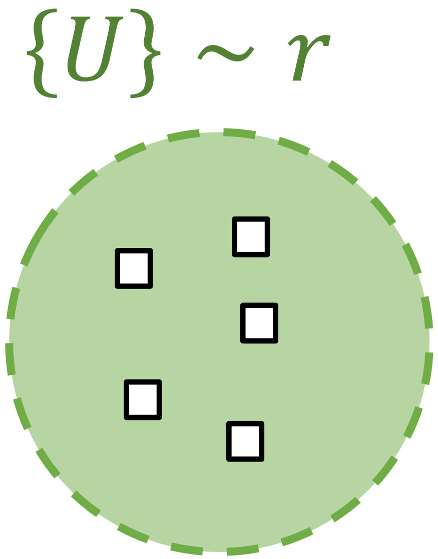
Part 2: Correlated ensembles [2401.10874]

See also [Bacchio 2305.07932]

Flow an ensemble

→ $\{U\}$ and $\{f(U)\}$ are correlated

This is useful!



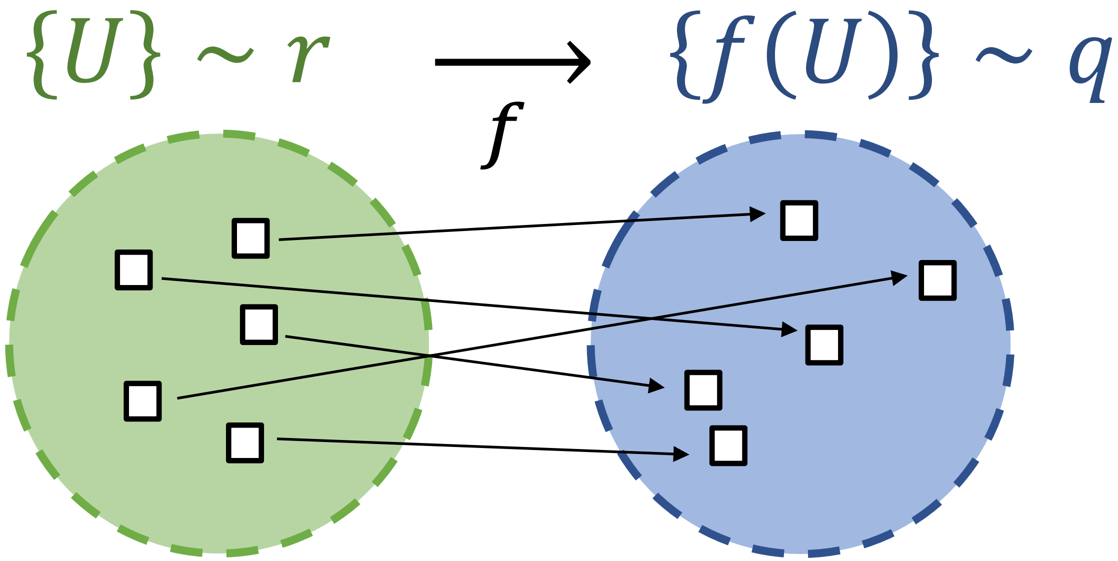
Part 2: Correlated ensembles [2401.10874]

See also [Bacchio 2305.07932]

Flow an ensemble

→ $\{U\}$ and $\{f(U)\}$ are correlated

This is useful!



