Flow-based sampling for lattice gauge theories

Fernando Romero-López

BNL

February 17th

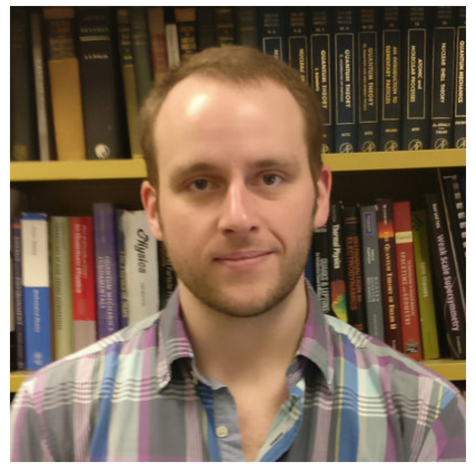
Massachusetts Institute of Technology

fernando@mit.edu

Collaboration (non-exhaustive)

Massachusetts Institute of Technology

- Phiala Shanahan



- Dan Hackett

- Fernando Romero-Lopez

- Ryan Abbot

- Julian Urban

- Denis Boyda

Argonne NATIONAL LABORATORY

Massachusetts Institute of Technology

- Kyle Cranmer

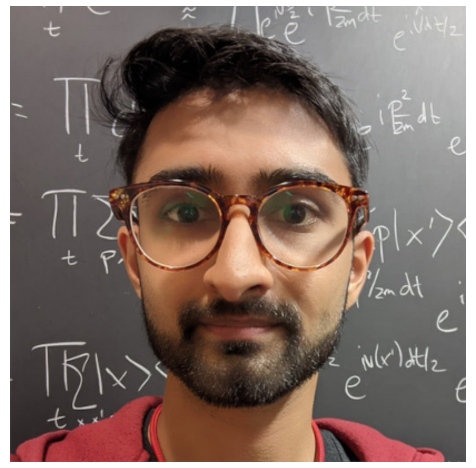
WISCONSIN UNIVERSITY OF WISCONSIN–MADISON

NEW YORK UNIVERSITY

- Michael Albergo

u^b

b UNIVERSITÄT BERN



- Gurtej Kanwar

DeepMind

- Sébastien Racanière

- Danilo Rezende

- Ali Razavi

- Aleksandar Botev

- Alex Matthews

Outline

1. Introduction to Lattice QCD
2. Critical slowing down in Lattice QCD
3. Flow-based sampling for Lattice Field Theory
4. Flow-based sampling for gauge theories
5. Flows for fermionic gauge theories

Introduction to Lattice QCD

Quantum Chromodynamics

> Quantum chromodynamics is conceptually simple. Its realization in nature, however, is usually very complex.
>
> Frank Wilczek

Quantum Chromodynamics

> Quantum chromodynamics is conceptually simple. Its realization in nature, however, is usually very complex.
>
> Frank Wilczek

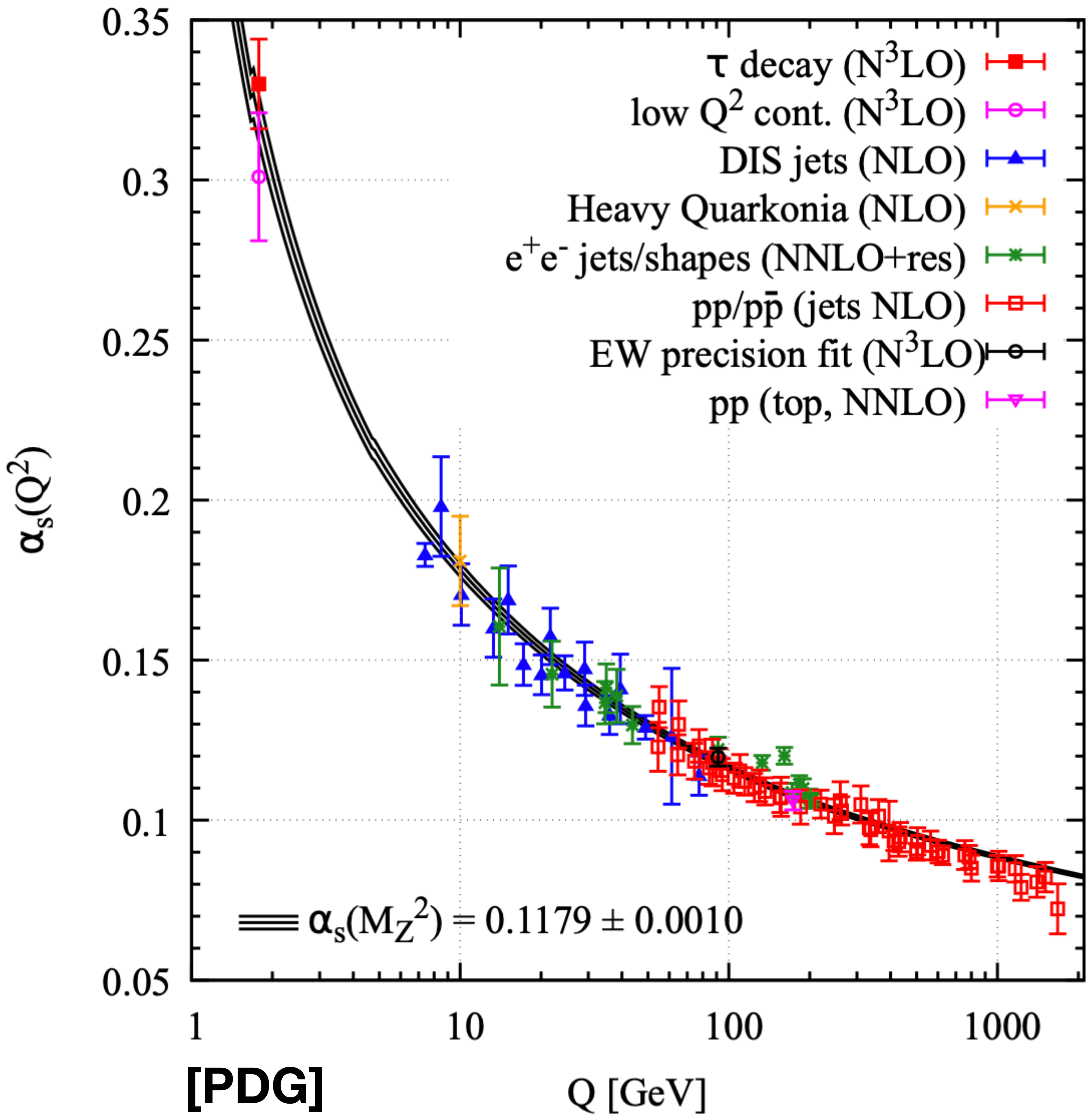
$$\mathscr{L}_{QCD} = \sum_i^{N_f} \bar{q}_i \left(D_\mu \gamma^\mu + m_i \right) q_i + \frac{1}{4 g_s^2} G^a_{\mu\nu} G_a^{\mu\nu}$$

Quantum Chromodynamics

> Quantum chromodynamics is conceptually simple. Its realization in nature, however, is usually very complex.
>
> Frank Wilczek

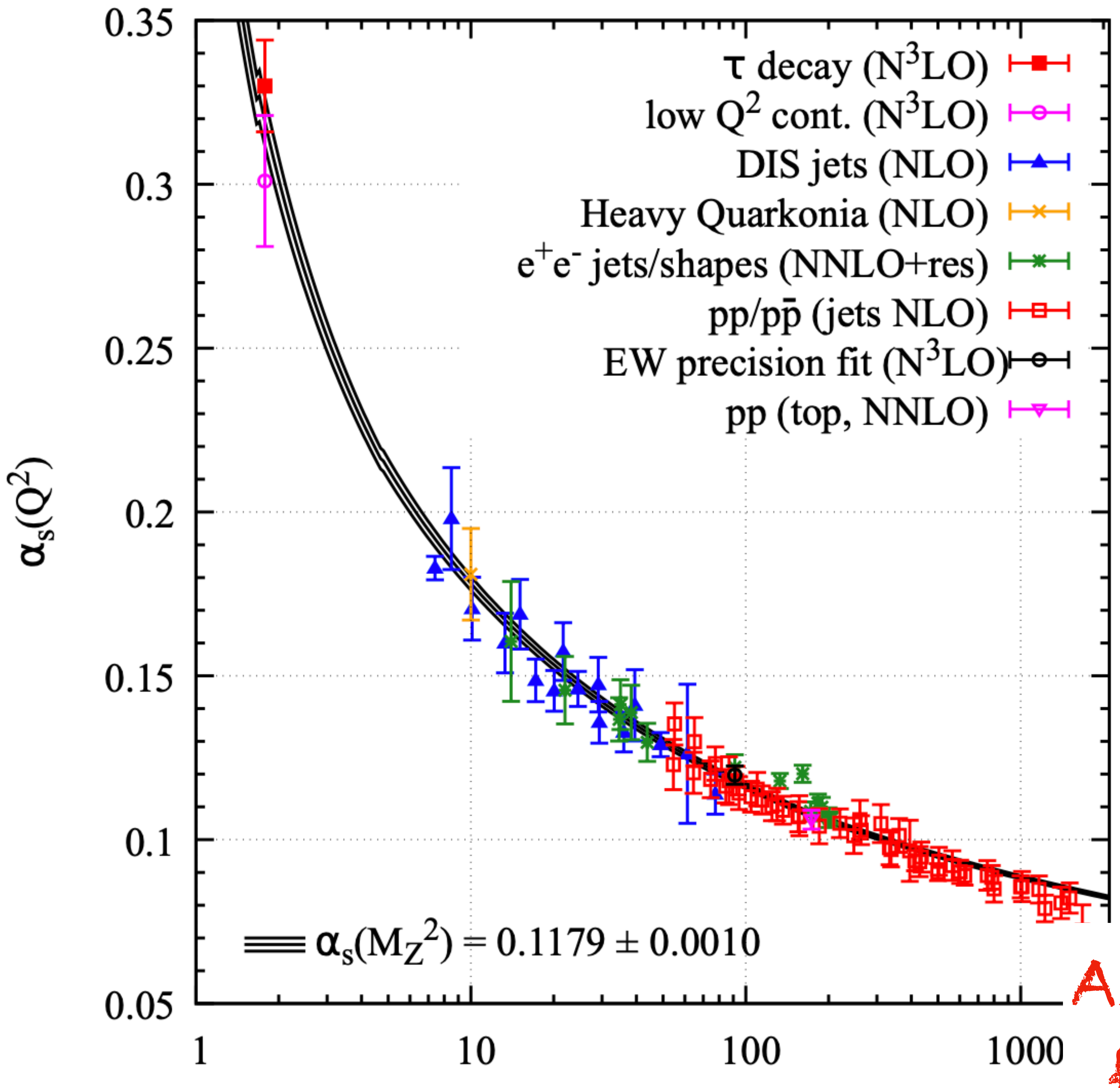
$$\mathcal{L}_{QCD} = \sum_i^{N_f} \bar{q}_i \left(D_\mu \gamma^\mu + m_i \right) q_i + \frac{1}{4 g_s^2} G^a_{\mu\nu} G_a^{\mu\nu}$$

Quantum Chromodynamics



[PDG]

Quantum Chromodynamics



[PDG]

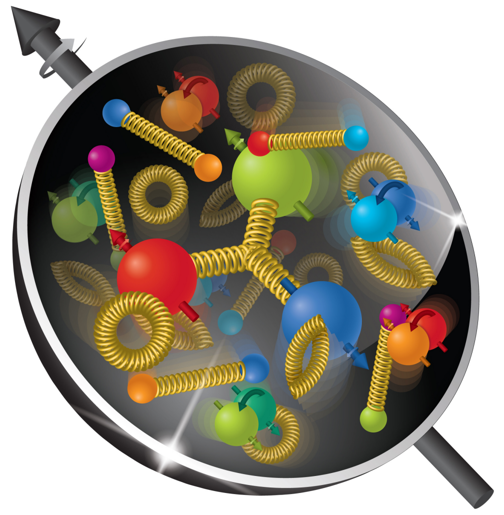
Asymptotic freedom

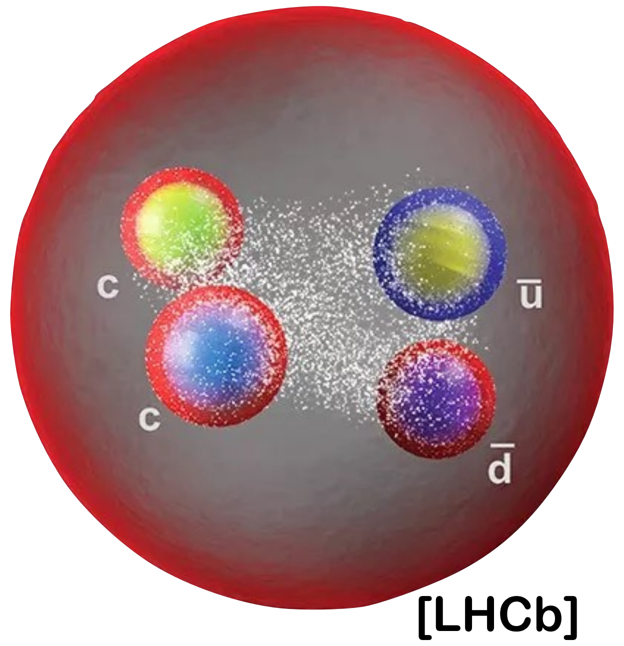
Perturbation theory:

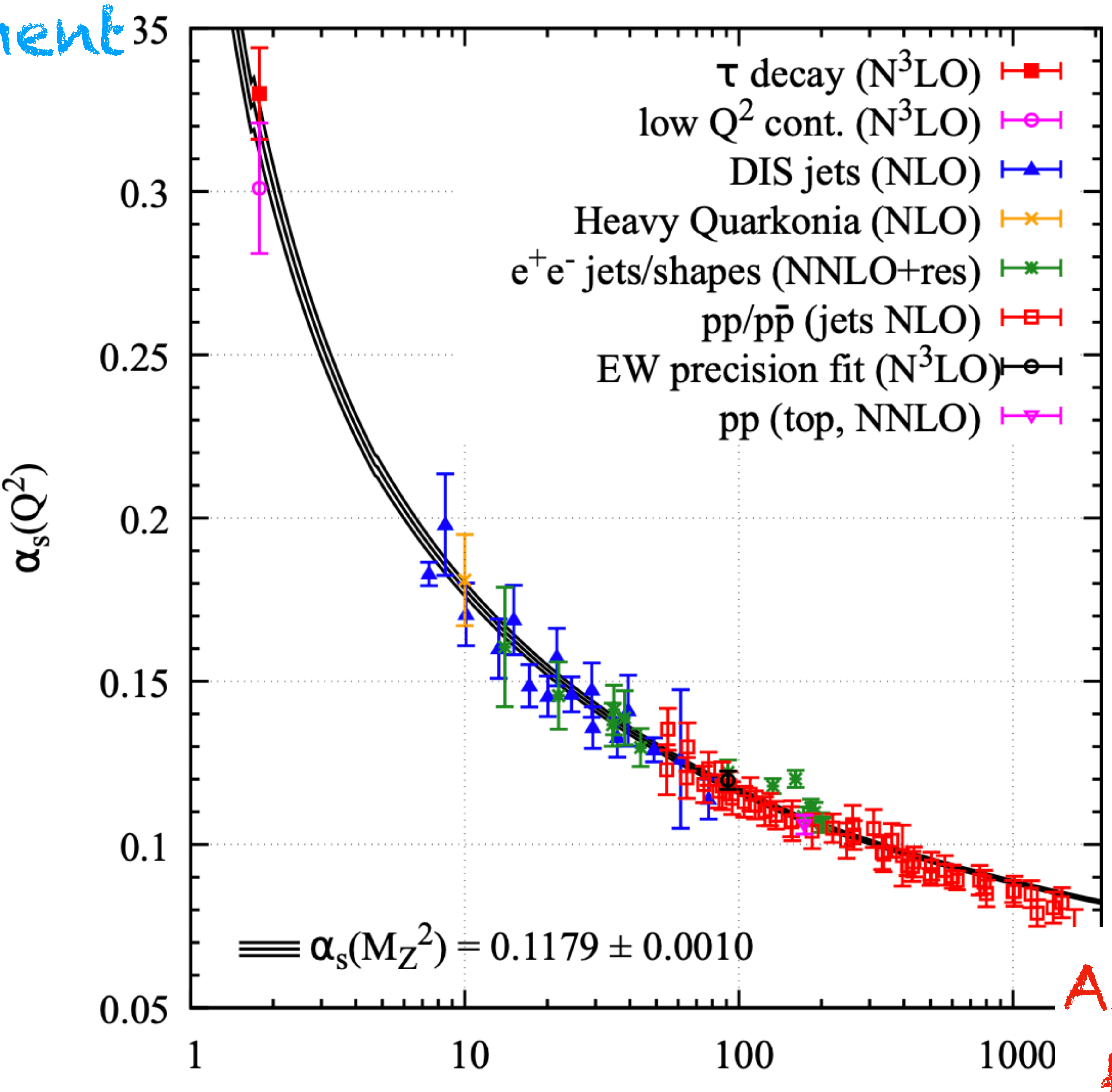
$$\mathcal{O} = \mathcal{O}_0 + \alpha_s \mathcal{O}_1 + \alpha_s^2 \mathcal{O}_2 + \ldots$$

Quantum Chromodynamics

Confinement







[PDG]

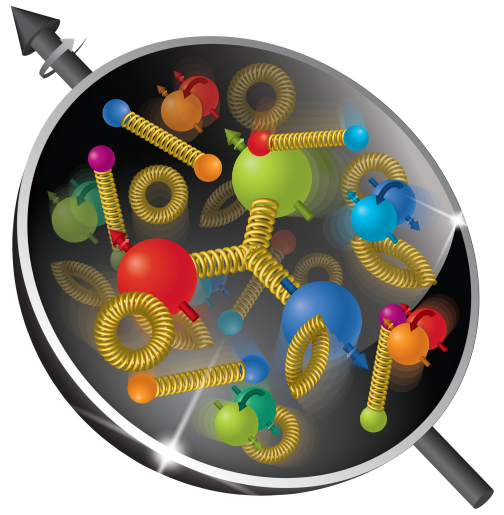
Asymptotic freedom

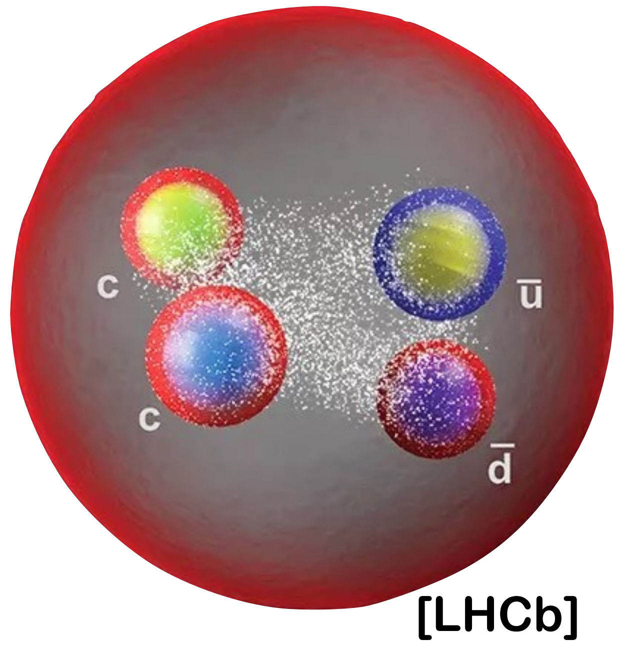
Perturbation theory:

$$\mathcal{O} = \mathcal{O}_0 + \alpha_s \mathcal{O}_1 + \alpha_s^2 \mathcal{O}_2 + \ldots$$

Quantum Chromodynamics

Confinement

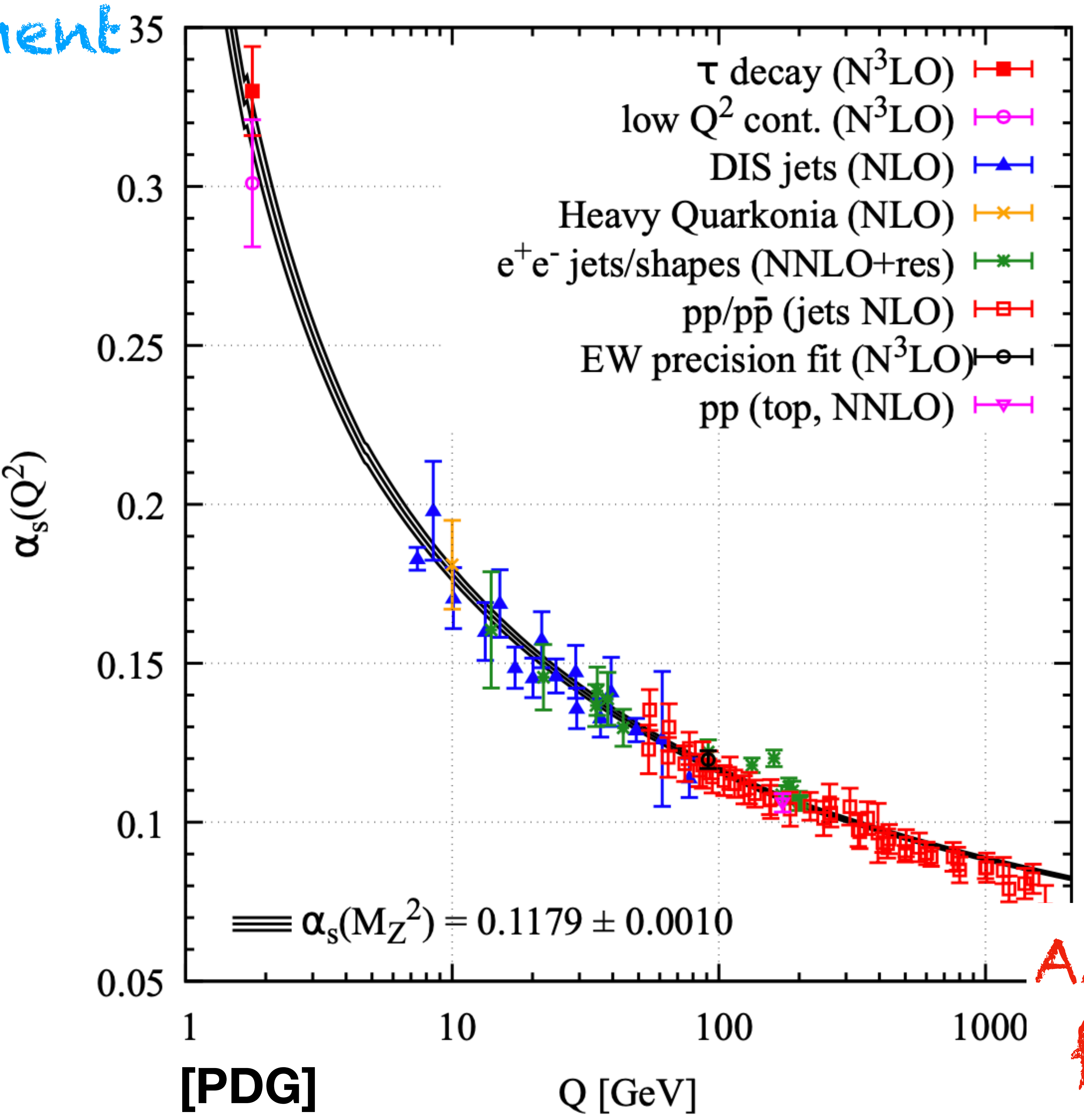




[LHCb]

Non-perturbative at low energies

Lattice QCD



[PDG]

Asymptotic freedom

Perturbation theory:

$$\mathcal{O} = \mathcal{O}_0 + \alpha_s \mathcal{O}_1 + \alpha_s^2 \mathcal{O}_2 + \ldots$$

Fernando Romero-López, MIT

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

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

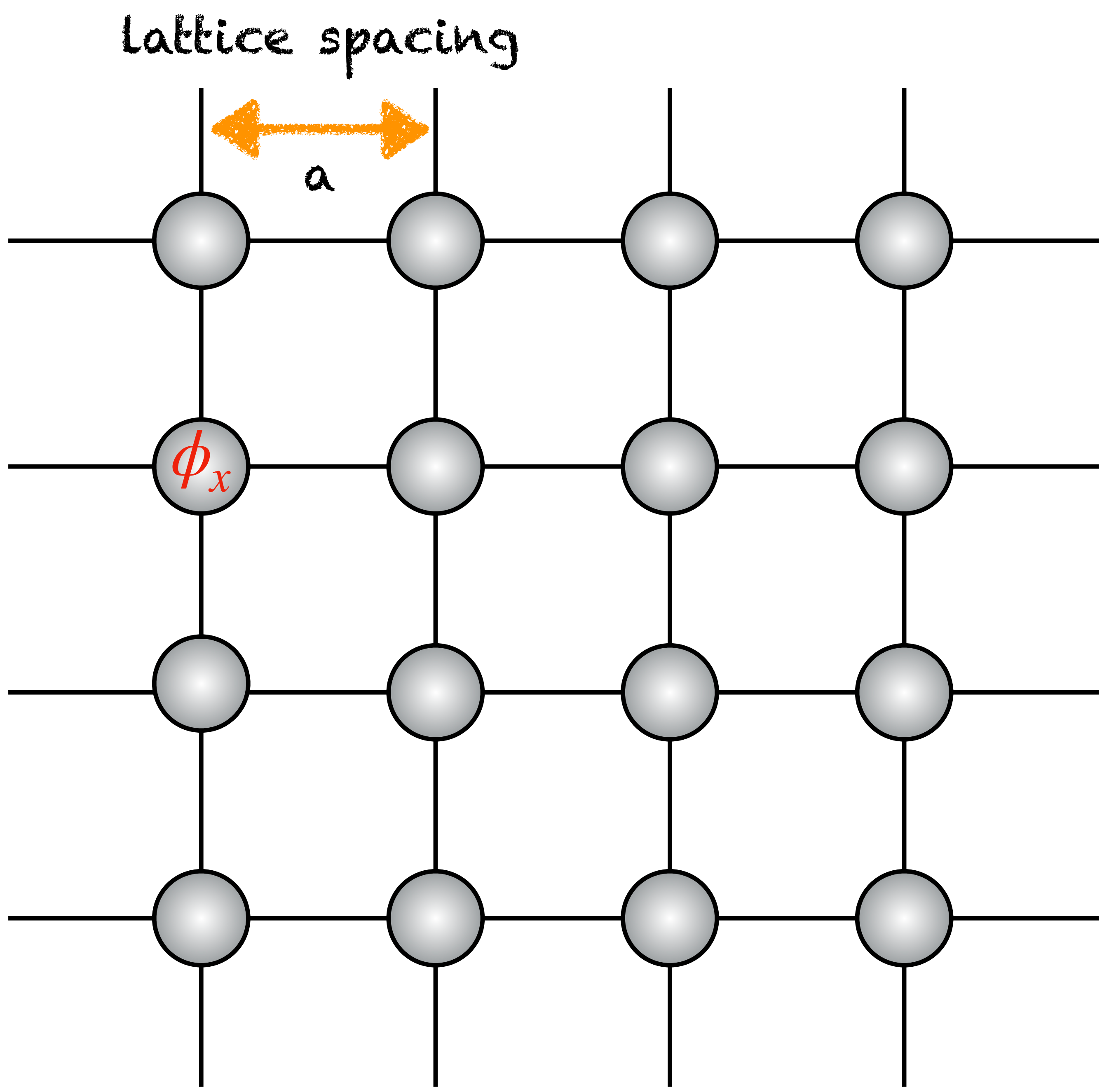
Boltzmann factor

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



Fernando Romero-López, MIT

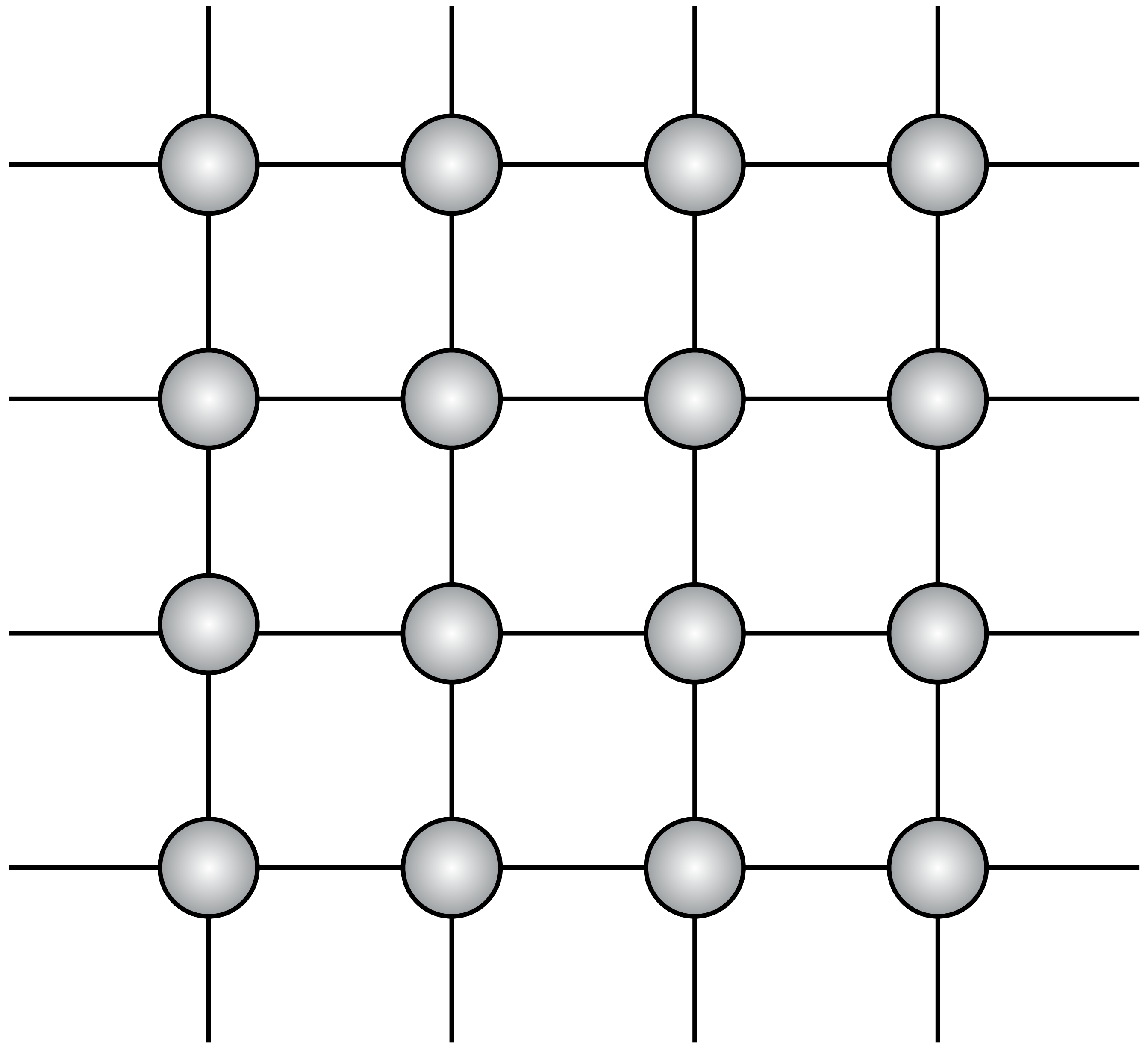
Lattice QCD

- **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

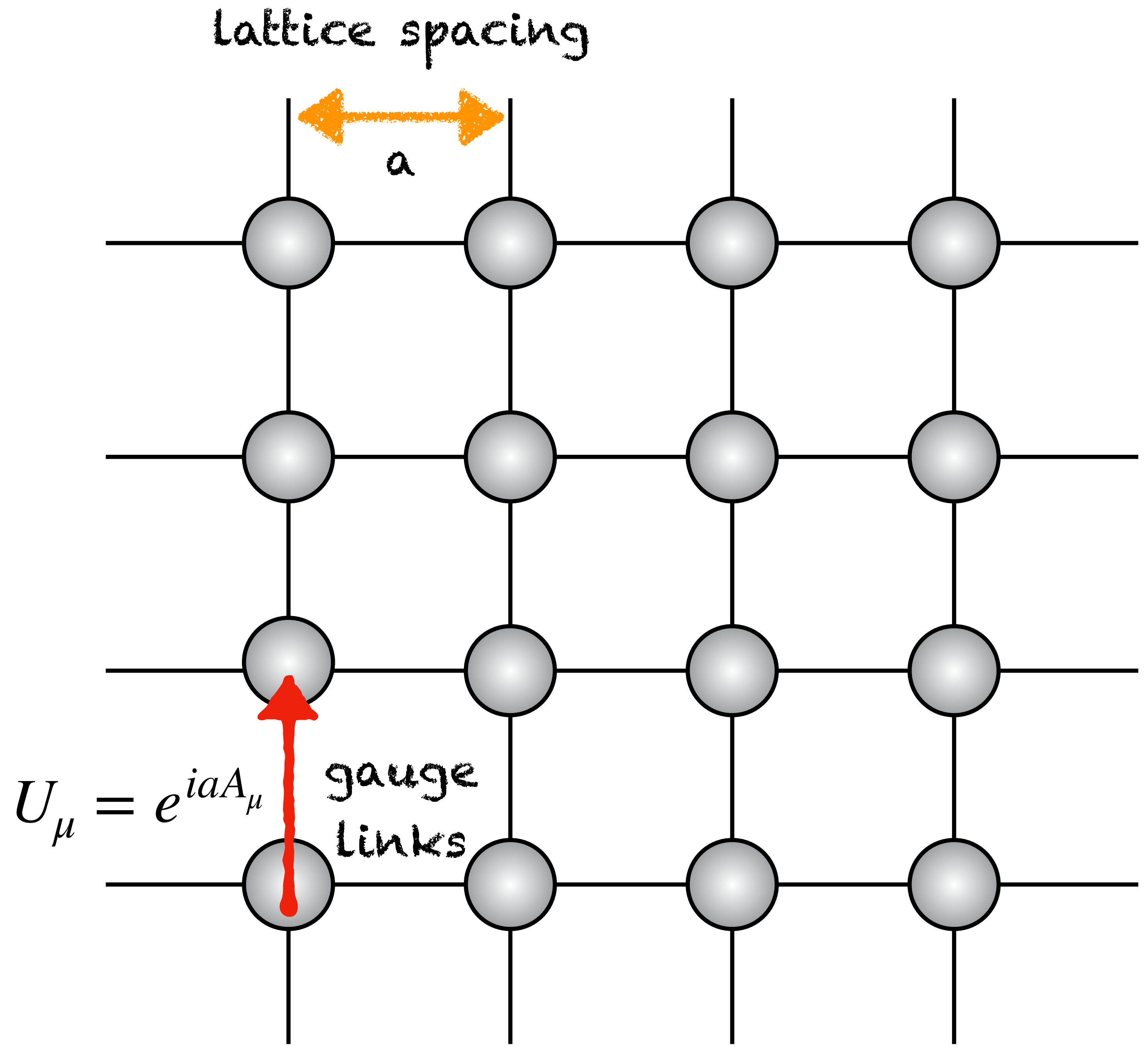
$$\mathcal{Z} = \int D\psi\, D\bar{\psi}\, DA\, e^{-S_E(\psi,\bar{\psi},A_\mu)}$$

Boltzmann factor

  - Discretize gauge fields and fermion fields:

    →Under control but technical

    →Need continuum limit



Fernando Romero-López, MIT

Lattice QCD

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

- **Lattice QCD partition function**

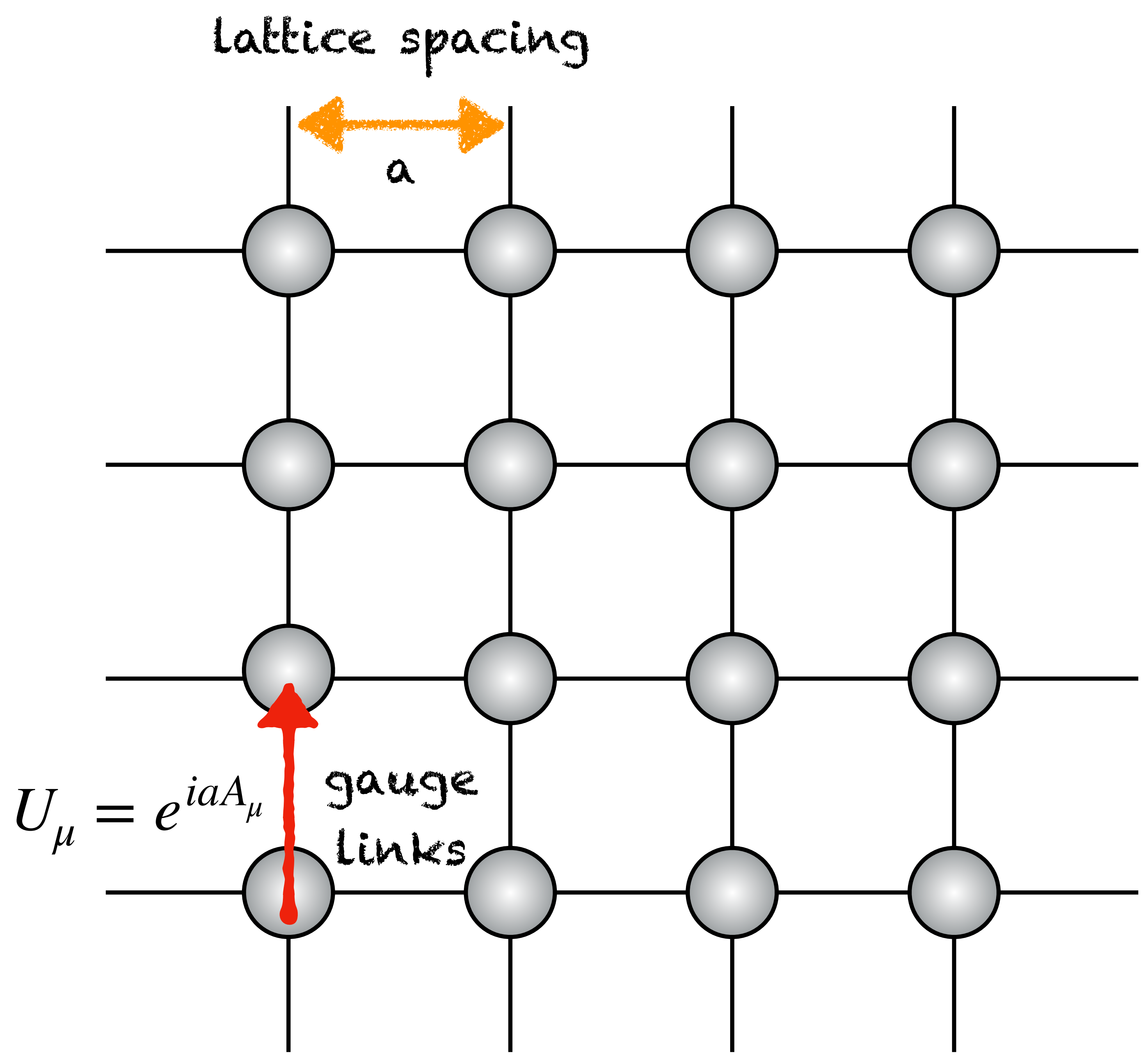
$$\mathcal{Z} = \int D\psi\, D\bar{\psi}\, DA e^{-S_E(\psi,\bar{\psi},A_\mu)}$$

Boltzmann factor

- **Discretize gauge fields and fermion fields:**
  - →Under control but technical
  - →Need continuum limit

- **Compute observables as expectation values**

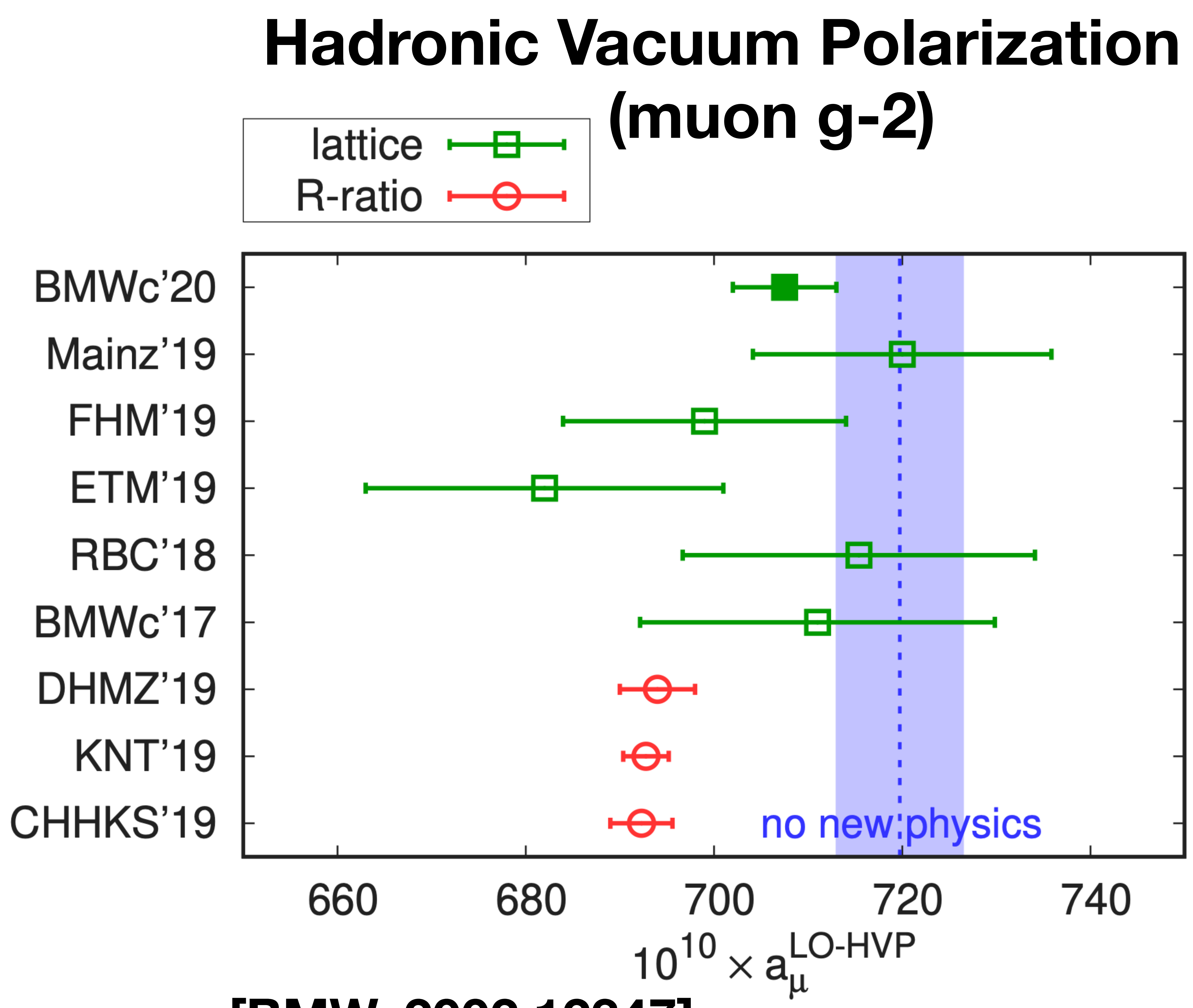
$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int D\psi D\bar{\psi} DA\, \mathcal{O}\, e^{-S(\psi,\bar{\psi},A_\mu)}$$



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

Fernando Romero-López, MIT

Many different applications

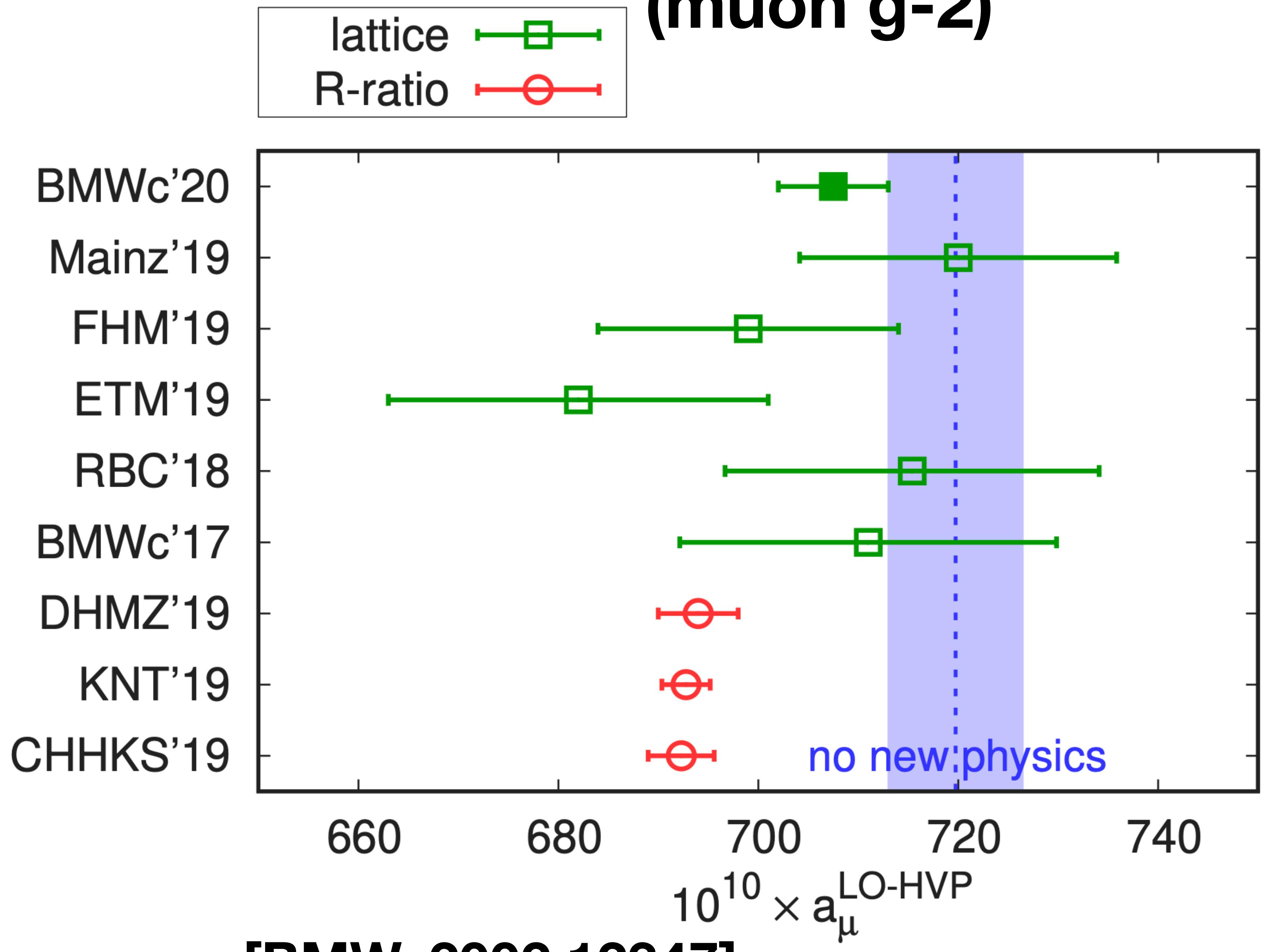


[BMW, 2002.12347]

Fernando Romero-López, MIT

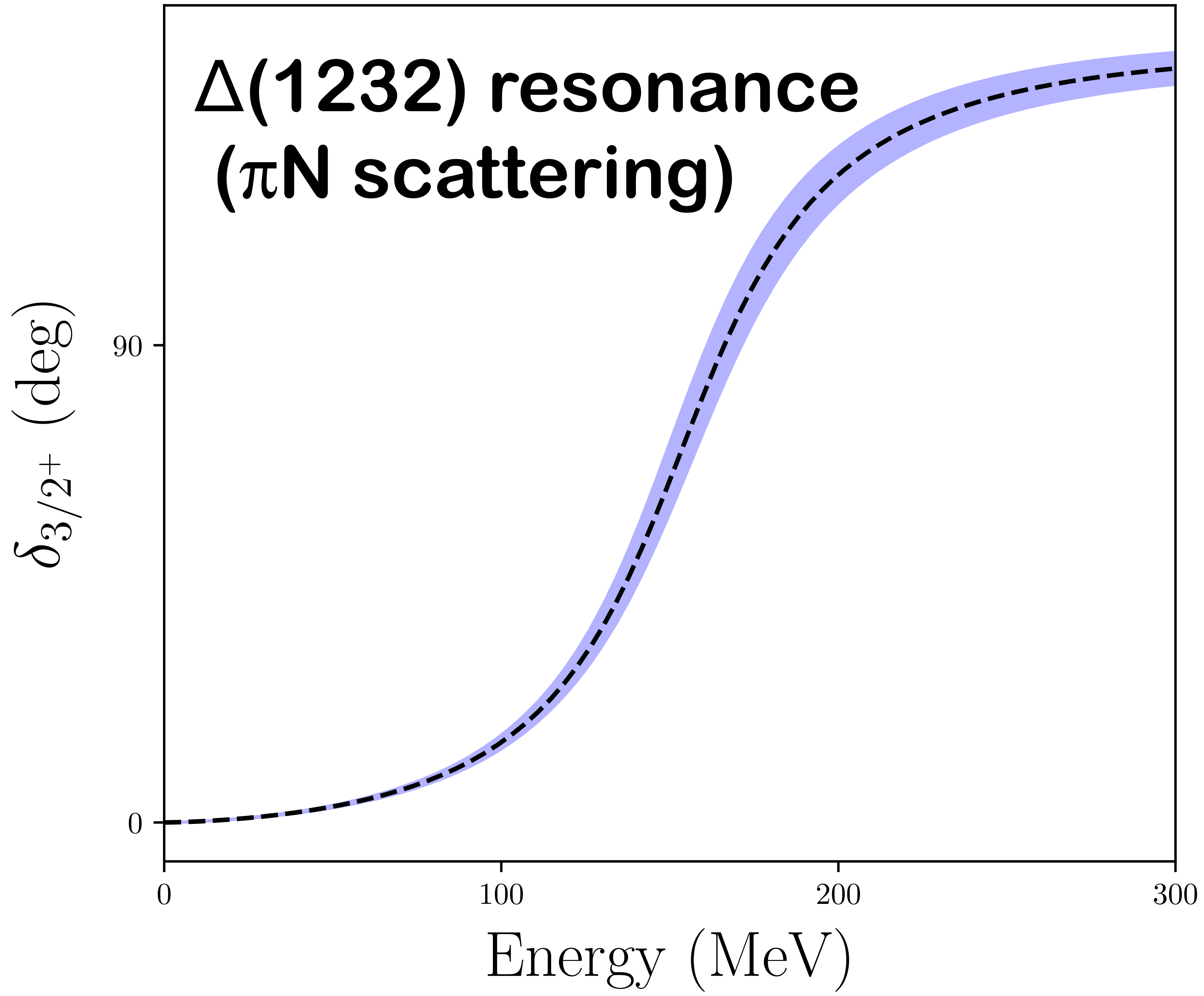
Many different applications

Hadronic Vacuum Polarization (muon g-2)



[BMW, 2002.12347]

Scattering properties of hadrons



[Bulava, Hanlon, Hörz, Morningstar, Nicholson, FRL, Skinner, Vranas, Walker-Loud, 2208.03867]

Critical slowing down in Lattice QCD

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

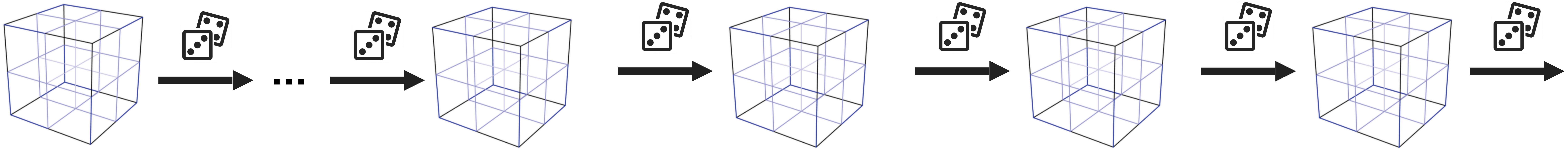
Fernando Romero-López, MIT

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



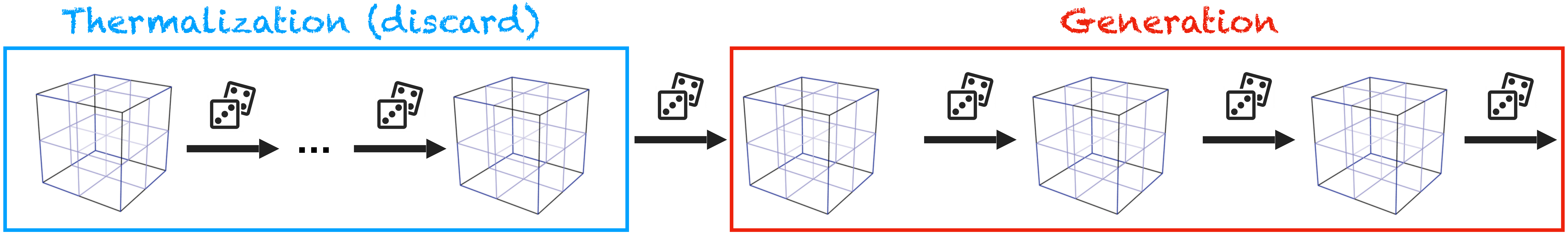
Fernando Romero-López, MIT

Markov Chain Monte Carlo

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

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

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



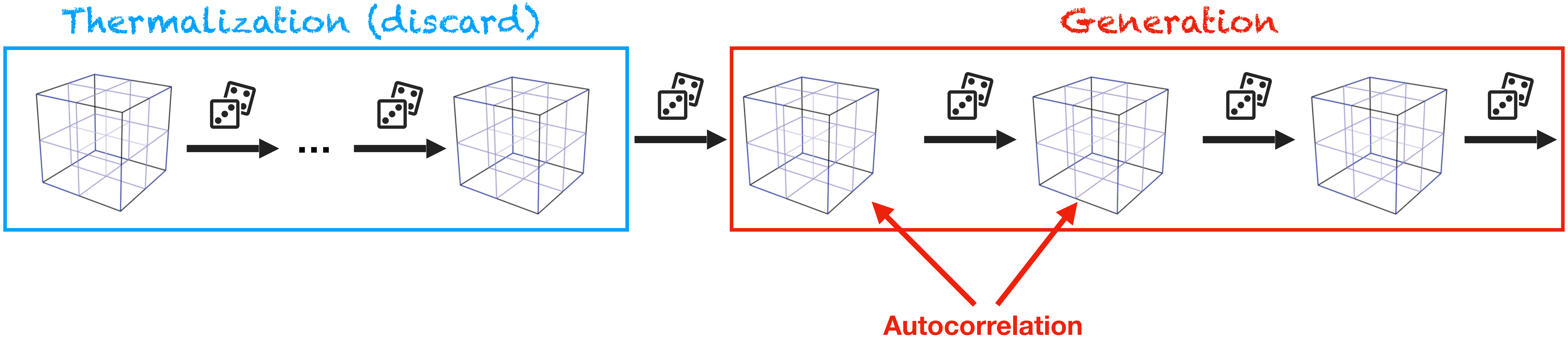
Fernando Romero-López, MIT

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



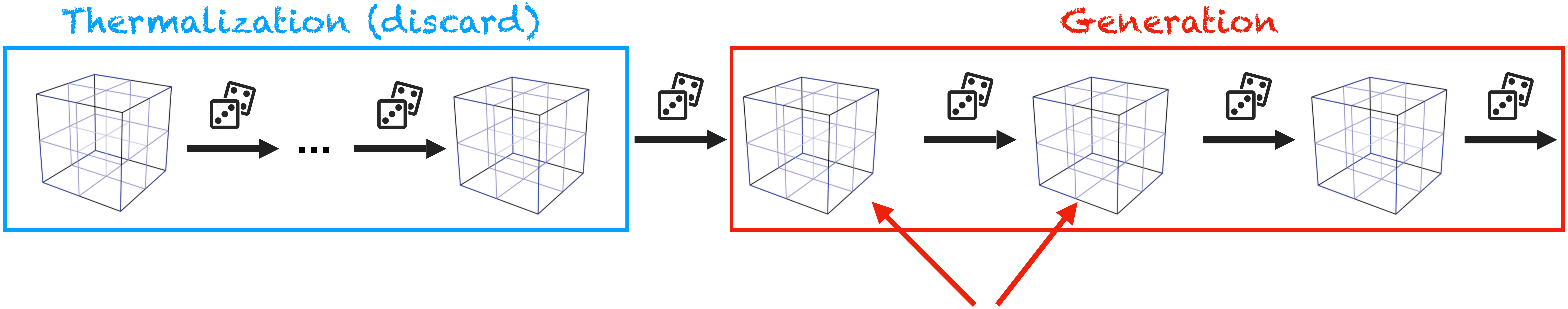
Fernando Romero-López, MIT

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



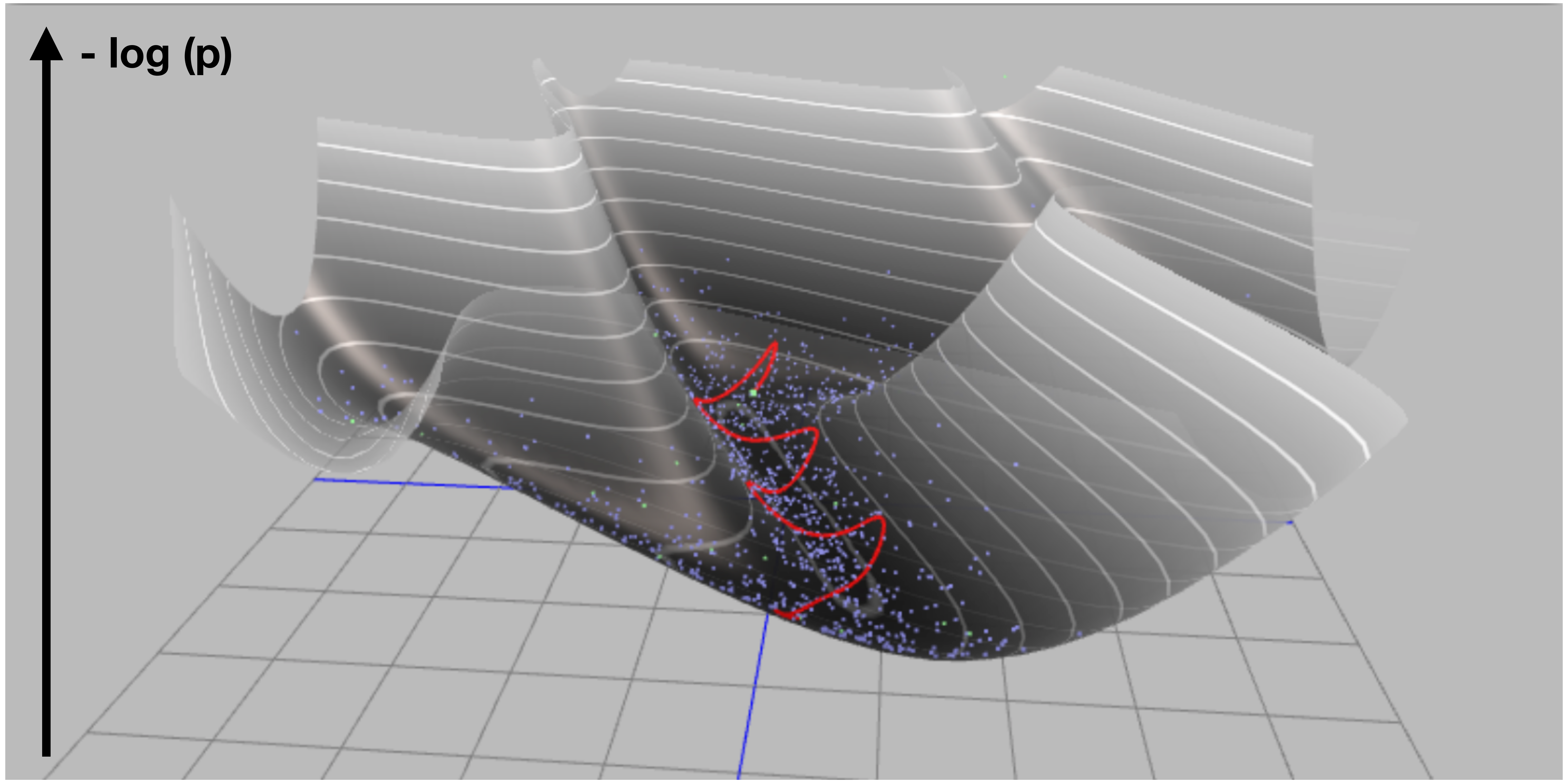
Autocorrelation $\equiv$ **Computational cost**

Fernando Romero-López, MIT

Generation of gauge configs

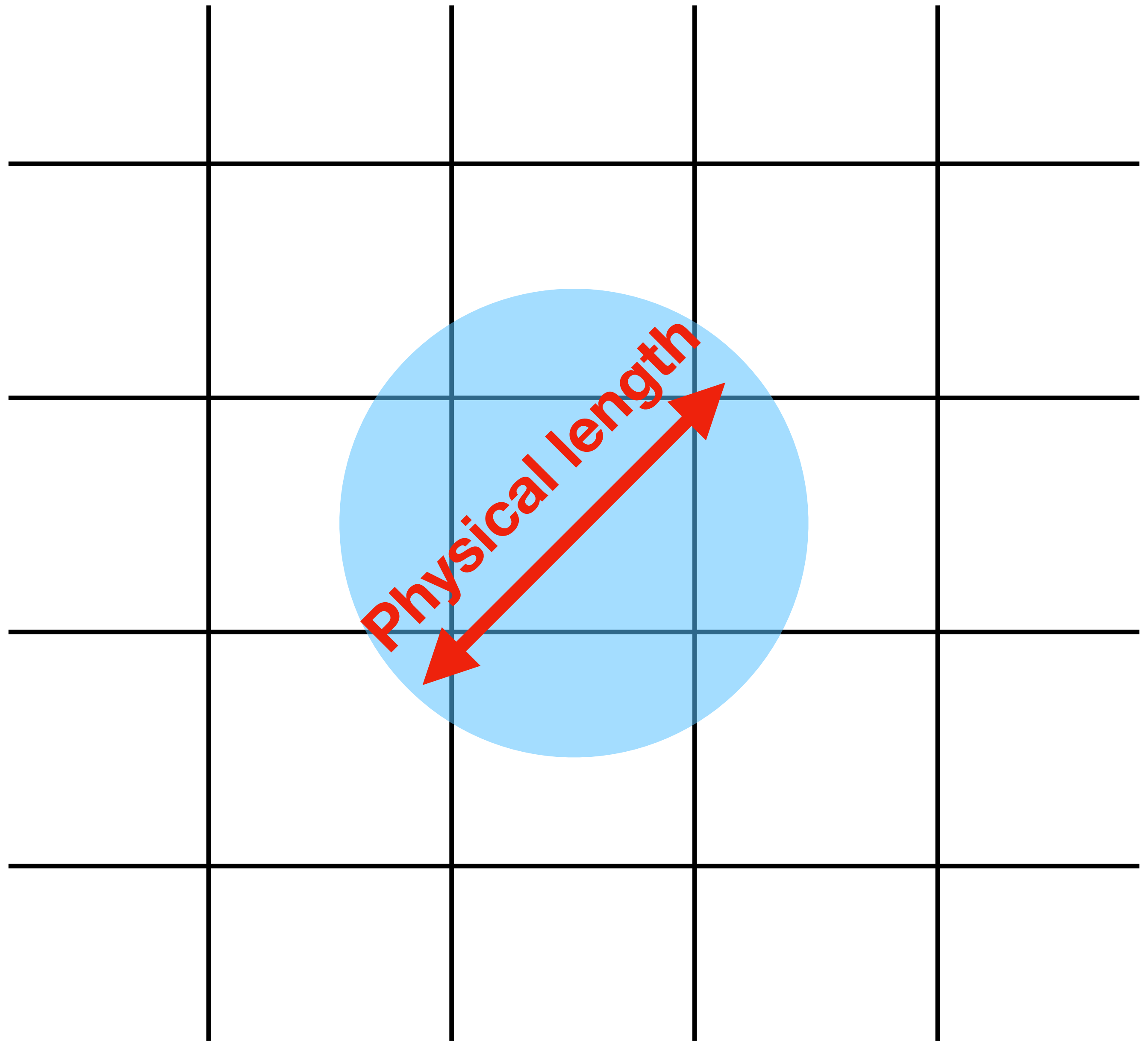
Hybrid/Hamiltonian Monte Carlo (HMC)

Molecular dynamics + Markov Chain Monte Carlo



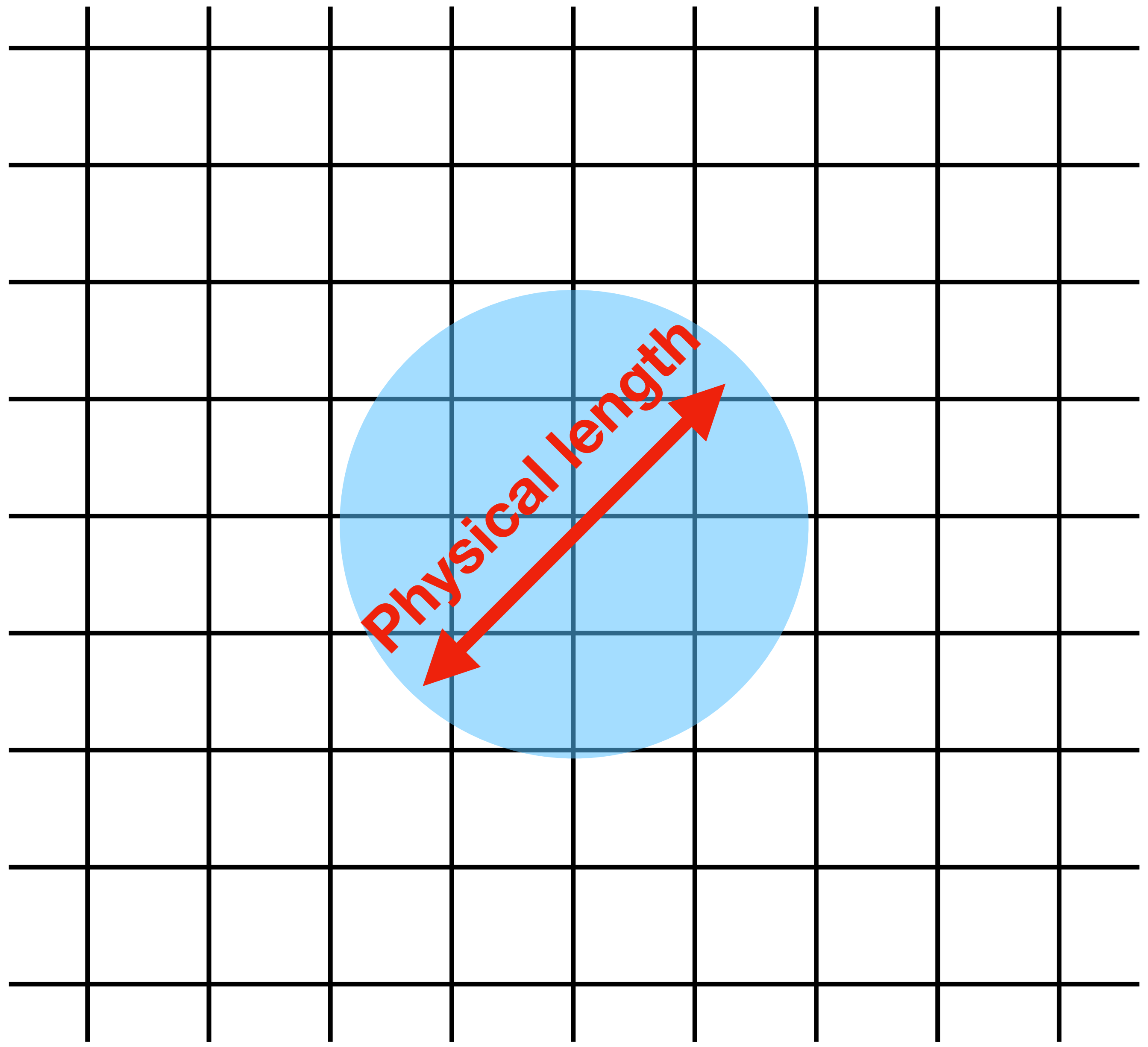
The challenging continuum limit

- Remove discretization effects by taking the continuum limit.



The challenging continuum limit

- Remove discretization effects by taking the continuum limit.

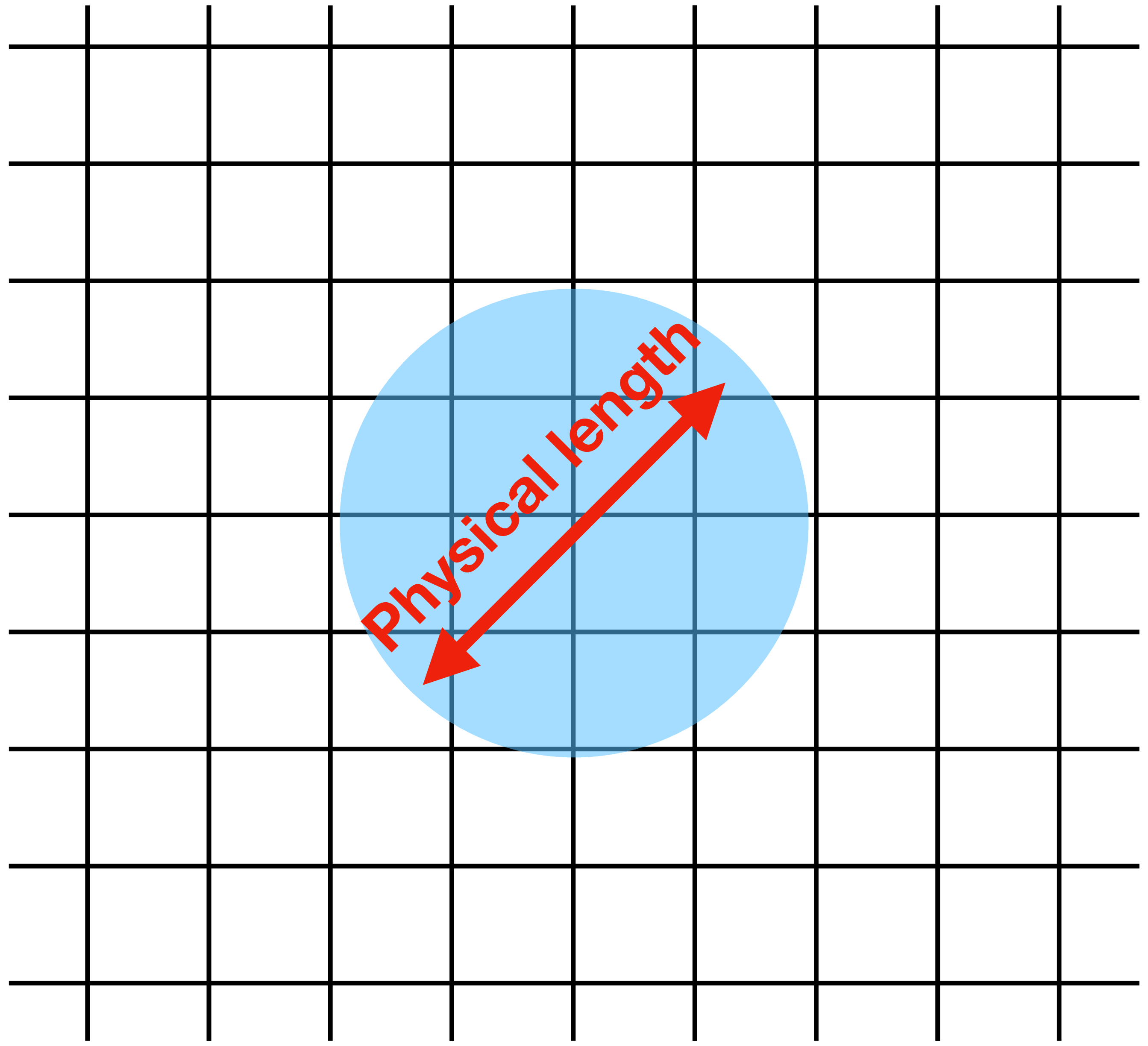


Lattice spacing ⟶ 0

Fernando Romero-López, MIT

The challenging continuum limit

- Remove discretization effects by taking the continuum limit.



