Gauge-equivariant flow models for sampling in LFT with pseudofermions

Fernando Romero-López

Lattice 2022 @ Bonn

August 8th

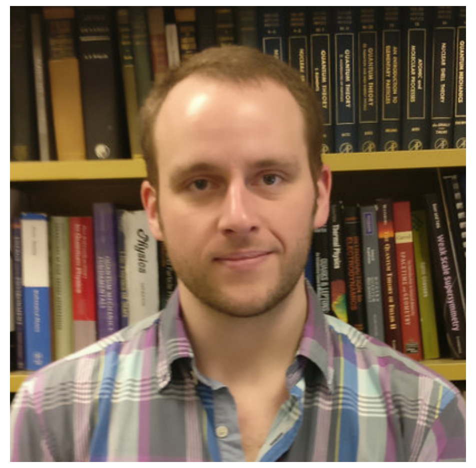
Massachusetts Institute of Technology

fernando@mit.edu

Collaboration (non-exhaustive)

Massachusetts Institute of Technology

- Phiala Shanahan



- Dan Hackett

- Fernando Romero-Lopez

- Ryan Abbot

- Denis Boyda

Argonne NATIONAL LABORATORY

Massachusetts Institute of Technology

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

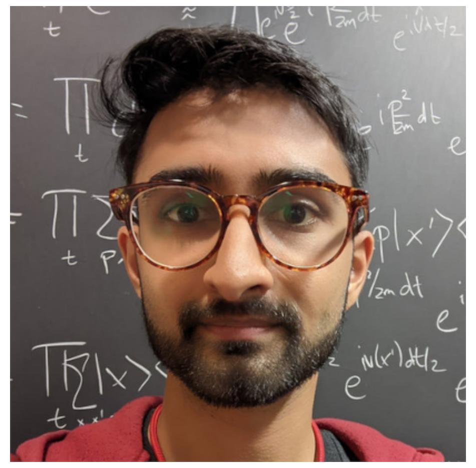
- Julian Urban

NEW YORK UNIVERSITY

- Kyle Cranmer

- Michael Albergo

UNIVERSITÄT BERN



- Gurtej Kanwar

DeepMind

- Sébastien Racanière

- Danilo Rezende

This talk is based on:

Flow-based sampling in the lattice Schwinger model at criticality

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

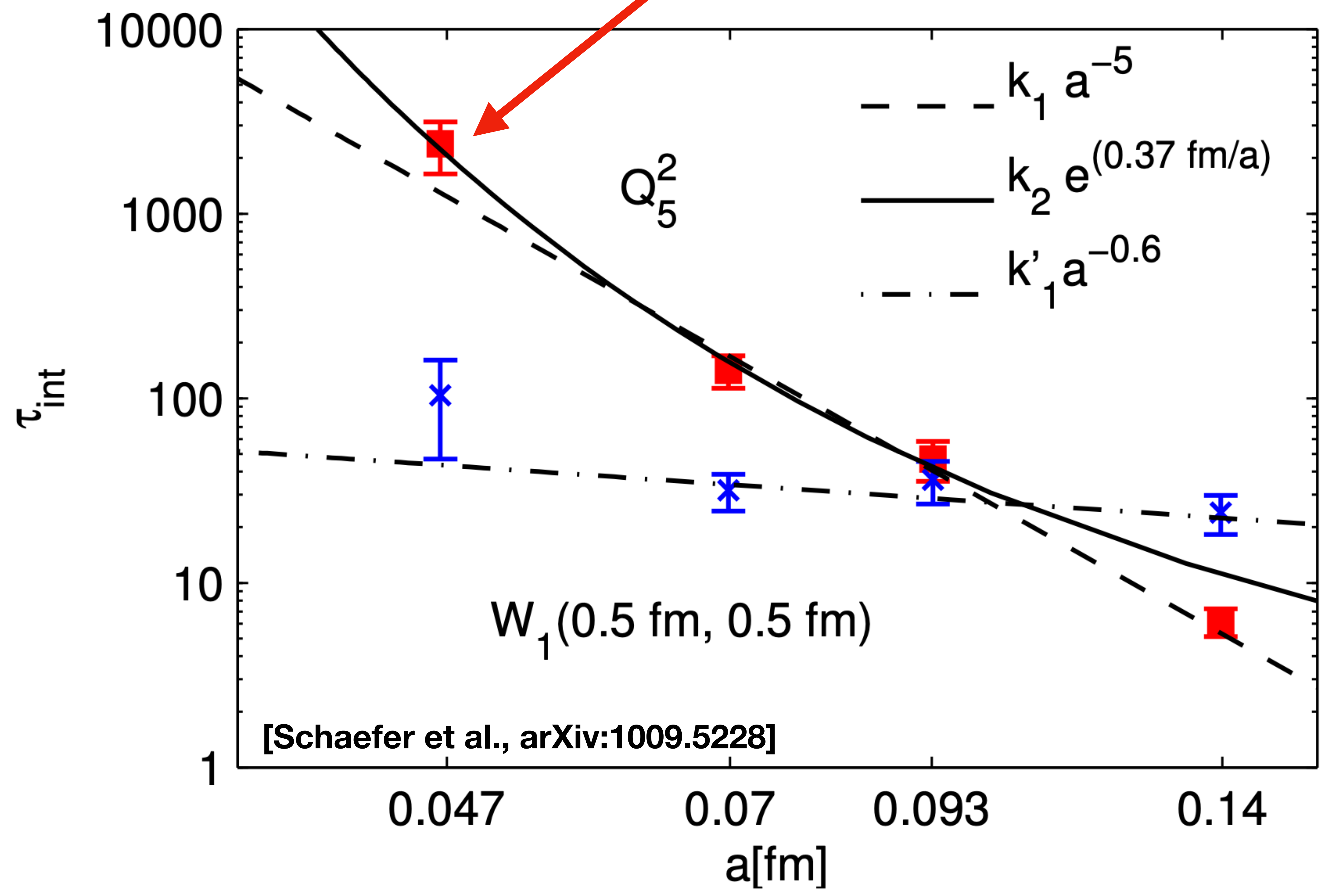
Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions

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

Introduction

- Critical slowing down is a well-know obstacle to extend the reach of state-of-the-art lattice QCD calculations
- Flow models are a promising tool, and have been able to overcome critical slowing down in some 2D theories
  [Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan 2003.06413]
  [Finkenrath 2201.02216]
  [Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Urban 2202.11712]
- Several technical challenges remain for at-scale QCD
  - Compare: HMC has been optimized for the last ~30 years!

