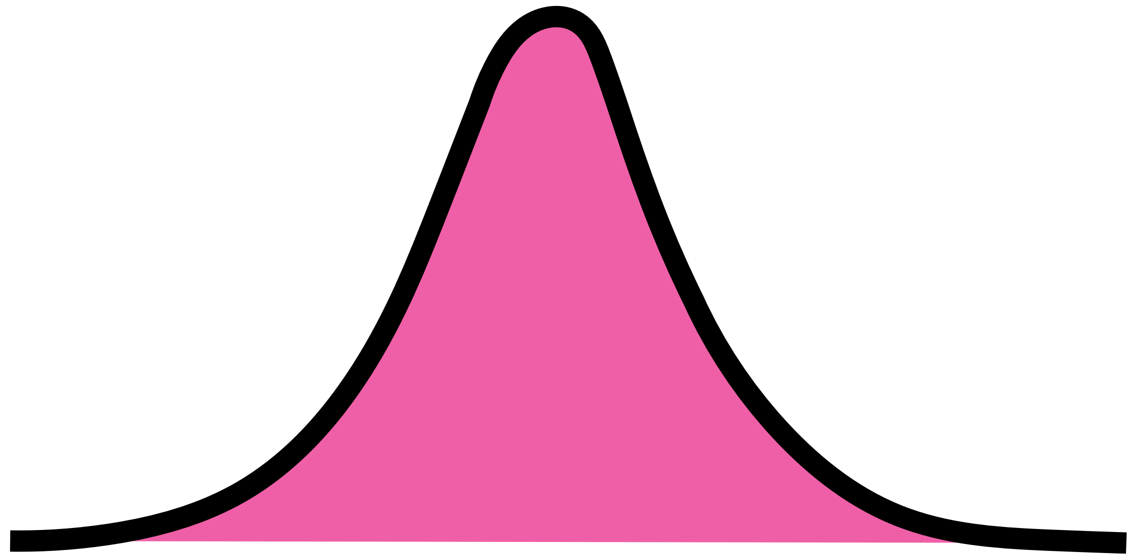
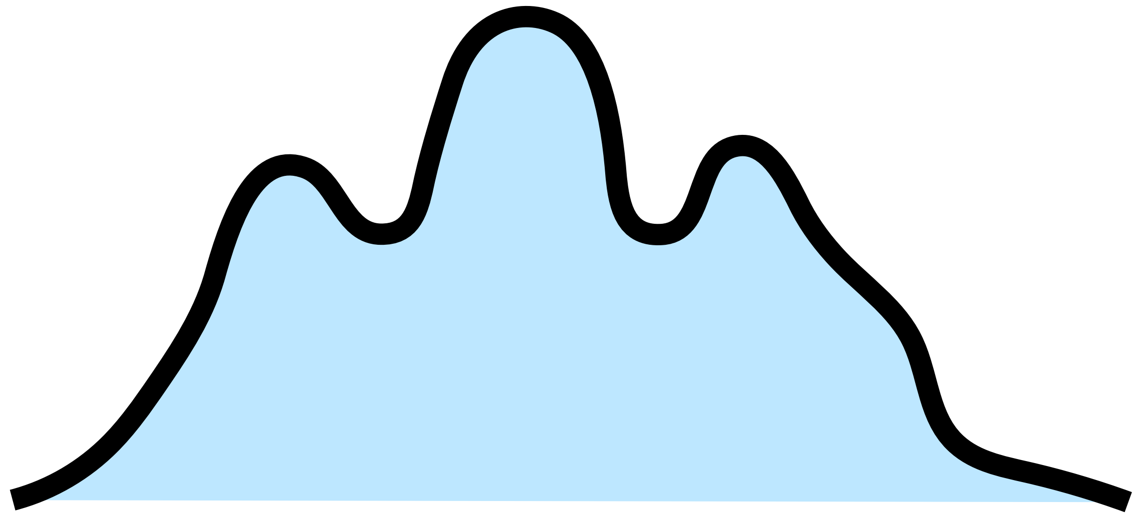
Applications of flow models to the generation of correlated lattice QCD ensembles

Fernando Romero-López

Lattice Field Theory and Machine Learning
NTU, Taipei…virtually :(

December 6th





fernando.romero-lopez@unibe.ch

Collaboration (non-exhaustive)

Massachusetts Institute of Technology

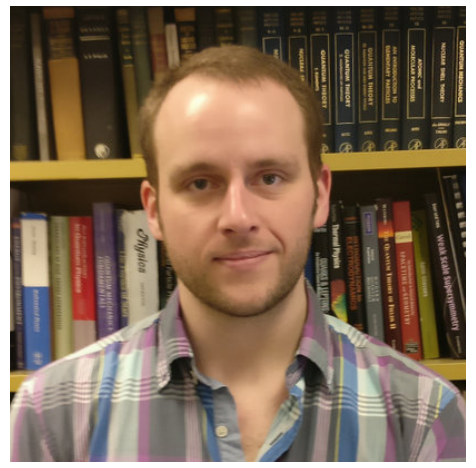
- Phiala Shanahan

- Ryan Abbot

- Julian Urban

- Denis Boyda

Fermilab



- Dan Hackett

- Kyle Cranmer

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

- Michael Albergo

UNIVERSITÄT BERN

- Fernando Romero-Lopez

DeepMind

- Sébastien Racanière

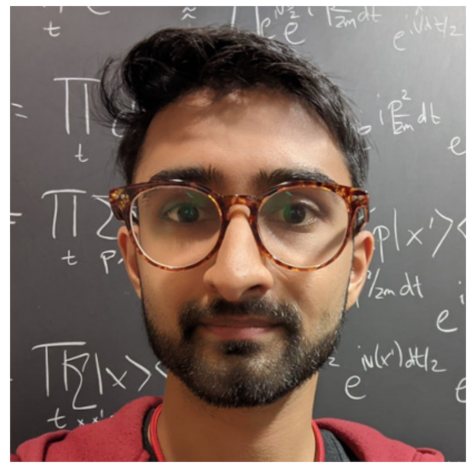
- Danilo Rezende

- Ali Razavi

- Aleksandar Botev

- Alex Matthews

THE UNIVERSITY OF EDINBURGH



- Gurtej Kanwar

Outline

1. Introduction
2. Flows for correlated ensembles
3. Numerical demonstrations
4. Hadron structure in dynamical QCD
5. Bonus: Transformed Replica Exchange (T-REX)
6. Conclusion

Introduction

Flows and lattice QCD

- Lattice Field Theory is a numerical first-principles treatment of the generic QFT
    - Significant progress in computing QCD observables at hadronic energies.

Fernando Romero-López, Uni Bern

Flows and Lattice QCD

- Lattice Field Theory is a numerical first-principles treatment of the generic QFT
  - Significant progress in computing QCD observables at hadronic energies.
- Lattice QCD can be formulated as a sampling problem:
  - Path integral in Euclidean or imaginary time: statistical meaning

$$\mathcal{Z} = \int D\phi \, e^{-S_E(\phi)}$$, where $$\underbrace{S_E(\phi) = \int d^4x \, \mathscr{L}_E(\phi)}_{\text{Euclidean action}}$$

Flows and Lattice QCD

- Lattice Field Theory is a numerical first-principles treatment of the generic QFT
  - Significant progress in computing QCD observables at hadronic energies.
- Lattice QCD can be formulated as a sampling problem:
  - Path integral in Euclidean or imaginary time: statistical meaning

$$\mathcal{Z} = \int D\phi \, e^{-S_E(\phi)} \quad \text{, where} \quad \underbrace{S_E(\phi) = \int d^4x \, \mathcal{L}_E(\phi)}_{\text{Euclidean action}}$$

- Increasing interest in applying generative flow models to LQCD
  - Can flow models reduce computational costs?

Fernando Romero-López, Uni Bern

The main problem

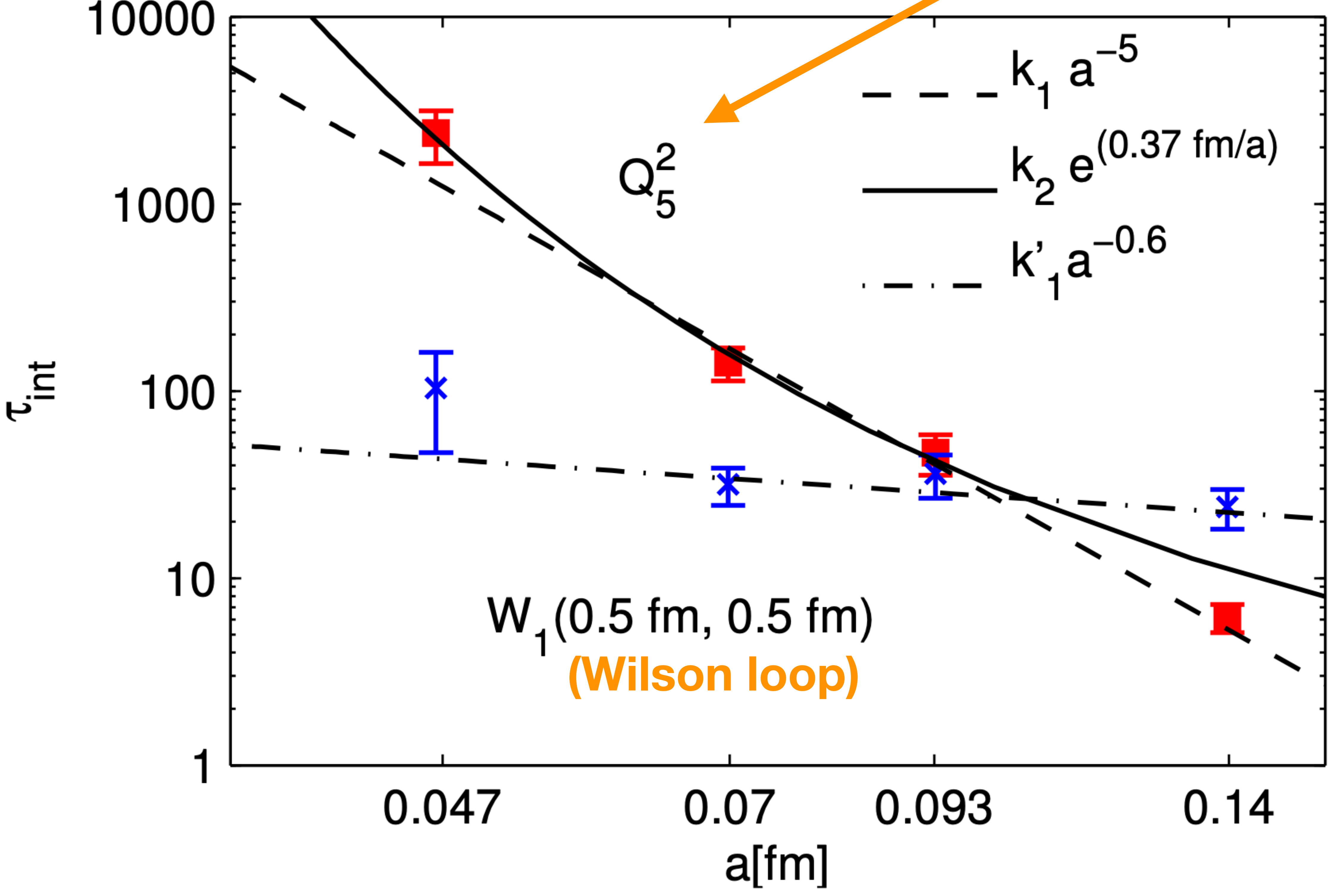
Topological charge

[Schaefer et al., arXiv:1009.5228]

autocorrelation time

$\simeq$

computational cost



Continuum limit

Computational cost of generating independent samples "explodes" towards the continuum limit:
Critical Slowing Down

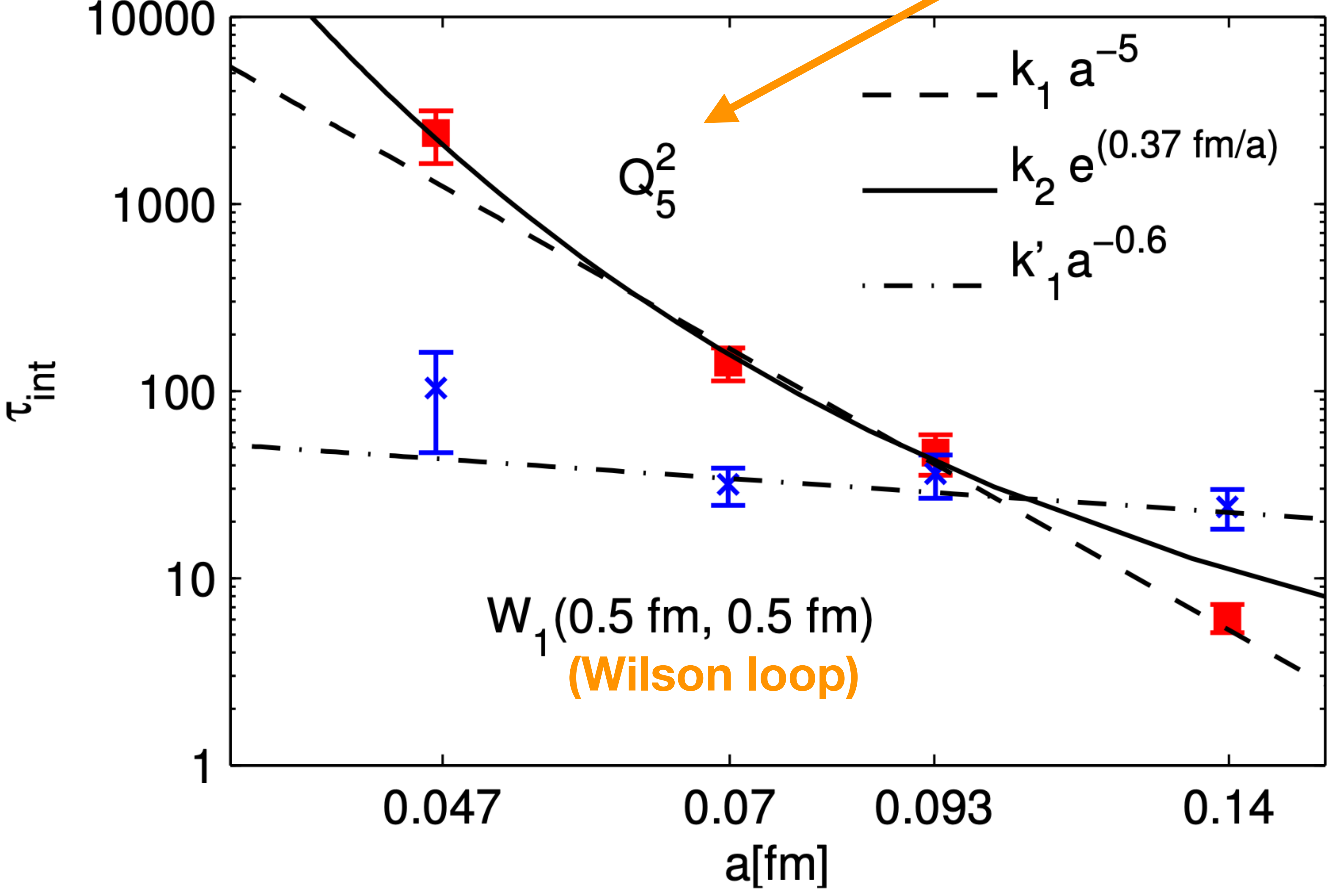
Fernando Romero-López, Uni Bern

The main problem

Topological charge

[Schaefer et al., arXiv:1009.5228]

autocorrelation time
$\simeq$
computational cost



(Wilson loop)

