Applications of machine-learned flows to lattice QCD

Dan Hackett (MIT → FNAL)

Probing the Frontiers of Nuclear Physics with AI at the EIC

Stony Brook / CFNS

September 25, 2023

Collaborators (non-exhaustive)


Phiala Shanahan


Denis Boyda


Ryan Abbott

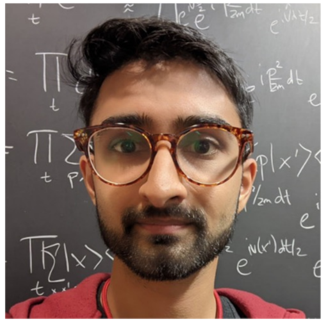

Julian Urban


Fernando Romero-López

NEW YORK UNIVERSITY


Michael Albergo

UNIVERSITÄT BERN


Gurtej Kanwar


Kyle Cranmer

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

DeepMind


Sébastien Racanière


Danilo Rezende


Alex Matthews


Aleksandar Botev


Ali Razavi

Quantum Chromodynamics (QCD)

Part of the Standard Model of particle physics

QFT that describes the strong force

- Dynamics of quarks & gluons
- $SU(3)$ gauge theory

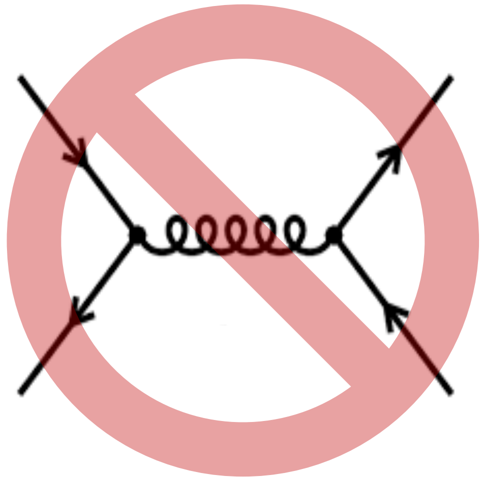
Pen-and-paper calculations don't work

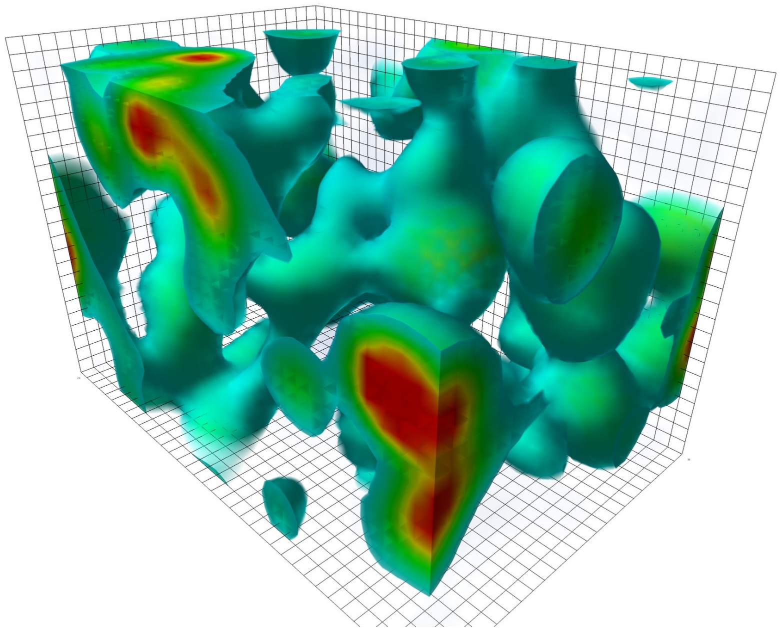
- Strongly interacting QFT
- Non-perturbative dynamics

Instead: lattice QCD!

- Evaluate discretized QCD path integral w/ (MC)MC

Integral part of nuclear/particle theory toolkit

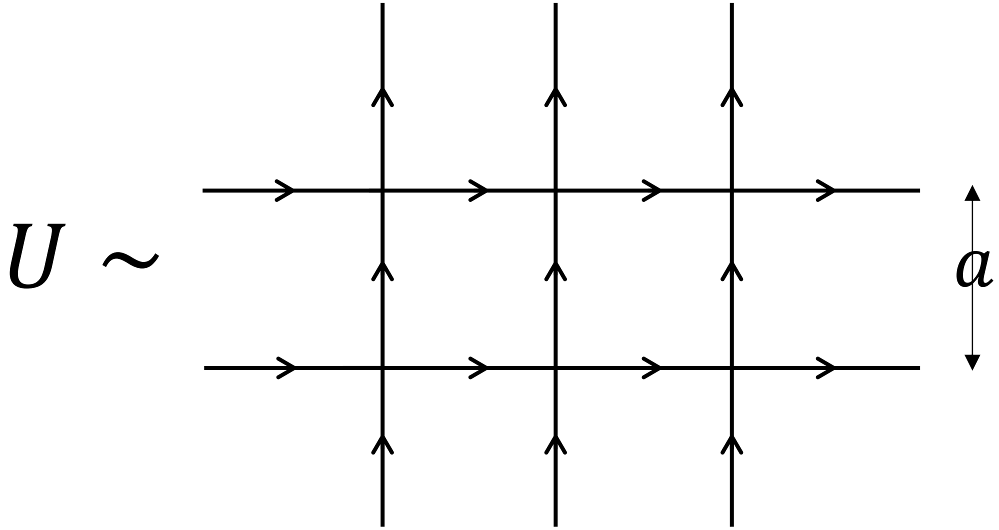




Visualizations of Quantum Chromodynamics

Lattice QCD

Discretized gluon (gauge) field



Setup:

- (3+1)d QFT ⇒ 4d Stat Mech
- Restrict to finite volume
- Discretize spacetime

Use Monte Carlo to (path) integrate over discretized fields:

$$\langle \mathcal{O} \rangle = \int d[U] \frac{1}{Z} e^{-S[U]} \mathcal{O}(U) \equiv \int d[U]\, p(U)\, \mathcal{O}(U) \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(U_i)$$

→ Numerical QFT on supercomputers!

Workflow of a lattice calculation

1. Configuration generation

   Sample ensembles of gauge fields w/ MCMC

   …for different lattice spacings, m_q, V

2. Observable measurements

   Evaluate n-point functions on e/a config

   → compute expectations

3. Analysis

   Extract hadronic properties (masses, matrix elements, etc)

   Extrapolate to physical limit

Workflow of a lattice calculation

1. Configuration generation

   Sample ensembles of gauge fields w/ MCMC

   …for different lattice spacings, m_q, V

2. Observable measurements

   Evaluate n-point functions on e/a config

   → compute expectations

Expensive! (items 1–2)

3. Analysis

   Extract hadronic properties (masses, matrix elements, etc)

   Extrapolate to physical limit

Workflow of a lattice calculation

1. Configuration generation

   Sample ensembles of gauge fields w/ MCMC

   …for different lattice spacings, m_q, V

2. Observable measurements

   Evaluate n-point functions on e/a config

   → compute expectations

3. Analysis

   Extract hadronic properties (masses, matrix elements, etc)

   Extrapolate to physical limit

Accelerate with ML?

Expensive!

Topological freezing

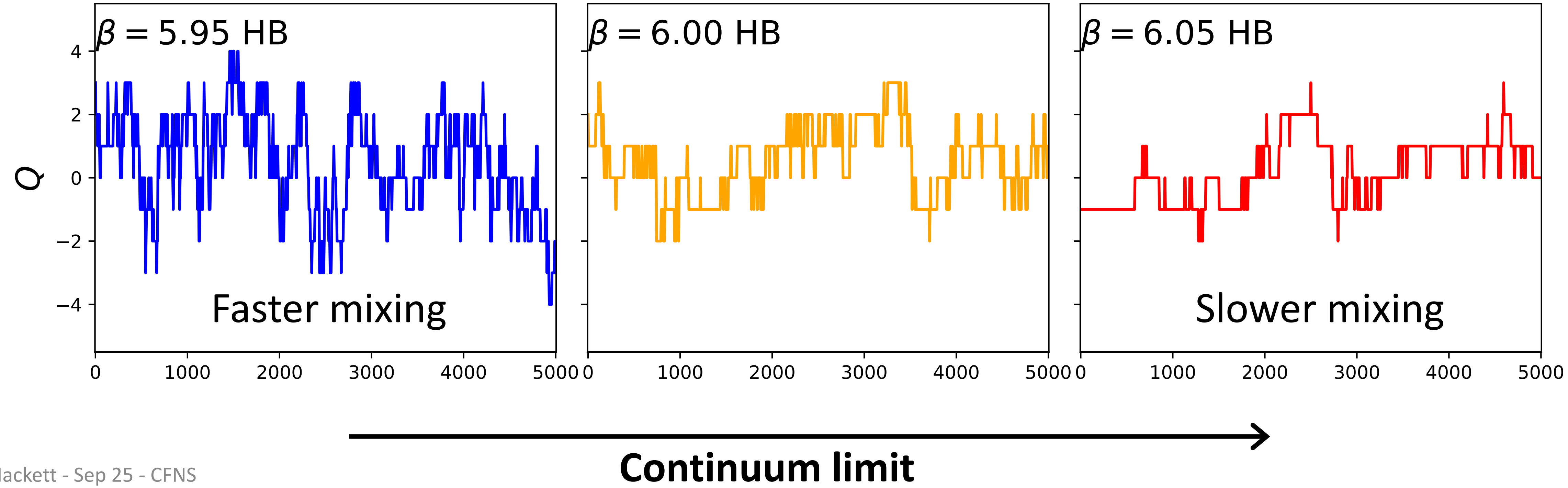
Problem: gauge field distribution is multimodal

Must sample different **topological sectors**

Exponential slowdown of tunneling between sectors w/ standard MCMC algos

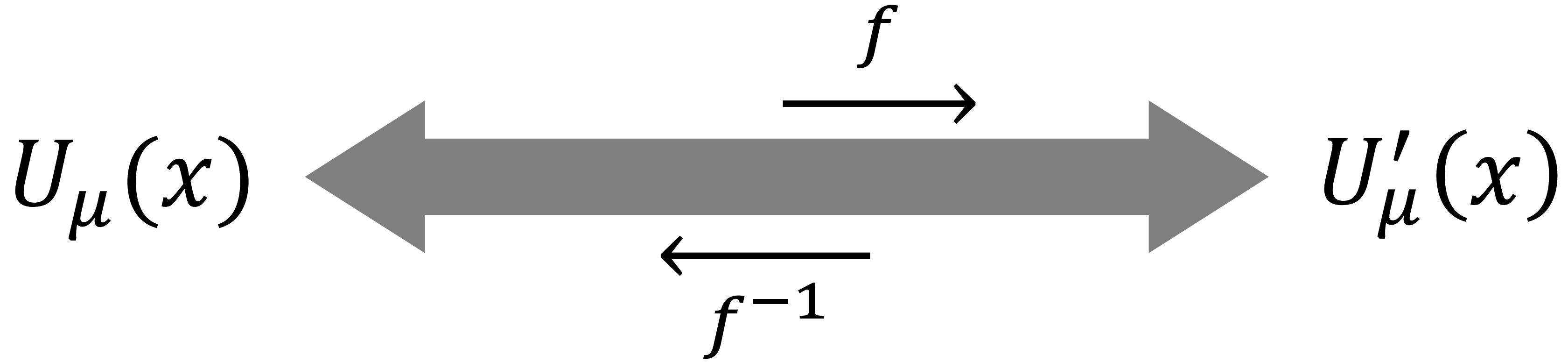
Can result in *effective* loss of ergodicity [2202.1172]