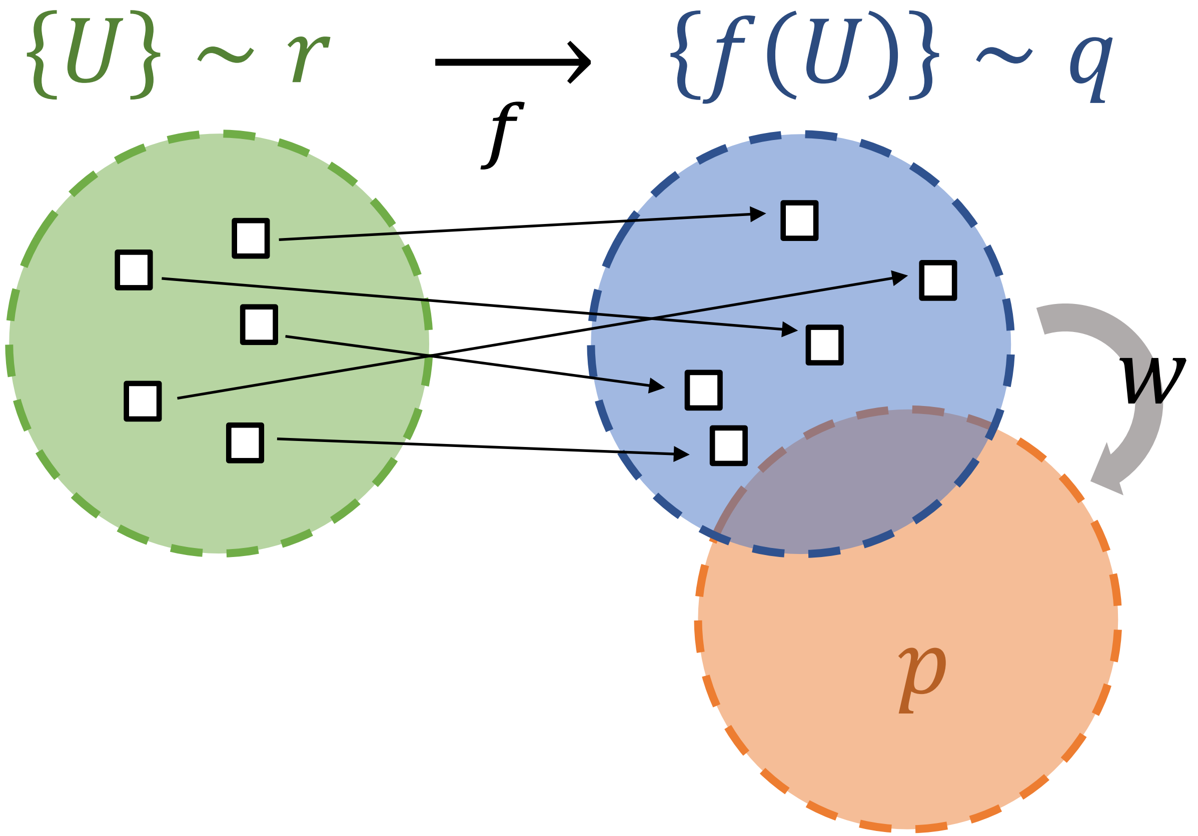
Part 2: Correlated ensembles [2401.10874]

See also [Bacchio 2305.07932]

Flow an ensemble

→ $\{U\}$ and $\{f(U)\}$ are correlated

This is useful!

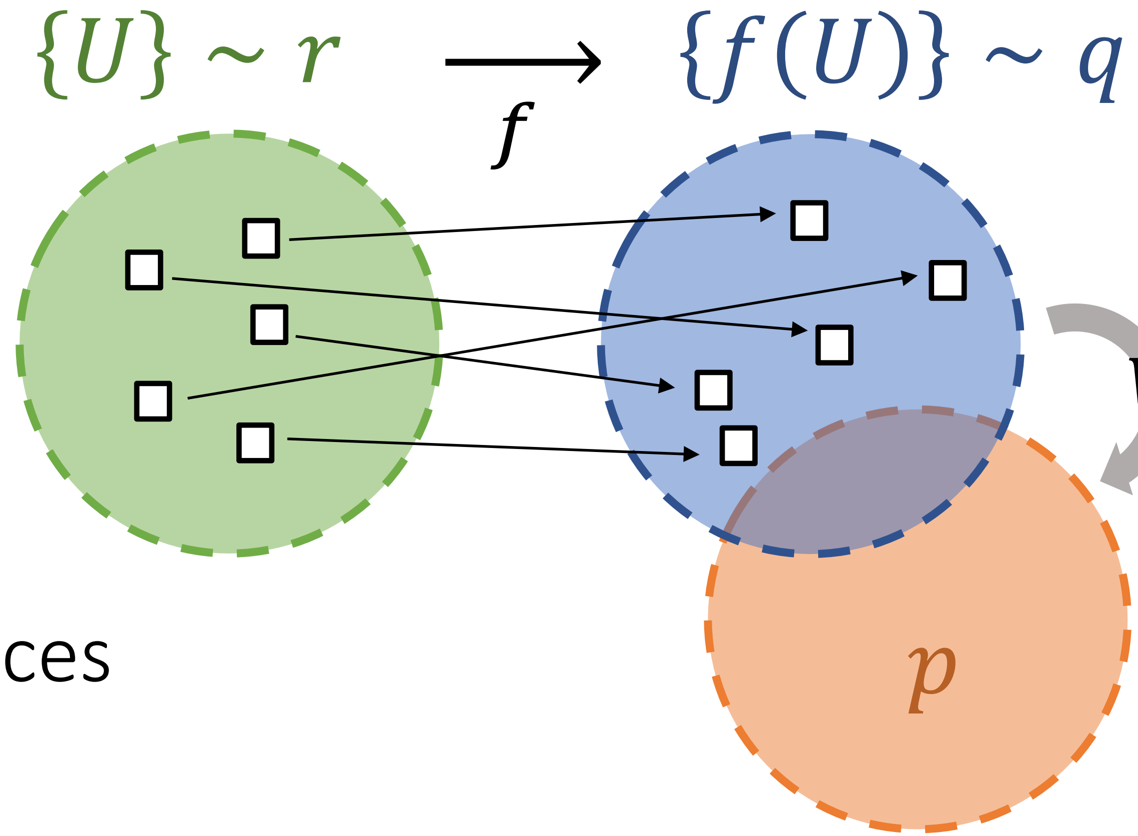


Part 2: Correlated ensembles [2401.10874]

