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

Fernando Romero-López

ECT*

May 28th

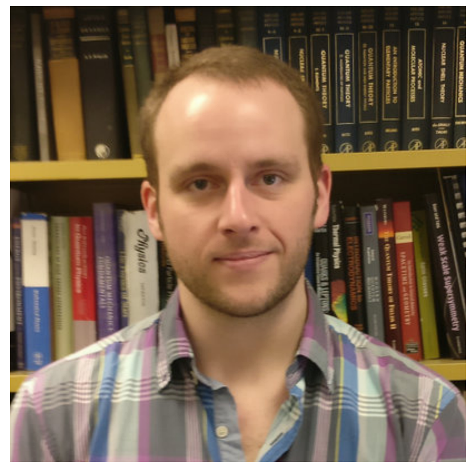
Massachusetts Institute of Technology

fernando@mit.edu

Collaboration (non-exhaustive)

Massachusetts Institute of Technology

- Phiala Shanahan



- Dan Hackett

- Fernando Romero-Lopez

- Ryan Abbot

- Julian Urban

- Denis Boyda

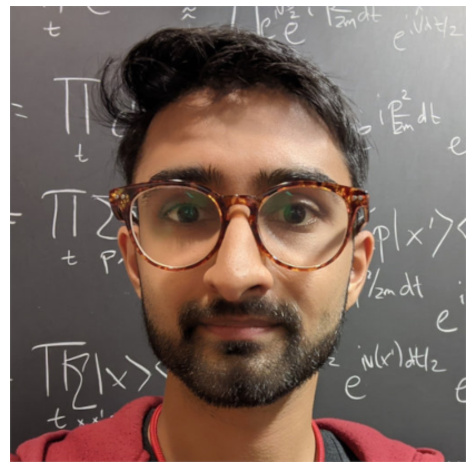
- Kyle Cranmer

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

NEW YORK UNIVERSITY

- Michael Albergo

UNIVERSITÄT BERN



- Gurtej Kanwar

DeepMind

- Sébastien Racanière

- Danilo Rezende

- Ali Razavi

- Aleksandar Botev

- Alex Matthews

Outline

1. Introduction
2. Flows for correlated ensembles
3. Numerical demonstrations
4. Bonus: Transformed Replica Exchange (T-REX)
5. 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, MIT

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 $$\underbrace{S_E(\phi) = \int d^4x \, \mathscr{L}_E(\phi)}_{\text{Euclidean action}}$$

Fernando Romero-López, MIT

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 $S_E(\phi) = \int d^4x \, \mathcal{L}_E(\phi)$

Euclidean action

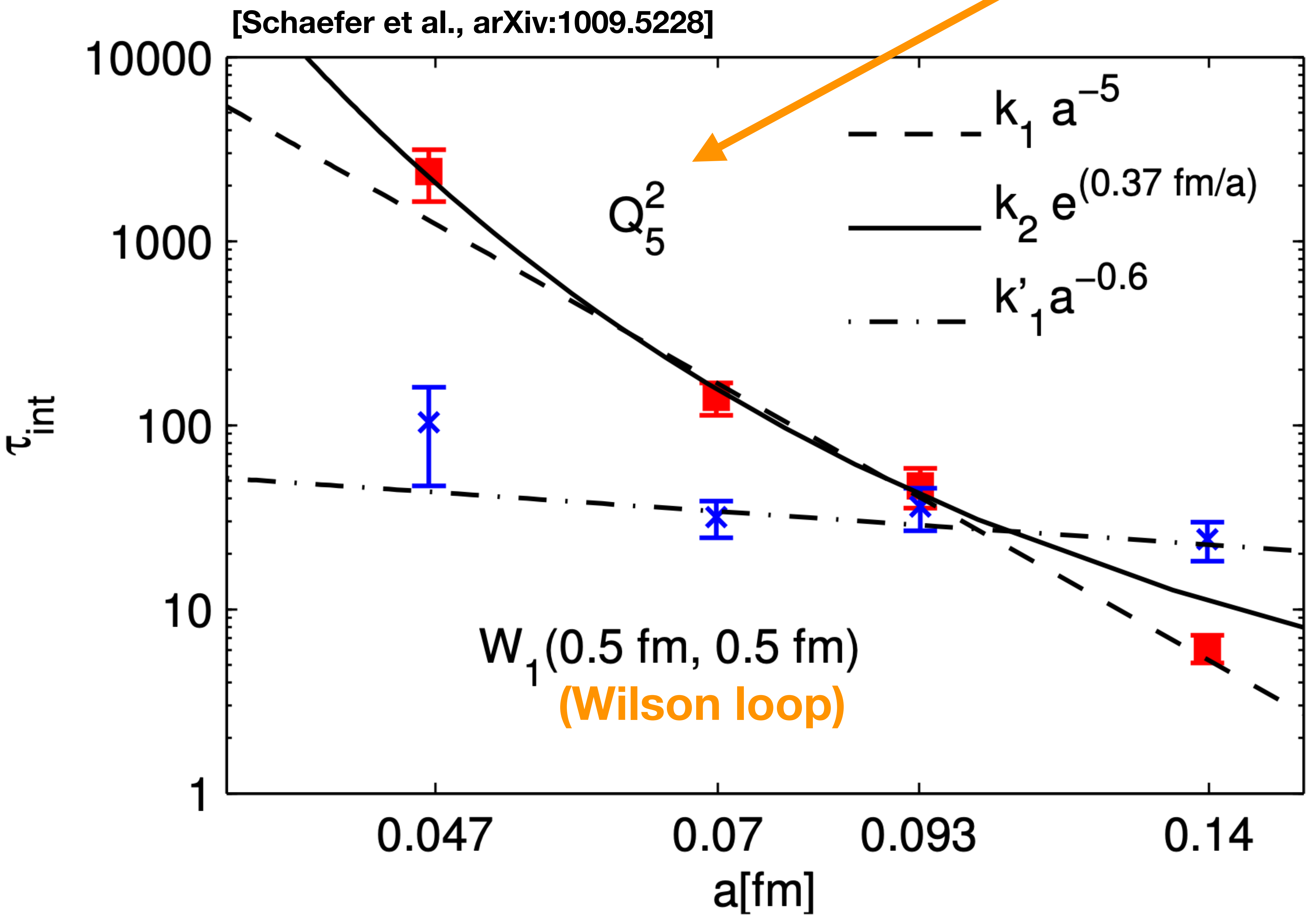
- Increasing interest in applying generative flow models to LQCD
  - Can flow models reduce the associated costs of generating configurations?

Fernando Romero-López, MIT

The main problem

Topological charge

autocorrelation time $\simeq$ computational cost



Continuum limit

Computational cost of generating independent samples “explodes” towards the continuum limit: **Critical Slowing Down**

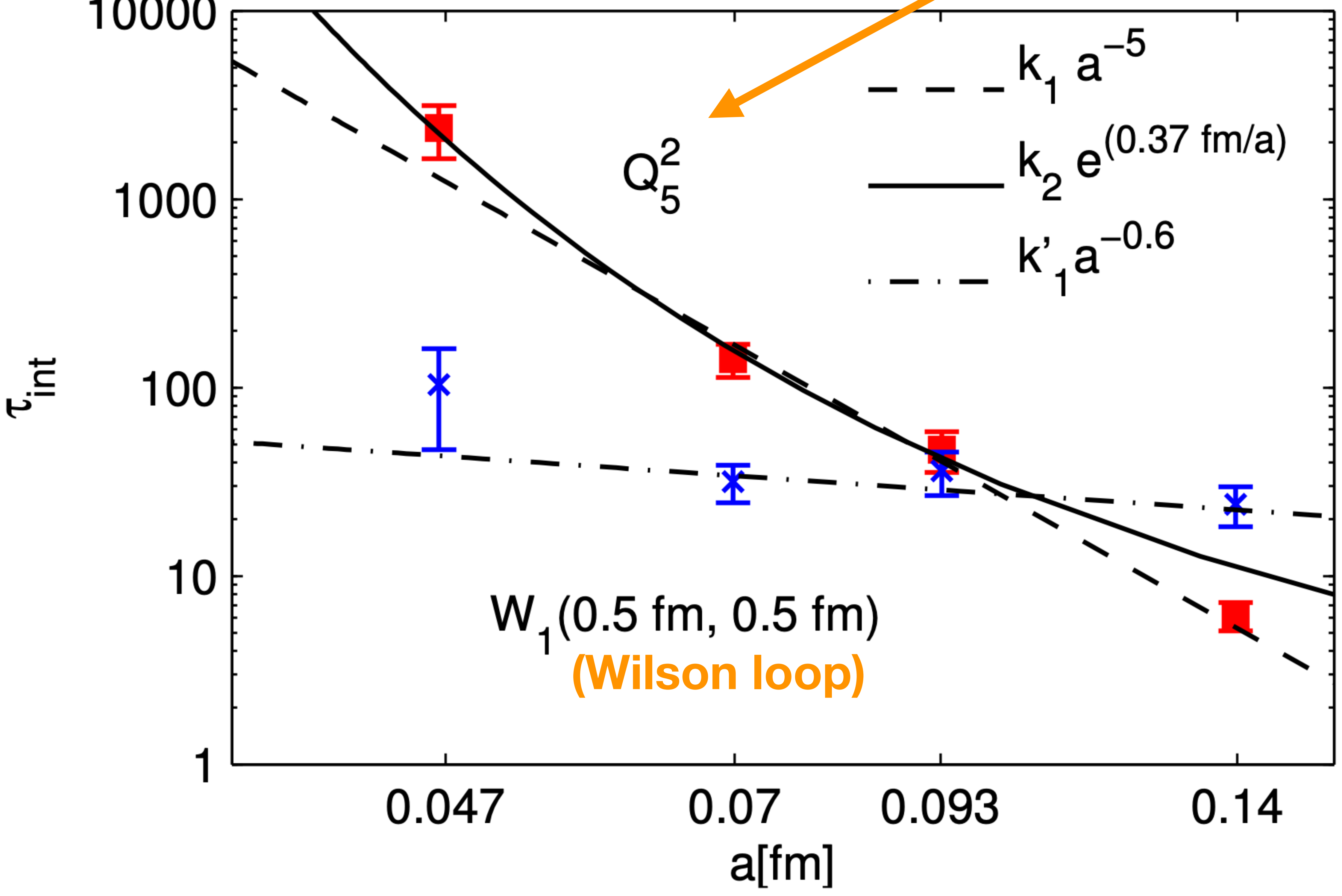
Fernando Romero-López, MIT

The main problem

Topological charge

[Schaefer et al., arXiv:1009.5228]

autocorrelation time
≃
computational cost



Continuum limit

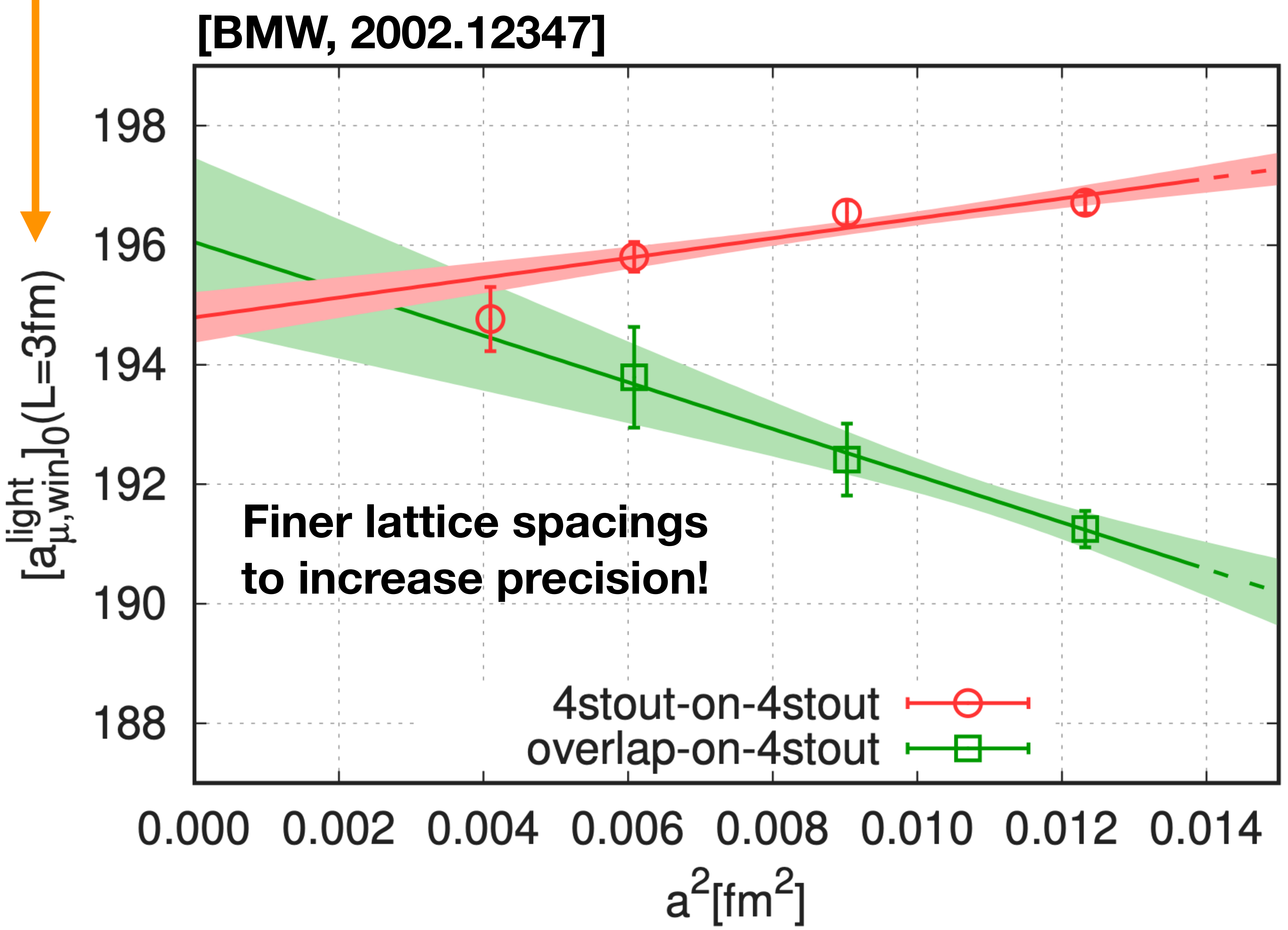
Computational cost of generating independent samples "explodes" towards the continuum limit: Critical Slowing Down

Can flows help?

Fernando Romero-López, MIT

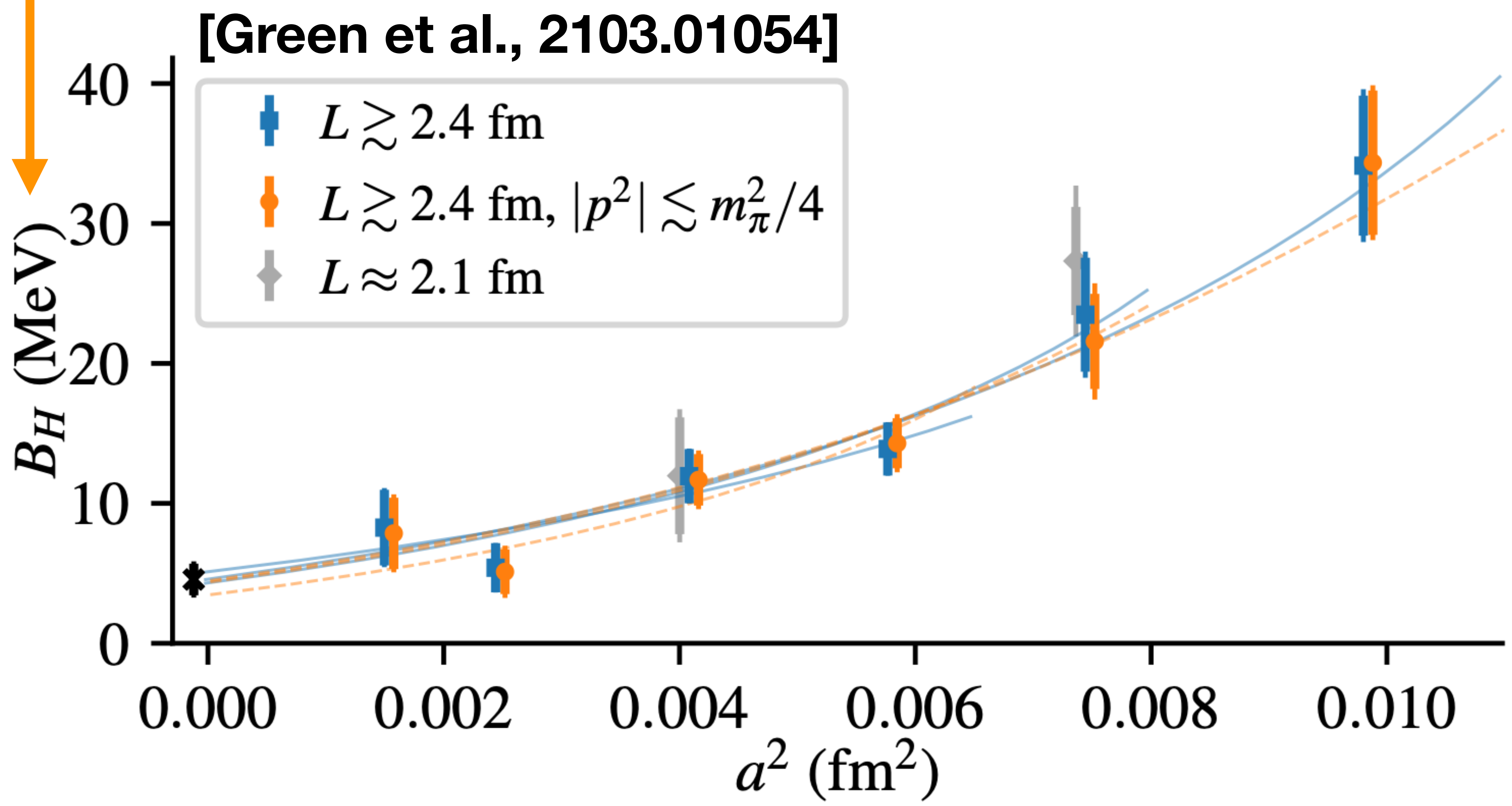
The need for a continuum limit

HVP of muon magnetic moment



Continuum limit

Binding energy of H dibaryon



Continuum limit

Where do we stand?

(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]

…

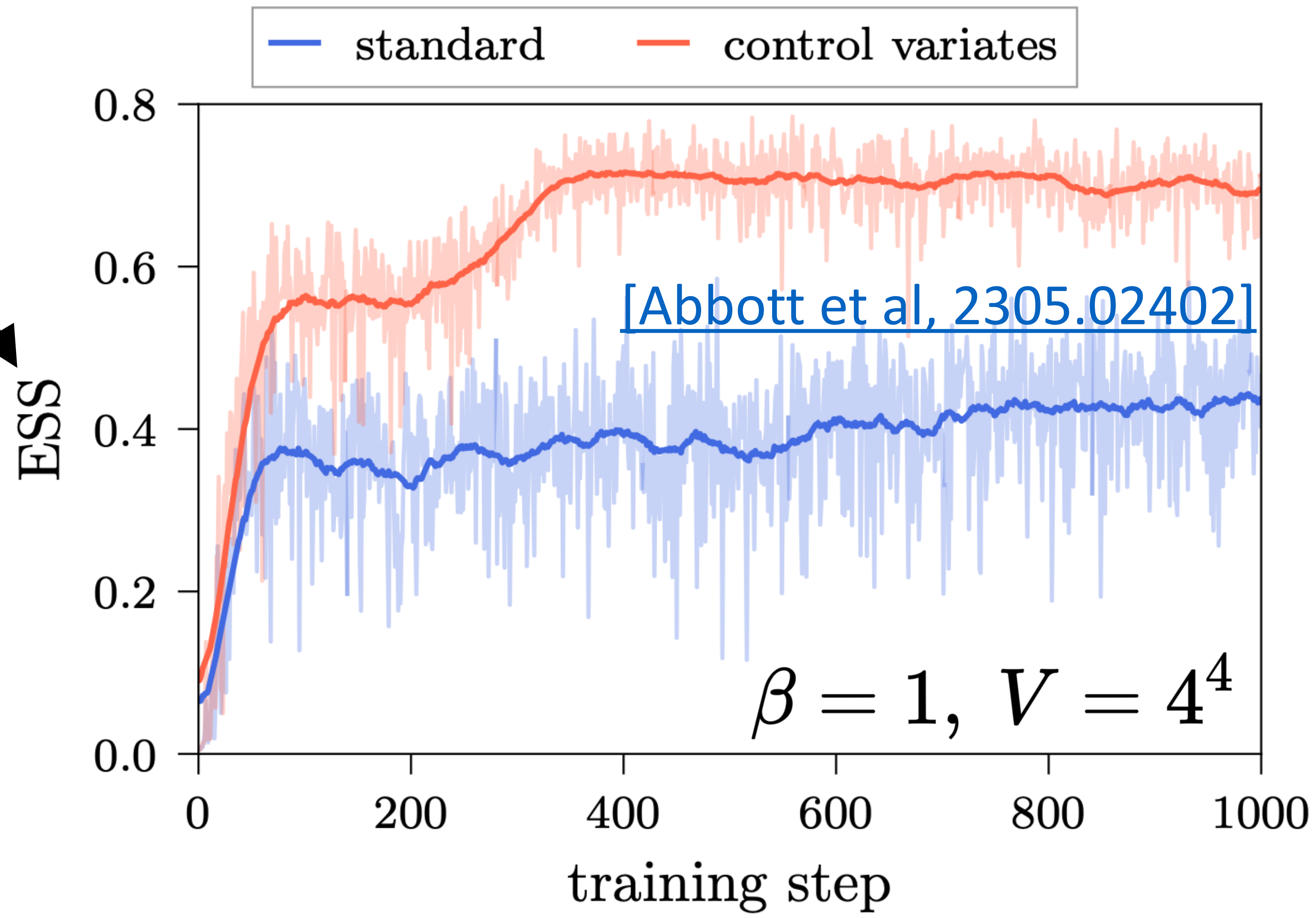
Still some developments are needed for at-scale QCD

Fernando Romero-López, MIT

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

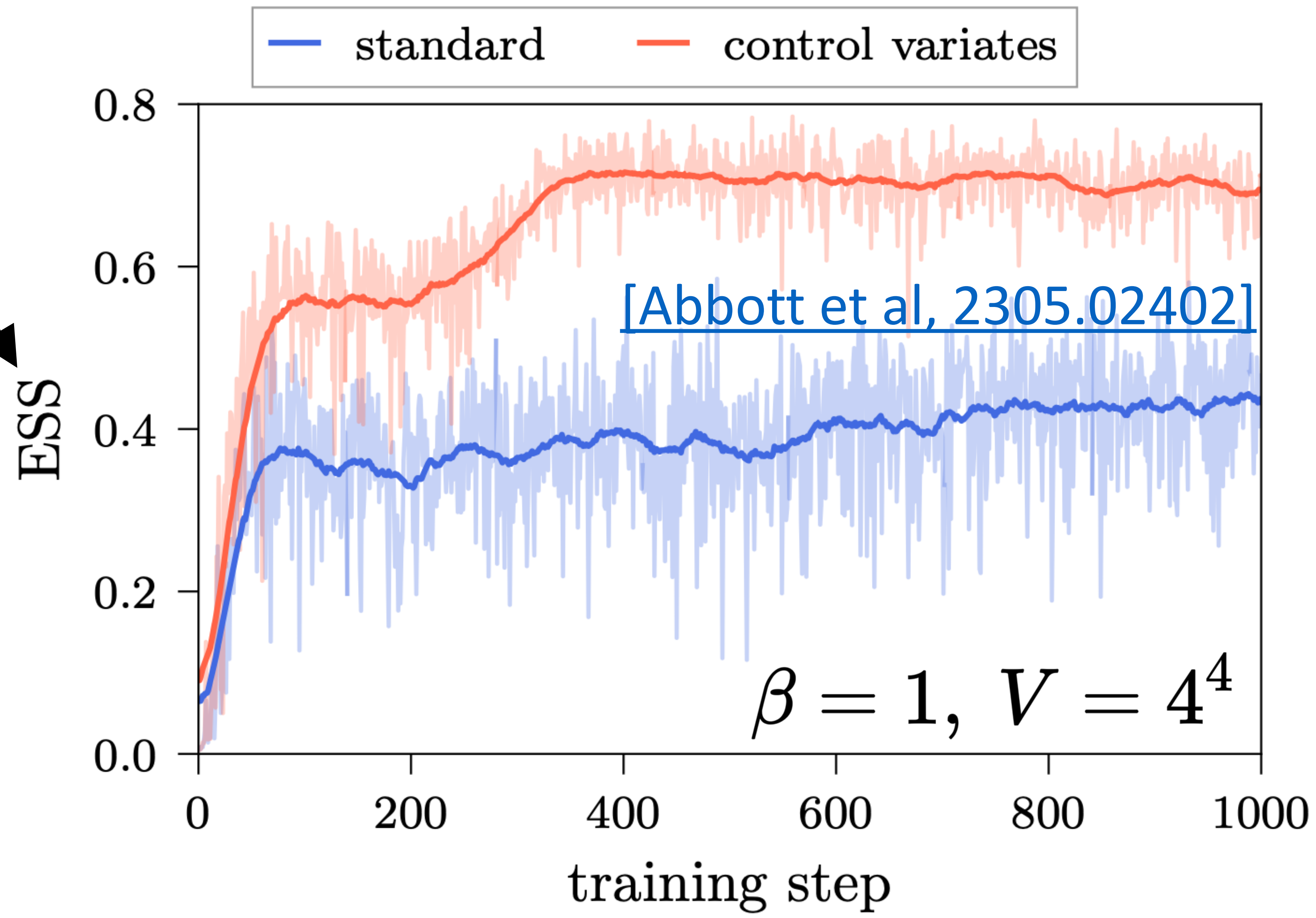


Fernando Romero-López, MIT

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)

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]

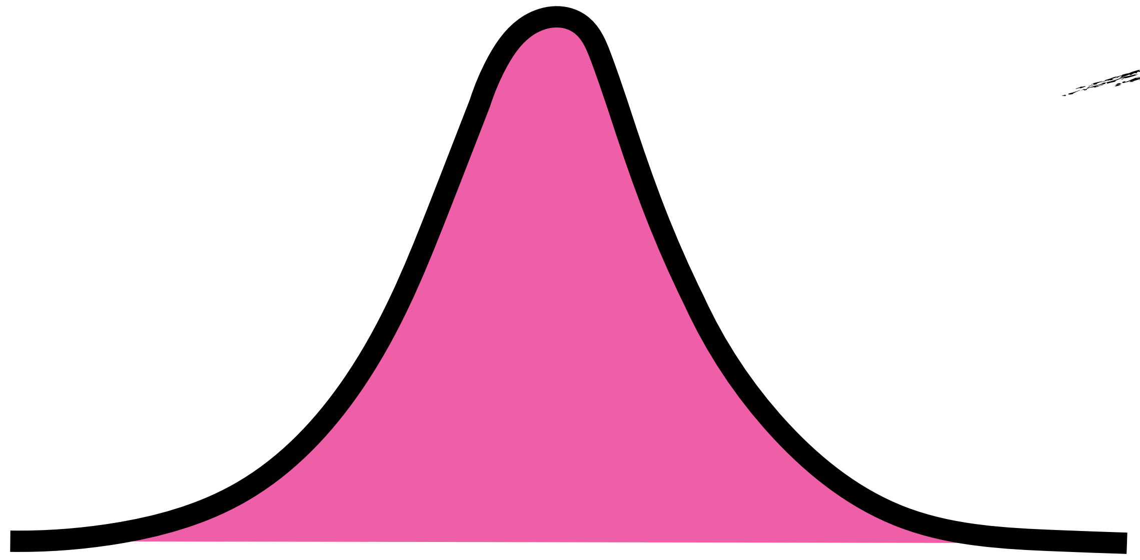
Generative flow models

[Rezende, Mohamed, 1505.05770]

$f(z)$

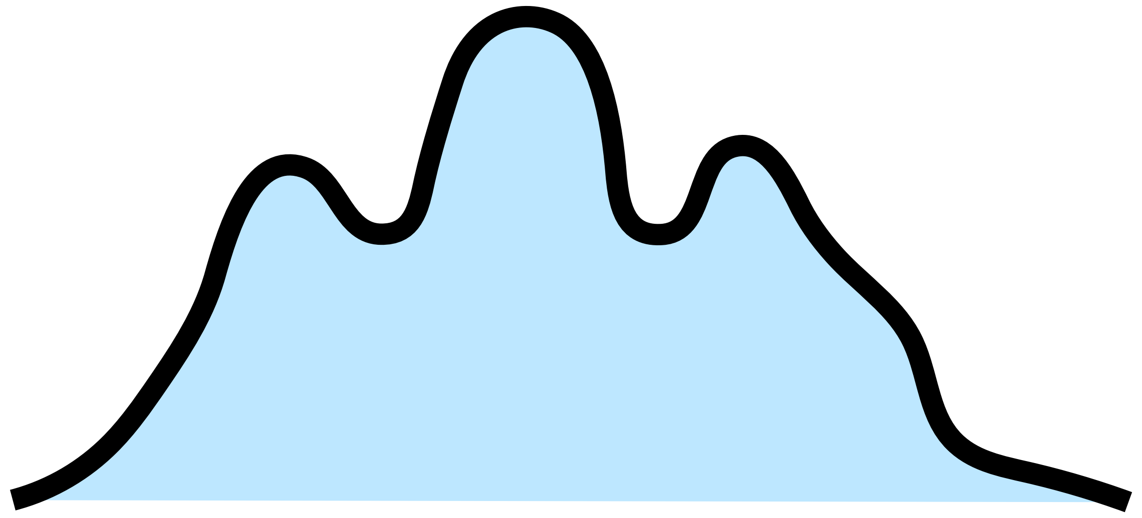
tractable Jacobian
&
invertible

$r(z)$



easy-to-sample
distribution
(prior)

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



approximates
target distribution
(model)

"Trainable change of variables"

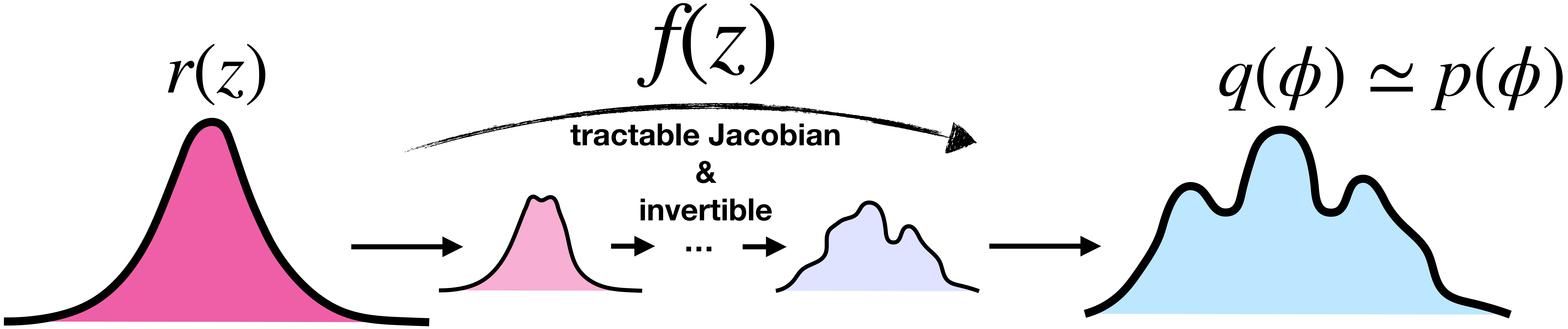
Model probability

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

Fernando Romero-López, MIT

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"

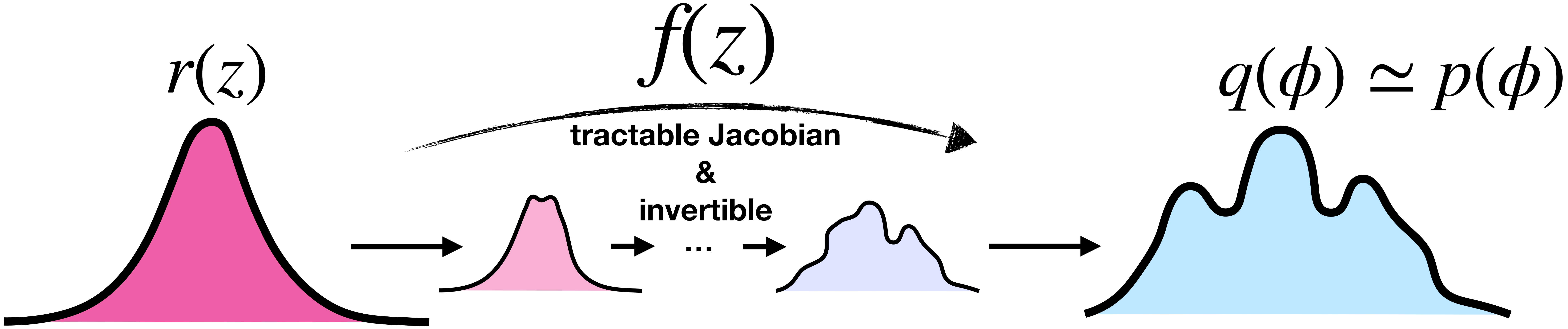
Model probability

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

Fernando Romero-López, MIT

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"

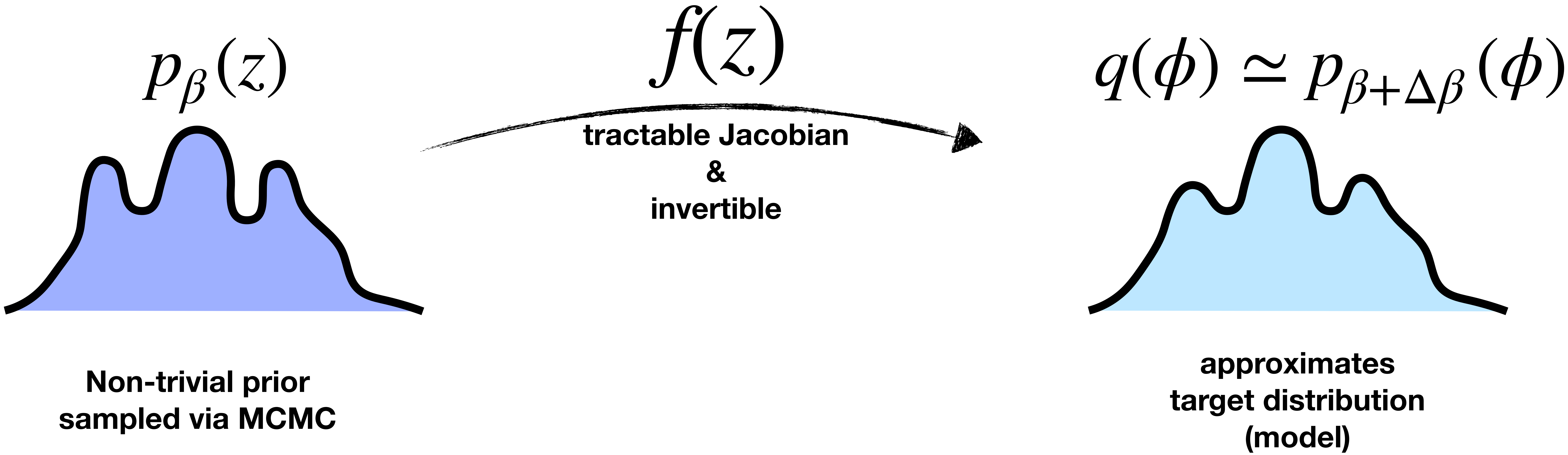
Model probability

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

❗ Trained models are not perfect.