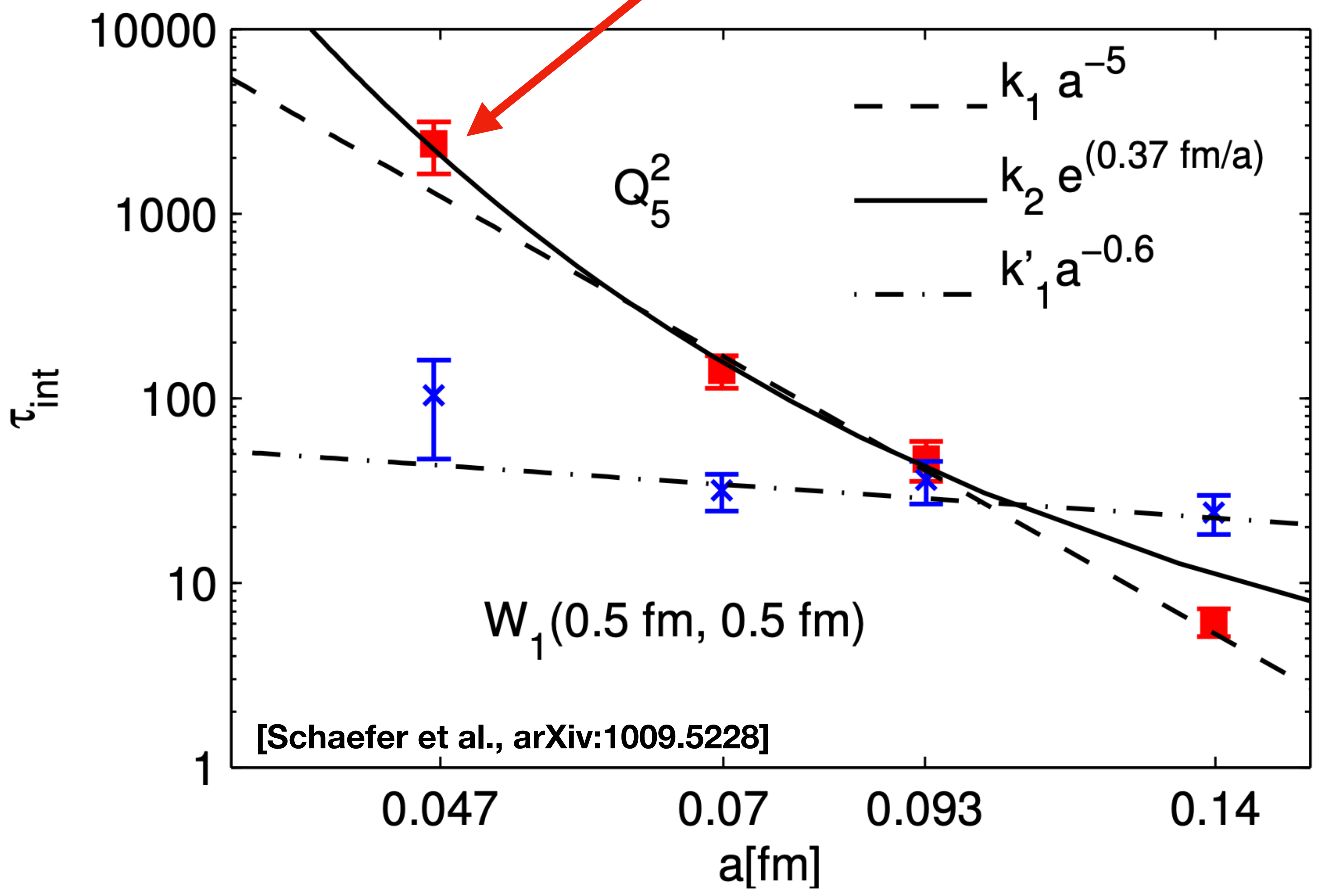
autocorrelation time of the topological charge



Introduction

- Critical slowing down is a well-know obstacle to extend the reach of state-of-the-art lattice QCD calculations
- Flow models are a promising tool, and have been able to overcome critical slowing down in some 2D theories
  [Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan 2003.06413]
  [Finkenrath 2201.02216]
  [Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Urban 2202.11712]
- Several technical challenges remain for at-scale QCD
  - Compare: HMC has been optimized for the last ~30 years!

autocorrelation time of the topological charge



This talk:

Sampling in gauge theories using <u>pseudofermions</u>

Generative flow models

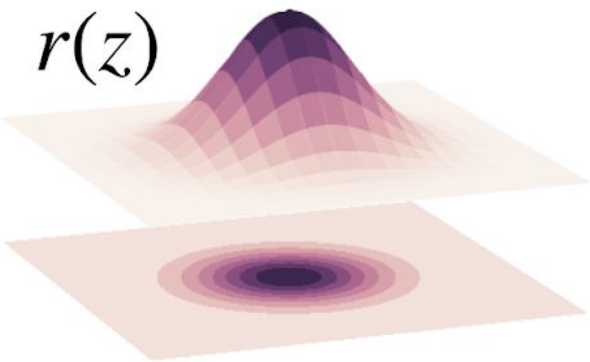
[Rezende, Mohamed, arXiv:1505.05770]

Invertible & tractable Jacobian

$$f(z)$$

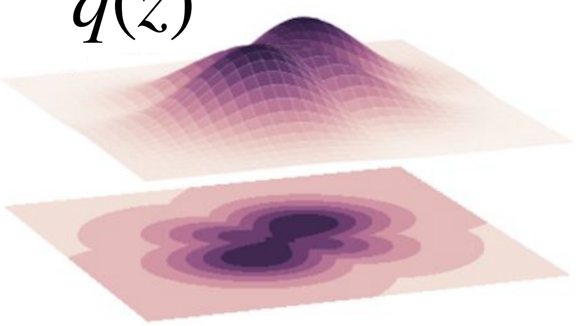
$z \longrightarrow \phi$

$r(z)$



easily sampled distribution (prior)

$q(z)$



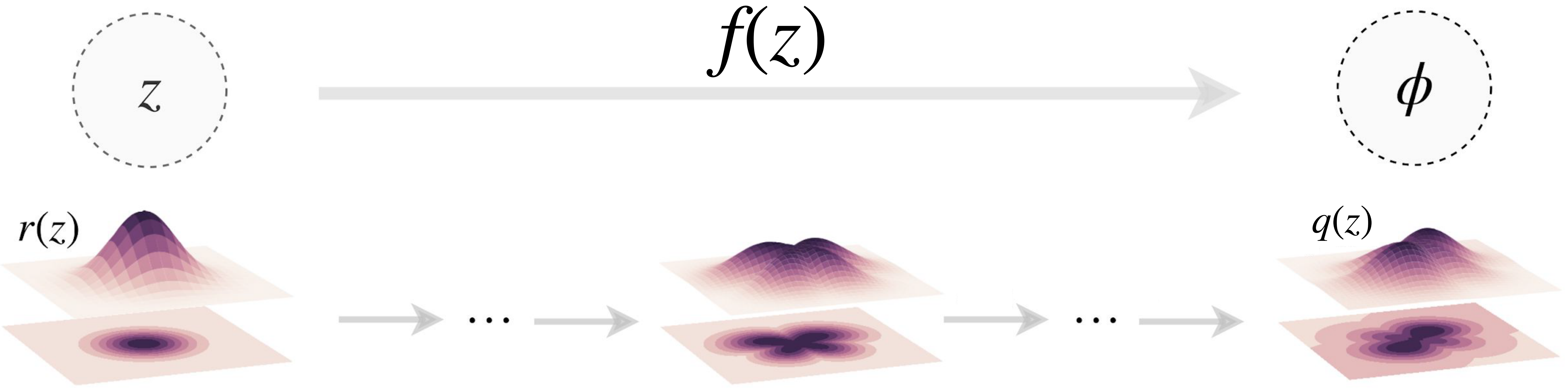
approximates target distribution (model)

$$q \simeq p = \frac{1}{\mathcal{Z}} e^{-S}$$

Generative flow models

[Rezende, Mohamed, arXiv:1505.05770]

Invertible & tractable Jacobian



easily sampled distribution (prior)

approximates target distribution (model)

$$q \simeq p = \frac{1}{\mathcal{Z}} e^{-S}$$

$f(z)$ is composed of many simple layers

parametrized by neural networks (trainable and expressive)

Model probability

$$\log q(z) = \log r(z) + \sum_i \log \det J_i$$

Fernando Romero-López, MIT

Training & using flows

- Train to minimize e.g. Kullback-Leibler divergence:

$$D_{\mathrm{KL}}(q||p) = \int d\phi \, q(\phi)[\log q(U) + S(\phi)] + (\mathrm{const})$$

- Self-training: 1. Draw samples from the model to measure sample mean of $[\log q(U) + S(\phi)]$

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

  [Kingma, Ba, arXiv:1412.6980]

Training & using flows

- Train to minimize e.g. Kullback-Leibler divergence:

$$D_{\mathrm{KL}}(q\|p) = \int d\phi \, q(\phi)[\log q(U) + S(\phi)] + (\mathrm{const})$$

- Self-training: 1. Draw samples from the model to measure sample mean of $[\log q(U) + S(\phi)]$

  2. Gradient-based methods to optimize model parameters (e.g. Adam optimizer)
  [Kingma, Ba, arXiv:1412.6980]

! Trained models are not perfect but exactness is essential.