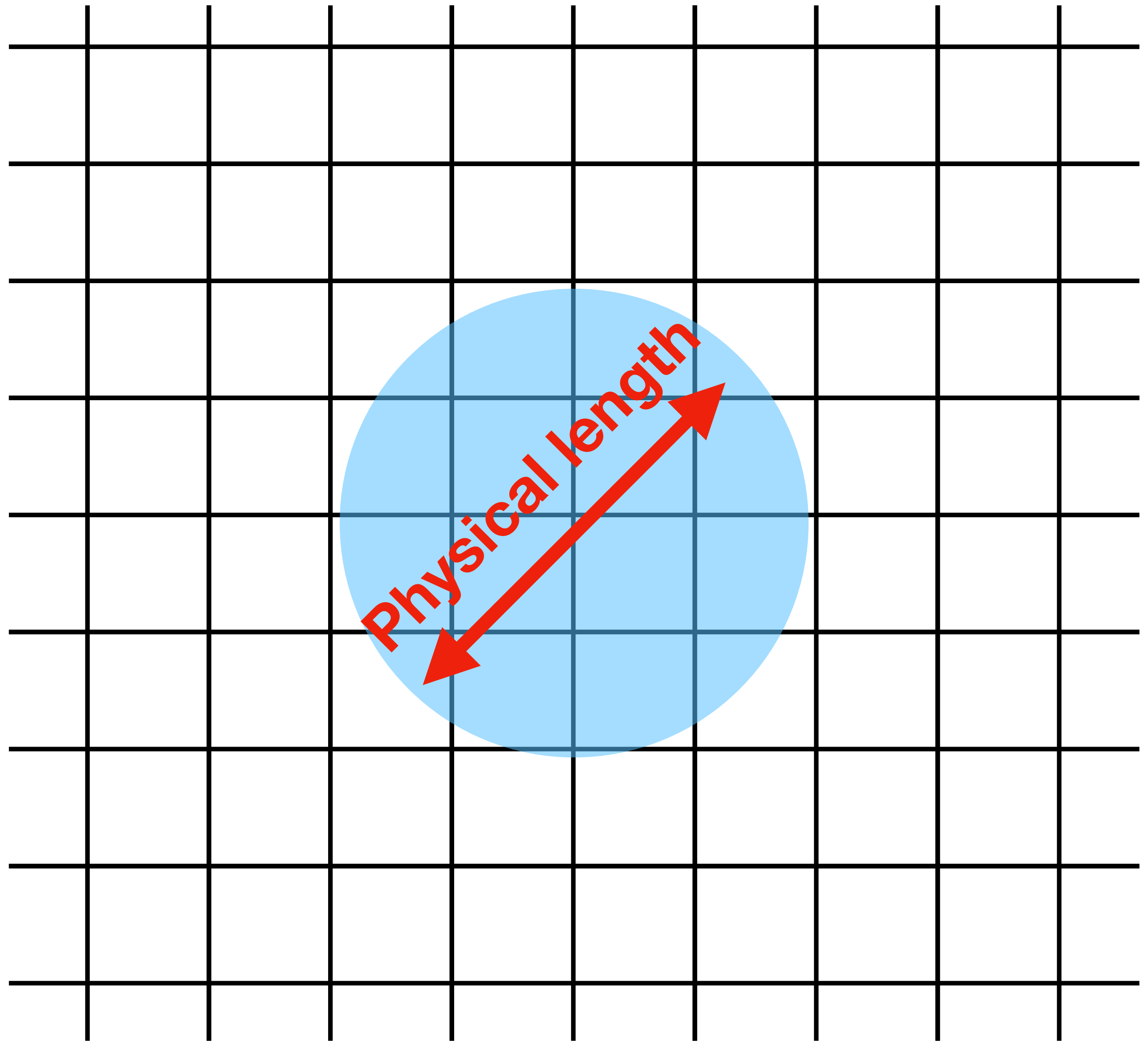
Lattice spacing $\longrightarrow 0$

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

Fernando Romero-López, MIT

The challenging continuum limit

- Remove discretization effects by taking the continuum limit.



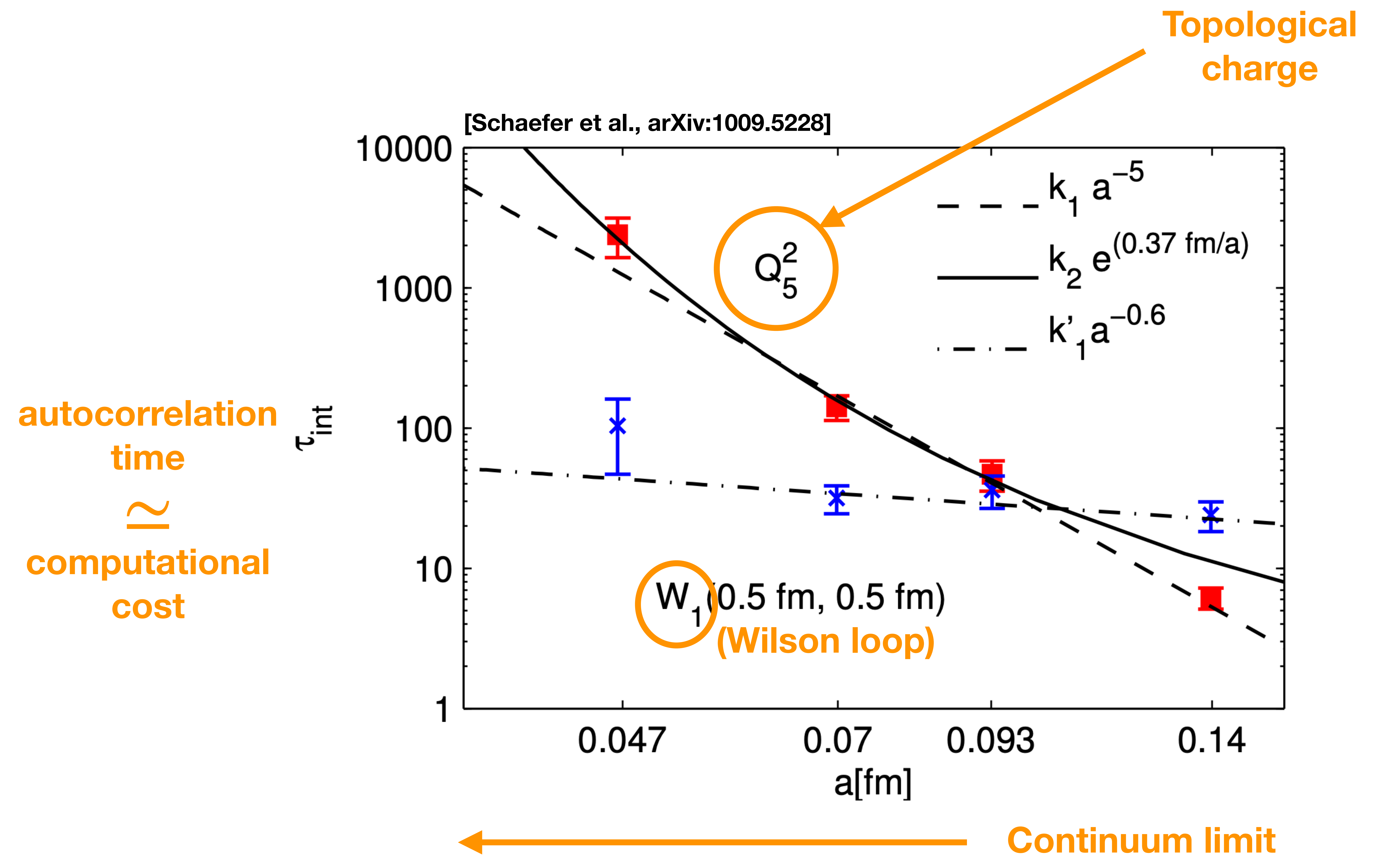
Lattice spacing $\longrightarrow$ 0

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

> **"Critical slowing-down"**
> of generation of uncorrelated samples

Fernando Romero-López, MIT

Critical slowing down

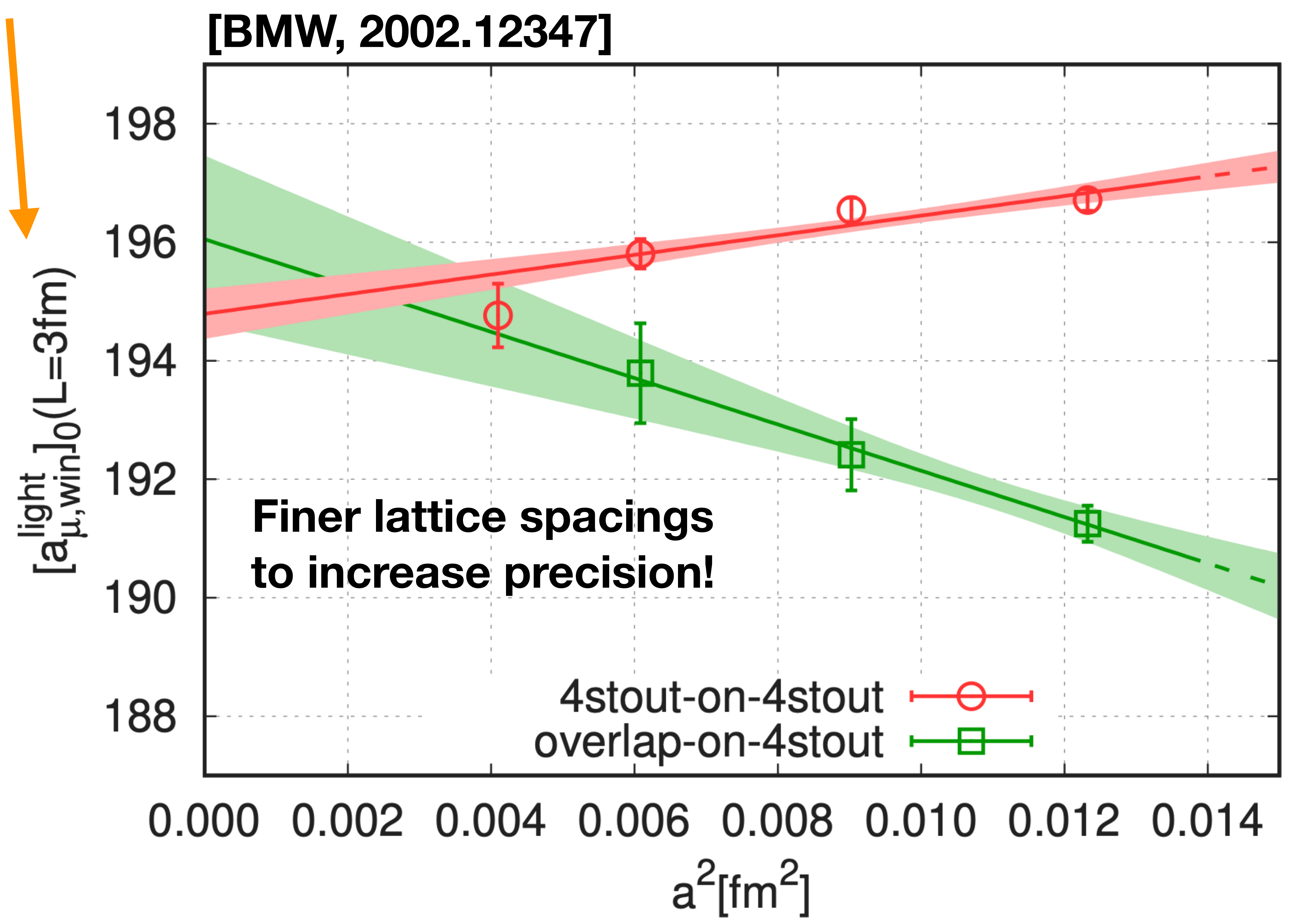


CSD is a well-know obstacle to extend the reach of state-of-the-art lattice QCD calculations

Fernando Romero-López, MIT

Critical slowing down

HVP of muon
magnetic moment



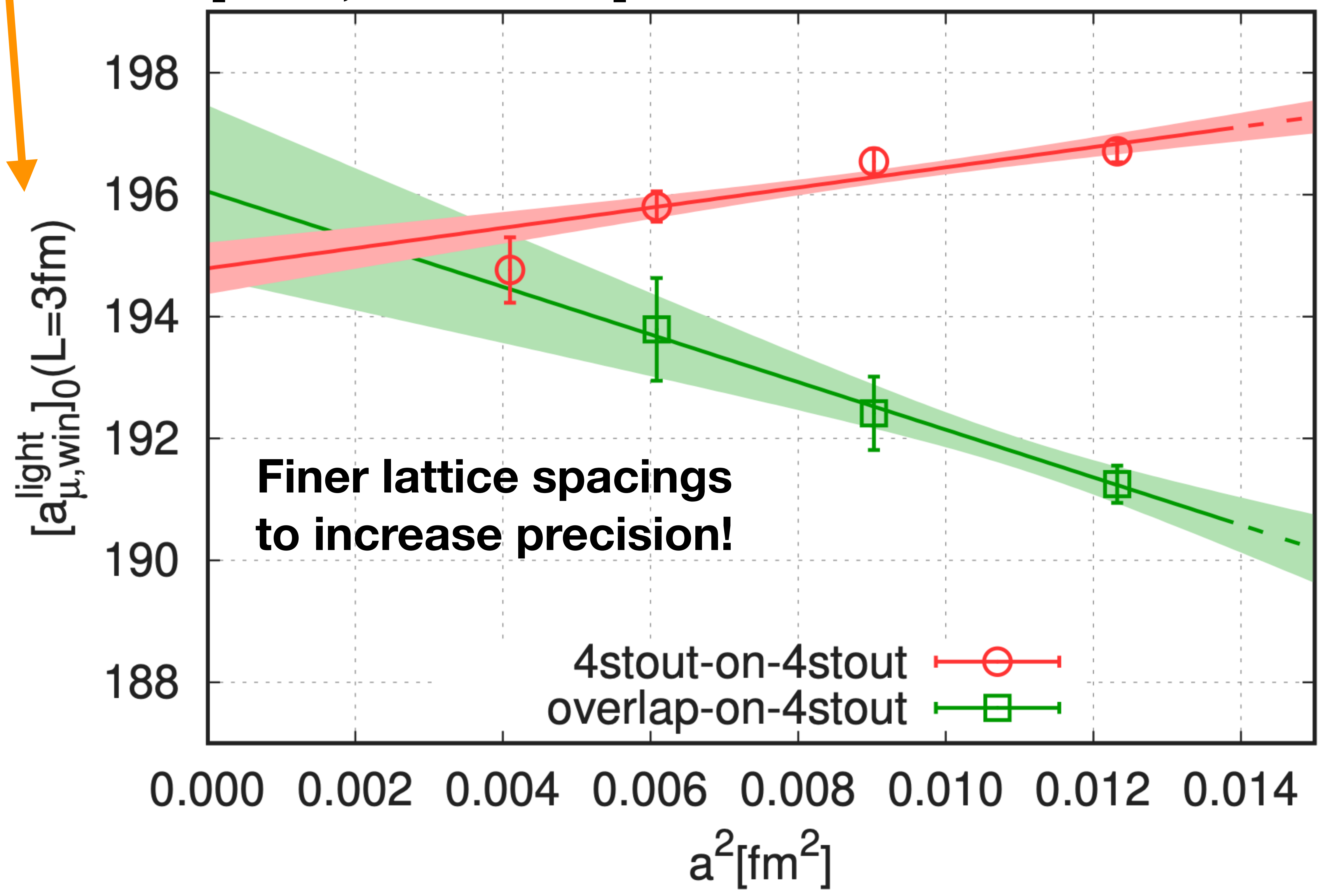
Continuum limit

CSD is a well-know obstacle to extend the reach of state-of-the-art lattice QCD calculations

Critical slowing down

HVP of muon magnetic moment

[BMW, 2002.12347]



Continuum limit

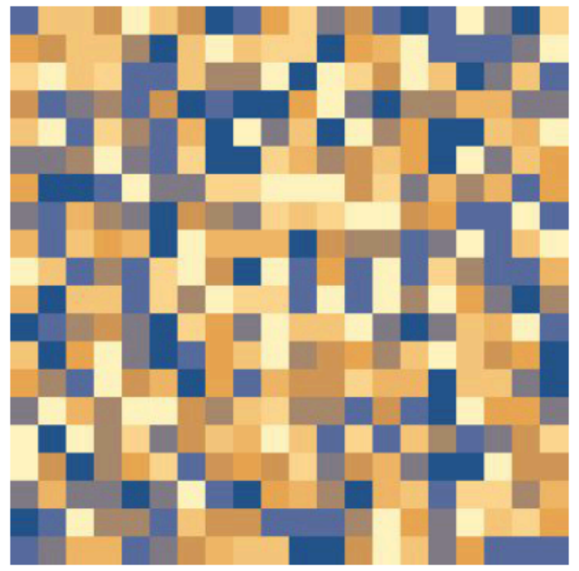
CSD is a well-know obstacle to extend the reach of state-of-the-art lattice QCD calculations

Can ML help?

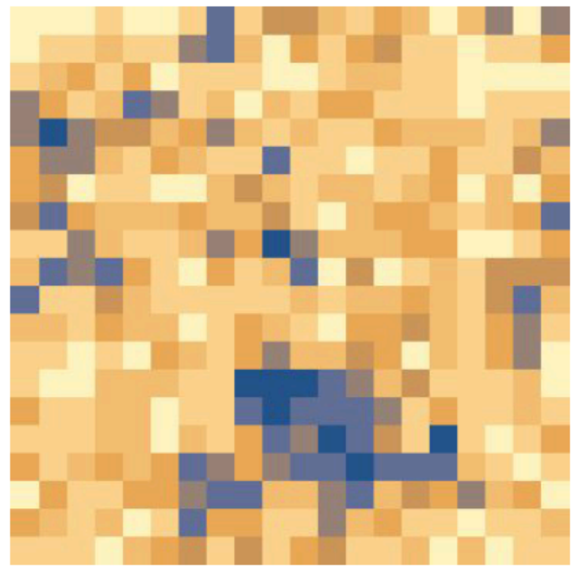
Flow-based sampling for Lattice Field Theories

Test case: scalar theory

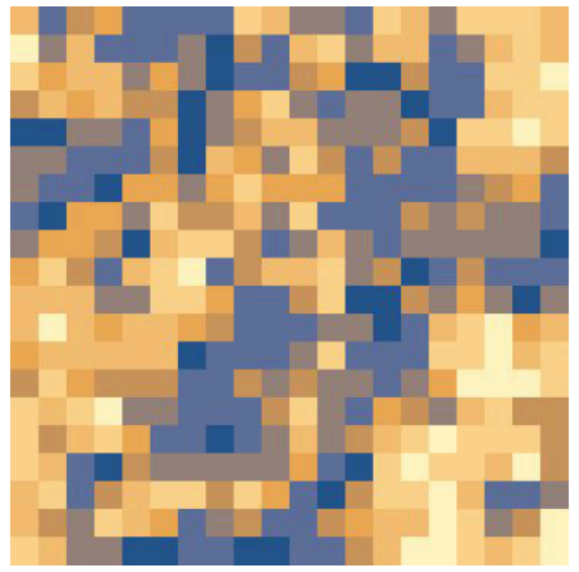
- A real scalar field per lattice site in a 2D lattice. $\phi(x) \in (-\infty, +\infty)$



unlikely
(log prob = -6107)



likely
(log prob = 22)



likely
(log prob = 5)

Field configurations with probability:

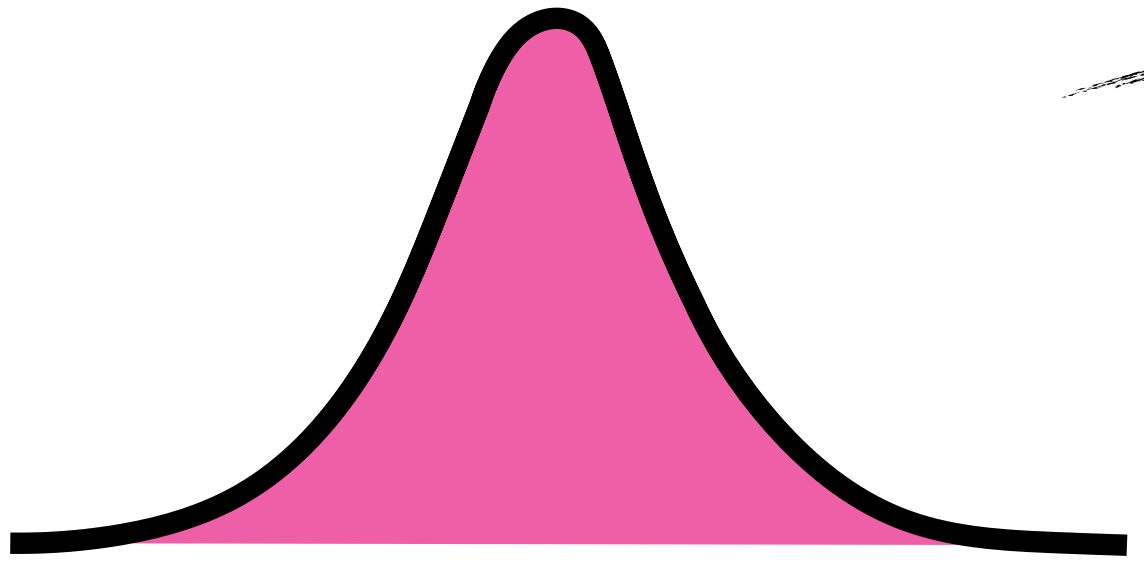
$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

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

Generative flow models

[Rezende, Mohamed, 1505.05770]

$r(z)$

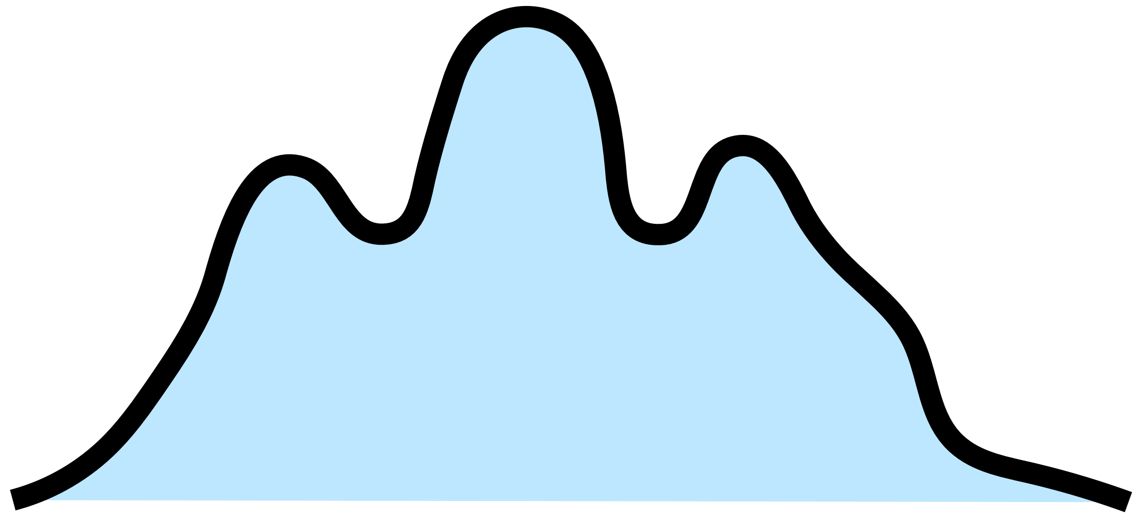


easily sampled
distribution
(prior)

$f(z)$

tractable Jacobian
&
invertible

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



approximates
target distribution
(model)

“Trainable change of variables”

Model probability

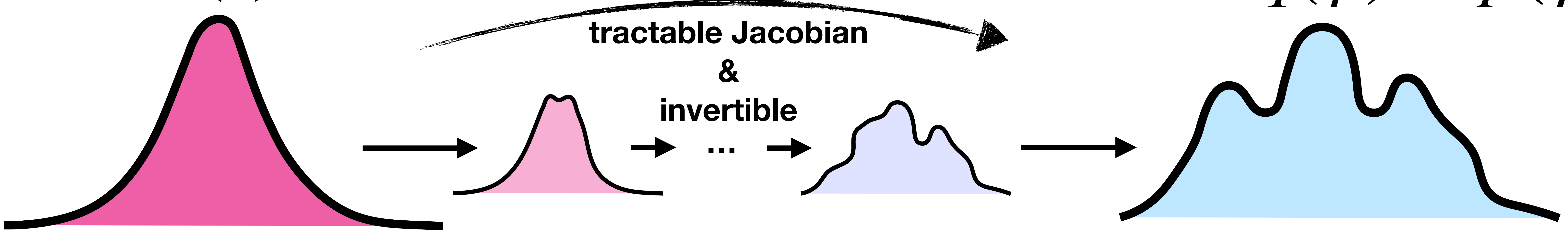
$$q(\phi) = r(z) \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$$

Generative flow models

[Rezende, Mohamed, 1505.05770]

$r(z)$ $f(z)$ $q(\phi) \simeq p(\phi)$

tractable Jacobian
&
invertible



easily sampled
distribution
(prior)

approximates
target distribution
(model)

parametrized by neural networks
(trainable and expressive)

composed of
many simple layers

“Trainable change of variables”

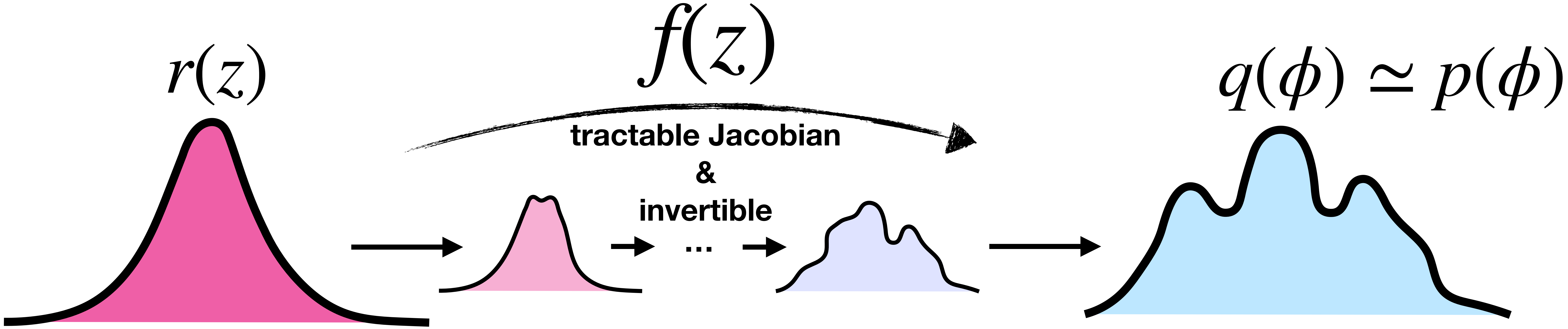
Model probability

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

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

❗ Trained models are not perfect.

☑ But exact sampling can be recovered via Markov Chain

"Trainable change of variables"

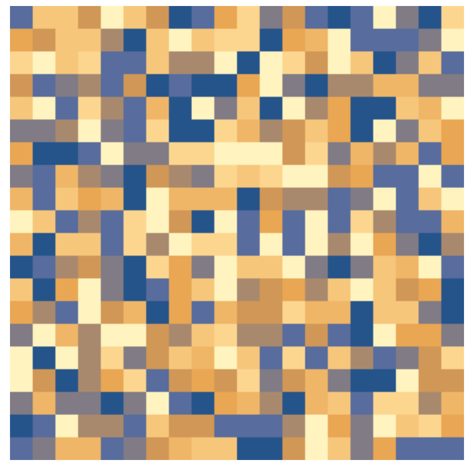
Model probability

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

Fernando Romero-López, MIT

Implementing flows for scalar theory

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



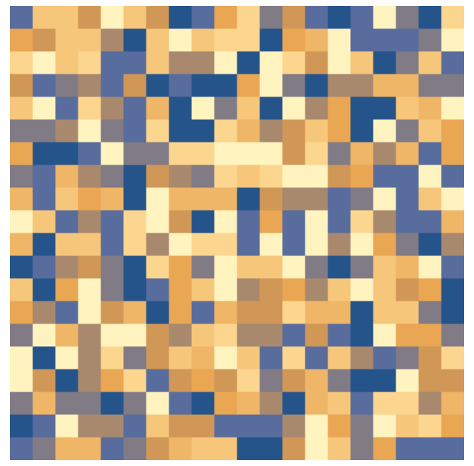
→**Efficient flow models need tractable Jacobians.**

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

Fernando Romero-López, MIT

Implementing flows for scalar theory

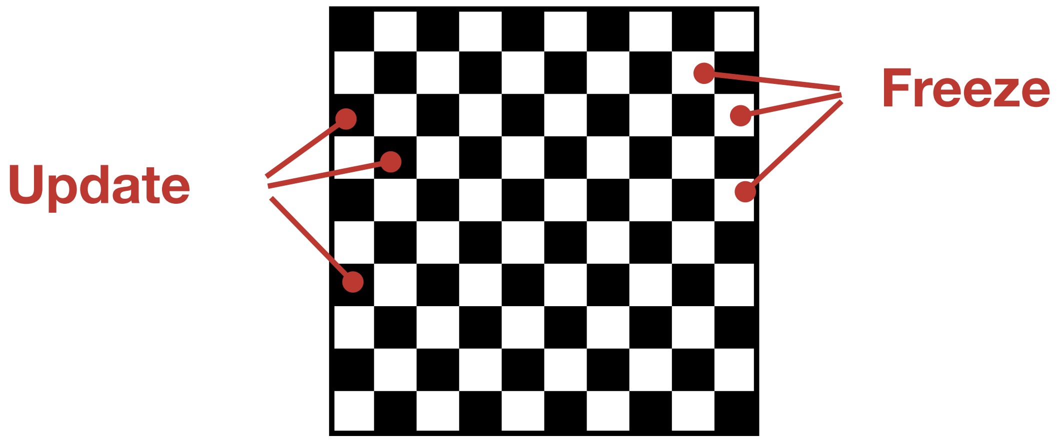
$\phi(x) \in (-\infty, +\infty)$



→**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.

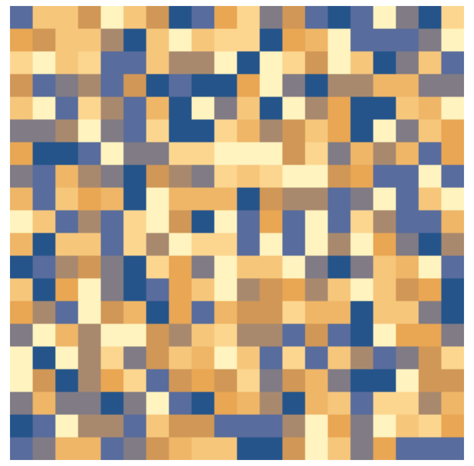


$$\phi' = \begin{cases} \phi'_{\text{frozen}} = \phi_{\text{frozen}} \\ \phi'_{\text{active}} = h(\phi_{\text{frozen}}) \times \phi_{\text{active}} + t \end{cases}$$

conditioning

Implementing flows for scalar theory

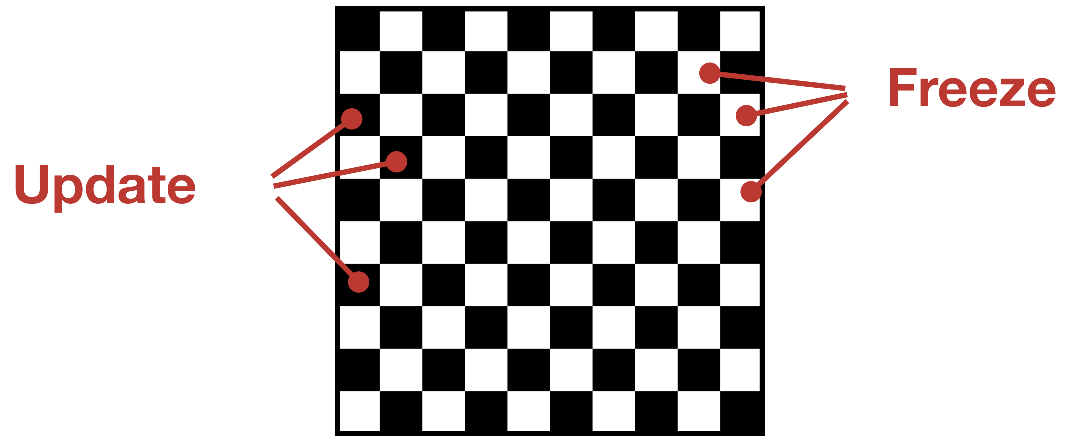
$\phi(x) \in (-\infty, +\infty)$



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

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

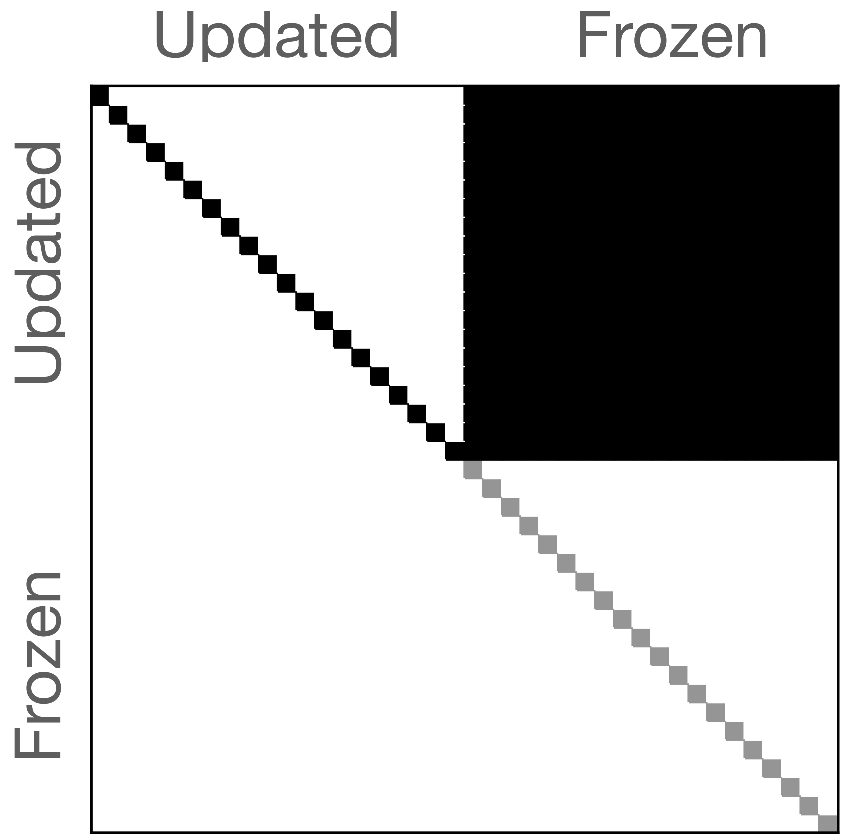
"Masking pattern" m defines subsets.