☑ But exact sampling can be recovered via Markov Chain

Flows for correlated ensembles



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

Fernando Romero-López, MIT

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

action parameter

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

- 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…

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 $\Delta\alpha = \epsilon$

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

Fernando Romero-López, MIT

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 $\Delta\alpha$
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, MIT

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 $\Delta\alpha$

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, MIT

The architecture

- Use an equivariant flow architecture based on the Gradient Flow

"Residual layers"

[Abbott et al, 2305.02402]
[see also S. Bacchio et al, 2212.08469]

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

Untraced Wilson loops that start and end at x

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

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 S. Bacchio et al, 2212.08469]

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

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

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, MIT

The architecture

- Use an equivariant flow architecture based on the Gradient Flow

"Residual layers"

[Abbott et al, 2305.02402]
[see also S. Bacchio et al, 2212.08469]

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

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

Untraced Wilson loops that start and end at x

trainable

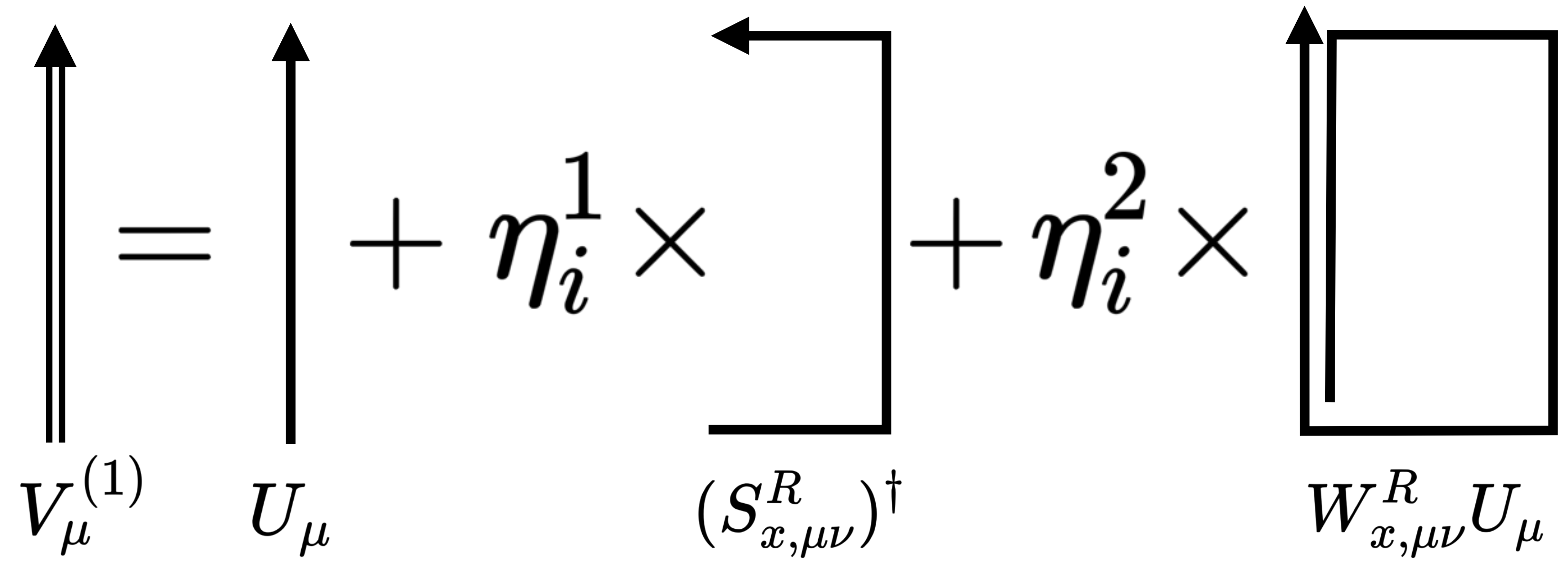
Traceless-antihermitian projection

- 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



[See also talk by U. Wenger/K. Holland]
[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]

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

solve numerically with infinitesimal steps

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]

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

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]

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

Qualitatively induces a change in the lattice spacing $\beta \longrightarrow \beta + \Delta\beta$

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

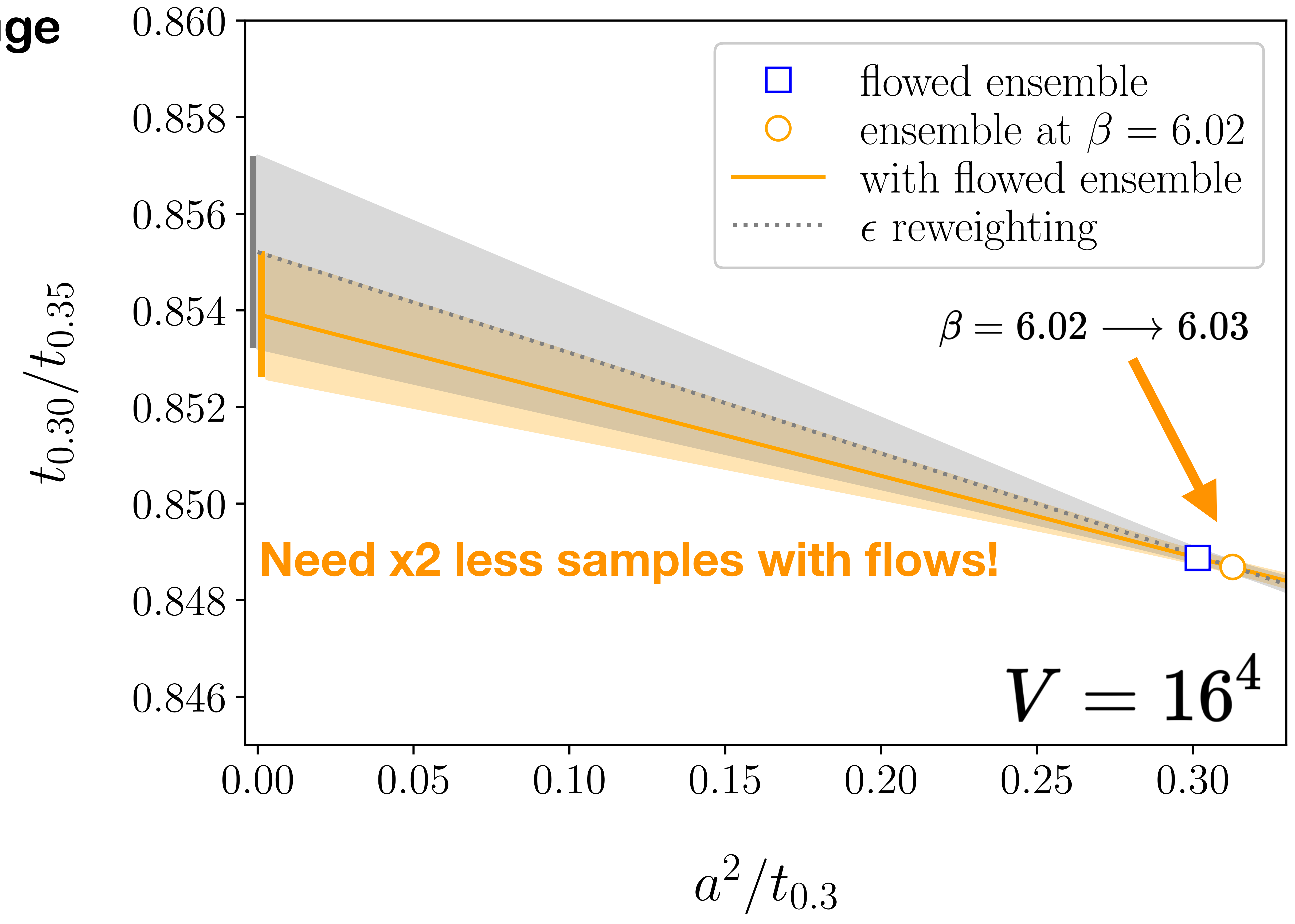
- Derivative of an observable with respect to lattice spacing is useful in constraining the continuum limit

- **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:**

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



Fernando Romero-López, MIT

Hadron structure

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

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

"Feynman-Hellmann theorem"

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, MIT

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!

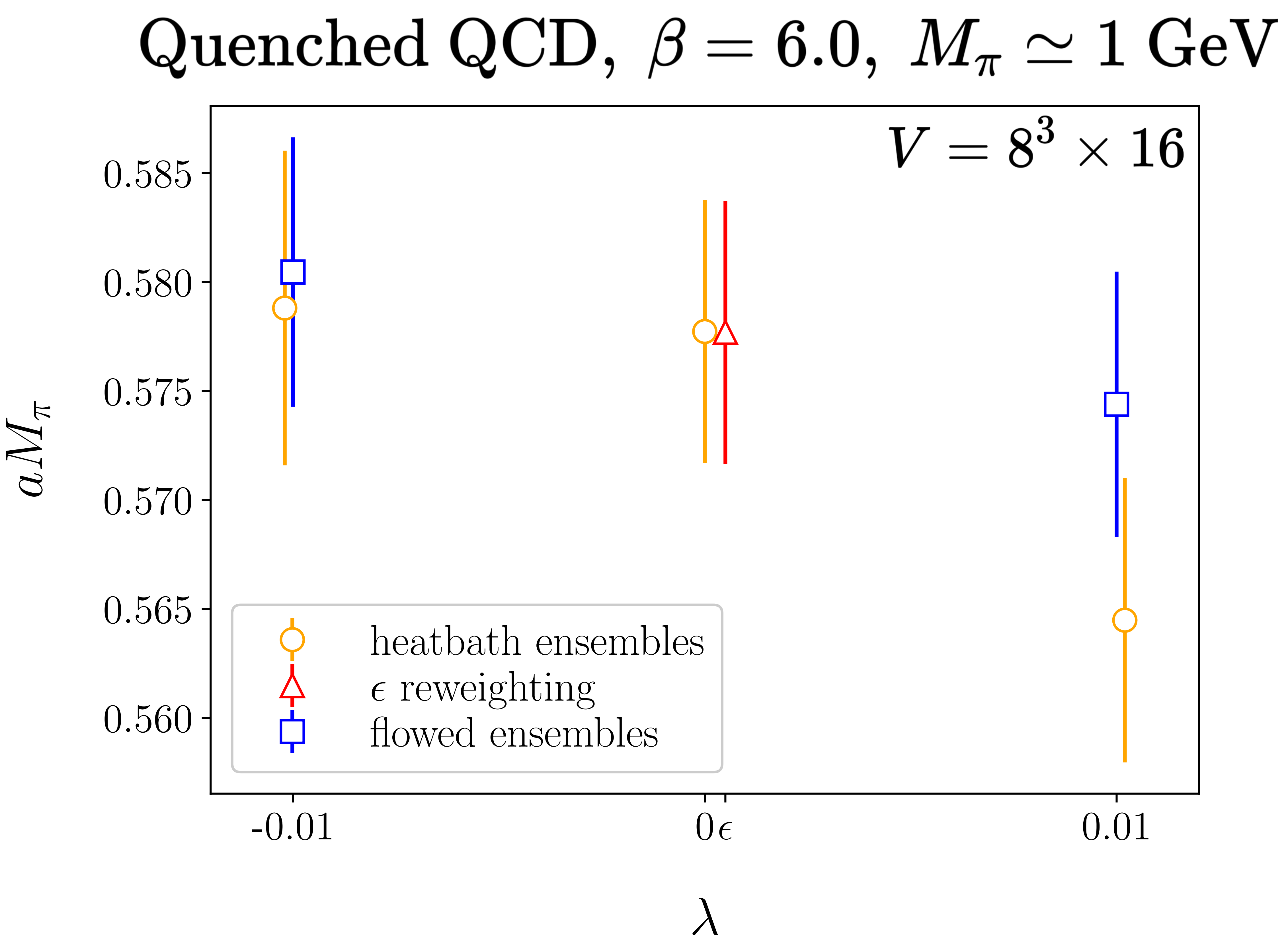
$$\boxed{S_\lambda = -\frac{\beta}{N_c}(1+\lambda)\,\mathrm{Re}\,\mathrm{Tr}\sum_i U_{i0} - \frac{\beta}{N_c}(1-\lambda)\,\mathrm{Re}\,\mathrm{Tr}\sum_{i<j} U_{ij}} \longrightarrow$$

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

Fernando Romero-López, MIT

Feynman-Hellmann results

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



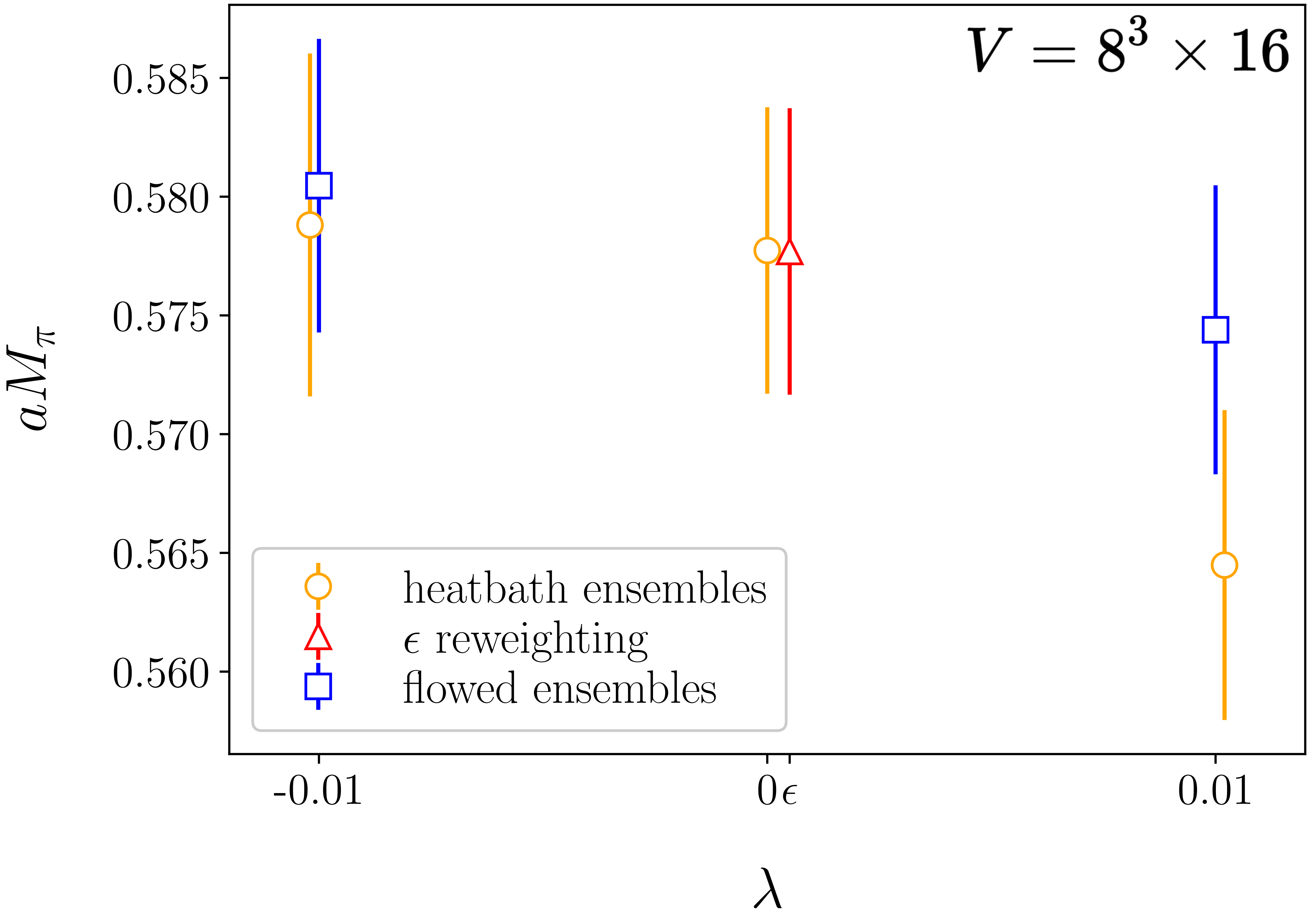
Same setup as [QCDSF, arXiv:1205.6410]

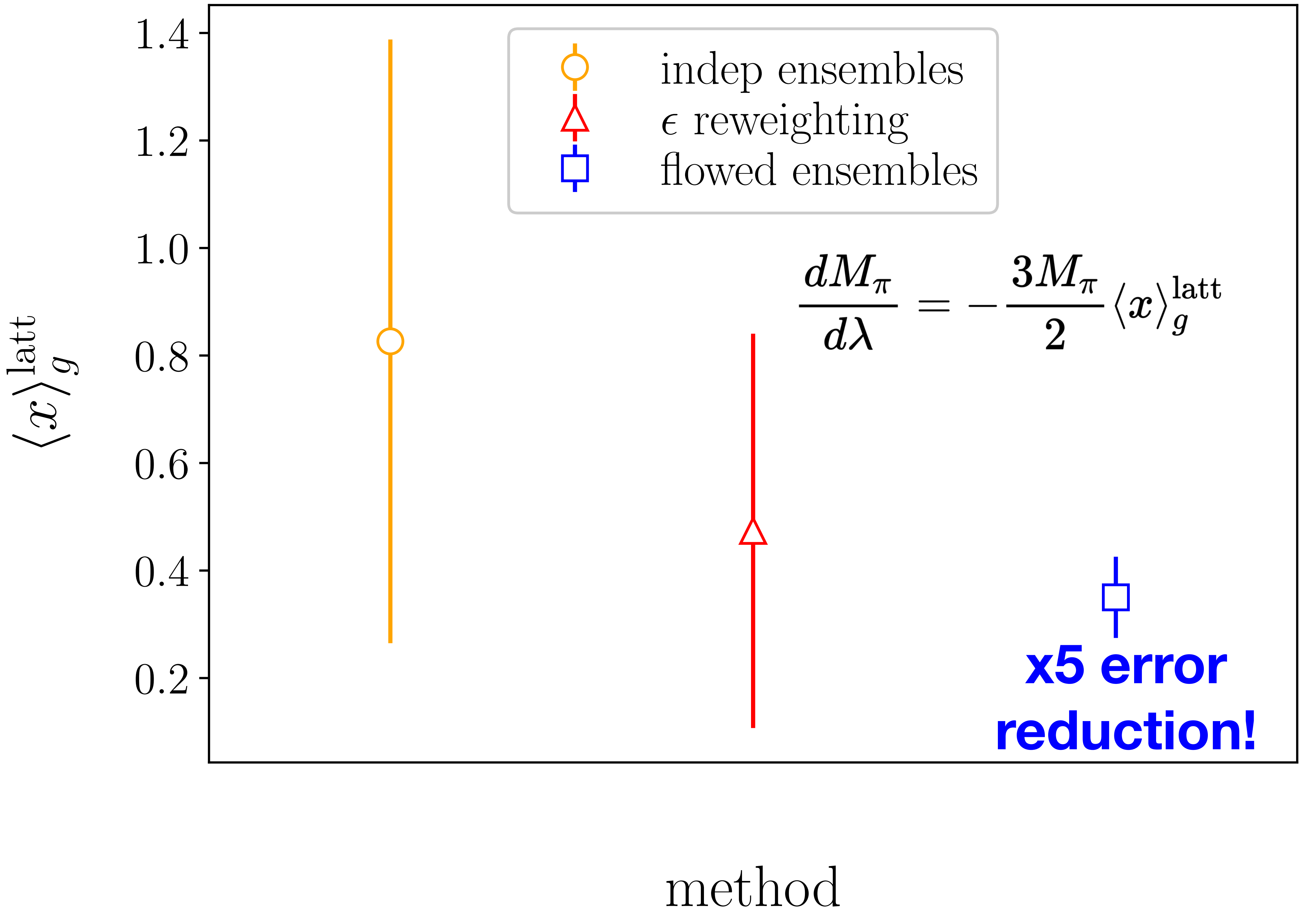
Fernando Romero-López, MIT

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]

Fernando Romero-López, MIT

Quark mass dependence

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

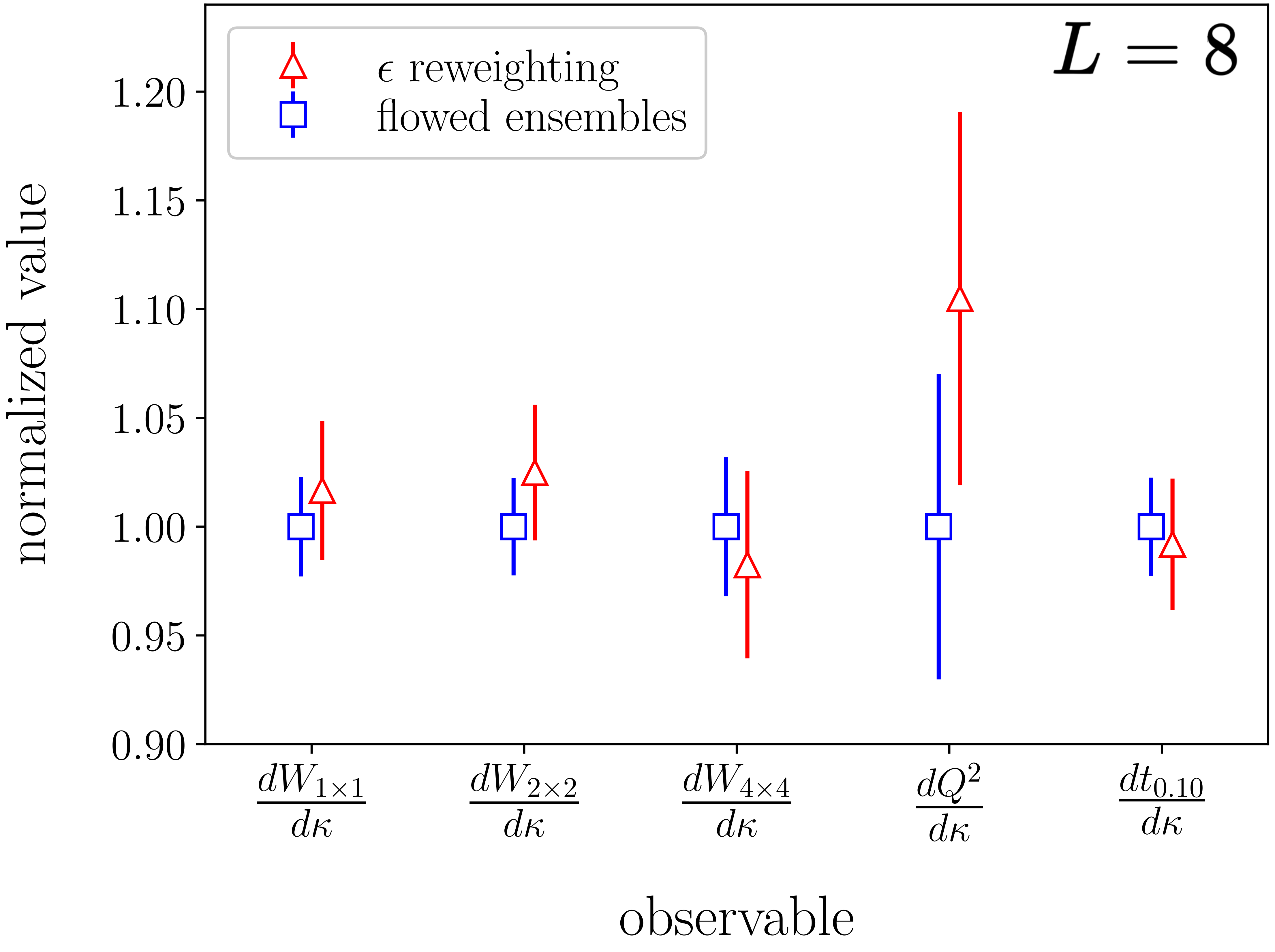
- $N_f = 2$ **QCD with "exact determinant"**

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

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

- **As an example, compute**

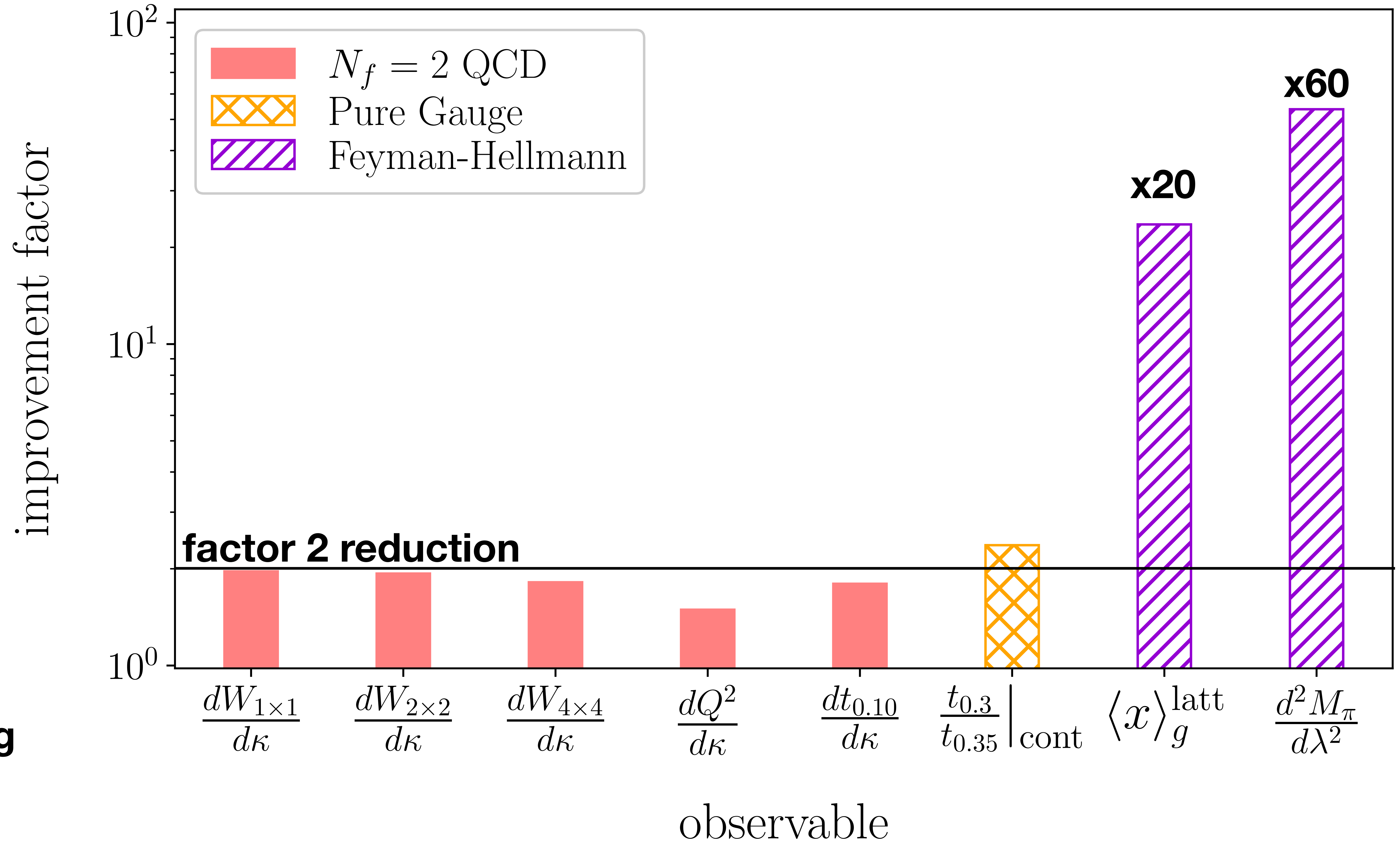
$$\frac{d\langle \mathcal{O} \rangle}{d\kappa} = \frac{\langle \mathcal{O} \rangle_{\kappa_1} - \langle \mathcal{O} \rangle_{\kappa_2}}{\Delta\kappa}$$



Fernando Romero-López, MIT

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…

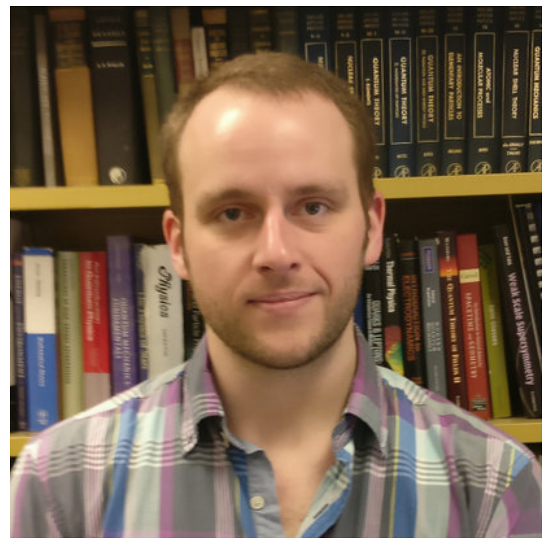
Fernando Romero-López, MIT

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”.

Fernando Romero-López, MIT

Transformed Replica EXchange (T-REX)

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

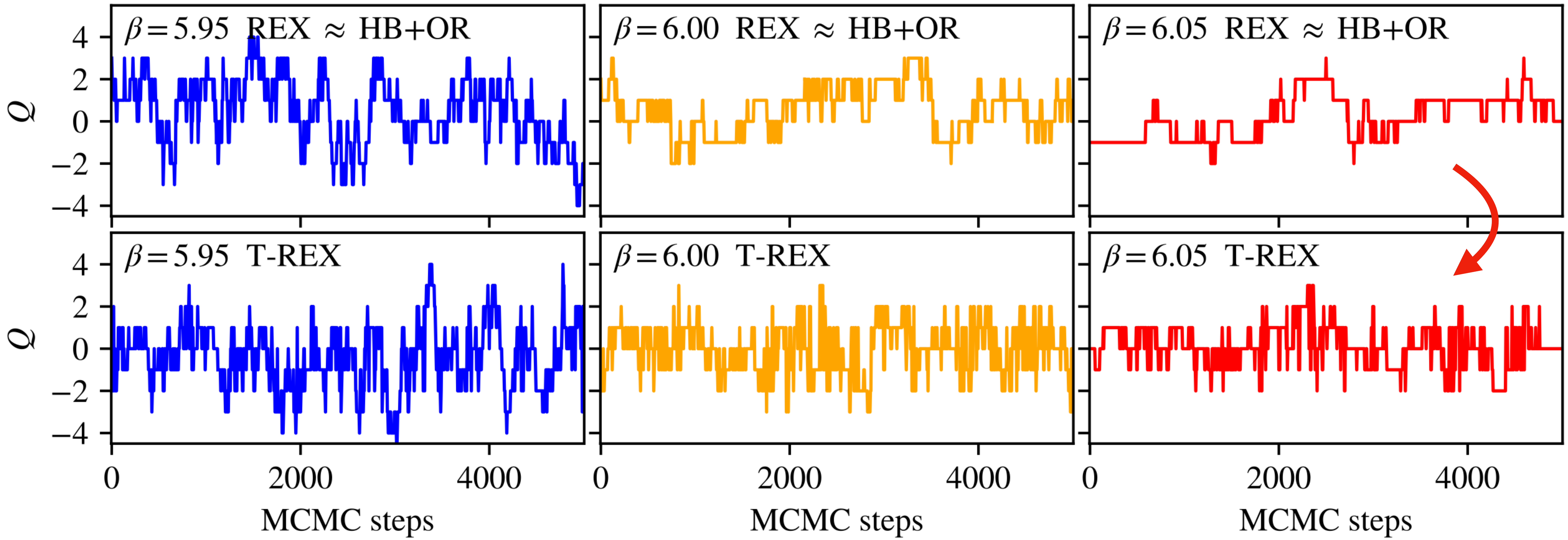
$$
\begin{tikzcd}
\color{blue}{\beta_1 :} & U_1^{(n)} \arrow[r, blue] & U_1^{(n+1)} \arrow[dr, blue] & U_1^{(n+2)} = f(U_0^{(n+1)}) \\
\color{red}{\beta_0 :} & U_0^{(n)} \arrow[r, red] & U_0^{(n+1)} \arrow[ur, red] & U_0^{(n+2)} = f^{-1}(U_1^{(n+1)})
\end{tikzcd}
$$

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

Example

Topology mixing slows down

$V = 12^4$

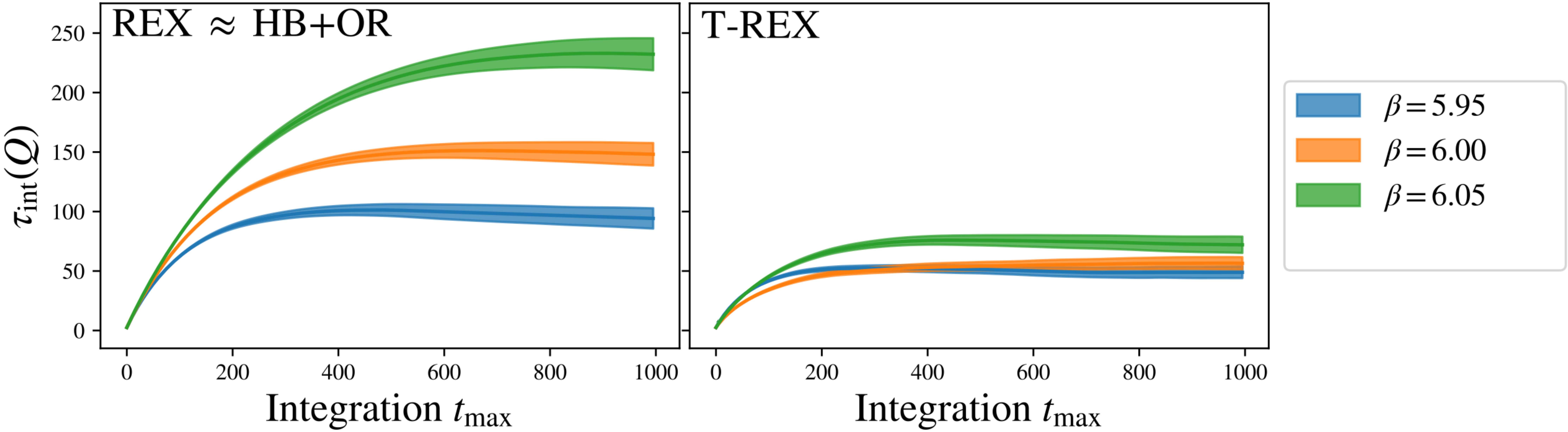


Faster topology mixing

Fernando Romero-López, MIT

MCMC history

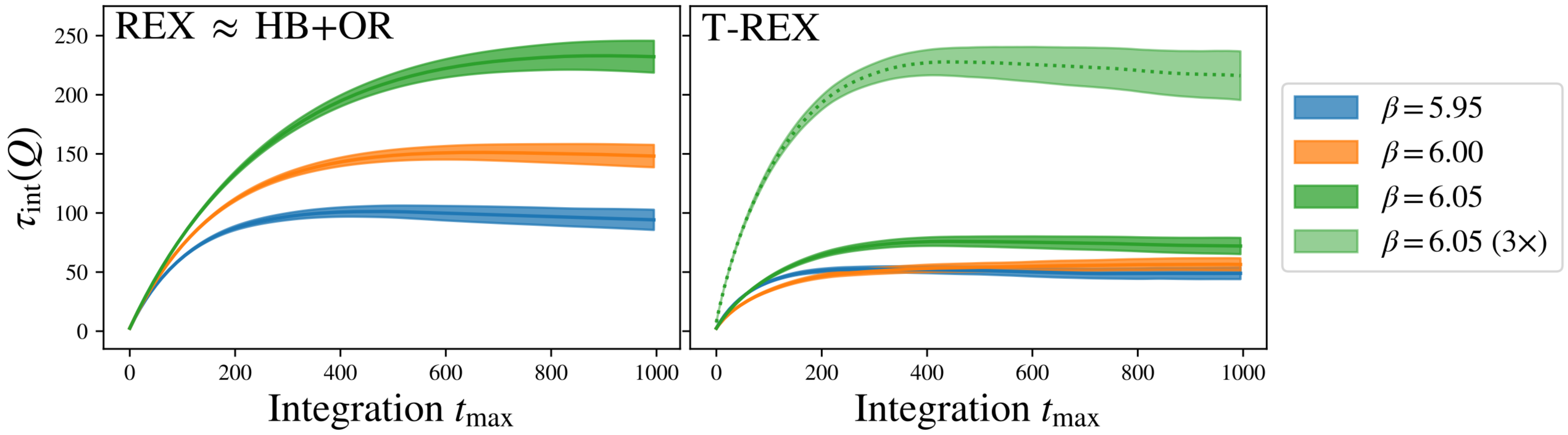
- All integrated autocorrelation times reduce significantly



Fernando Romero-López, MIT

MCMC history

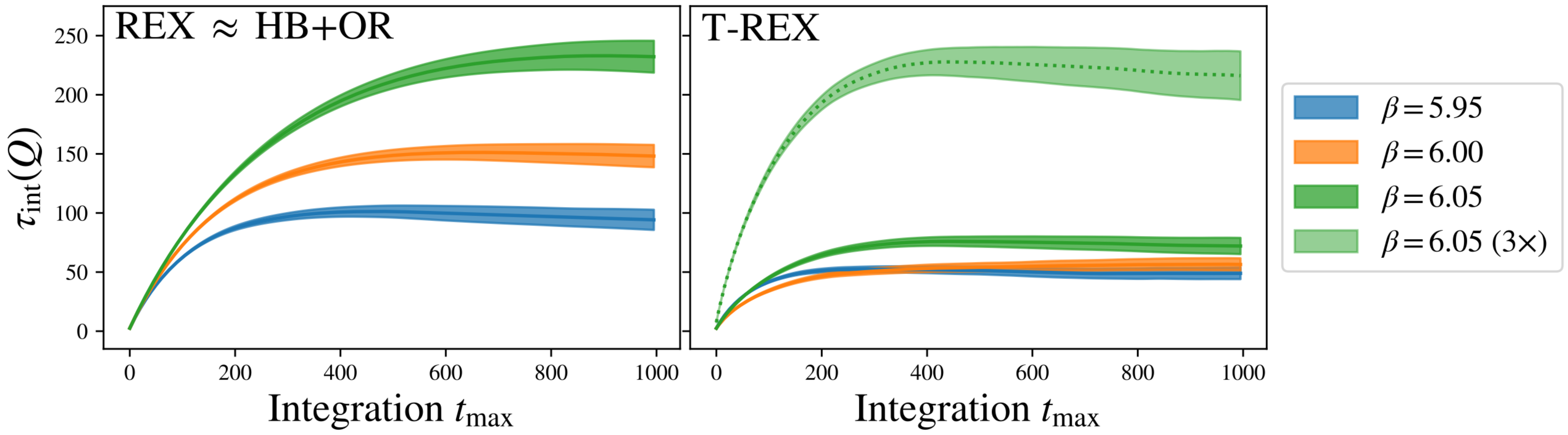
- All integrated autocorrelation times reduce significantly



Fernando Romero-López, MIT

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

Fernando Romero-López, MIT

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, MIT

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?