Continuum limit

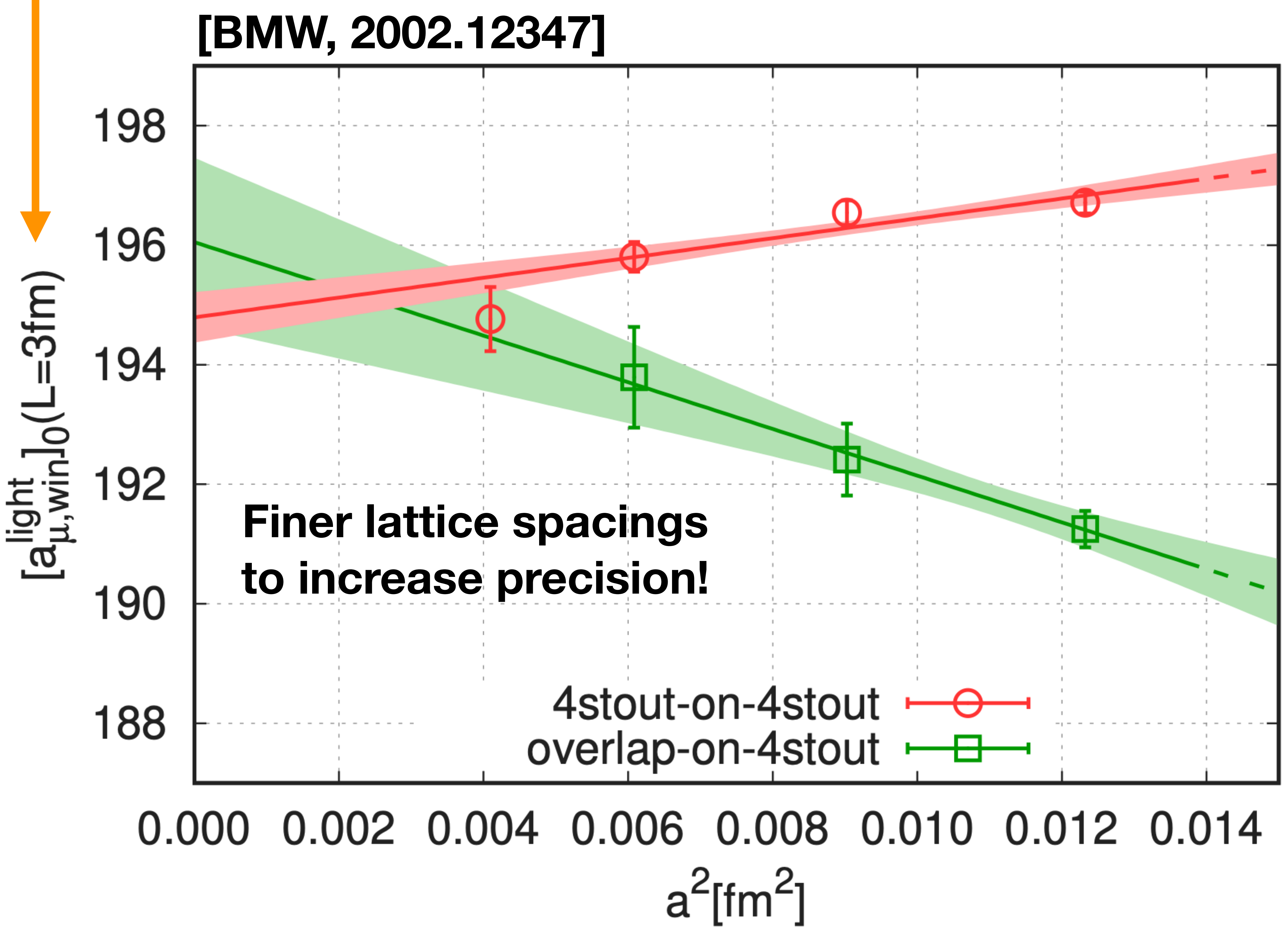
Computational cost of generating independent samples "explodes" towards the continuum limit:
Critical Slowing Down

Can flows help?

Fernando Romero-López, Uni Bern

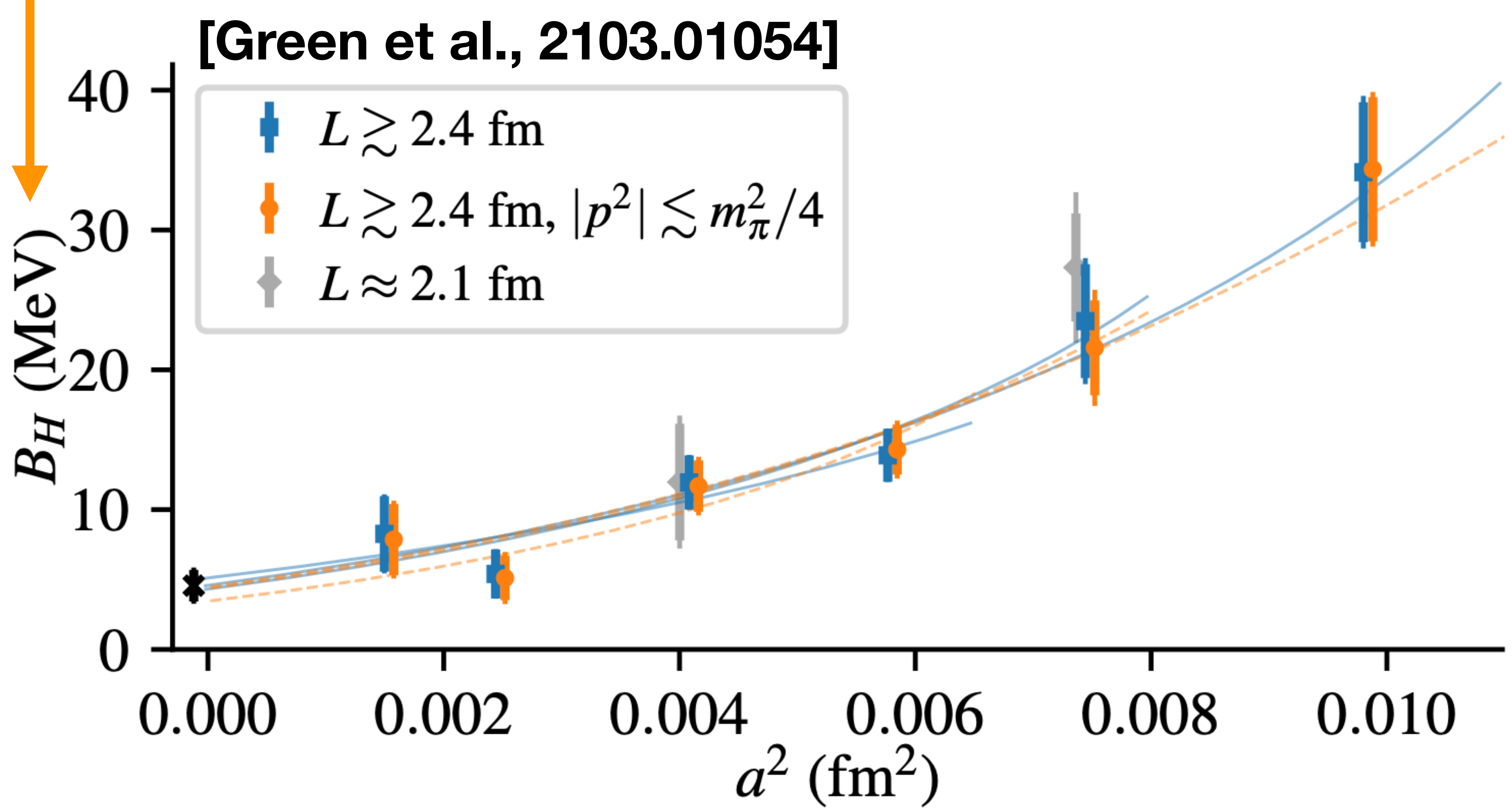
The need for a continuum limit

HVP of muon magnetic moment



Continuum limit

Binding energy of H dibaryon



Continuum limit

Where do we stand?

(in our collaboration)

(1+1)d real scalar field theory

[Albergo, Kanwar, Shanahan 1904.12072]
[Hackett, Hsieh, Albergo, Boyda, JW Chen, KF Chen, Cranmer, Kanwar, Shananan 2107.00734]

(1+1)d Abelian gauge theory

[Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan 2003.06413]

(1+1)d non-Abelian gauge theory

[Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan 2003.06413]

(1+1)d Yukawa model

i.e. real scalar field theory + fermions

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

Schwinger model i.e. (1+1)d QED

[Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Urban 2202.11712]

2D fermionic gauge theories with pseudofermions

[Abbott, Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Tian, Urban 2207.08945]

QCD/SU(3) in the strong-coupling region

[Abbott et al, 2208.03832] [Abbott et al, 2305.02402]

…

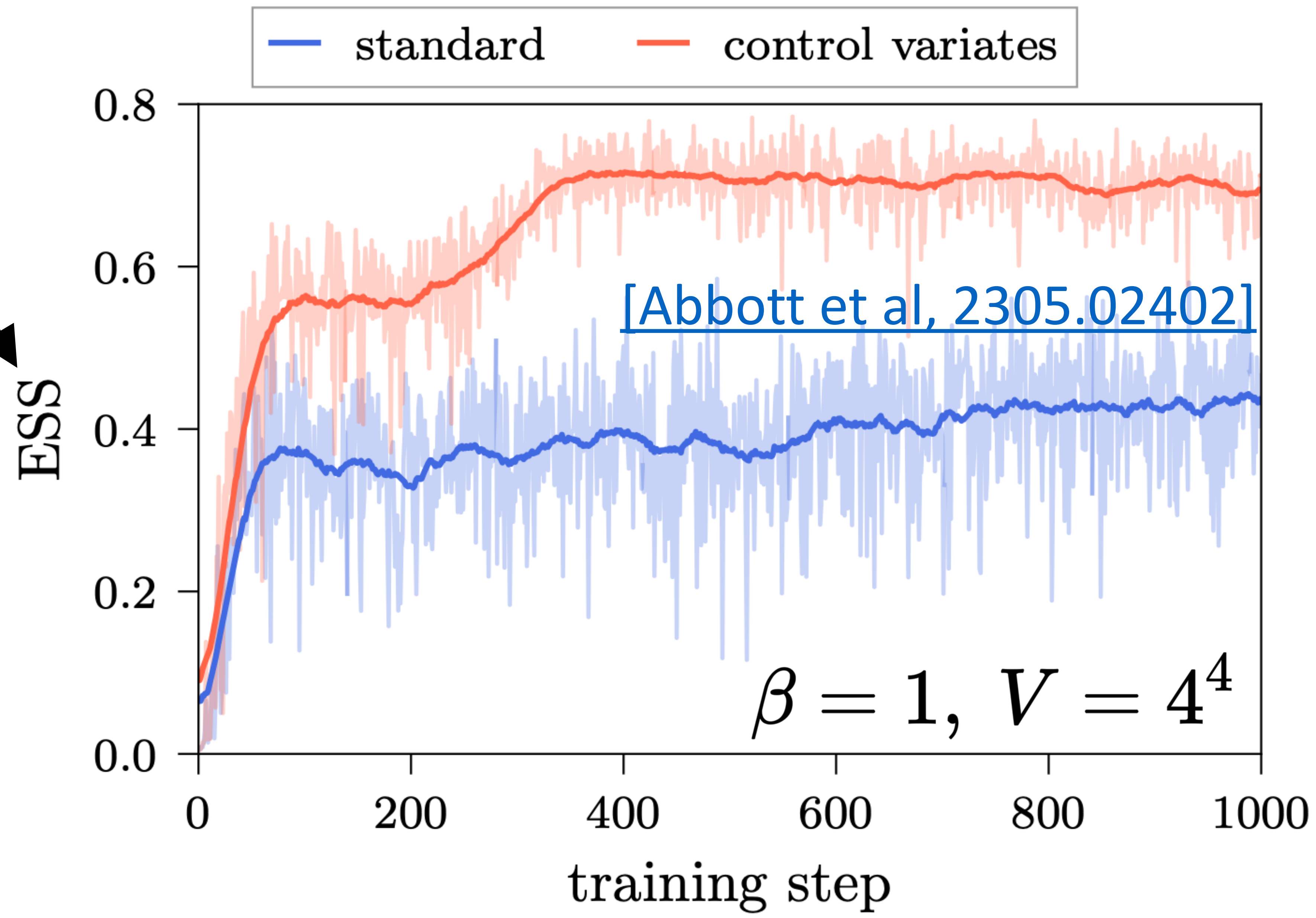
Still some developments are needed for at-scale QCD