→ Apparent convergence to wrong answers at achievable sample sizes



A New Tool: Machine-learned Flows

f: learned, invertible (diffeomorphic) map between gauge field configs



…with many tunable parameters → ML

…maybe equivariant w/r/t symmetries g of interest $f\big(g(U)\big) = g(f(U))$

…with a tractable Jacobian determinant $J_f(U) = \left|\det \frac{\partial f(U)}{\partial U}\right|$

A New Tool: Machine-learned Flows

Flows are “bridges” between different distributions/theories/actions

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

Exact bridge between r and q

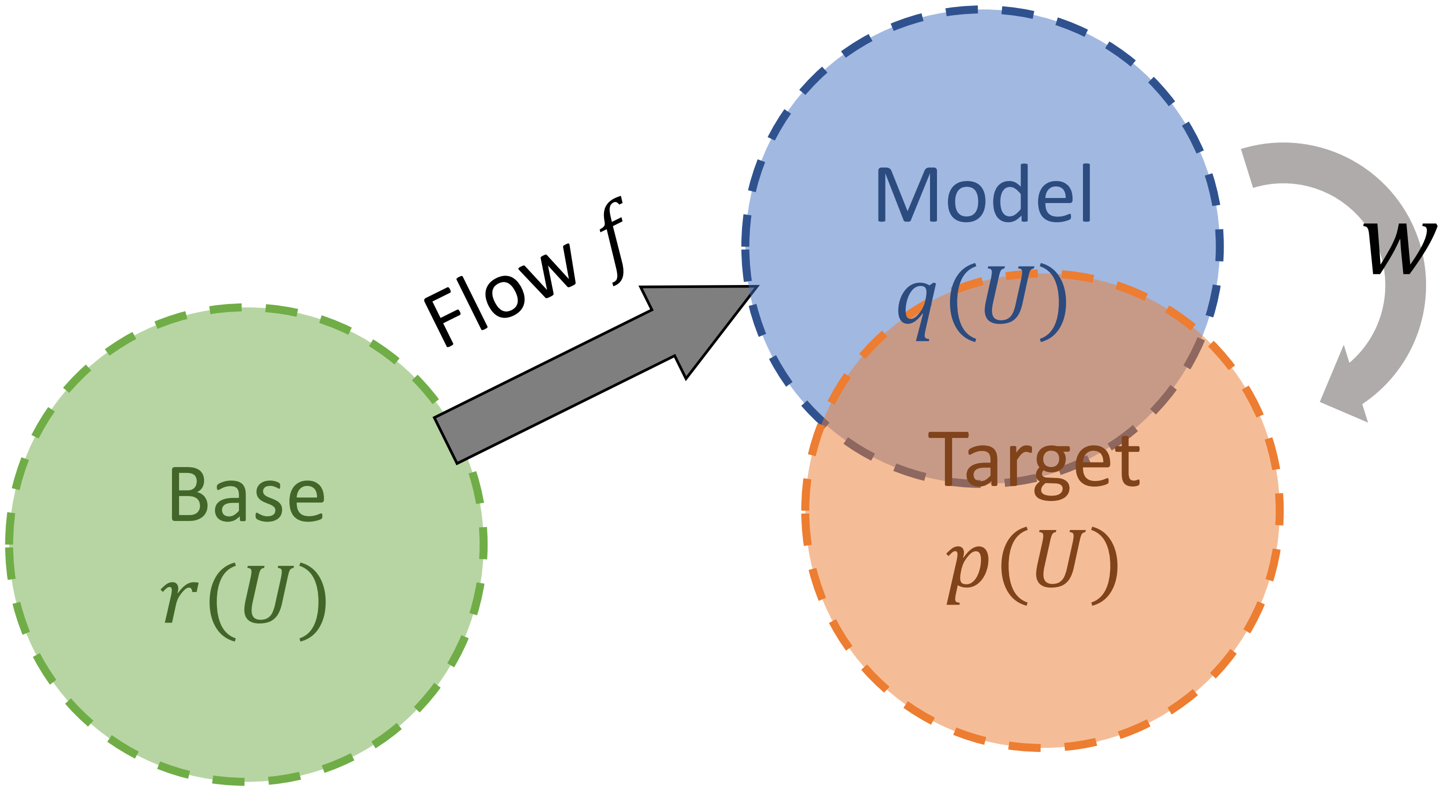
Choose r, but flow induces q

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

→ *Approximate* bridge between r and p

(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 $q \to p$

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

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

Or, ~ equivalently, Metropolize

→ Flow-based MCMC

a.k.a. neural MCMC

a.k.a. …

Progress so far

Flows for LQFT (scalar field theories)

[1904.12072] [2107.00734] [2211.07541]

Gauge-equivariant flows

U(1) [2003.06413]

SU(N) [2003.06413] [2305.0242]

Flows for fermionic theories

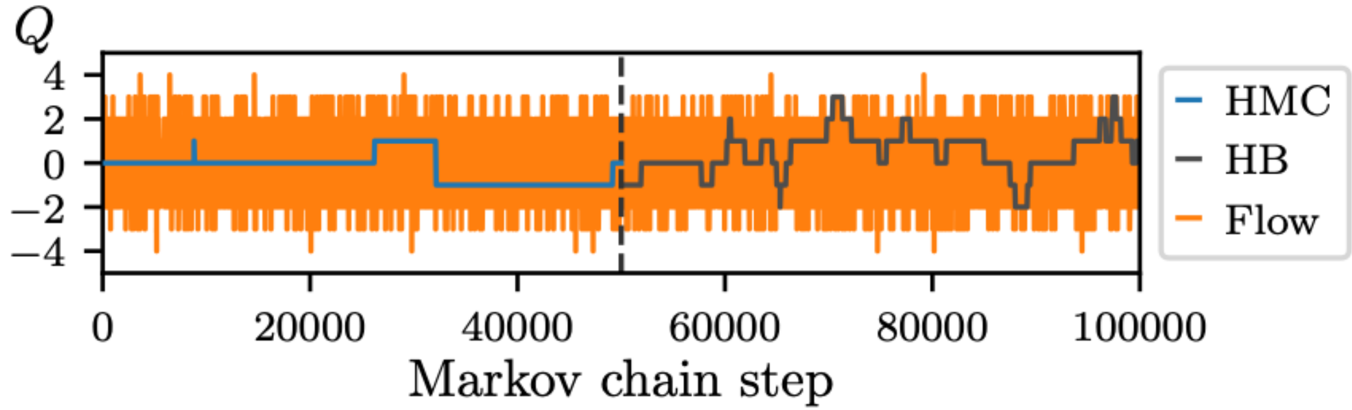
Yukawa model [2106.05934]

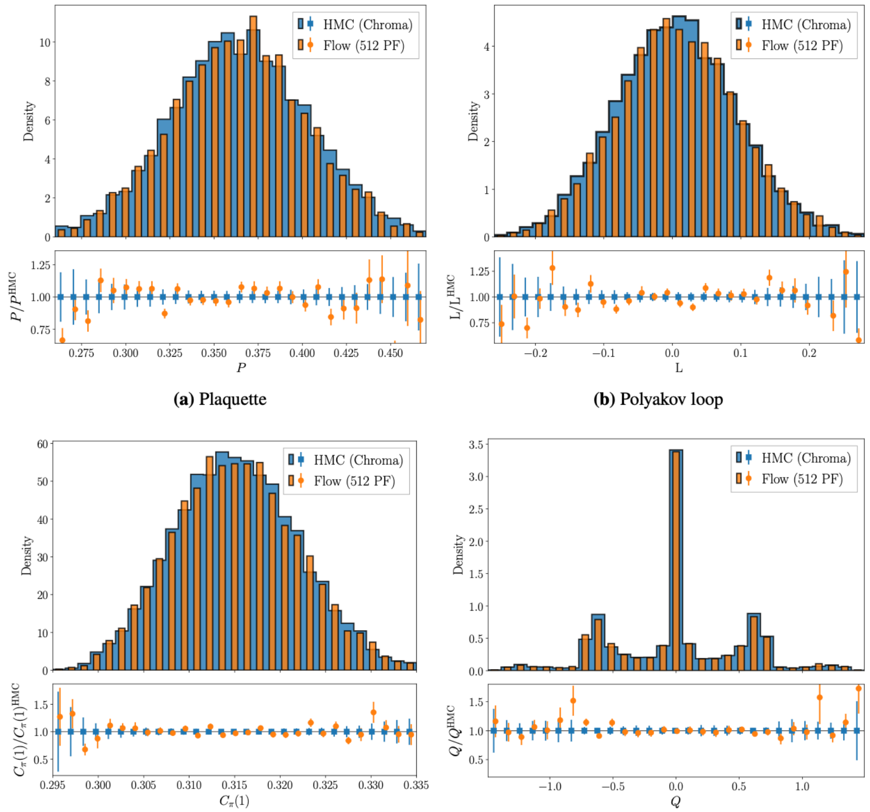
Schwinger model [2202.11712]

Stochastic methods for fermions [2207.08945]

→ First demonstration for QCD! [2208.03832]

Result: Improved sampling in (1+1)d U(1) gauge theory





(a) Plaquette

(b) Polyakov loop

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

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

Progress so far

Flows for LQFT (scalar field theories)

[1904.12072] [2107.00734] [2211.07541]

Gauge-equivariant flows

U(1) [2003.06413]

SU(N) [2003.06413] [2305.0242]

Flows for fermionic theories

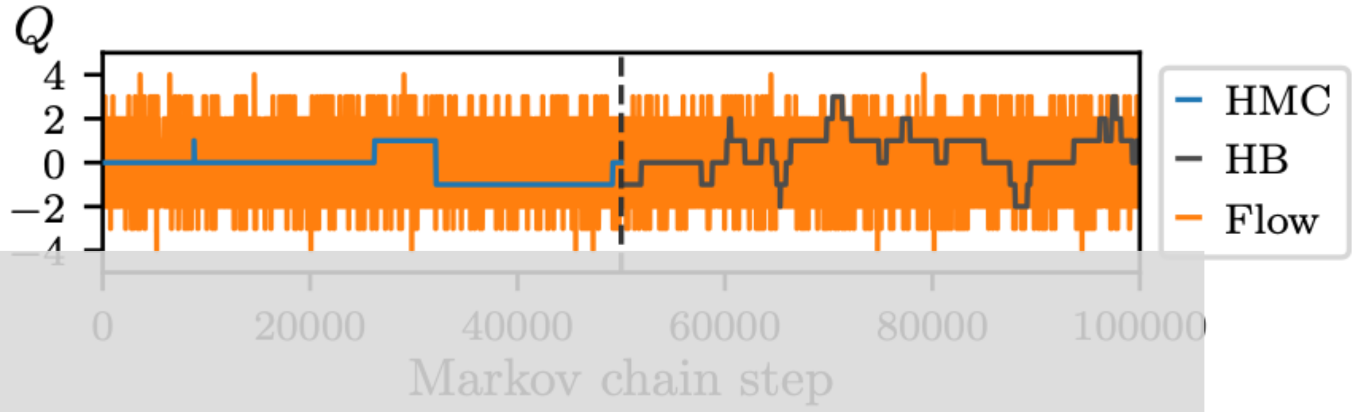
Yukawa model [2106.05934]