Fernando Romero-López, MIT

Back-up

Lattice Field Theory

- LFT is the first-principles treatment of the generic QFT
  - Path integral

$$\mathcal{Z} = \int D\phi \, e^{-iS(\phi)}$$

Lattice Field Theory

- LFT is the first-principles treatment of the generic QFT

  - Path integral in Euclidean or imaginary time: statistical meaning

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

Boltzmann factor

Euclidean action

Lattice Field Theory

- LFT is the first-principles treatment of the generic QFT

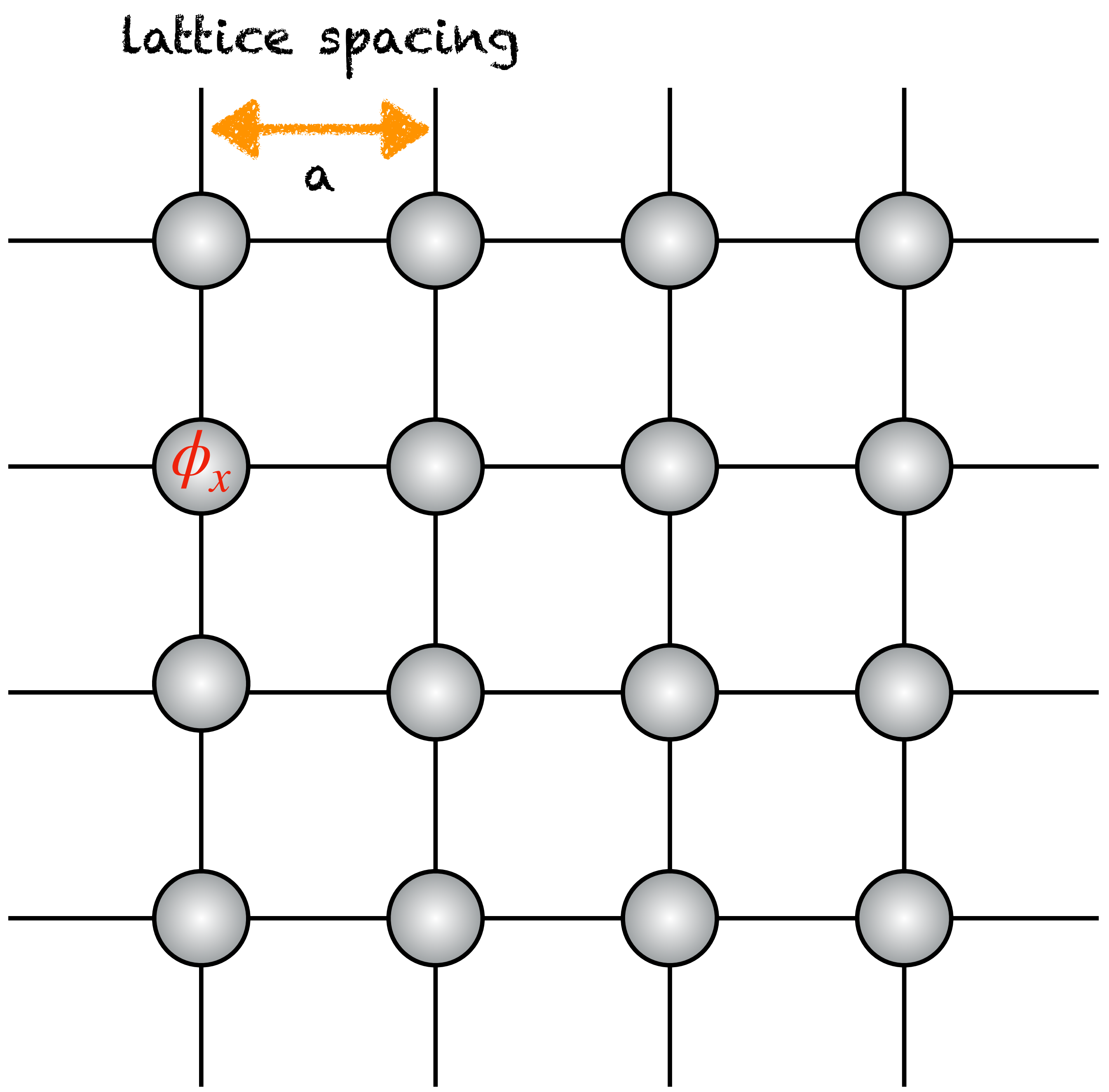
  - Path integral in Euclidean or imaginary time: statistical meaning

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

Boltzmann factor

Euclidean action

  - Discretize quantum fields (real scalars):



Continuum: $$S_E = \int d^4x \left[\frac{1}{2} \partial_\mu \phi(x)^2 + \frac{m^2}{2} \phi(x)^2 + \lambda \phi(x)^4 \right]$$

Lattice: $$S_E = a^4 \sum_x \frac{1}{2a^2} (\phi_{x+\mu} - \phi_x)^2 + \frac{m^2}{2} \phi_x^2 + \lambda \phi_x^4$$

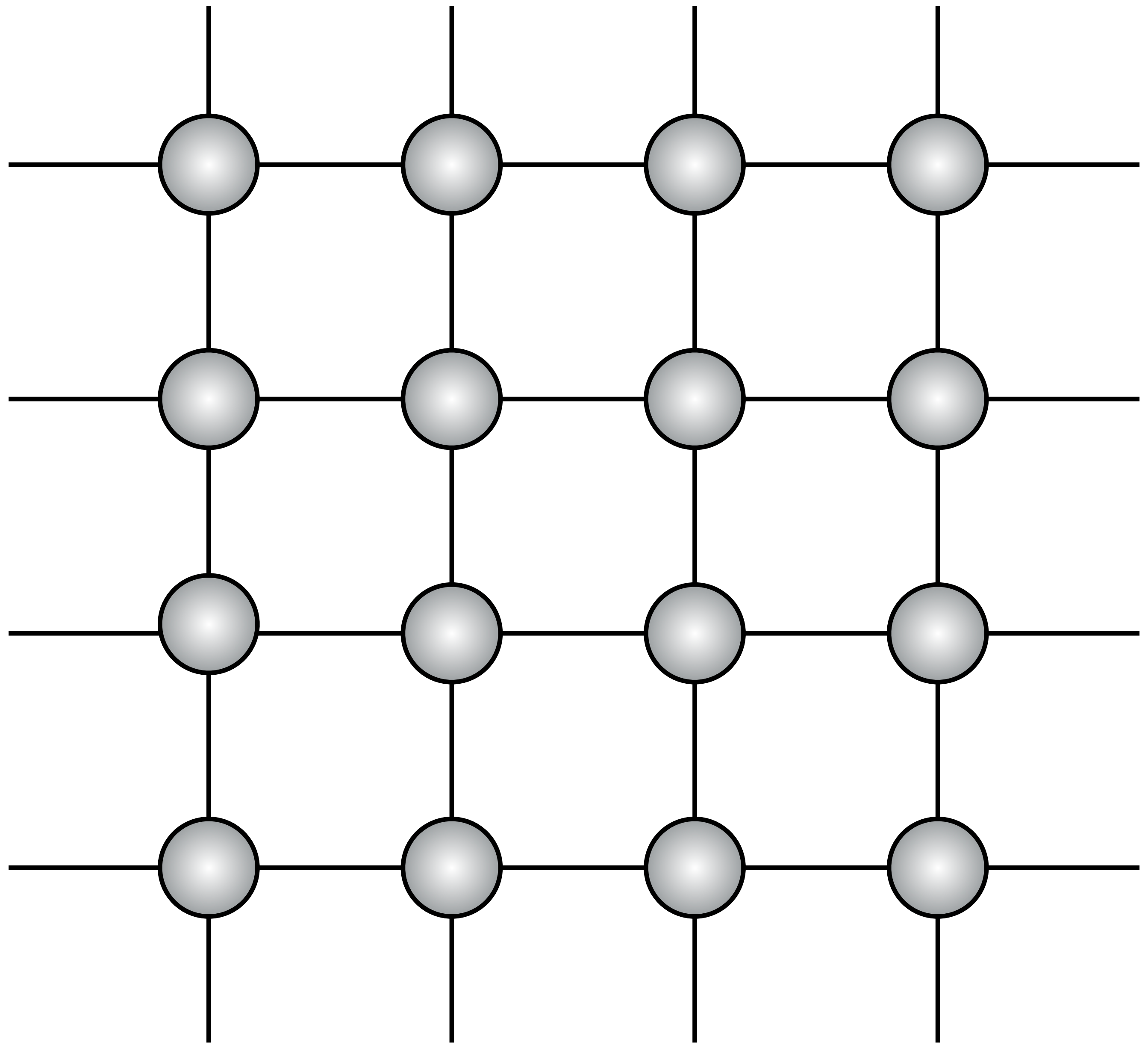
Lattice QCD

- LFT applied to QCD can be used to solve the dynamics of the strong interaction at hadronic energies

  - Lattice QCD partition function

$$\mathcal{Z} = \int D\psi\, D\bar{\psi}\, DA\, e^{-S_E(\psi,\bar{\psi},A_\mu)}$$

Boltzmann factor



Lattice QCD

- LFT applied to QCD can be used to solve the dynamics of the strong interaction at hadronic energies

  - Lattice QCD partition function

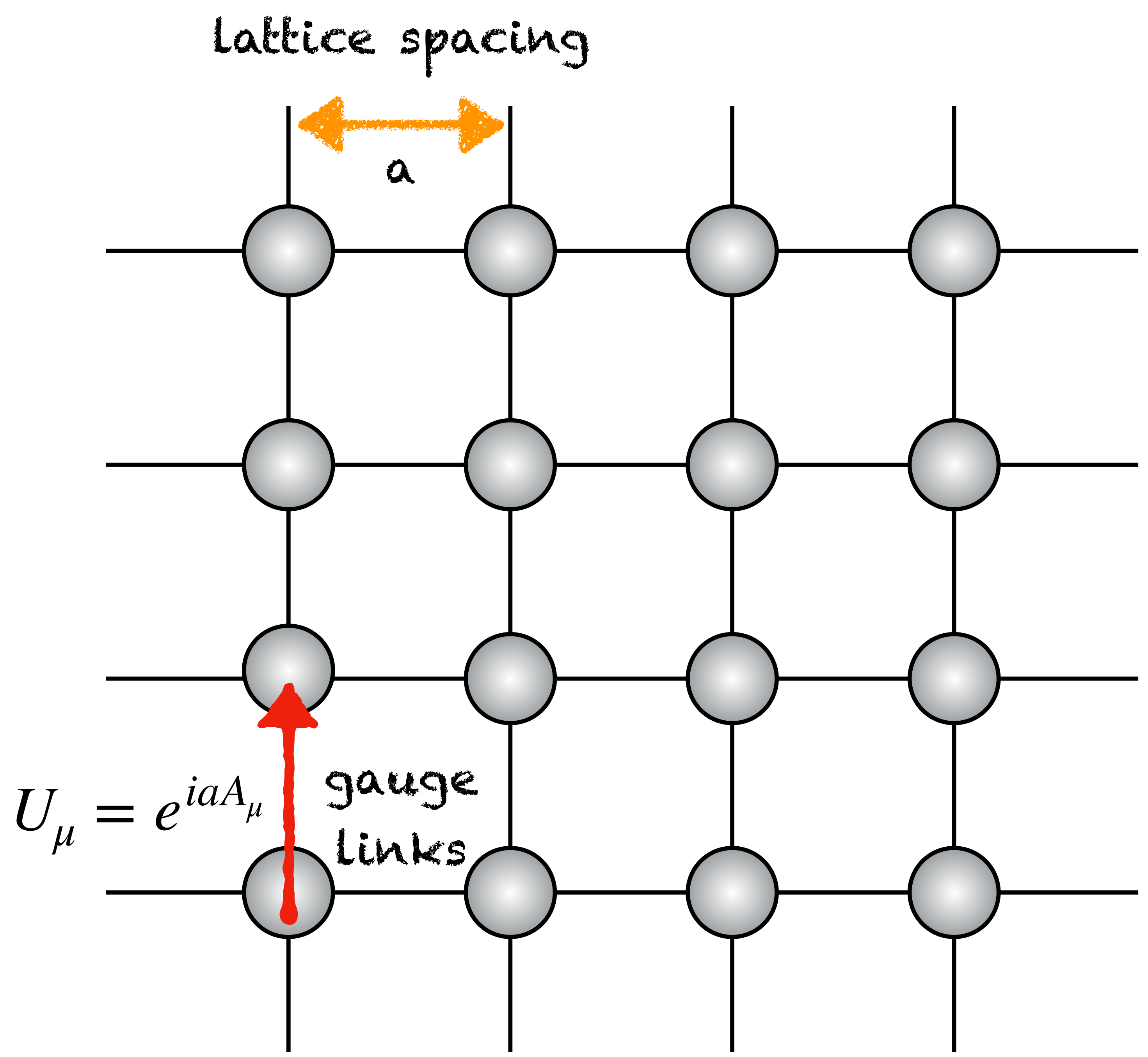
$$\mathcal{Z} = \int D\psi\, D\bar{\psi}\, DA\, e^{-S_E(\psi,\bar{\psi},A_\mu)}$$

Boltzmann factor

  - Discretize gauge fields and fermion fields:

    →Under control but technical

    →Need continuum limit



Lattice QCD

- LFT applied to QCD can be used to solve the dynamics of the strong interaction at hadronic energies

- Lattice QCD partition function

$$\mathcal{Z} = \int D\psi D\bar{\psi} DA e^{-S_E(\psi,\bar{\psi},A_\mu)}$$

Boltzmann factor

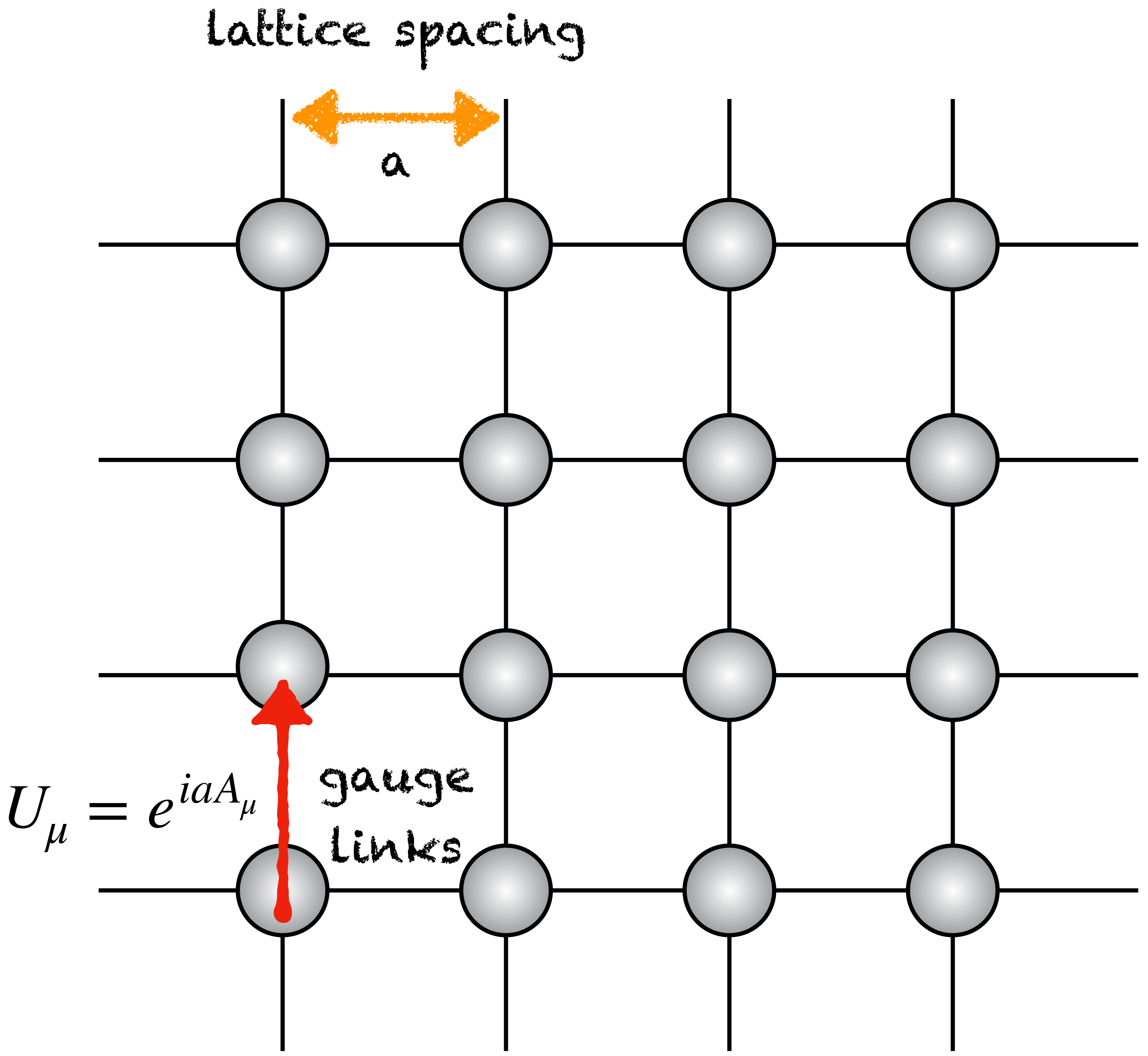
- Discretize gauge fields and fermion fields:

  →Under control but technical

  →Need continuum limit

- Compute observables as expectation values

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int D\psi D\bar{\psi} DA\, \mathcal{O}\, e^{-S(\psi,\bar{\psi},A_\mu)}$$



An integral over many variables:
$2 \times 3^2 \times 4 \times L^4 \simeq 10^{10}$

Fernando Romero-López, MIT

Markov Chain Monte Carlo

Lattice QCD is sampling problem over a very large number of variables

Markov Chain Monte Carlo

Lattice QCD is sampling problem over a very large number of variables

→**Generate field configurations:** $$\{U_i\} \sim p(U) = \frac{e^{-S_E(U)}}{\mathcal{Z}}$$

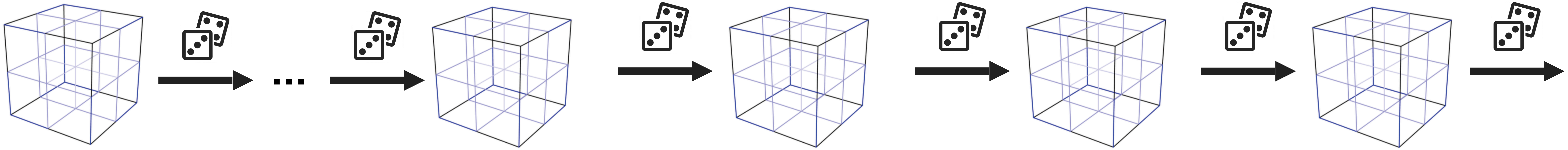
→**Compute observables:** $$\langle \mathcal{O} \rangle \simeq \sum_{i=0}^{N_{conf}} O(U_i)$$

Markov Chain Monte Carlo

Lattice QCD is sampling problem over a very large number of variables

→**Generate field configurations:** $\{U_i\} \sim p(U) = \dfrac{e^{-S_E(U)}}{\mathcal{Z}}$

→**Compute observables:** $\langle \mathcal{O} \rangle \simeq \sum_{i=0}^{N_{conf}} O(U_i)$



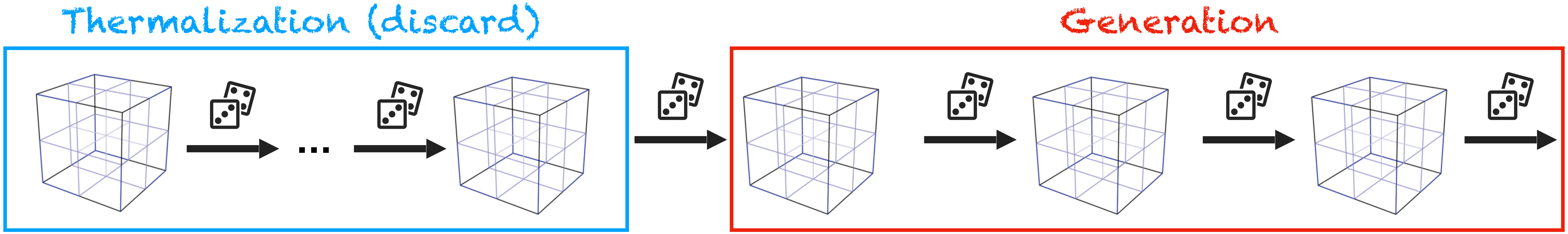
Fernando Romero-López, MIT

Markov Chain Monte Carlo

Lattice QCD is sampling problem over a very large number of variables

→Generate field configurations: $\{U_i\} \sim p(U) = \frac{e^{-S_E(U)}}{\mathcal{Z}}$

→Compute observables: $\langle \mathcal{O} \rangle \simeq \sum_{i=0}^{N_{conf}} O(U_i)$



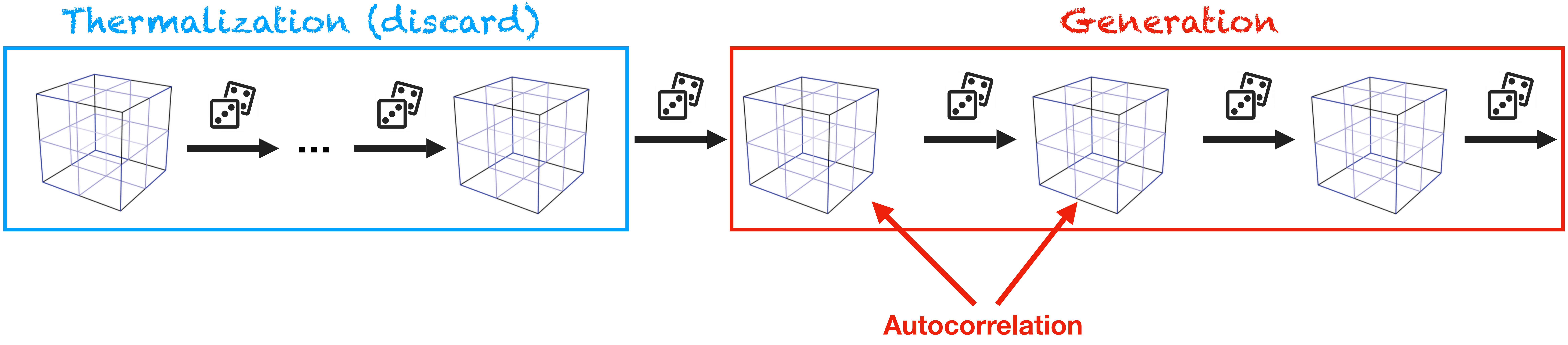
Fernando Romero-López, MIT

Markov Chain Monte Carlo

Lattice QCD is sampling problem over a very large number of variables

→**Generate field configurations:** $\{U_i\} \sim p(U) = \dfrac{e^{-S_E(U)}}{\mathcal{Z}}$

→**Compute observables:** $\langle \mathcal{O} \rangle \simeq \sum_{i=0}^{N_{conf}} O(U_i)$



Fernando Romero-López, MIT

Markov Chain Monte Carlo

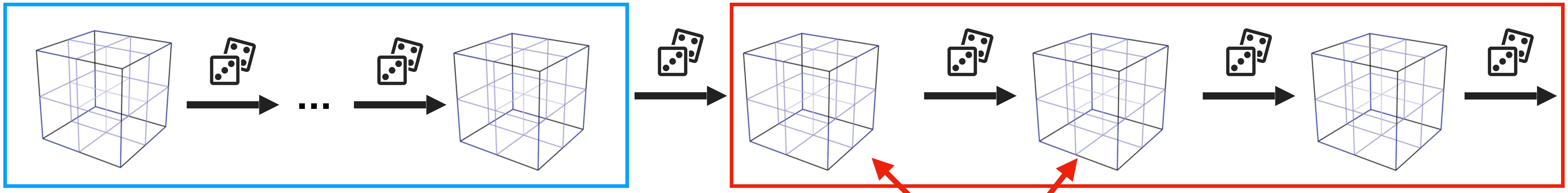
Lattice QCD is sampling problem over a very large number of variables

→Generate field configurations: $\{U_i\} \sim p(U) = \dfrac{e^{-S_E(U)}}{\mathcal{Z}}$

→Compute observables: $\langle \mathcal{O} \rangle \simeq \sum_{i=0}^{N_{conf}} O(U_i)$

Thermalization (discard)

Generation



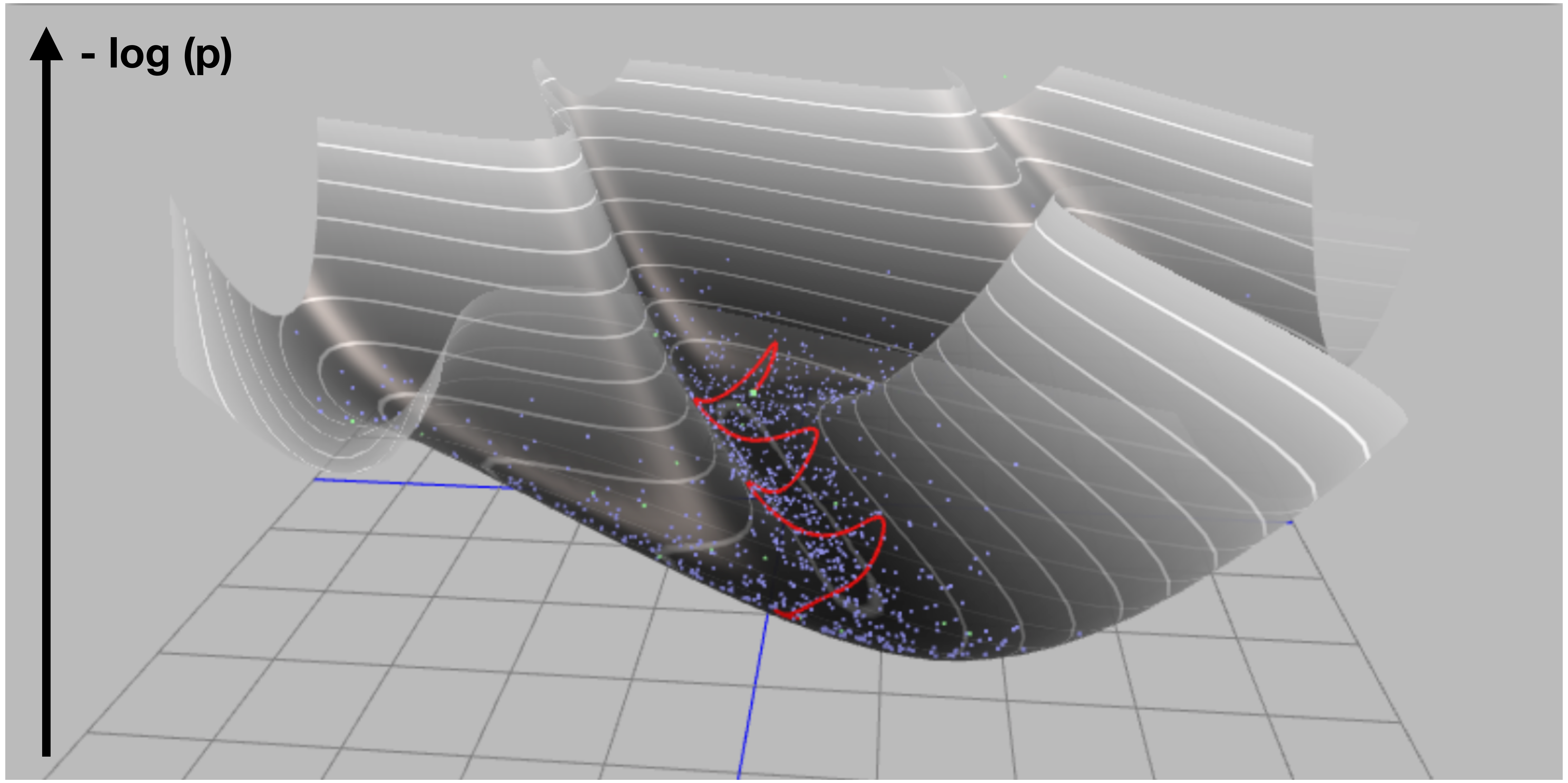
Autocorrelation $\equiv$ Computational cost

Fernando Romero-López, MIT

Generation of gauge configs

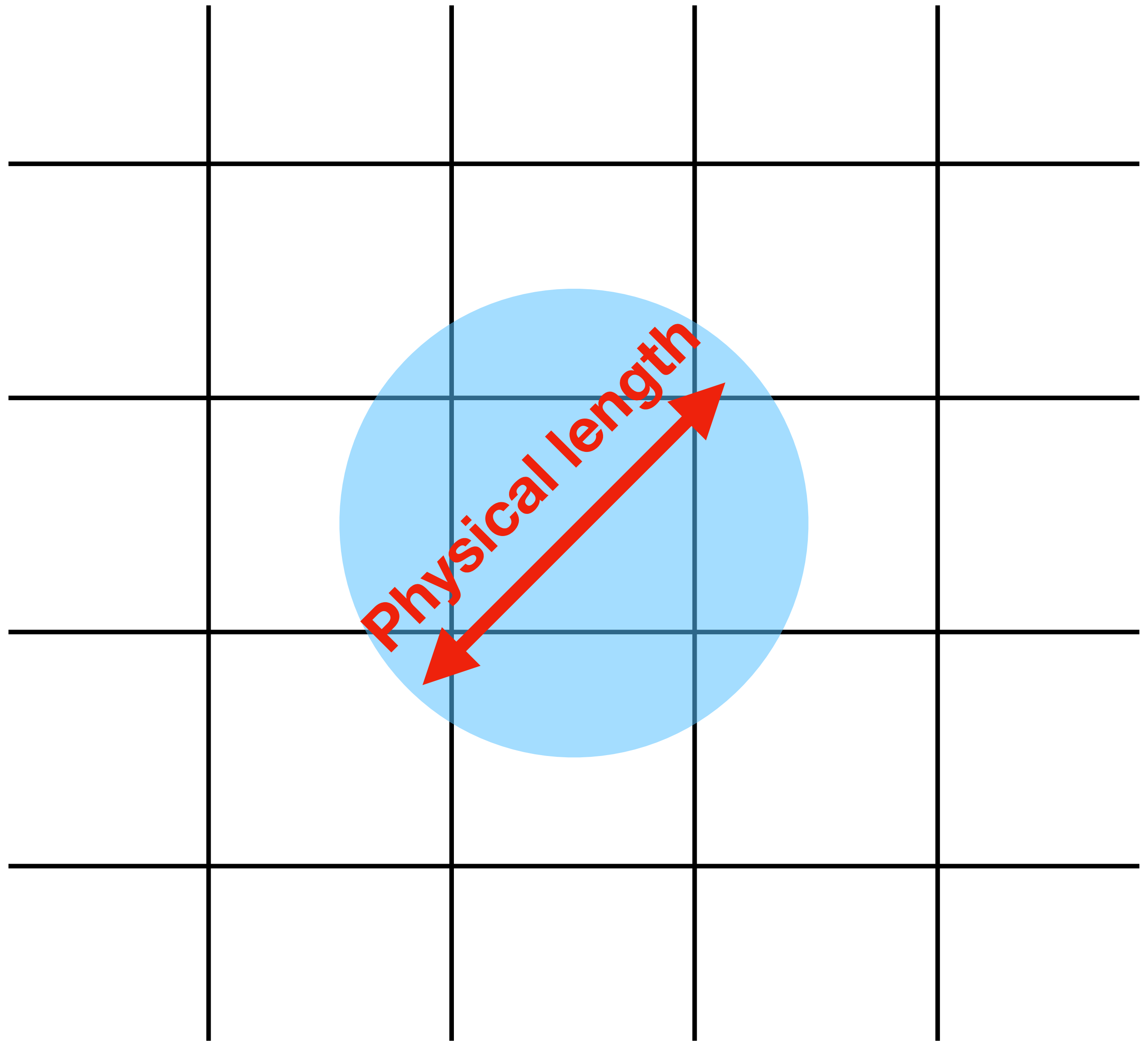
Hybrid/Hamiltonian Monte Carlo (HMC)

Molecular dynamics + Markov Chain Monte Carlo



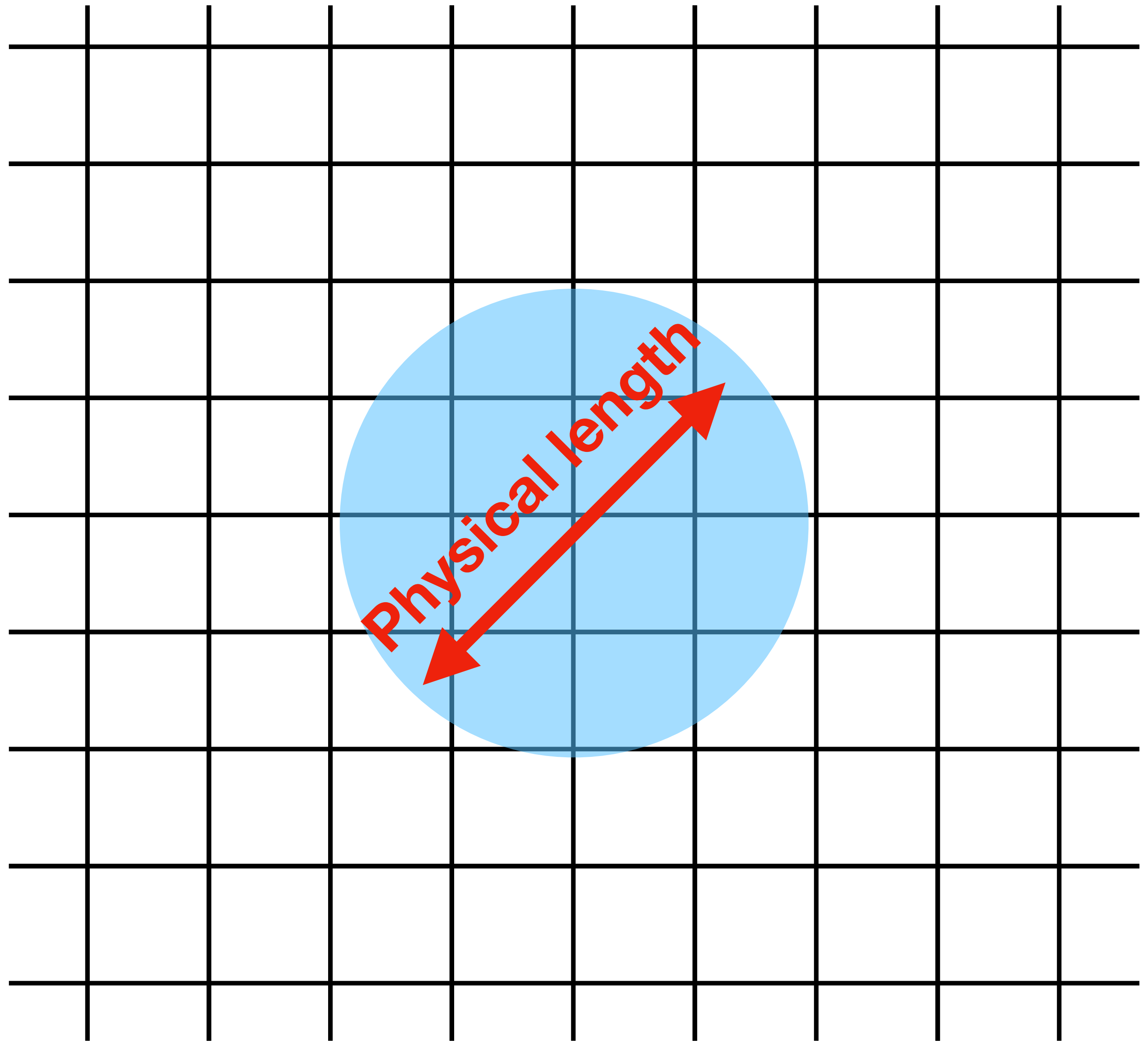
The challenging continuum limit

- Remove discretization effects by taking the continuum limit.



The challenging continuum limit

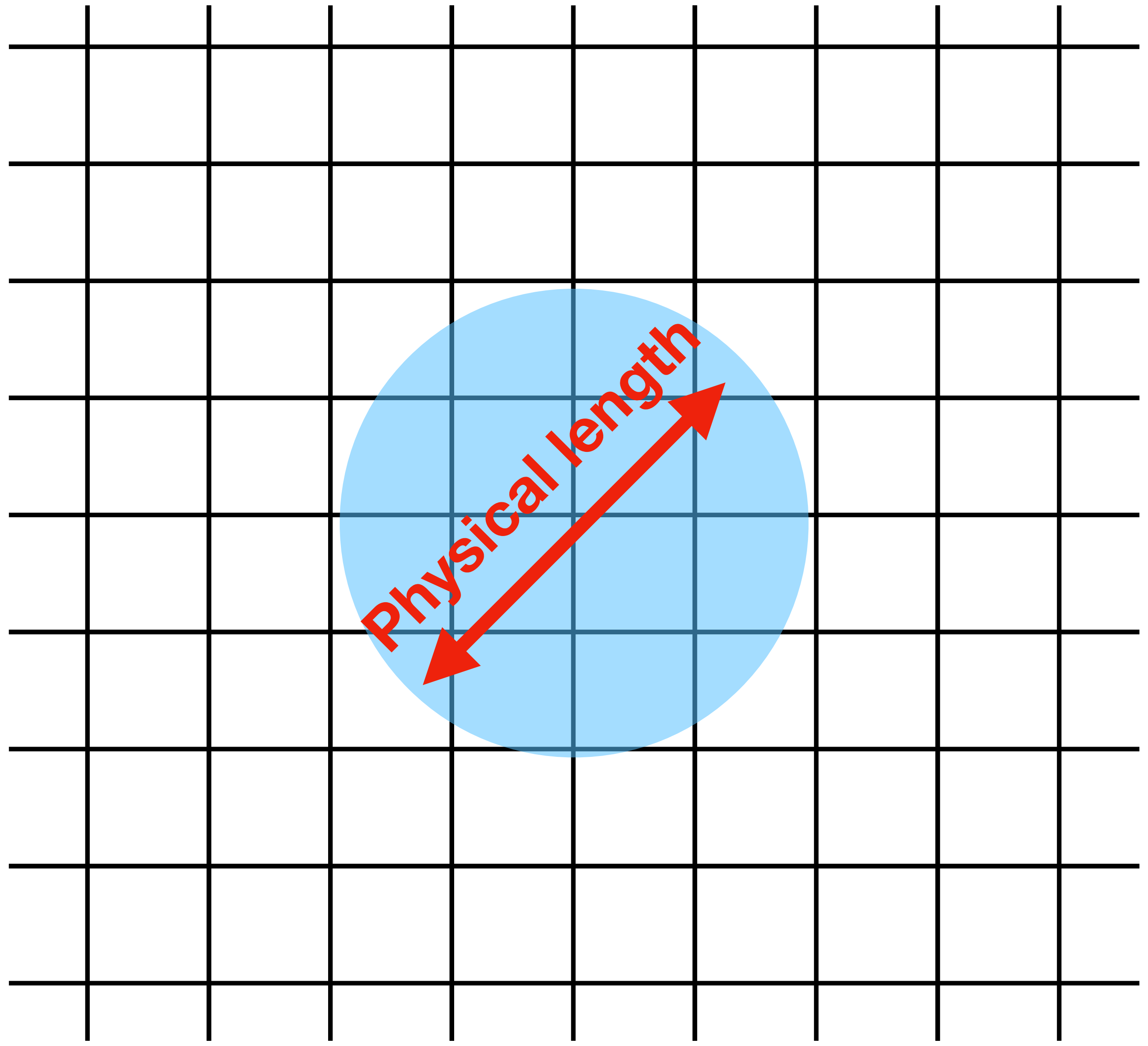
- Remove discretization effects by taking the continuum limit.



Lattice spacing $\longrightarrow$ 0

The challenging continuum limit

- Remove discretization effects by taking the continuum limit.