Fernando Romero-López, Uni Bern

Flows in 4D

- Already dealing with 4D gauge theories.
  - Direct sampling remains hard
  - Need very high-quality models to reach large volumes
    (Naive volume scaling is exponential)

model quality

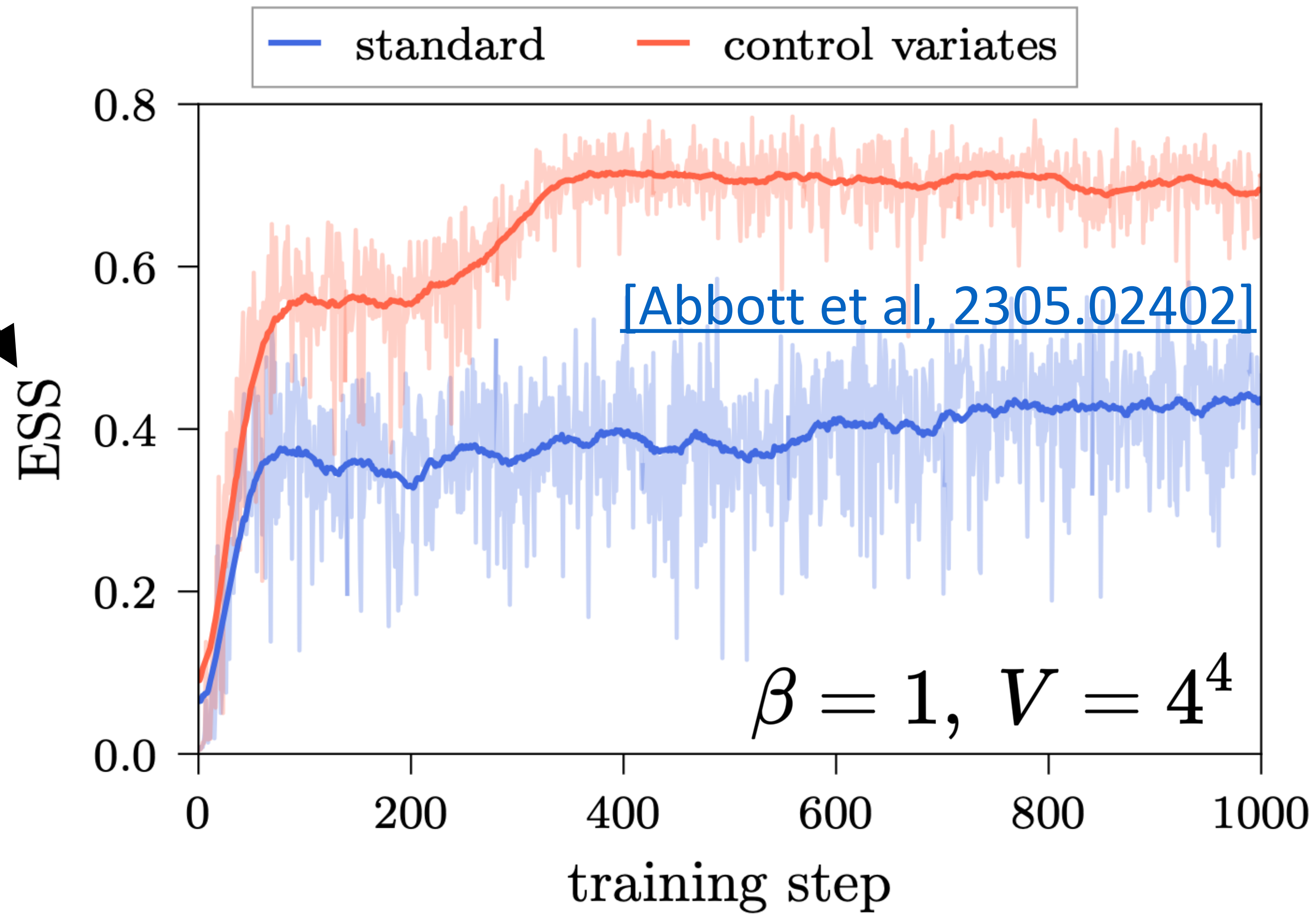


Fernando Romero-López, Uni Bern

Flows in 4D

- Already dealing with 4D gauge theories.
  - Direct sampling remains hard
  - Need very high-quality models to reach large volumes
    (Naive volume scaling is exponential)

model quality



- Instead, explore applications with smaller gap between theories: $\beta \longrightarrow \beta + \Delta\beta$
  - Can be useful for observable evaluation (and potentially sampling)

Fernando Romero-López, Uni Bern

Flows for the generation of correlated ensembles

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

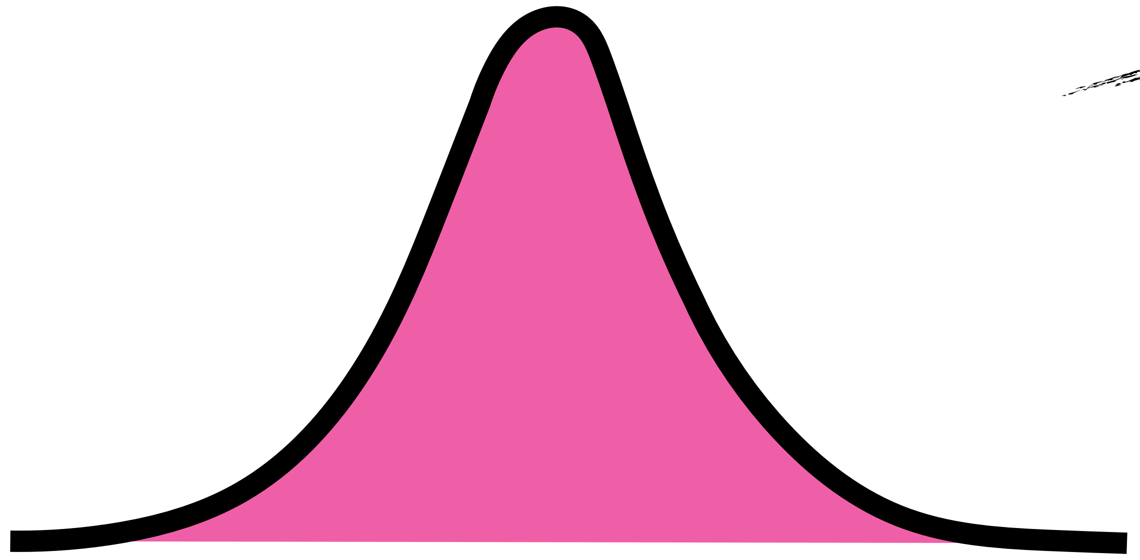
Ryan Abbott,[1,2] Aleksandar Botev,[3] Denis Boyda,[1,2] Daniel C. Hackett,[4,1,2] Gurtej Kanwar,[5] Sébastien Racanière,[3]
Danilo J. Rezende,[3] Fernando Romero-López,[1,2] Phiala E. Shanahan,[1,2] and Julian M. Urban[1,2]

[arXiv:2401.10874]

Generative flow models

[Rezende, Mohamed, 1505.05770]

$r(z)$

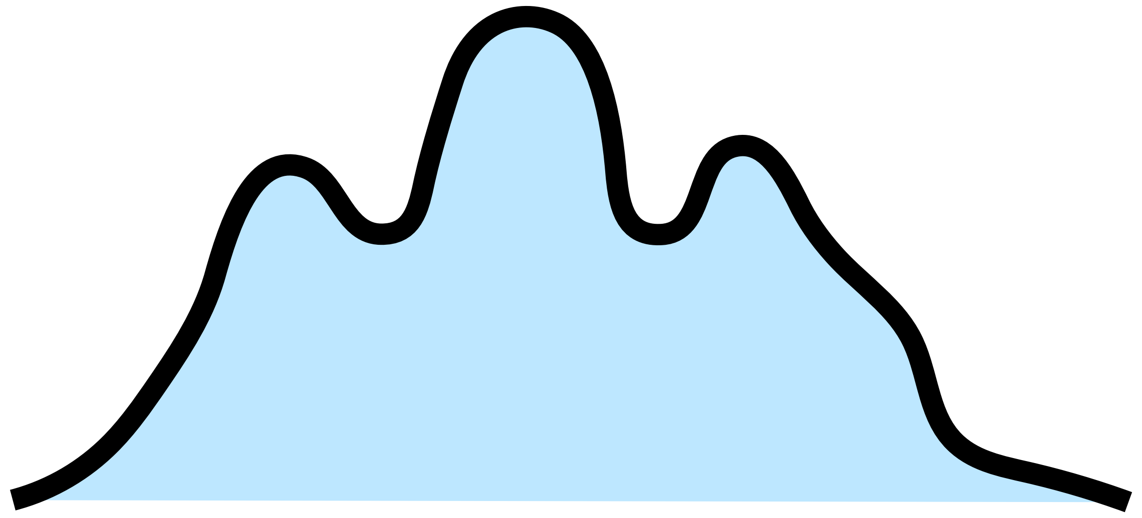


easy-to-sample
distribution
(prior)

$$\xrightarrow{f(z)}$$

tractable Jacobian
&
invertible

$q(\phi) \simeq p(\phi)$



approximates
target distribution
(model)

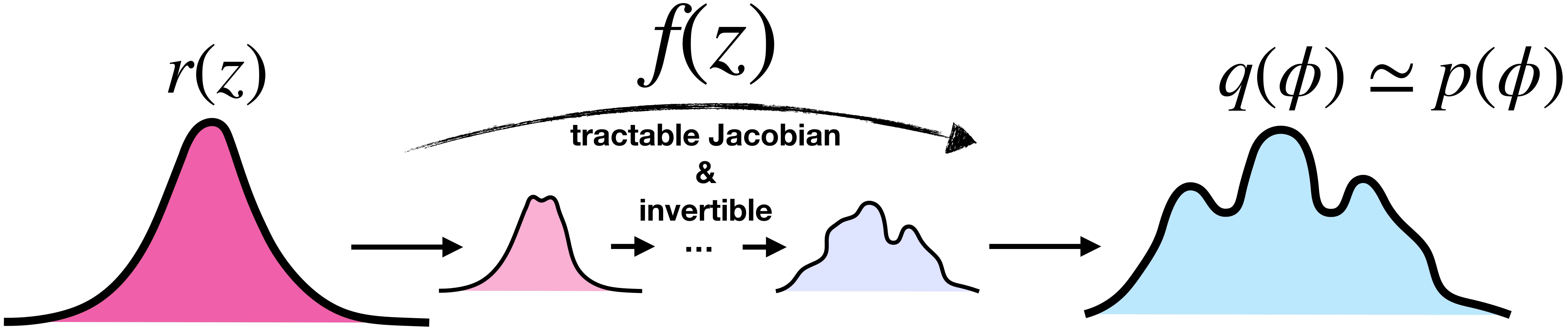
"Trainable change of variables"

Model probability

$$q(\phi) = r(z) \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$$

Generative flow models

[Rezende, Mohamed, 1505.05770]



easy-to-sample distribution (prior)

approximates target distribution (model)

parametrized by neural networks (trainable and expressive) → **composed of many simple layers**

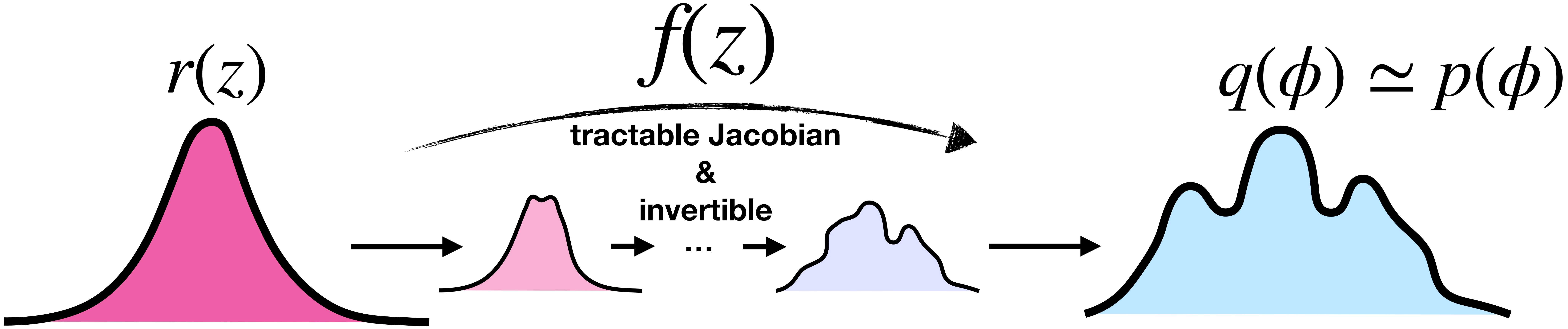
"Trainable change of variables"

Model probability

$$q(\phi) = r(z) \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$$

Generative flow models

[Rezende, Mohamed, 1505.05770]



easy-to-sample
distribution
(prior)

approximates
target distribution
(model)

parametrized by neural networks
(trainable and expressive)

composed of
many simple layers

"Trainable change of variables"

Model probability

$$q(\phi) = r(z)\left|\det \frac{\partial f(z)}{\partial z}\right|^{-1}$$

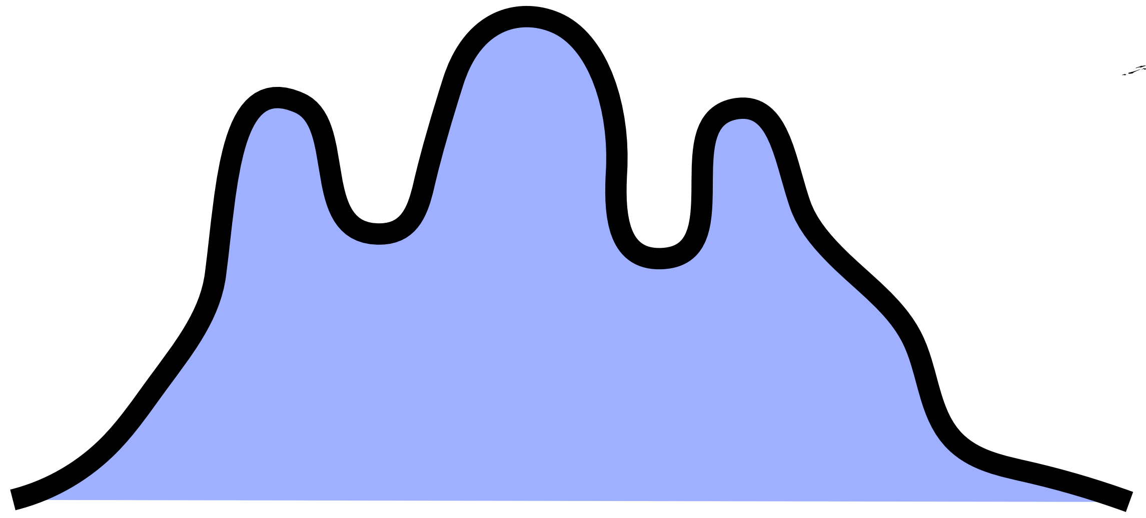
❗ Trained models are not perfect.