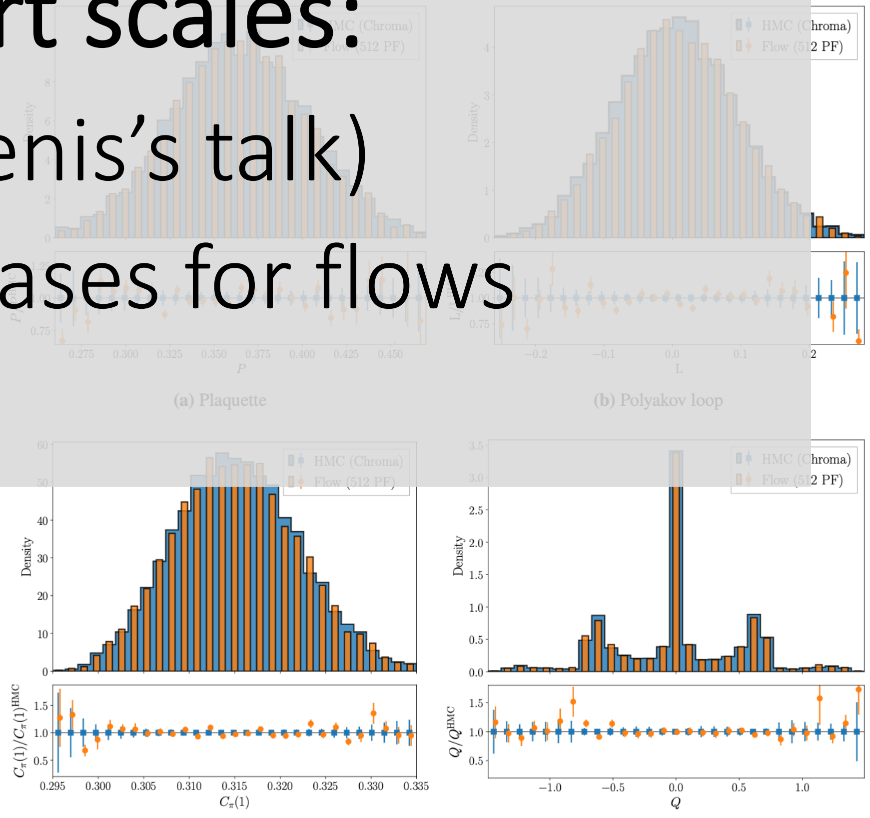
Schwinger model [2202.11712]

Stochastic methods for fermions [2207.08945]

→ First demonstration for QCD! [2208.03832]

Result: Improved sampling in (1+1)d U(1) gauge theory





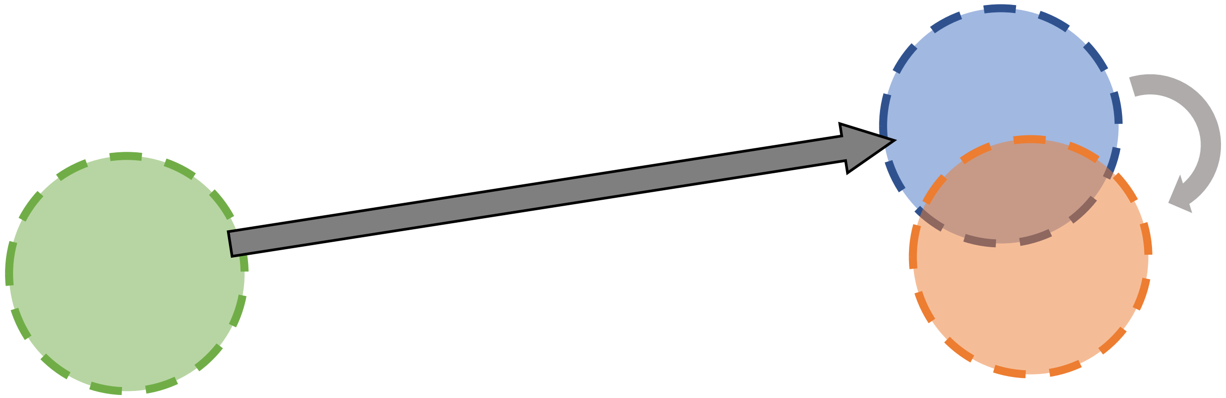
(c) Pion correlation function at $x_0 = 1$

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

Towards state-of-the-art scales:

1. Better QCD models (Denis's talk)
2. Different algorithmic use cases for flows

Applications



Basic idea: don't flow as far

→ 3x demos

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

Avoid fermionic complications (for now)

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

"Residual flows" [2305.02402]

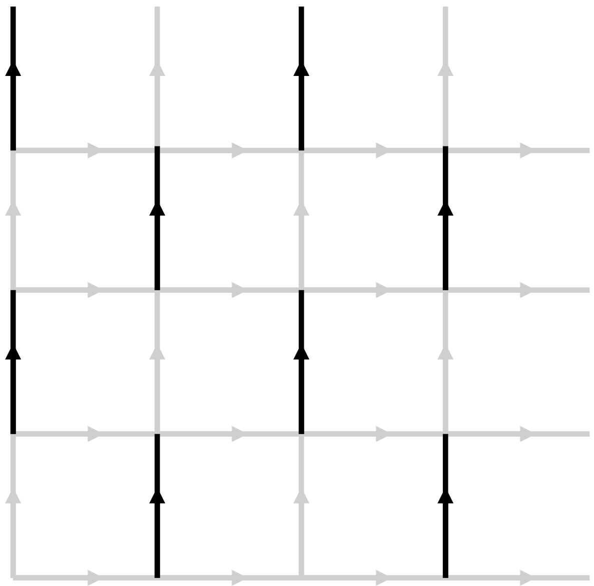
~ ODE flow + variable partitioning (cf. coupling layers)

→ tractable/inexpensive exact Jacobian

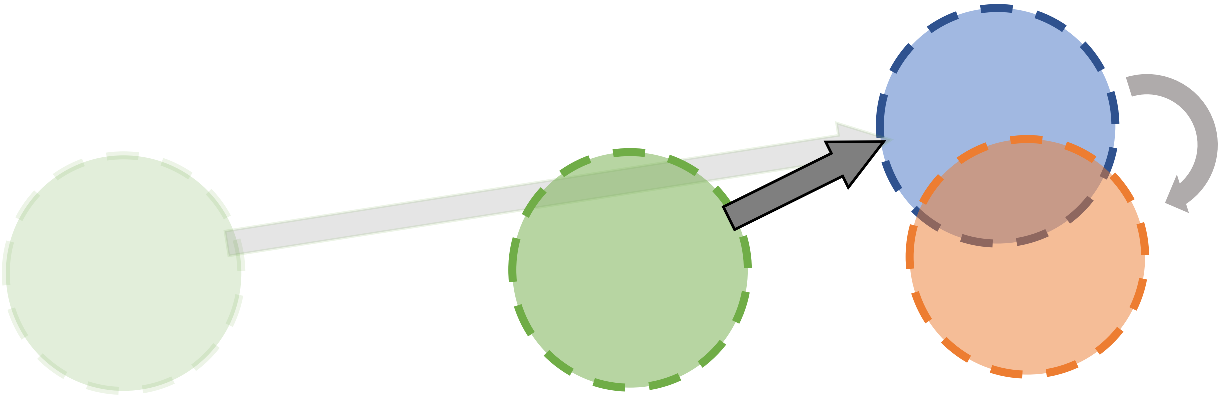
Features:

Gauge equivariant flows

Translationally invariant ↔ volume transferrable



Applications



Basic idea: don't flow as far

→ 3x demos

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

Avoid fermionic complications (for now)

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

"Residual flows" [2305.02402]

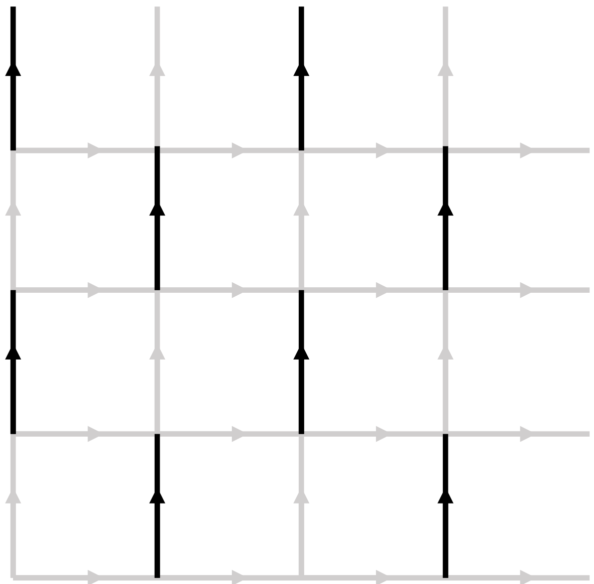
∼ ODE flow + variable partitioning (cf. coupling layers)

→ tractable/inexpensive exact Jacobian

Features:

Gauge equivariant flows

Translationally invariant ↔ volume transferrable



App 1: Correlated ensembles

Flow an ensemble

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

This is useful!

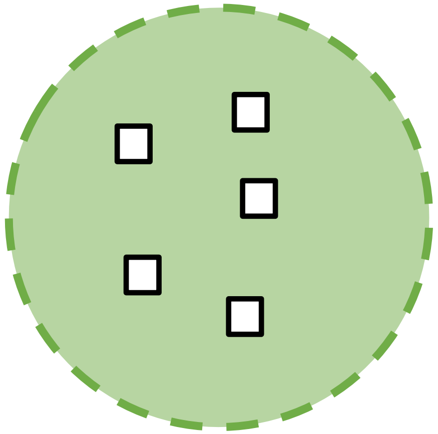
e.g. for noise cancellation in differences

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

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

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

$\{U\} \sim r$



Application: Feynman-Hellmann $\quad S \rightarrow S + \lambda\, O \quad \left.\frac{\partial E_h}{\partial \lambda}\right|_{\lambda=0} \sim \langle h|O|h\rangle$

(Complication: involves fits for E_h, but same idea)

See also [Bacchio 2305.07932]