Recover exactness by e.g. forming a Markov Chain with accept/reject steps

Flows for fermonic gauge theories

- **Several examples of flow models for pure gauge theories**

  [Kanwar et al, 2003.06413] [Boyda et al, 2008.05456]

- **For fermions: straightforward approach consists on integrating them out**

$$S_E(U) = -\beta \sum \underbrace{\mathrm{Re}\, P(x)}_{\text{(Plaquette)}} - \underbrace{\log \det D[U]^\dagger D[U]}_{\text{(fermion determinant)}}$$

Fernando Romero-López, MIT

Flows for fermonic gauge theories

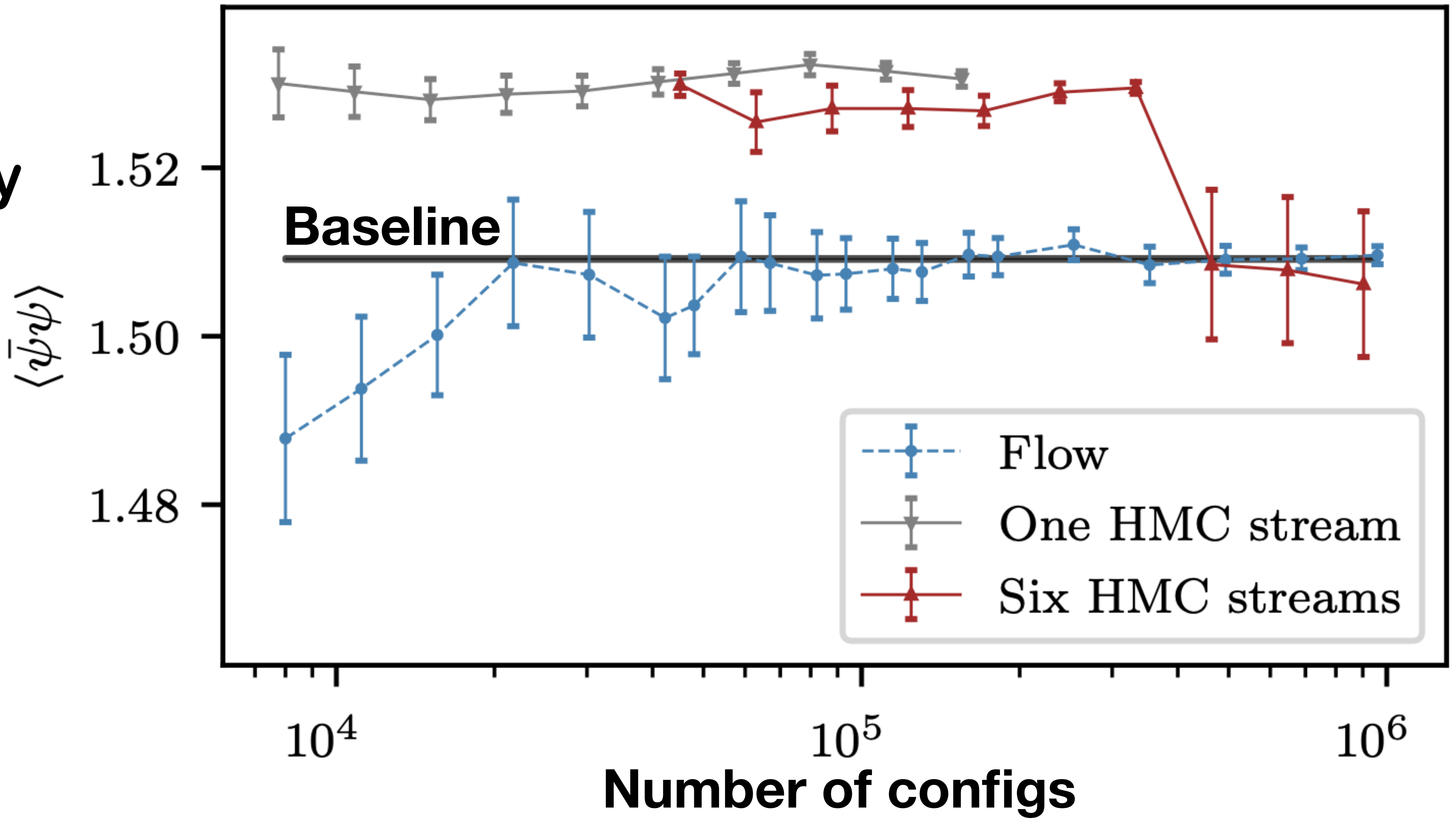
- Several examples of flow models for pure gauge theories
  [Kanwar et al, 2003.06413] [Boyda et al, 2008.05456]

- For fermions: straightforward approach consists on integrating them out

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

(Plaquette) (fermion determinant)

Chiral condensate



- Successfully applied to the Schwinger model at criticality
  [Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Urban 2202.11712]

✗ HMC shows biased results with underestimated errors

✓ Flow-based sampling provides correct results

Flows for fermonic gauge theories

- Several examples of flow models for pure gauge theories
  [Kanwar et al, 2003.06413]   [Boyda et al, 2008.05456]

- For fermions: straightforward approach consists on integrating them out

$$S_E(U) = -\beta \underbrace{\sum \mathrm{Re}\, P(x)}_{\text{(Plaquette)}} - \underbrace{\log \det D[U]^\dagger D[U]}_{\text{(fermion determinant)}}$$

- Successfully applied to the Schwinger model at criticality
  [Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Urban 2202.11712]

✗ HMC shows biased results with underestimated errors

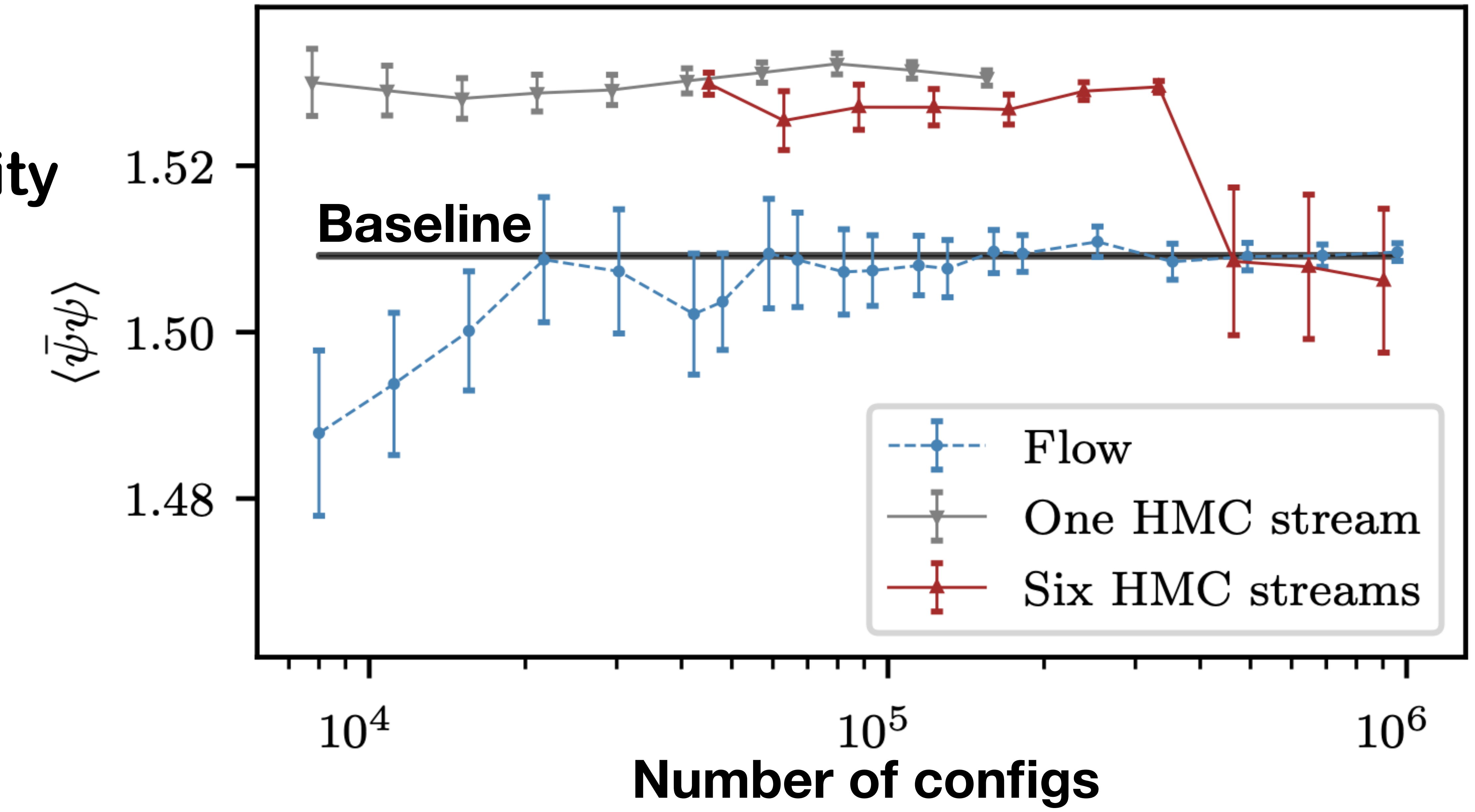
✓ Flow-based sampling provides correct results