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

conditioning

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

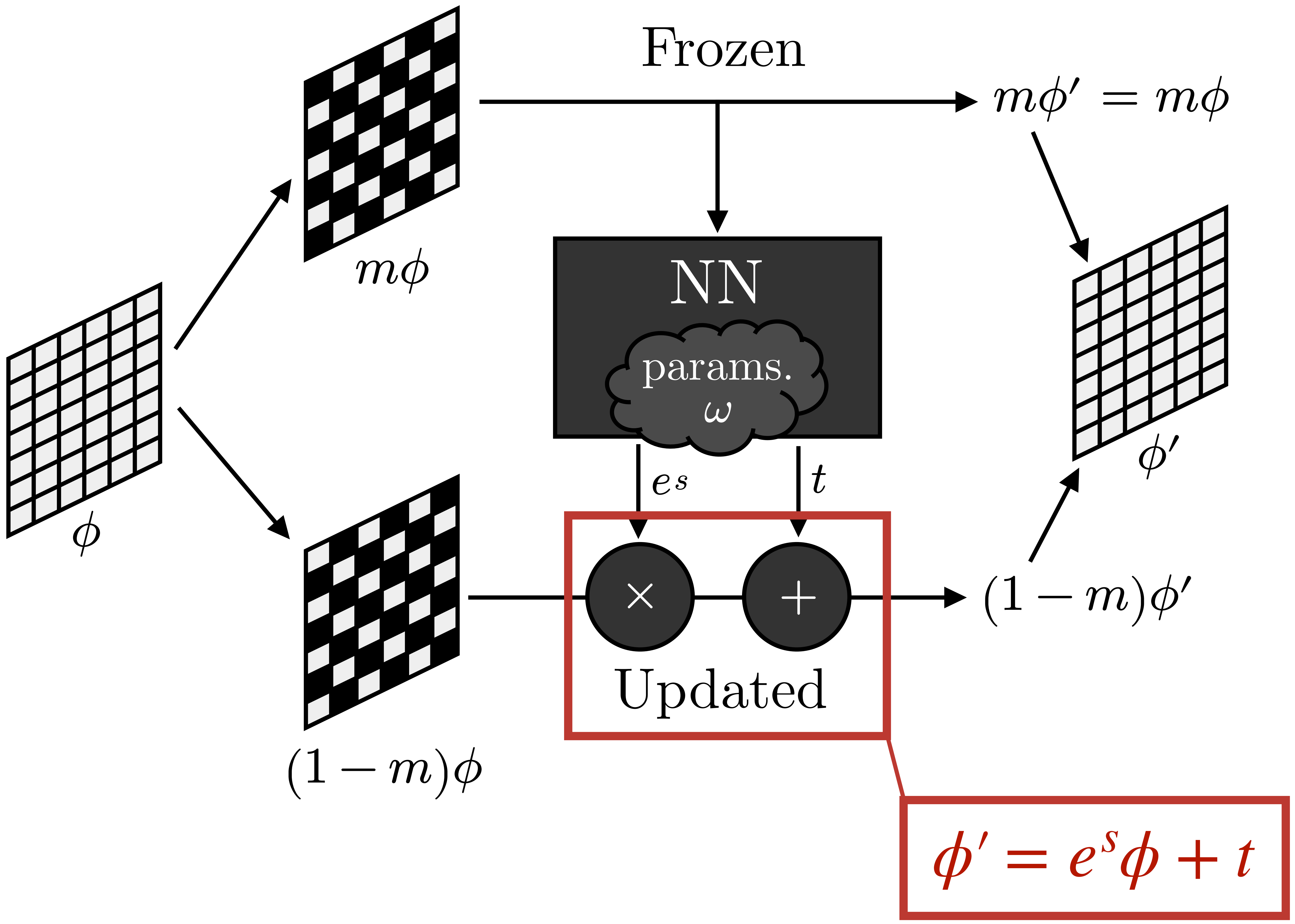


Fernando Romero-López, MIT

Implementing flows for scalar theory

Non-volume preserving coupling layer:

[Dinh, Sohl-Dickstein, Bengio 1605.08803]



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

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

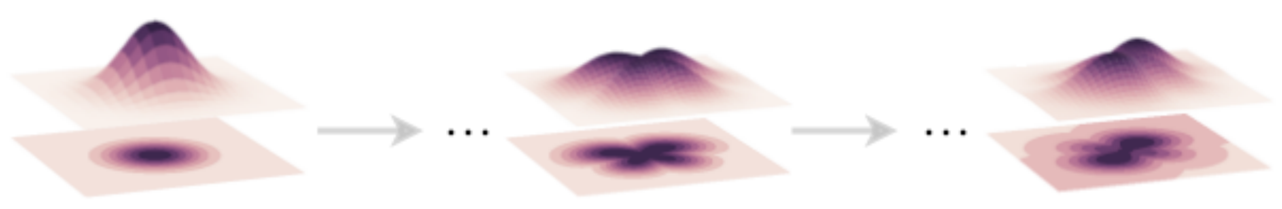
- Self-training:

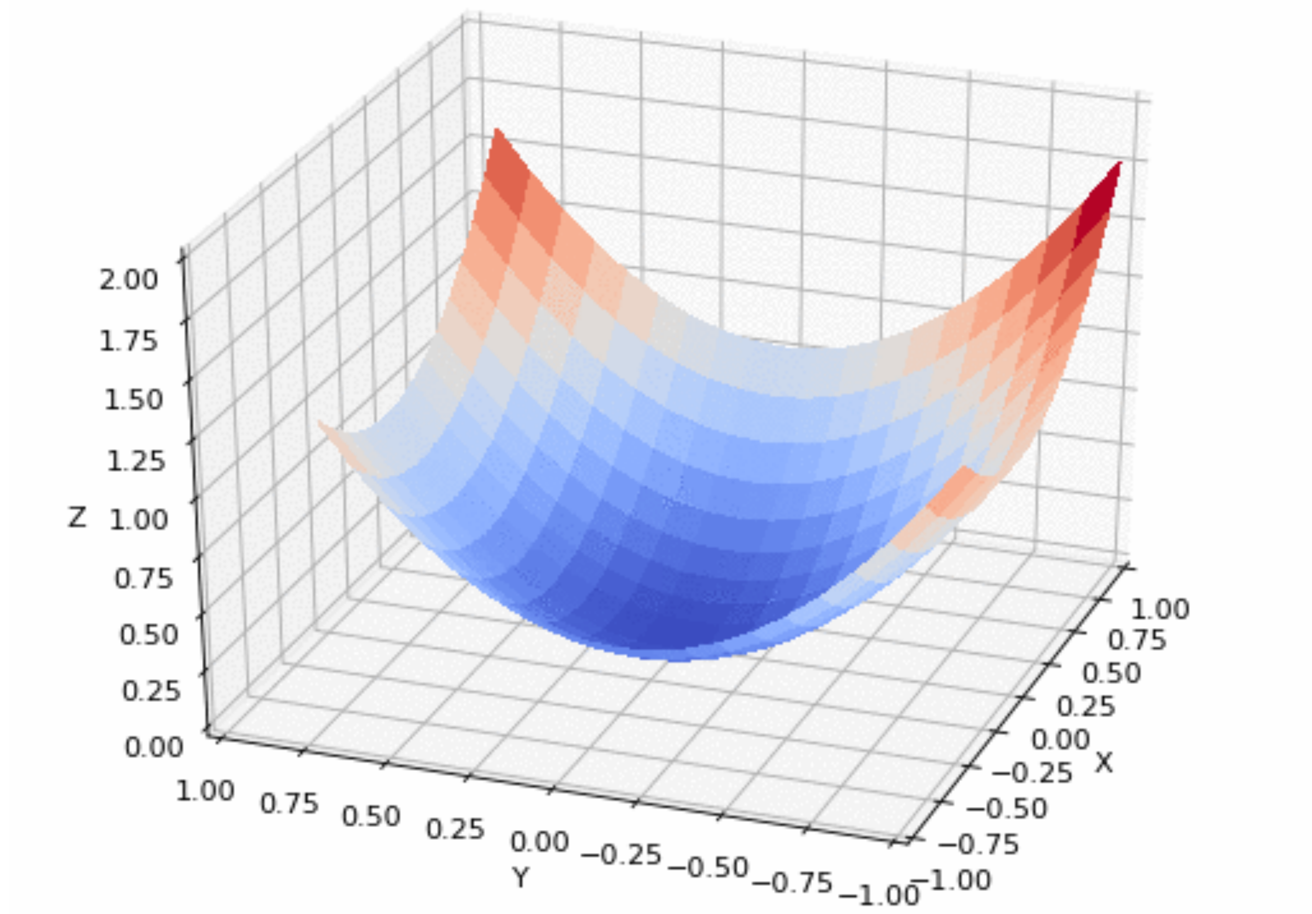
    1. Draw samples from the model to measure sample mean of $\log[q(\phi)/p(\phi)]$

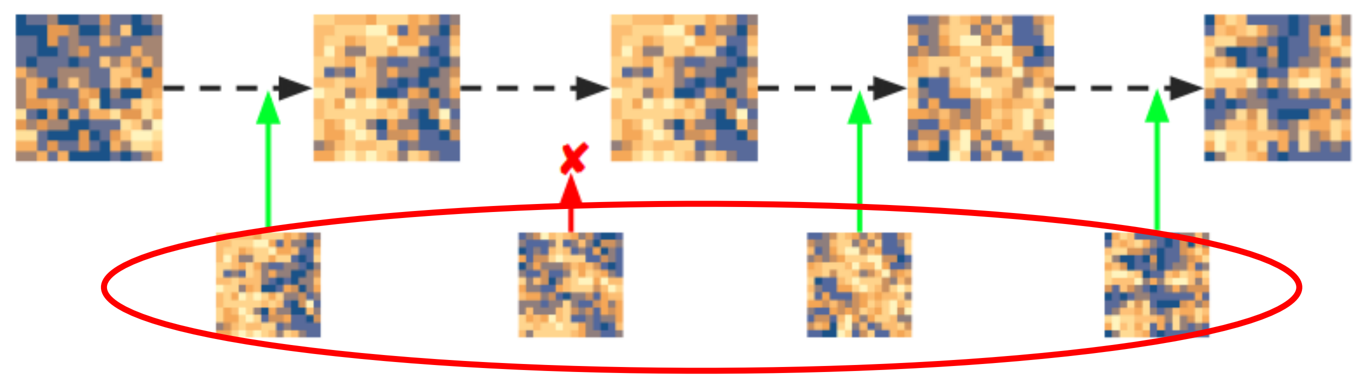
    2. Gradient-based methods to optimize model parameters (e.g. Adam optimizer)

[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?

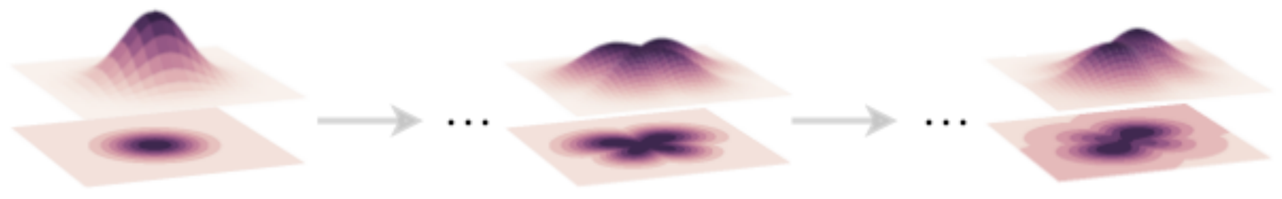
Save trained model

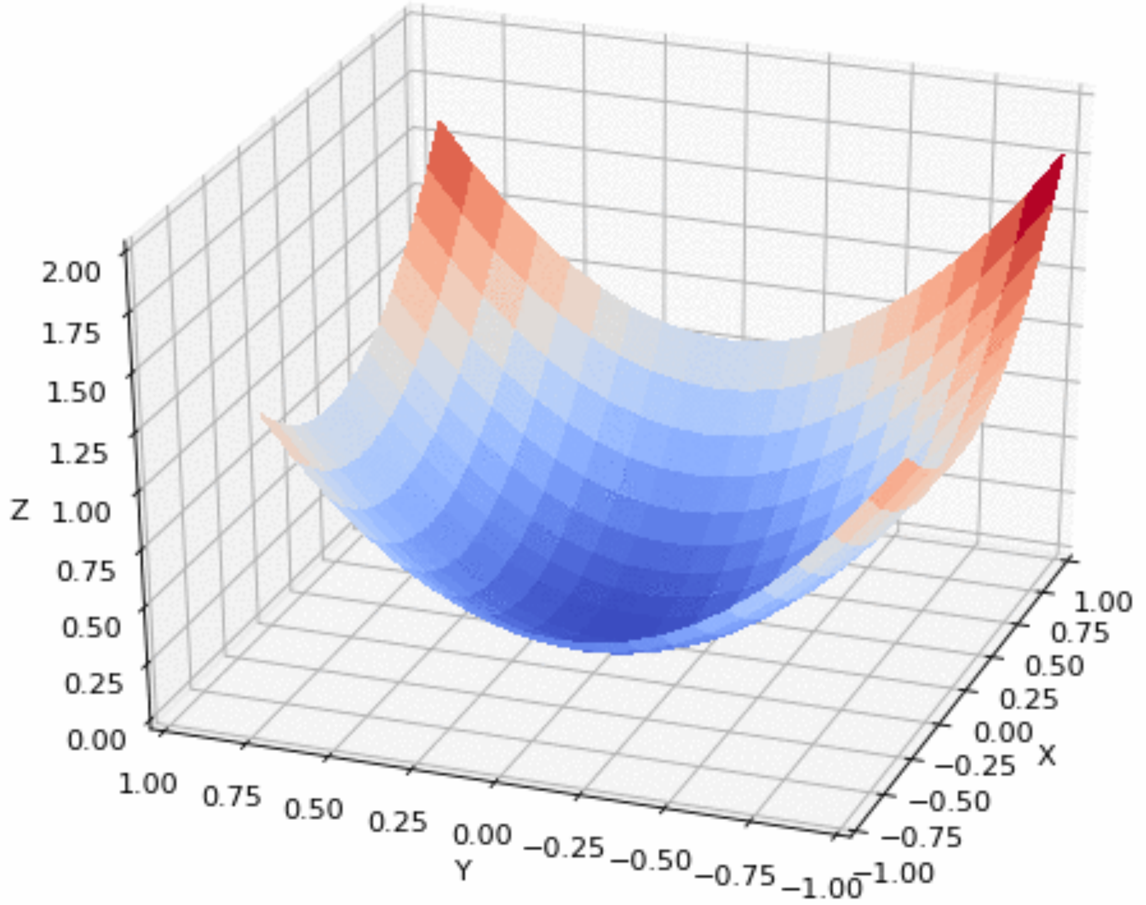
Markov chain using samples from model

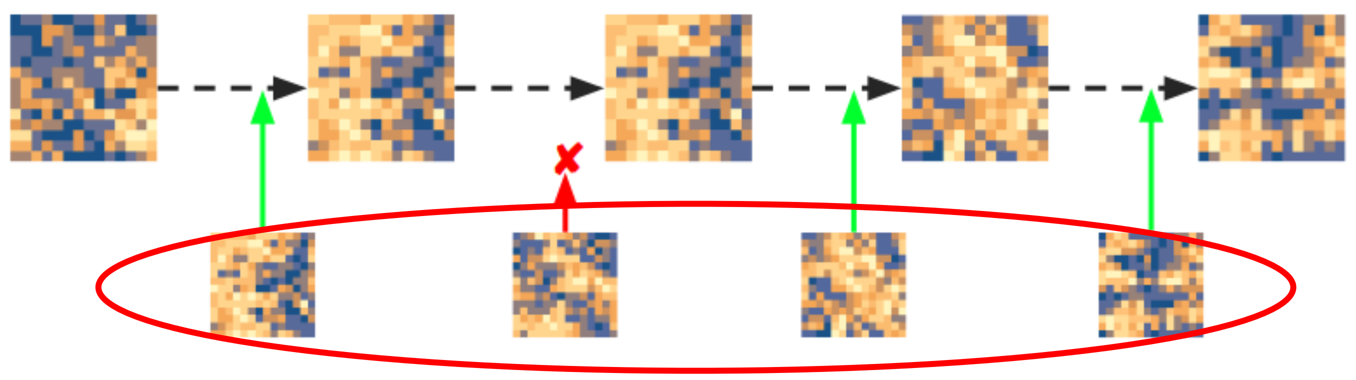
[credits: G. Kanwar]

Fernando Romero-López, MIT

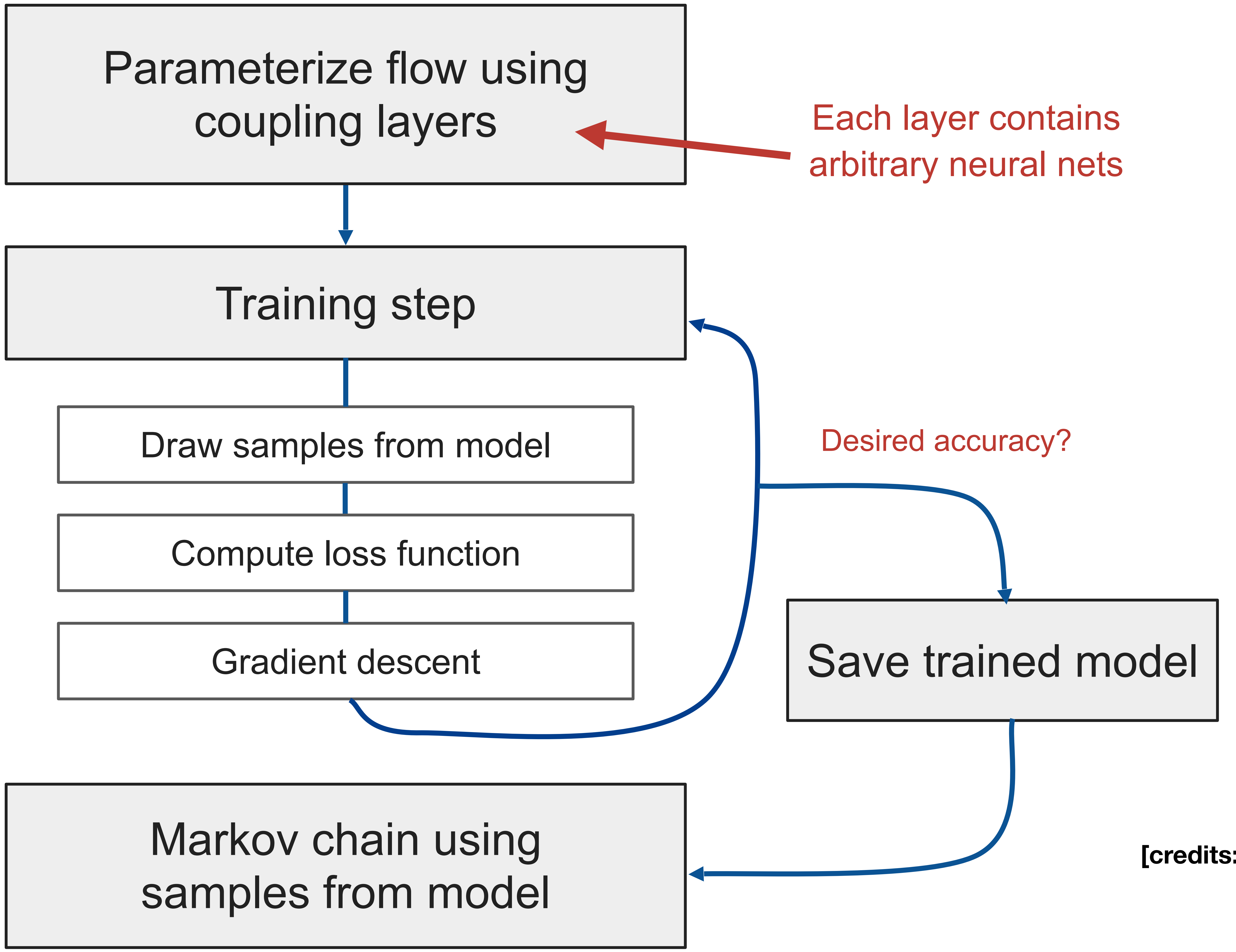
Lattice QFT via flow models







generating samples is
"embarrassingly parallel"

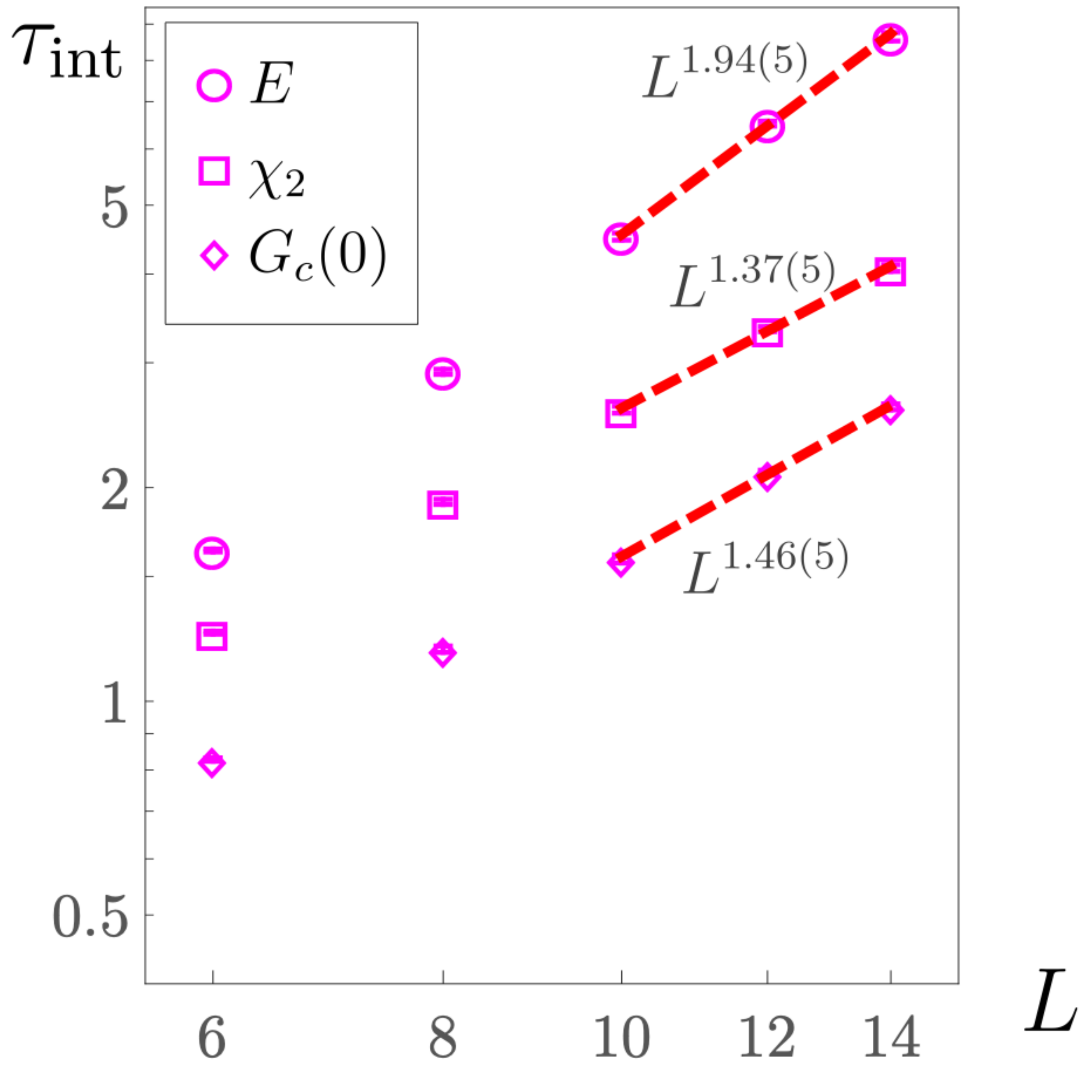


[credits: G. Kanwar]

Fernando Romero-López, MIT

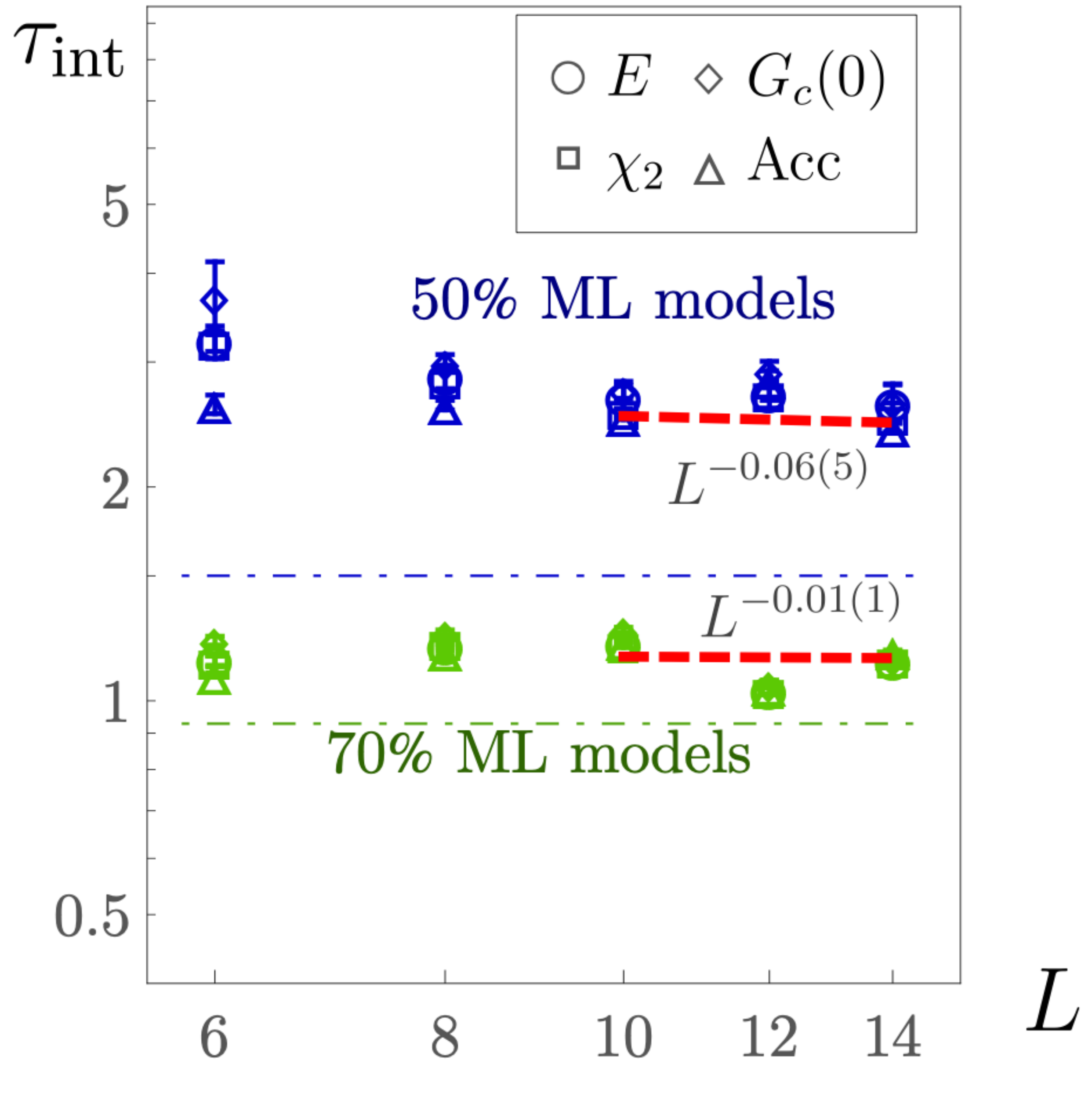
Scalar theory

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



(b) Local Metropolis ensembles

Conventional approaches slow down



(c) Flow-based MCMC ensembles

No slow-down!

[Albergo, Kanwar, Shanahan 1904.12072]

Flow-based sampling for gauge theories

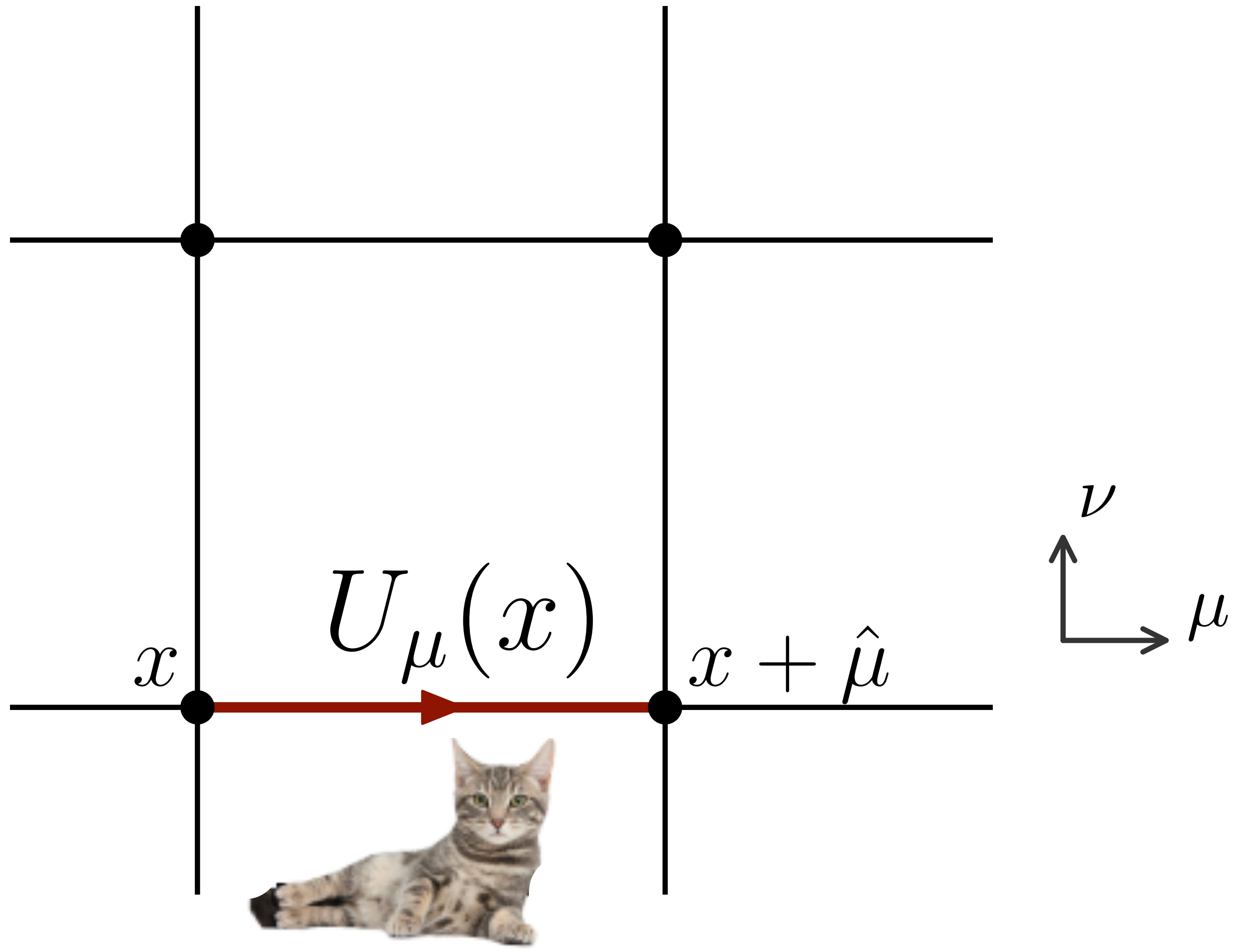
Lattice U(1) gauge symmetry

- 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

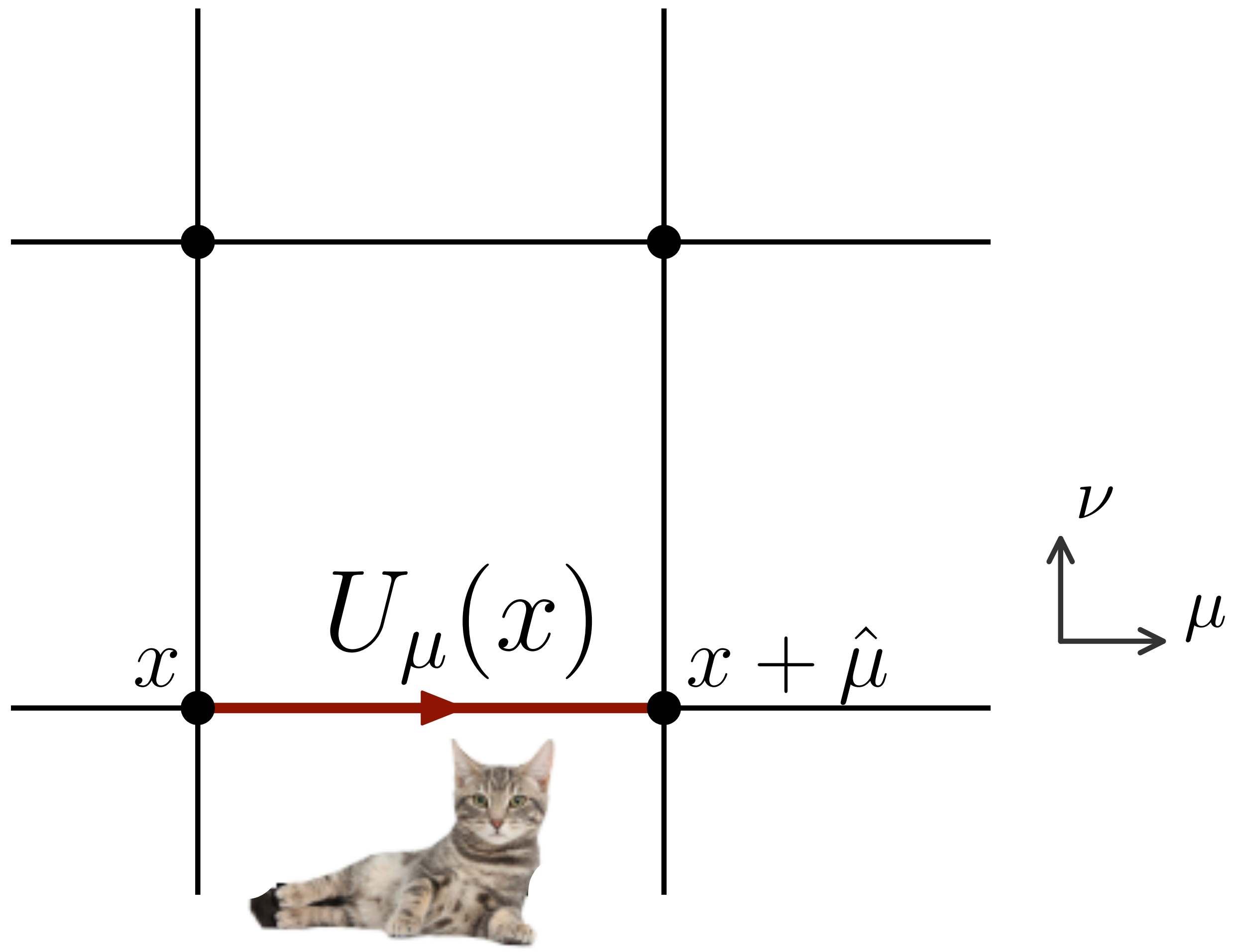
Lattice U(1) gauge symmetry

- Gauge variables are the gauge links

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

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

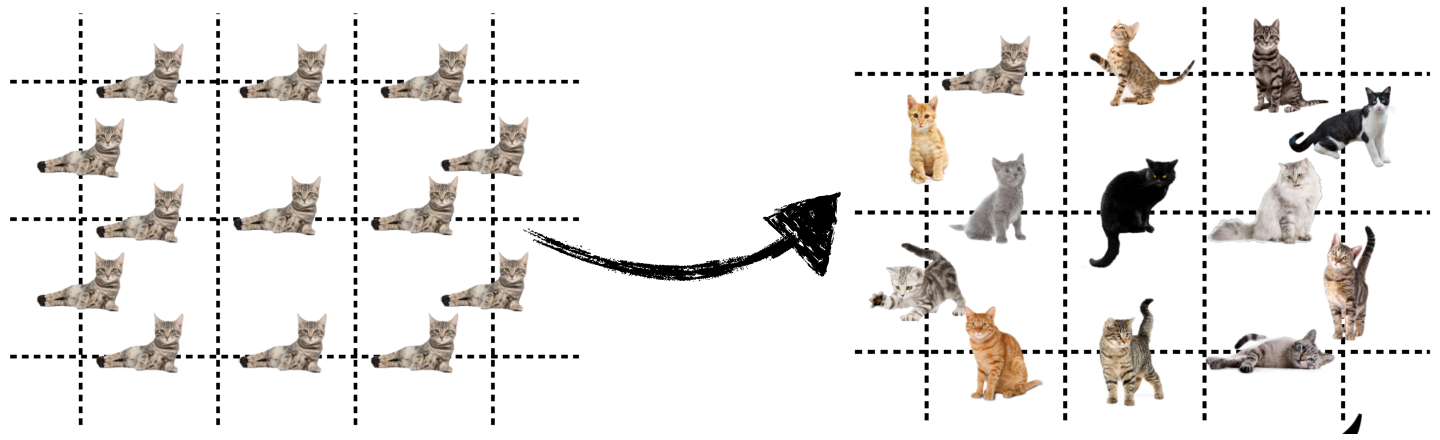
$$agA_\mu(x) \in [0, 2\pi)$$



- 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

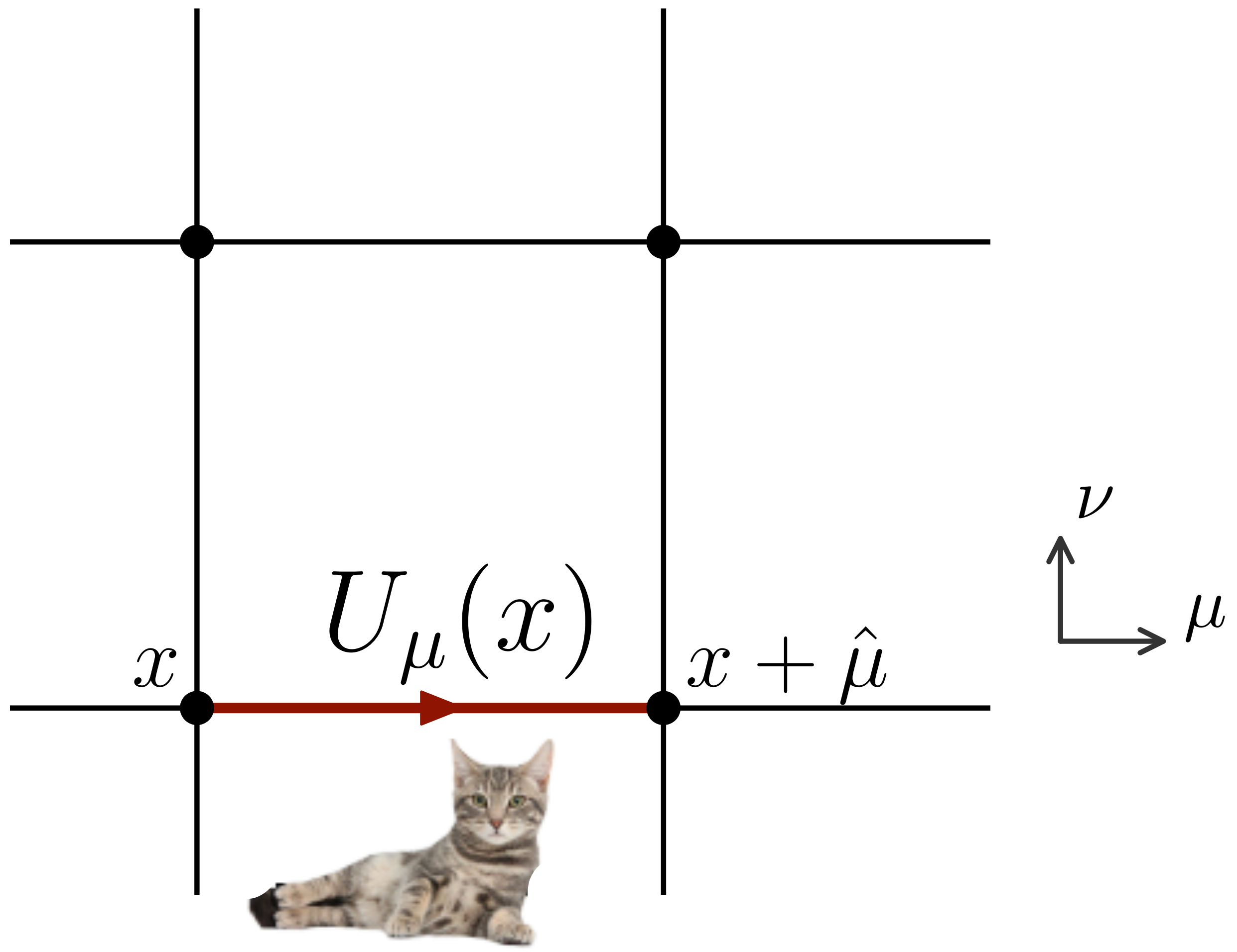
Lattice U(1) gauge symmetry

- Gauge variables are the gauge links

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

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

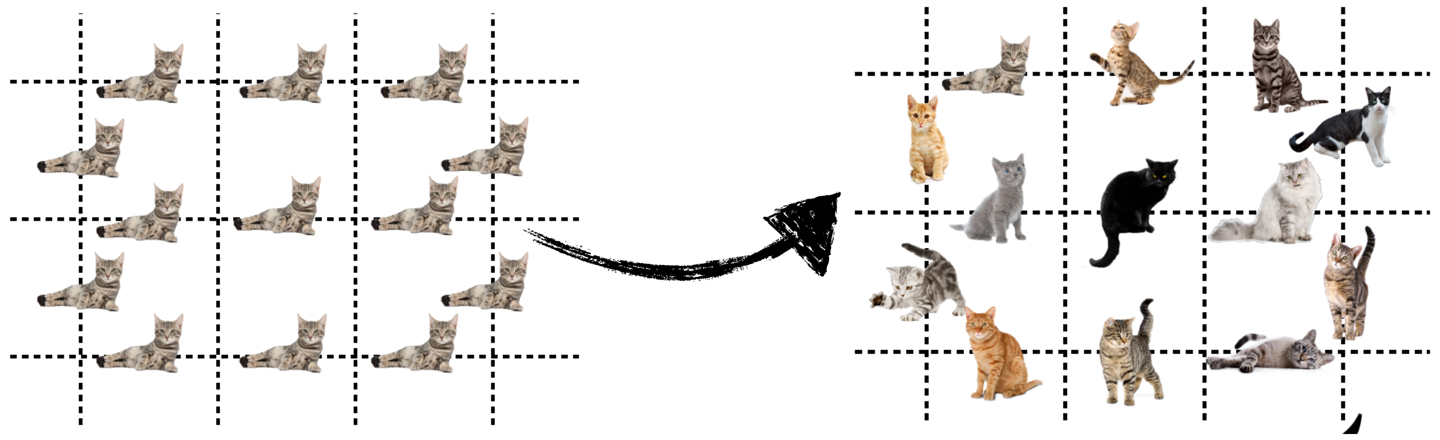
$$agA_\mu(x) \in [0, 2\pi)$$



- A gauge transformation is:

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

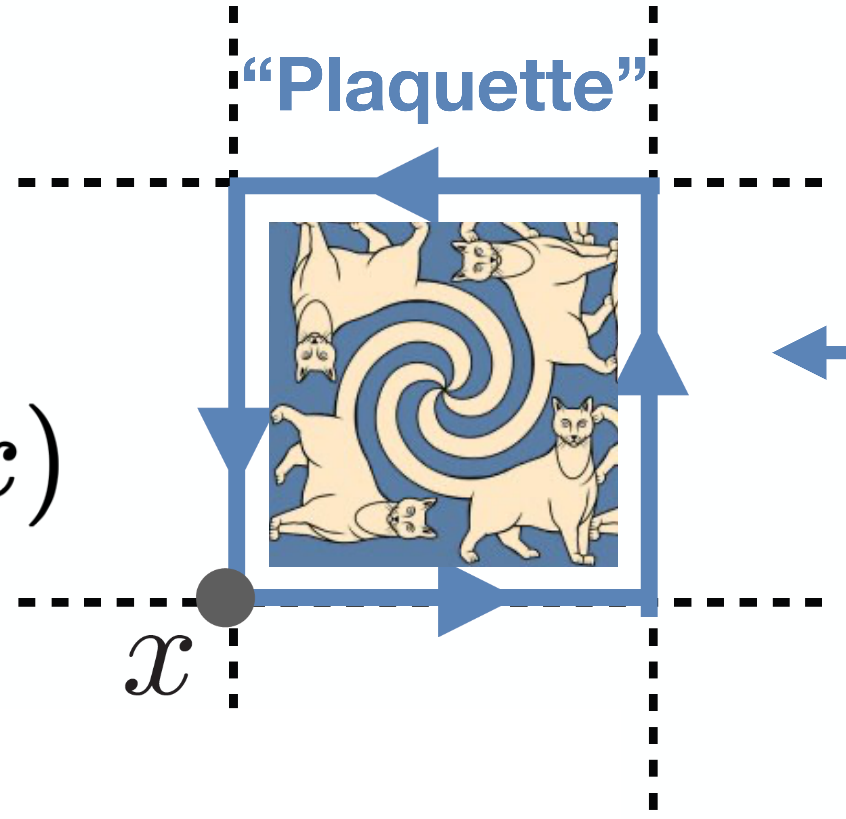
$$\text{for all } \Omega(x) \in \mathrm{U}(1)$$



- Pure gauge lattice action:

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

inverse gauge coupling

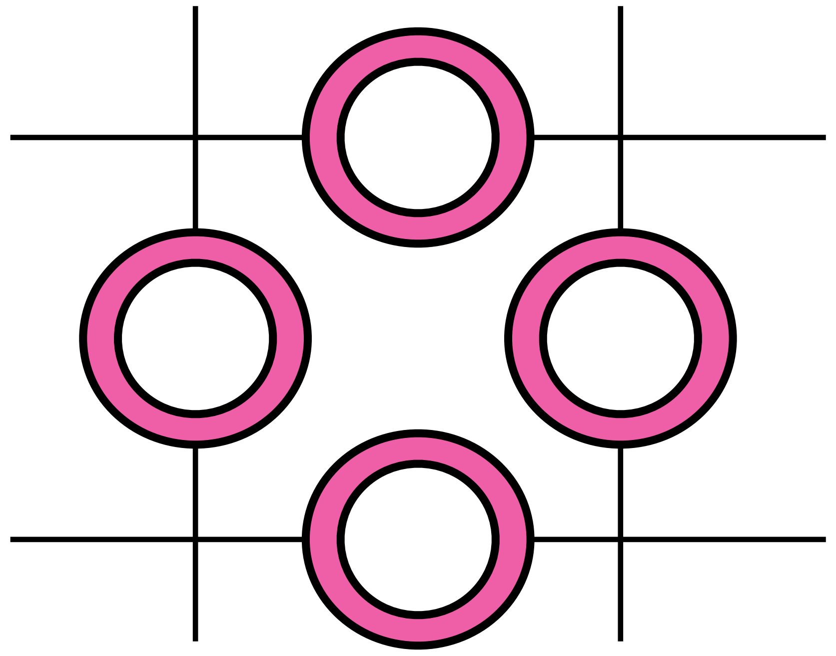


Gauge invariant
$P_{\mu\nu} \in U(1)$

Fernando Romero-López, MIT

Flows for gauge theories

$r(z)$

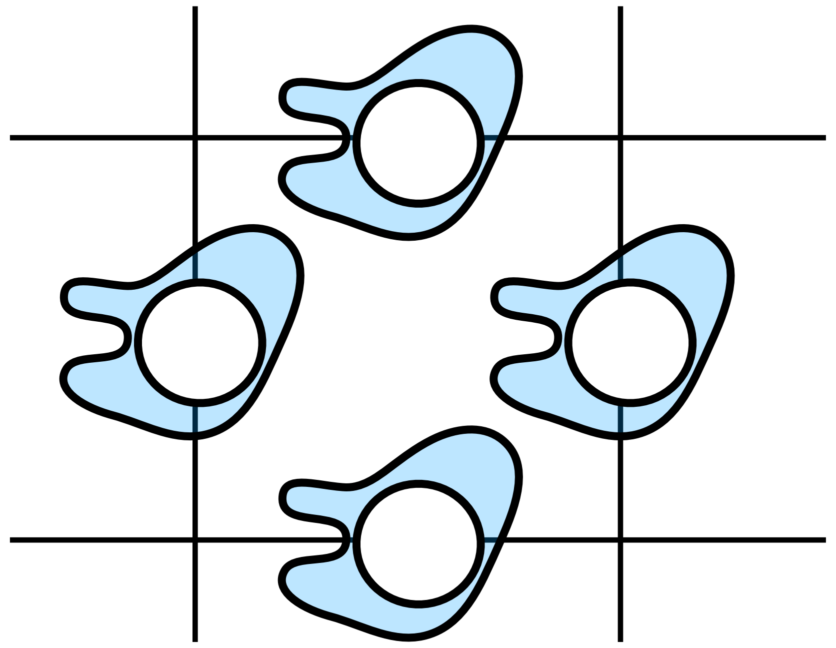


easily sampled distribution (prior)

$f(z)$

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

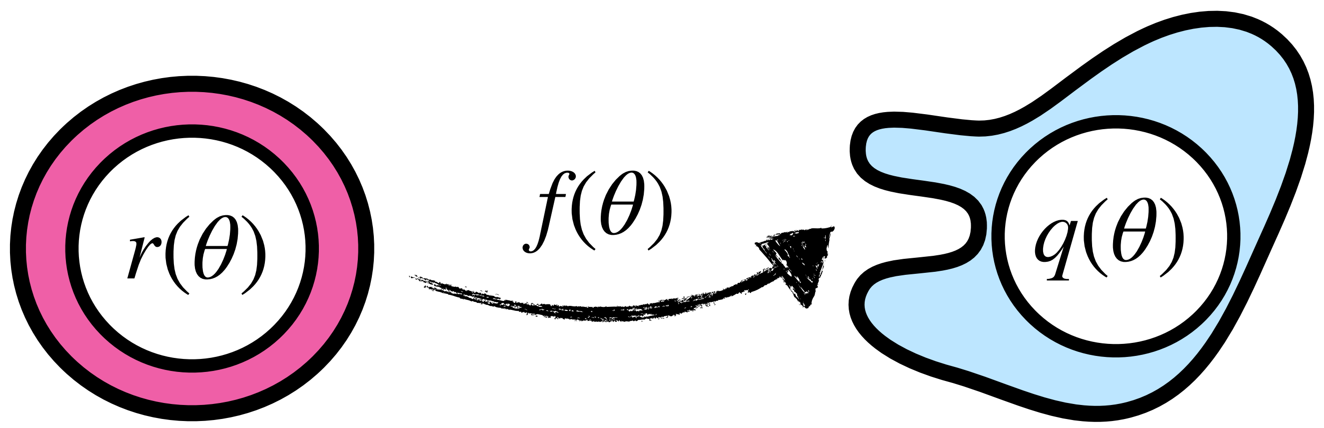
Model probability

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

Flows on compact variables

- **Flows on compact connected manifolds:**

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



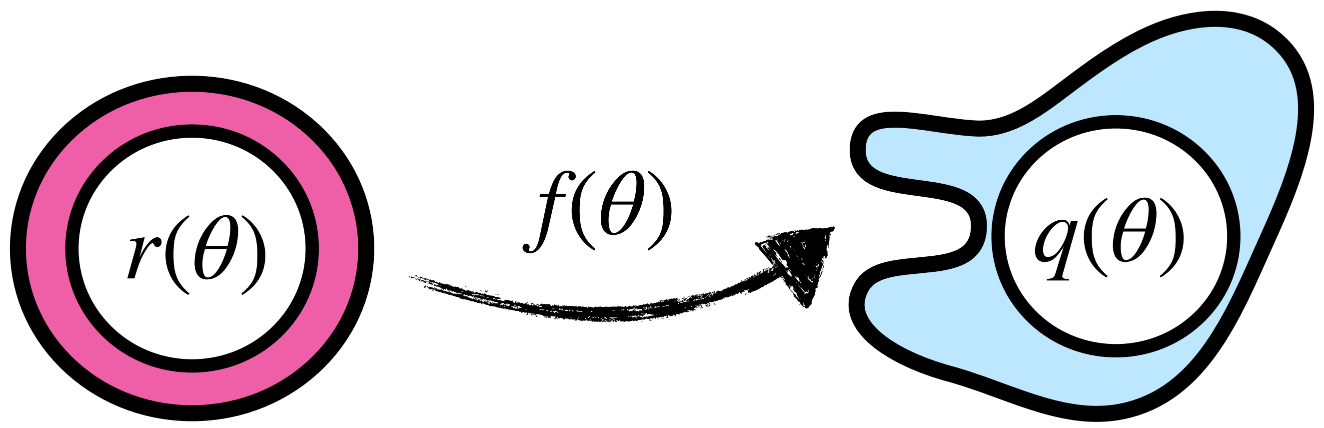
[Rezende, Papamakarios, Racanière , Albergo, Kanwar, Shanahan, Cranmer 2002.02428]

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:

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

monotonic, invertible

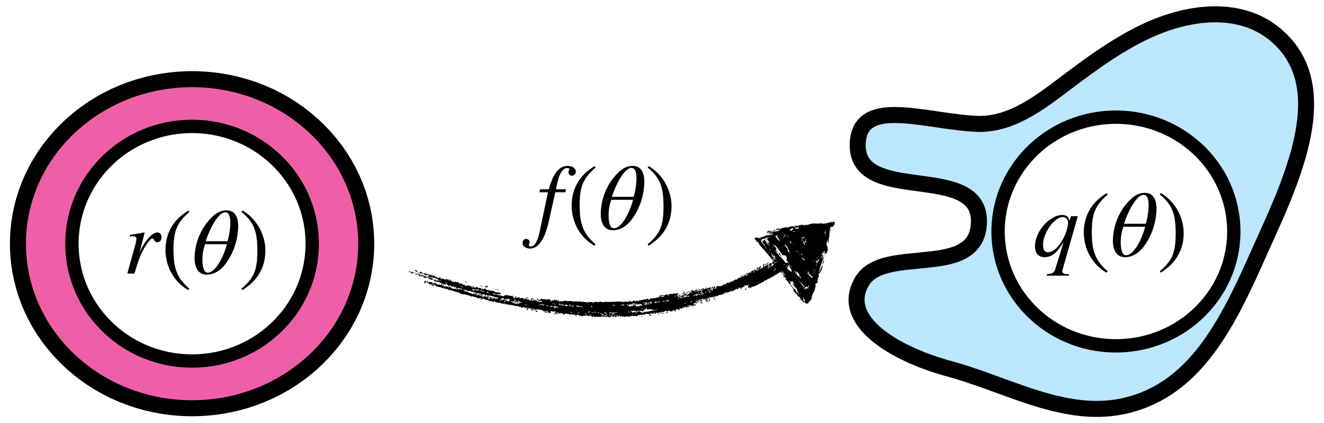
Fernando Romero-López, MIT

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]

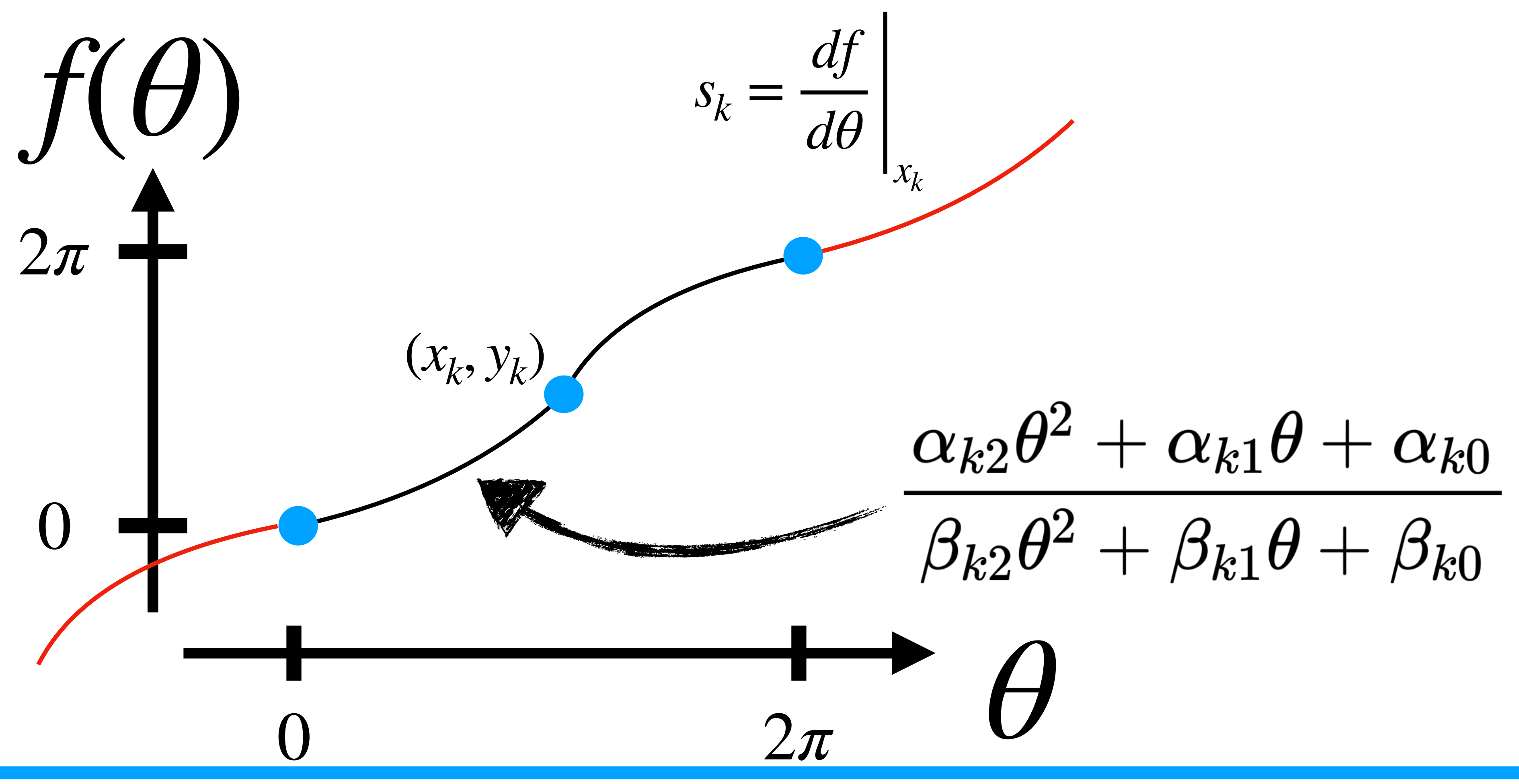


Diffeomorphism if:

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

monotonic, invertible

Circular splines:



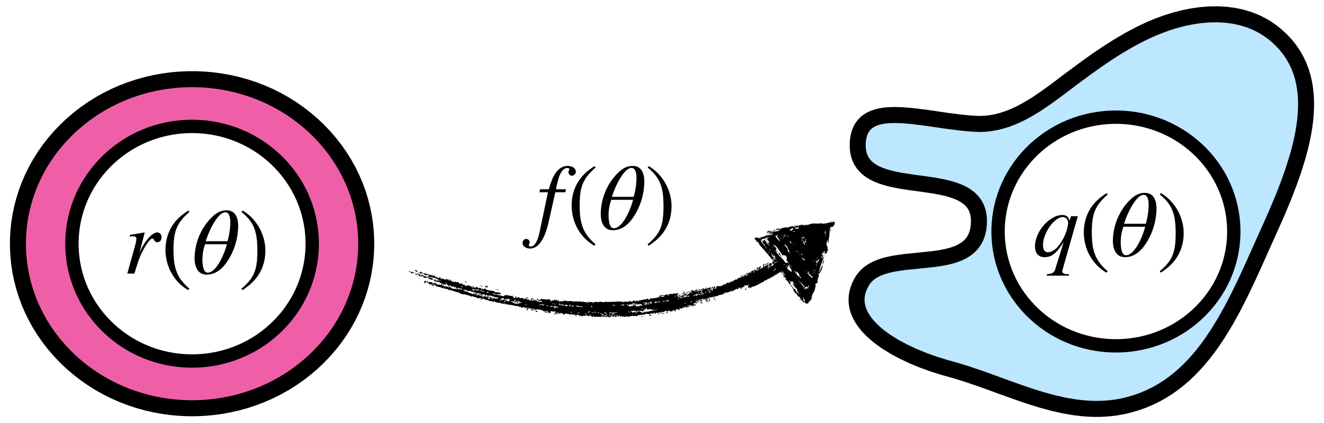
Fernando Romero-López, MIT

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]



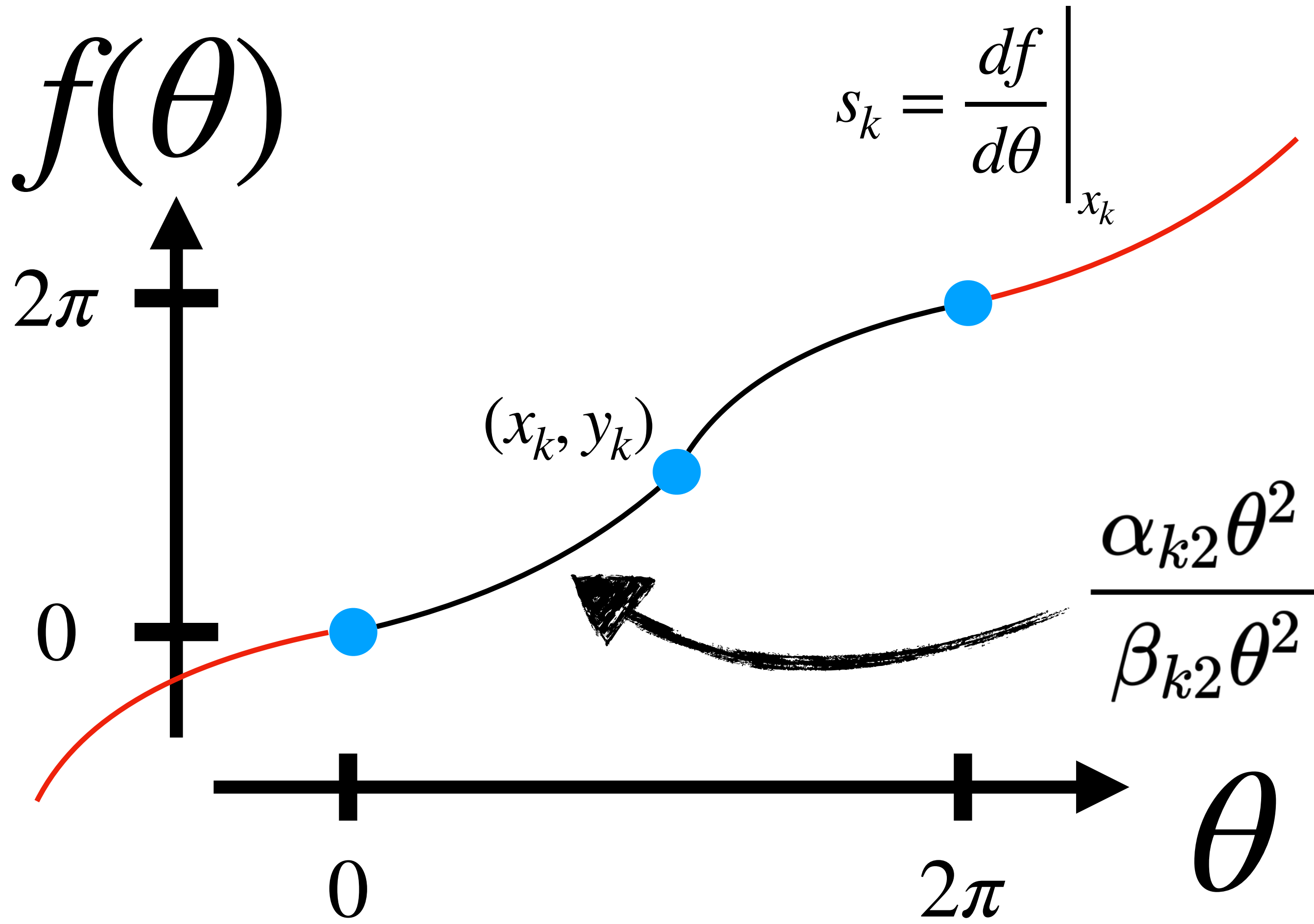
Diffeomorphism if:

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

monotonic, invertible

Circular splines:

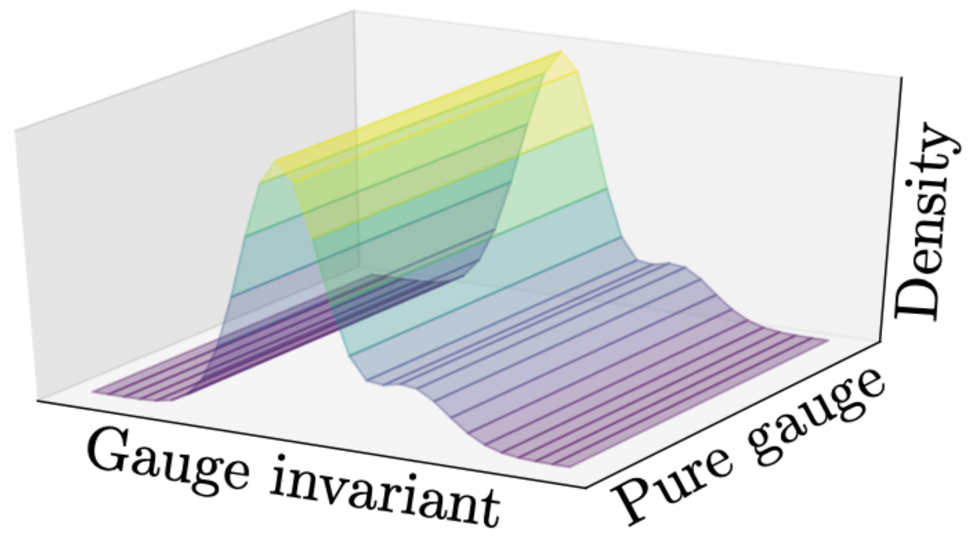
- Trainable positions and slopes
- Expressivity with more knots



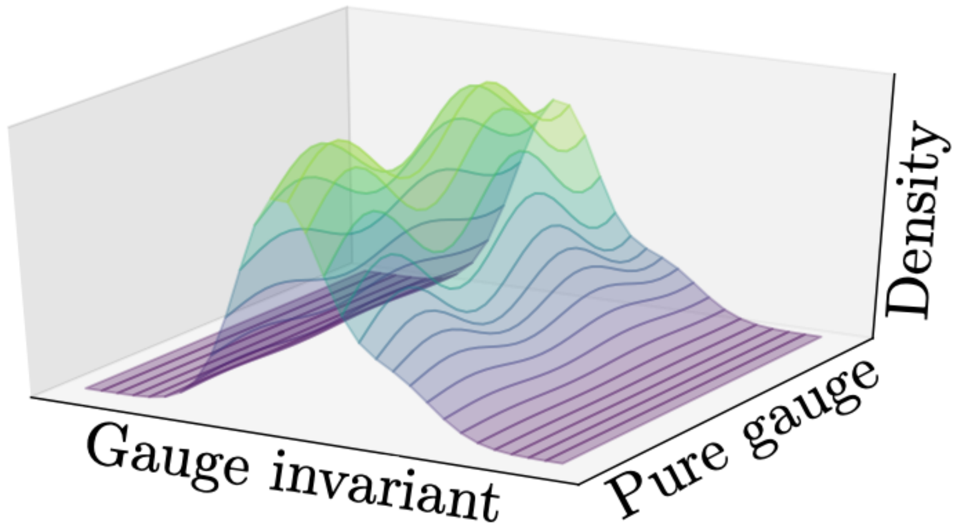
Incorporating symmetries

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

Gauge symmetries:



True distribution



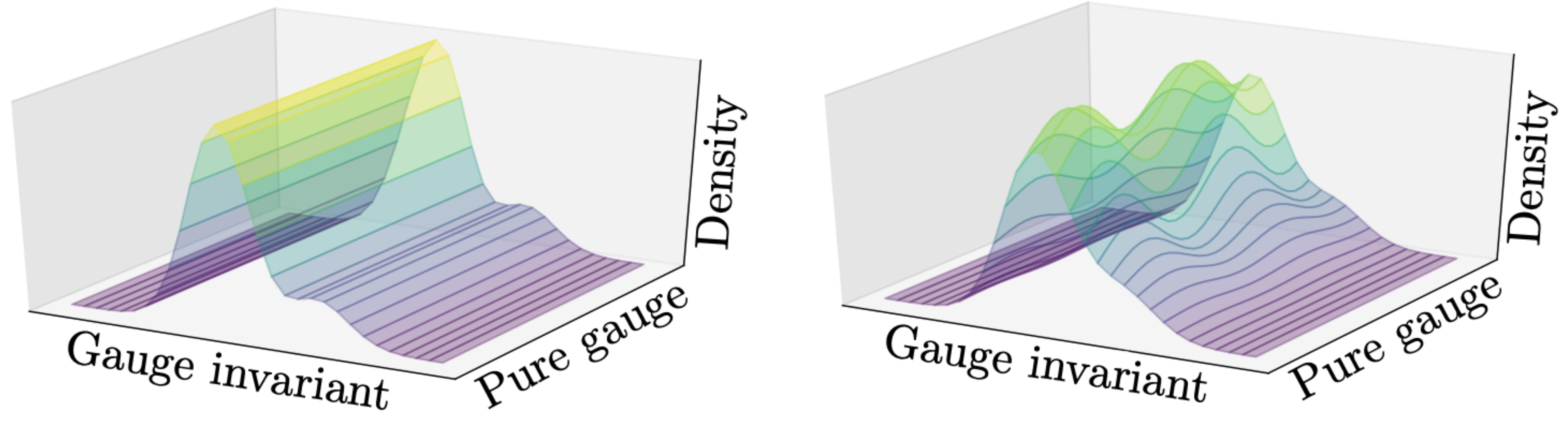
Learned by naive ML

Fernando Romero-López, MIT

Incorporating symmetries

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

Gauge symmetries:

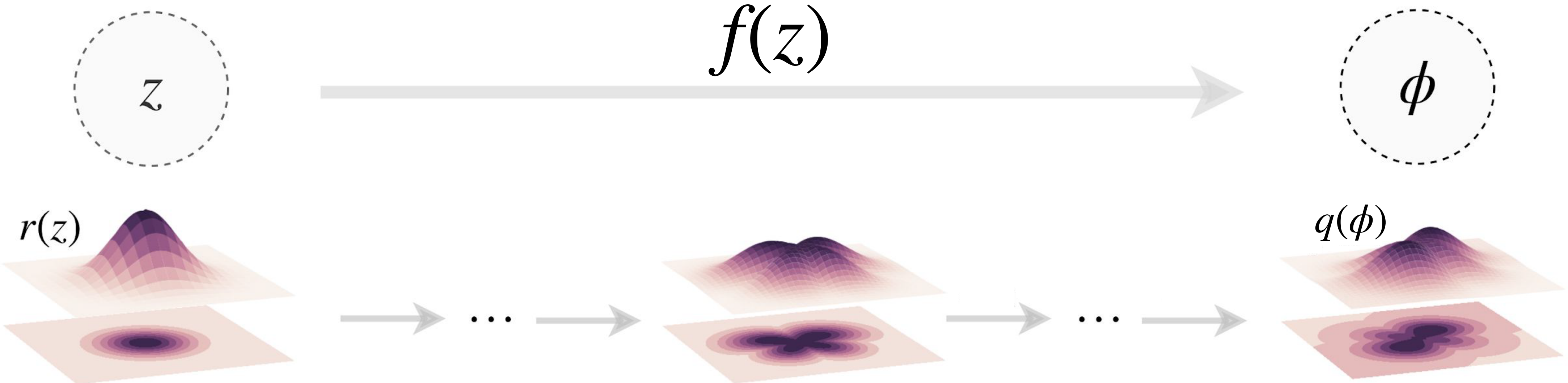


True distribution **Learned by naive ML**

General approach:

1. Invariant base distribution

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

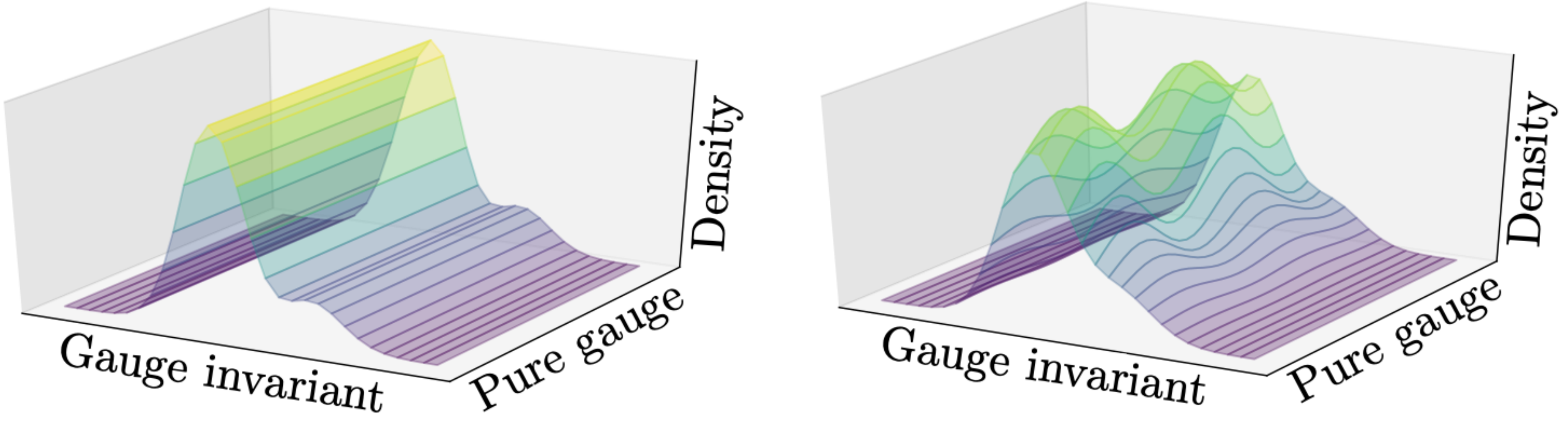


Fernando Romero-López, MIT

Incorporating symmetries

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

Gauge symmetries:



True distribution | **Learned by naive ML**

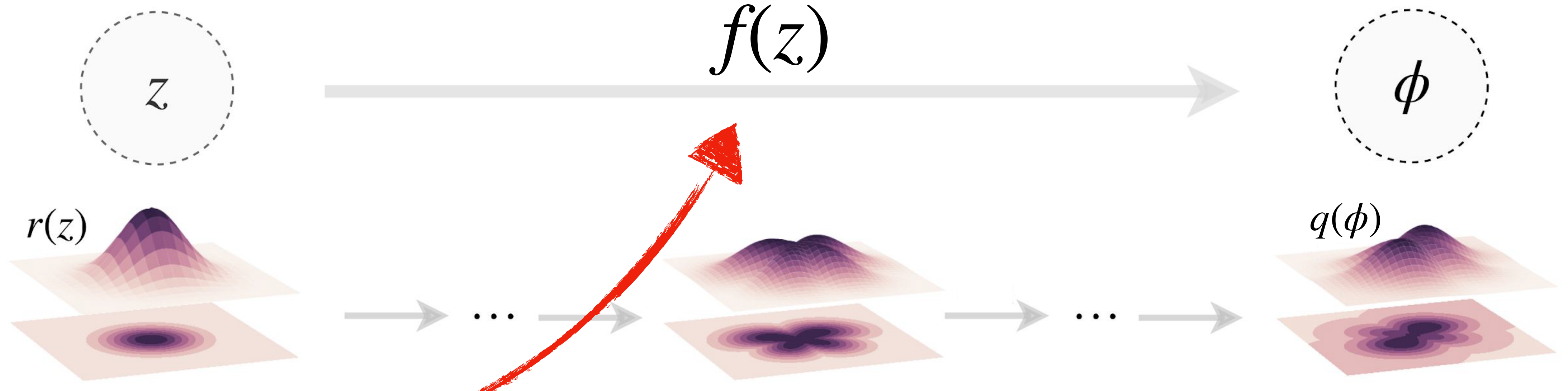
General approach:

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

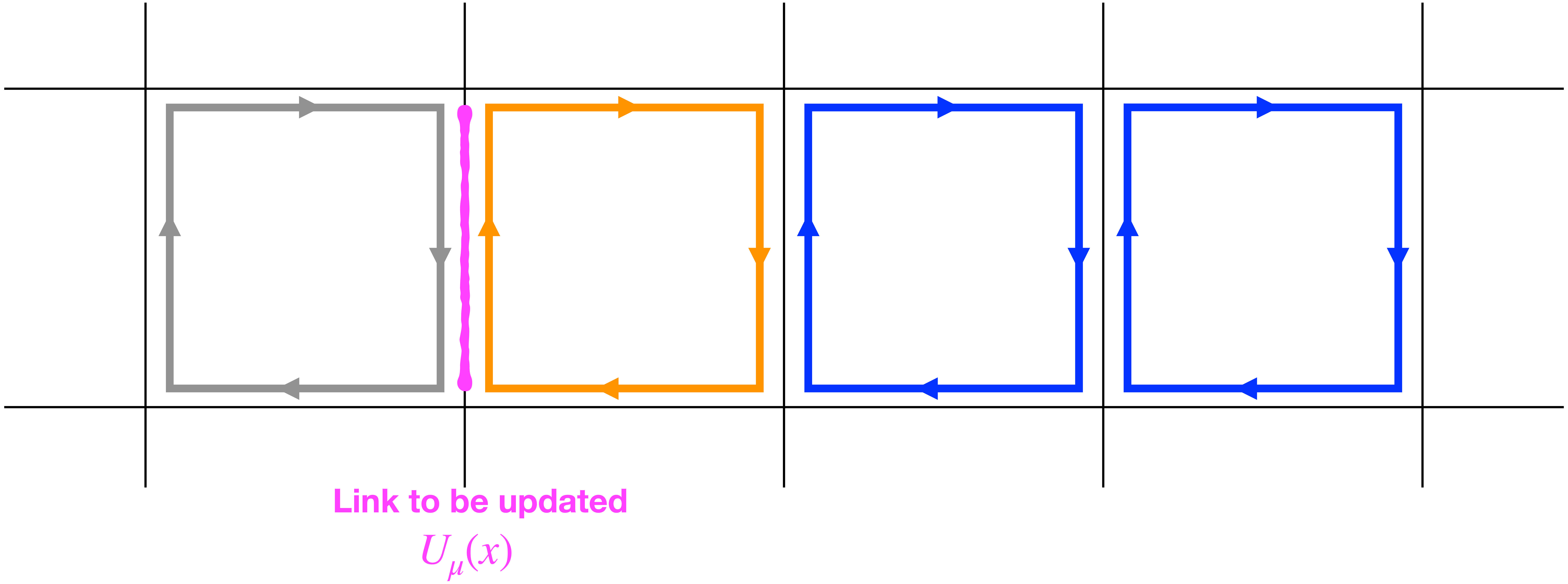
- **Gauge invariant prior + gauge-equivariant transformations:**

$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$
$$g\big(\Omega(x)U_\mu(x)\Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$

Flows with gauge variables

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

$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$
$$g\big(\Omega(x) U_\mu(x) \Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$



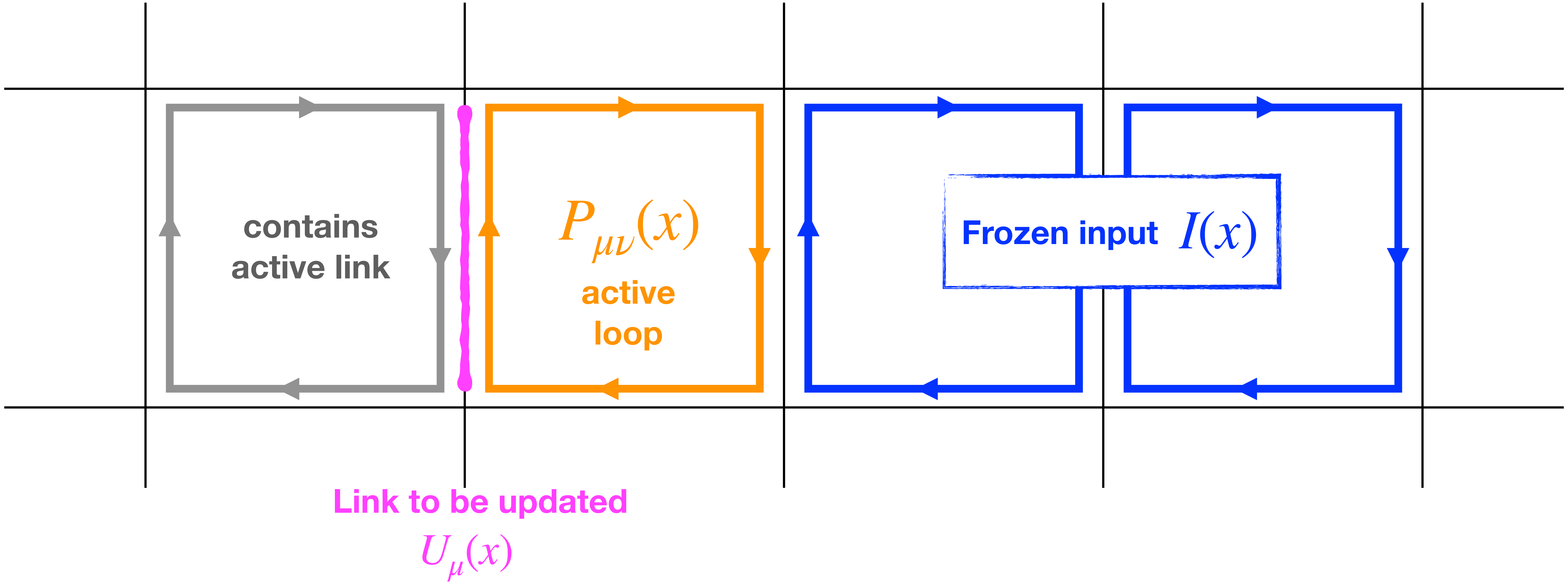
Fernando Romero-López, MIT

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

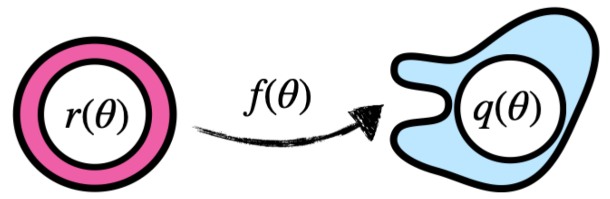
- **Gauge invariant prior + gauge-equivariant transformations:**

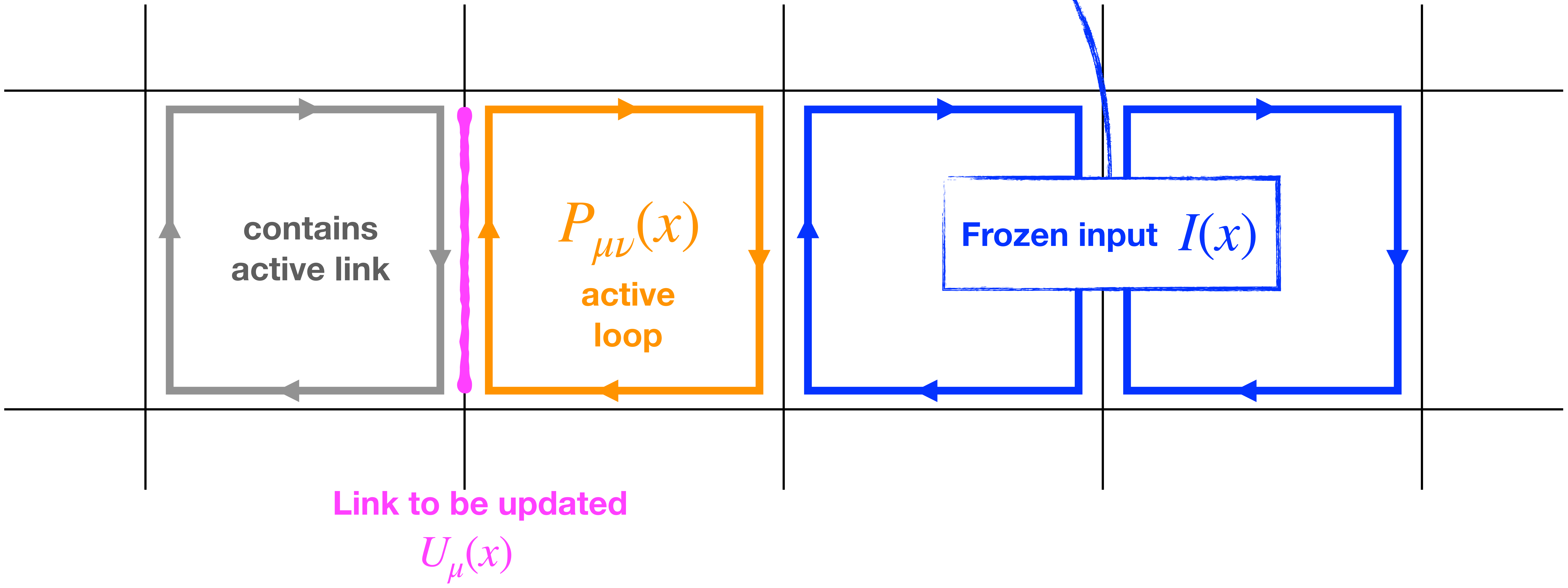
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

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

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

e.g. circular splines





Fernando Romero-López, MIT

Flows with gauge variables

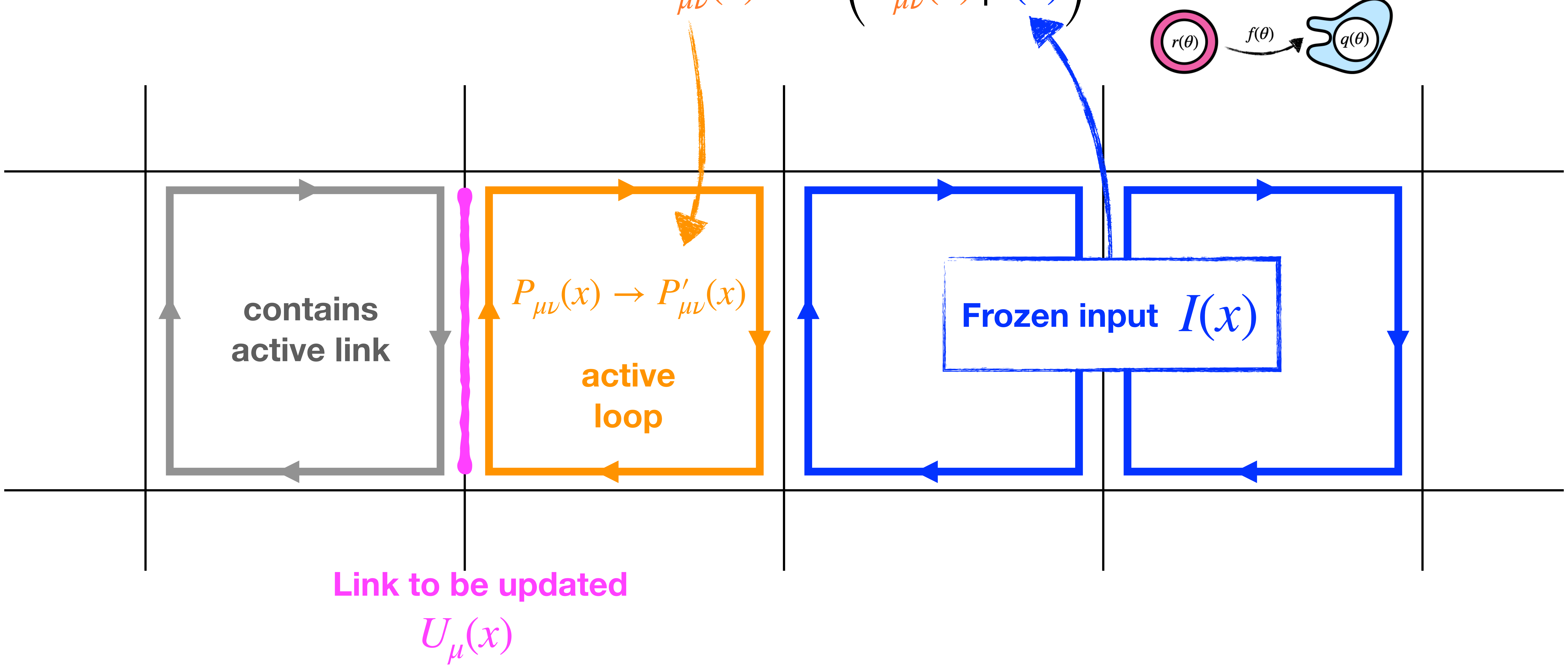
- **Gauge invariant prior + gauge-equivariant transformations:**

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

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

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

e.g. circular splines



Fernando Romero-López, MIT

Flows with gauge variables

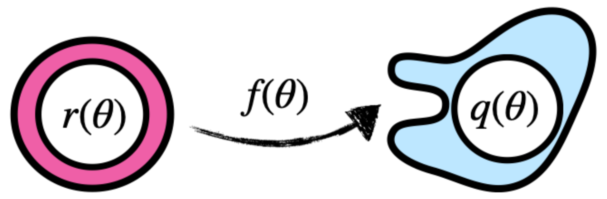
- **Gauge invariant prior + gauge-equivariant transformations:**

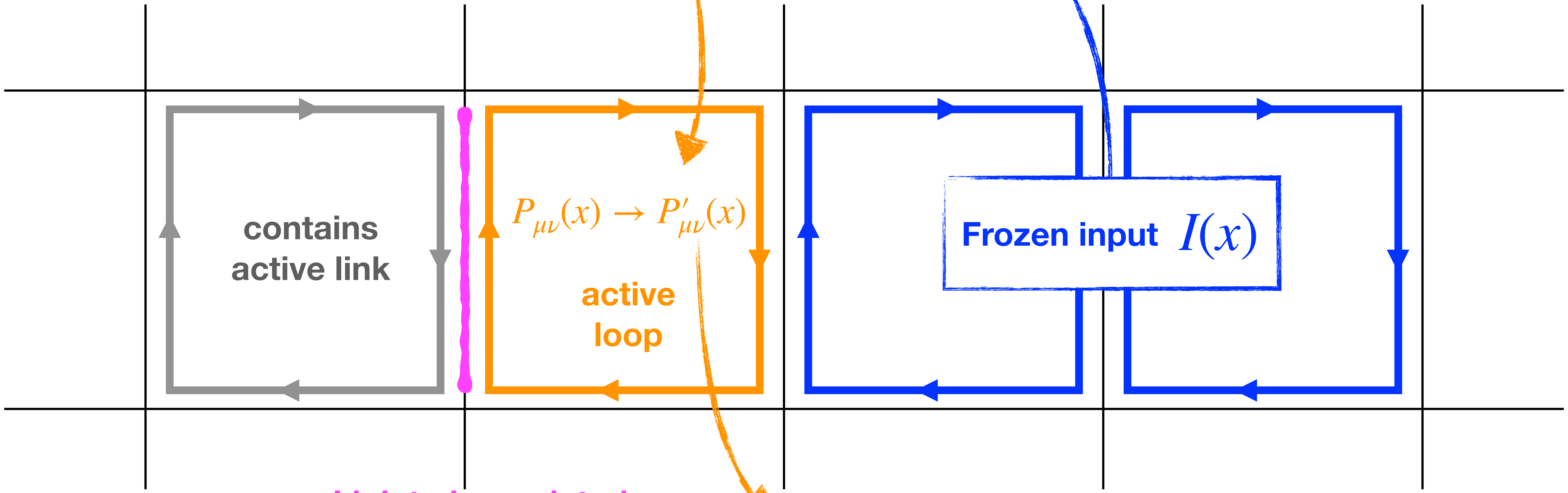
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

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

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

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

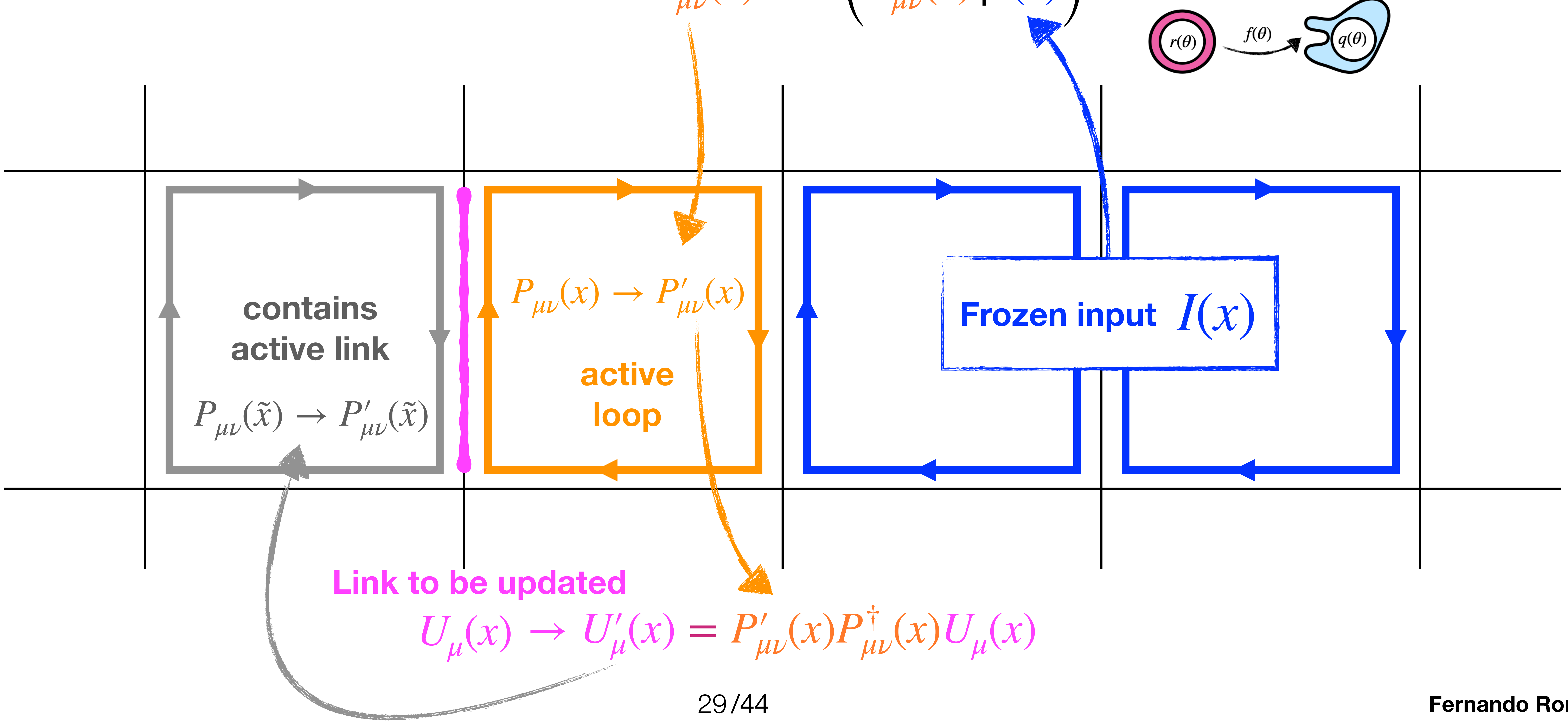
- **Gauge invariant prior + gauge-equivariant transformations:**

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

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

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

e.g. circular splines



Link to be updated

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

Flows with gauge variables

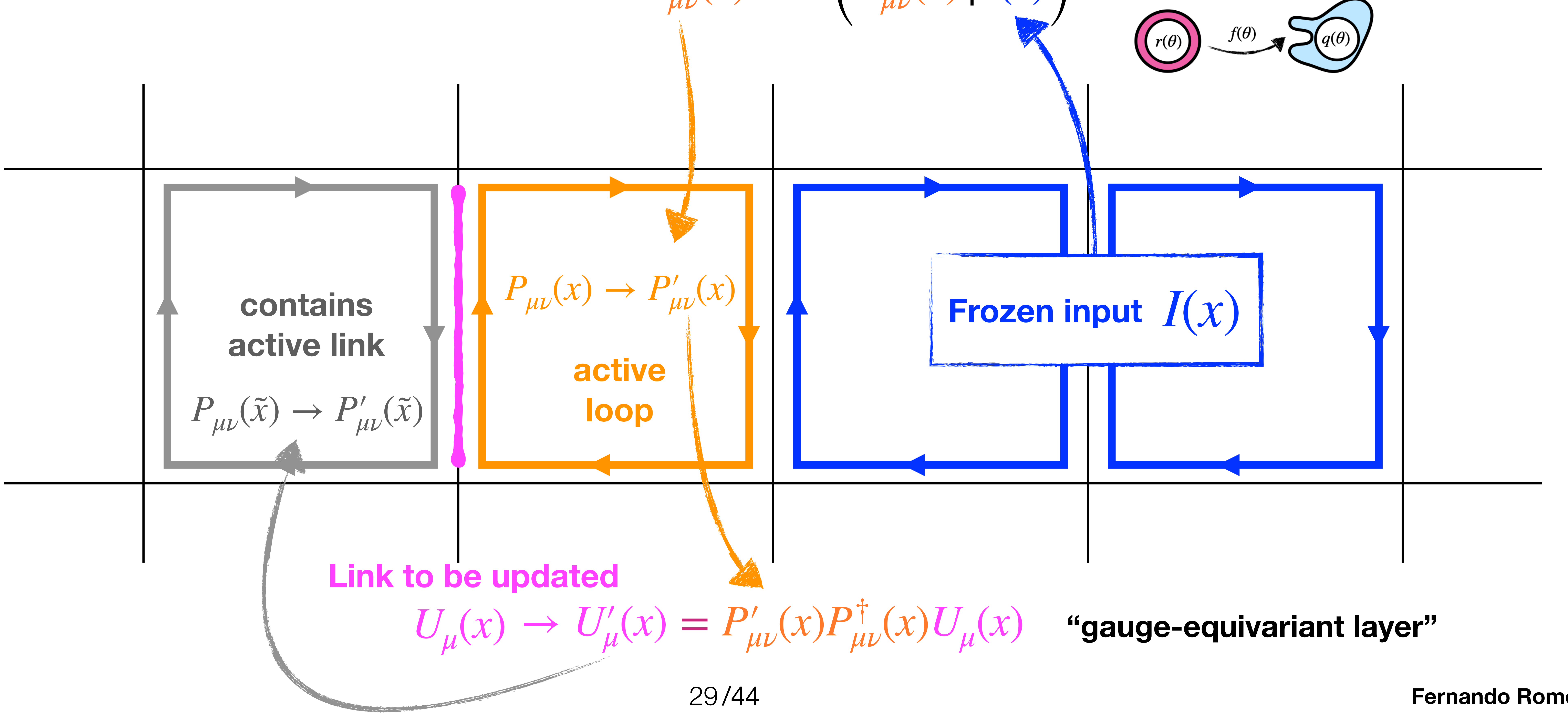
- **Gauge invariant prior + gauge-equivariant transformations:**

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

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

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

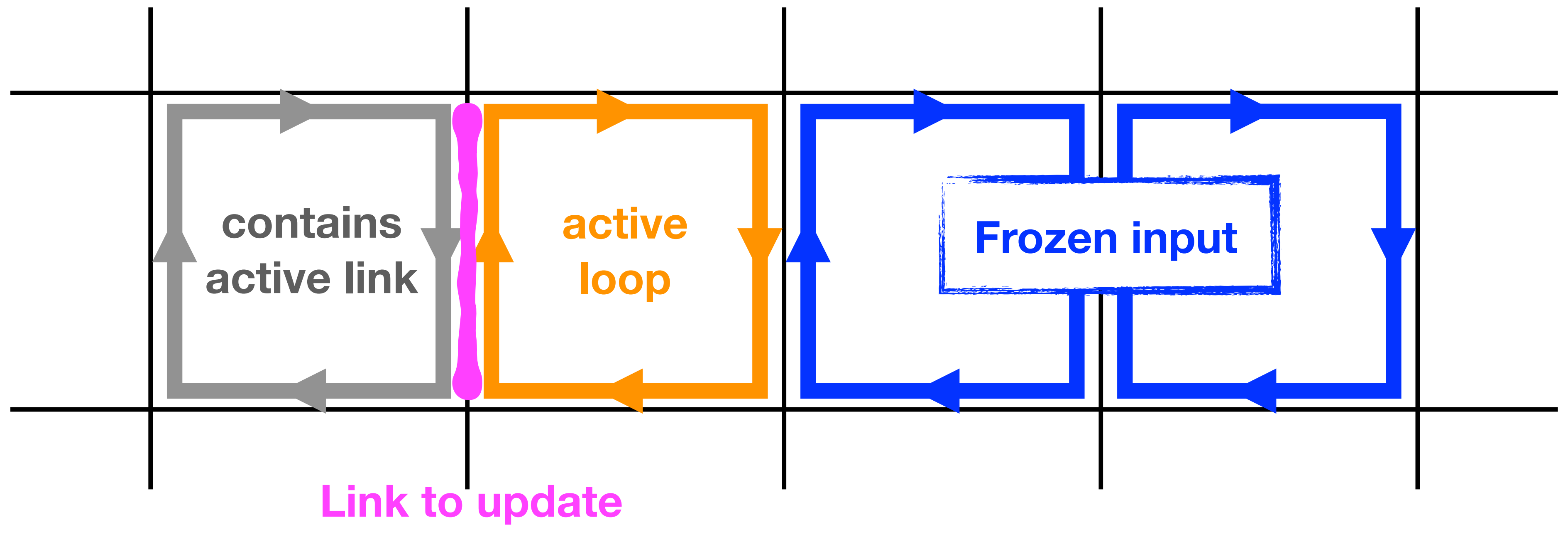
e.g. circular splines



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

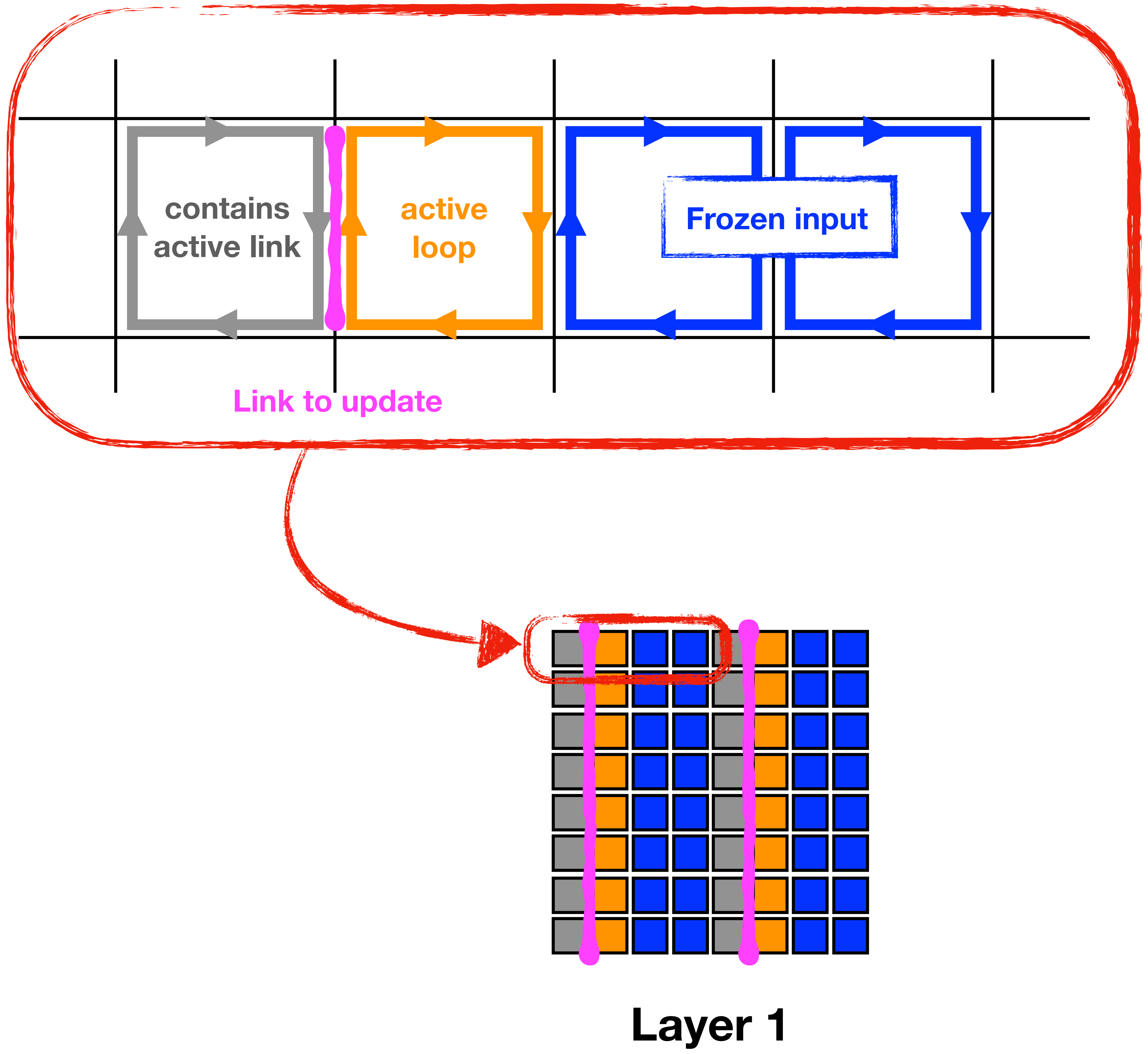
"gauge-equivariant layer"