☑ But exact sampling can be recovered via Markov Chain

Fernando Romero-López, Uni Bern

Flows for correlated ensembles

$p_\beta(z)$

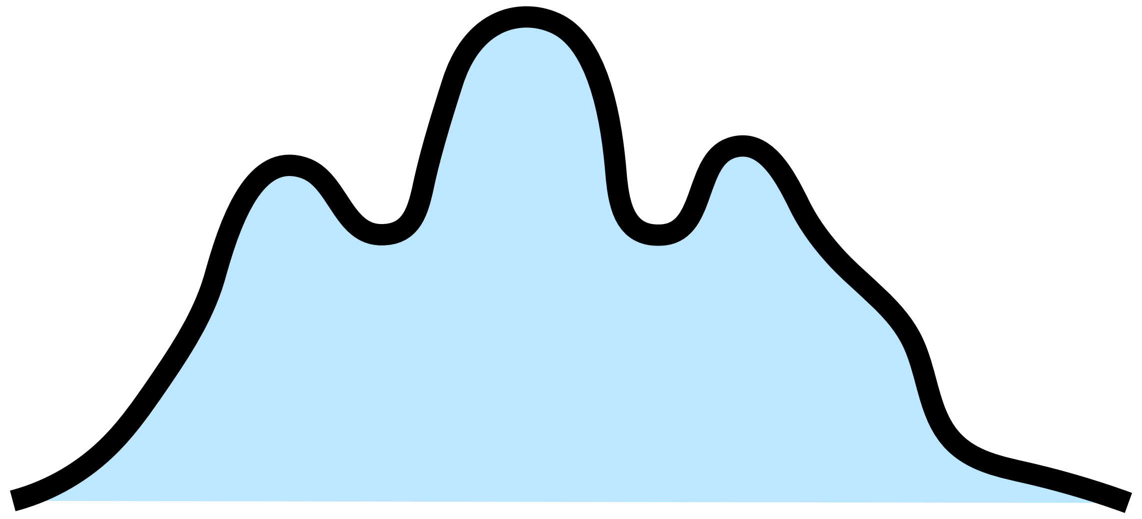


Non-trivial prior
sampled via MCMC

$f(z)$

tractable Jacobian
&
invertible

$q(\phi) \simeq p_{\beta+\Delta\beta}(\phi)$



approximates
target distribution
(model)

- ☑ Theories are "closer" and current flow models are able to effectively bridge between them

$$\mathrm{ESS} \simeq x^{\Delta\beta}$$

Derivative observables

- In lattice QCD there are many examples where derivatives with respect to action parameters are useful

$$\frac{d\langle \mathcal{O} \rangle}{d\alpha} \simeq \frac{\langle \mathcal{O} \rangle_{\alpha_1} - \langle \mathcal{O} \rangle_{\alpha_2}}{\Delta\alpha}$$

action parameter

Derivative observables

- In lattice QCD there are many examples where derivatives with respect to action parameters are useful

$$\frac{d\langle \mathcal{O} \rangle}{d\alpha} \simeq \frac{\langle \mathcal{O} \rangle_{\alpha_1} - \langle \mathcal{O} \rangle_{\alpha_2}}{\Delta \alpha}$$

action parameter

- Continuum limit, e.g., constraining the slope of a continuum extrapolation
- Matrix element using Feynman-Hellmann techniques: sigma terms, hadron structure.
- QCD + QED, e.g., derivative with respect to electromagnetic coupling
- Derivatives with respect to chemical potential, theta term… (caveat: sign problem)

Computing derivatives

- **How to compute derivative observables?**

$$\frac{d\langle \mathcal{O}\rangle}{d\alpha} \simeq \frac{\langle \mathcal{O}\rangle_{\alpha_1} - \langle \mathcal{O}\rangle_{\alpha_2}}{\Delta\alpha}$$

action parameter

Computing derivatives

- **How to compute derivative observables?** $$\frac{d\langle\mathcal{O}\rangle}{d\alpha} \simeq \frac{\langle\mathcal{O}\rangle_{\alpha_1} - \langle\mathcal{O}\rangle_{\alpha_2}}{\Delta\alpha}$$ action parameter

1. Independent ensembles:

- **Compute $\langle\mathcal{O}\rangle_{\alpha_i}$ on independent Markov Chains.**

Computing derivatives

- How to compute derivative observables? $$\frac{d\langle\mathcal{O}\rangle}{d\alpha} \simeq \frac{\langle\mathcal{O}\rangle_{\alpha_1} - \langle\mathcal{O}\rangle_{\alpha_2}}{\Delta\alpha}$$ action parameter

1. Independent ensembles:

- Compute $\langle\mathcal{O}\rangle_{\alpha_i}$ on independent Markov Chains.

2. Epsilon-reweighting

- Compute difference on a single ensemble using reweighing at $\Delta\alpha = \epsilon$

$$\langle\mathcal{O}\rangle_{\alpha_1} - \langle\mathcal{O}\rangle_{\alpha_1+\epsilon} = \langle\mathcal{O} - w_\epsilon\mathcal{O}\rangle_{\alpha_1} \longleftarrow w_\epsilon = p_{\alpha_1+\epsilon}/p_{\alpha_1}$$

Computing derivatives

- How to compute derivative observables? $$\frac{d\langle\mathcal{O}\rangle}{d\alpha} \simeq \frac{\langle\mathcal{O}\rangle_{\alpha_1} - \langle\mathcal{O}\rangle_{\alpha_2}}{\Delta\alpha}$$ action parameter

1. Independent ensembles:

- Compute $\langle\mathcal{O}\rangle_{\alpha_i}$ on independent Markov Chains.

2. Epsilon-reweighting

- Compute difference on a single ensemble using reweighing at $\Delta\alpha = \epsilon$

$$\langle\mathcal{O}\rangle_{\alpha_1} - \langle\mathcal{O}\rangle_{\alpha_1+\epsilon} = \langle\mathcal{O} - w_\epsilon\mathcal{O}\rangle_{\alpha_1} \longleftarrow w_\epsilon = p_{\alpha_1+\epsilon}/p_{\alpha_1}$$

3. Using Flows

- Create a "correlated ensemble" using a flow and compute the difference

$$\langle\mathcal{O}(U) - w(f(U))\mathcal{O}(f(U))\rangle_{\alpha_1}$$ where $$w = p/q$$ $$w(f(U)) \simeq 1$$

[see also S. Bacchio, 2305.07932]

Computing derivatives

- How to compute derivative observables? $$\frac{d\langle\mathcal{O}\rangle}{d\alpha} \simeq \frac{\langle\mathcal{O}\rangle_{\alpha_1} - \langle\mathcal{O}\rangle_{\alpha_2}}{\Delta\alpha}$$ action parameter

1. Independent ensembles: $\langle\mathcal{O}\rangle_{\alpha_1} - \langle\mathcal{O}\rangle_{\alpha_2}$

- One can use large $\Delta\alpha$, at the cost of $O(\Delta\alpha)$ effects in the derivative
- Statistical errors add in quadrature: signal only visible at large $\Delta\alpha$

2. Epsilon-reweighting $\langle\mathcal{O}\rangle_{\alpha_1} - \langle\mathcal{O}\rangle_{\alpha_1+\epsilon} = \langle\mathcal{O} - w_\epsilon\mathcal{O}\rangle_{\alpha_1}$

- Uncertainties benefit from correlated cancellations
- Errors increase very rapidly with $\Delta\alpha = \epsilon$

3. Using Flows $\langle\mathcal{O}(U) - w(f(U))\mathcal{O}(f(U))\rangle_{\alpha_1}$

- Uncertainties benefit from correlated cancellations
- Can go to larger $\Delta\alpha$ than with ϵ reweighing

Fernando Romero-López, Uni Bern

Computing derivatives

- How to compute derivative observables? $$\frac{d\langle\mathcal{O}\rangle}{d\alpha} \simeq \frac{\langle\mathcal{O}\rangle_{\alpha_1} - \langle\mathcal{O}\rangle_{\alpha_2}}{\Delta\alpha}$$ ← action parameter

1. **Independent ensembles:** $\langle\mathcal{O}\rangle_{\alpha_1} - \langle\mathcal{O}\rangle_{\alpha_2}$
   - One can use large $\Delta\alpha$, at the cost of $O(\Delta\alpha)$ effects in the derivative
   - Statistical errors add in quadrature: signal only visible at large $\Delta\alpha$

2. **Epsilon-reweighting** $\langle\mathcal{O}\rangle_{\alpha_1} - \langle\mathcal{O}\rangle_{\alpha_1+\epsilon} = \langle\mathcal{O} - w_\epsilon\mathcal{O}\rangle_{\alpha_1}$
   - Uncertainties benefit from correlated cancellations
   - Errors increase very rapidly with $\Delta\alpha = \epsilon$

3. **Using Flows** $\langle\mathcal{O}(U) - w(f(U))\mathcal{O}(f(U))\rangle_{\alpha_1}$
   - Uncertainties benefit from correlated cancellations
   - Can go to larger $\Delta\alpha$ than with ϵ reweighing

→ **Best of both worlds!**

Fernando Romero-López, Uni Bern

The architecture

- Use an equivariant flow architecture based on the Gradient Flow "Residual layers"

[Abbott et al, 2305.02402]
See also: [Bacchio et al, 2212.08469]
[Gerdes et al, 2410.13161]
[Nagai, Tomiya, 2103.11965]

$$\boxed{U'_\mu(x) = e^{F(U)} U_\mu(x)}$$

$$F = \sum_i \delta_i \, \mathrm{P}\left(W^i{}_{\mu\nu}\right)$$

Untraced Wilson loops that start and end at x

trainable

Traceless-antihermitian projection

The architecture

- Use an equivariant flow architecture based on the Gradient Flow "Residual layers"

[Abbott et al, 2305.02402]
See also: [Bacchio et al, 2212.08469]
[Gerdes et al, 2410.13161]
[Nagai, Tomiya, 2103.11965]

$$\boxed{U'_\mu(x) = e^{F(U)} U_\mu(x)}$$

$$F = \sum_i \delta_i \, \mathrm{P}\left(W^i{}_{\mu\nu}\right)$$

Untraced Wilson loops that start and end at x

trainable

Traceless-antihermitian projection

- Split lattice in active + frozen variables, and update only active (upper triangular Jacobian)

Fernando Romero-López, Uni Bern

The architecture

- Use an equivariant flow architecture based on the Gradient Flow

"Residual layers"

[Abbott et al, 2305.02402]
See also: [Bacchio et al, 2212.08469]
[Gerdes et al, 2410.13161]
[Nagai, Tomiya, 2103.11965]

$$\boxed{U'_\mu(x) = e^{F(U)} U_\mu(x)}$$

$$F = \sum_i \delta_i \, \mathrm{P}\left(W^i{}_{\mu\nu}\right)$$

Untraced Wilson loops that start and end at x

trainable

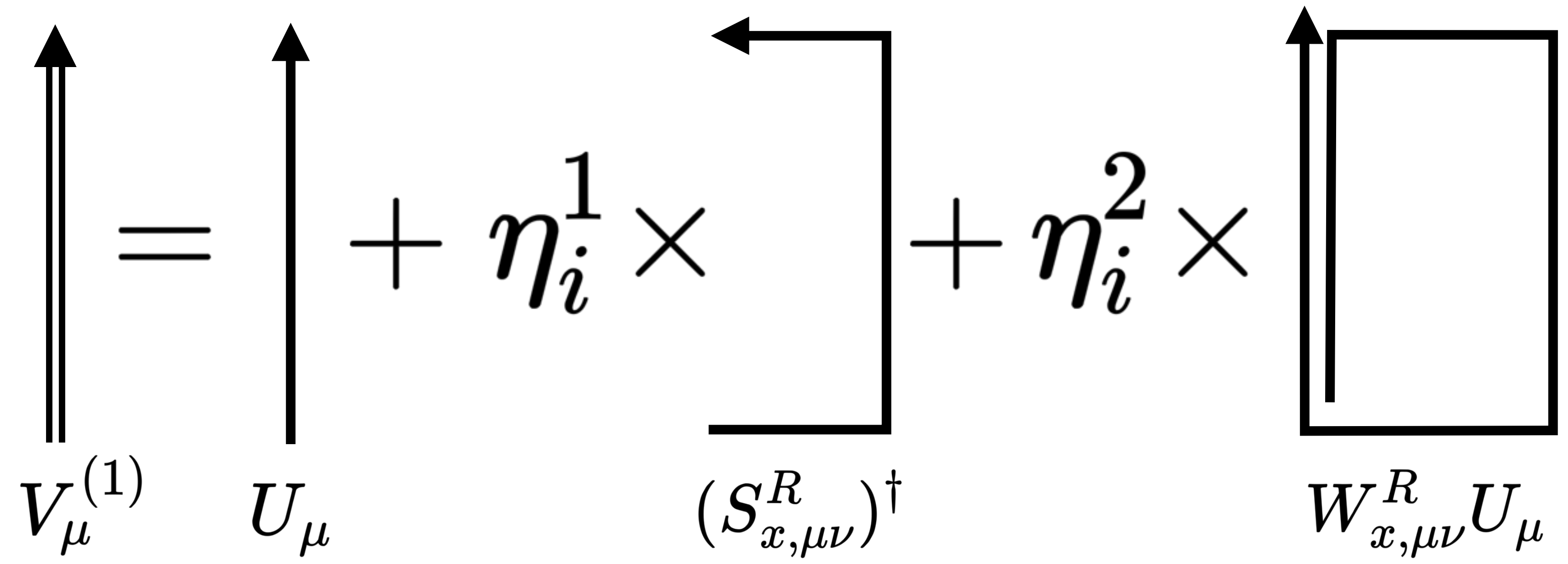
Traceless-antihermitian projection

- Split lattice in active + frozen variables, and update only active (upper triangular Jacobian)