! Evaluation of the fermion determinant is expensive.
Not feasible for QCD-scale calculations!

- Scalable approach: use stochastic determinant estimators!

→ Pseudofermions!

Chiral condensate



Fernando Romero-López, MIT

Recap of Pseudofermions

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

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

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

- **Joint target distribution**

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

- **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]

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"

$f_m(\chi)$

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

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

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"

$f_m(\chi)$

existing gauge architectures

[Kanwar et al, 2003.06413]

[Boyda et al, 2008.05456]

$f_c(z \mid U)$

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

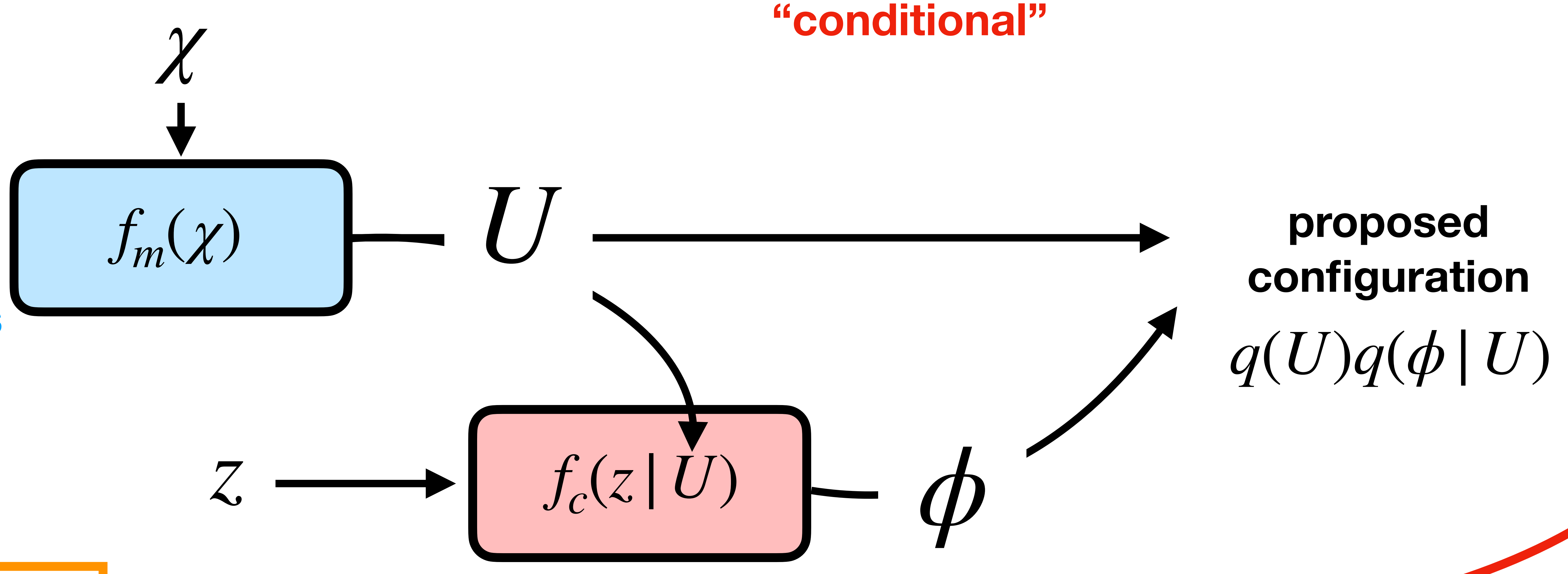
$$p(U,\phi) = p(U)p(\phi \mid U)$$

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

"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]

new conditional architectures

Different flow models to approximate marginal and conditional distributions

[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:

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

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

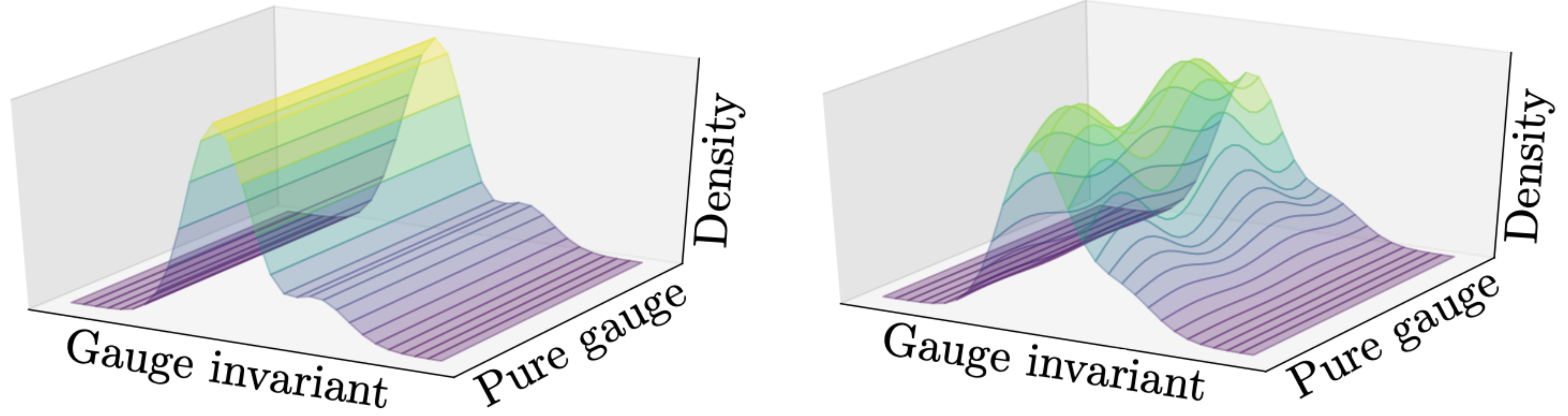
Gauge-equivariant conditional models

- 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

- Gauge equivariance simplifies the problem:



True distribution

Learned by naive ML

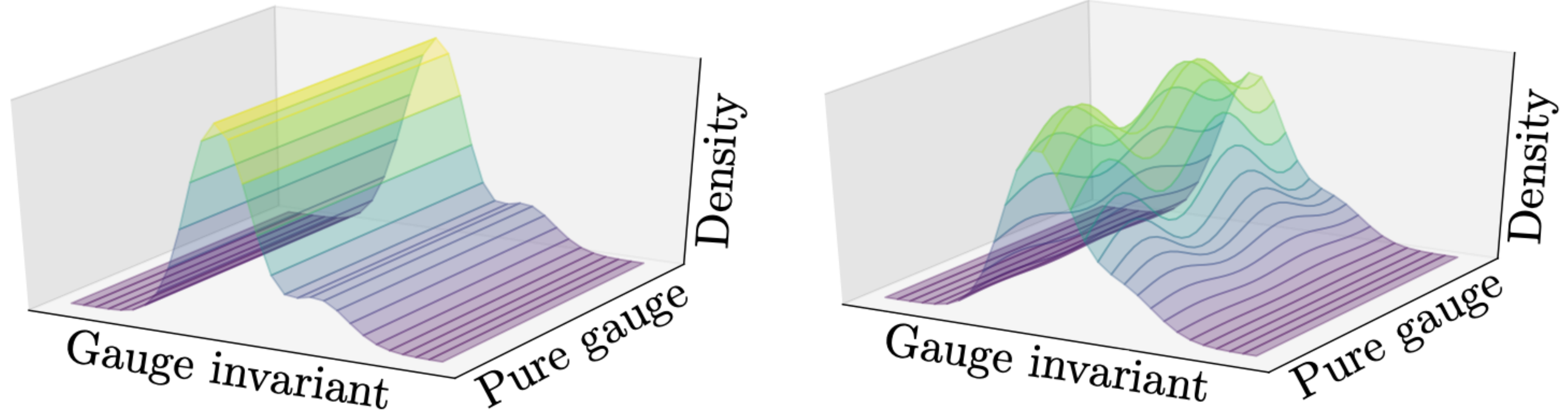
Gauge-equivariant conditional models

- 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

- Gauge equivariance simplifies the problem:



True distribution   Learned by naive ML

- 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

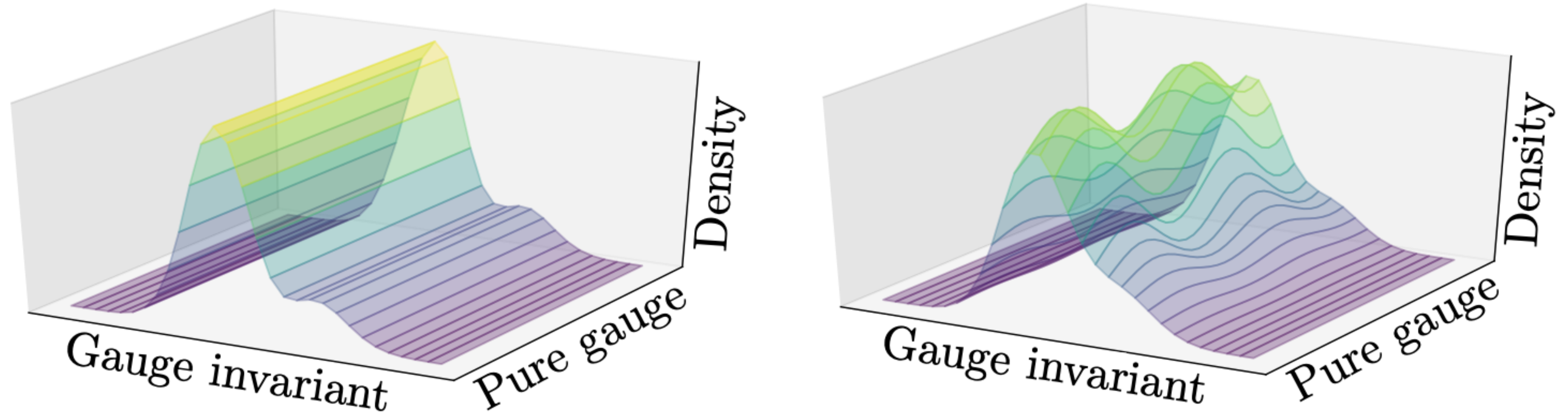
Gauge-equivariant conditional models

- 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

- Gauge equivariance simplifies the problem:



True distribution　　Learned by naive ML

- 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

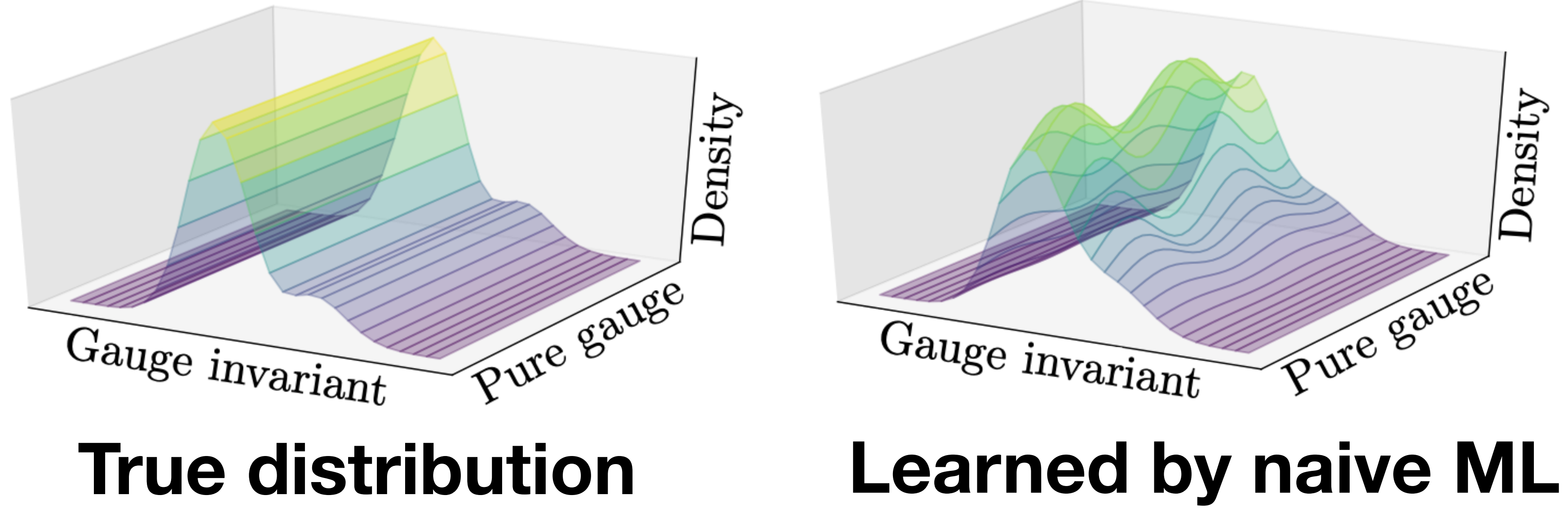
Gauge-equivariant conditional models

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

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

approximates

- Gauge equivariance simplifies the problem:



- 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

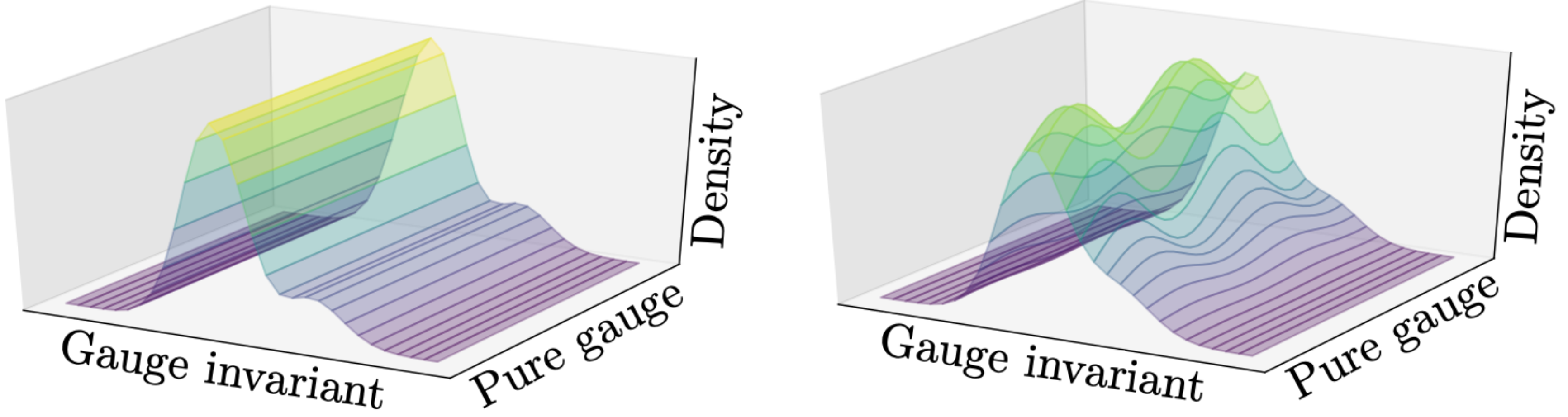
Gauge-equivariant conditional models

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

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

approximates

- Gauge equivariance simplifies the problem:



True distribution **Learned by naive ML**

- 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

Sketch of pseudofermion layers

"active"

ϕ_a

$\phi \to \phi_a$

$\phi \to \phi_f$

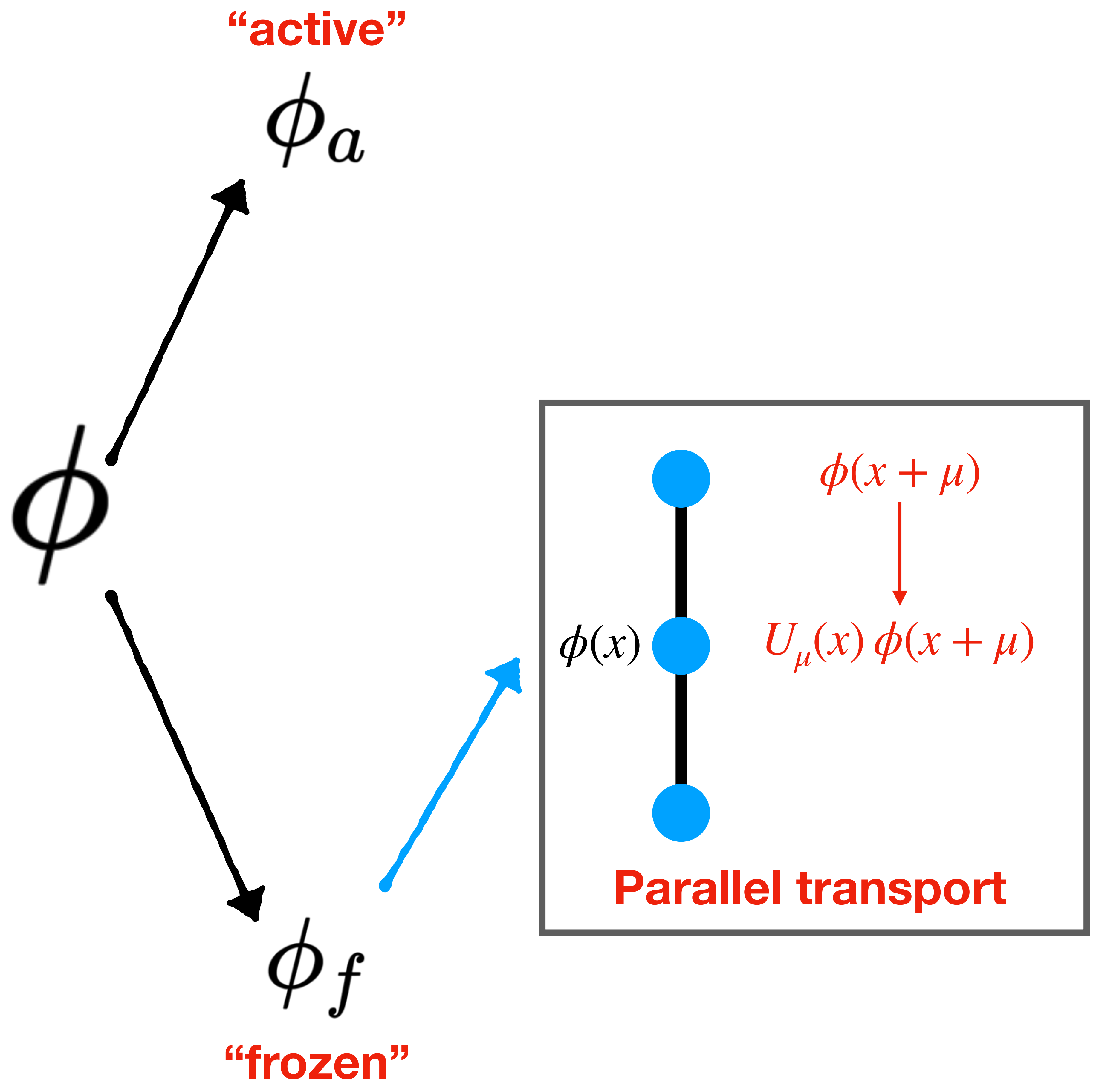
ϕ_f

"frozen"