App 1: Correlated ensembles

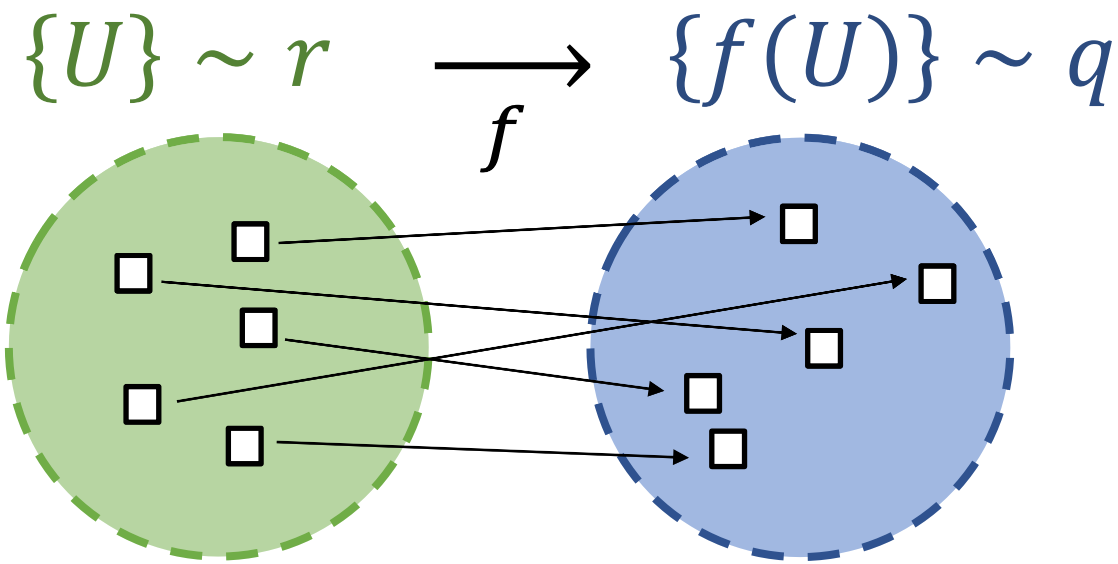
Flow an ensemble

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

This is useful!

e.g. for noise cancellation in differences

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



Application: Feynman-Hellmann $\qquad S \to S + \lambda\, O \qquad \left.\frac{\partial E_h}{\partial \lambda}\right|_{\lambda=0} \sim \langle h|O|h\rangle$

(Complication: involves fits for E_h, but same idea)

See also [Bacchio 2305.07932]

App 1: Correlated ensembles

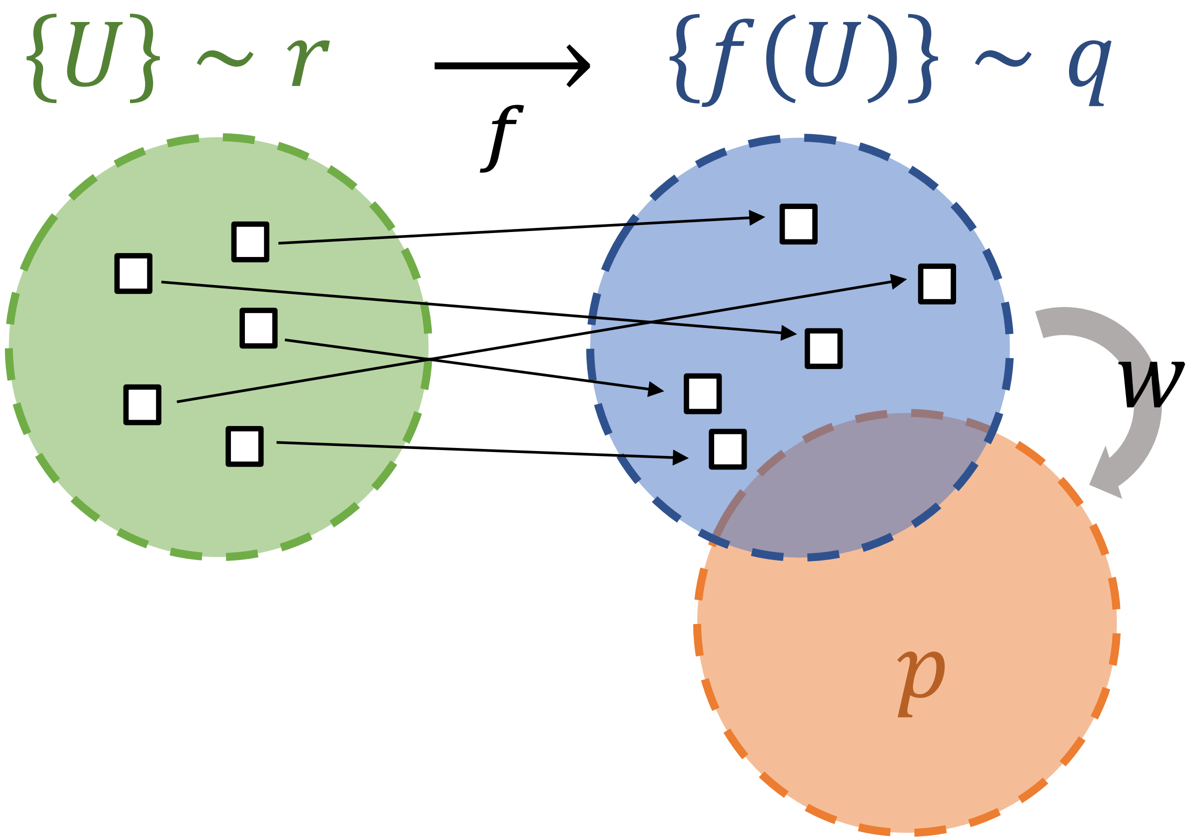
Flow an ensemble

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

This is useful!

e.g. for noise cancellation in differences

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



Application: Feynman-Hellmann $\quad S \rightarrow S + \lambda\, O \quad \left.\frac{\partial E_h}{\partial \lambda}\right|_{\lambda=0} \sim \langle h|O|h \rangle$

(Complication: involves fits for E_h, but same idea)

See also [Bacchio 2305.07932]

App 1: Pion $\langle x \rangle_g$ w/ flowed Feynman-Hellmann

[QCDSF-UKQCD 1205.6410]

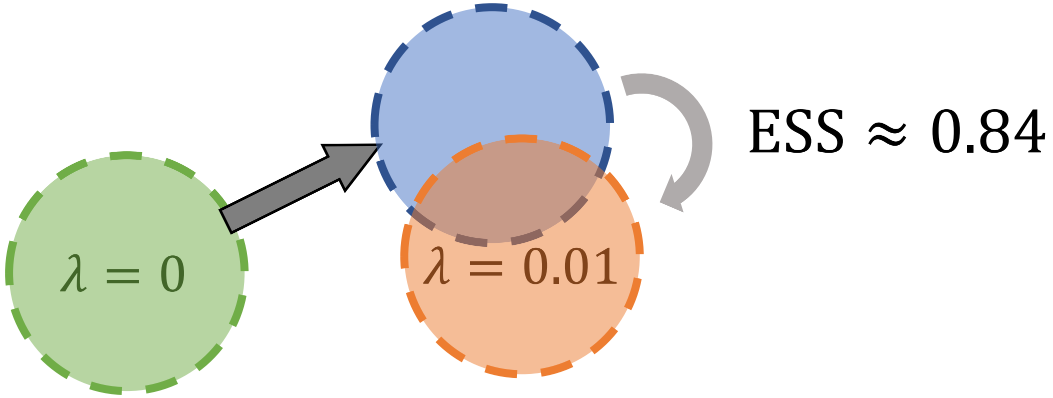
Spatial-temporal anisotropy: $\delta S = -\lambda \frac{\beta}{N_c} \left[\sum_i P_{ti} - \sum_{i<j} P_{ij} \right]$

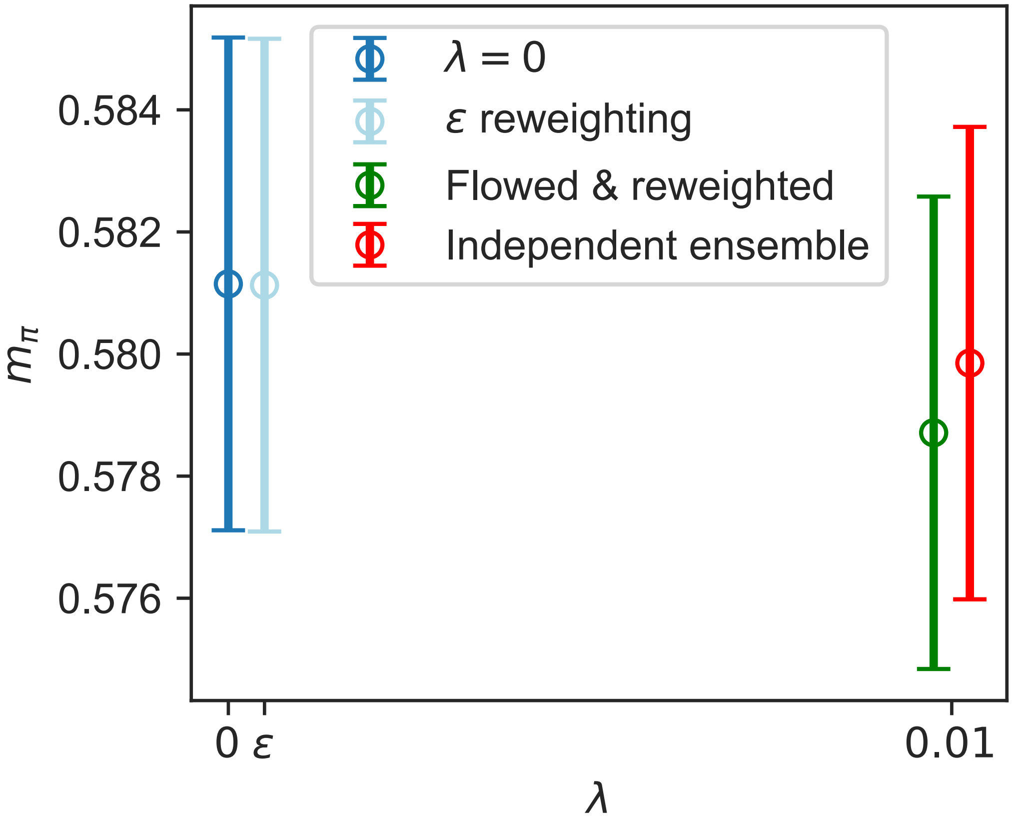
$$\langle x \rangle_g^{\text{lat}} = -\frac{2}{3m} \frac{\partial m}{\partial \lambda}\Big|_{\lambda=0} \approx -\frac{2}{3m} \frac{1}{\lambda} [m(\lambda) - m(0)]$$

Parameters:

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

Flow:





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

App 1: Pion $\langle x \rangle_g$ w/ flowed Feynman-Hellmann

[QCDSF-UKQCD 1205.6410]

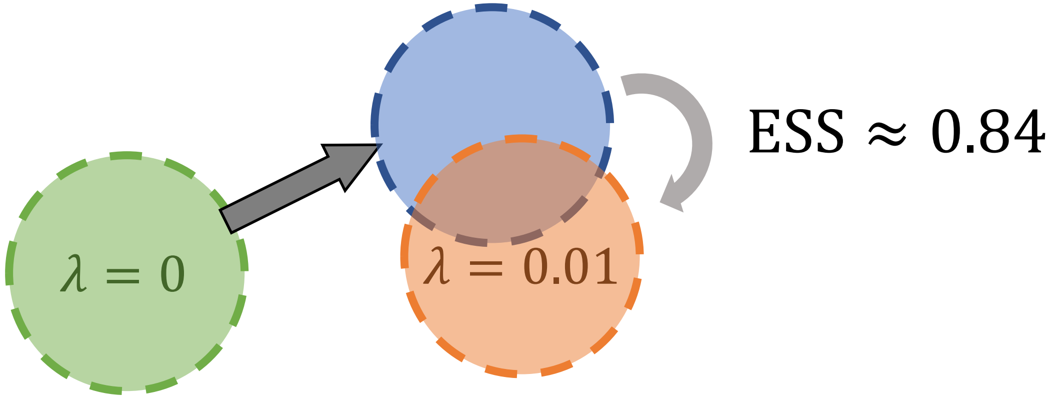
Spatial-temporal anisotropy: $\delta S = -\lambda \frac{\beta}{N_c} \left[\sum_i P_{ti} - \sum_{i<j} P_{ij} \right]$