- Build arbitrary loops "convoluting" the frozen links

$$F(V)$$

Force built from convoluted links



[Similar to L-CNN, Favoni et al, 2012.12901]

Connection to gradient flow

- **Gradient flow minimizes the action on a gauge configuration by solving the differential equation**
  [M. Lüscher, arXiv:1006.4518]

$$\dot{B}_\mu = -\frac{\delta S(U)}{dB_\mu} \xrightarrow{\text{solve numerically with infinitesimal steps}} U_\mu^{(i+1)} = e^{\epsilon F(U^{(i)})} U_\mu^{(i)}$$

Connection to gradient flow

- Gradient flow minimizes the action on a gauge configuration by solving the differential equation
  [M. Lüscher, arXiv:1006.4518]

$$\dot{B}_\mu = -\frac{\delta S(U)}{dB_\mu} \xrightarrow{\text{solve numerically with infinitesimal steps}} U_\mu^{(i+1)} = e^{\epsilon F(U^{(i)})} U_\mu^{(i)}$$

- The residual layers act as a single and finite step of "generalized" gradient flow

$$\boxed{U'_\mu(x) = e^{F(U)} U_\mu(x)}$$

$$F = \epsilon \times \sum_{\mu \neq \nu} \mathrm{P}(W_{\mu\nu}^{1\times 1})$$

Connection to gradient flow

- Gradient flow minimizes the action on a gauge configuration by solving the differential equation
  [M. Lüscher, arXiv:1006.4518]

$$\dot{B}_\mu = -\frac{\delta S(U)}{dB_\mu} \xrightarrow{\hspace{3em}} U_\mu^{(i+1)} = e^{\epsilon F(U^{(i)})} U_\mu^{(i)}$$

solve numerically with infinitesimal steps

- The residual layers act as a single and finite step of "generalized" gradient flow

$$\boxed{U'_\mu(x) = e^{F(U)} U_\mu(x)} \qquad F = \epsilon \times \sum_{\mu \neq \nu} \mathrm{P}(W_{\mu\nu}^{1\times 1})$$

- Close to the original M. Lüscher trivializing map proposal [M. Lüscher, arXiv:0907.5491]

$$\log J = -\frac{4}{3}\epsilon \sum_{\mu\neq\nu} \mathrm{tr}\, W_{\mu\nu}^{1\times 1} + O(\epsilon^2) \xrightarrow{\hspace{3em}}$$

Qualitatively induces a change in the lattice spacing

$$\beta \longrightarrow \beta + \Delta\beta$$

Fernando Romero-López, Uni Bern

Numerical demonstrations

Model	Prior type	Parameters	Target type	Parameters	Train ESS	Eval. vol.	ESS
A	Pure Gauge SU(3)	$\beta = 6.02$	Pure Gauge SU(3)	$\beta = 6.03$	99.72%	16^4	67%
B1	Pure Gauge SU(3)	$\beta = 6.00$	Feynman-Hellmann	$\beta = 6.00, \lambda = +0.01$	99.4%	16×8^3	84%
B2	Pure Gauge SU(3)	$\beta = 6.00$	Feynman-Hellmann	$\beta = 6.00, \lambda = -0.01$	99.4%	16×8^3	84%
C	$N_f = 2$ QCD	$\beta = 5.60, \kappa = 0.153$	$N_f = 2$ QCD	$\beta = 5.60, \kappa = 0.1545$	99.2%	8^4	48%

Continuum Limit

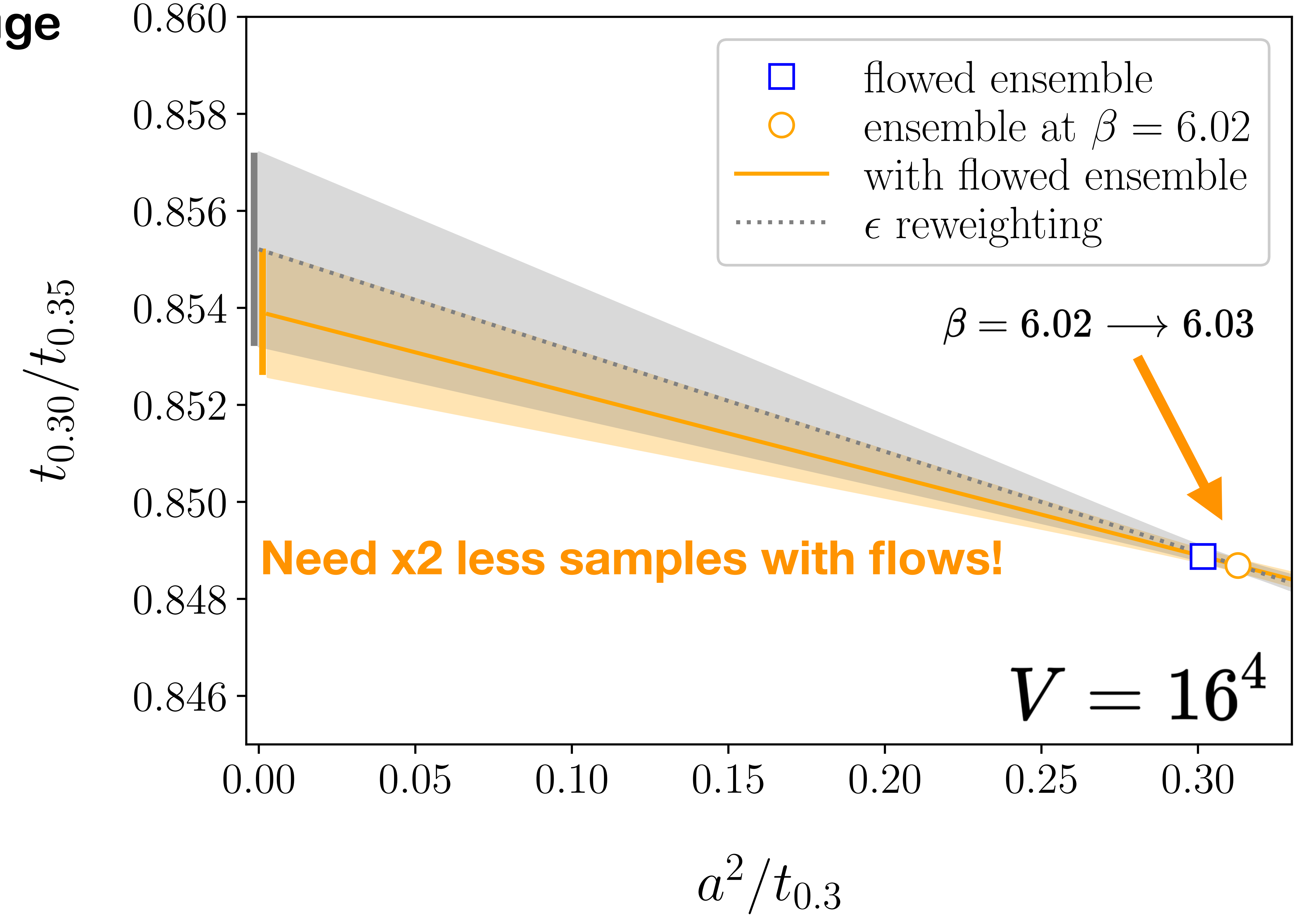
- Derivative of an observable with respect to lattice spacing is useful in constraining the continuum limit

- **Example: gradient flow scales in SU(3) pure gauge**

$$k_1 = \frac{d(t_{0.3}/t_{0.35})}{d(a^2/t_{0.3})}$$

- **Extrapolate to the continuum as:**

$$\left.\frac{t_{0.3}}{t_{0.35}}\right|_{\mathrm{lat}} = \left.\frac{t_{0.3}}{t_{0.35}}\right|_{\mathrm{cont}} + k_1 \frac{a^2}{t_{0.3}} + \cdots$$



Hadron structure

- **Computation of hadronic matrix elements can be formulated as a derivative**

$$S_\lambda = S + \lambda \mathcal{O} \longrightarrow \langle \pi | \mathcal{O} | \pi \rangle = \frac{1}{2M_\pi} \frac{dM_\pi}{d\lambda}\Big|_{\lambda=0}$$

"Feynman-Hellmann theorem"

Fernando Romero-López, Uni Bern

Hadron structure

- Computation of hadronic matrix elements can be formulated as a derivative

$$S_\lambda = S + \lambda \mathcal{O} \longrightarrow \langle \pi | \mathcal{O} | \pi \rangle = \frac{1}{2M_\pi} \frac{dM_\pi}{d\lambda}\Big|_{\lambda=0}$$

"Feynman-Hellmann theorem"

- If the operator is the gluon energy-momentum tensor, it leads to the gluon momentum fraction

$$\mathcal{O} = -\frac{\beta}{N_c} \mathrm{Tr}\,\mathrm{Re}\left(\sum_i U_{i0} - \sum_{i<j} U_{ij} \right) \longrightarrow \frac{dM_\pi}{d\lambda} = -\frac{3M_\pi}{2} \langle x \rangle_g^{\mathrm{latt}}$$

Fernando Romero-López, Uni Bern

Hadron structure

- Computation of hadronic matrix elements can be formulated as a derivative

$$S_\lambda = S + \lambda \mathcal{O} \longrightarrow \langle \pi | \mathcal{O} | \pi \rangle = \frac{1}{2M_\pi} \frac{dM_\pi}{d\lambda}\bigg|_{\lambda=0}$$

"Feynman-Hellmann theorem"

- If the operator is the gluon energy-momentum tensor, it leads to the gluon momentum fraction

$$\mathcal{O} = -\frac{\beta}{N_c} \mathrm{Tr}\,\mathrm{Re}\left(\sum_i U_{i0} - \sum_{i<j} U_{ij} \right) \longrightarrow \frac{dM_\pi}{d\lambda} = -\frac{3M_\pi}{2} \langle x \rangle_g^{\mathrm{latt}}$$

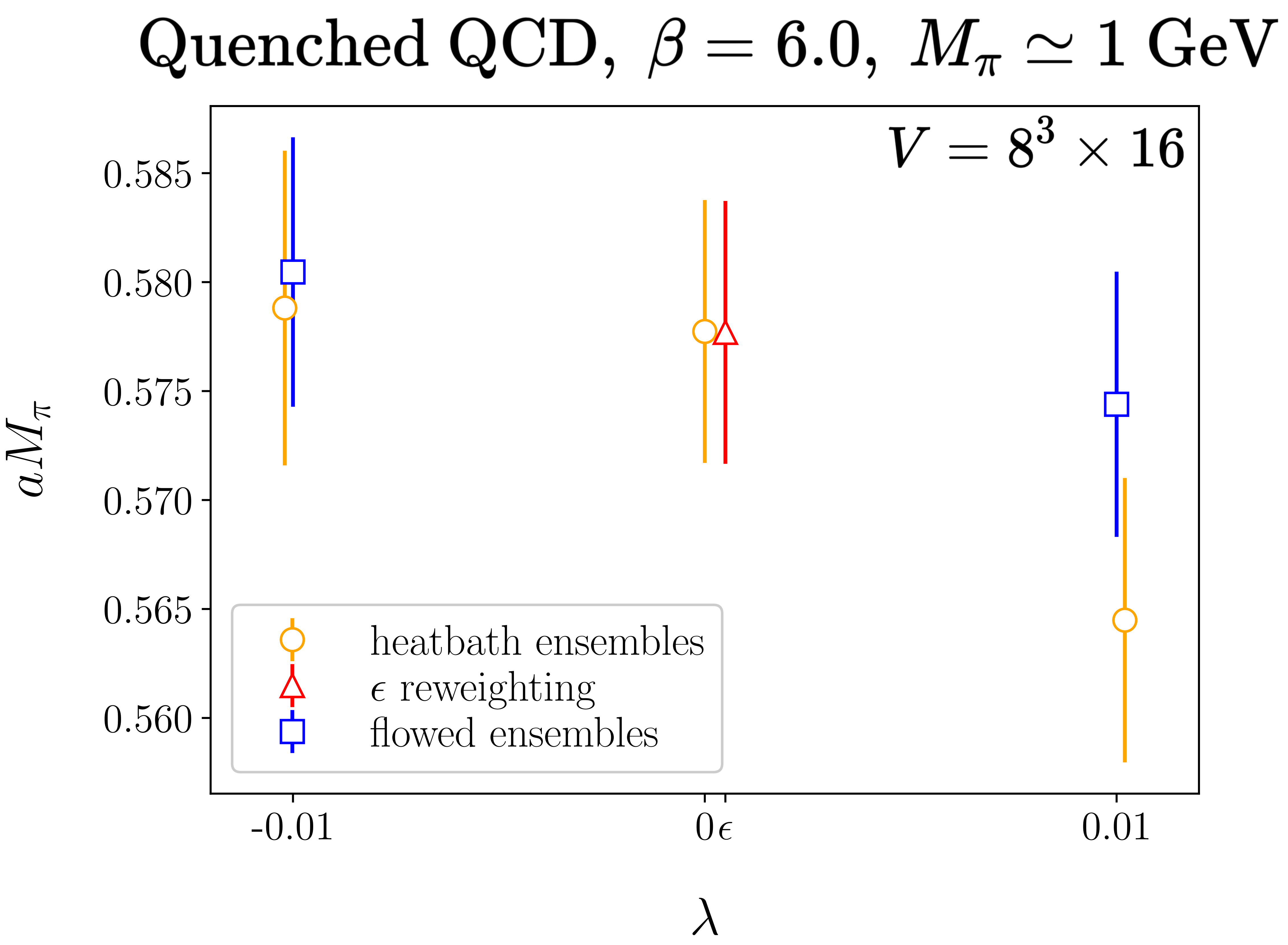
- The gauge action becomes just an anisotropic target!