variable partitioning
ensures tractible Jacobian

Fernando Romero-López, MIT

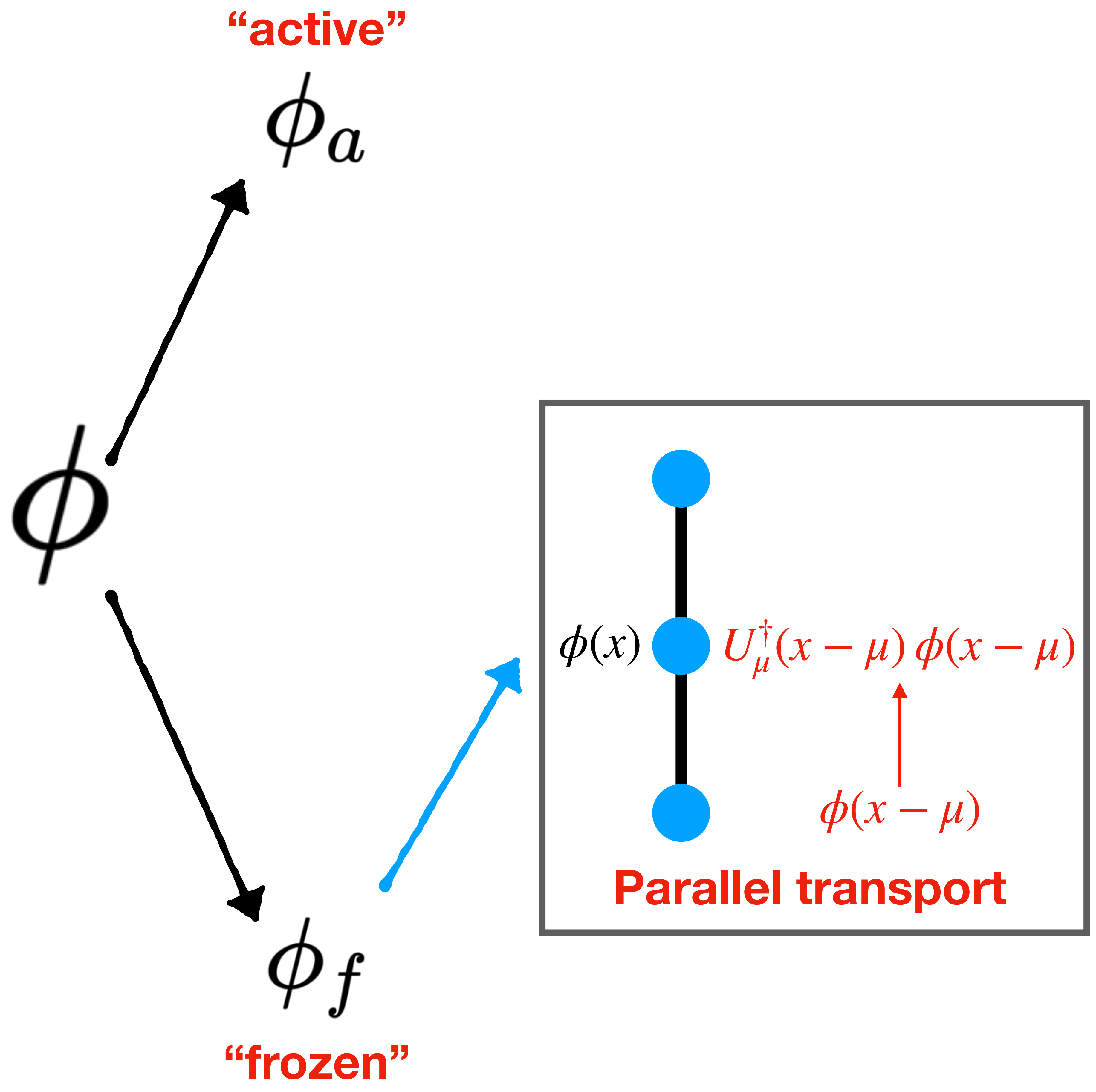
Sketch of pseudofermion layers



variable partitioning
ensures tractible Jacobian

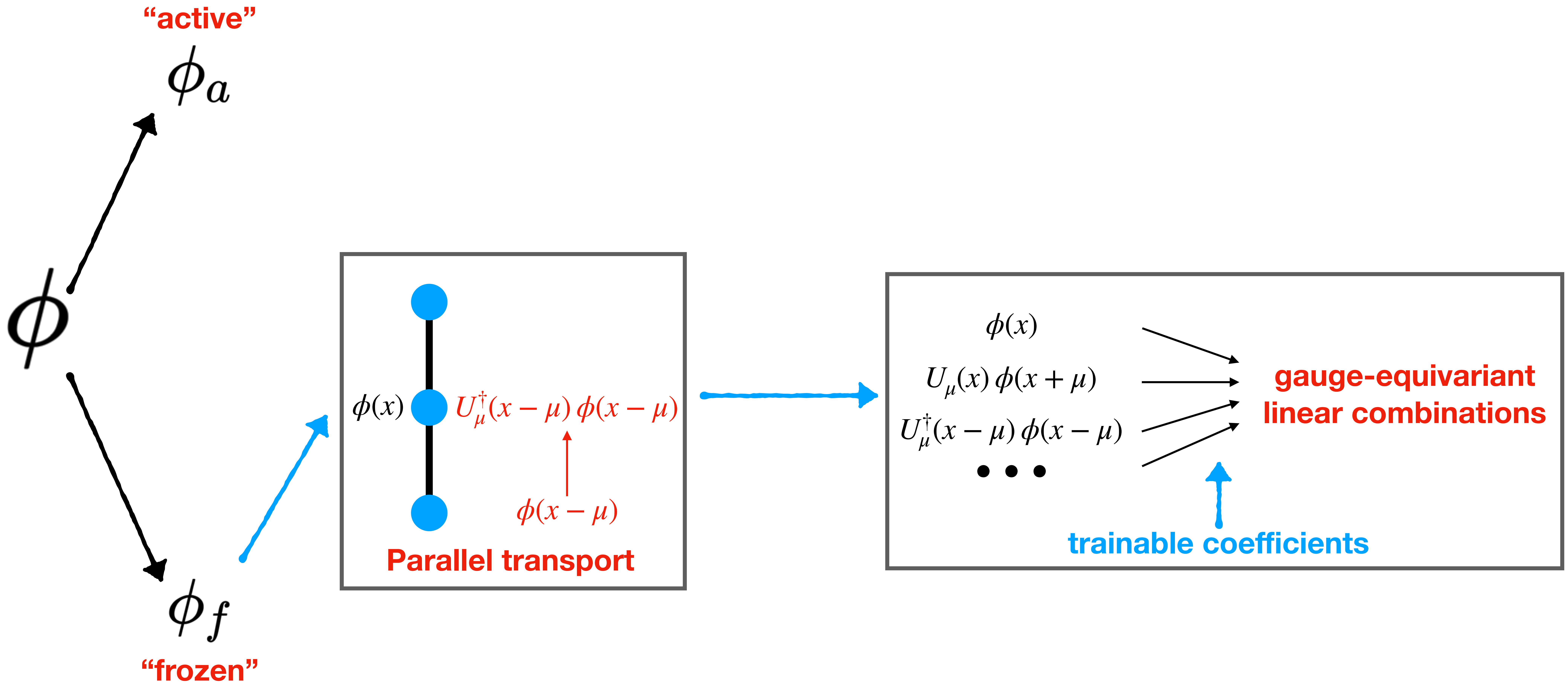
Fernando Romero-López, MIT

Sketch of pseudofermion layers



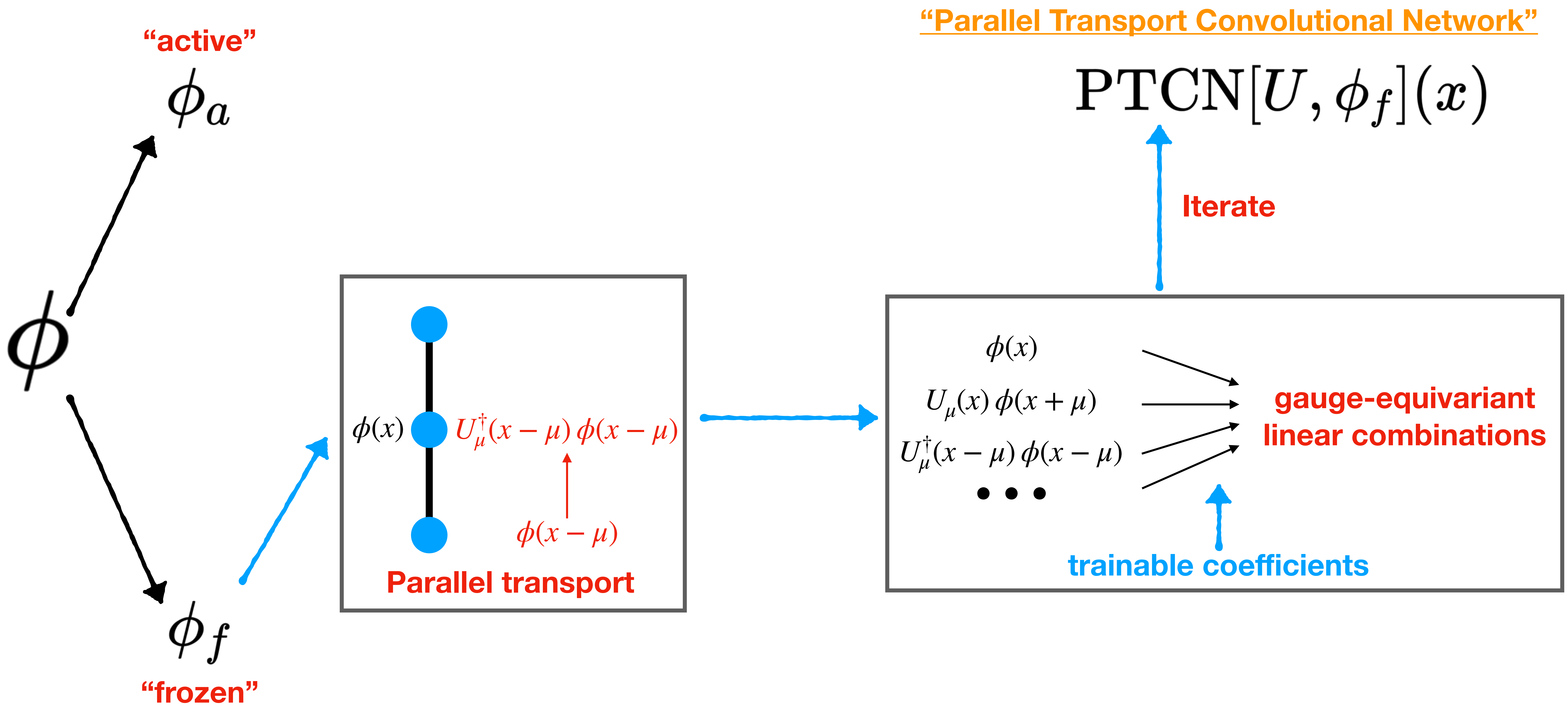
variable partitioning
ensures tractible Jacobian