Flow an ensemble

→ $\{U\}$ and $\{f(U)\}$ are correlated

This is useful!



e.g. for noise cancellation in differences

$$\langle O \rangle_p - \langle O \rangle_r$$

$$= \langle wO \rangle_q - \langle O \rangle_r$$

$$= \langle w(f(U))\, O(f(U)) - O(U) \rangle_{U \sim r}$$

Derivative observables

Improve $\langle O \rangle_p - \langle O \rangle_r$ in derivatives w/r/t action parameters:

$$\frac{d\langle O \rangle}{d\beta} \approx \frac{1}{\delta\beta}\left[\langle O \rangle_{\beta+\delta\beta} - \langle O \rangle_{\beta}\right]$$

Applications:

- Constraints on extrapolations to continuum/chiral/… limits
- Feynman-Hellmann $\quad S \to S + \lambda\, O \quad \left.\frac{\partial E_h}{\partial \lambda}\right|_{\lambda=0} \sim \langle h|O|h\rangle$

e.g.: nucleon-pion sigma term $\sigma_{\pi N} = m_q \langle N|\bar{q}q|N\rangle = m_q \frac{\partial M_N}{\partial m_q}$

e.g.: gluon momentum fraction $\langle x \rangle_g = \frac{1}{2m^2}\langle h(0)|T^{00}(0)|h(0)\rangle$ for hadron h

$\langle x \rangle_g = -\frac{2}{3m}\left.\frac{\partial m}{\partial \lambda}\right|_{\lambda=0}$ where $\delta S = -\lambda\, \frac{1}{2}[-E^2 + B^2]$

Demo: Pion $\langle x \rangle_g$ w/ flowed Feynman-Hellmann

[QCDSF-UKQCD 1205.6410]

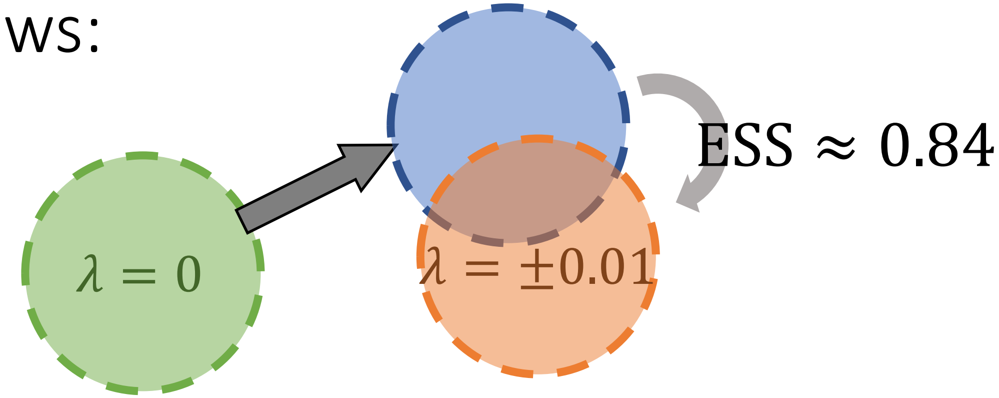
Spatial-temporal anisotropy: $\delta S = -\lambda \frac{\beta}{N_c}\left[\sum_i P_{ti} - \sum_{i<j} P_{ij}\right]$