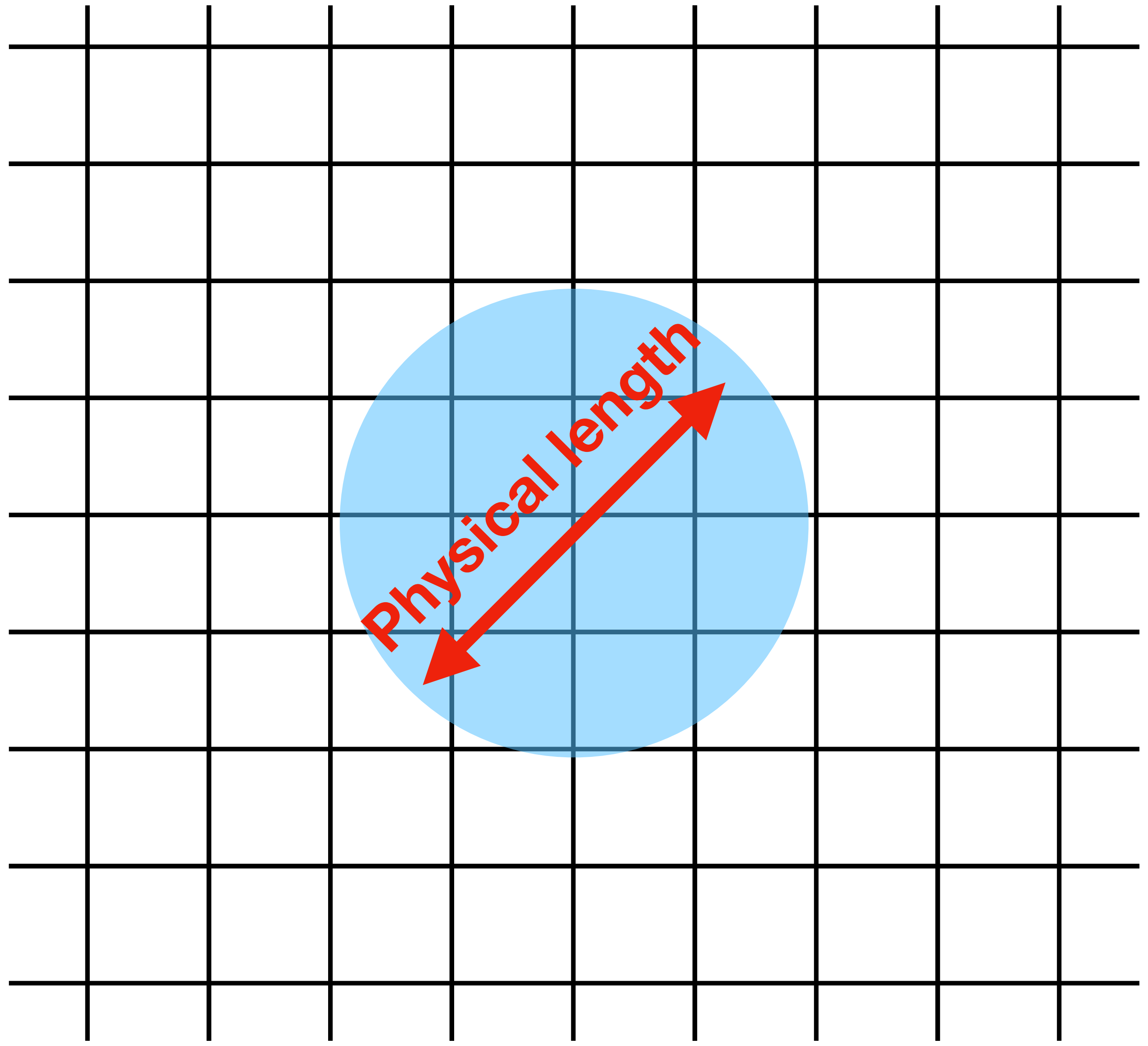


Lattice spacing ⟶ 0

Number of updates to change fixed physical length scale ⟶ ∞

The challenging continuum limit

- Remove discretization effects by taking the continuum limit.



Lattice spacing $\longrightarrow 0$

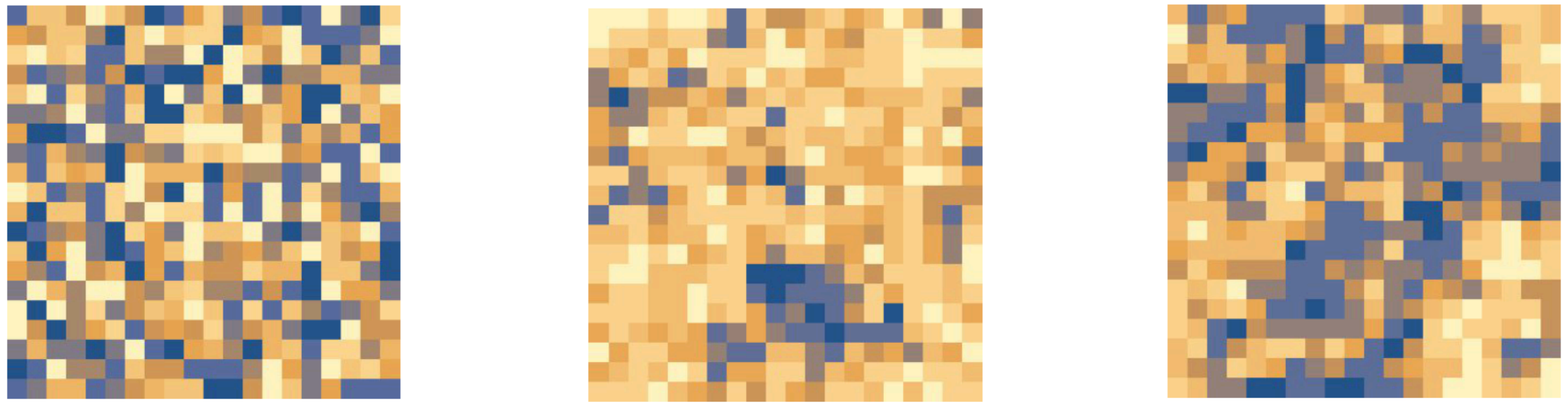
Number of updates to change fixed physical length scale $\longrightarrow \infty$

"Critical slowing-down"
of generation of uncorrelated samples

Fernando Romero-López, MIT

Test case: scalar theory

- A real scalar field per lattice site in a 2D lattice. $\phi(x) \in (-\infty, +\infty)$



unlikely (log prob = -6107) | likely (log prob = 22) | likely (log prob = 5)

Field configurations with probability:

$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

- Lattice action:

$$S = a^4 \sum_x \frac{1}{2a^2}(\phi_{x+\mu} - \phi_x)^2 + \frac{m^2}{2}\phi_x^2 + \lambda\phi_x^4$$

Fernando Romero-López, MIT

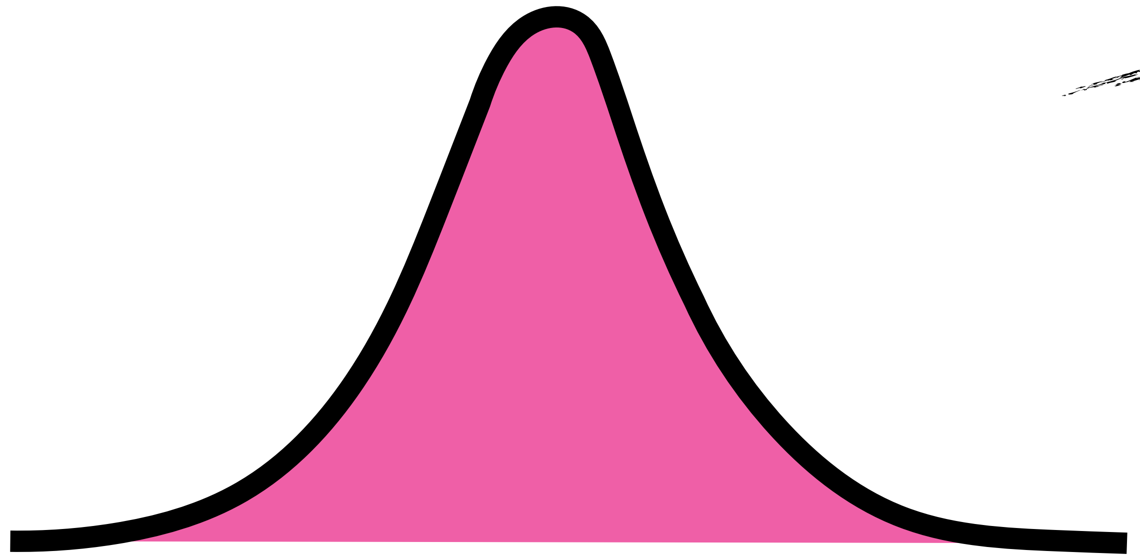
Generative flow models

[Rezende, Mohamed, 1505.05770]

$f(z)$

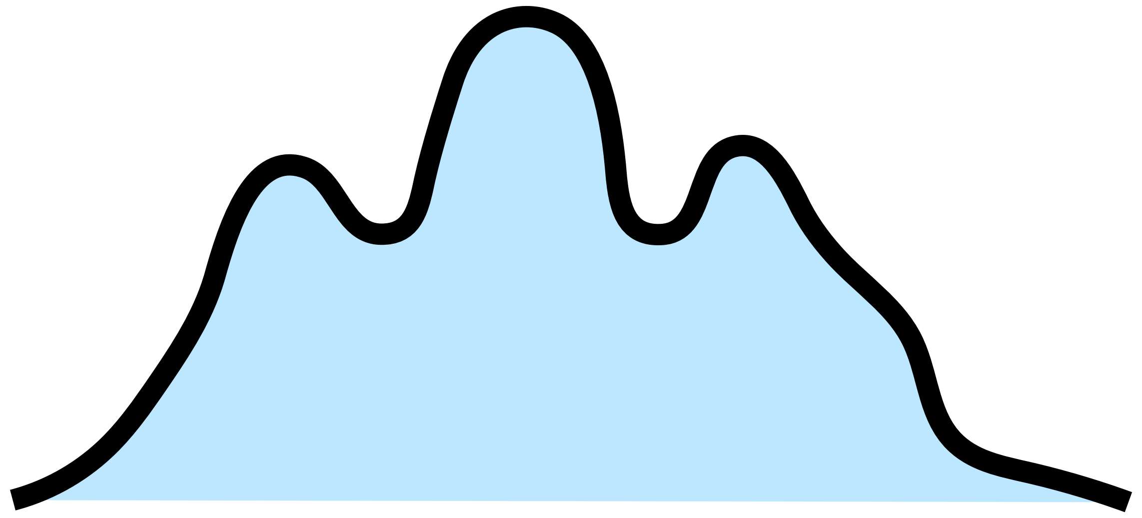
tractable Jacobian
&
invertible

$r(z)$



easily sampled
distribution
(prior)

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



approximates
target distribution
(model)

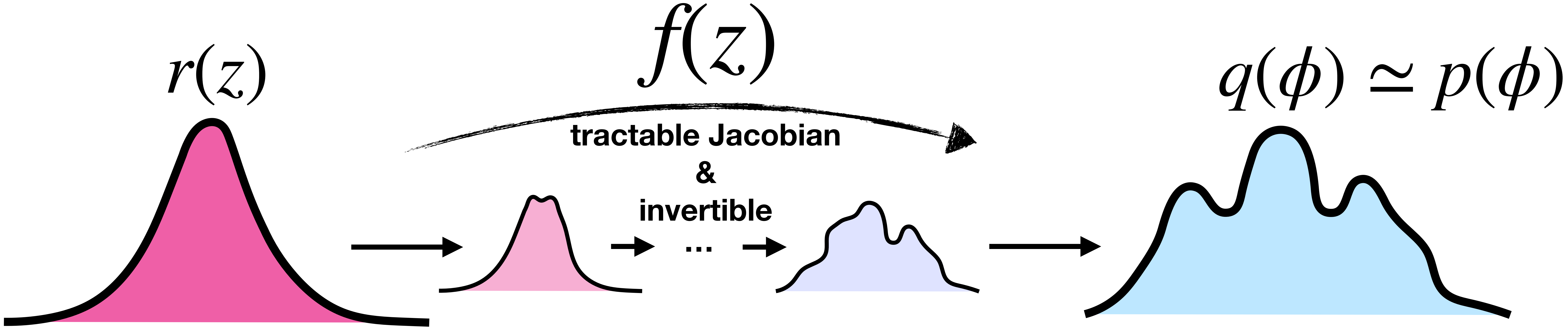
“Trainable change of variables”

Model probability

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

Generative flow models

[Rezende, Mohamed, 1505.05770]



easily sampled distribution (prior)

approximates target distribution (model)

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

"Trainable change of variables"

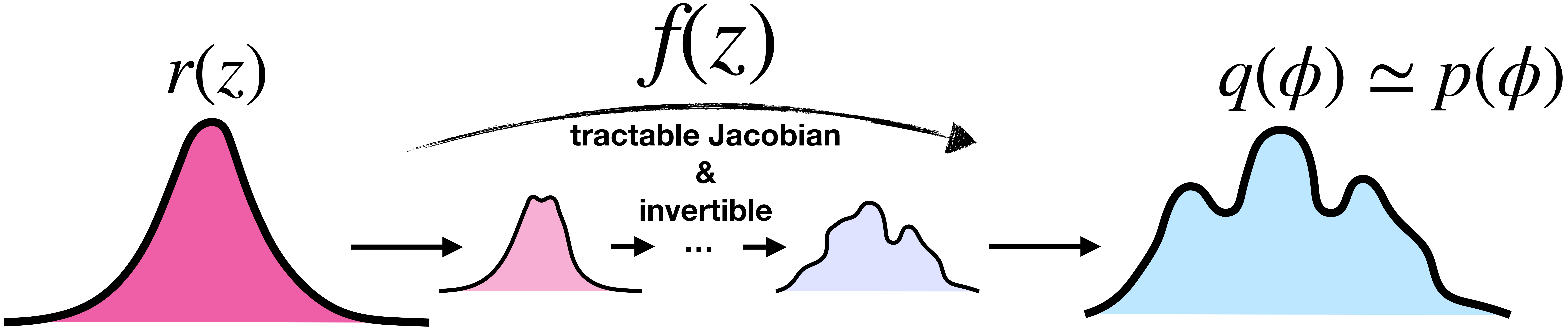
Model probability

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

Fernando Romero-López, MIT

Generative flow models

[Rezende, Mohamed, 1505.05770]



easily sampled
distribution
(prior)

approximates
target distribution
(model)

parametrized by neural networks
(trainable and expressive)

→ composed of
many simple layers

"Trainable change of variables"

Model probability

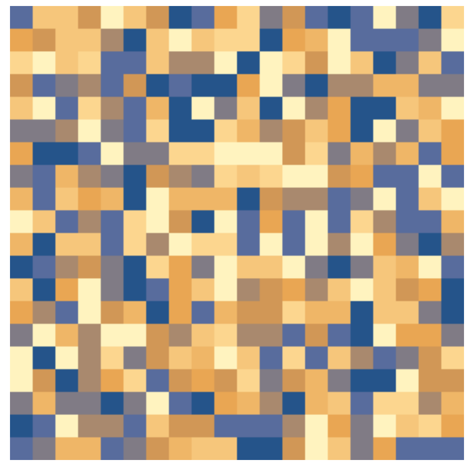
$$q(\phi) = r(z)\left|\det \frac{\partial f(z)}{\partial z}\right|^{-1}$$

❗ Trained models are not perfect.

☑ But exact sampling can be recovered via Markov Chain

Implementing flows for scalar theory

$\phi(x) \in (-\infty, +\infty)$



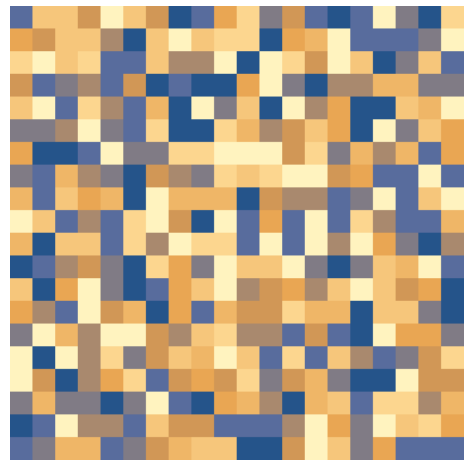
→**Efficient flow models need tractable Jacobians.**

Idea: Each layer acts on a subset of components, conditioned only on the complimentary subset.

Fernando Romero-López, MIT

Implementing flows for scalar theory

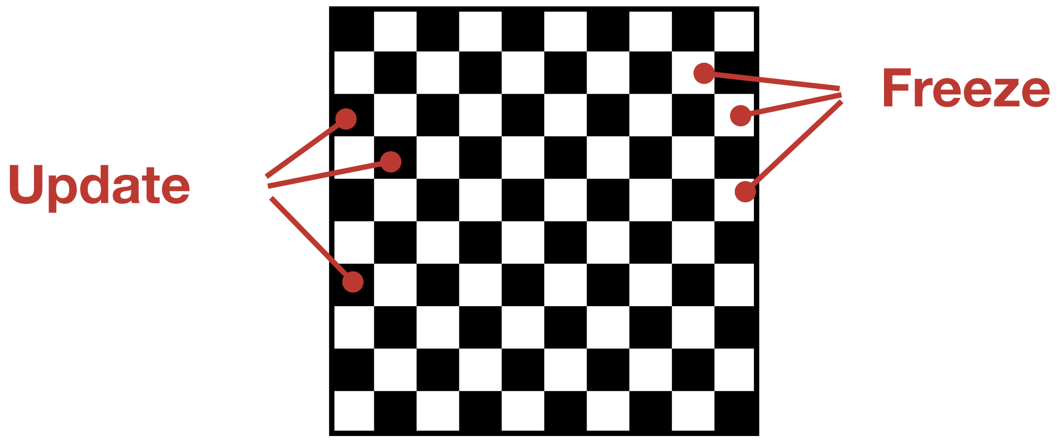
$\phi(x) \in (-\infty, +\infty)$



→**Efficient flow models need tractable Jacobians.**

Idea: Each layer acts on a subset of components, conditioned only on the complimentary subset.

"Masking pattern" m defines subsets.

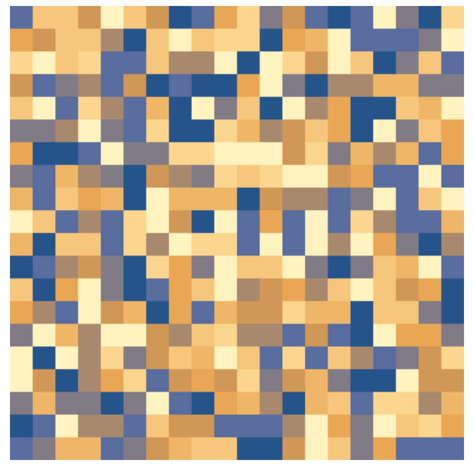


$$\phi' = \begin{cases} \phi'_{\text{frozen}} = \phi_{\text{frozen}} \\ \phi'_{\text{active}} = h(\phi_{\text{frozen}}) \times \phi_{\text{active}} + t \end{cases}$$

conditioning

Implementing flows for scalar theory

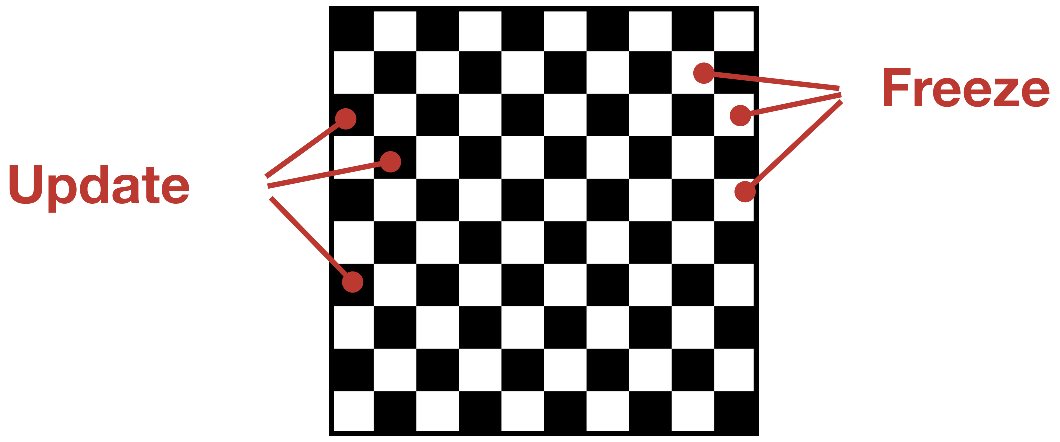
$\phi(x) \in (-\infty, +\infty)$



→**Efficient flow models need tractable Jacobians.**

Idea: Each layer acts on a subset of components, conditioned only on the complimentary subset.

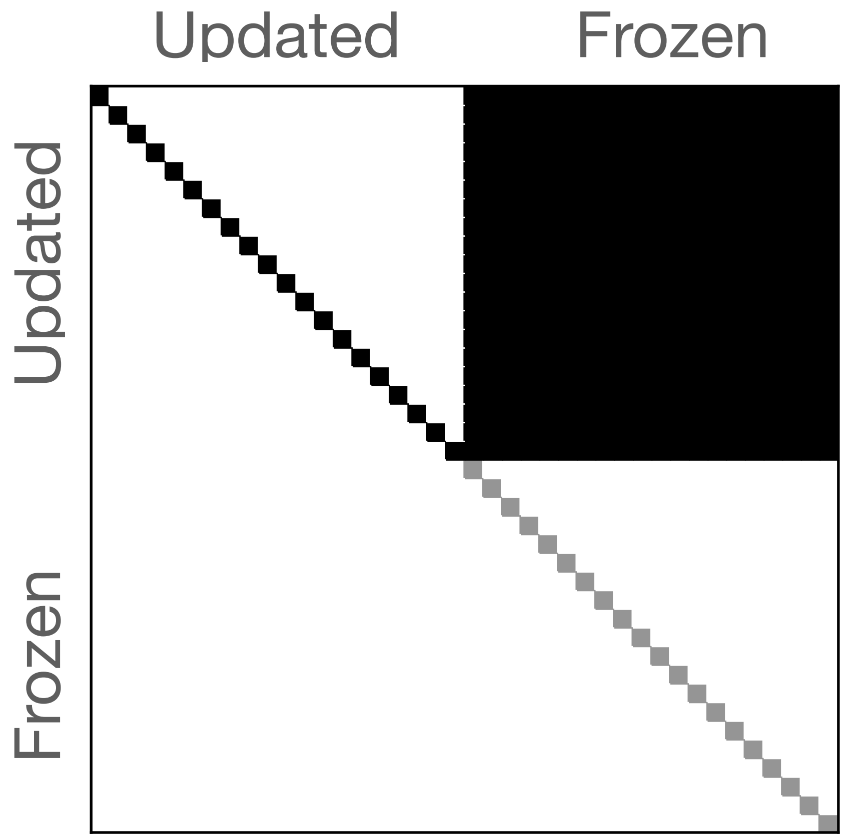
"Masking pattern" m defines subsets.



$$\phi' = \begin{cases} \phi'_{\text{frozen}} = \phi_{\text{frozen}} \\ \phi'_{\text{active}} = h(\phi_{\text{frozen}}) \times \phi_{\text{active}} + t \end{cases}$$

conditioning

Jacobian $= \dfrac{\partial \phi'}{\partial \phi} =$

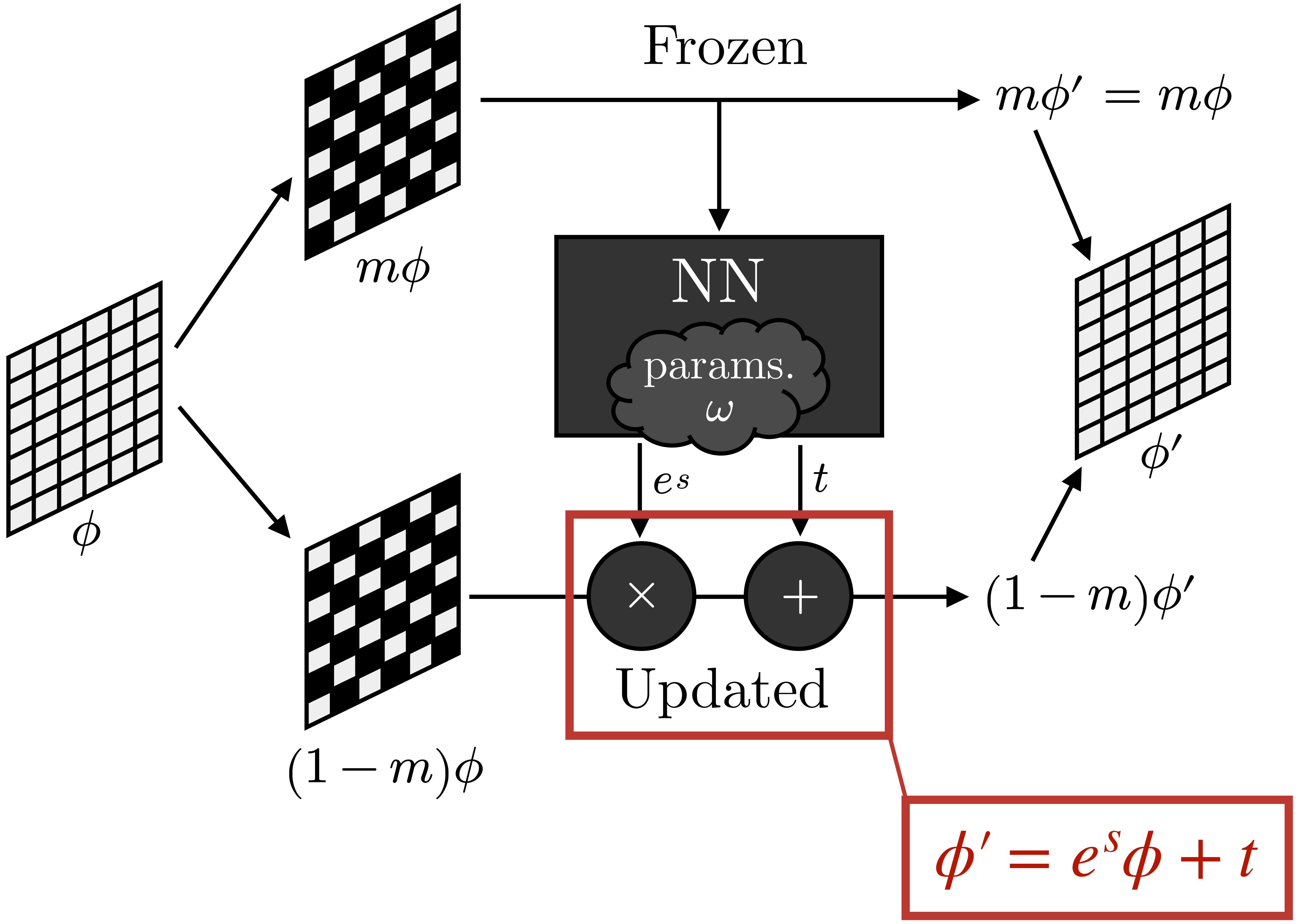


Fernando Romero-López, MIT

Implementing flows for scalar theory

Non-volume preserving coupling layer:

[Dinh, Sohl-Dickstein, Bengio 1605.08803]



Fernando Romero-López, MIT

Training the models

- Target distribution know up to a constant:

$$p[\phi(x)] \sim e^{-S[\phi(x)]}$$

- Train to minimize Kullback-Leibler divergence:

$$D_{\mathrm{KL}}(q\|p) = \sum_{\mathrm{samples}} \log[q(\phi)/p(\phi)]$$

Measures deviation between model and target

Training the models

- Target distribution know up to a constant:

$$p[\phi(x)] \sim e^{-S[\phi(x)]}$$

- Train to minimize Kullback-Leibler divergence:

$$D_{\mathrm{KL}}(q\|p) = \sum_{\text{samples}} \log[q(\phi)/p(\phi)]$$

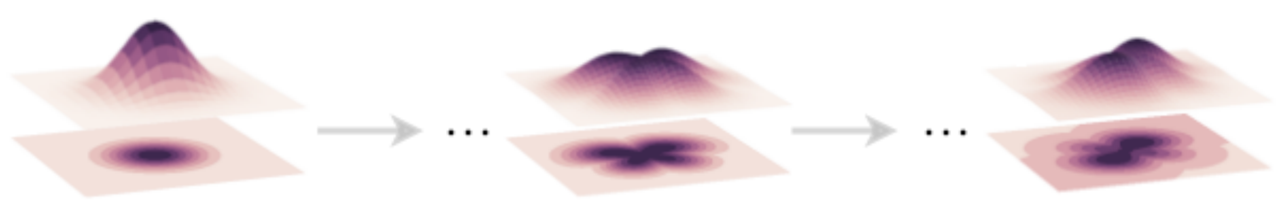
Measures deviation between model and target

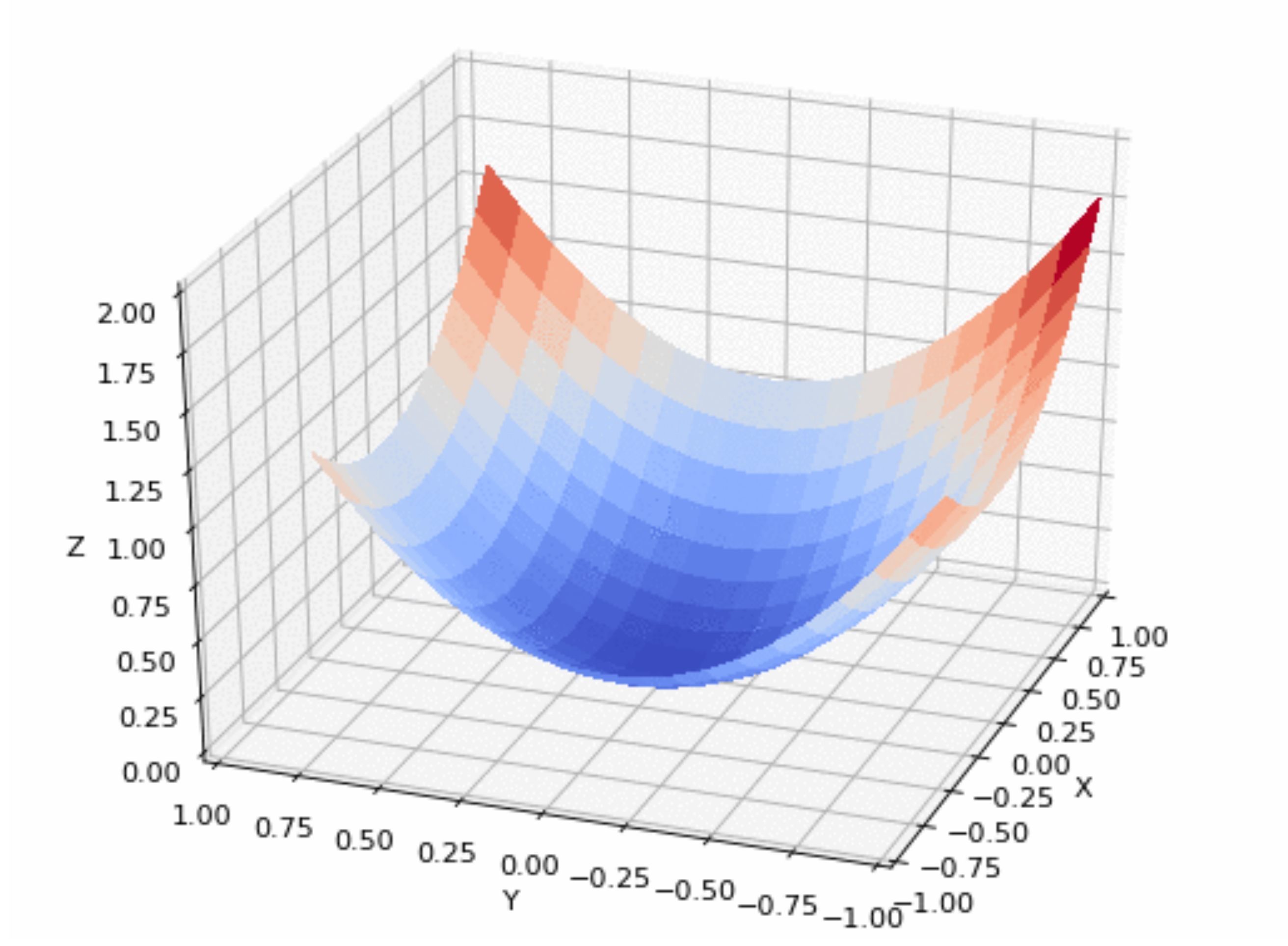
- Self-training:

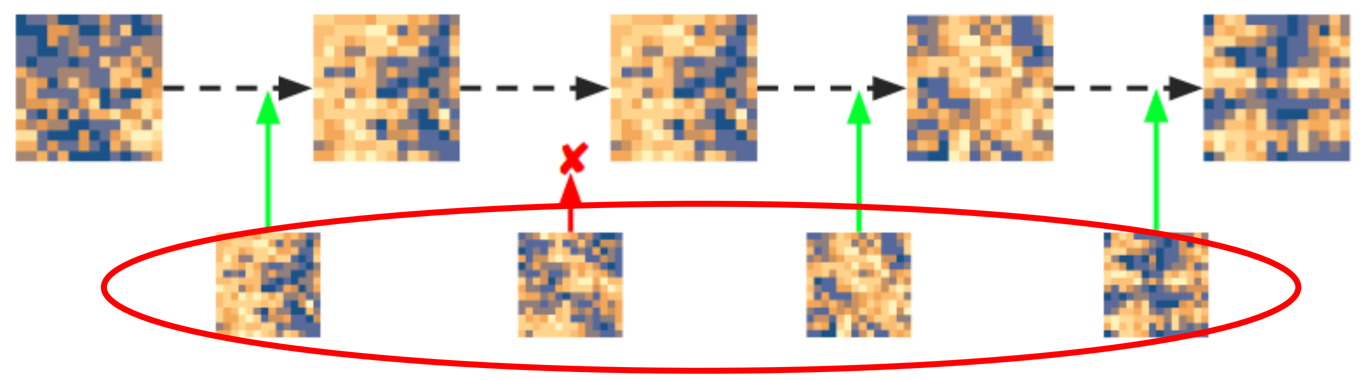
1. Draw samples from the model to measure sample mean of $\log[q(\phi)/p(\phi)]$
2. Gradient-based methods to optimize model parameters (e.g. Adam optimizer)

[Kingma, Ba, arXiv:1412.6980]

Lattice QFT via flow models







generating samples is "embarrassingly parallel"

Parameterize flow using coupling layers

Each layer contains arbitrary neural nets

Training step

Draw samples from model

Compute loss function

Gradient descent

Desired accuracy?

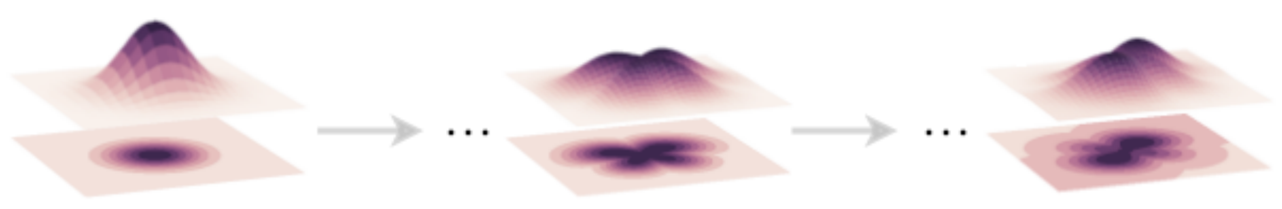
Save trained model

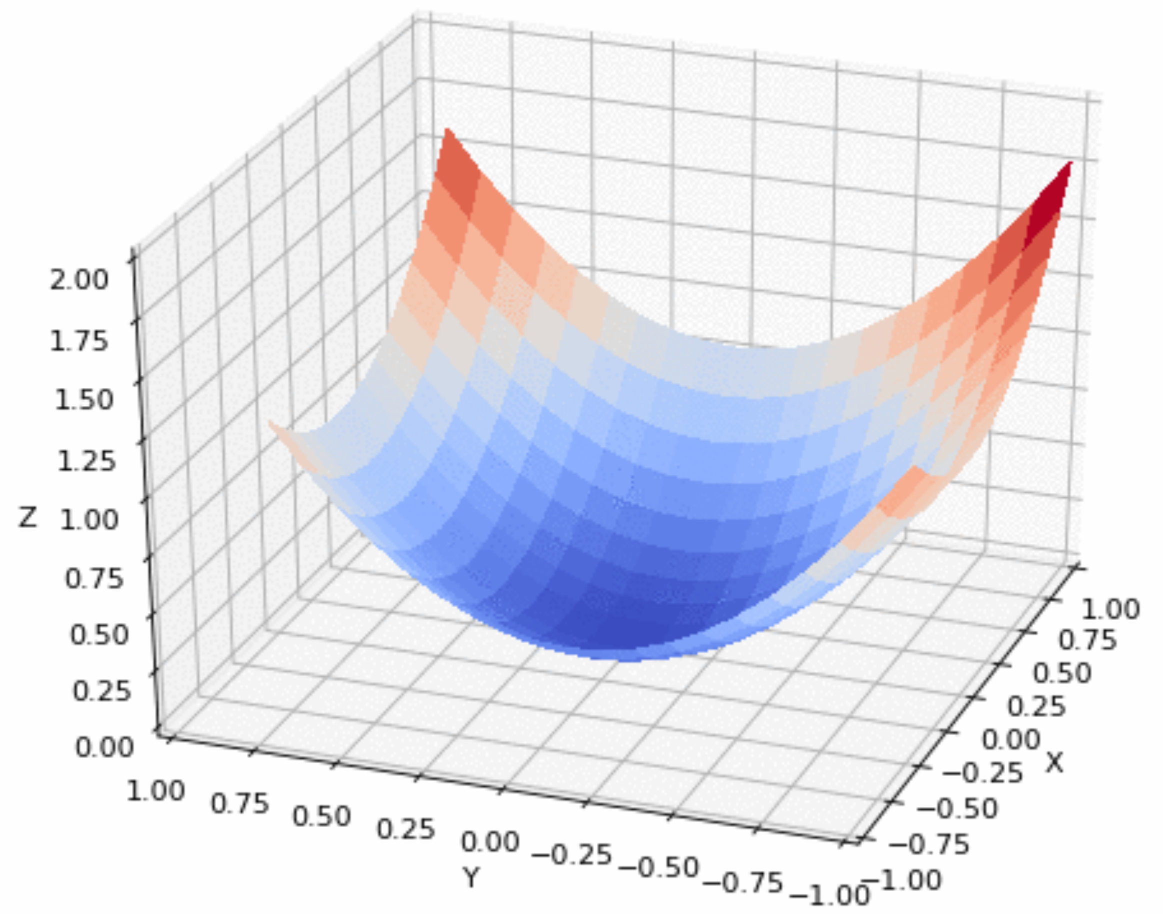
Markov chain using samples from model

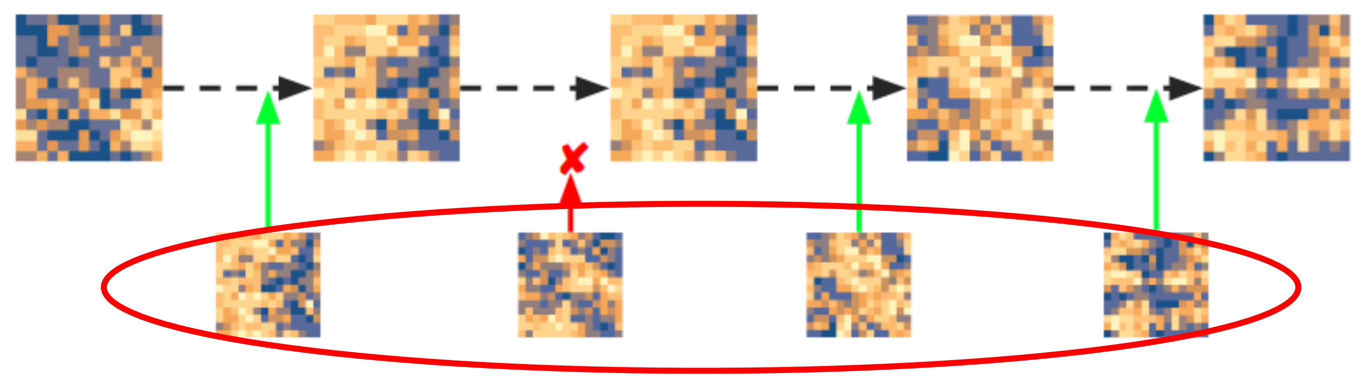
[credits: G. Kanwar]

Fernando Romero-López, MIT

Lattice QFT via flow models







generating samples is "embarrassingly parallel"

Parameterize flow using coupling layers

Each layer contains arbitrary neural nets

Training step

Draw samples from model

Compute loss function

Gradient descent

Desired accuracy?