Gauge-equivariant models



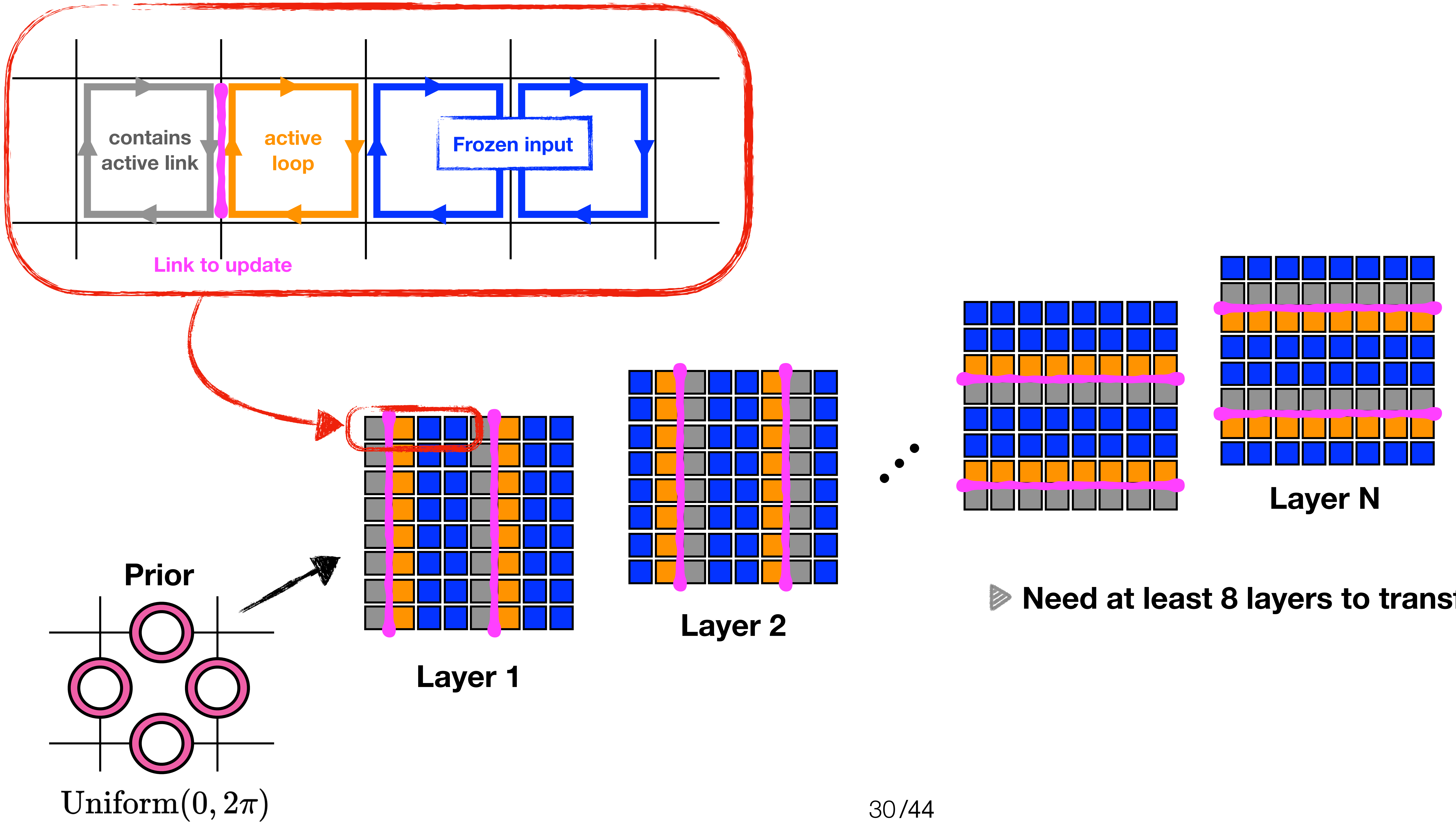
Fernando Romero-López, MIT

Gauge-equivariant models



Fernando Romero-López, MIT

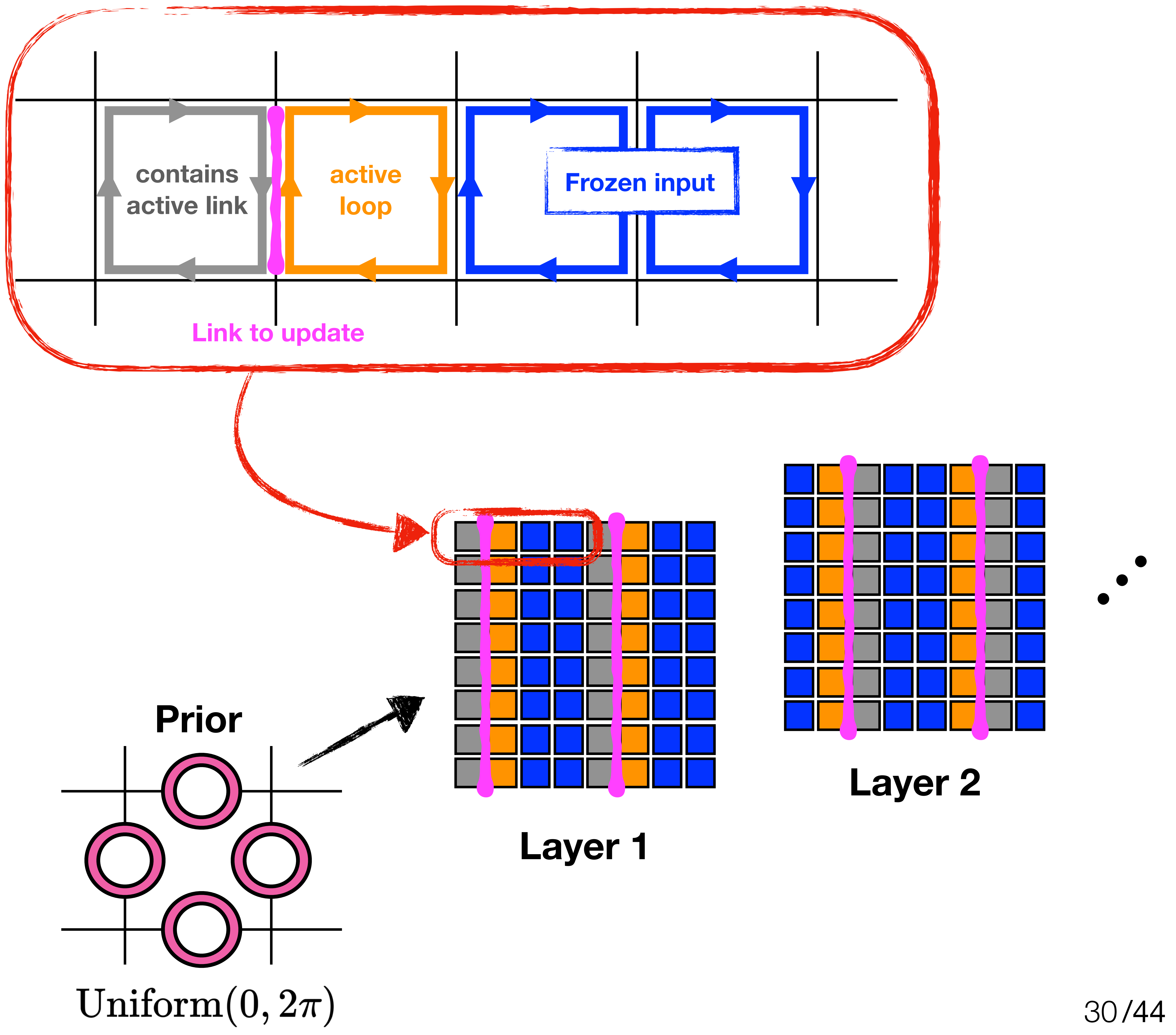
Gauge-equivariant models

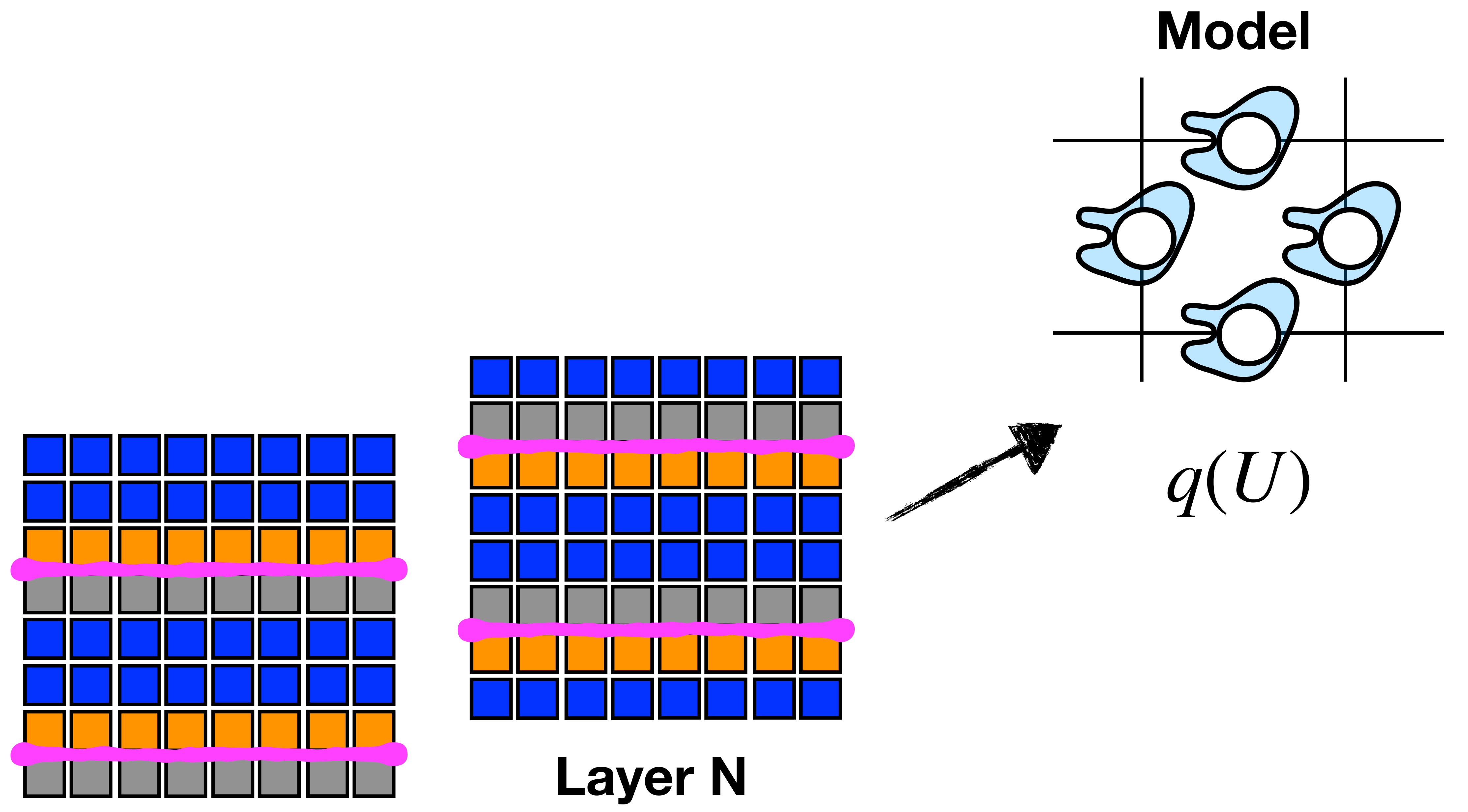


- Need at least 8 layers to transform all links

Fernando Romero-López, MIT

Gauge-equivariant models



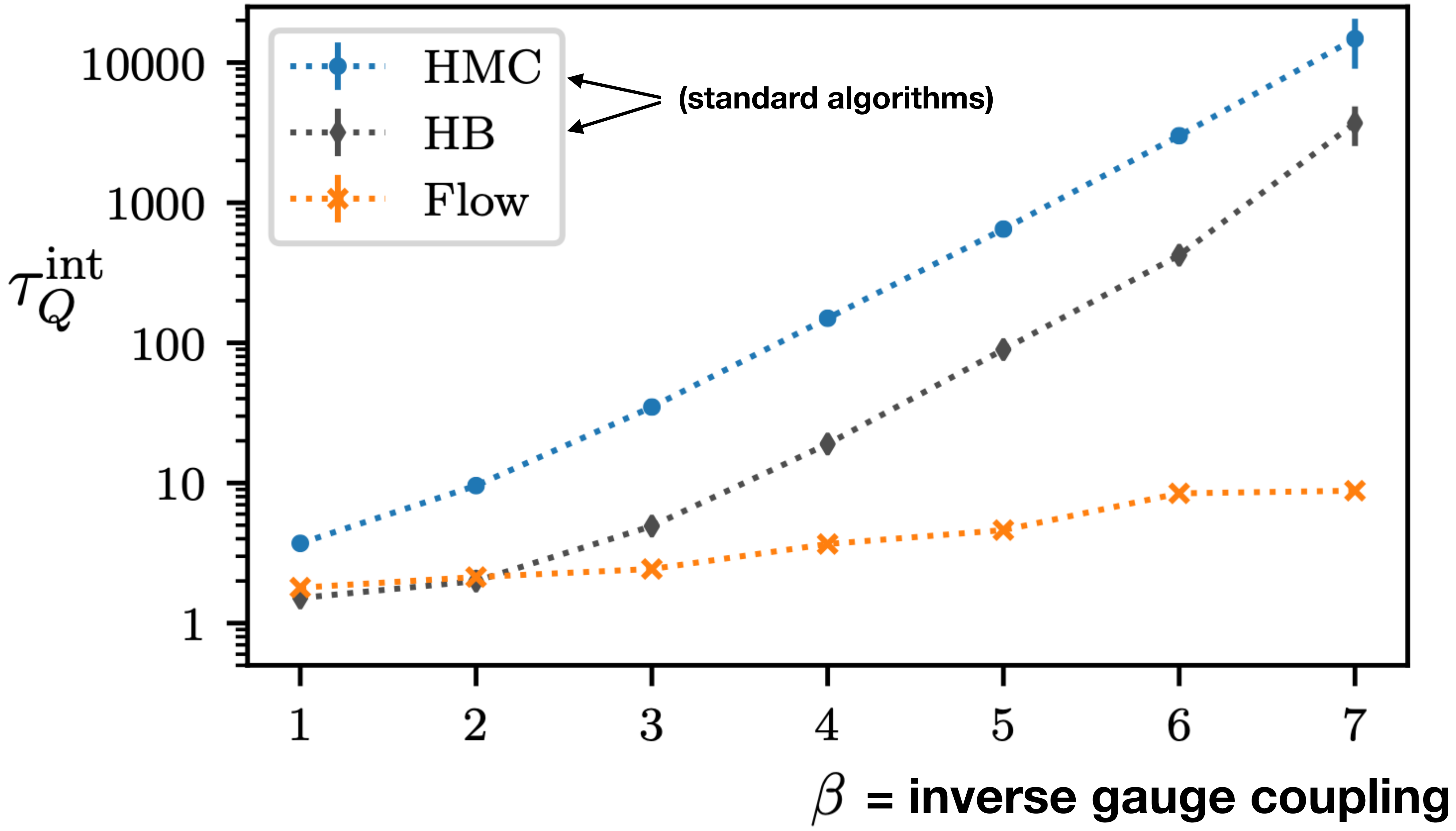


- Need at least 8 layers to transform all links
- ✓ Similar approach can be applied to SU(N)
  [Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan 2003.06413]

Fernando Romero-López, MIT

Results in 2D U(1) theory

Cost per independent sample



Autocorrelation time
of the topological charge

Continuum limit

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

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

Fernando Romero-López, MIT

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 N_f=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

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 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**

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

Fermionic theories

- **The fermionic part of the action**

Dirac-Wilson operator

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

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

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

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

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

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!

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

"exact determinant"

(gauge part)

(assuming N_f=2)

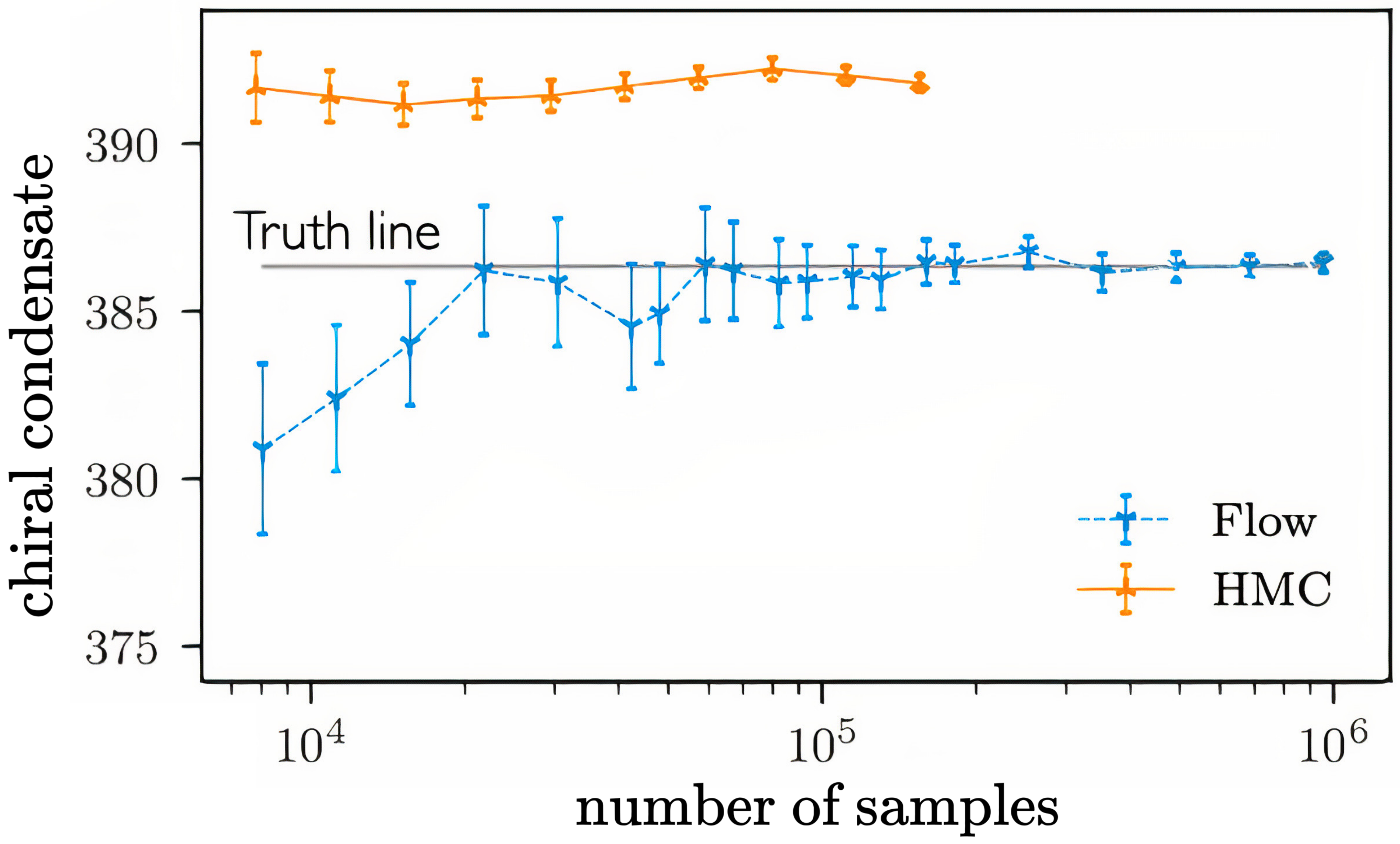
Fernando Romero-López, MIT

The Schwinger model at criticality

Critical parameters:

- **Vanishing fermion mass**
- **Diverging correlation length**

! Hardest to simulate in standard approaches



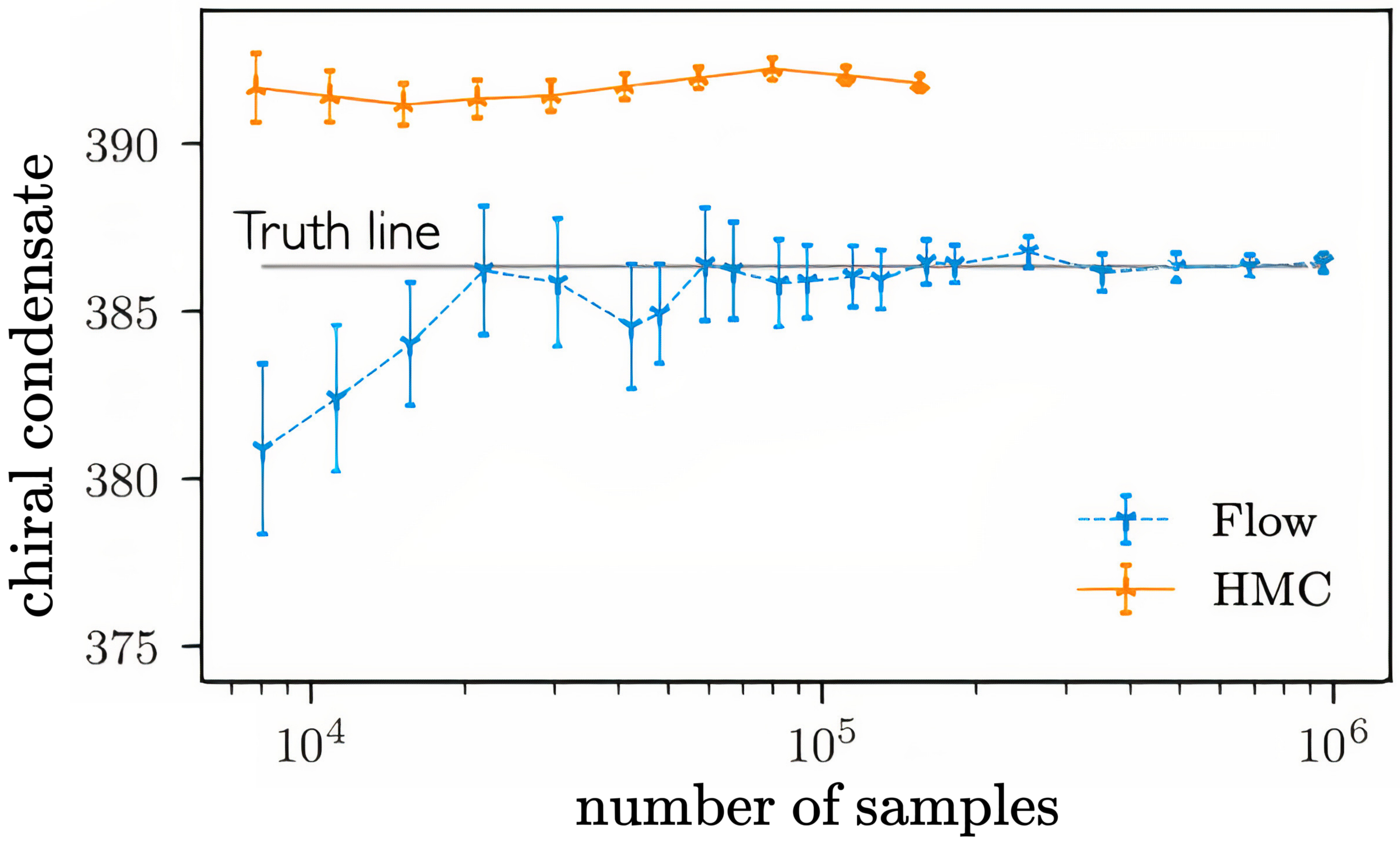
[Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Urban 2202.11712]

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]

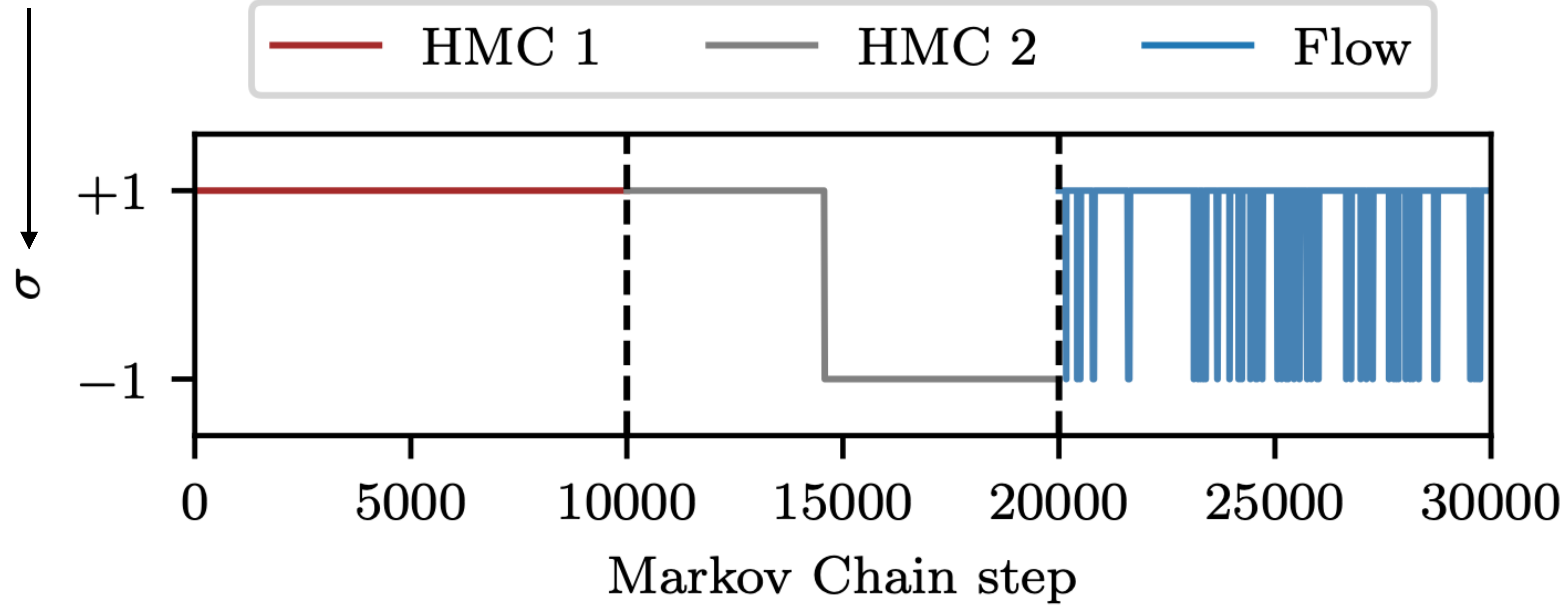
The Schwinger model at criticality

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.**

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

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!

Fernando Romero-López, MIT

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

Fernando Romero-López, MIT

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

- Joint target distribution

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

- 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”

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"

$$\boxed{f_m(\chi)}$$

$$\boxed{f_c(z \mid U)}$$

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"

$\boxed{f_m(\chi)}$

existing gauge architectures

[Kanwar et al, 2003.06413]

[Boyda et al, 2008.05456]

$\boxed{f_c(z \mid U)}$

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

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

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

"marginal"

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

"conditional"

χ

$f_m(\chi)$ — U

$f_c(z \mid U)$

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

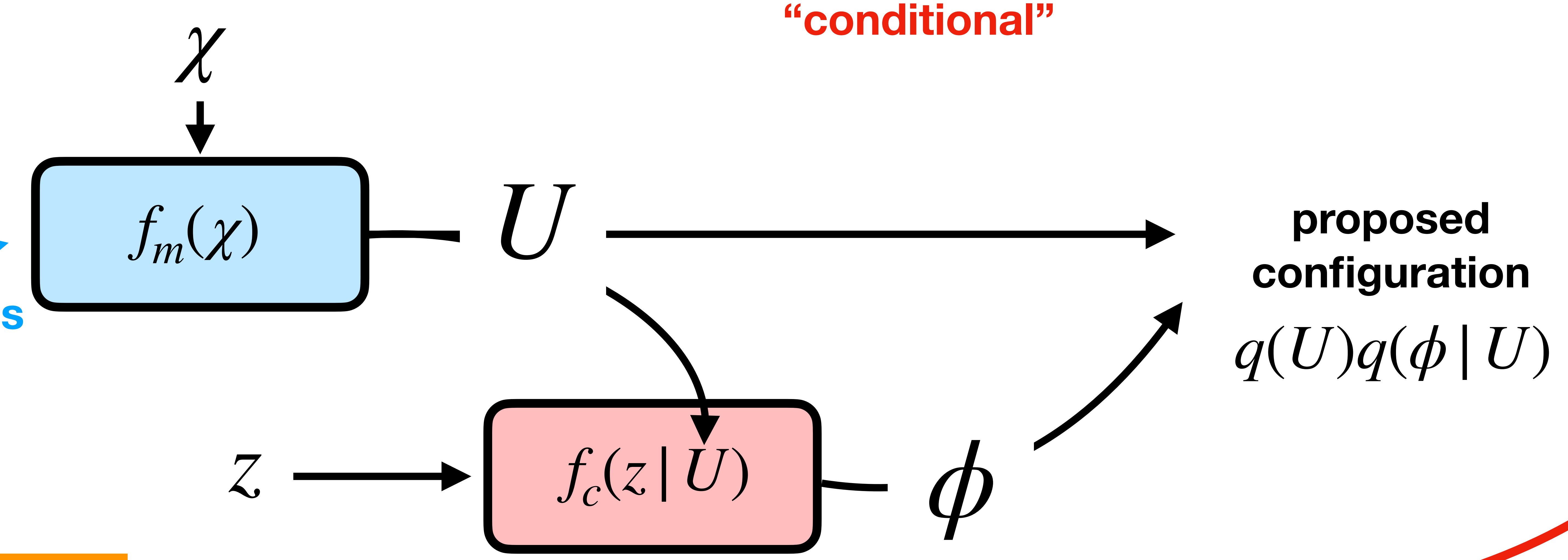
$$p(U, \phi) = p(U)p(\phi \mid U)$$

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

"marginal"

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

"conditional"



existing gauge architectures

[Kanwar et al, 2003.06413]

[Boyda et al, 2008.05456]

Different flow models to approximate marginal and conditional distributions

new conditional architectures

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

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

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:**

$$\phi'(x) = A(U)\,\phi(x) + B(U)\,\underbrace{U_\mu(x)\phi(x+\mu)}_{\text{parallel-transported neighbor}}$$

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:

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

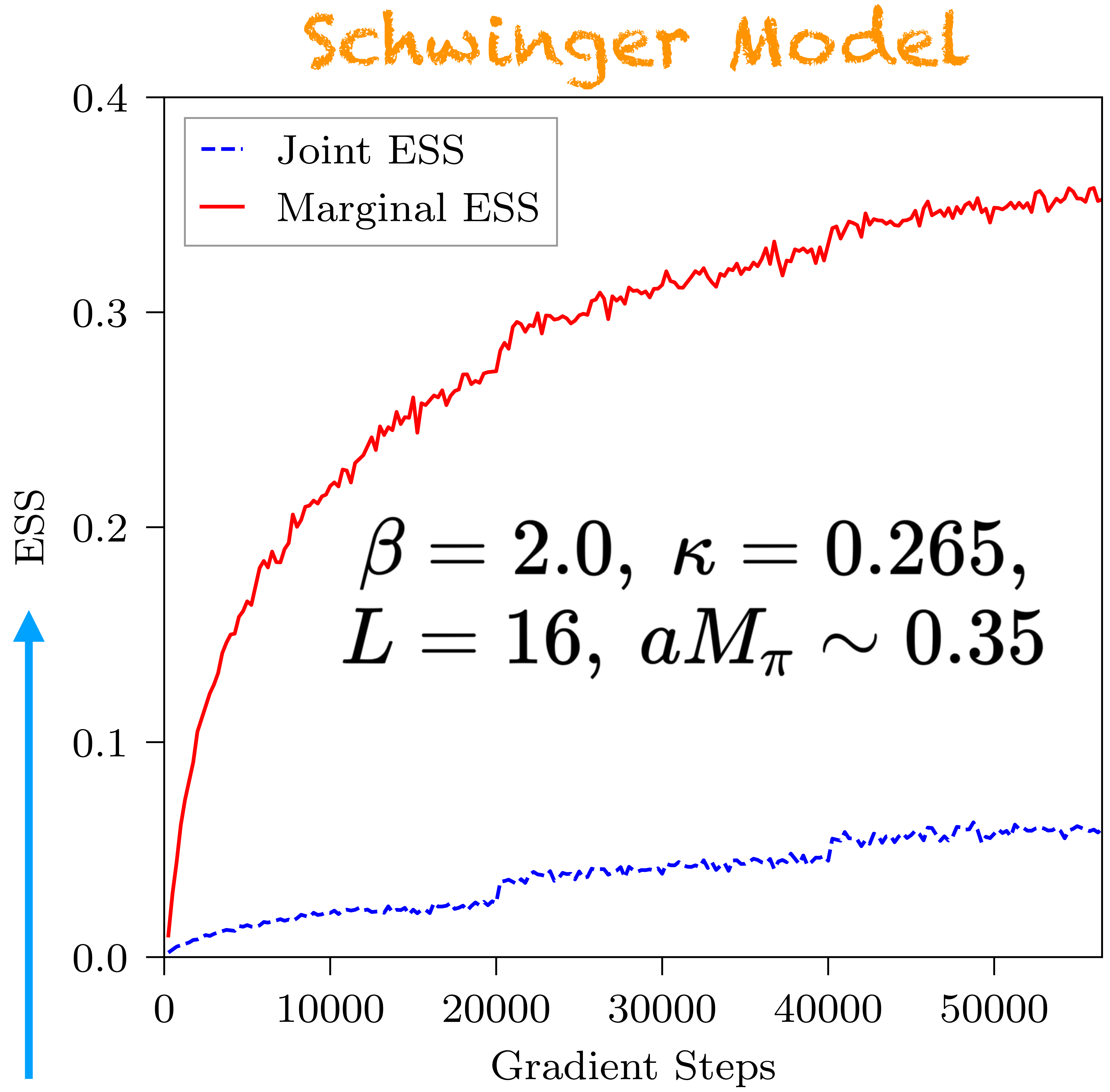
parallel-transported neighbor

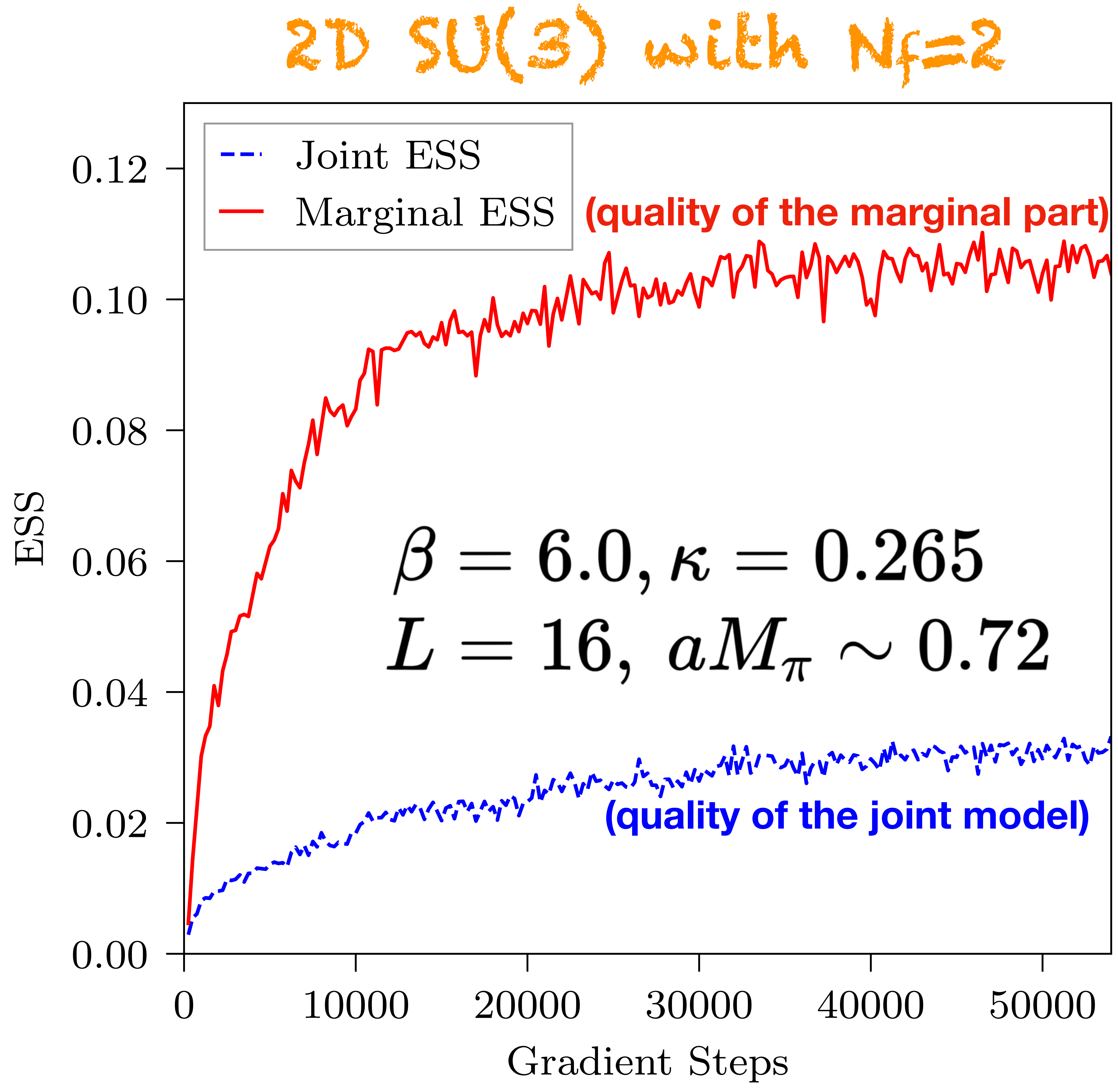
(gauge equivariance)

NN outputs
(gauge-invariant inputs)

Fernando Romero-López, MIT

Examples of flow models





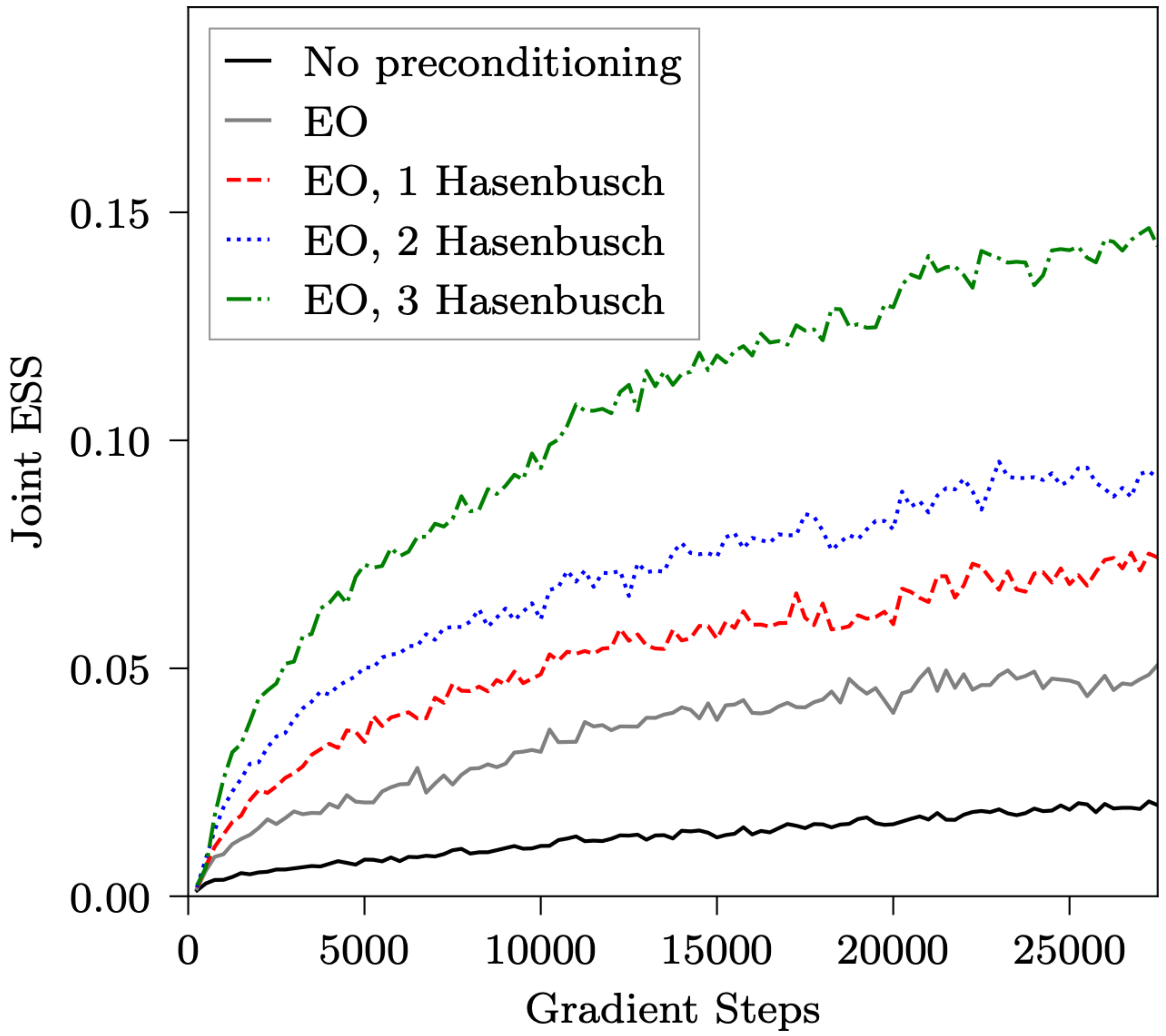
"Effective sample size" ≡ model quality
(ESS=1 for a perfect model)

[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

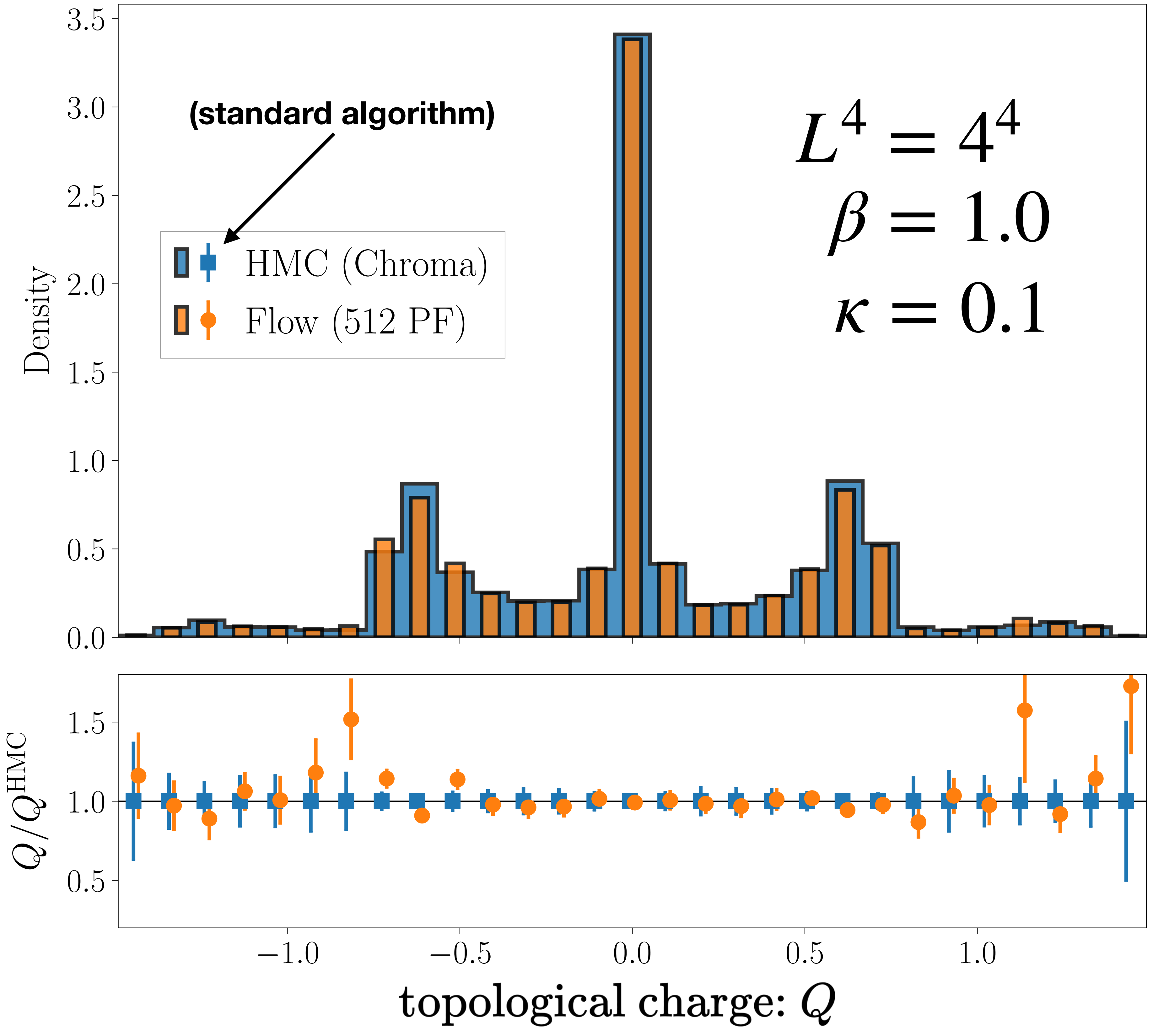
Can we ML generate lattice QCD?