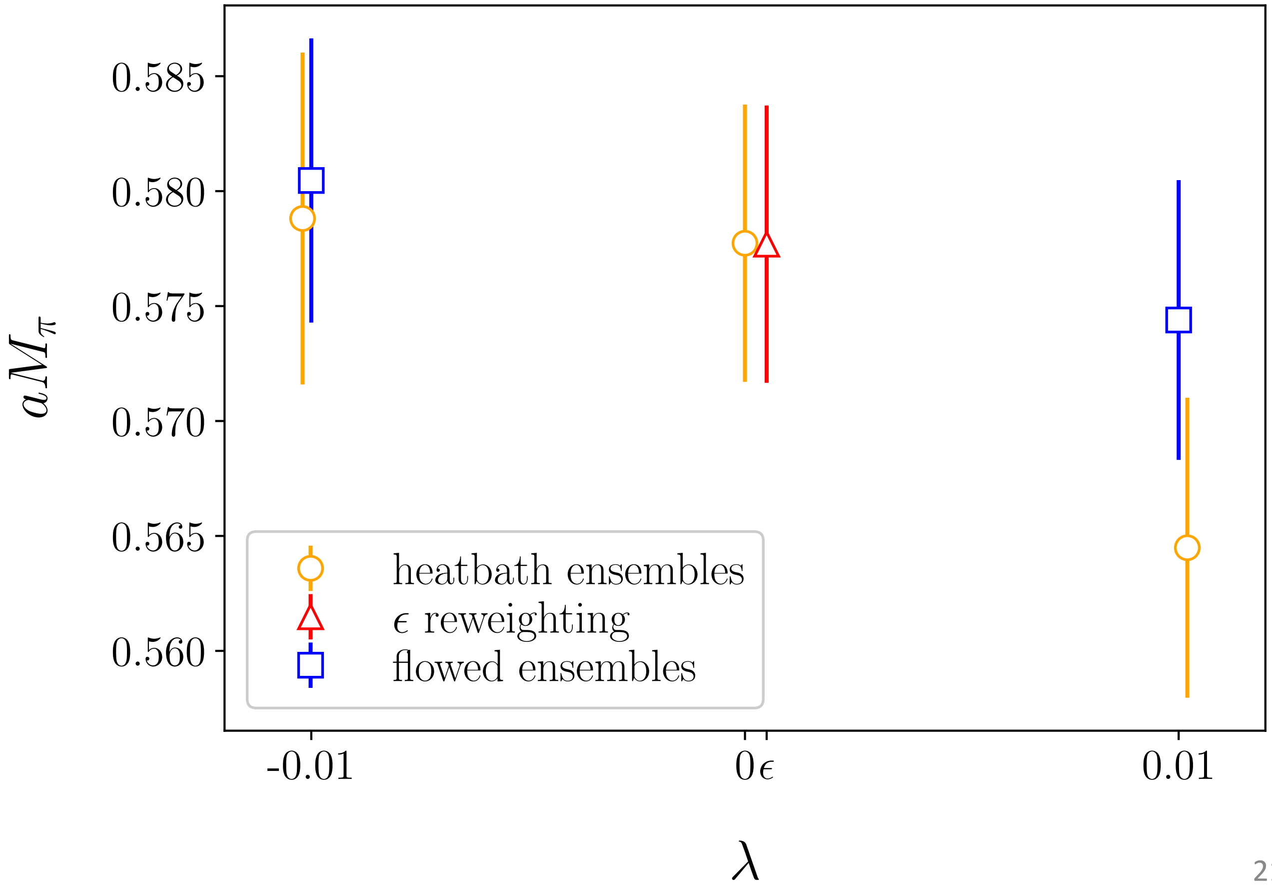
$$\langle x \rangle_g^{\text{latt}} = -\frac{2}{3m}\frac{\partial m}{\partial \lambda}\Big|_{\lambda=0} \approx -\frac{2}{3m}\frac{1}{\lambda}[m(\lambda) - m(0)]$$

Parameters:

$8^3 \times 16 \quad \beta = 6 \quad \kappa = 0.132$ (quenched)

Flows:





Compute $m(\lambda)$ from $\langle w\, C^{2\text{pt}} \rangle_q$

Demo: Pion $\langle x \rangle_g$ w/ flowed Feynman-Hellmann

[QCDSF-UKQCD 1205.6410]

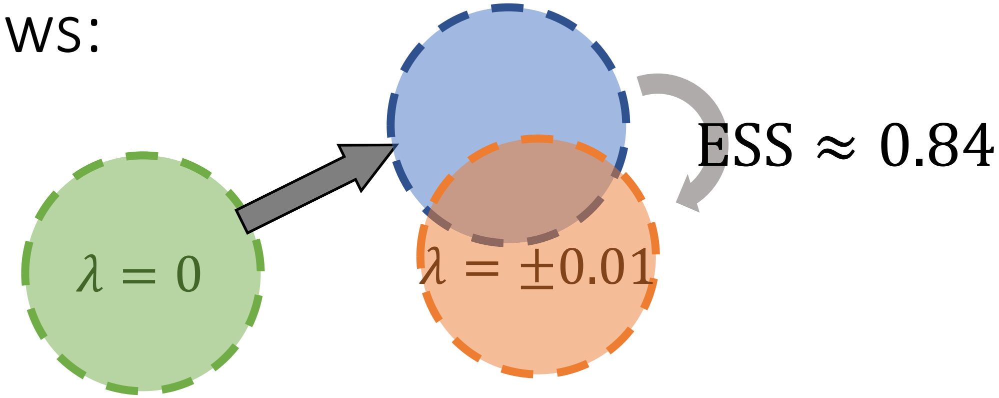
Spatial-temporal anisotropy: $\delta S = -\lambda \frac{\beta}{N_c} \left[\sum_i P_{ti} - \sum_{i<j} P_{ij} \right]$