Save trained model

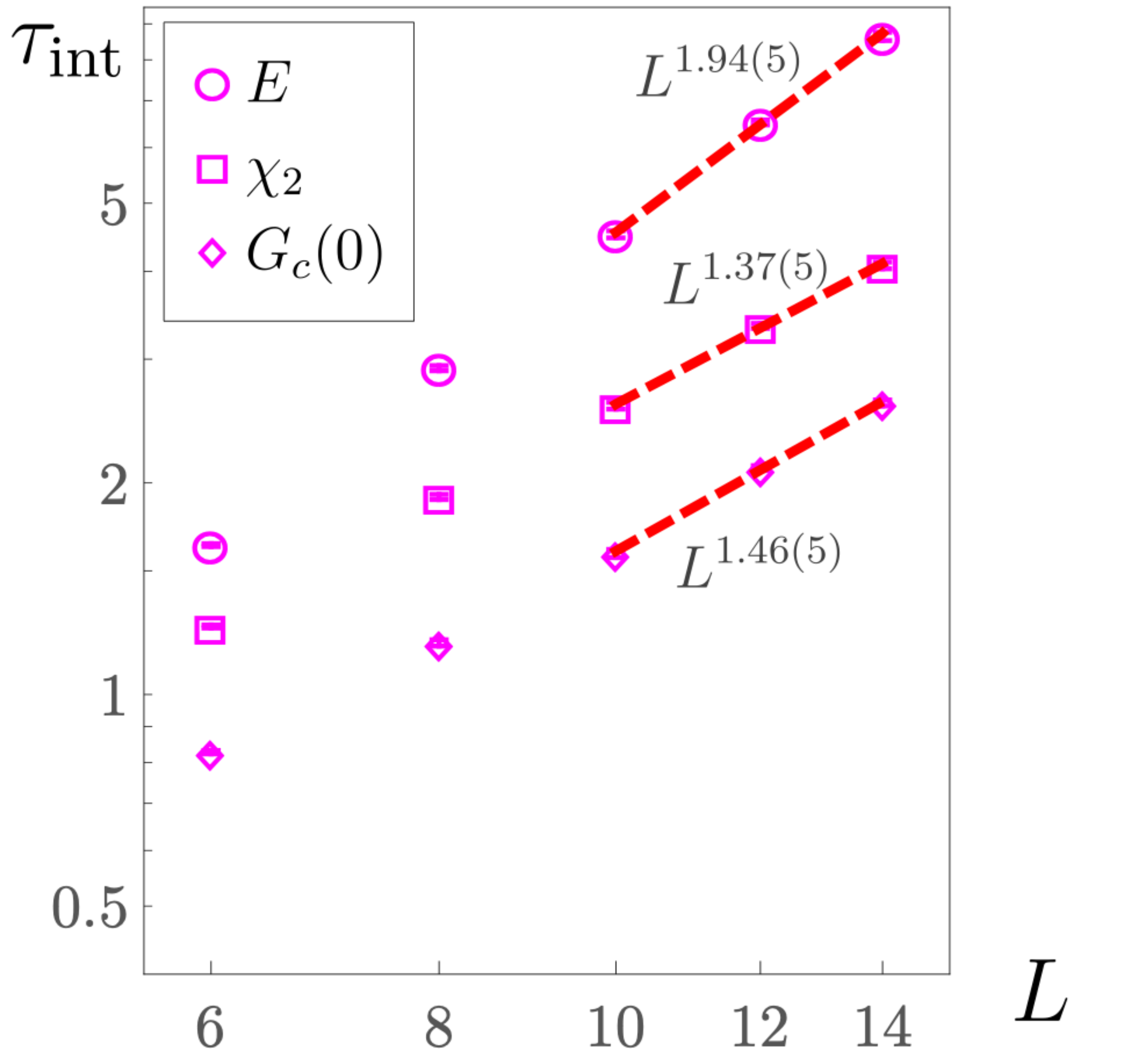
Markov chain using samples from model

[credits: G. Kanwar]

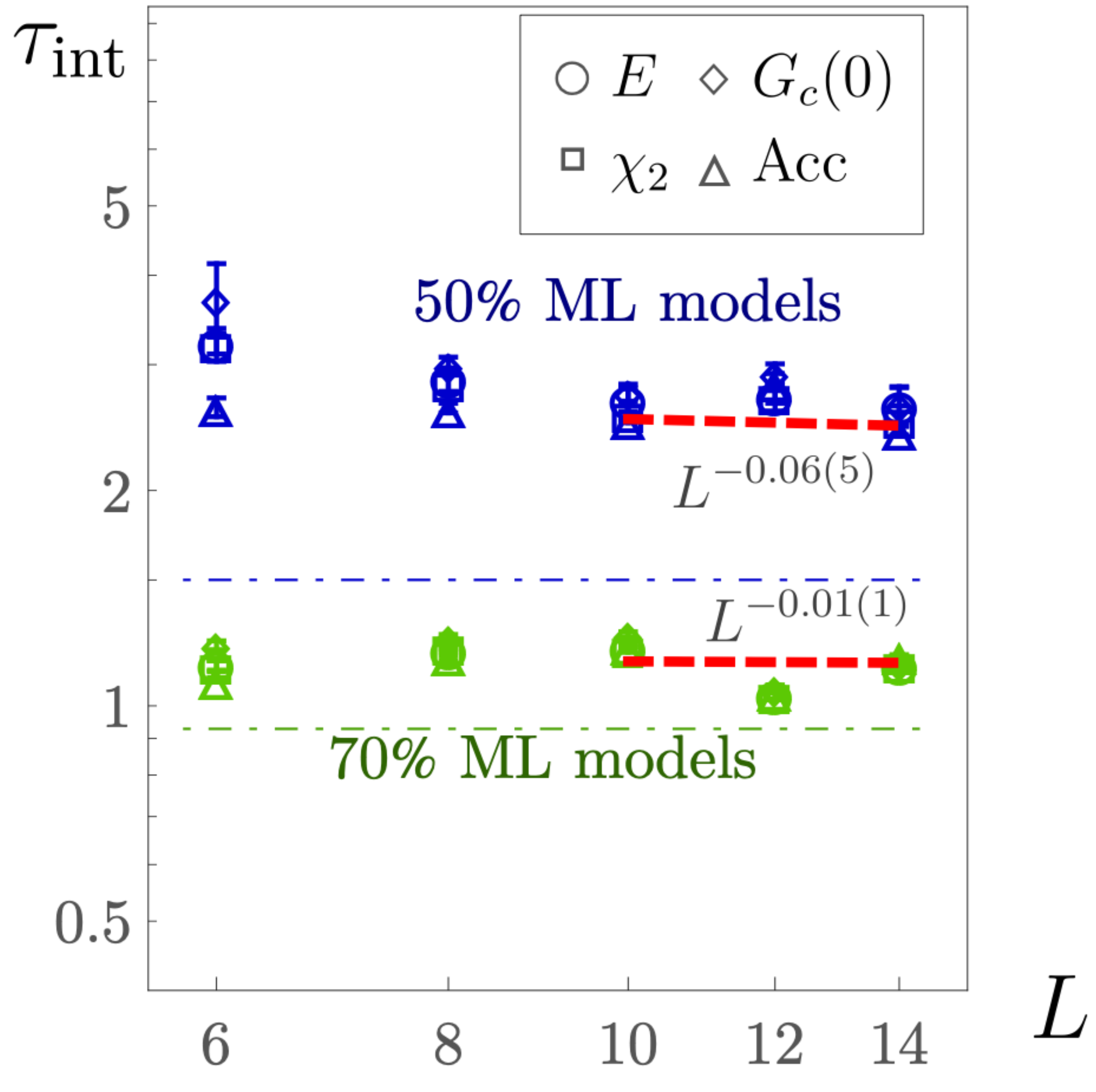
Fernando Romero-López, MIT

Scalar theory

No critical slowing down at the cost of up-front training



(b) Local Metropolis ensembles



(c) Flow-based MCMC ensembles

Conventional approaches slow down

No slow-down!

[Albergo, Kanwar, Shanahan 1904.12072]

Flow-based sampling for gauge theories

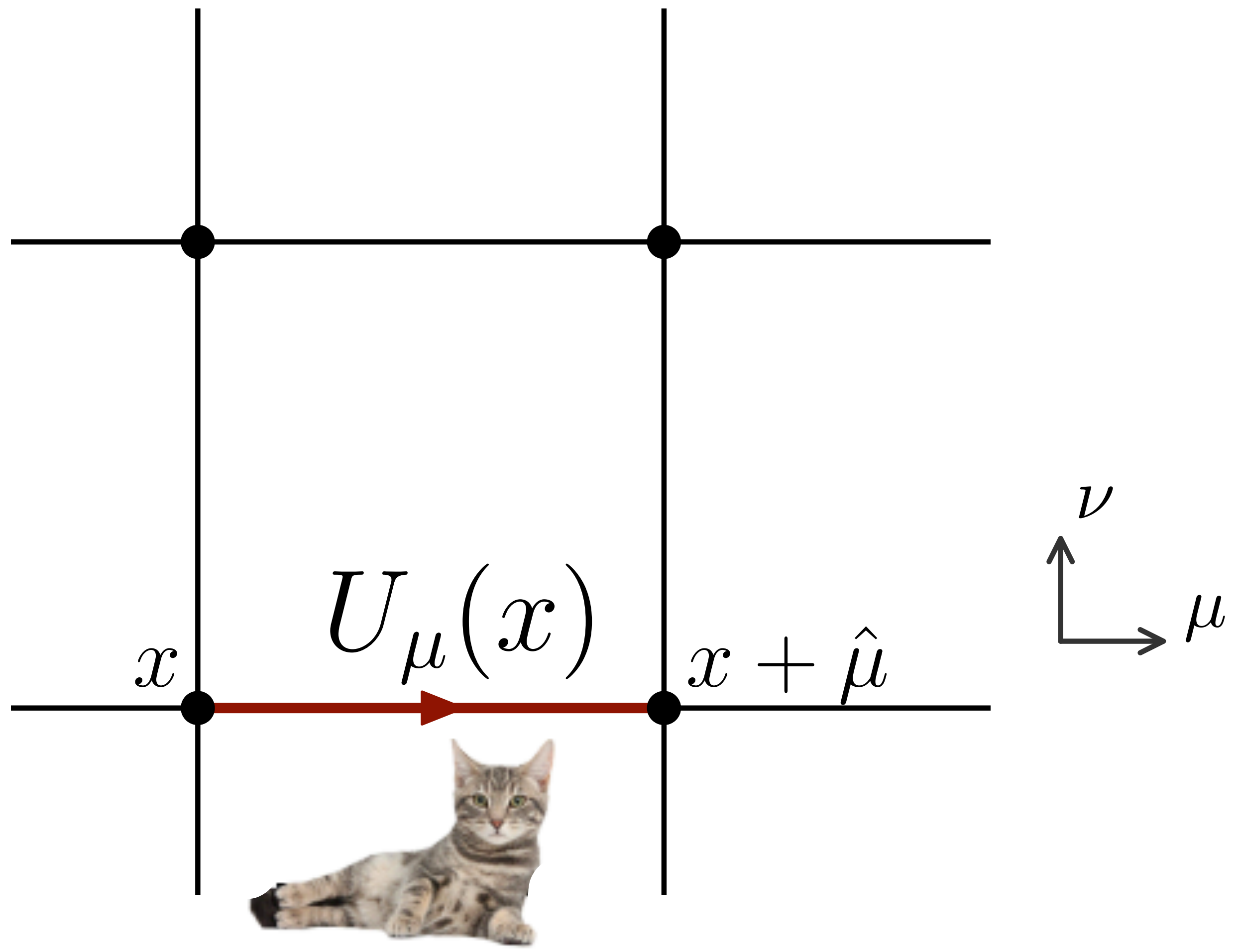
Lattice U(1) gauge symmetry

- Gauge variables are the gauge links

$$U_\mu(x) \in \mathrm{U}(1)$$

$$U_\mu(x) = e^{iagA_\mu(x)}$$

$$agA_\mu(x) \in [0, 2\pi)$$



Fernando Romero-López, MIT

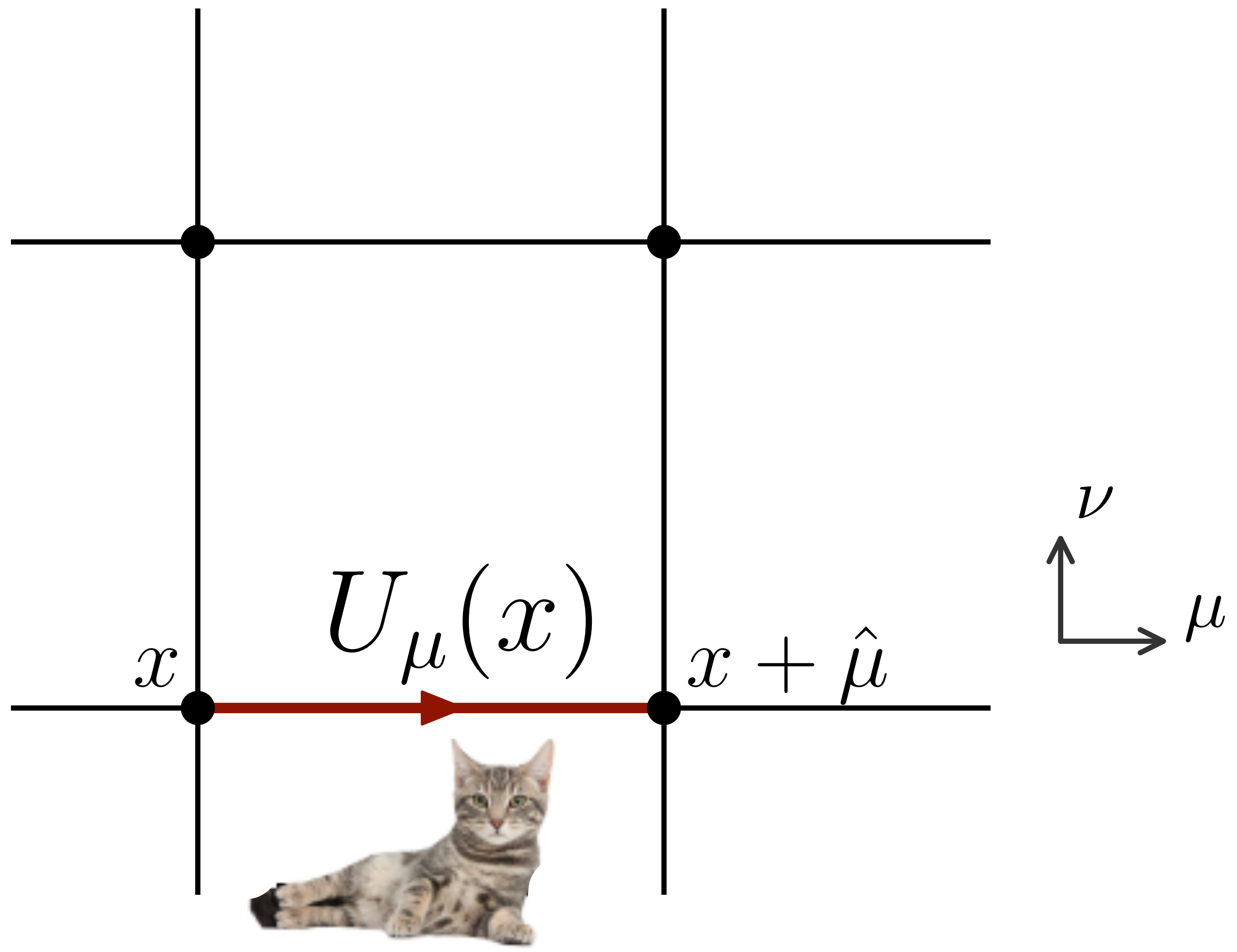
Lattice U(1) gauge symmetry

- Gauge variables are the gauge links

$$U_\mu(x) \in \mathrm{U}(1)$$

$$U_\mu(x) = e^{iagA_\mu(x)}$$

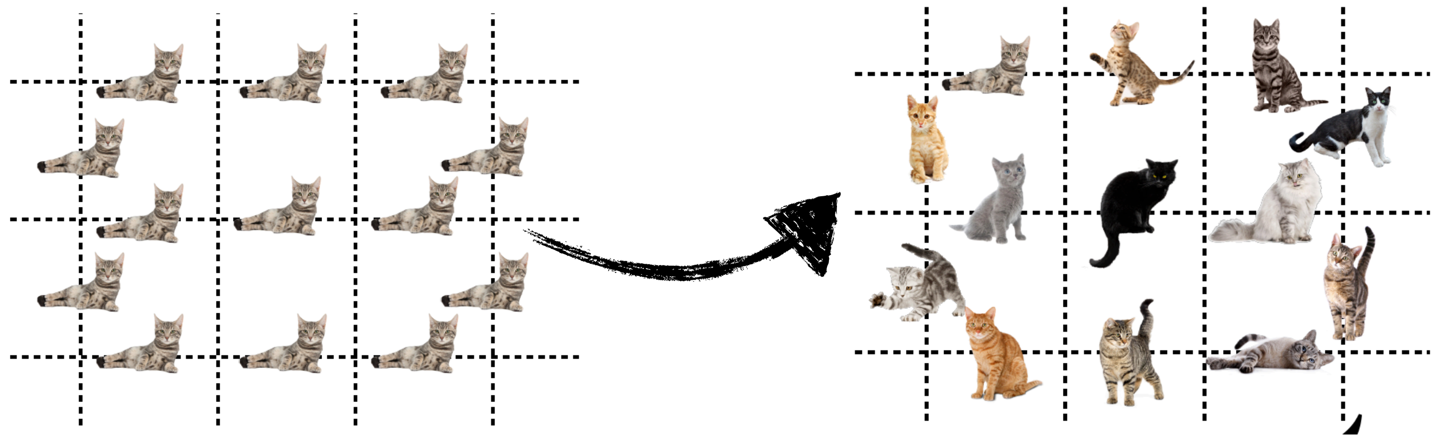
$$agA_\mu(x) \in [0, 2\pi)$$



- A gauge transformation is:

$$U_\mu(x) \to \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu})$$

$$\text{for all } \; \Omega(x) \in \mathrm{U}(1)$$



Fernando Romero-López, MIT

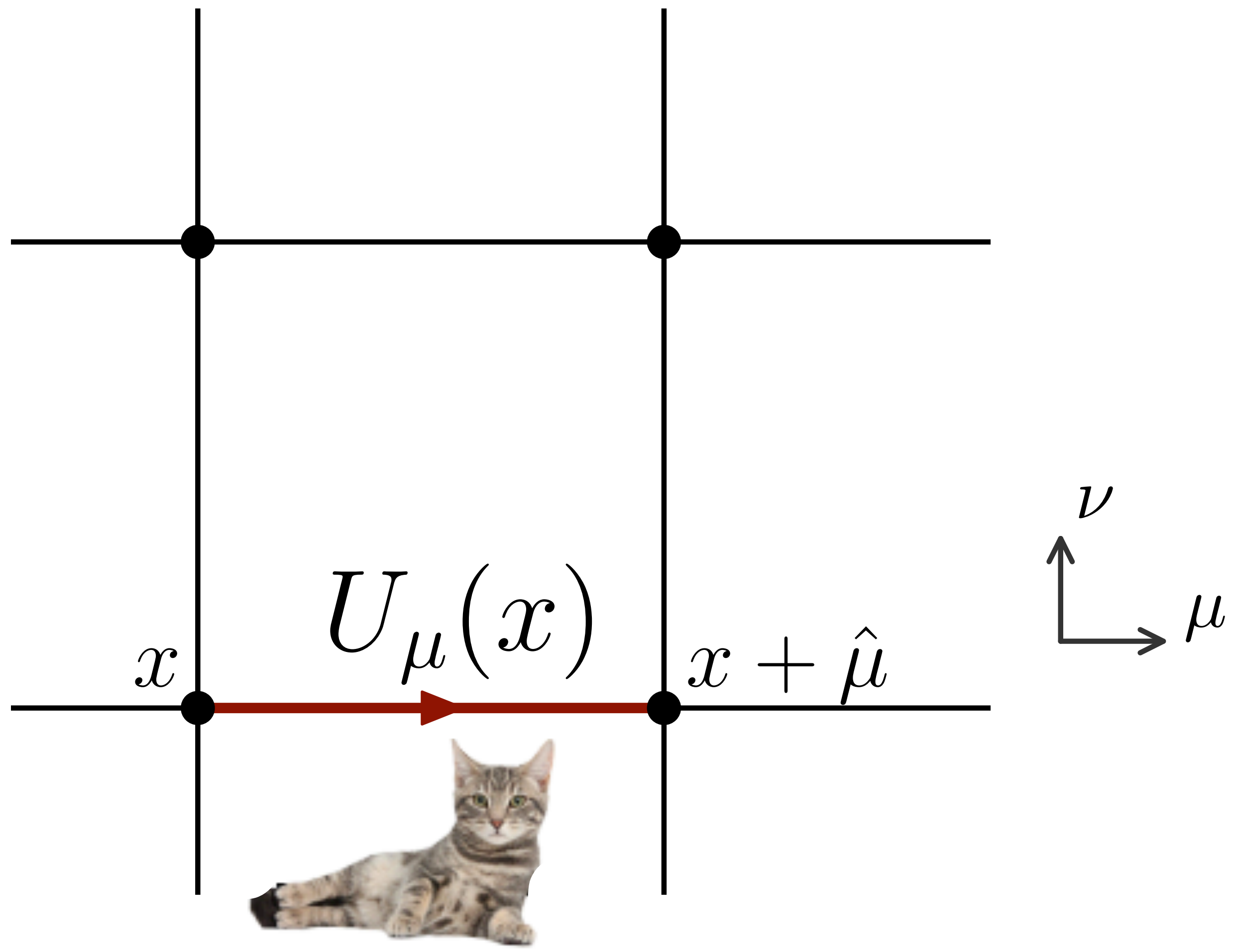
Lattice U(1) gauge symmetry

- **Gauge variables are the gauge links**

$$U_\mu(x) \in \mathrm{U}(1)$$

$$U_\mu(x) = e^{iagA_\mu(x)}$$

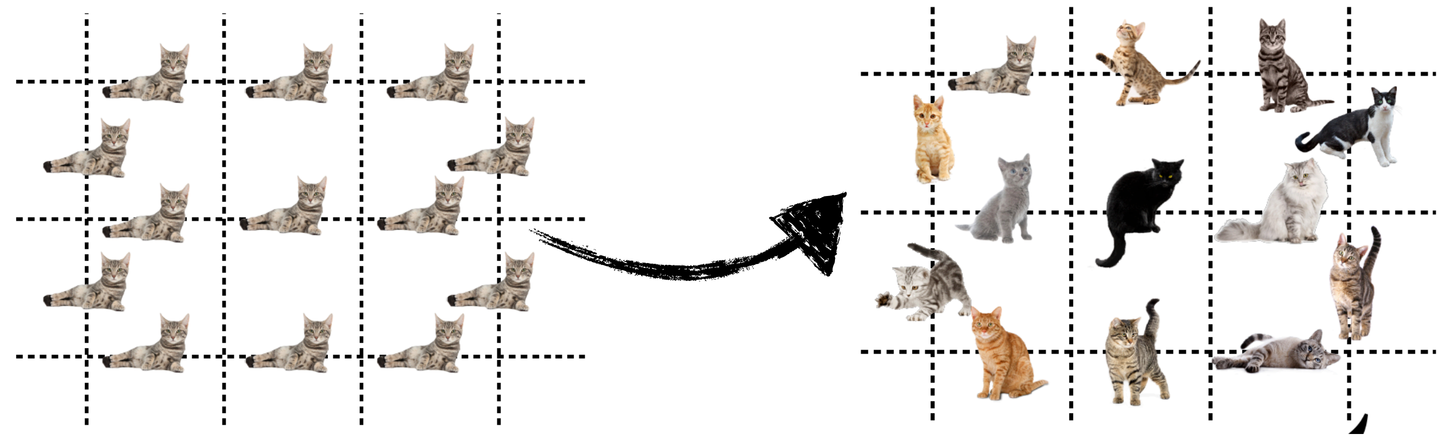
$$agA_\mu(x) \in [0, 2\pi)$$



- **A gauge transformation is:**

$$U_\mu(x) \to \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu})$$

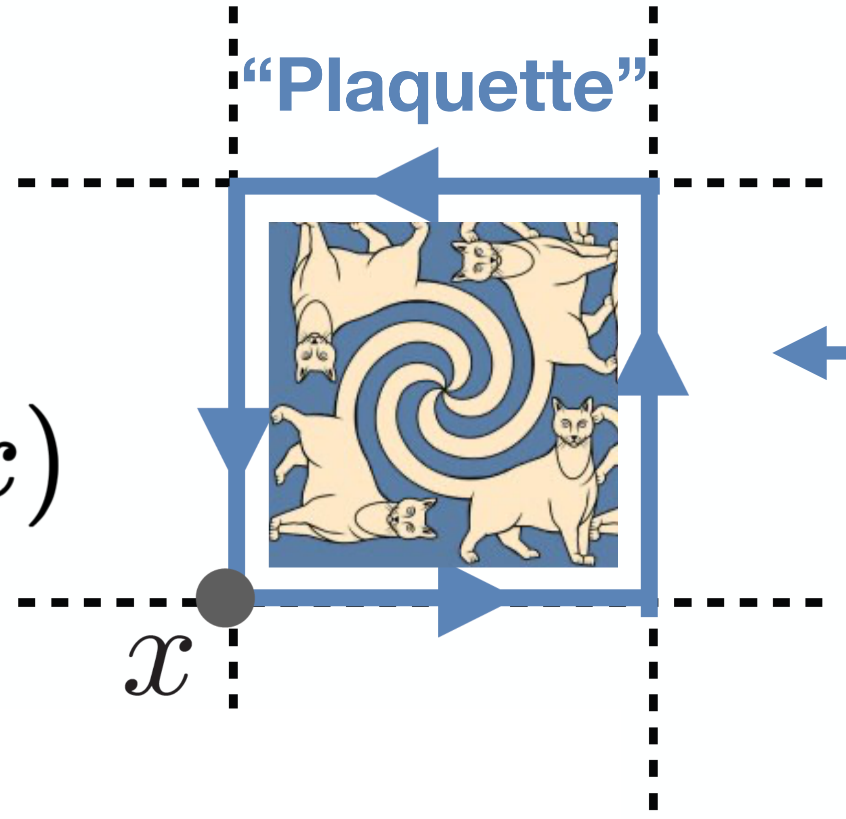
$$\text{for all } \Omega(x) \in \mathrm{U}(1)$$



- **Pure gauge lattice action:**

$$S_E(U) = -\beta \sum_x \mathrm{Re}\, P_{\mu\nu}(x)$$

inverse
gauge coupling

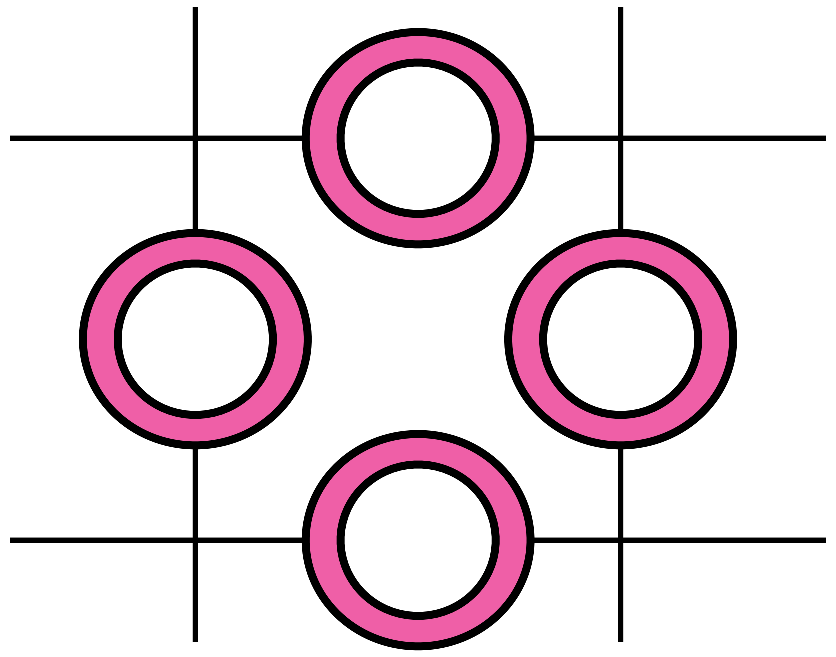


Gauge invariant
$P_{\mu\nu} \in U(1)$

Fernando Romero-López, MIT

Flows for gauge theories

$r(z)$

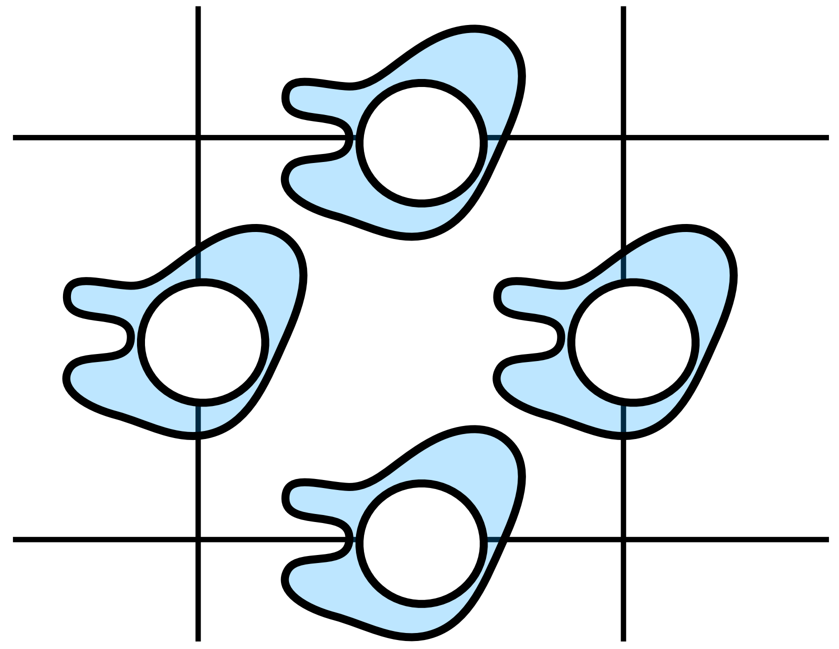


easily sampled distribution (prior)

$f(z)$

tractable Jacobian
&
invertible

$q(U) \simeq p = \frac{1}{\mathcal{Z}} e^{-S_E[U]}$



approximates target distribution (model)

- Gauge variables on compact connected manifolds
- Gauge symmetry should be included

Model probability

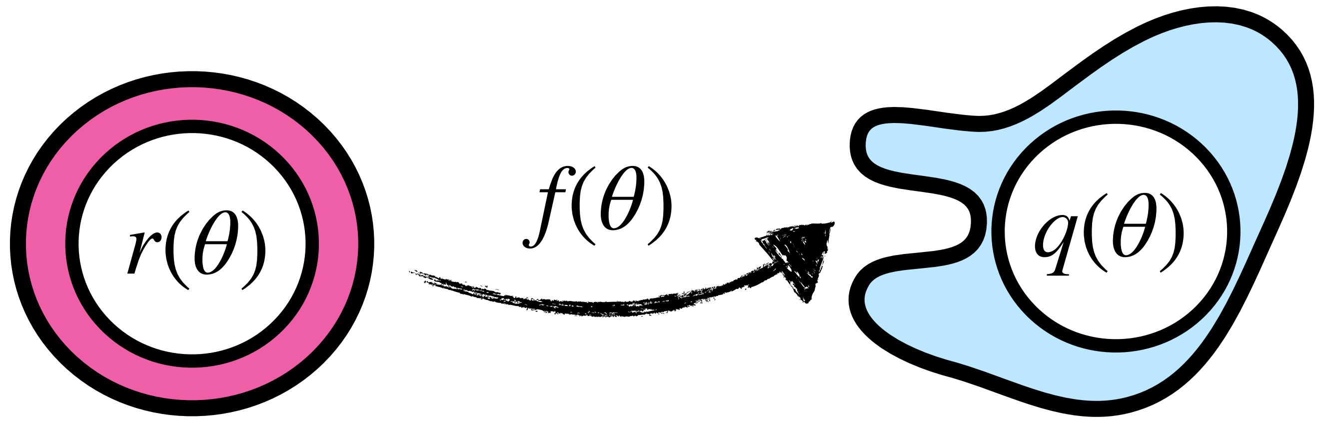
$$q(U) = r(z) \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$$

Flows on compact variables

- **Flows on compact connected manifolds:**

$$U_\mu(x) = \exp(i\theta) \in \mathrm{U}(1)$$

[Rezende, Papamakarios, Racanière , Albergo, Kanwar, Shanahan, Cranmer 2002.02428]

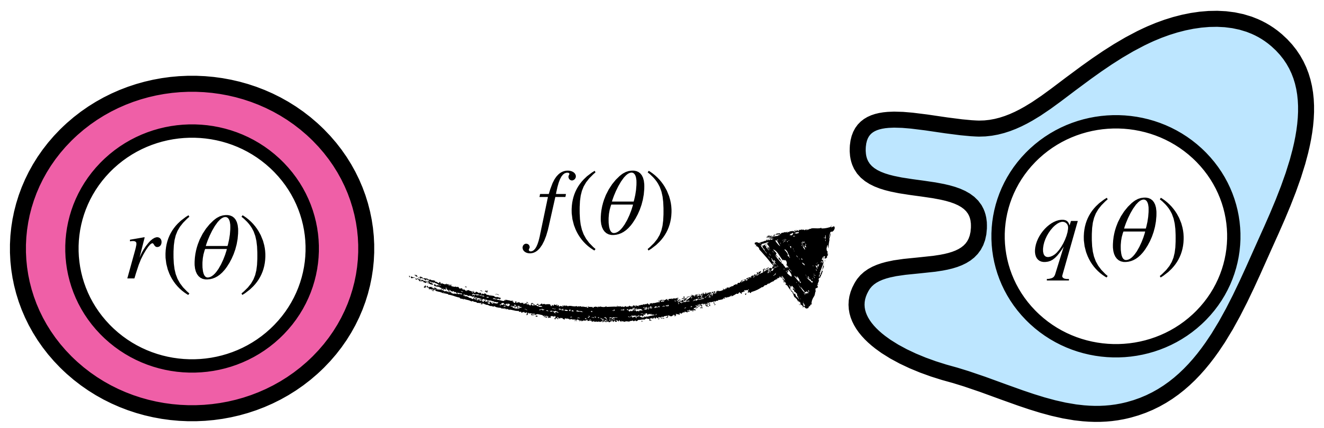


Fernando Romero-López, MIT

Flows on compact variables

- **Flows on compact connected manifolds:**

$$U_\mu(x) = \exp(i\theta) \in \mathrm{U}(1)$$



[Rezende, Papamakarios, Racanière , Albergo, Kanwar, Shanahan, Cranmer 2002.02428]

Diffeomorphism if:

$$f(0) = 0,$$
$$f(2\pi) = 2\pi,$$
$$\nabla f(\theta) > 0,$$
$$\nabla f(\theta)|_{\theta=0} = \nabla f(\theta)|_{\theta=2\pi}$$

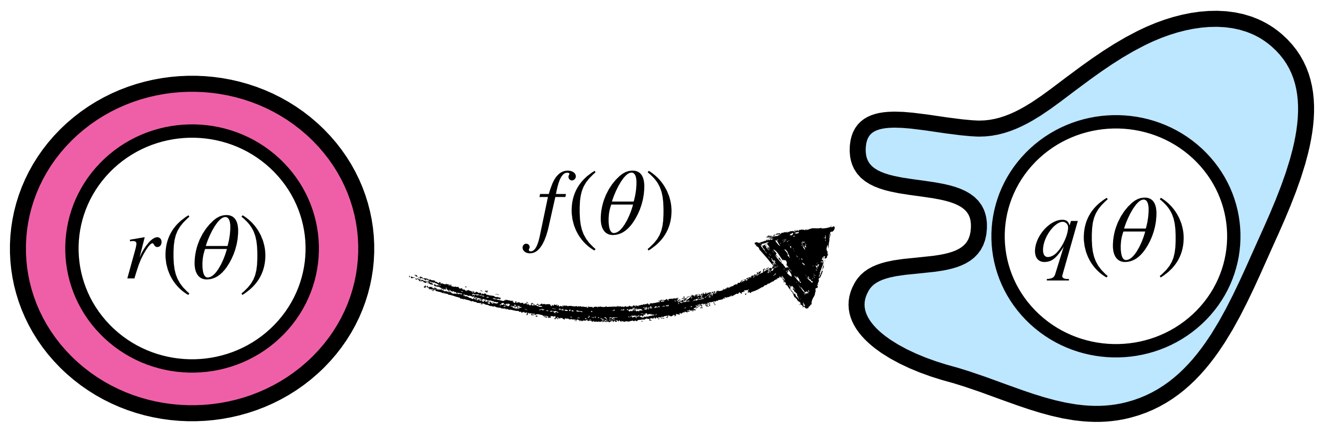
monotonic, invertible

Flows on compact variables

- **Flows on compact connected manifolds:**

$$U_\mu(x) = \exp(i\theta) \in \mathrm{U}(1)$$

[Rezende, Papamakarios, Racanière , Albergo, Kanwar, Shanahan, Cranmer 2002.02428]