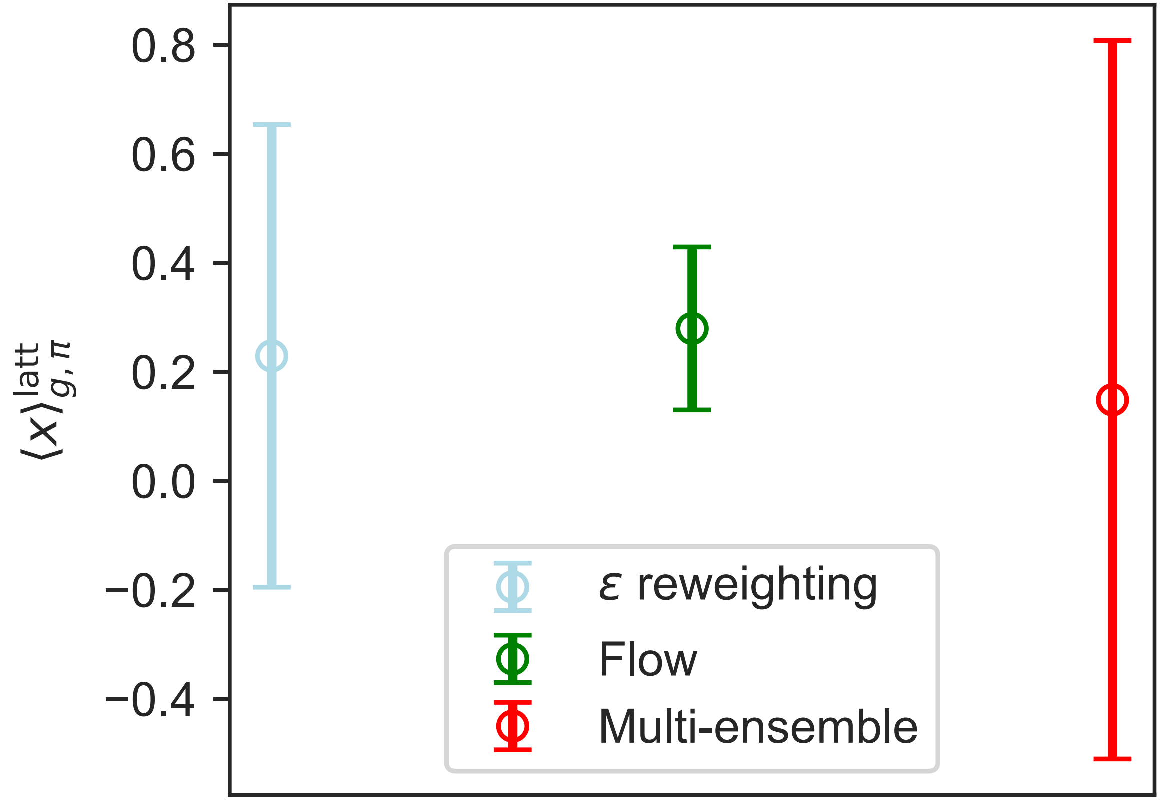
$$\langle x \rangle_g^{\text{lat}} = -\frac{2}{3m} \frac{\partial m}{\partial \lambda}\bigg|_{\lambda=0} \approx -\frac{2}{3m} \frac{1}{\lambda} [m(\lambda) - m(0)]$$

Parameters:

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

Flow:





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

App 1: Pion $\langle x \rangle_g$ w/ flowed Feynman-Hellmann

[QCDSF-UKQCD 1205.6410]

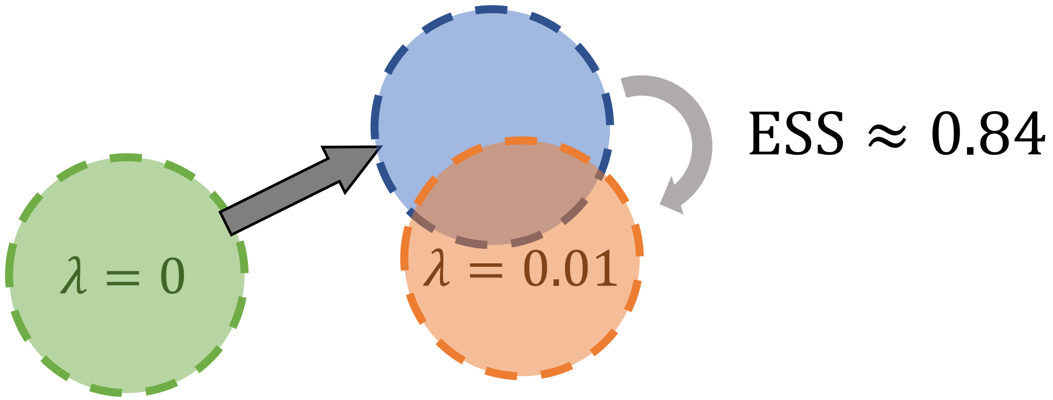
Spatial-temporal anisotropy: $\delta S = -\lambda \frac{\beta}{N_c} \left[\sum_i P_{ti} - \sum_{i<j} P_{ij} \right]$

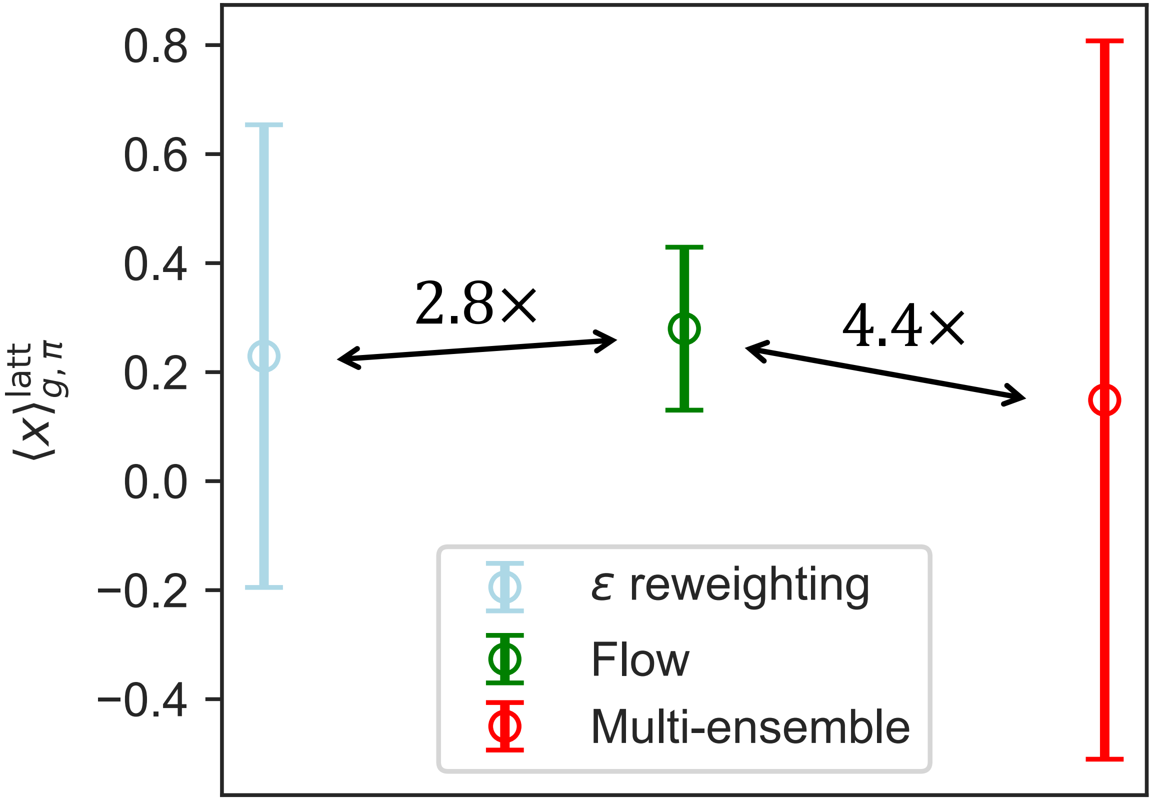
$$\langle x \rangle_g^{\text{lat}} = -\frac{2}{3m} \frac{\partial m}{\partial \lambda}\bigg|_{\lambda=0} \approx -\frac{2}{3m} \frac{1}{\lambda} [m(\lambda) - m(0)]$$

Parameters:

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

Flow:



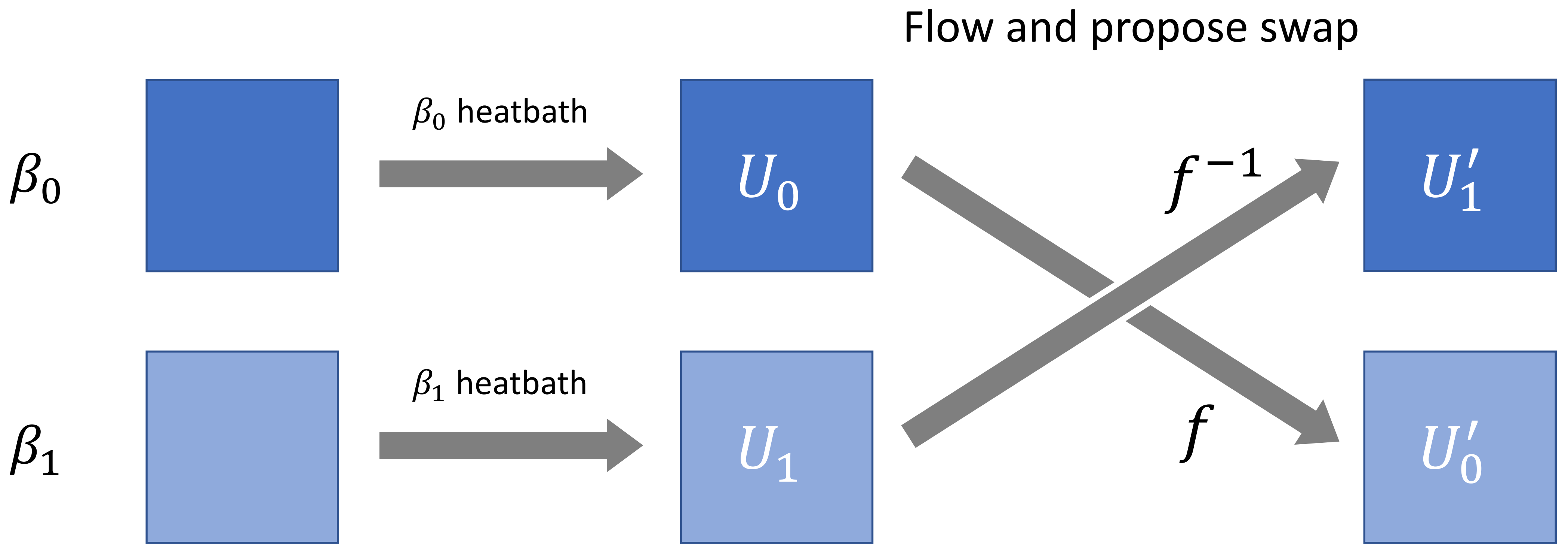


Compute $m(\lambda)$ from $\langle w\, C^{\text{2pt}} \rangle_q$

App 2: Transformed Replica EXchange (T-REX)