$$\boxed{S_\lambda = -\frac{\beta}{N_c}(1+\lambda)\,\mathrm{Re}\,\mathrm{Tr} \sum_i U_{i0} - \frac{\beta}{N_c}(1-\lambda)\,\mathrm{Re}\,\mathrm{Tr} \sum_{i<j} U_{ij}} \longrightarrow$$

Train from from
$\lambda = 0$ to non-zero λ

Feynman-Hellmann results

- **Results in quenched QCD using central finite differences for derivatives**

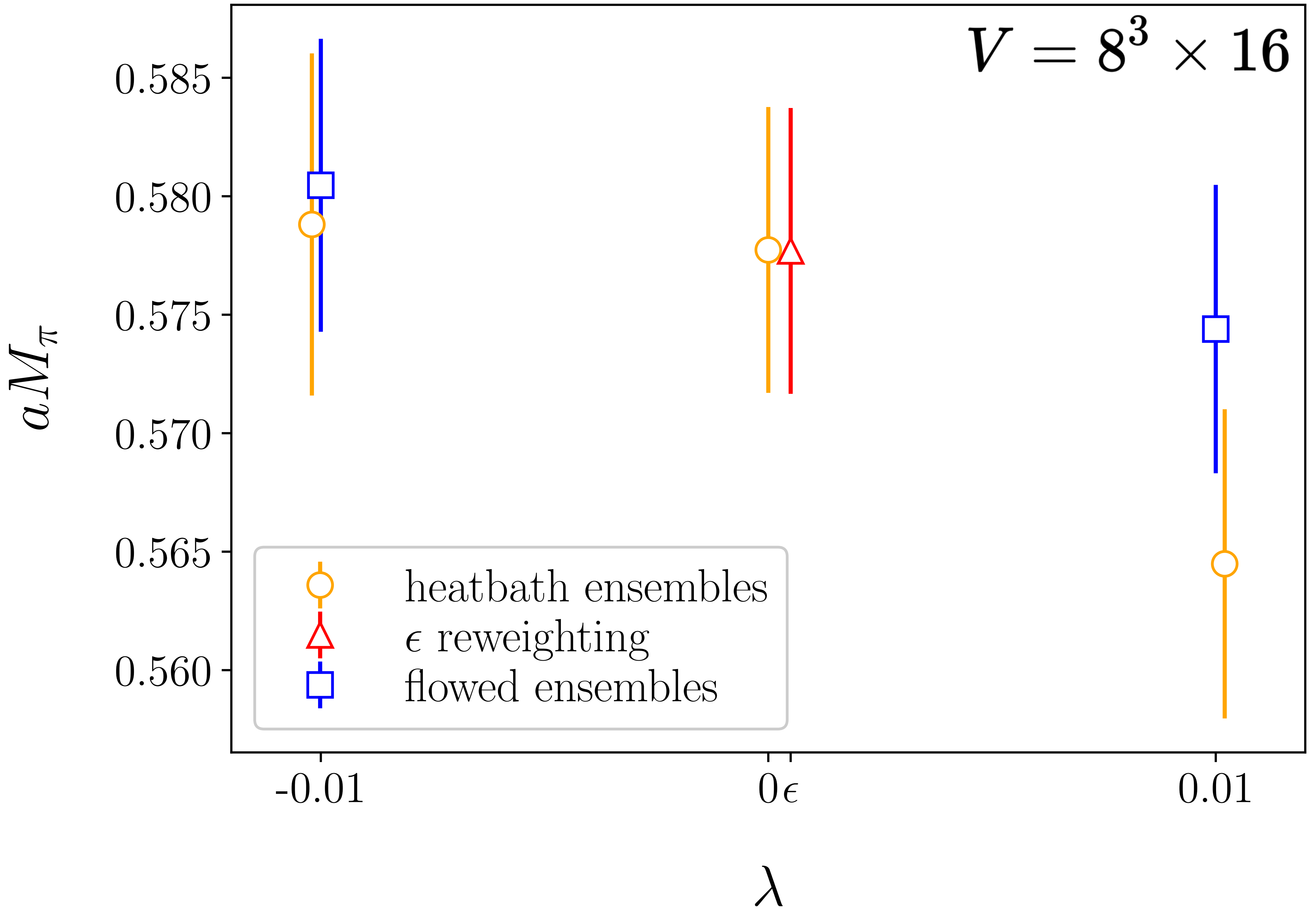


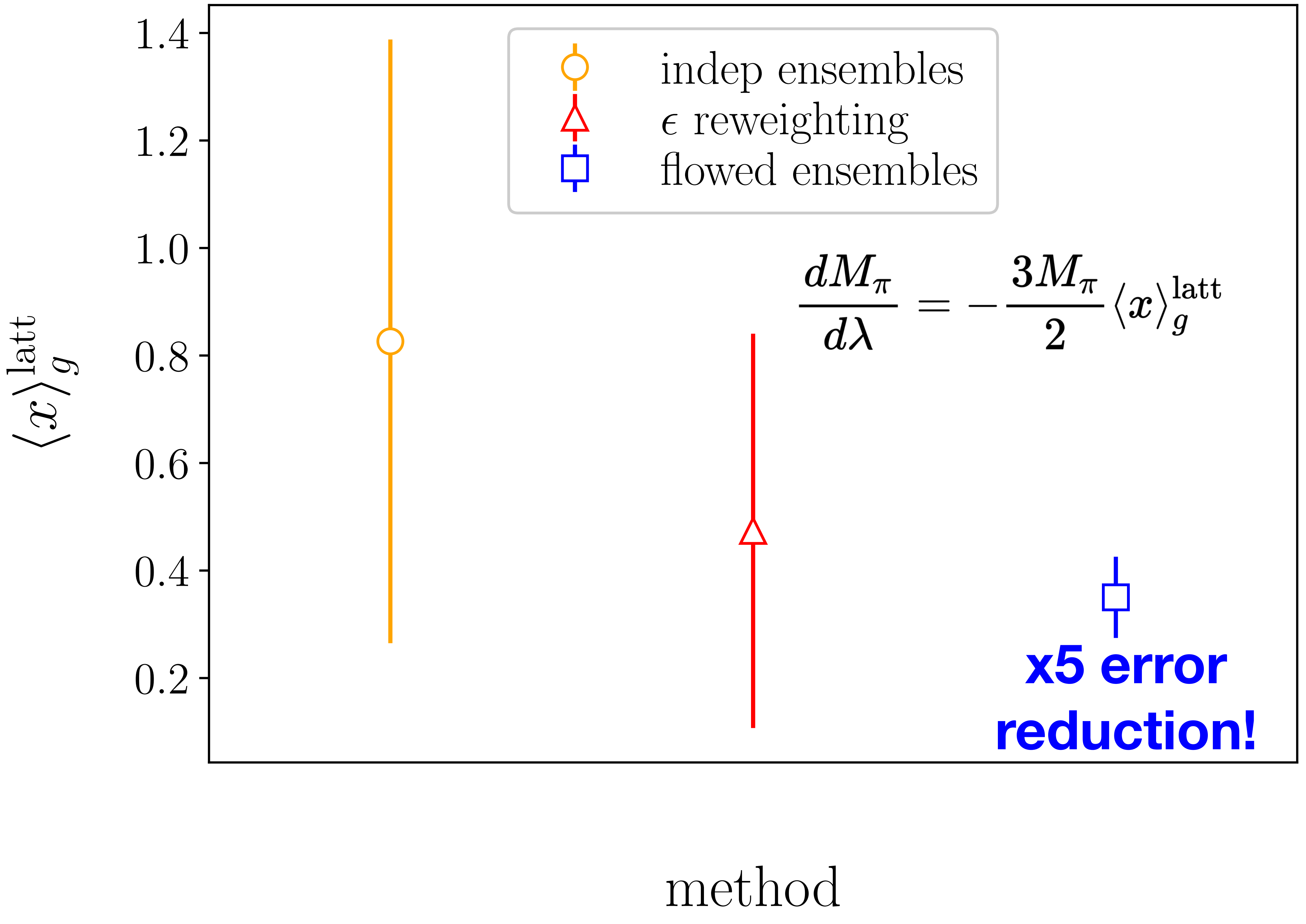
Same setup as [QCDSF, arXiv:1205.6410]

Feynman-Hellmann results

- **Results in quenched QCD using central finite differences for derivatives**

Quenched QCD, $\beta = 6.0,\ M_\pi \simeq 1$ GeV





Same setup as [QCDSF, arXiv:1205.6410]

Quark mass dependence

- Dependence of observables with respect to quark masses is useful for tuning, or e.g. sigma terms.