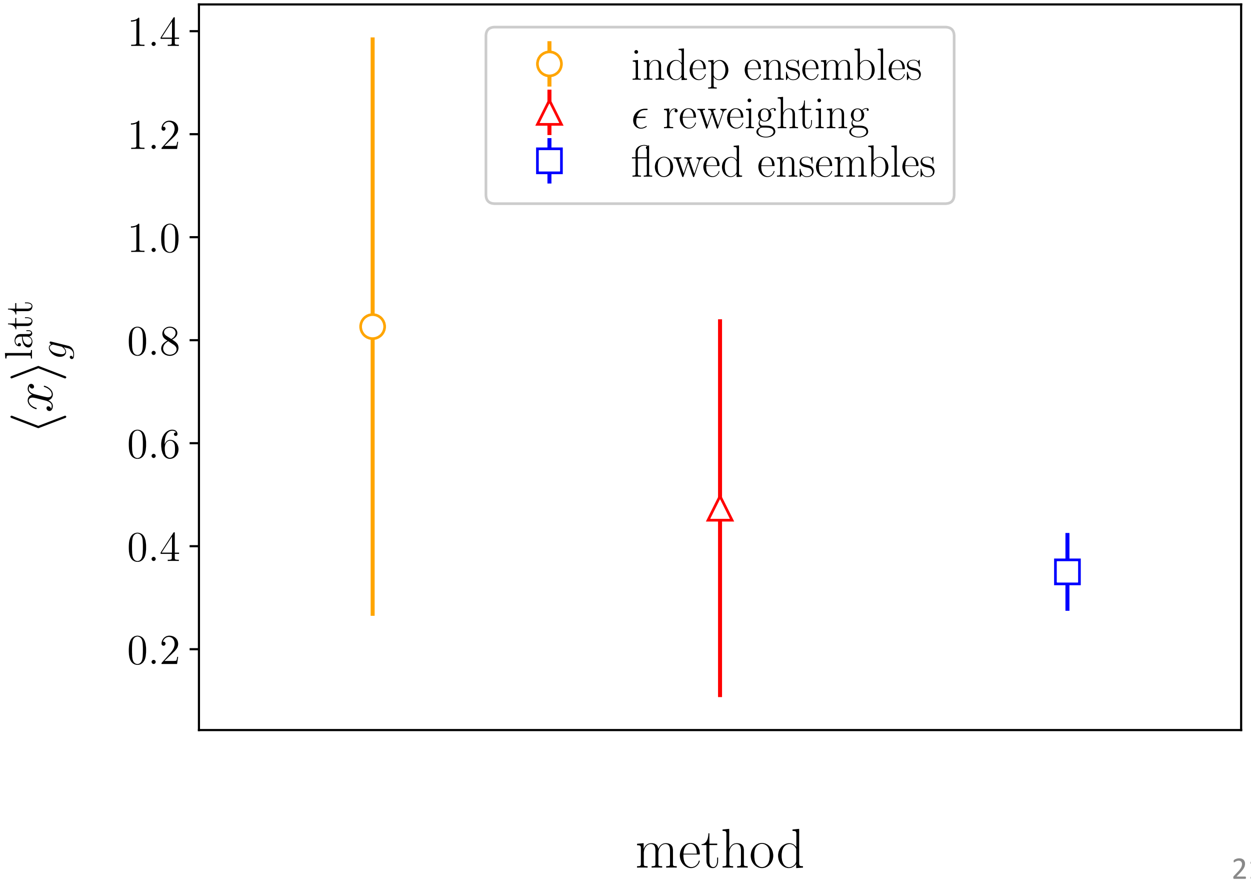
$$\langle x \rangle_g^{\mathrm{latt}} = -\frac{2}{3m} \frac{\partial m}{\partial \lambda}\bigg|_{\lambda=0} \approx -\frac{2}{3m} \frac{1}{\lambda} [m(\lambda) - m(0)]$$

Parameters:

$8^3 \times 16 \quad \beta = 6 \quad \kappa = 0.132$ (quenched)

Flows:





Compute $m(\lambda)$ from $\langle w \, C^{2\mathrm{pt}} \rangle_q$

Demo: Pion $\langle x \rangle_g$ w/ flowed Feynman-Hellmann

[QCDSF-UKQCD 1205.6410]