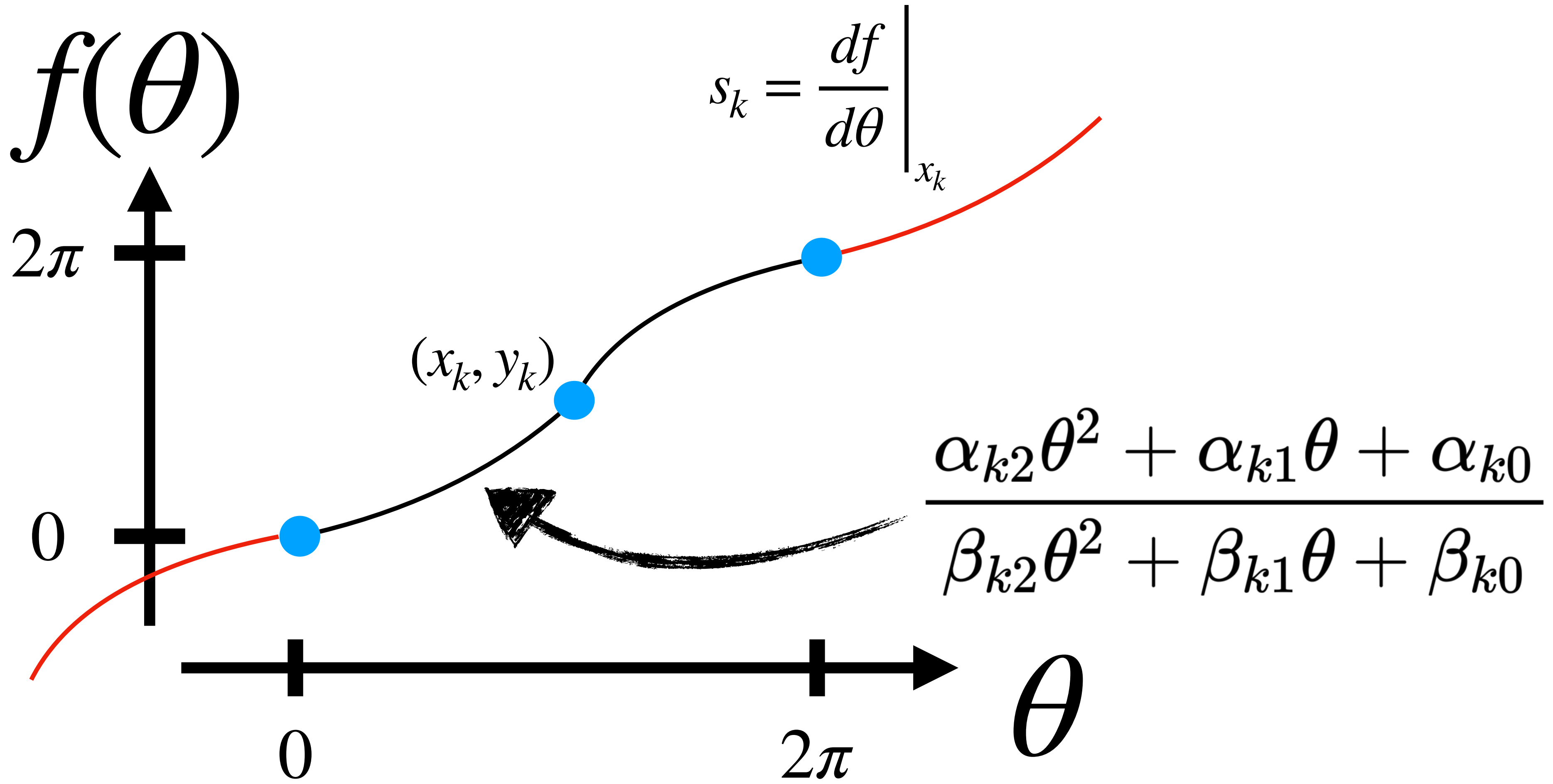


Diffeomorphism if:

$$f(0) = 0,$$
$$f(2\pi) = 2\pi,$$
$$\nabla f(\theta) > 0,$$
$$\nabla f(\theta)|_{\theta=0} = \nabla f(\theta)|_{\theta=2\pi}$$

monotonic, invertible

Circular splines:



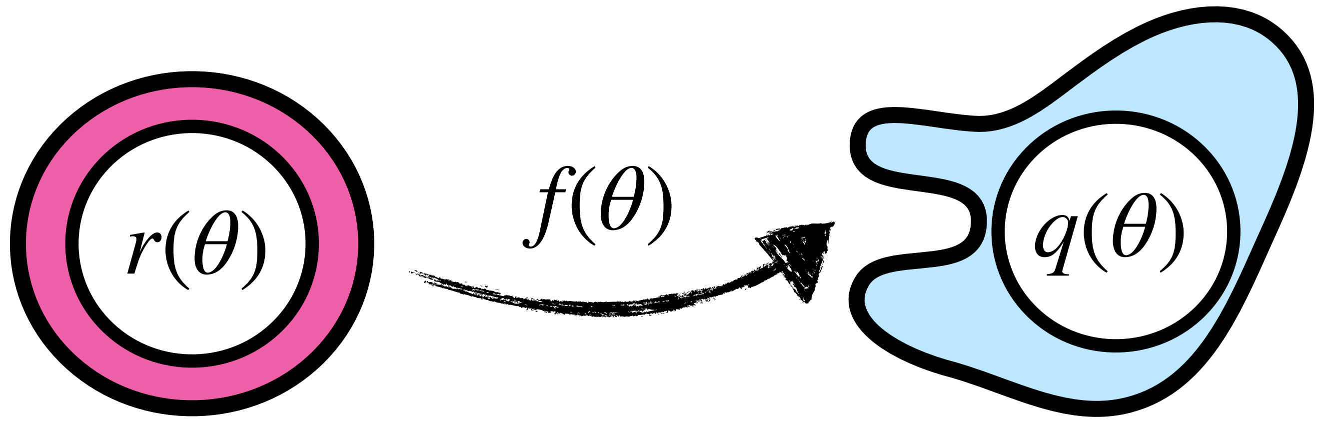
Fernando Romero-López, MIT

Flows on compact variables

- **Flows on compact connected manifolds:**

$$U_\mu(x) = \exp(i\theta) \in \mathrm{U}(1)$$

[Rezende, Papamakarios, Racanière , Albergo, Kanwar, Shanahan, Cranmer 2002.02428]



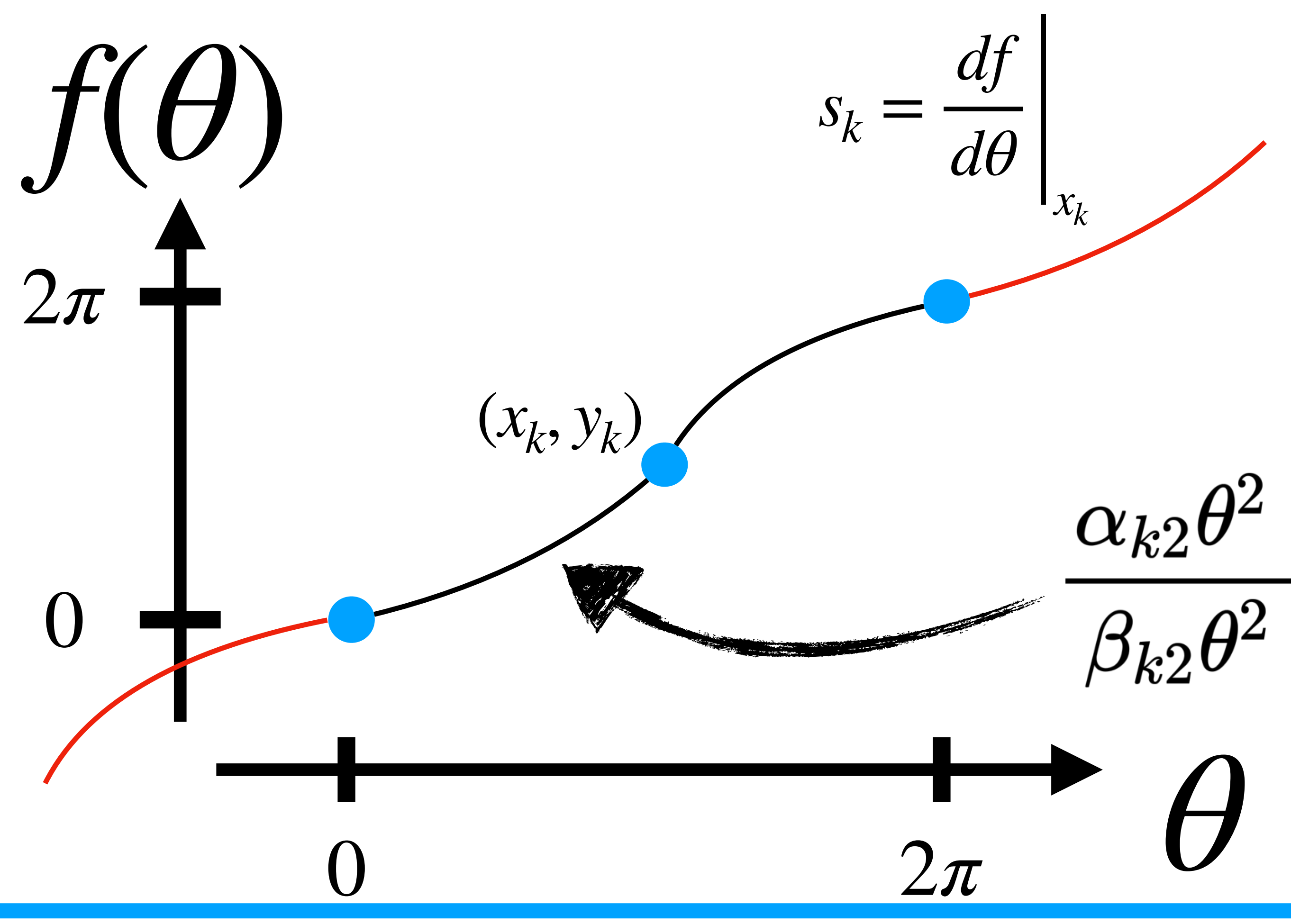
Diffeomorphism if:

$$\begin{aligned} f(0) &= 0, \\ f(2\pi) &= 2\pi, \\ \nabla f(\theta) &> 0, \\ \nabla f(\theta)|_{\theta=0} &= \nabla f(\theta)|_{\theta=2\pi} \end{aligned}$$

monotonic, invertible

Circular splines:

- Trainable positions and slopes
- Expressivity with more knots

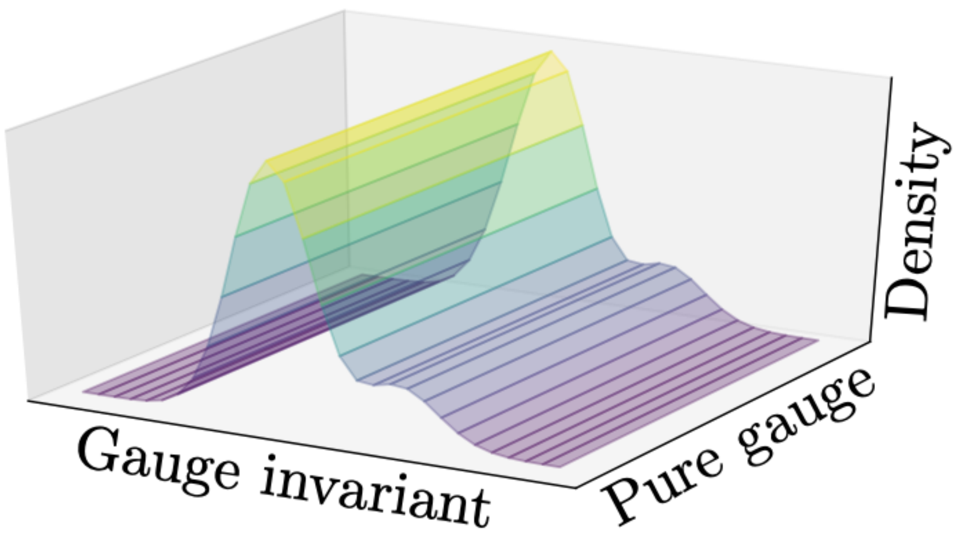


$$\frac{\alpha_{k2}\theta^2 + \alpha_{k1}\theta + \alpha_{k0}}{\beta_{k2}\theta^2 + \beta_{k1}\theta + \beta_{k0}}$$

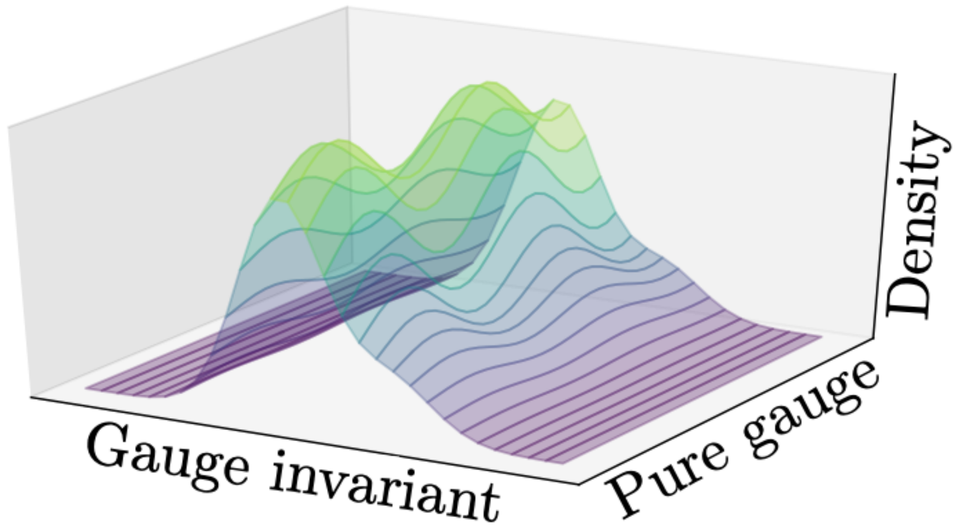
Incorporating symmetries

- Not essential, but:
  - ✓ Reduces complexity of training
  - ✓ Reduces parameter count

Gauge symmetries:



True distribution



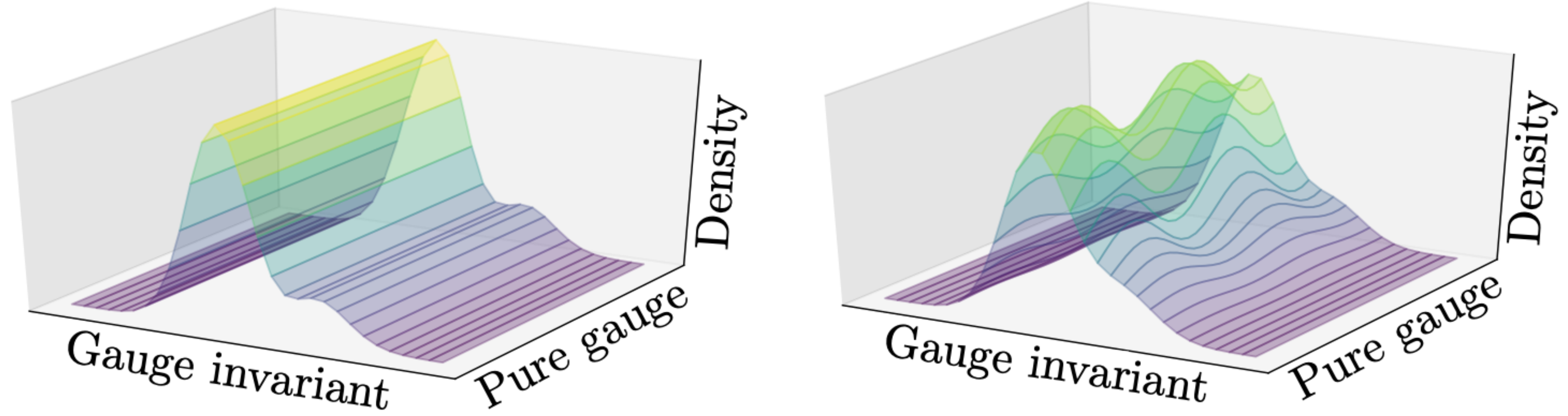
Learned by naive ML

Fernando Romero-López, MIT

Incorporating symmetries

- Not essential, but:
  - ✓ Reduces complexity of training
  - ✓ Reduces parameter count

Gauge symmetries:

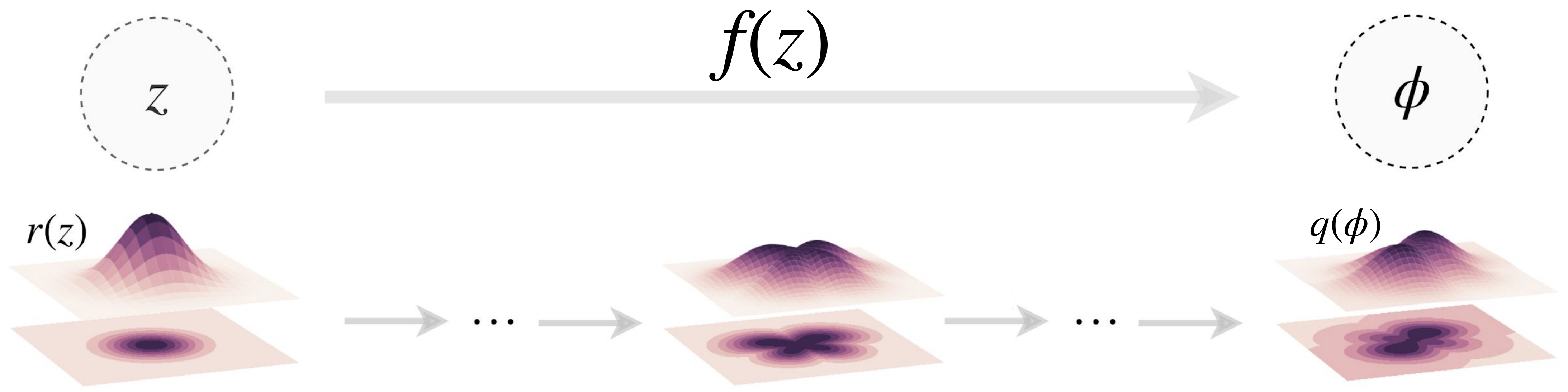


True distribution **Learned by naive ML**

General approach:

1. Invariant base distribution

$$r(z) = r(t(z))$$

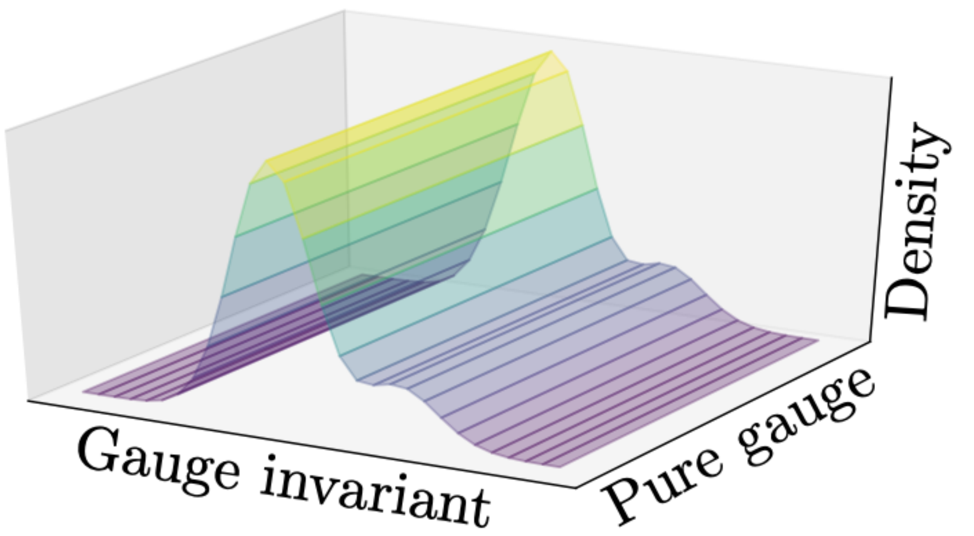


Fernando Romero-López, MIT

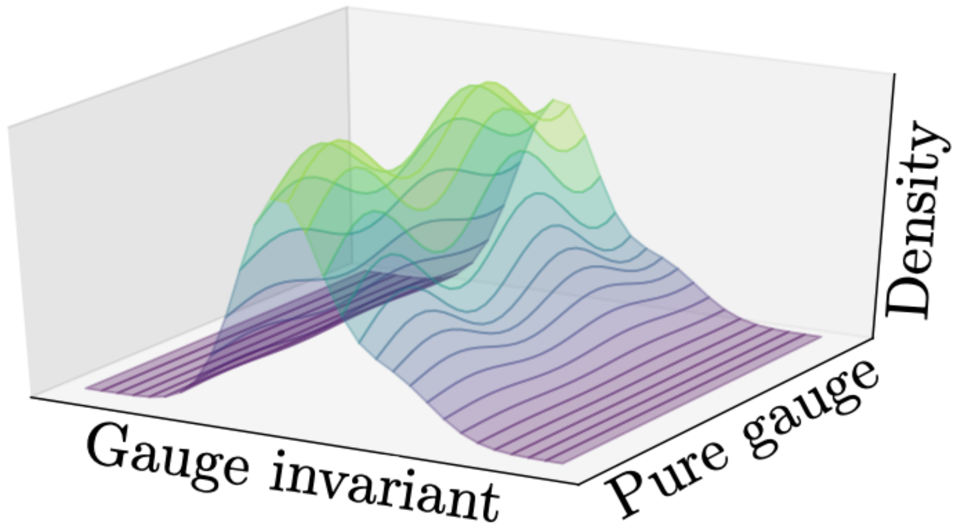
Incorporating symmetries

- Not essential, but:
  - ✓ Reduces complexity of training
  - ✓ Reduces parameter count

Gauge symmetries:



True distribution



Learned by naive ML

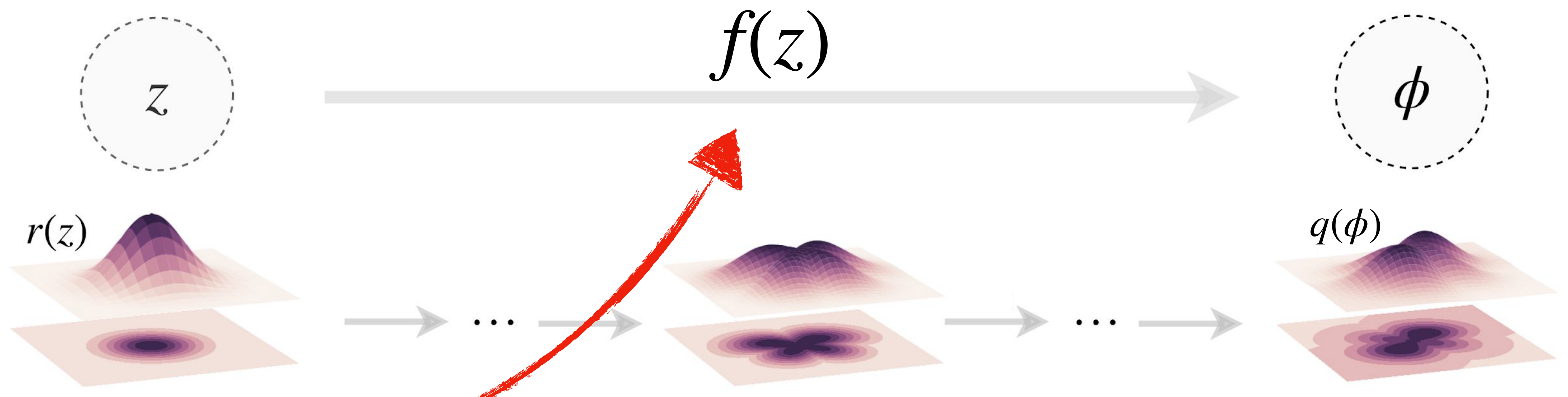
General approach:

1. Invariant base distribution

$$r(z) = r(t(z))$$

2. Equivariant flow

$$f(t(\phi)) = t(f(\phi))$$



Fernando Romero-López, MIT

Flows with gauge variables

- **Gauge invariant prior + gauge-equivariant transformations:**

$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

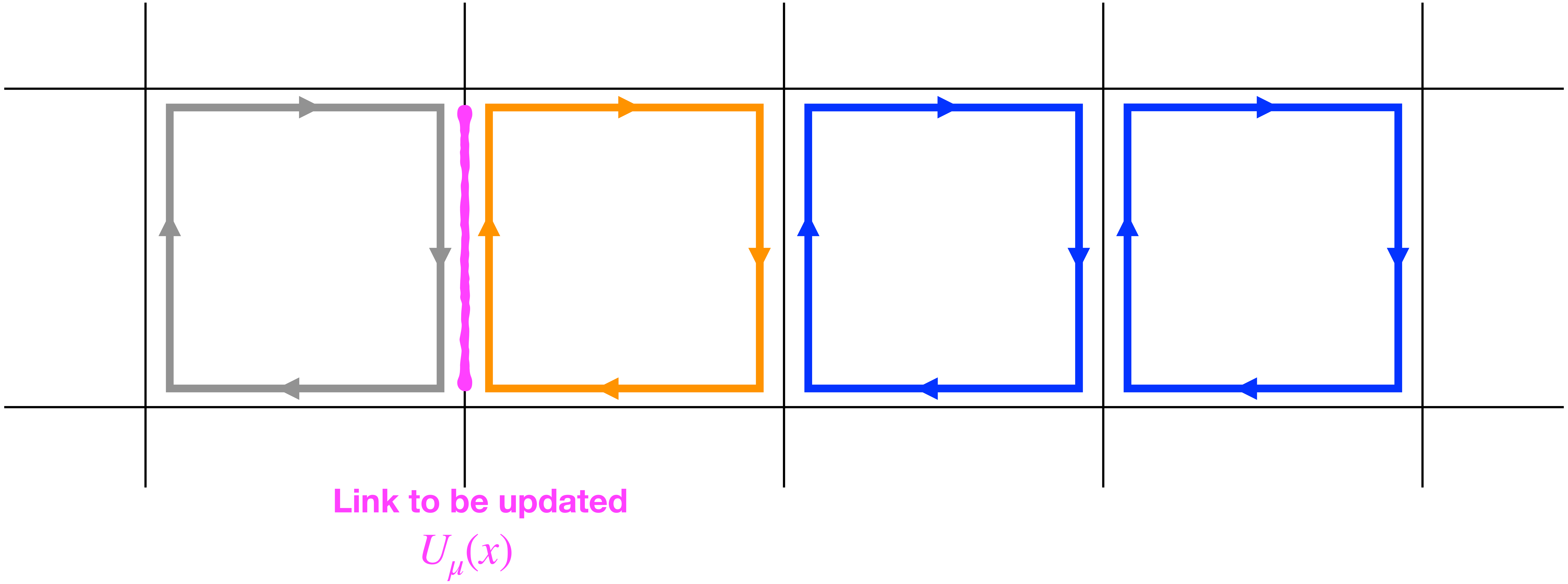
$$g\big(\Omega(x) U_\mu(x) \Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$

Flows with gauge variables

- **Gauge invariant prior + gauge-equivariant transformations:**

$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

$$g\big(\Omega(x)U_\mu(x)\Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$



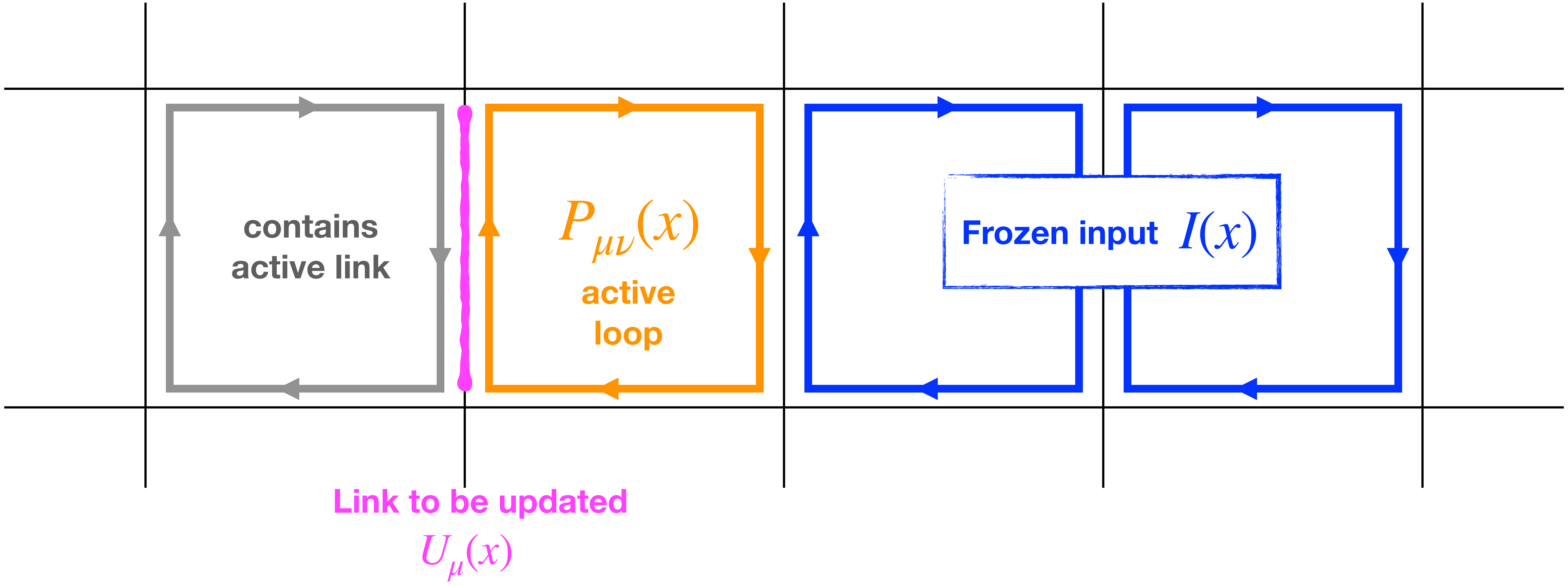
Fernando Romero-López, MIT

Flows with gauge variables

- **Gauge invariant prior + gauge-equivariant transformations:**

$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

$$g\big(\Omega(x)U_\mu(x)\Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$



Fernando Romero-López, MIT

Flows with gauge variables

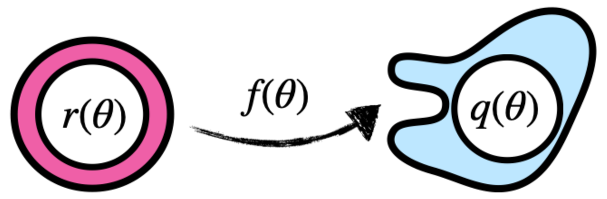
- **Gauge invariant prior + gauge-equivariant transformations:**

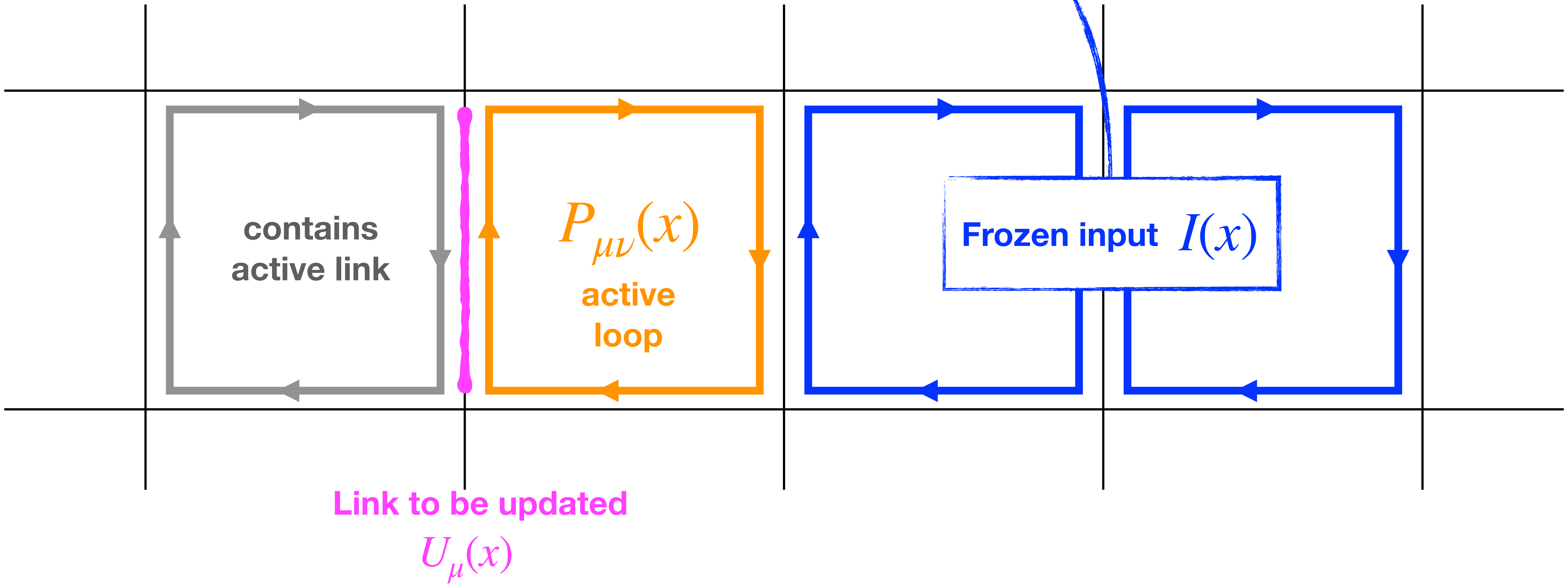
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

$$g\big(\Omega(x)U_\mu(x)\Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$

$$P'_{\mu\nu}(x) = h\left(P_{\mu\nu}(x)\,|\, I(x)\right)$$

e.g. circular splines





Fernando Romero-López, MIT

Flows with gauge variables

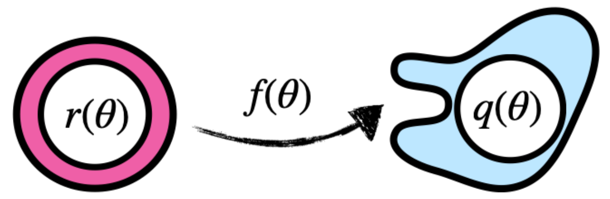
- **Gauge invariant prior + gauge-equivariant transformations:**

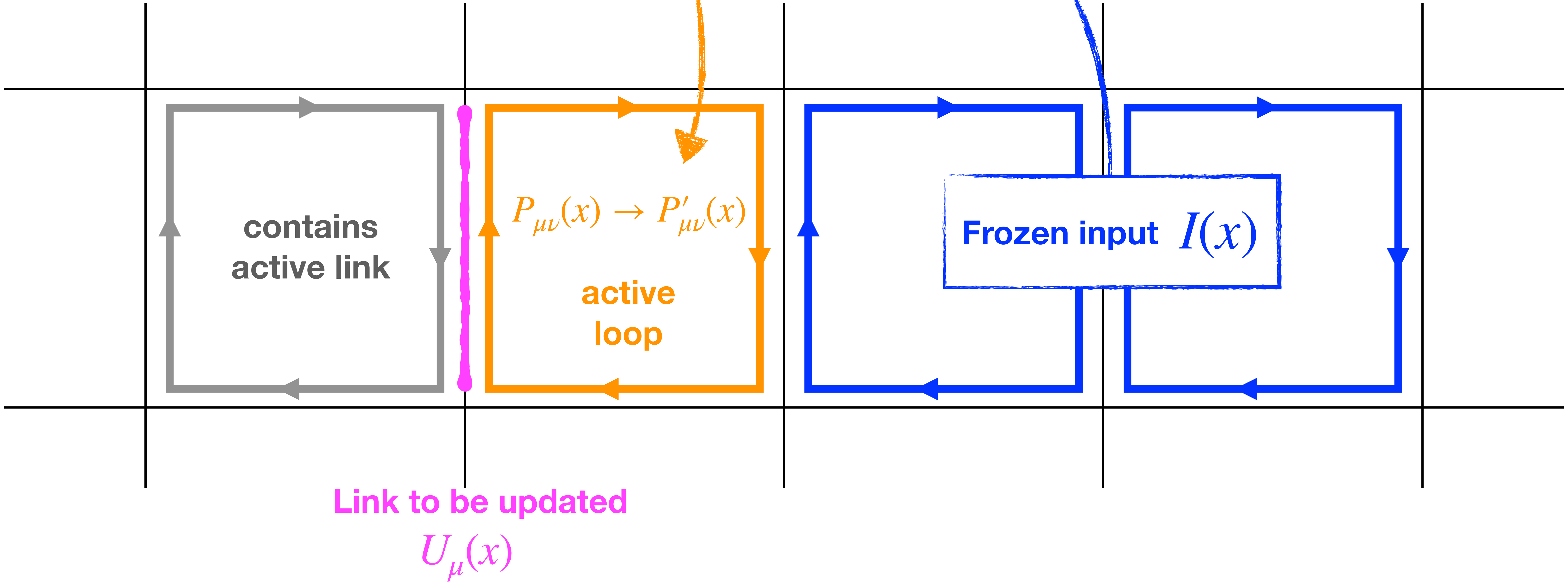
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

$$g\big(\Omega(x)U_\mu(x)\Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$

$$P'_{\mu\nu}(x) = h\left(P_{\mu\nu}(x)\,|\,I(x)\right)$$

e.g. circular splines





Fernando Romero-López, MIT

Flows with gauge variables

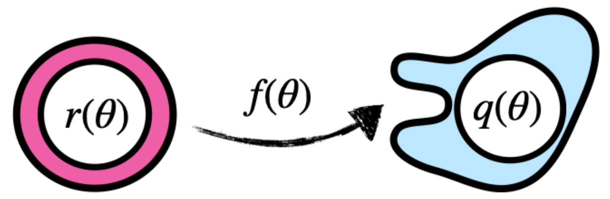
- **Gauge invariant prior + gauge-equivariant transformations:**