- $N_f = 2$ **QCD with "exact determinant"**

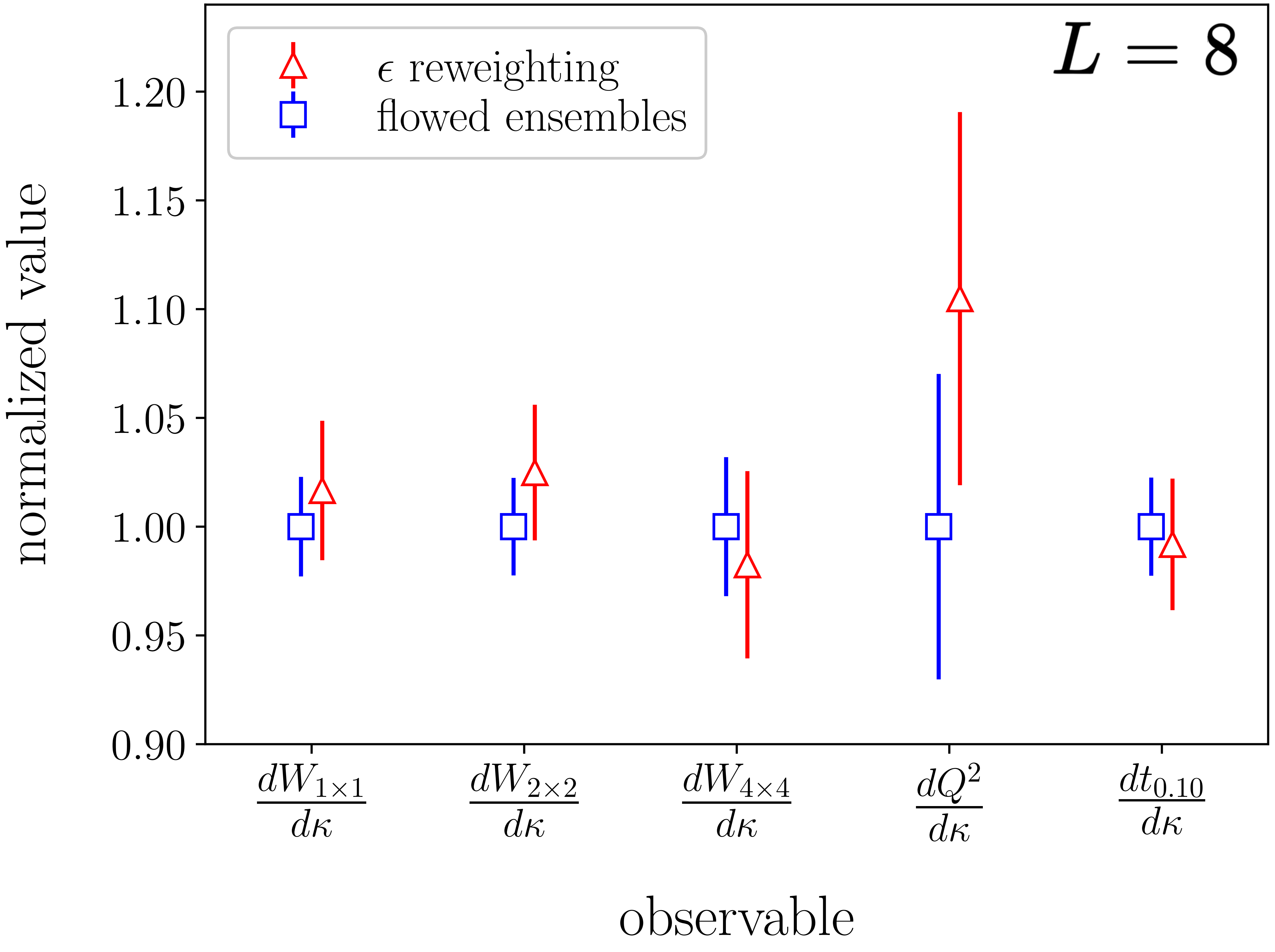
$$\beta = 5.6, \ \kappa_1 = 0.1530$$

$$\downarrow$$

$$\beta = 5.6, \ \kappa_2 = 0.1545$$

- **As an example, compute**

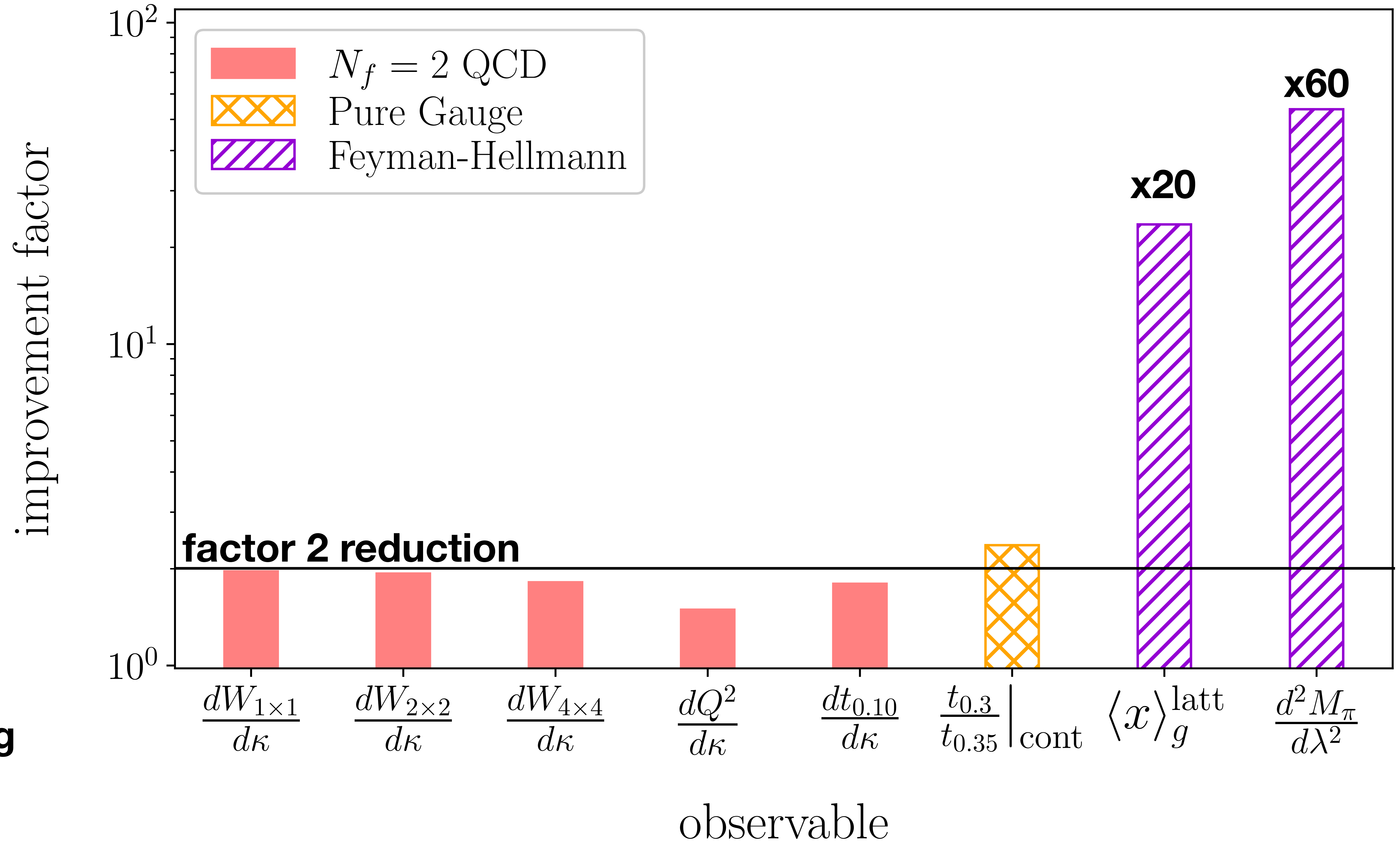
$$\frac{d\langle \mathcal{O} \rangle}{d\kappa} = \frac{\langle \mathcal{O} \rangle_{\kappa_1} - \langle \mathcal{O} \rangle_{\kappa_2}}{\Delta\kappa}$$



Fernando Romero-López, Uni Bern

Summary of results

- Overall, number of required samples for a error goal decreases



Ratio of variance with respect to epsilon reweighing

Comparison at fixed number of samples

A more exhaustive comparison needs training costs, flow evaluation costs, observable evaluation costs…

Towards hadron structure in dynamical QCD

The setup

- Consider N_f=2 QCD with twisted-mass fermions
  - Tree-level improved gauge action
  - Lattice spacing $a = 0.10\ \mathrm{fm}$
  - Pion mass $M_\pi = 520\ \mathrm{MeV}$
  - Target volume $12^3 \times 24$
- Compute matrix elements of the gluon part of the Energy-Momentum tensor:

$$\mathcal{O} = -\frac{\beta}{N_c} \mathrm{Tr}\,\mathrm{Re}\left(\sum_i U_{i0} - \sum_{i<j} U_{ij}\right)$$

- Training can be done at small volume V=4^4 with exact fermion determinant

$$\text{train ESS} = 99.6\%$$
$$(\text{c.f. baseline ESS} = 93.7\%)$$

Pseudofermions

- At the target volume, cannot evaluate the fermion determinant

- For the flow reweighting factors, need to evaluate stochastically the ratio of determinants

$$\frac{\det DD^\dagger[f(U)]}{\det DD^\dagger[U]} = \int D\phi \exp^{-\phi^\dagger (MM^\dagger)^{-1}\phi}$$

$$M = D[f(U)]D^{-1}[U]$$

Pseudofermions

- At the target volume, cannot evaluate the fermion determinant
- For the flow reweighting factors, need to evaluate stochastically the ratio of determinants

$$\frac{\det DD^\dagger[f(U)]}{\det DD^\dagger[U]} = \int D\phi \exp^{-\phi^\dagger (MM^\dagger)^{-1}\phi}$$

$$M = D[f(U)]D^{-1}[U]$$

- Can actually add a pseudofermion model on top of the gauge flow!

$$\phi'(x) = A(U)\,\phi(x) + B(U)\,U_\mu(x)\phi(x+\mu) \longrightarrow \phi' \simeq M\phi$$

Trainable

parallel-transported neighbor

Fernando Romero-López, Uni Bern

Pseudofermions

- At the target volume, cannot evaluate the fermion determinant

- For the flow reweighting factors, need to evaluate stochastically the ratio of determinants

$$\frac{\det DD^\dagger[f(U)]}{\det DD^\dagger[U]} = \int D\phi \exp^{-\phi^\dagger (MM^\dagger)^{-1}\phi}$$

$$M = D[f(U)]D^{-1}[U]$$

- Can actually add a pseudofermion model on top of the gauge flow!

$$\phi'(x) = A(U)\,\phi(x) + B(U)\,U_\mu(x)\phi(x+\mu) \longrightarrow \phi' \simeq M\phi$$

Trainable

parallel-transported neighbor

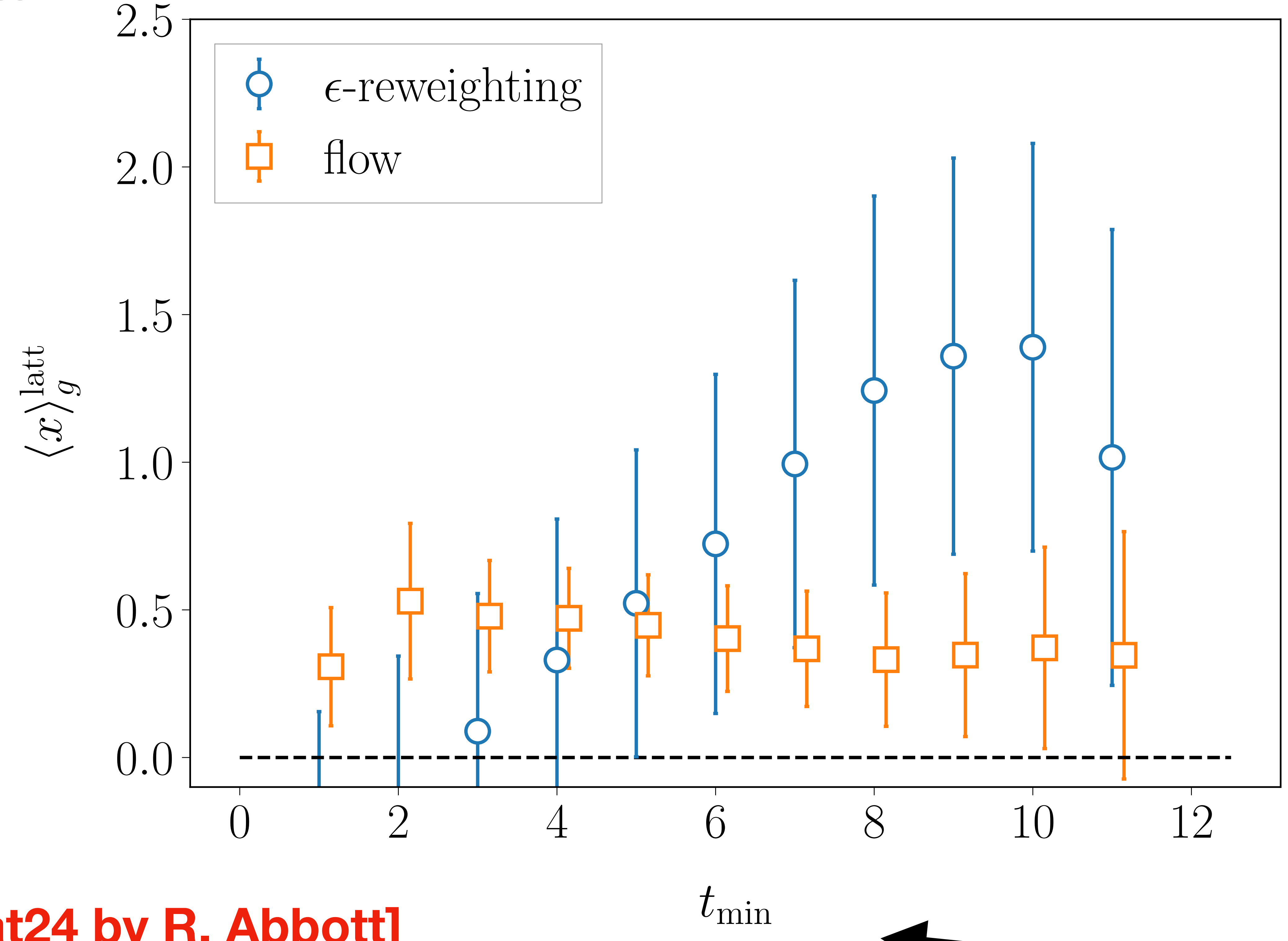
- This leads to an increase of the ESS in the target volume:

$$\text{ESS(stoch ratio det)} = 45\%$$
$$\text{ESS(PF flow)} = 50\%$$

Fernando Romero-López, Uni Bern

Preliminary results

$$\frac{dM_\pi}{d\lambda} = -\frac{3M_\pi}{2}\langle x\rangle_g^{\mathrm{latt}}$$



Error reduction of about 2x-3x

Need 5-10 times less configs!

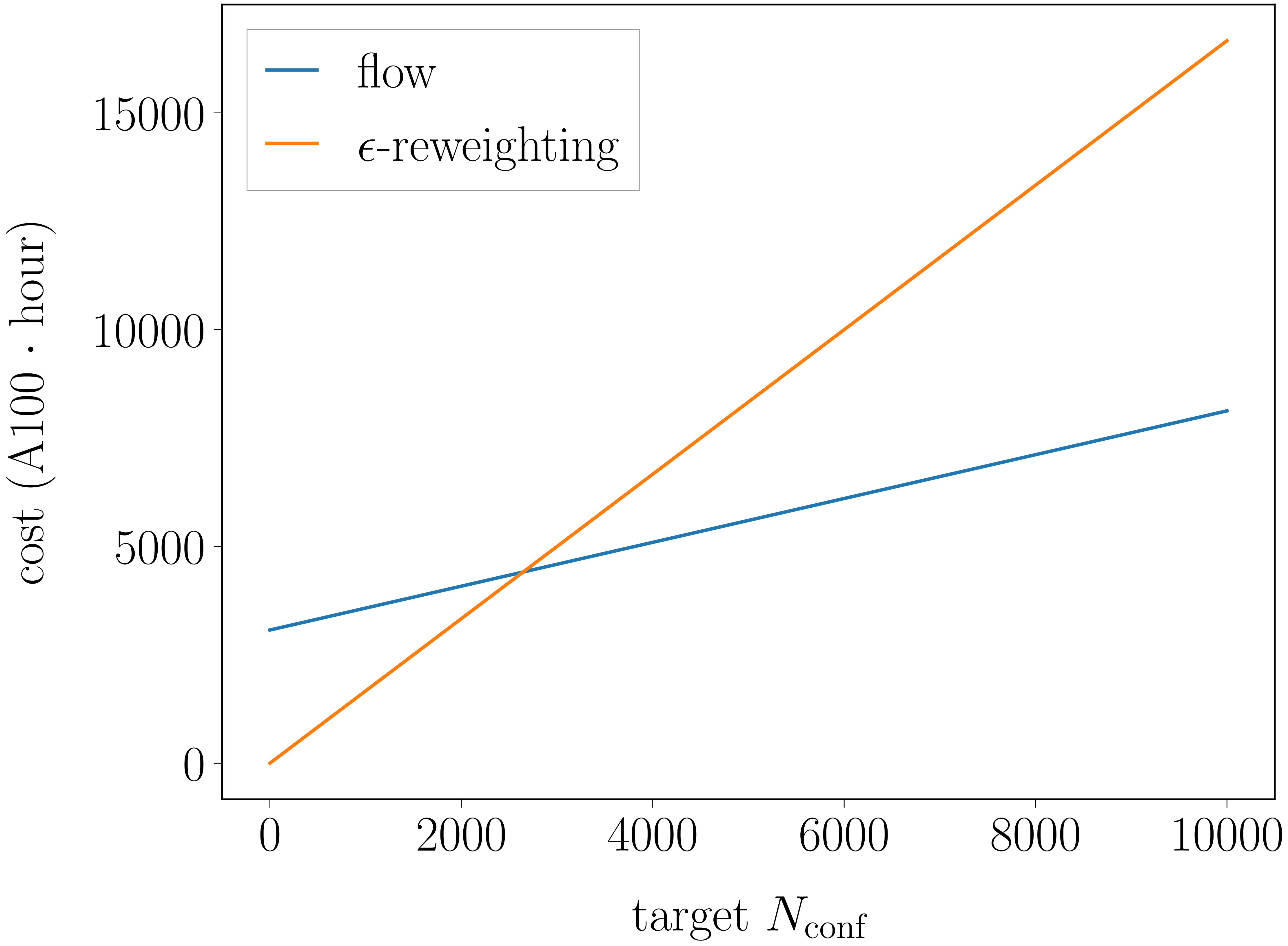
[Talk presented at Lat24 by R. Abbott]