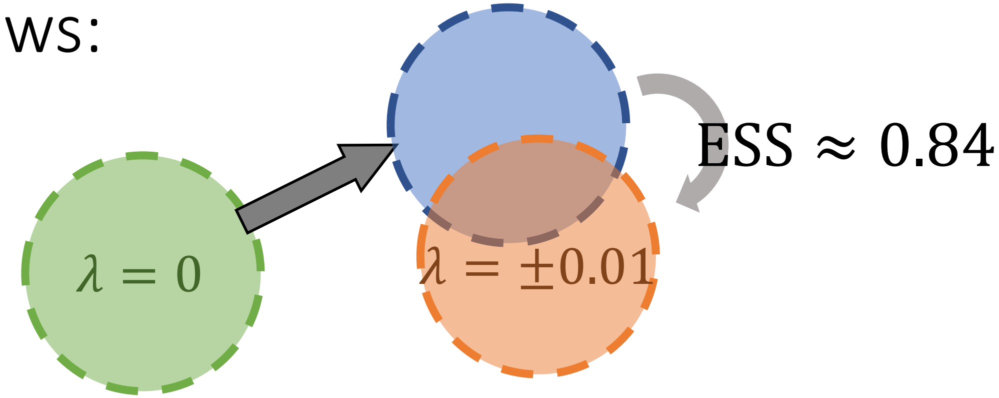
Spatial-temporal anisotropy: $\delta S = -\lambda \frac{\beta}{N_c}\left[\sum_i P_{ti} - \sum_{i<j} P_{ij}\right]$

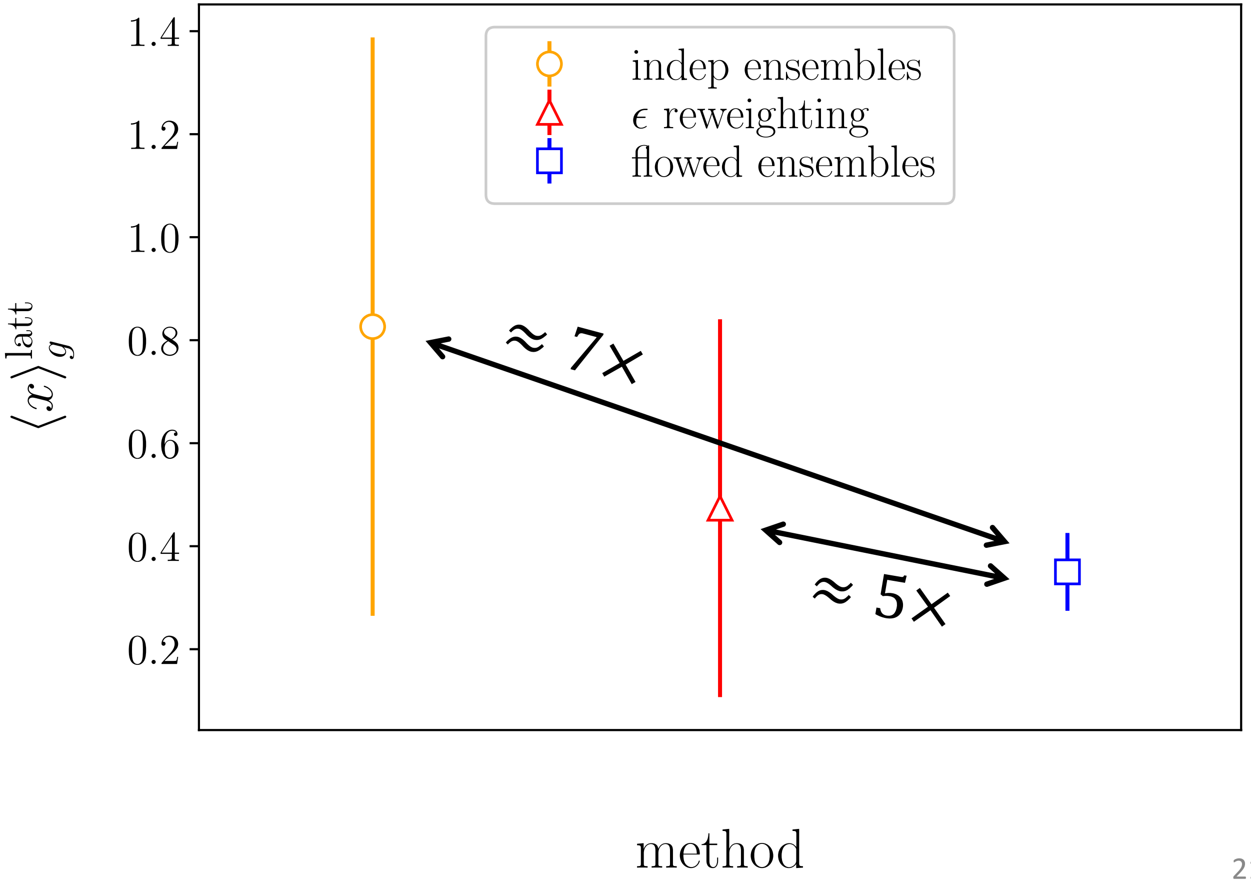
$$\langle x \rangle_g^{\text{latt}} = -\frac{2}{3m}\frac{\partial m}{\partial \lambda}\bigg|_{\lambda=0} \approx -\frac{2}{3m}\frac{1}{\lambda}[m(\lambda) - m(0)]$$