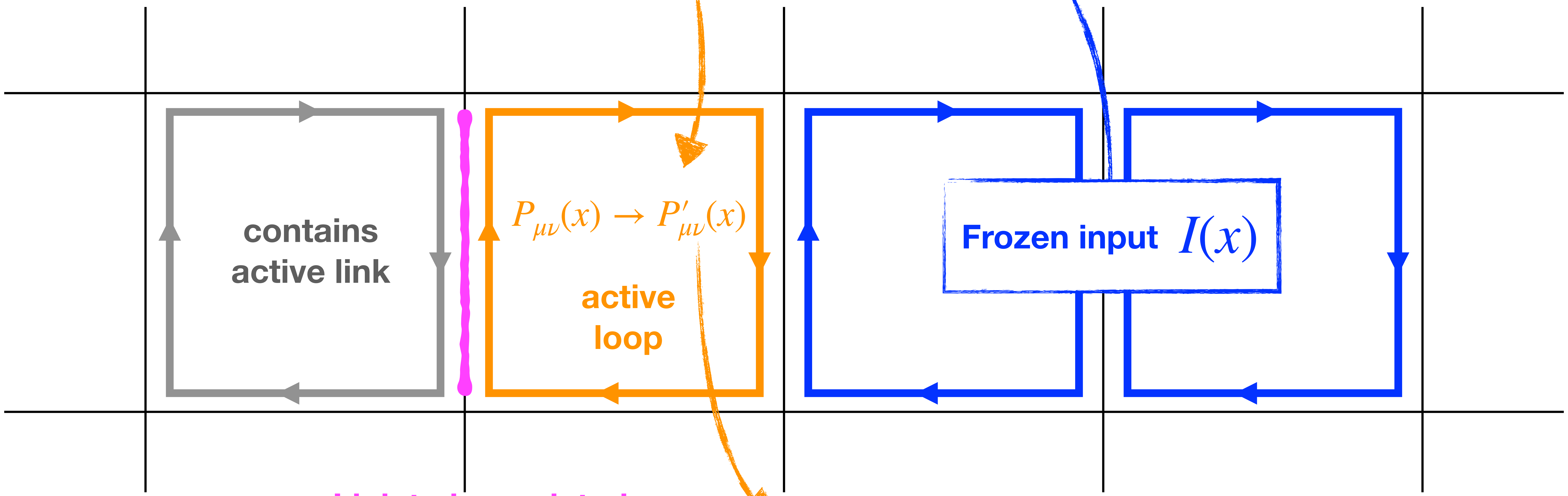
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

$$g\big(\Omega(x)U_\mu(x)\Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$

$$P'_{\mu\nu}(x) = h\left(P_{\mu\nu}(x)\,|\,I(x)\right)$$

e.g. circular splines





Link to be updated

$$U_\mu(x) \to U'_\mu(x) = P'_{\mu\nu}(x)P^\dagger_{\mu\nu}(x)U_\mu(x)$$

Fernando Romero-López, MIT

Flows with gauge variables

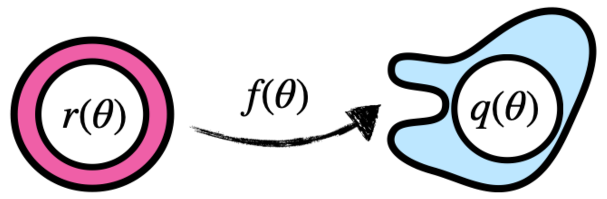
- **Gauge invariant prior + gauge-equivariant transformations:**

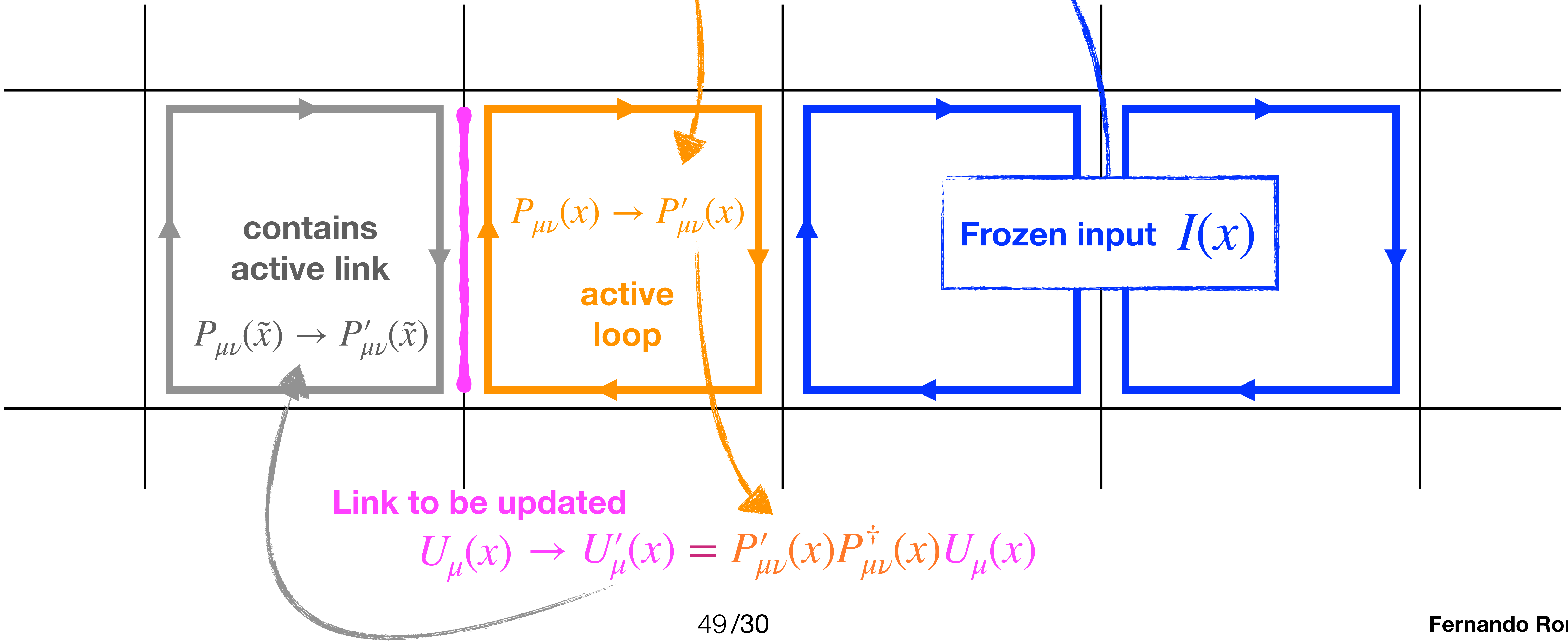
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

$$g\big(\Omega(x) U_\mu(x) \Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$

$$P'_{\mu\nu}(x) = h\left(P_{\mu\nu}(x) \,|\, I(x)\right)$$

e.g. circular splines





$$U_\mu(x) \to U'_\mu(x) = P'_{\mu\nu}(x) P^\dagger_{\mu\nu}(x) U_\mu(x)$$

Fernando Romero-López, MIT

Flows with gauge variables

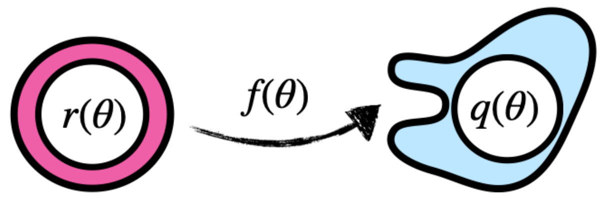
- **Gauge invariant prior + gauge-equivariant transformations:**

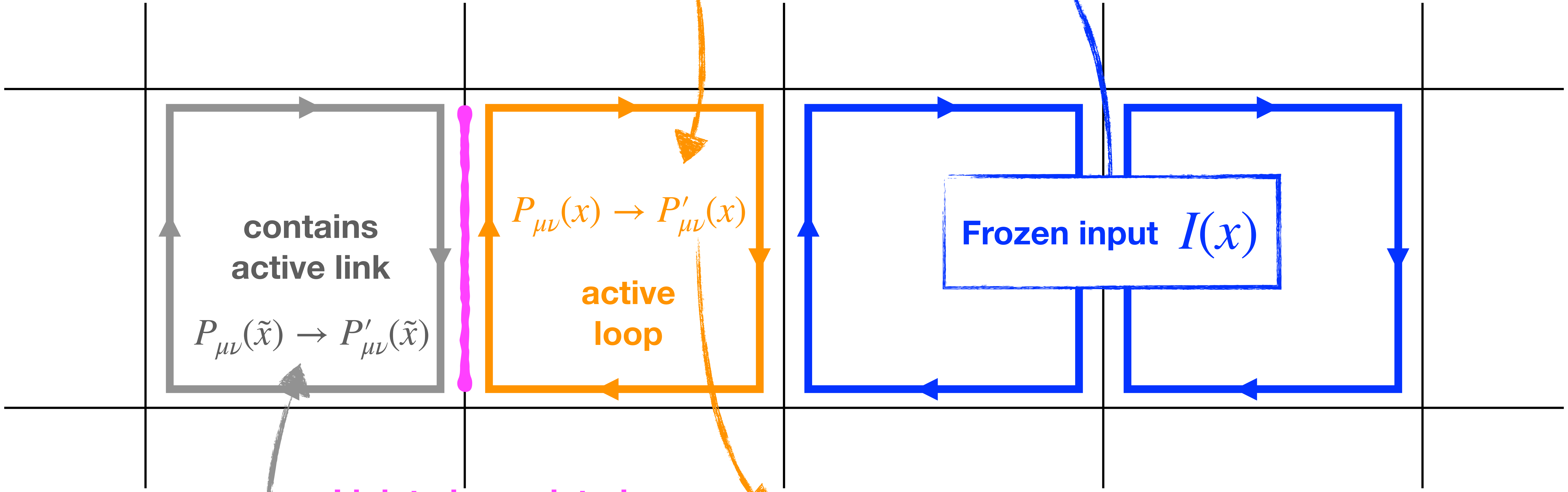
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

$$g\big(\Omega(x)U_\mu(x)\Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$

$$P'_{\mu\nu}(x) = h\left(P_{\mu\nu}(x) \,|\, I(x)\right)$$

e.g. circular splines



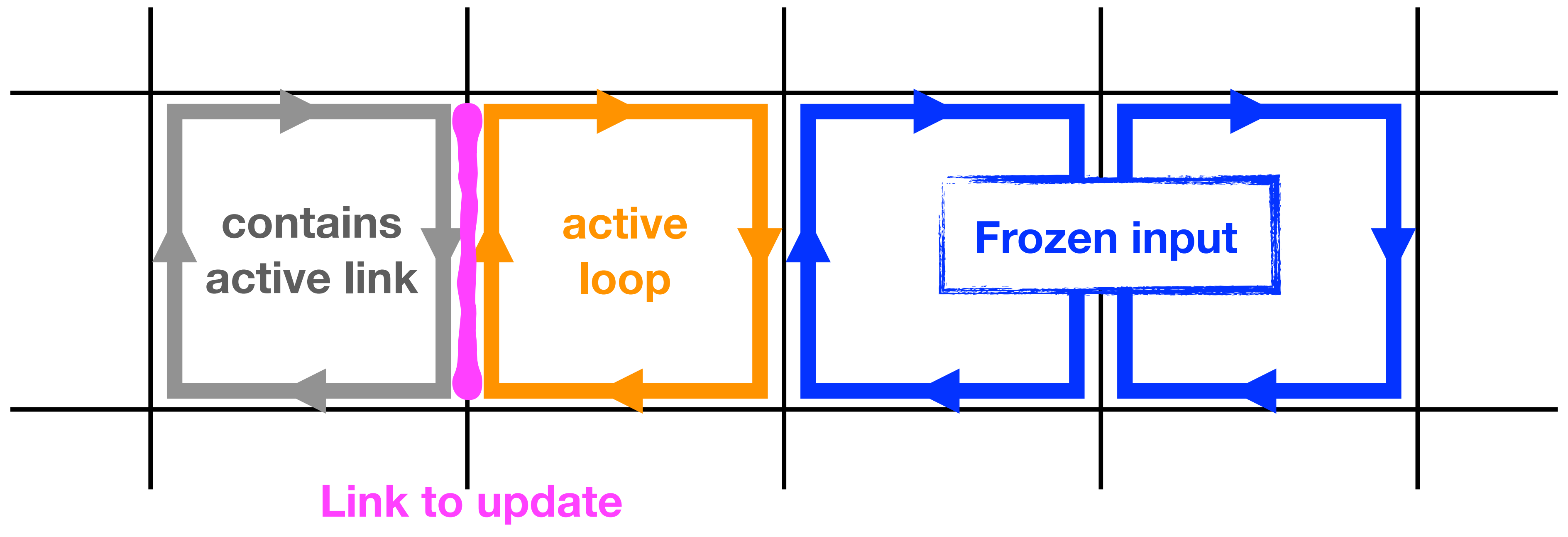


Link to be updated

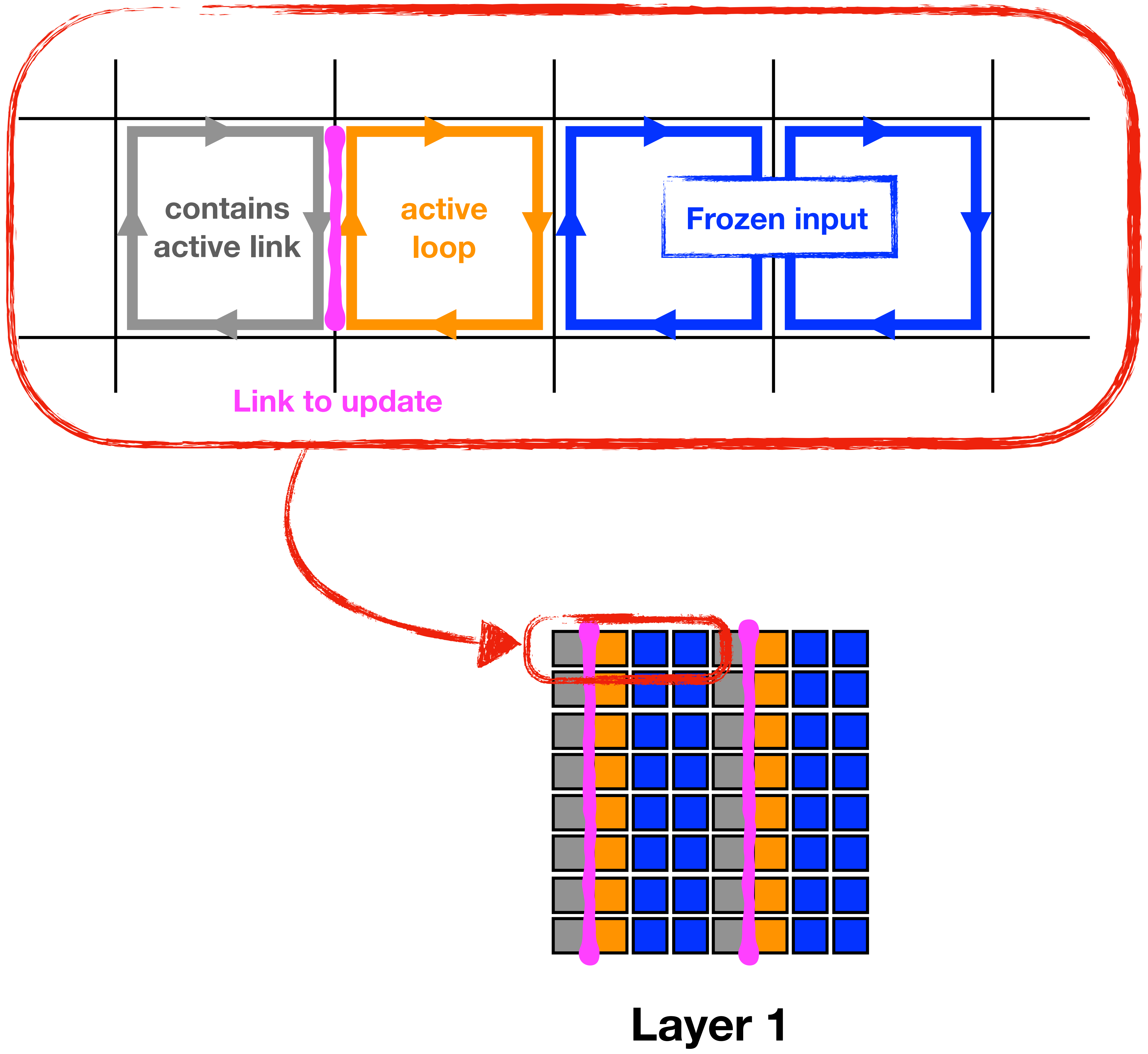
$$U_\mu(x) \to U'_\mu(x) = P'_{\mu\nu}(x) P^\dagger_{\mu\nu}(x) U_\mu(x)$$

"gauge-equivariant layer"

Gauge-equivariant models

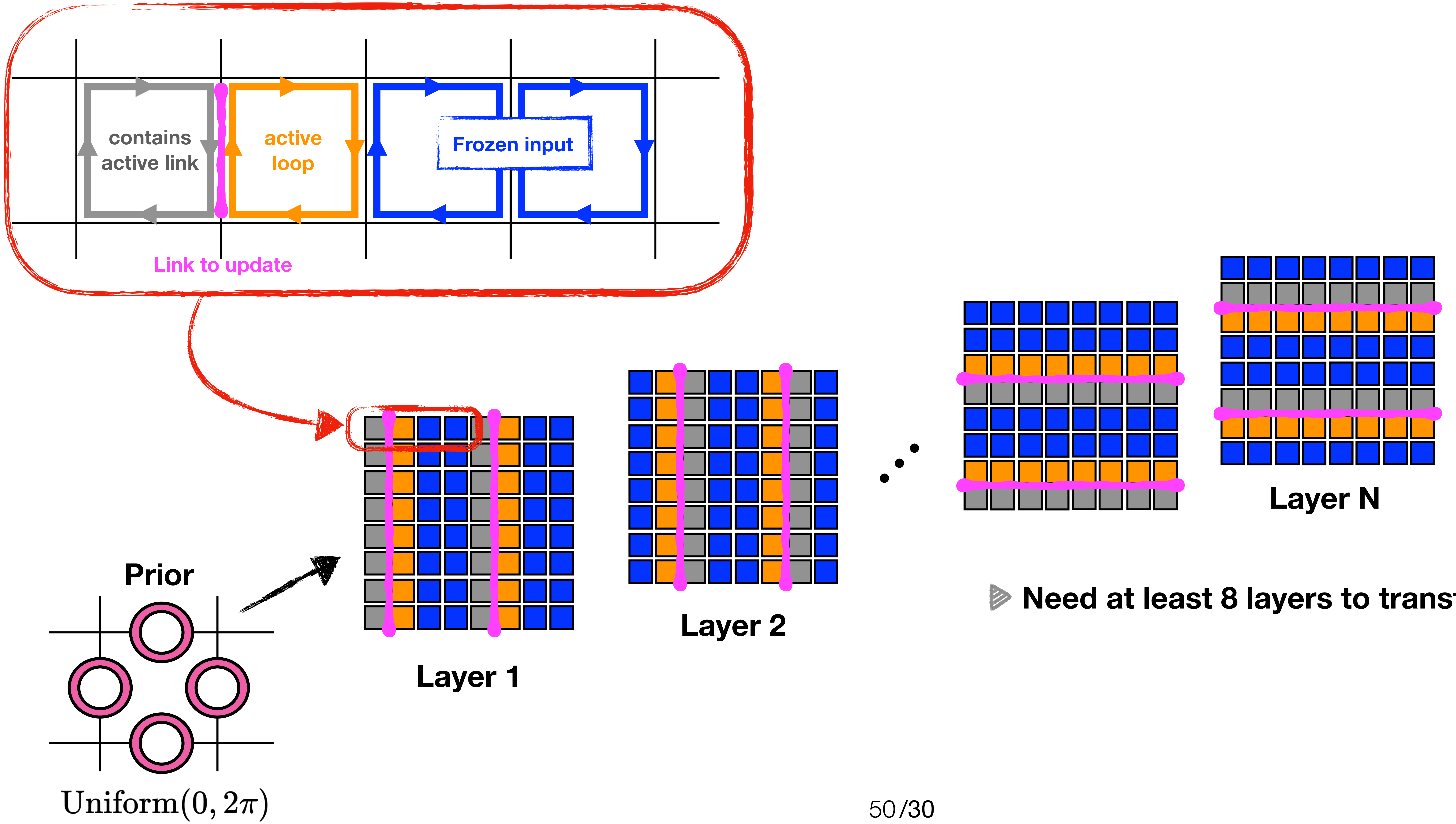


Gauge-equivariant models



Fernando Romero-López, MIT

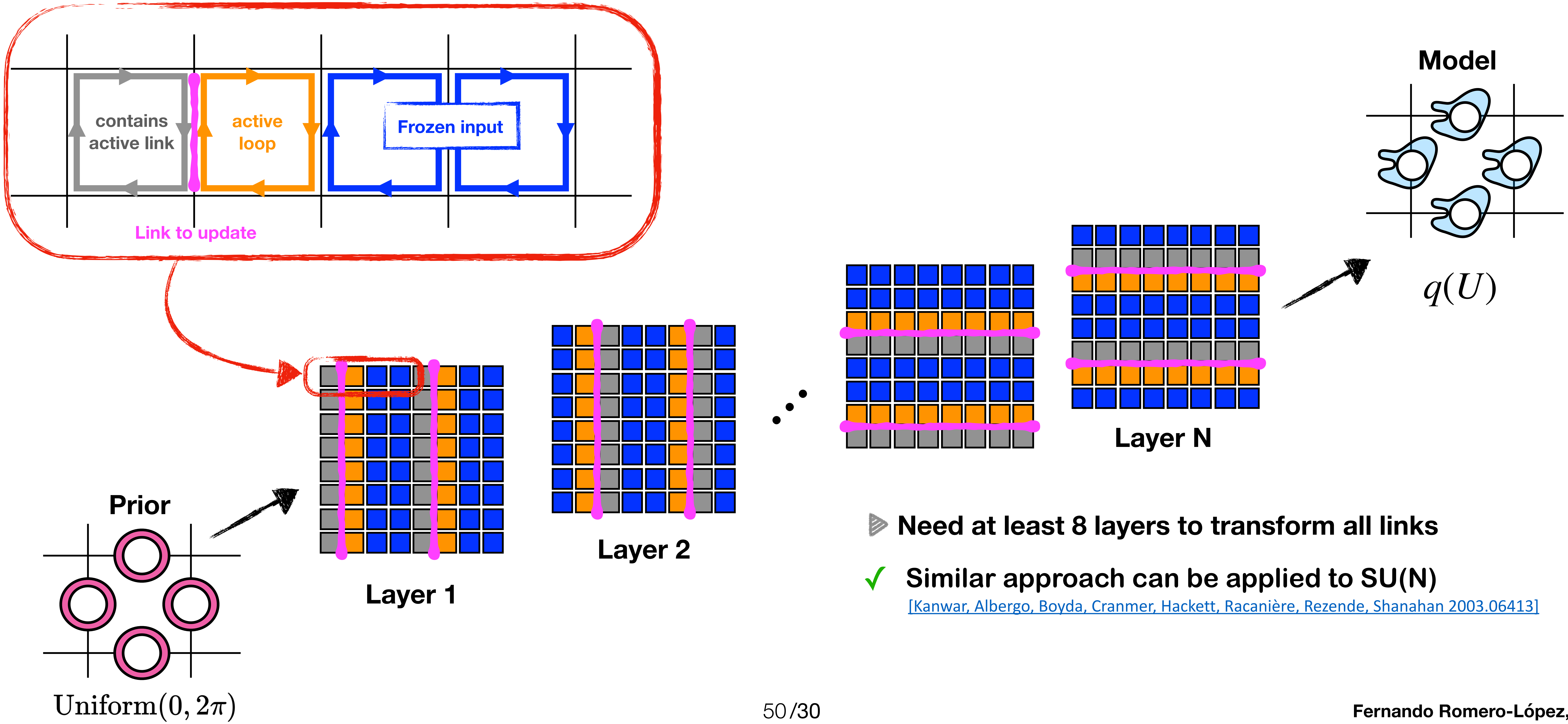
Gauge-equivariant models



- Need at least 8 layers to transform all links

Fernando Romero-López, MIT

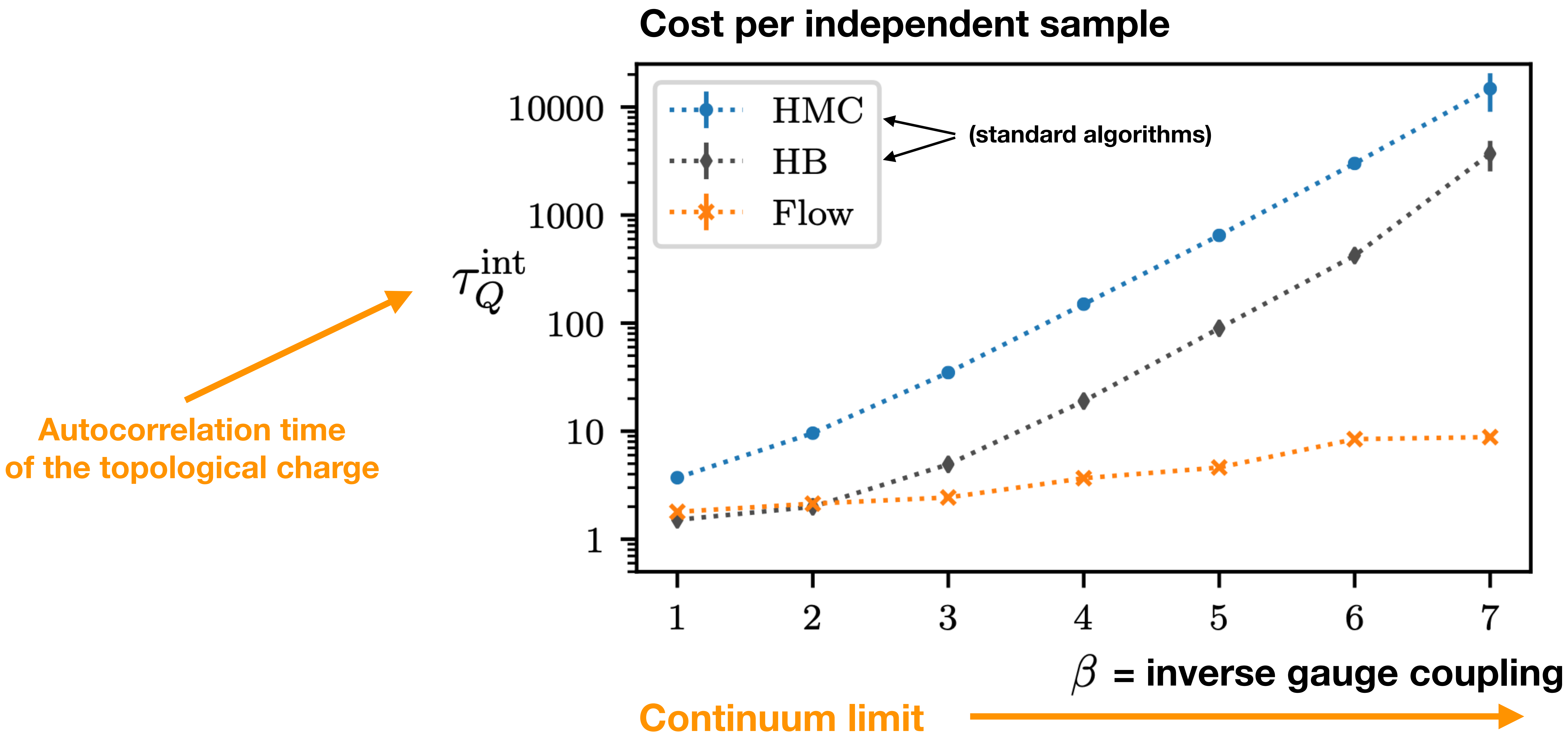
Gauge-equivariant models



- Need at least 8 layers to transform all links
- ✓ Similar approach can be applied to SU(N)
  [Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan 2003.06413]

Fernando Romero-López, MIT

Results in 2D U(1) theory



✓ **Flow-based sampling has no topological freezing**

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

Flows for fermionic gauge theories

The Schwinger Model

QED in 1+1 dimensions with Nf=2 fermions

$$\mathcal{L} = -\frac{1}{4g^2} F_{\mu\nu} F^{\mu\nu} + \sum_{f=1,2} \bar{\psi}_f (i\gamma^\mu D_\mu - m)\psi_f$$

Fernando Romero-López, MIT

The Schwinger Model

QED in 1+1 dimensions with Nf=2 fermions

$$\mathcal{L} = -\frac{1}{4g^2} F_{\mu\nu} F^{\mu\nu} + \sum_{f=1,2} \bar{\psi}_f (i\gamma^\mu D_\mu - m)\psi_f$$

Toy model for QCD

1. Confinement
2. Chiral symmetry breaking
3. Topology

Machine Learning challenges

1. Gauge symmetry
2. Fermion degrees of freedom
3. Long-range correlations

Fermionic theories

- **The fermionic part of the action**

Dirac-Wilson operator

$$S_{\text{ferm}}(U,\psi,\bar\psi) = \sum_{f=1}^{N_f}\sum_{x,y}\bar\psi_f^{\beta}(y)D[U](y,x)^{\beta\alpha}\psi_f^{\alpha}(x)$$

$$D[U](y,x)^{\beta\alpha} = \delta(y-x)\delta^{\beta\alpha} - \kappa\sum_{\mu=0,1}\left\{[1-\sigma_\mu]^{\beta\alpha}U_\mu(y)\delta(y-x+\hat\mu) + [1+\sigma_\mu]^{\beta\alpha}U_\mu^\dagger(y-\hat\mu)\delta(y-x-\hat\mu)\right\},$$

$$\dim D[U] = \text{volume}\times\text{spin}\times\text{gauge}$$

Fernando Romero-López, MIT

Fermionic theories

- **The fermionic part of the action**

Dirac-Wilson operator

$$S_{\text{ferm}}(U,\psi,\bar{\psi}) = \sum_{f=1}^{N_f} \sum_{x,y} \bar{\psi}^{\beta}_{f}(y) D[U](y,x)^{\beta\alpha} \psi^{\alpha}_{f}(x)$$

$$D[U](y,x)^{\beta\alpha} = \delta(y-x)\delta^{\beta\alpha} - \kappa \sum_{\mu=0,1} \Big\{ [1-\sigma_\mu]^{\beta\alpha} U_\mu(y)\delta(y-x+\hat{\mu}) + [1+\sigma_\mu]^{\beta\alpha} U^{\dagger}_{\mu}(y-\hat{\mu})\delta(y-x-\hat{\mu}) \Big\} \,,$$

$$\dim D[U] = \text{volume} \times \text{spin} \times \text{gauge}$$

- **Fermionic degrees of freedom can be integrated out**

$$\int \mathcal{D}[U]\mathcal{D}[\psi,\bar{\psi}] e^{-S_{\text{gauge}}(U)} e^{-S_{\text{ferm}}(\psi,\bar{\psi},U)}$$

Fermionic theories

- **The fermionic part of the action**

Dirac-Wilson operator

$$S_{\text{ferm}}(U,\psi,\bar\psi) = \sum_{f=1}^{N_f}\sum_{x,y}\bar\psi_f^{\beta}(y)D[U](y,x)^{\beta\alpha}\psi_f^{\alpha}(x)$$

$$D[U](y,x)^{\beta\alpha} = \delta(y-x)\delta^{\beta\alpha} - \kappa\sum_{\mu=0,1}\left\{[1-\sigma_\mu]^{\beta\alpha}U_\mu(y)\delta(y-x+\hat\mu) + [1+\sigma_\mu]^{\beta\alpha}U_\mu^\dagger(y-\hat\mu)\delta(y-x-\hat\mu)\right\},$$

$$\dim D[U] = \text{volume}\times\text{spin}\times\text{gauge}$$

- **Fermionic degrees of freedom can be integrated out**

$$\int\mathcal{D}[U]\mathcal{D}[\psi,\bar\psi]e^{-S_{\text{gauge}}(U)}\boxed{e^{-S_{\text{ferm}}(\psi,\bar\psi,U)}} \longrightarrow \boxed{\int\mathcal{D}[\psi,\bar\psi]e^{-S_{\text{ferm}}(\psi,\bar\psi,U)} = \prod_{f=1}^{N_f}\det D_f(U)}$$

Fernando Romero-López, MIT

Fermionic theories

- The fermionic part of the action

Dirac-Wilson operator

$$S_{\text{ferm}}(U,\psi,\bar\psi) = \sum_{f=1}^{N_f}\sum_{x,y}\bar\psi_f^{\beta}(y)D[U](y,x)^{\beta\alpha}\psi_f^{\alpha}(x)$$

$$D[U](y,x)^{\beta\alpha} = \delta(y-x)\delta^{\beta\alpha} - \kappa\sum_{\mu=0,1}\left\{[1-\sigma_\mu]^{\beta\alpha}U_\mu(y)\delta(y-x+\hat\mu) + [1+\sigma_\mu]^{\beta\alpha}U_\mu^\dagger(y-\hat\mu)\delta(y-x-\hat\mu)\right\},$$

$$\dim D[U] = \text{volume}\times\text{spin}\times\text{gauge}$$

- Fermionic degrees of freedom can be integrated out

$$\int\mathcal{D}[U]\mathcal{D}[\psi,\bar\psi]e^{-S_{\text{gauge}}(U)}\boxed{e^{-S_{\text{ferm}}(\psi,\bar\psi,U)}} \longrightarrow \boxed{\int\mathcal{D}[\psi,\bar\psi]e^{-S_{\text{ferm}}(\psi,\bar\psi,U)} = \prod_{f=1}^{N_f}\det D_f(U)}$$

- Full action can be expressed only in terms of gauge variables: use gauge flow architectures!

$$S_E(U) = \underbrace{-\beta\sum_x \mathrm{Re}\,P(x)}_{\text{(gauge part)}} - \underbrace{\log\det D[U]^\dagger D[U]}_{\text{(assuming } N_f=2)}$$

"exact determinant"

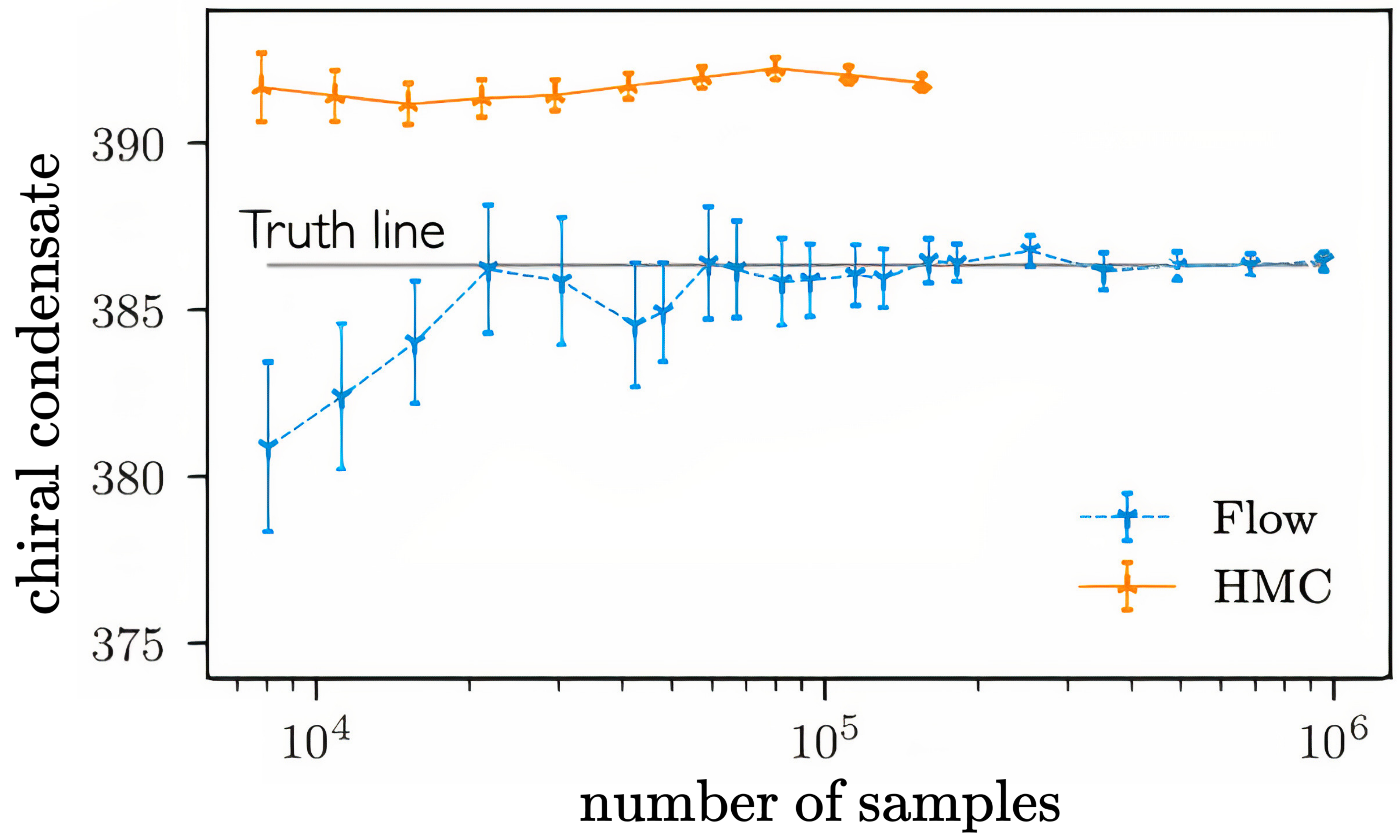
Fernando Romero-López, MIT

The Schwinger model at criticality

Critical parameters:

- Vanishing fermion mass
- Diverging correlation length

! Hardest to simulate in standard approaches



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

The Schwinger model at criticality

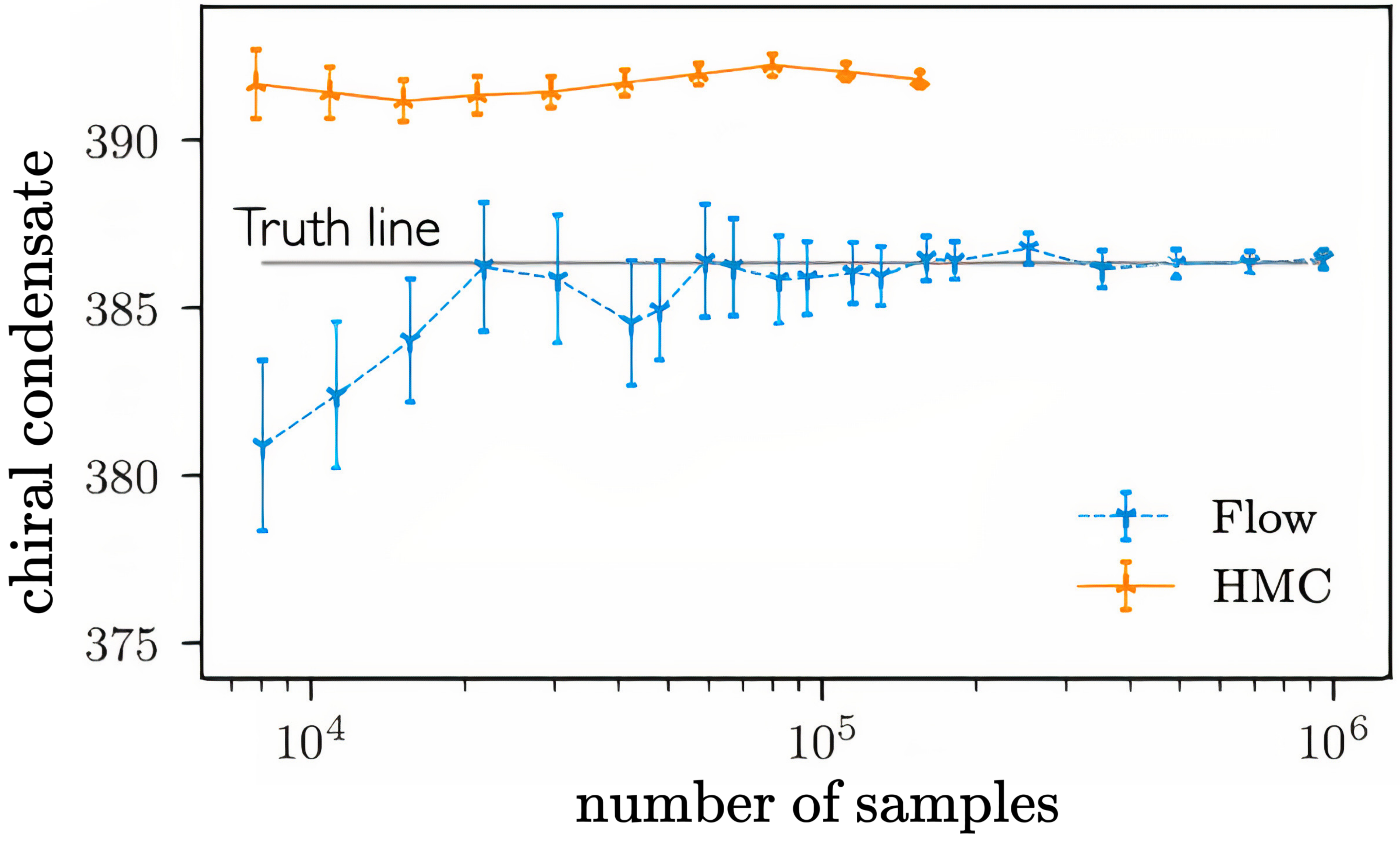
Critical parameters:

- Vanishing fermion mass
- Diverging correlation length

! Hardest to simulate in standard approaches

✗ HMC shows biased results with underestimated errors

✓ Flow-based sampling provides correct results



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

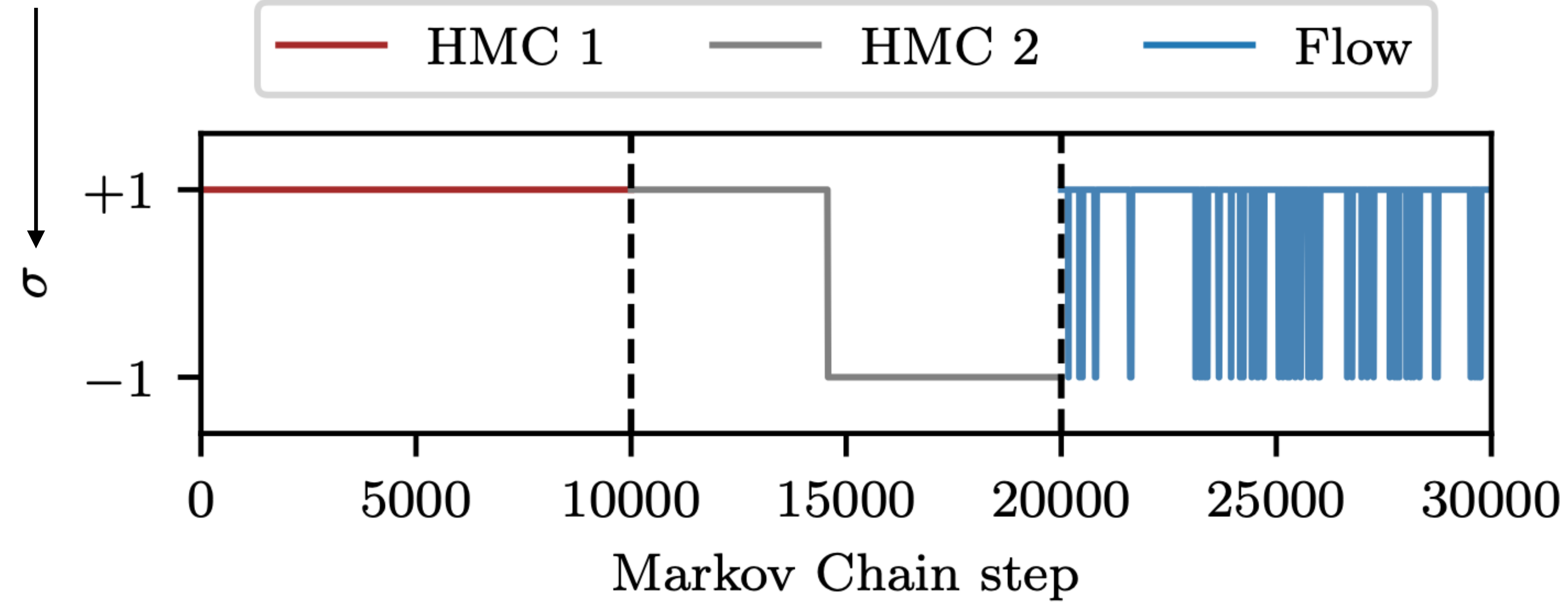
The Schwinger model at criticality

Critical parameters:

- Vanishing fermion mass
- Diverging correlation length

! Hardest to simulate in standard approaches

(topological observable)
sign of $\det D[U]$



✗ HMC shows biased results with underestimated errors

✓ Flow-based sampling provides correct results

✓ Flow-based sampling mitigates topology freezing even at criticality

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

The cost of the determinant

- Evaluation of the fermion determinant is expensive.

$$S_E(U) = -\beta \sum_x \mathrm{Re}\, P(x) - \log\det D[U]^\dagger D[U]$$

Fernando Romero-López, MIT

The cost of the determinant

- Evaluation of the fermion determinant is expensive.

$$S_E(U) = -\beta \sum_x \mathrm{Re}\, P(x) - \log\det D[U]^\dagger D[U]$$

- The cost of determinant computation is $\sim O(n^3)$

- In QCD, the Dirac operator is a matrix of dimension:

$$n = 2 \times 4 \times N_c \times L^3 \times T \sim 10^9$$

Not feasible for QCD-scale calculations!

The cost of the determinant

- Evaluation of the fermion determinant is expensive.

$$S_E(U) = -\beta \sum_x \mathrm{Re}\, P(x) - \log\det D[U]^\dagger D[U]$$

- The cost of determinant computation is $\sim O(n^3)$

- In QCD, the Dirac operator is a matrix of dimension:

$$n = 2 \times 4 \times N_c \times L^3 \times T \sim 10^9$$

Not feasible for QCD-scale calculations!

- Scalable approach: use stochastic determinant estimators: Pseudofermions!

Pseudofermions

- Stochastically estimate determinant using auxiliary degrees of freedom

Assuming N_f=2

$$\det DD^\dagger = \frac{1}{Z_{\mathcal{N}}} \int \mathcal{D}\phi \, e^{-\phi^\dagger (DD^\dagger)^{-1} \phi}$$

Pseudofermions

Pseudofermions

- Stochastically estimate determinant using auxiliary degrees of freedom

Assuming $N_f=2$ $\longrightarrow$

$$\det DD^\dagger = \frac{1}{Z_\mathcal{N}} \int \mathcal{D}\phi \, e^{-\phi^\dagger (DD^\dagger)^{-1} \phi}$$

Pseudofermions

- Joint target distribution

$$p(U,\phi) = \frac{1}{Z} e^{-S_g(U) - S_{\rm pf}(U,\phi)} \quad \text{with} \quad S_{\rm pf}(\phi, U) = \phi^\dagger \left[D(U) D^\dagger(U)\right]^{-1} \phi^\dagger$$

- Only need the Dirac operator applied to the PF field: scales linearly with the lattice volume

Fernando Romero-López, MIT

Joint flow models

$$p(U,\phi) = p(U)p(\phi \mid U)$$

$$p(U) \propto \det DD^\dagger(U) e^{-S_g(U)}$$

"marginal"

$$p(\phi|U) \propto \frac{1}{\det DD^\dagger(U)} e^{-S_{PF}(\phi|U)}$$

"conditional"

Fernando Romero-López, MIT

Joint flow models

$$p(U,\phi) = p(U)p(\phi \mid U)$$

$$p(U) \propto \det DD^\dagger(U) e^{-S_g(U)}$$

"marginal"

$$p(\phi|U) \propto \frac{1}{\det DD^\dagger(U)} e^{-S_{PF}(\phi|U)}$$

"conditional"

$f_m(\chi)$

$f_c(z \mid U)$

Different flow models to approximate marginal and conditional distributions

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

Joint flow models

$$p(U,\phi) = p(U)p(\phi \mid U)$$

$$p(U) \propto \det DD^{\dagger}(U)e^{-S_g(U)}$$

"marginal"

$$p(\phi|U) \propto \frac{1}{\det DD^{\dagger}(U)} e^{-S_{PF}(\phi|U)}$$

"conditional"

$f_m(\chi)$

existing gauge architectures

[Kanwar et al, 2003.06413]

[Boyda et al, 2008.05456]

$f_c(z \mid U)$

Different flow models to approximate marginal and conditional distributions

new conditional architectures

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

Joint flow models

$$p(U,\phi) = p(U)p(\phi \mid U)$$

$$p(U) \propto \det DD^\dagger(U) e^{-S_g(U)}$$

"marginal"

$$p(\phi|U) \propto \frac{1}{\det DD^\dagger(U)} e^{-S_{PF}(\phi|U)}$$

"conditional"

$$\chi \rightarrow \boxed{f_m(\chi)} \text{ --- } U \rightarrow \boxed{f_c(z \mid U)}$$

existing gauge architectures

[Kanwar et al, 2003.06413]

[Boyda et al, 2008.05456]

new conditional architectures

Different flow models to approximate marginal and conditional distributions

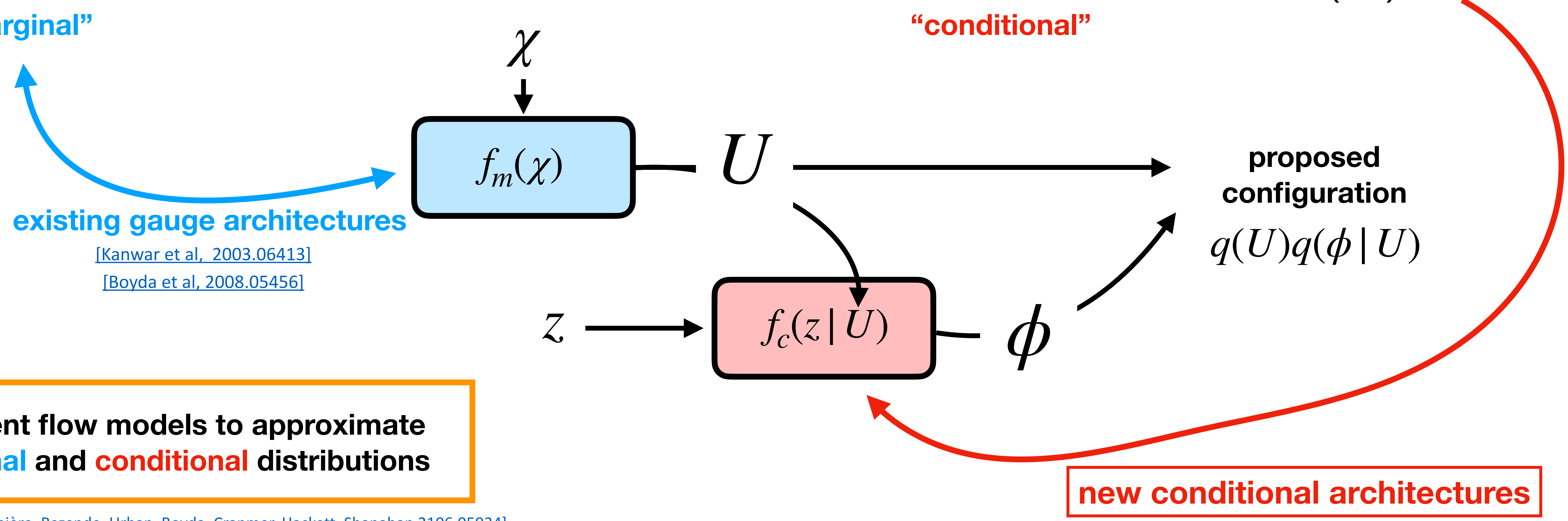
[Albergo, Kanwar, Racanière, Rezende, Urban, Boyda, Cranmer, Hackett, Shanahan 2106.05934]

Joint flow models

$$p(U,\phi) = p(U)p(\phi \mid U)$$

$$p(U) \propto \det DD^\dagger(U) e^{-S_g(U)}$$

$$p(\phi|U) \propto \frac{1}{\det DD^\dagger(U)} e^{-S_{PF}(\phi|U)}$$



Different flow models to approximate marginal and conditional distributions

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

Fernando Romero-López, MIT

Gauge-equivariant conditional models

- **Train a flow to map an uncorrelated gaussian into a correlated one:**

$$r(z) \propto e^{-z^\dagger z} \xrightarrow{f_c(z\,|\,U)} \boxed{q(\phi\,|\,U)} \propto e^{-\phi^\dagger A(U)\phi} \simeq e^{-\phi^\dagger \left(D(U)D^\dagger(U)\right)^{-1}\phi}$$

approximates

Fernando Romero-López, MIT

Gauge-equivariant conditional models

- Train a flow to map an uncorrelated gaussian into a correlated one:

$$r(z) \propto e^{-z^\dagger z} \xrightarrow{f_c(z\,|\,U)} \boxed{q(\phi\,|\,U)} \propto e^{-\phi^\dagger A(U)\phi} \simeq e^{-\phi^\dagger \left(D(U)D^\dagger(U)\right)^{-1}\phi}$$

approximates

- Construct flow with expressive gauge-equivariant linear transformations:

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

Fernando Romero-López, MIT

Gauge-equivariant conditional models

- **Train a flow to map an uncorrelated gaussian into a correlated one:**

$$r(z) \propto e^{-z^\dagger z} \xrightarrow{f_c(z\,|\,U)} \boxed{q(\phi\,|\,U)} \propto e^{-\phi^\dagger A(U)\phi} \simeq e^{-\phi^\dagger \left(D(U)D^\dagger(U)\right)^{-1}\phi}$$

approximates

- **Construct flow with expressive gauge-equivariant linear transformations:**

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

parallel-transported neighbor

Fernando Romero-López, MIT

Gauge-equivariant conditional models

- Train a flow to map an uncorrelated gaussian into a correlated one:

$$r(z) \propto e^{-z^\dagger z} \xrightarrow{f_c(z\,|\,U)} \boxed{q(\phi\,|\,U)} \propto e^{-\phi^\dagger A(U)\phi} \simeq e^{-\phi^\dagger \left(D(U)D^\dagger(U)\right)^{-1}\phi}$$

approximates

- Construct flow with expressive gauge-equivariant linear transformations:

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

parallel-transported neighbor

NN outputs (gauge-invariant inputs)

Fernando Romero-López, MIT

Gauge-equivariant conditional models

- Train a flow to map an uncorrelated gaussian into a correlated one:

$$\boxed{r(z) \propto e^{-z^\dagger z} \xrightarrow{f_c(z\,|\,U)} \boxed{q(\phi\,|\,U)} \propto e^{-\phi^\dagger A(U)\phi} \simeq e^{-\phi^\dagger \left(D(U)D^\dagger(U)\right)^{-1}\phi}}$$

approximates

- Construct flow with expressive gauge-equivariant linear transformations:

$$\phi'(x) = A(U)\,\phi(x) + B(U)\,U_\mu(x)\phi(x+\mu) \xrightarrow{\text{gauge trafo}} \phi'(x) \to \Omega(x)\,\phi'(x)$$

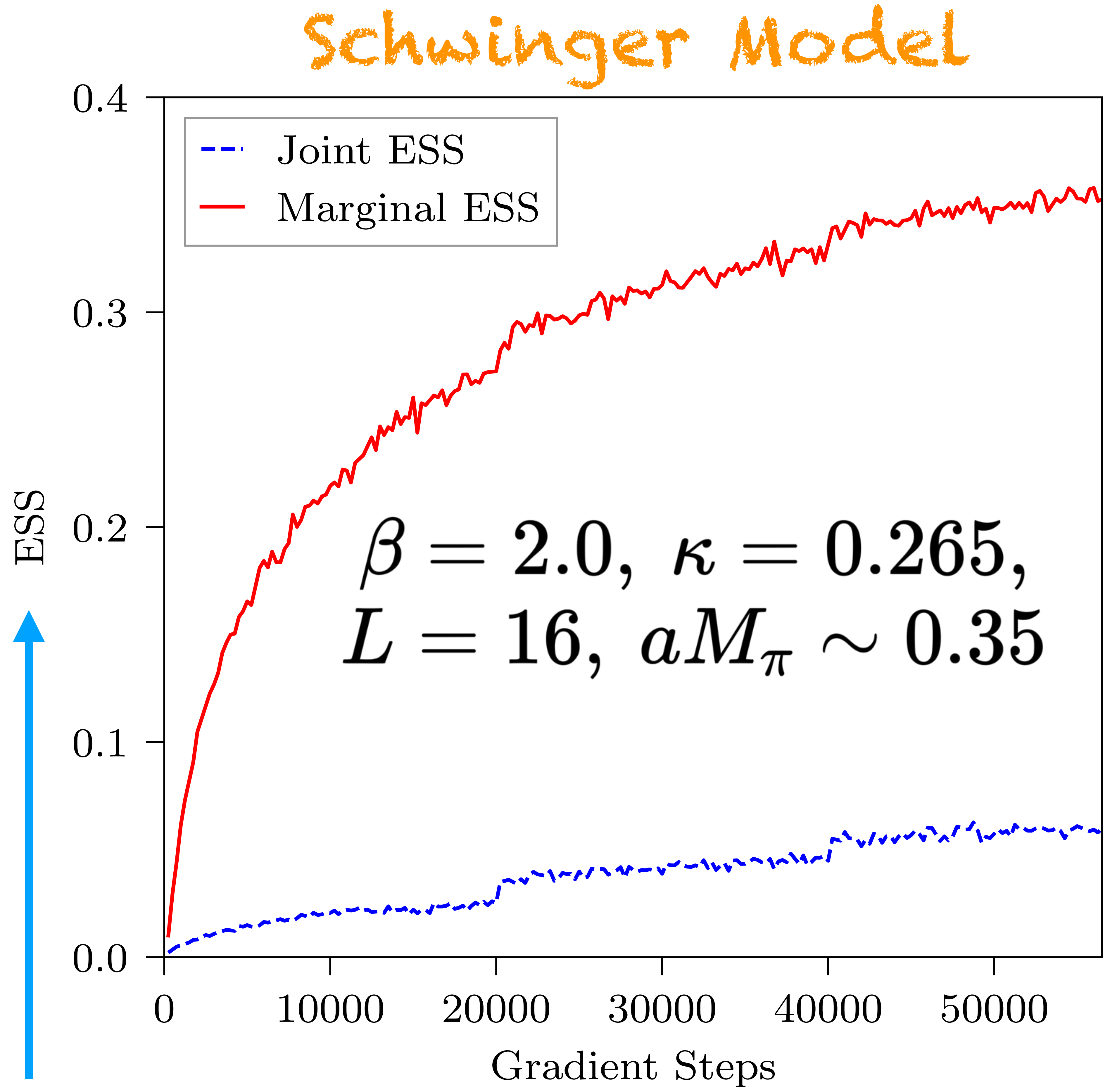
NN outputs
(gauge-invariant inputs)

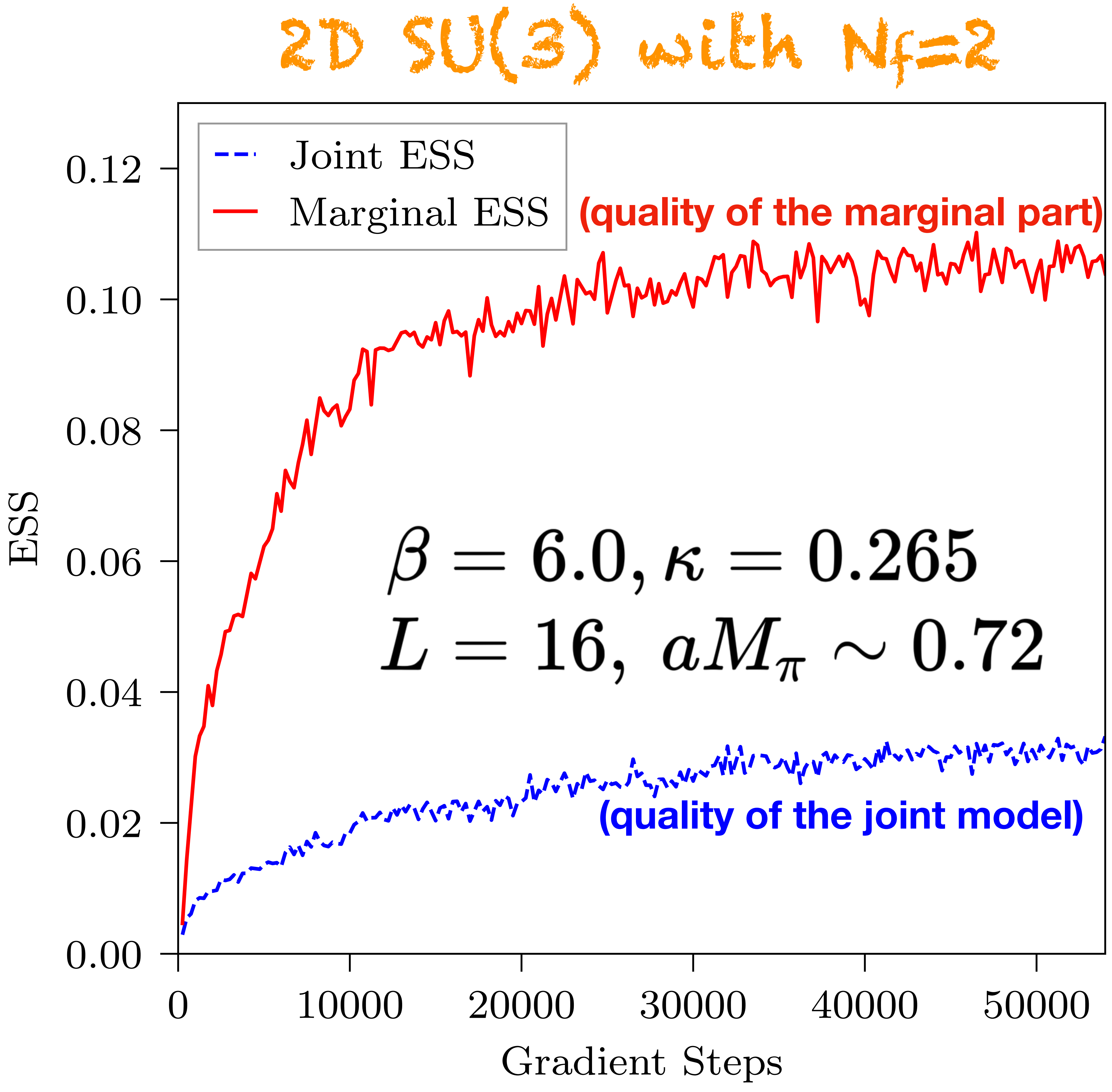
parallel-transported
neighbor

(gauge equivariance)

Fernando Romero-López, MIT

Examples of flow models





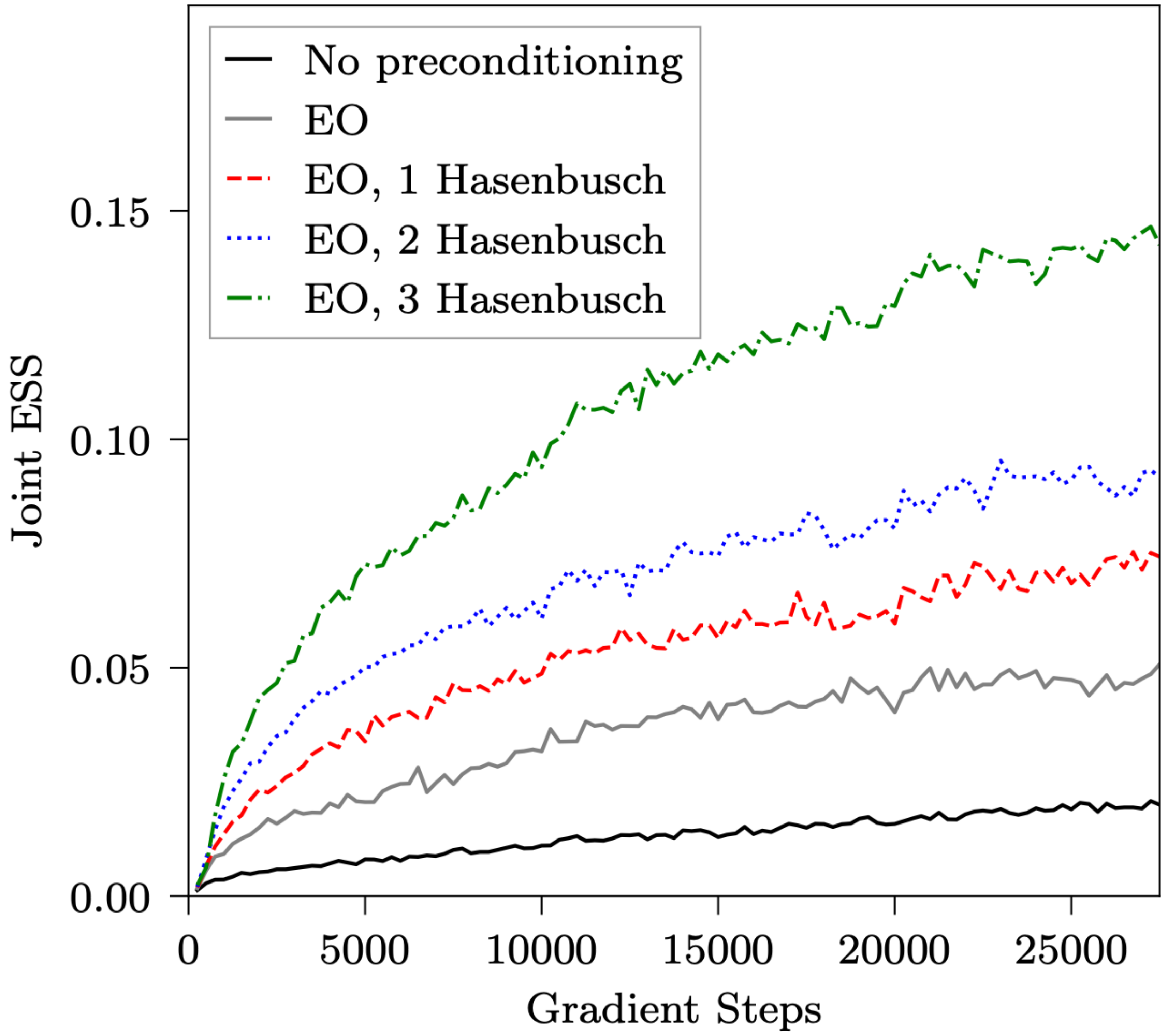
"Effective sample size" ≡ model quality
(ESS=1 for a perfect model)

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

Fernando Romero-López, MIT

Examples of flow models

- Use of preconditioners in flow models
- ✓ Even/odd and Hasenbusch improve model quality



(Examples in Schwinger model)

Fernando Romero-López, MIT

Can we ML generate lattice QCD?

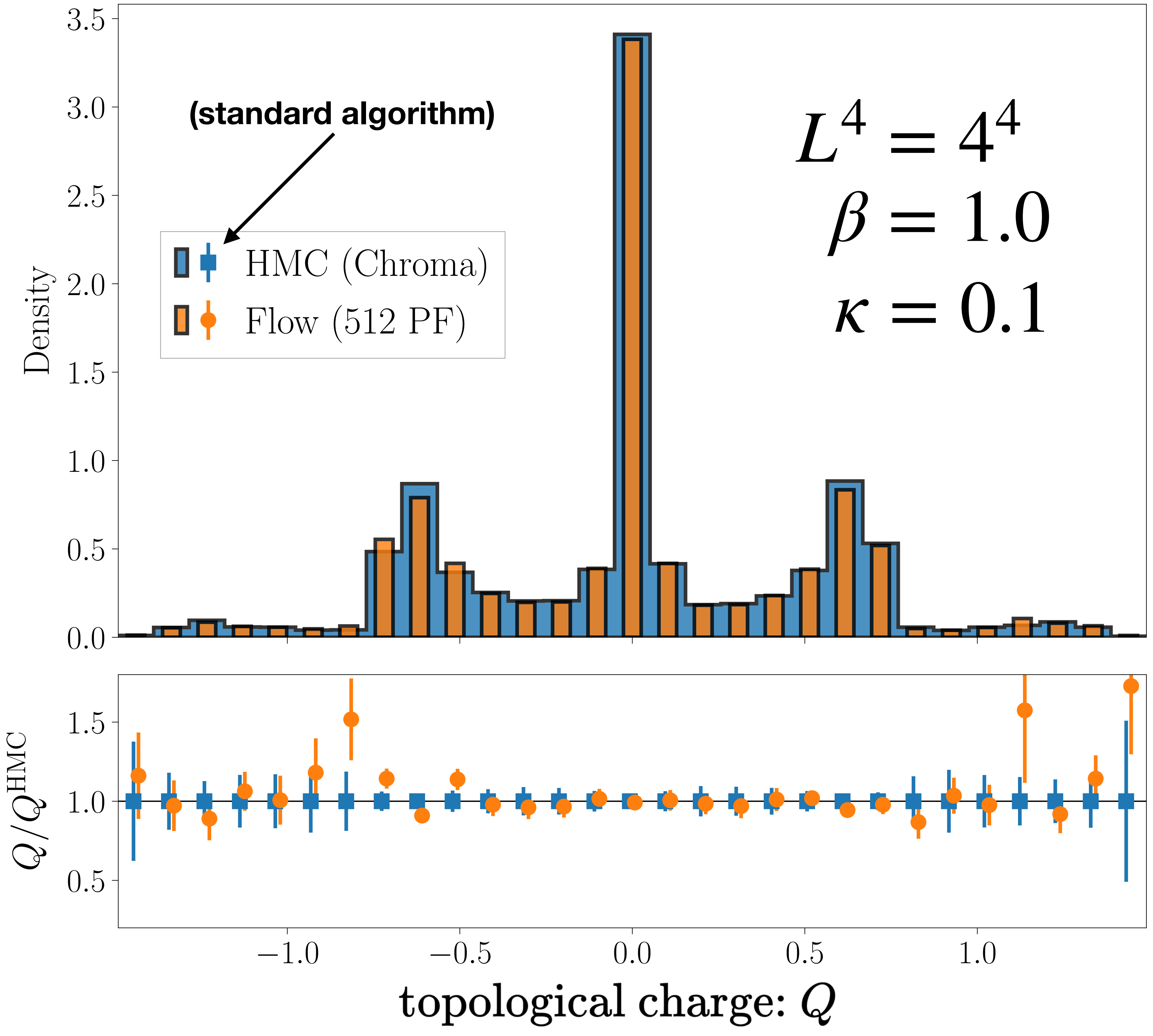
☑ **Overcoming critical slowing down in toy models: scalar theories, Schwinger model** [Abbott ...FRL... et al.] (several works)

Fernando Romero-López, MIT

Can we ML generate lattice QCD?

- ☑ **Overcoming critical slowing down in toy models: scalar theories, Schwinger model** [Abbott ...FRL... et al.] (several works)

- ☑ **Proof of principle in small-scale QCD**



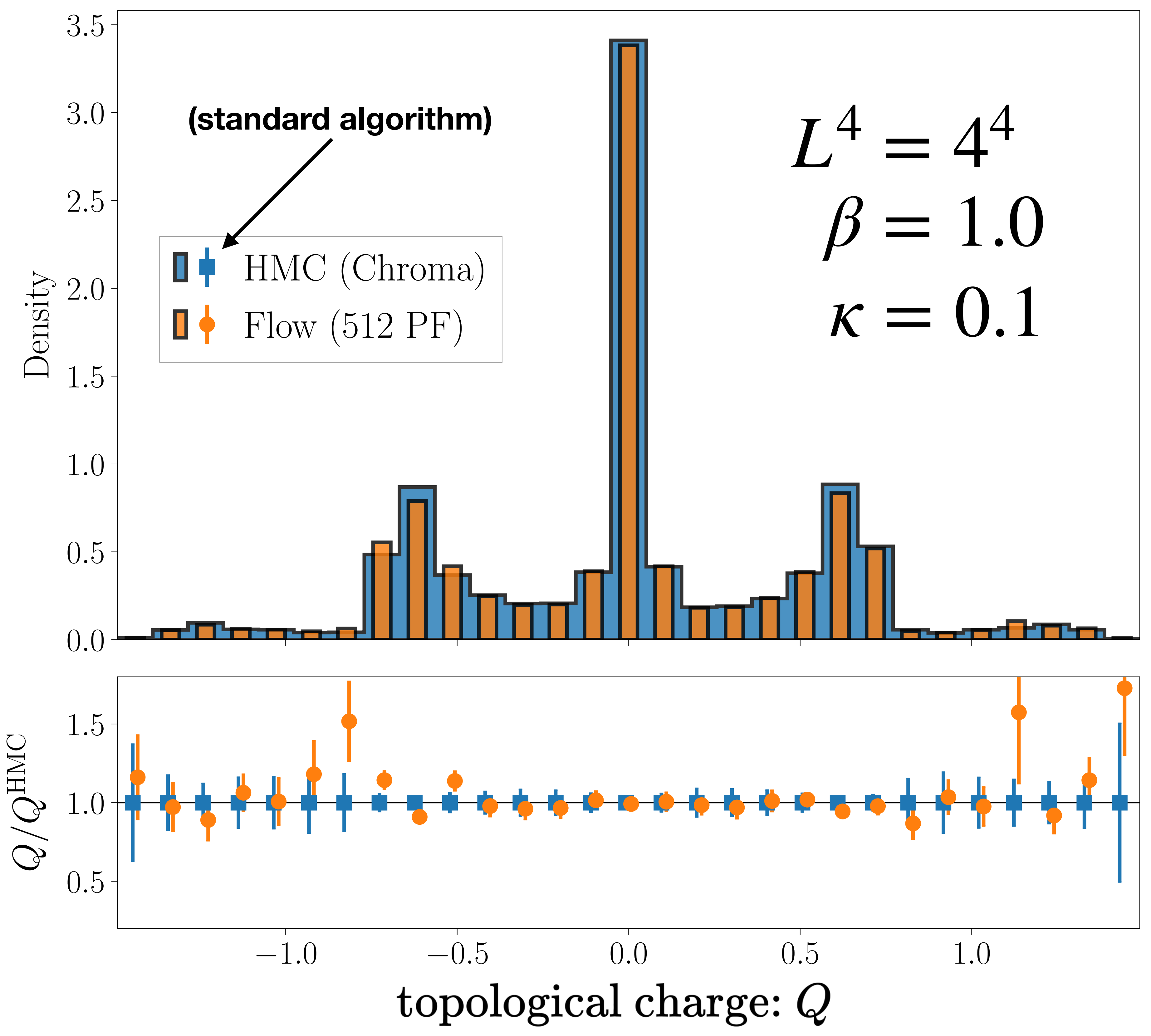
[Abbott, ... , FRL, ..., et al 2208.03832]

Fernando Romero-López, MIT

Can we ML generate lattice QCD?

- ☑ **Overcoming critical slowing down in toy models: scalar theories, Schwinger model** [Abbott ...FRL... et al.] (several works)

- ☑ **Proof of principle in small-scale QCD**



[Abbott, ... , FRL, ..., et al 2208.03832]

- ☐ **Will it work for state-of-the-art Lattice QCD?**

Aspects of scaling and scalability for flow-based sampling of lattice QCD

Ryan Abbott[1,2], Michael S. Albergo[3], Aleksandar Botev[6], Denis Boyda[4,1,2], Kyle Cranmer[5,3], Daniel C. Hackett[1,2], Alexander G. D. G. Matthews[6], Sébastien Racanière[6], Ali Razavi[6], Danilo J. Rezende[6], Fernando Romero-López[1,2], Phiala E. Shanahan[1,2] and Julian M. Urban[1,2]

[arXiv:2211.07541]

> "For flow-based methods, assessing scalability will require direct, experimental investigation of applications to QCD itself, which has only just begun."