Fit range: $[t_{min}, T/2]$

Costs

1. Using flows:

- Training costs: 100 hours in 16 A100s.
- Configuration generation (Chroma): 600s/config in 1 A100
- Flow application: 20s/config in 1 A100
- Measurements: 2 x 600s/config in 1 A100



2. Epsilon reweighting:

- Need x5 more configs
- Same generation costs
- Measurements only needed once

Fernando Romero-López, Uni Bern

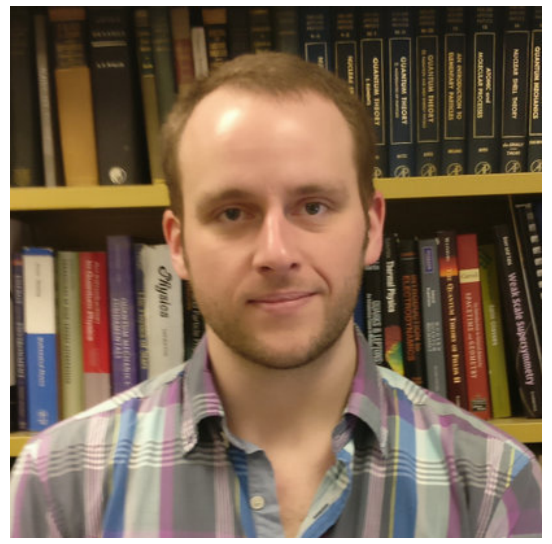
Bonus:
Transformed Replica EXchange (T-REX)

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]

arXiv:2404.11674



See talk @ latt23

Dan Hackett (FNAL)

Replica EXchange (REX)

- A known algorithm for lattice QCD is running several Markov Chains in parallel and proposing swaps
  [Hasenbusch, arXiv:1706.04443], [Bonanno et al, arXiv:2012.14000 & arXiv:2014.14151]

$$\beta_1 : \quad U_1^{(n)} \longrightarrow U_1^{(n+1)} \qquad U_1^{(n+2)} = U_0^{(n+1)}$$

$$\beta_0 : \quad U_0^{(n)} \longrightarrow U_0^{(n+1)} \qquad U_0^{(n+2)} = U_1^{(n+1)}$$

$$p_{\mathrm{acc}} = \min\left[1, \frac{p_0(U_1)p_1(U_0)}{p_0(U_0)p_1(U_1)}\right]$$

- Can accelerate mixing of topological sectors if one chain “moves faster”.

Transformed Replica EXchange (T-REX)

- Swapping of configurations can be combined with a flow to increase swap probability

$$\beta_1 : \quad U_1^{(n)} \longrightarrow U_1^{(n+1)} \qquad U_1^{(n+2)} = f(U_0^{(n+1)})$$

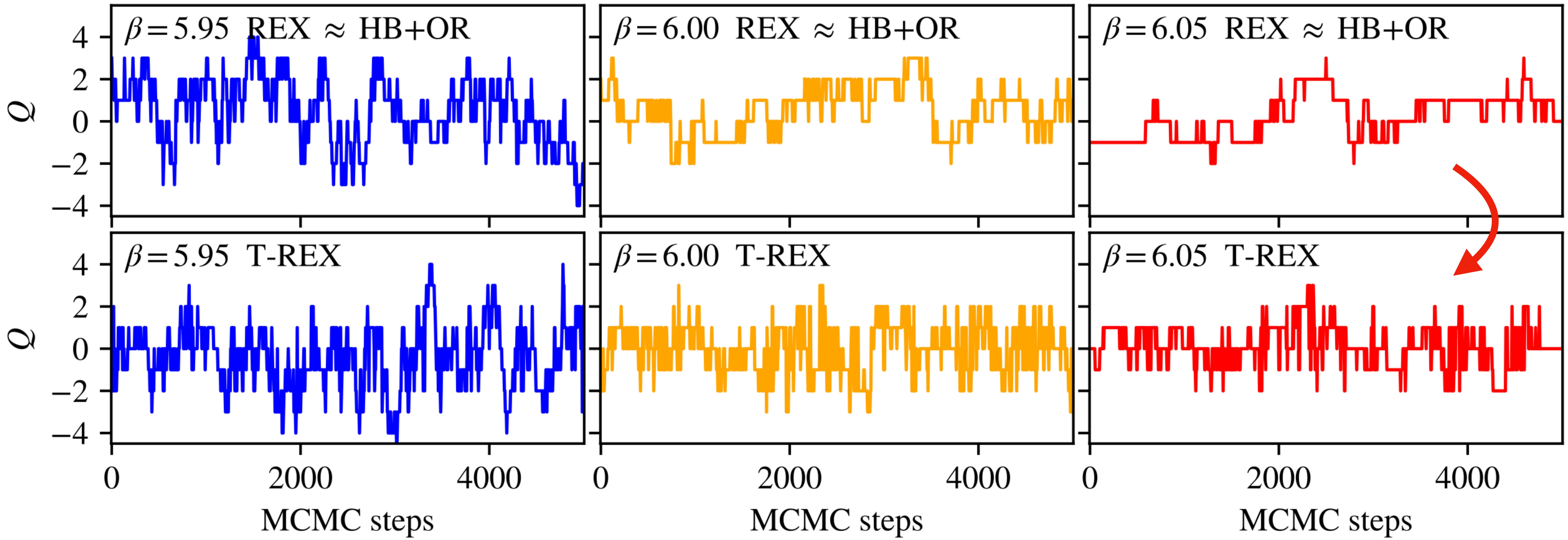
$$\beta_0 : \quad U_0^{(n)} \longrightarrow U_0^{(n+1)} \qquad U_0^{(n+2)} = f^{-1}(U_1^{(n+1)})$$

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

Example

Topology mixing slows down

$V = 12^4$

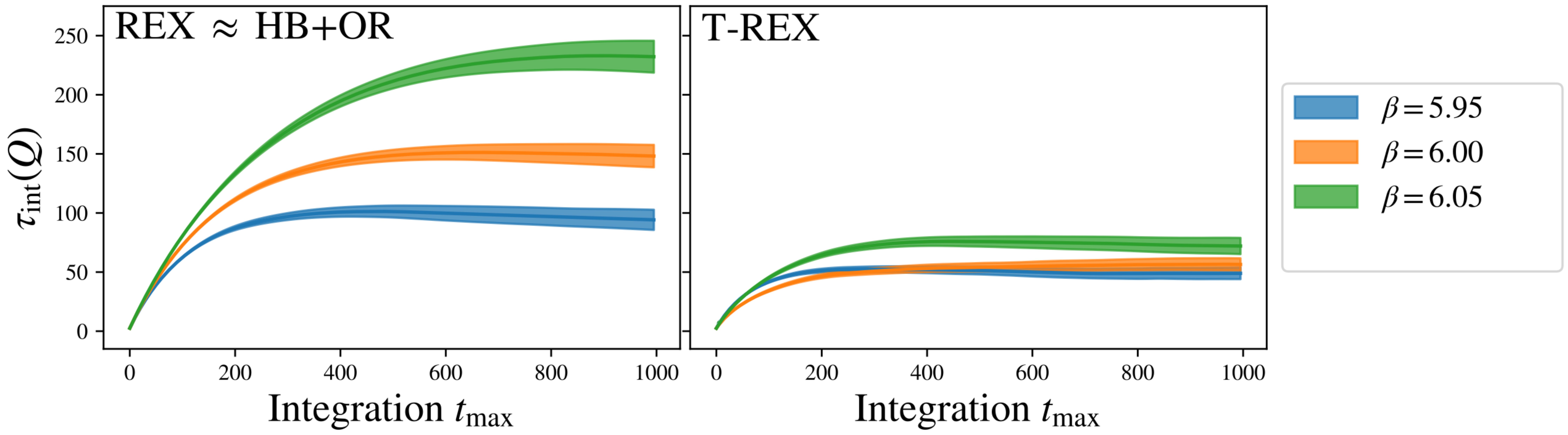


Faster topology mixing

Fernando Romero-López, Uni Bern

MCMC history

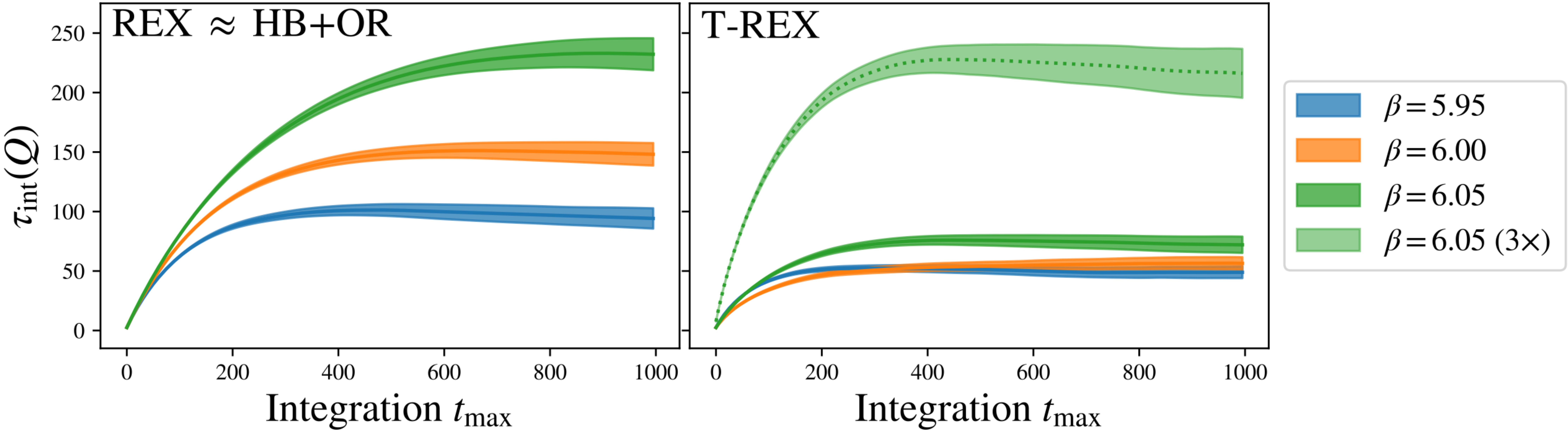
- All integrated autocorrelation times reduce significantly



Fernando Romero-López, Uni Bern

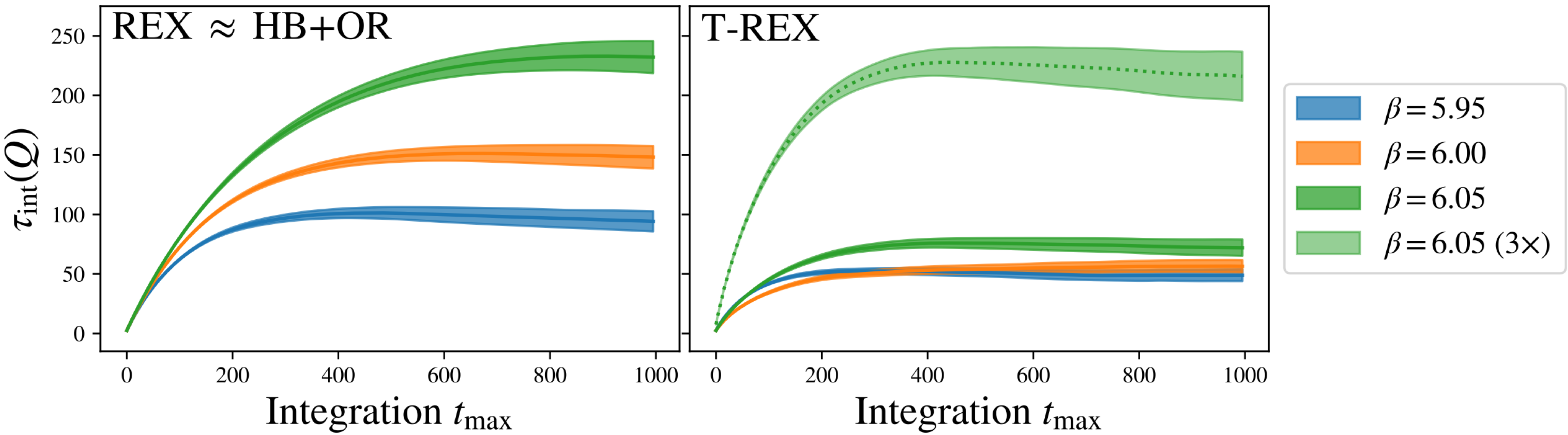
MCMC history

- All integrated autocorrelation times reduce significantly



MCMC history

- All integrated autocorrelation times reduce significantly



- Neglecting flow costs, computational advantage if one is interested in all three chains
- If only the “finest” ensembles is used, almost break even

Summary $ Outlook

Summary & Outlook

- ☑ Lattice QCD is the first-principle treatment of the strong interaction at hadronic energies
- ○ Flow-based sampling has the potential to accelerate sampling of QCD configurations

- ☑ Direct sampling remains challenging, but current flows can map effectively between nearby parameters
- ☑ Flow models can be used to compute derivative observables by generating “correlated ensembles”

- ☑ Promising numerical demonstrations in QCD / Yang Mills
- ○ Next steps: correlated ensembles at state-of-the-art QCD scales!

- ☑ Flows allow for increased acceptance rates in replica exchange: T-REX
- ○ Acceptance rate degrades with volume. Use an action with localized defects? What about fermions?

Fernando Romero-López, Uni Bern

Summary & Outlook

- ☑ Lattice QCD is the first-principle treatment of the strong interaction at hadronic energies
- ○ Flow-based sampling has the potential to accelerate sampling of QCD configurations

- ☑ Direct sampling remains challenging, but current flows can map effectively between nearby parameters
- ☑ Flow models can be used to compute derivative observables by generating “correlated ensembles”

- ☑ Promising numerical demonstrations in QCD / Yang Mills
- ○ Next steps: correlated ensembles at state-of-the-art QCD scales!

Thanks!

- ☑ Flows allow for increased acceptance rates in replica exchange: T-REX
- ○ Acceptance rate degrades with volume. Use an action with localized defects? What about fermions?