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

Simultaneously sample chains for different targets



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

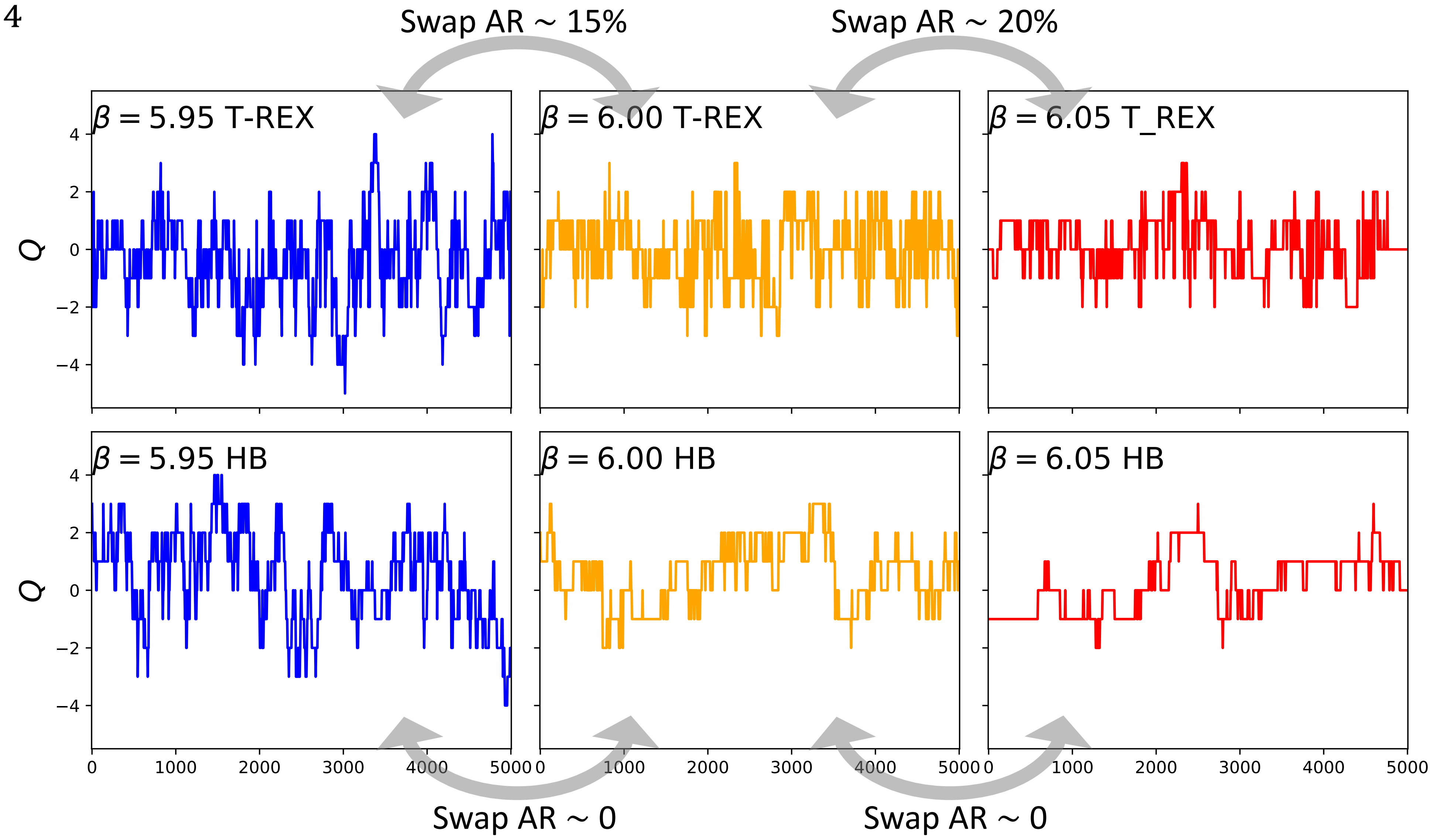
App 2: T-REX Results

Three target βs on 12^4

Two different flows

$5.95 \leftrightarrow 6$

$6 \leftrightarrow 6.05$



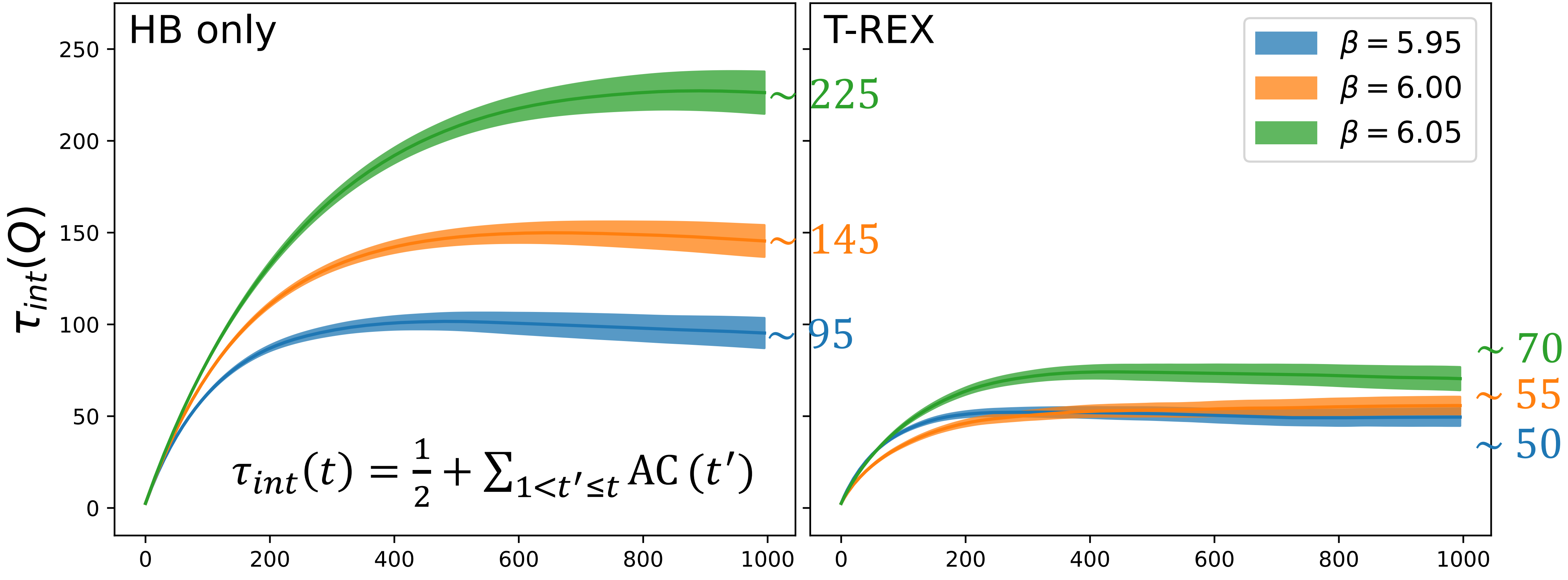
1 step = 5 HB + 2 OR, propose swaps every 5 steps

App 2: T-REX Results

Speed-up for multi-ensemble calculation

T-REX streams correlated (useful!)

Neglecting flow costs!



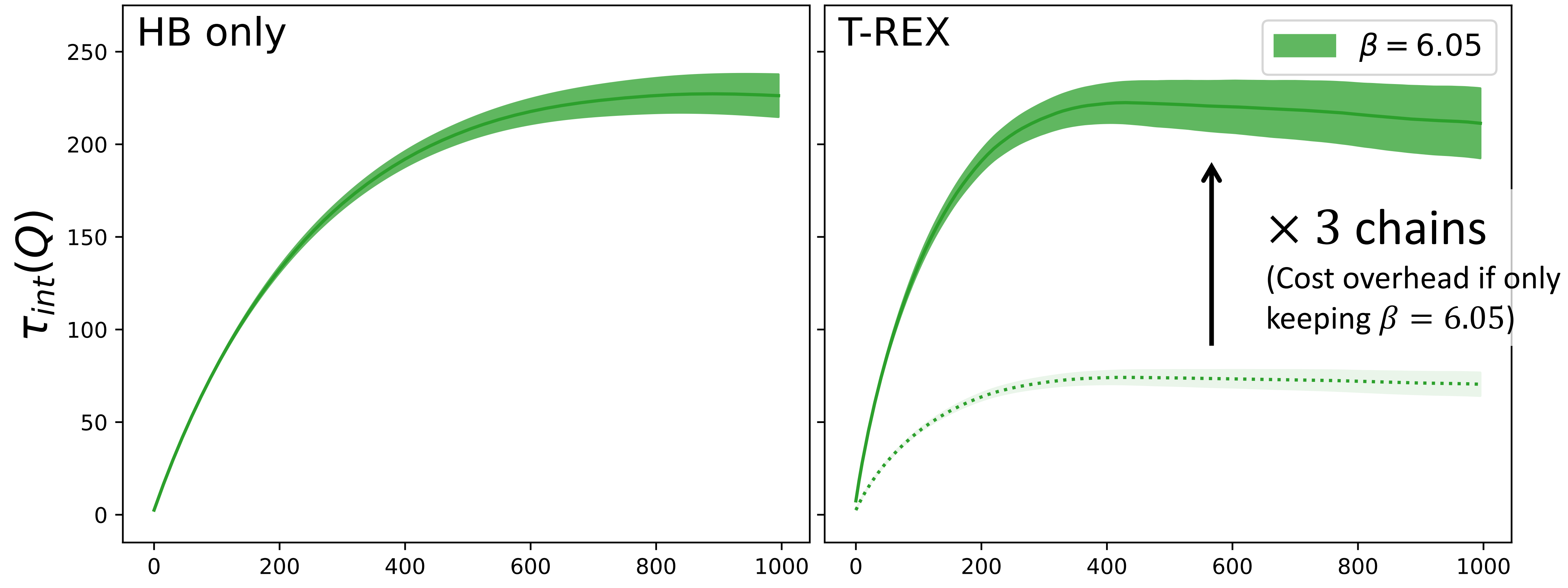
App 2: T-REX Results

Speed-up for multi-ensemble calculation

T-REX streams correlated (useful!)

Break-even for sampling $\beta = 6.05$

} Neglecting flow costs!