Parameters:

$8^3 \times 16 \quad \beta = 6 \quad \kappa = 0.132$ (quenched)

Flows:

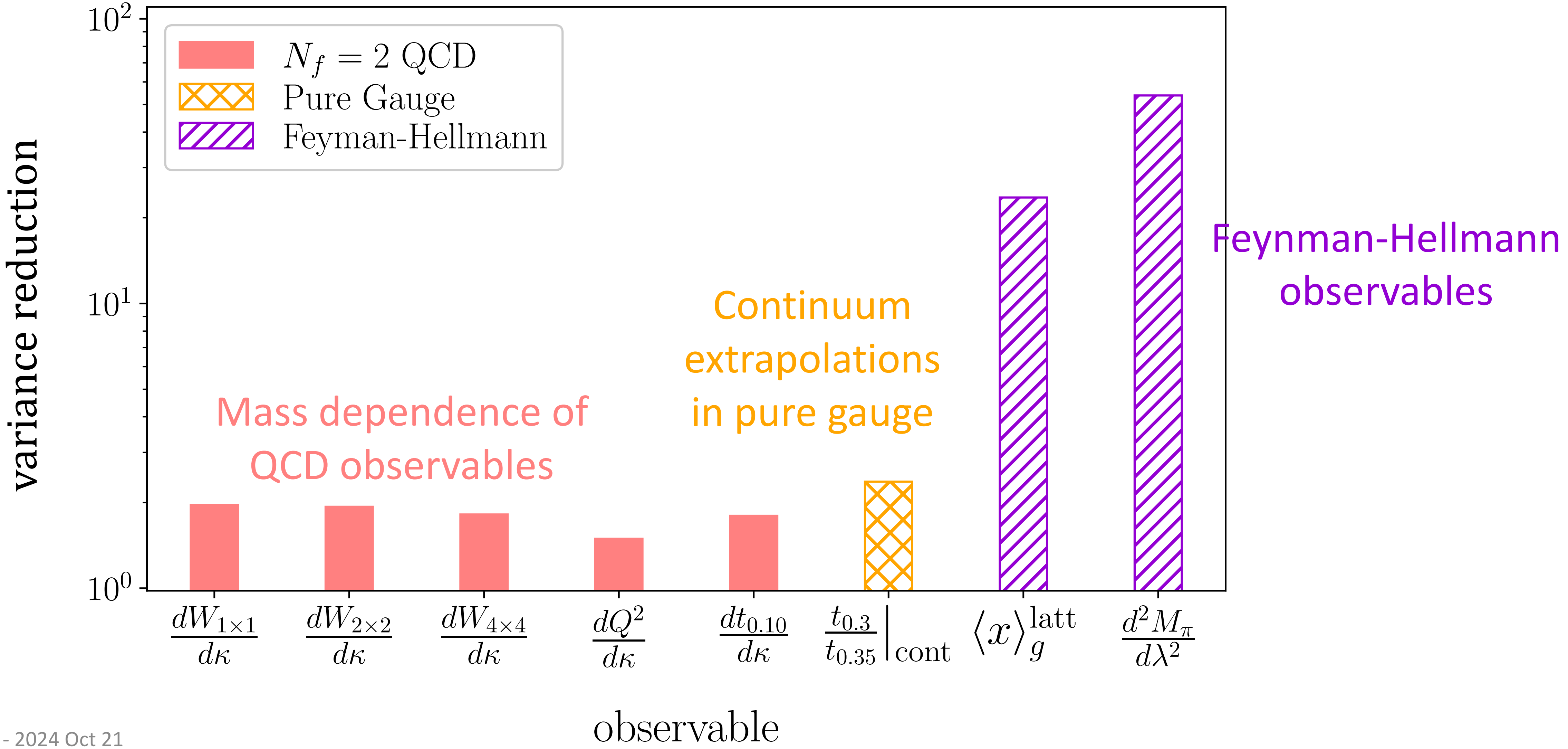




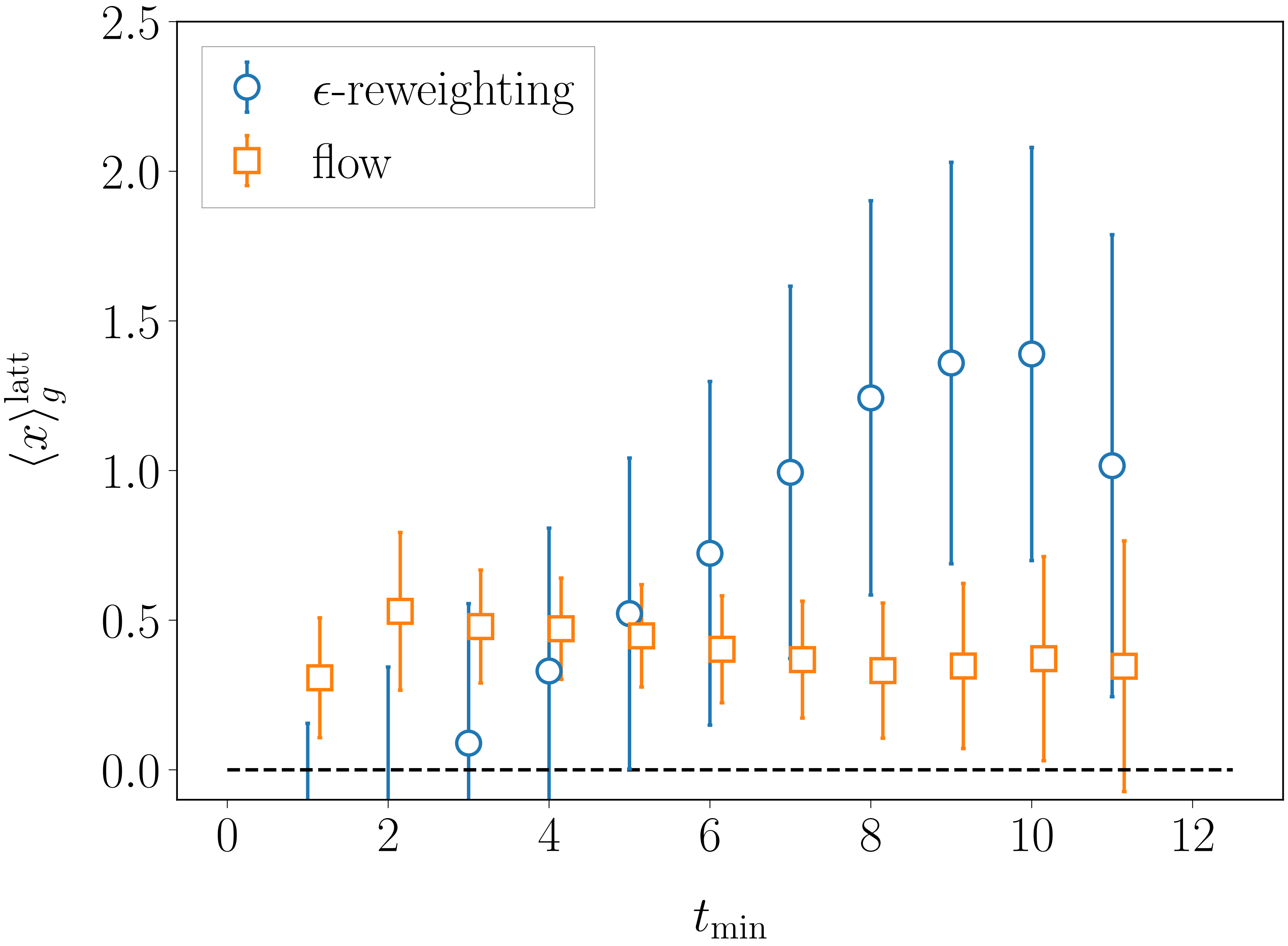
Compute $m(\lambda)$ from $\langle w\, C^{2\text{pt}} \rangle_q$

Demo: More derivative observables

Able to demonstrate significant variance reductions in several observable classes



WIP: extend to QCD



Conclusions

Volume scaling motivates hybrid approaches

- Opportunities for useful applications before full generative modeling possible
- "Hybrid approach" another way of incorporating a priori physics into the flow?

This talk:

- Accelerate sampling w/ REX methods
- Correlated ensembles (for derivative observables)

Many other possibilities!

Some other ideas:

- Learn defects to accelerate sampling?