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

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**

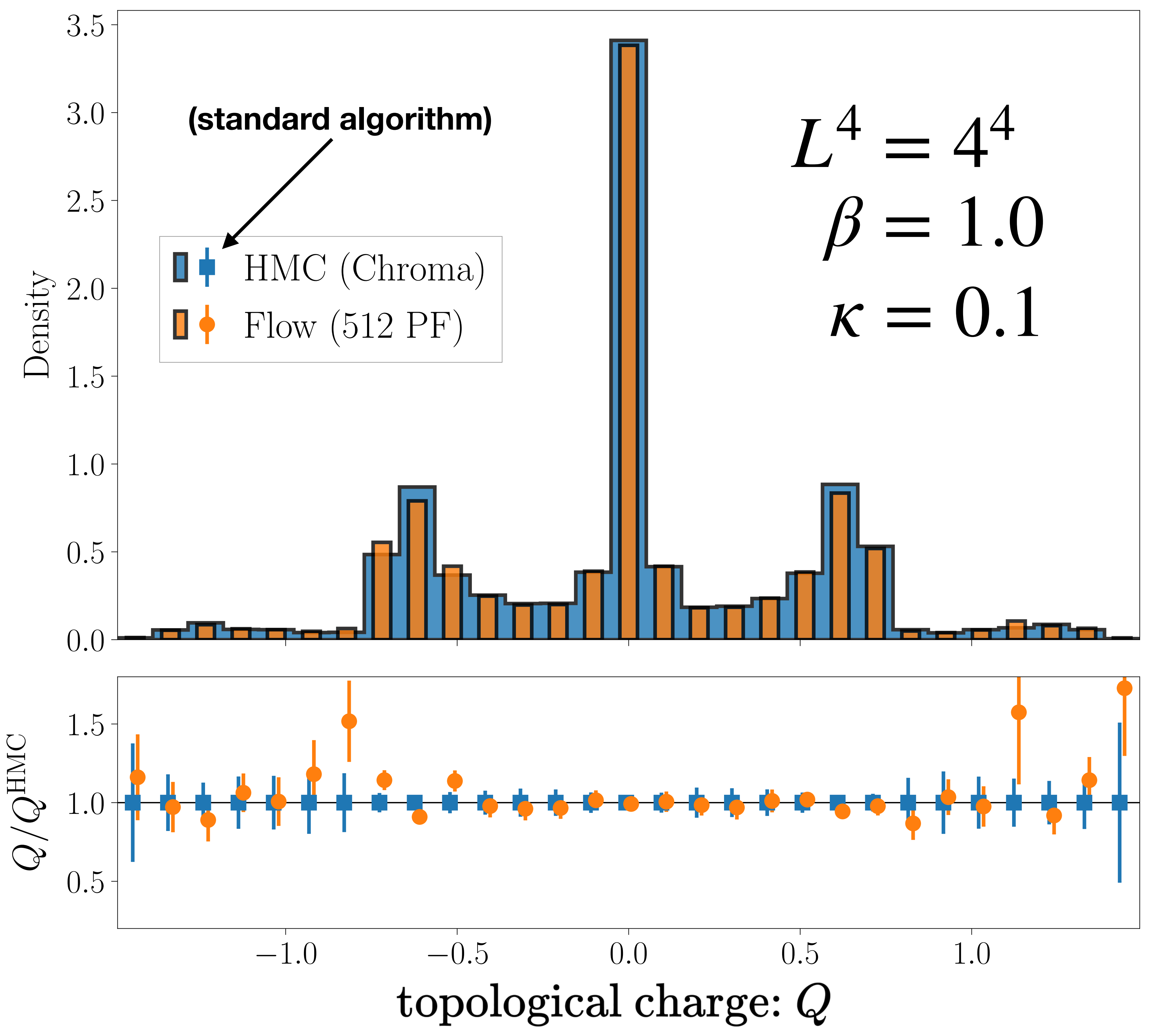


[Abbott, … , FRL, …, et al 2208.03832]

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."

Summary & Outlook

Summary & Outlook

- Lattice QCD is the first-principle treatment of the strong interaction at hadronic energies

- Continuum limit in lattice QCD is limited by the performance of algorithms (critical slowing down)

- Flow-based sampling has the potential to mitigate this problem

- Success has been demonstrated in 2D fermionic gauge theories: Schwinger model
  [Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Urban 2202.11712]

- Demonstration in QCD at small scale
  [Abbott et al, 2208.03832]

- Next steps: scale flow-based sampling to state-of-the-art QCD
  [Abbott et al, 2211.07541]

Summary & Outlook

- Lattice QCD is the first-principle treatment of the strong interaction at hadronic energies
- Continuum limit in lattice QCD is limited by the performance of algorithms (critical slowing down)
- Flow-based sampling has the potential to mitigate this problem
- Success has been demonstrated in 2D fermionic gauge theories: Schwinger model
  [Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Urban 2202.11712]
- Demonstration in QCD at small scale
  [Abbott et al, 2208.03832]
- Next steps: scale flow-based sampling to state-of-the-art QCD
  [Abbott et al, 2211.07541]

Thanks!

Back-up slides

The road to QCD

(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 in the strong-coupling region
[Abbott et al, 2208.03832]

…

State-of-the-art lattice QCD!

Fernando Romero-López, MIT

Recap of Pseudofermions

- **Action:** $$S(\psi,\bar\psi,U) = S_{\rm g}(U) + S_F(\psi,\bar\psi,U)$$

Gauge fields

Fermions

$$S_F(\psi,\bar\psi,U) = \sum_{f=1}^{N_f} \bar\psi_f D_f(U)\psi_f \longrightarrow \int \mathcal{D}[\psi,\bar\psi] e^{-S_F(\psi,\bar\psi,U)} = \prod_{f=1}^{N_f} \det D_f(U)$$

Fernando Romero-López, MIT

Recap of Pseudofermions

- **Action:** $$S(\psi,\bar\psi,U) = S_{\rm g}(U) + S_F(\psi,\bar\psi,U)$$

Gauge fields

Fermions

$$S_F(\psi,\bar\psi,U) = \sum_{f=1}^{N_f} \bar\psi_f D_f(U)\psi_f \longrightarrow \int \mathcal{D}[\psi,\bar\psi] e^{-S_F(\psi,\bar\psi,U)} = \prod_{f=1}^{N_f} \det D_f(U)$$

- **Evaluate 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

Recap of Pseudofermions

- **Action:** $S(\psi,\bar\psi,U) = S_{\rm g}(U) + S_F(\psi,\bar\psi,U)$

Gauge fields

Fermions

$$S_F(\psi,\bar\psi,U) = \sum_{f=1}^{N_f} \bar\psi_f D_f(U)\psi_f \longrightarrow \int \mathcal{D}[\psi,\bar\psi] e^{-S_F(\psi,\bar\psi,U)} = \prod_{f=1}^{N_f} \det D_f(U)$$

- **Evaluate 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**

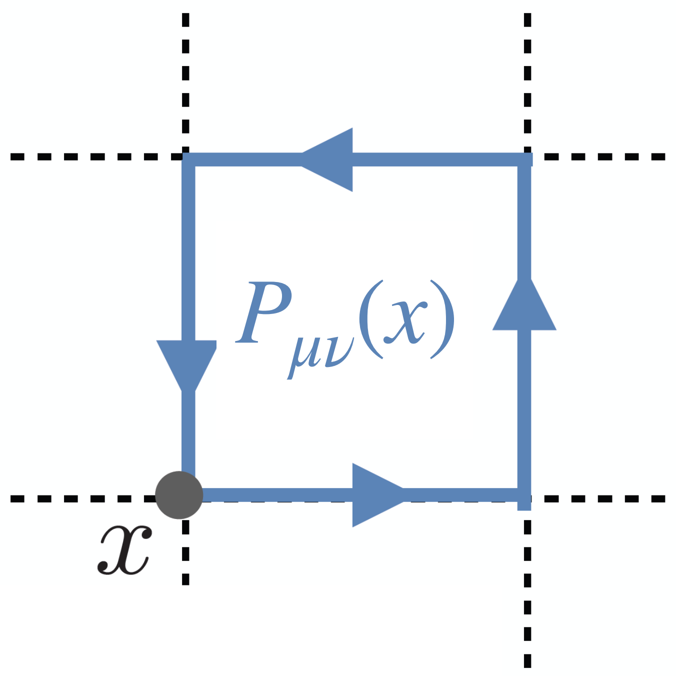
$$p(U,\phi) = \frac{1}{Z} e^{-S_g(U) - S_{\rm pf}(U,\phi)} \quad \text{with} \quad S_{\rm pf}(\phi,U) = \phi^\dagger \left[D(U)D^\dagger(U)\right]^{-1}\phi^\dagger$$

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

Fernando Romero-López, MIT

Flows on SU(3)

- **Gauge transformations maintain the eigenvalues of SU(3) matrices:**

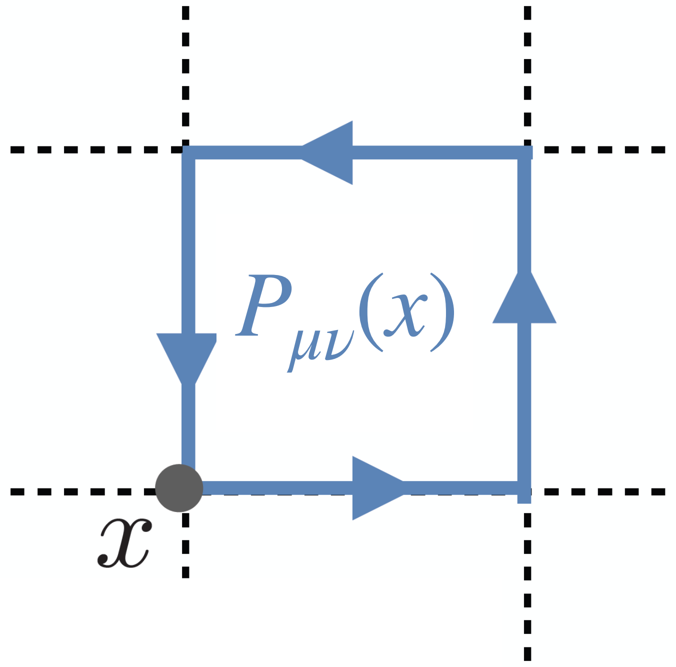


$$P_{\mu\nu} = X \begin{pmatrix} e^{i\theta_1} & 0 & 0 \\ 0 & e^{i\theta_2} & 0 \\ 0 & 0 & e^{i\theta_3} \end{pmatrix} X^\dagger$$

Fernando Romero-López, MIT

Flows on SU(3)

- **Gauge transformations maintain the eigenvalues of SU(3) matrices:**

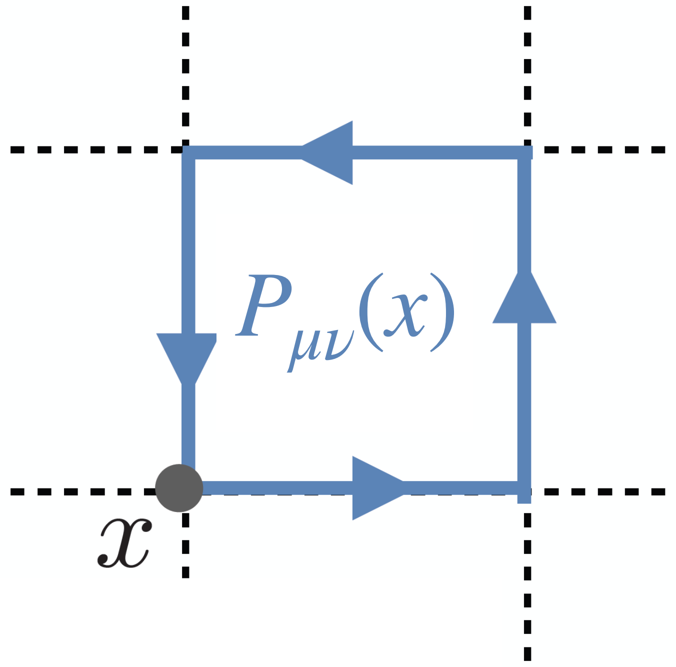


$$P_{\mu\nu} = X \begin{pmatrix} e^{i\theta_1} & 0 & 0 \\ 0 & e^{i\theta_2} & 0 \\ 0 & 0 & e^{i\theta_3} \end{pmatrix} X^\dagger$$

Use transformation on eigenvalues

Flows on SU(3)

- Gauge transformations maintain the eigenvalues of SU(3) matrices:



$$P_{\mu\nu} = X \begin{pmatrix} e^{i\theta_1} & 0 & 0 \\ 0 & e^{i\theta_2} & 0 \\ 0 & 0 & e^{i\theta_3} \end{pmatrix} X^\dagger$$

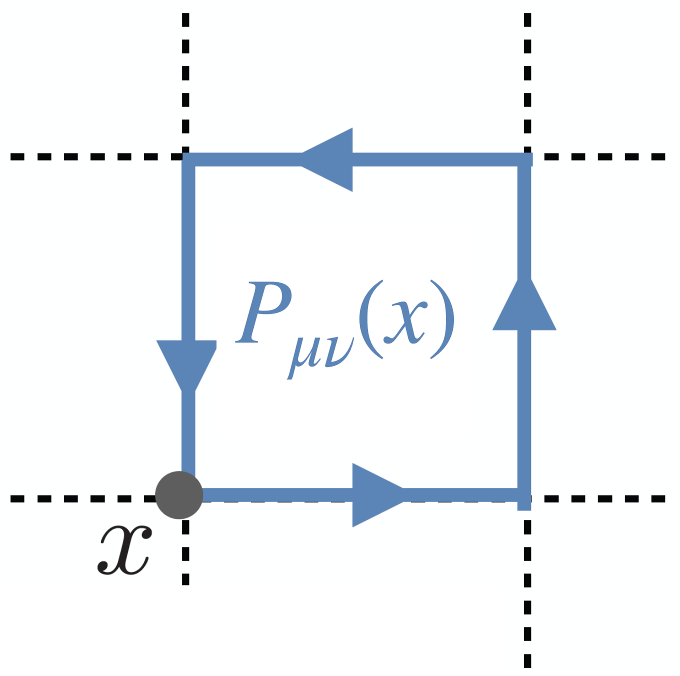
Use transformation on eigenvalues

- However, ordering of the eigenvalues is not unique: <u>need permutation equivariance</u>

Fernando Romero-López, MIT

Flows on SU(3)

- **Gauge transformations maintain the eigenvalues of SU(3) matrices:**



$$P_{\mu\nu} = X \begin{pmatrix} e^{i\theta_1} & 0 & 0 \\ 0 & e^{i\theta_2} & 0 \\ 0 & 0 & e^{i\theta_3} \end{pmatrix} X^\dagger$$

Use transformation on eigenvalues

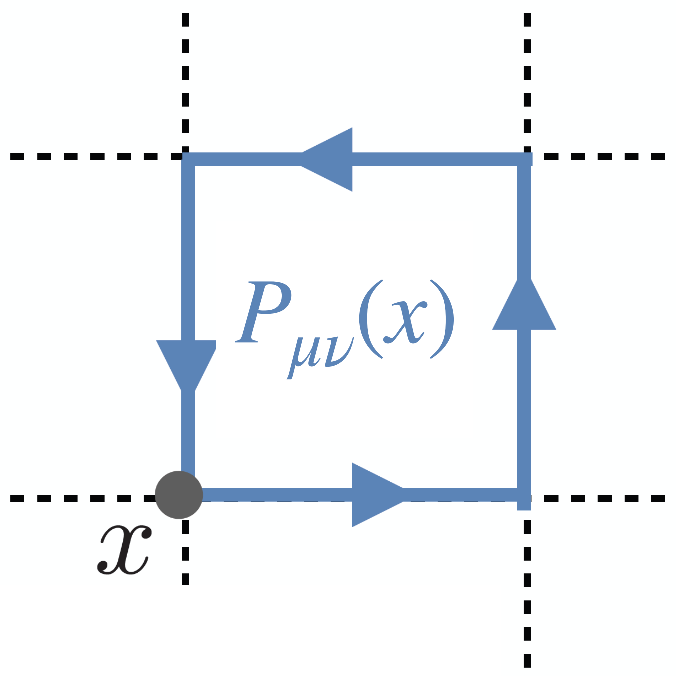
- **However, ordering of the eigenvalues is not unique: <u>need permutation equivariance</u>**
  - **Place eigenvalues in canonical ordering**
  - **Transform eigenvalues**
  - **Undo canonical ordering**

$$\mathrm{iscanon}(\theta_1, \theta_2, \theta_3) = \begin{cases} \theta_3 \geq \theta_2 \geq \theta_1 & \sum_i \theta_i = 0 \\ \theta_1 \geq \theta_3 \geq \theta_2 & \sum_i \theta_i = 2\pi \\ \theta_2 \geq \theta_1 \geq \theta_3 & \sum_i \theta_i = -2\pi \end{cases}$$

Fernando Romero-López, MIT

Flows on SU(3)

- Gauge transformations maintain the eigenvalues of SU(3) matrices:



$$P_{\mu\nu} = X \begin{pmatrix} e^{i\theta_1} & 0 & 0 \\ 0 & e^{i\theta_2} & 0 \\ 0 & 0 & e^{i\theta_3} \end{pmatrix} X^\dagger$$

Use transformation on eigenvalues

- However, ordering of the eigenvalues is not unique: <u>need permutation equivariance</u>

  - Place eigenvalues in canonical ordering
  - Transform eigenvalues
  - Undo canonical ordering

$$\mathrm{iscanon}(\theta_1, \theta_2, \theta_3) = \begin{cases} \theta_3 \geq \theta_2 \geq \theta_1 & \sum_i \theta_i = 0 \\ \theta_1 \geq \theta_3 \geq \theta_2 & \sum_i \theta_i = 2\pi \\ \theta_2 \geq \theta_1 \geq \theta_3 & \sum_i \theta_i = -2\pi \end{cases}$$

✓ Successfully applied to SU(3) in 2D

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

Fernando Romero-López, MIT

Exactness via Markov Chain

❗ **Trained models are not perfect but exactness is essential.**

Exactness via Markov Chain

❗ Trained models are not perfect but exactness is essential.

- Guarantee exactness by forming a Markov chain with accept/reject Metropolis-Hastings steps

Acceptance probability ⟶

$$A\left(\phi^{(i-1)}, \phi'\right) = \min\left(1, \frac{q\left(\phi^{(i-1)}\right)}{p\left(\phi^{(i-1)}\right)} \frac{p(\phi')}{q(\phi')}\right)$$

True distribution

Model distribution

Fernando Romero-López, MIT

Exactness via Markov Chain

! Trained models are not perfect but exactness is essential.

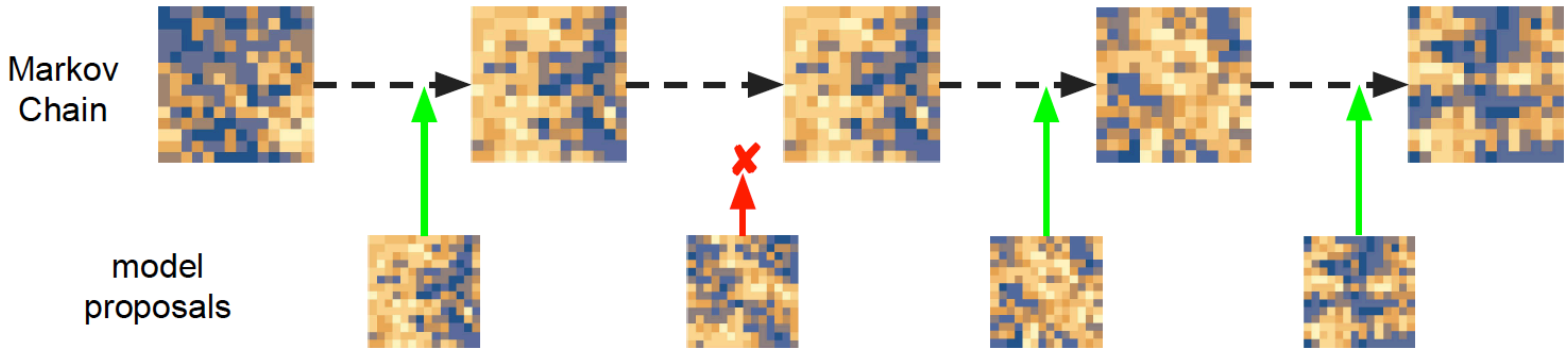
Guarantee exactness by forming a Markov chain with accept/reject Metropolis-Hastings steps

Acceptance probability $\longrightarrow$

$$A\left(\phi^{(i-1)}, \phi'\right) = \min\left(1, \frac{q\left(\phi^{(i-1)}\right)}{p\left(\phi^{(i-1)}\right)} \frac{p(\phi')}{q(\phi')}\right)$$

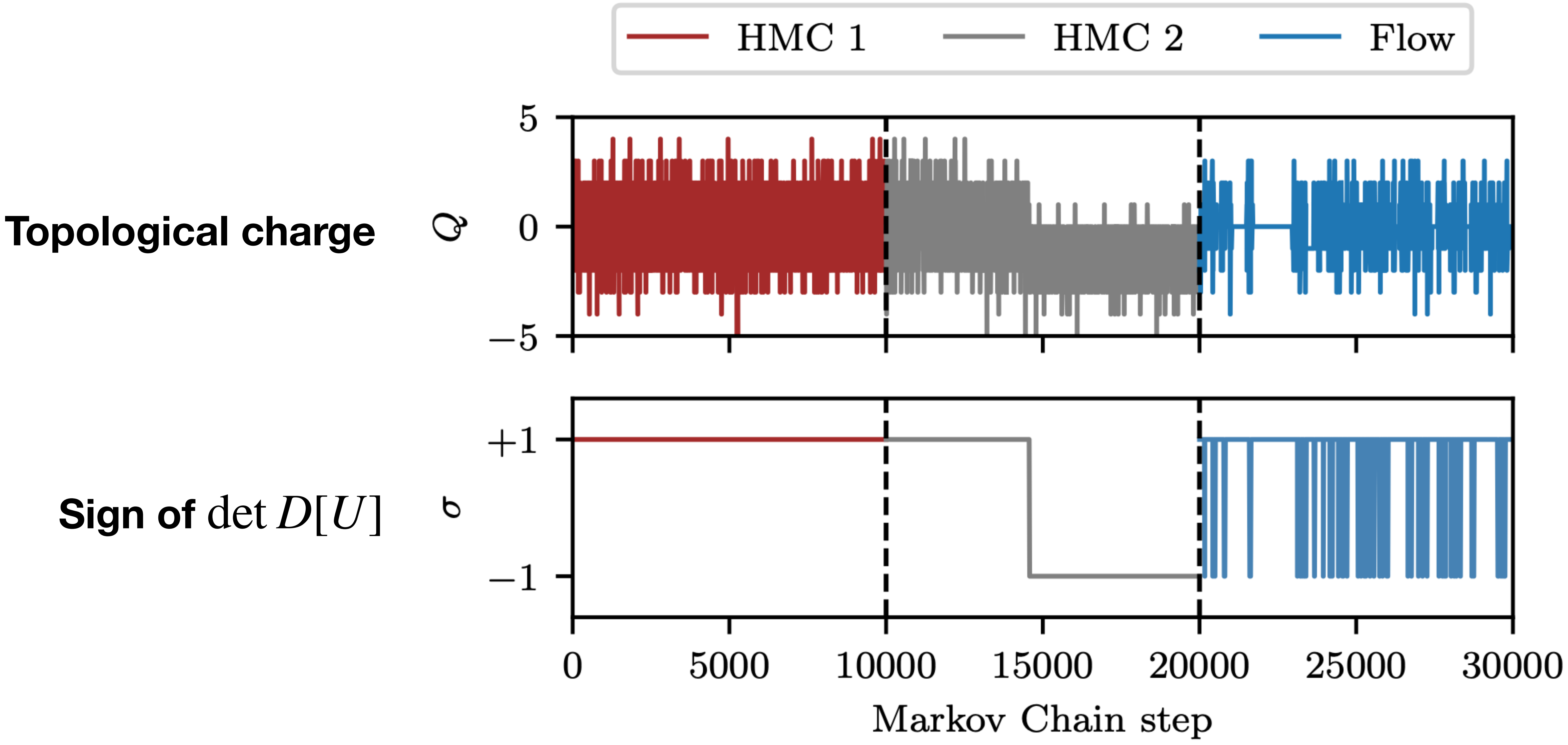
True distribution

Model distribution



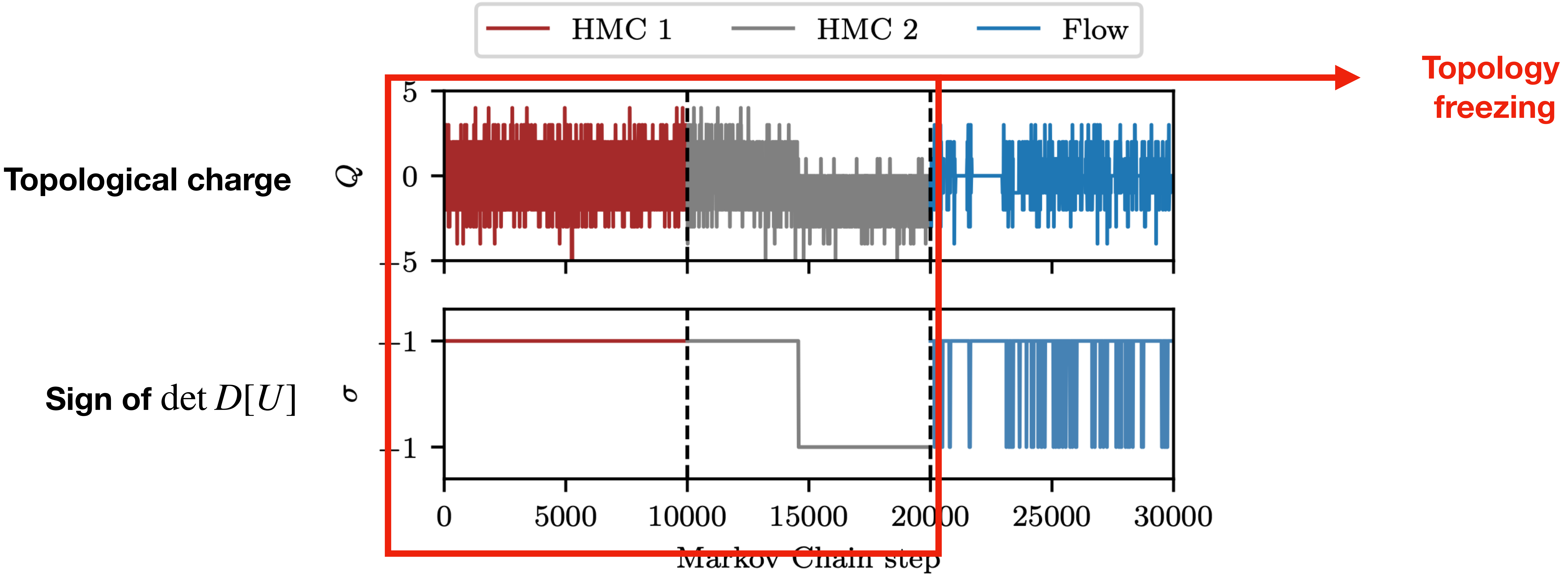
Fernando Romero-López, MIT

The Schwinger model at criticality



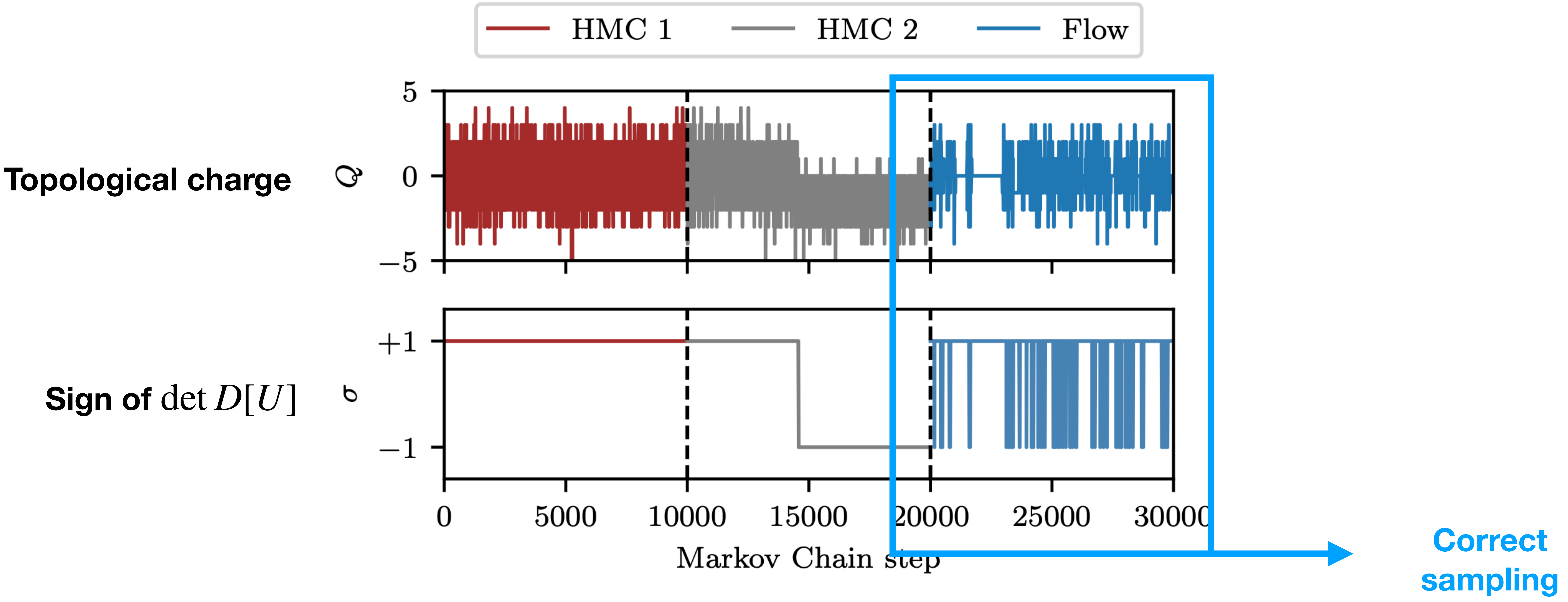
[Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Urban 2202.11712]

The Schwinger model at criticality



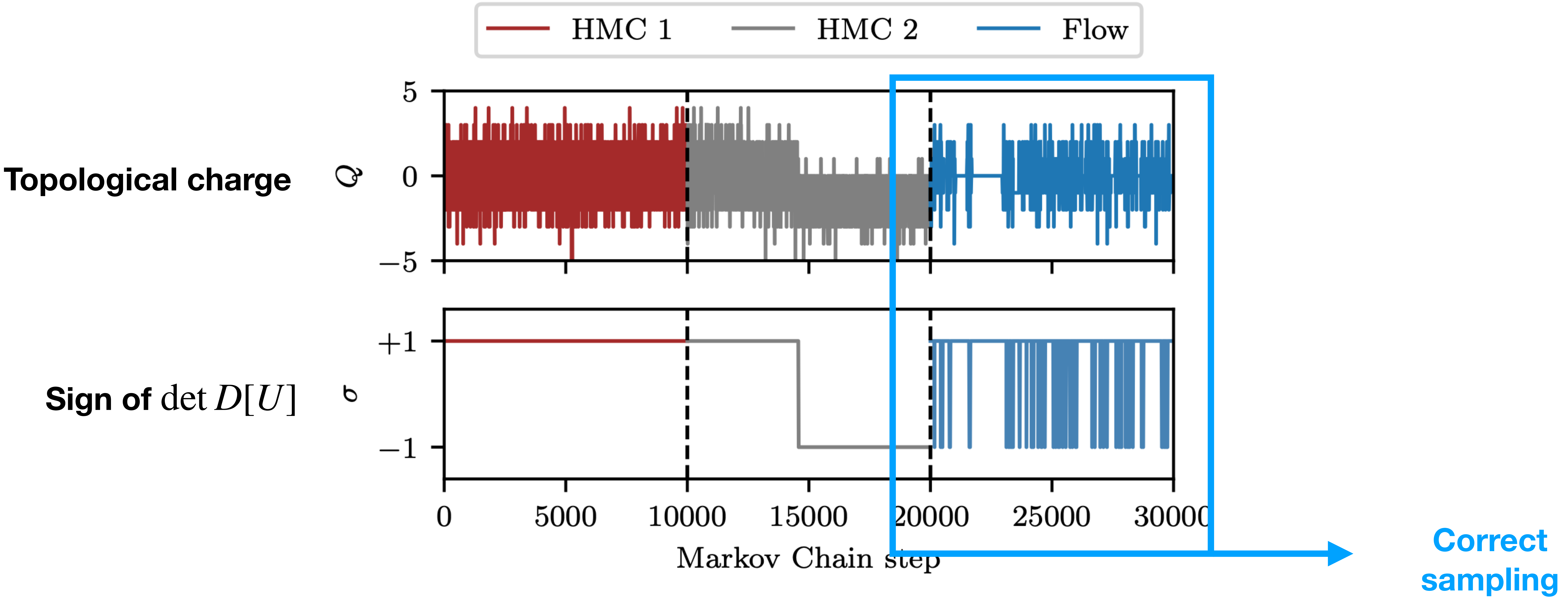
[Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Urban 2202.11712]

The Schwinger model at criticality



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

The Schwinger model at criticality



✓ **Flow-based sampling mitigates topology freezing even at criticality**

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