App 3: Parallel Tempering on Boundary Conditions (PTBC)

[Hasenbusch 1706.04443]

[Bonnano Bonati D'Elia 2012.14000]

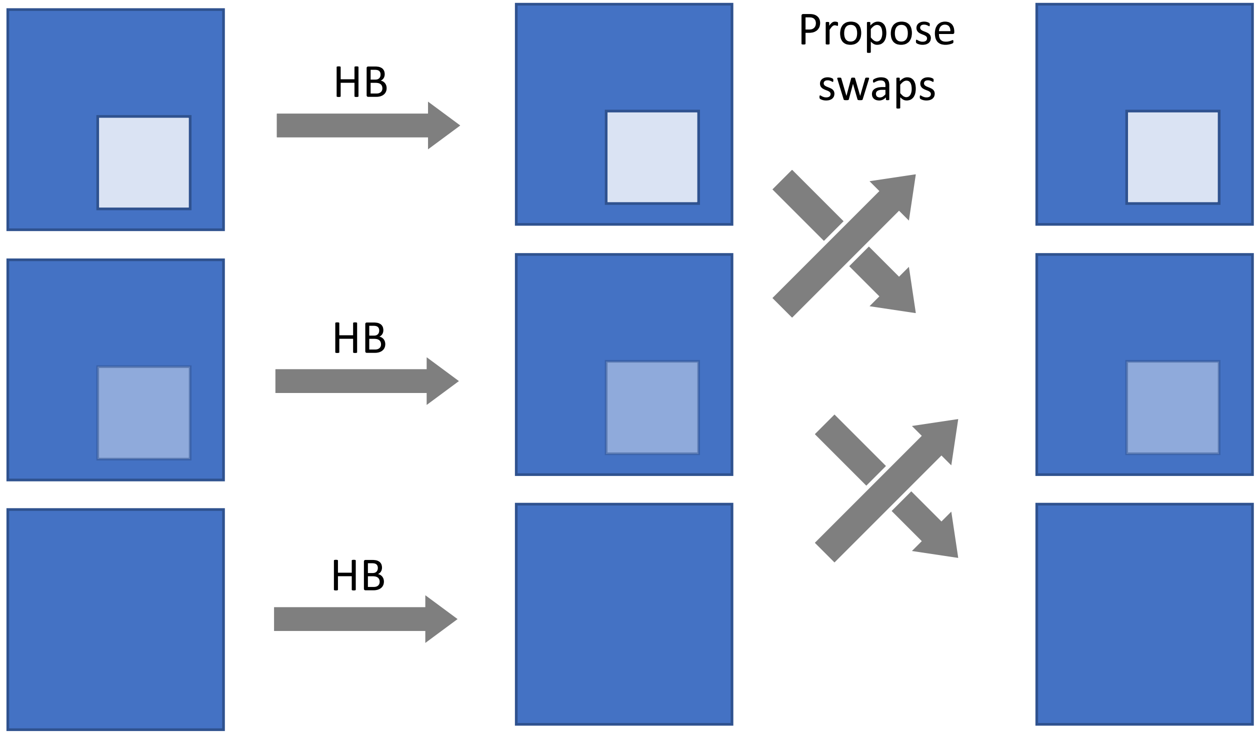
Introduce localized defect

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

"Poke a hole in the boundary"

→ Faster topological mixing

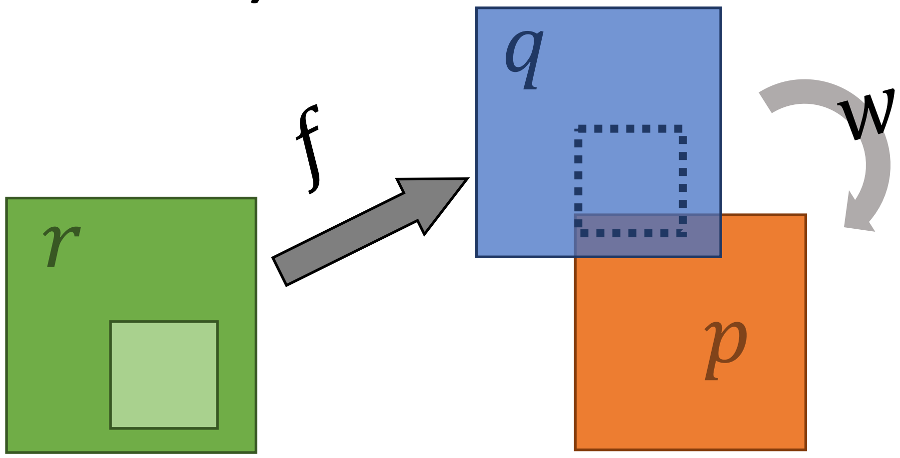
Remove defect w/ REX



App 3: Defect Repair Replica EXchange (DR-REX)

Train flow to repair defect

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

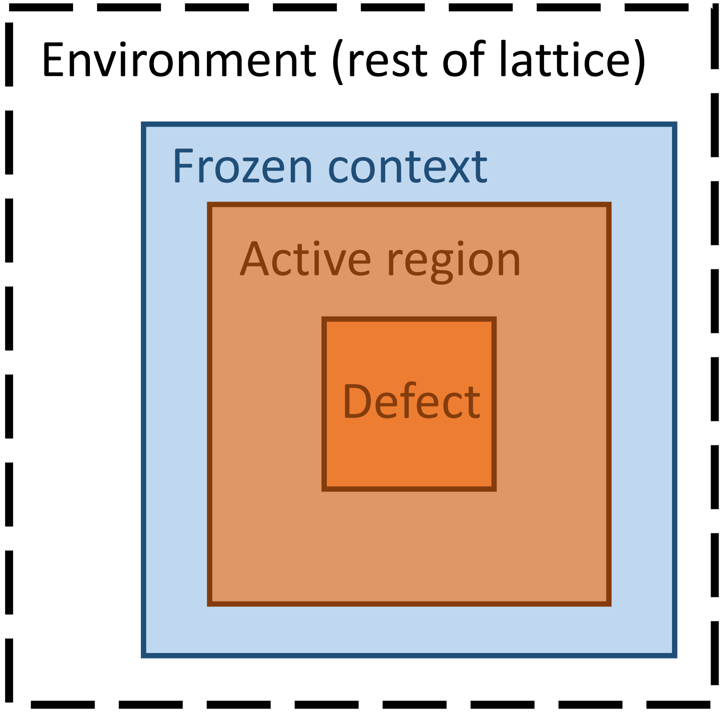


Defect has localized physical effects

Flow acts on subvolume

→ No ESS volume scaling

→ Volume-independent computational cost



App 3: DR-REX Results

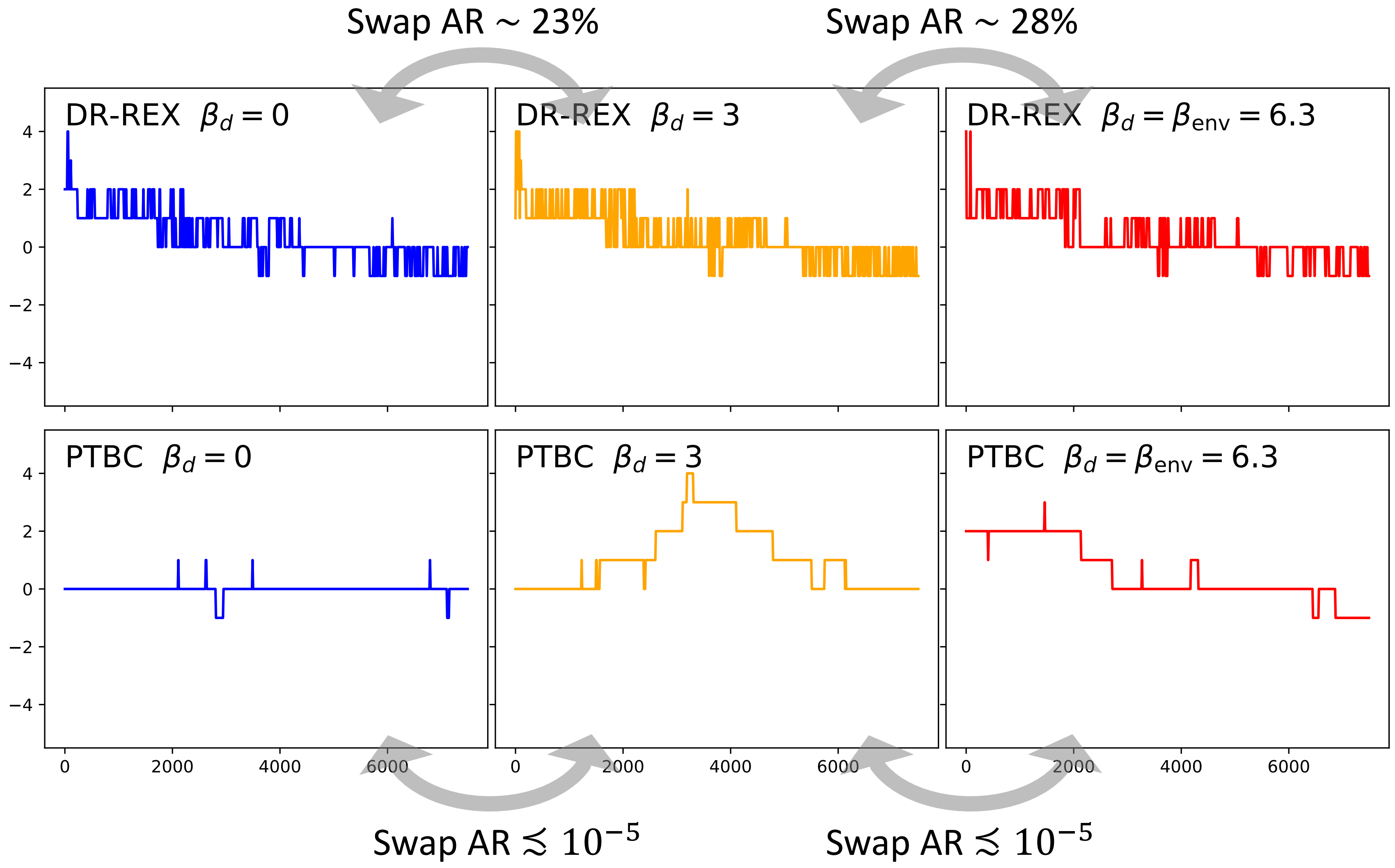
Target: $\beta = 6.3$ on 16^4

2^3 OBC defect

Two flows to repair

$$\beta_d = 0 \rightarrow 3 \rightarrow 6.3$$

Flows act on 8^4 subvolume



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

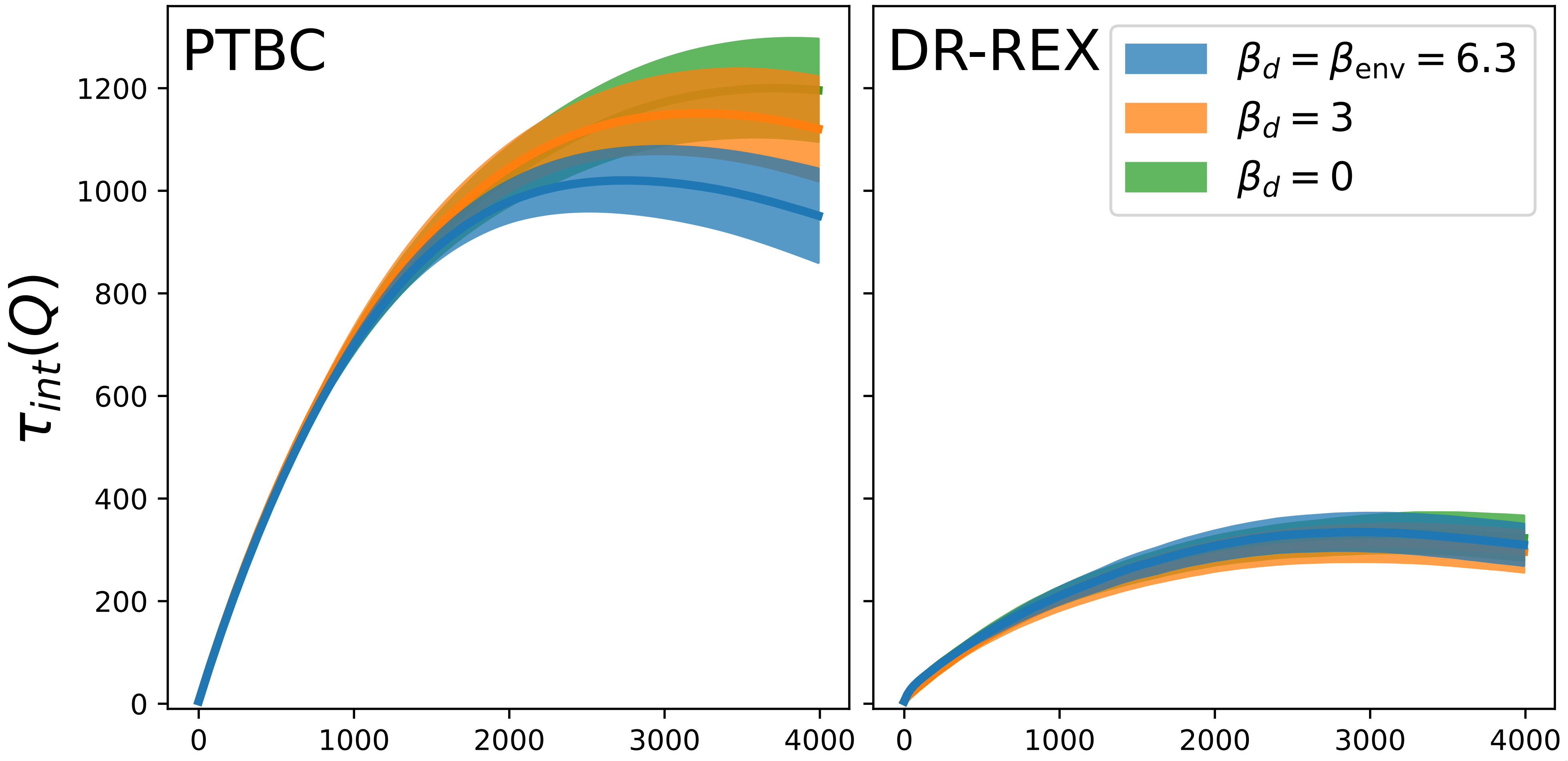
1 step = 1 HB + 5 OR, propose swaps every 10 steps

App 3: DR-REX Results

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

Two flows to repair a 2^3 OBC defect $\beta_d = 0 \rightarrow 3 \rightarrow 6.3$

Flows act on 8^4 subvolume

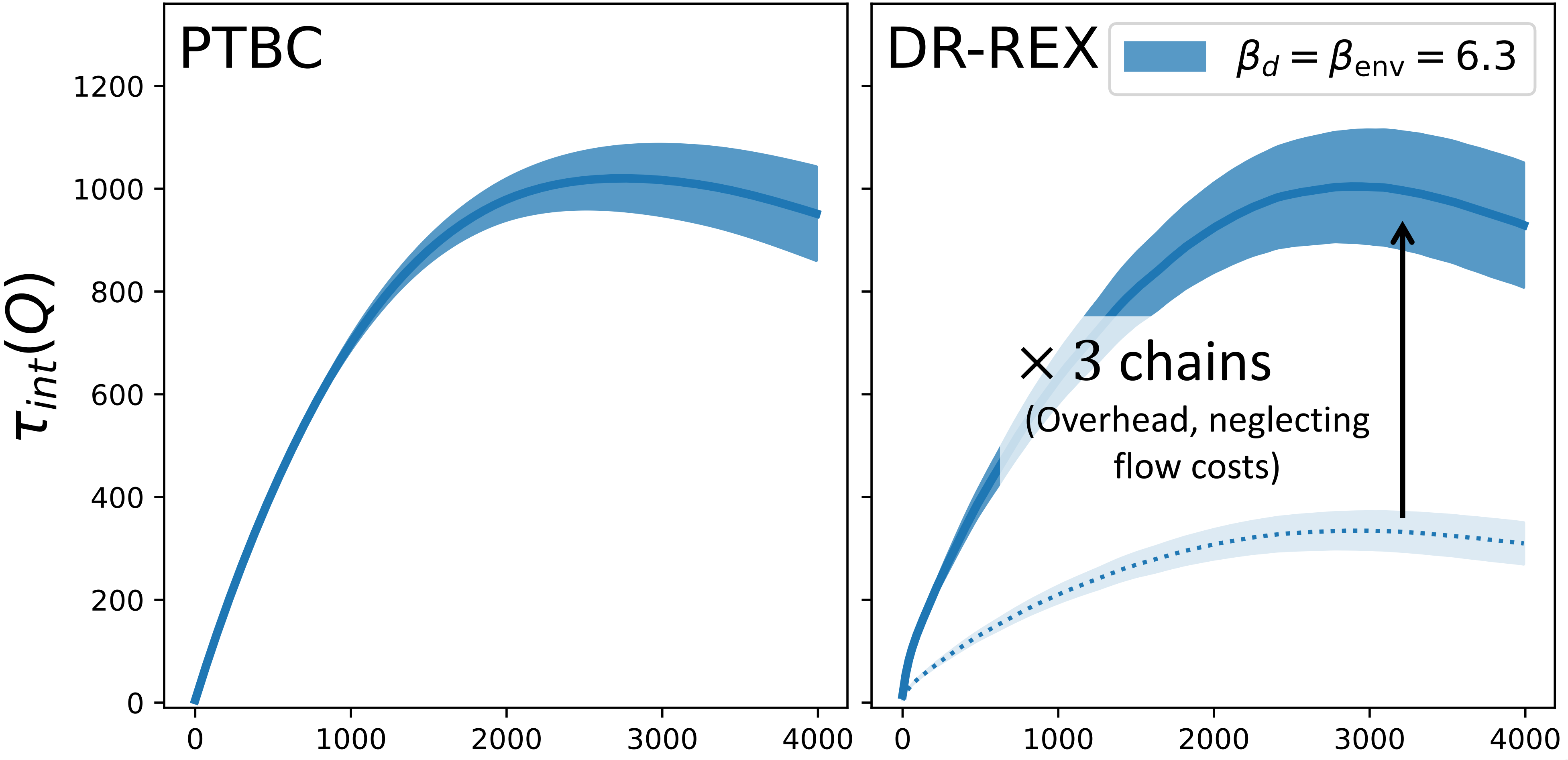


App 3: DR-REX Results

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

Two flows to repair a 2^3 OBC defect $\beta_d = 0 \to 3 \to 6.3$

Flows act on 8^4 subvolume

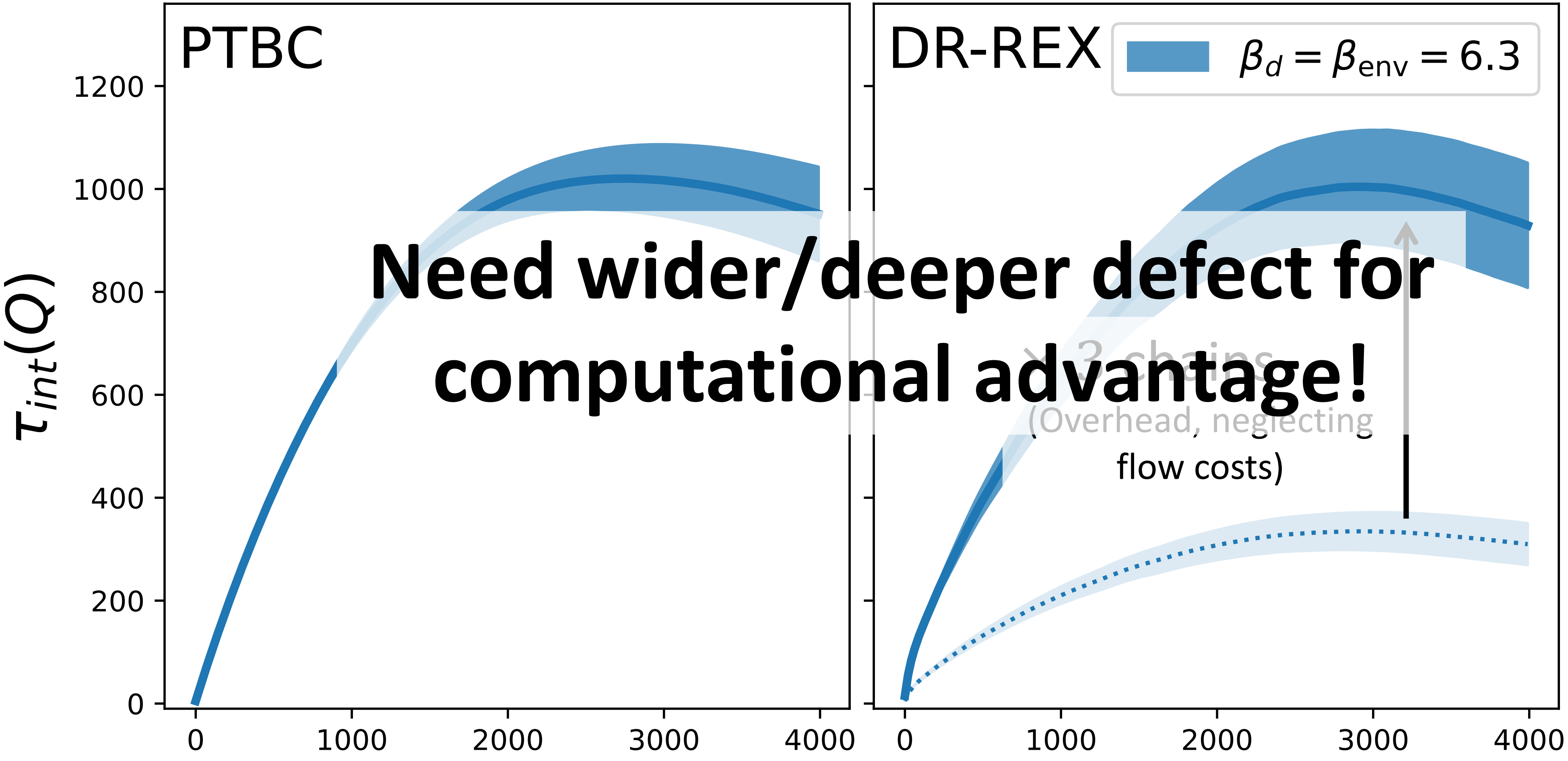


App 3: DR-REX Results

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

Two flows to repair a 2^3 OBC defect $\beta_d = 0 \to 3 \to 6.3$

Flows act on 8^4 subvolume



Need wider/deeper defect for computational advantage!

Outlook

Great potential for ML acceleration of lattice QCD calcs

Opportunities for useful applications before full generative modeling

REX makes natural use of flows

- Limiting case: direct sampling
  - See also: CRAFT [Matthews Arbel Rezende Doucet 2201.13117] → SNFs [Nada M 14:30]

Correlated ensembles / T-REX generally applicable in MC calcs

Straightforward generalizations to (pseudo)fermions / QCD

- Flows for fermions [2106.05934] PFs for gauge fields [2207.08945] QCD [2207.08945]

PTBC → DR-REX: very promising