Sketch of pseudofermion layers



variable partitioning
ensures tractible Jacobian

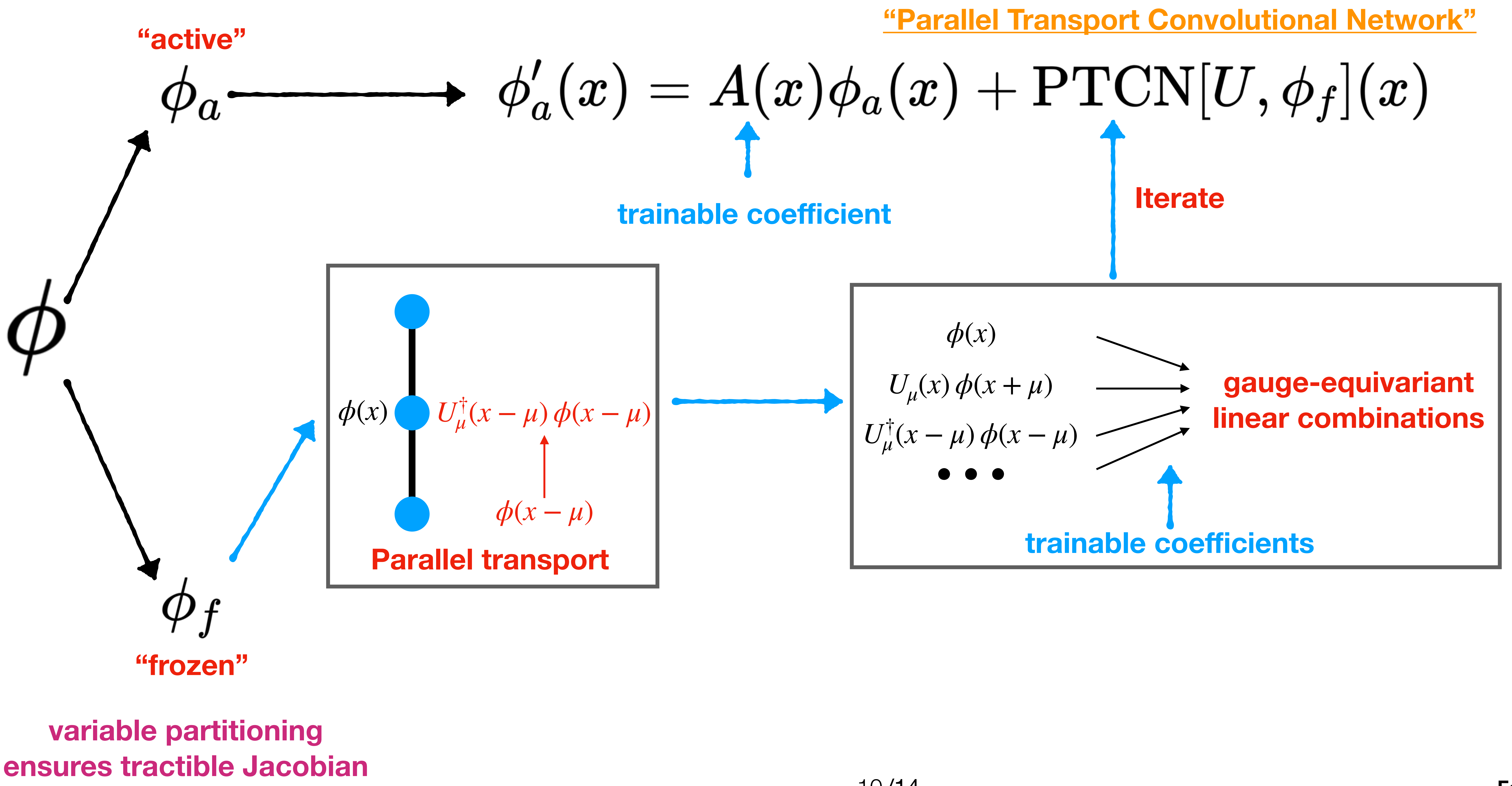
Sketch of pseudofermion layers



variable partitioning
ensures tractible Jacobian

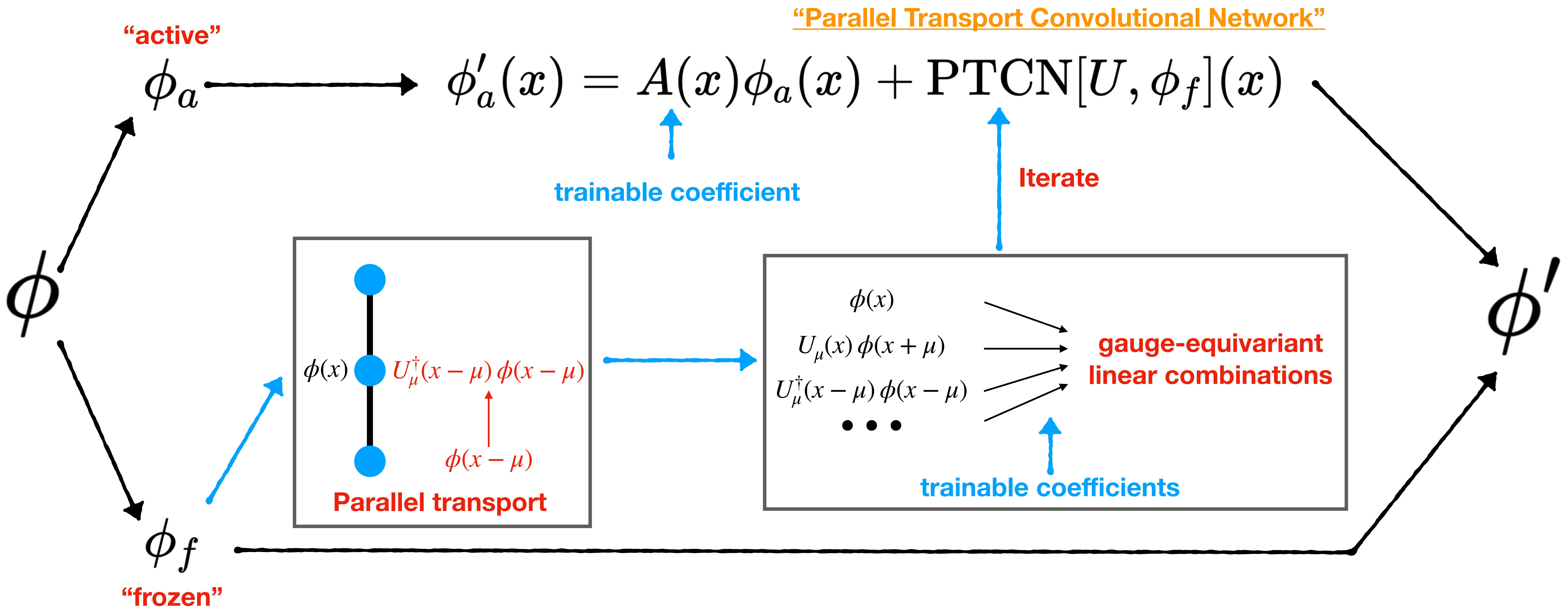
Fernando Romero-López, MIT

Sketch of pseudofermion layers



Fernando Romero-López, MIT

Sketch of pseudofermion layers

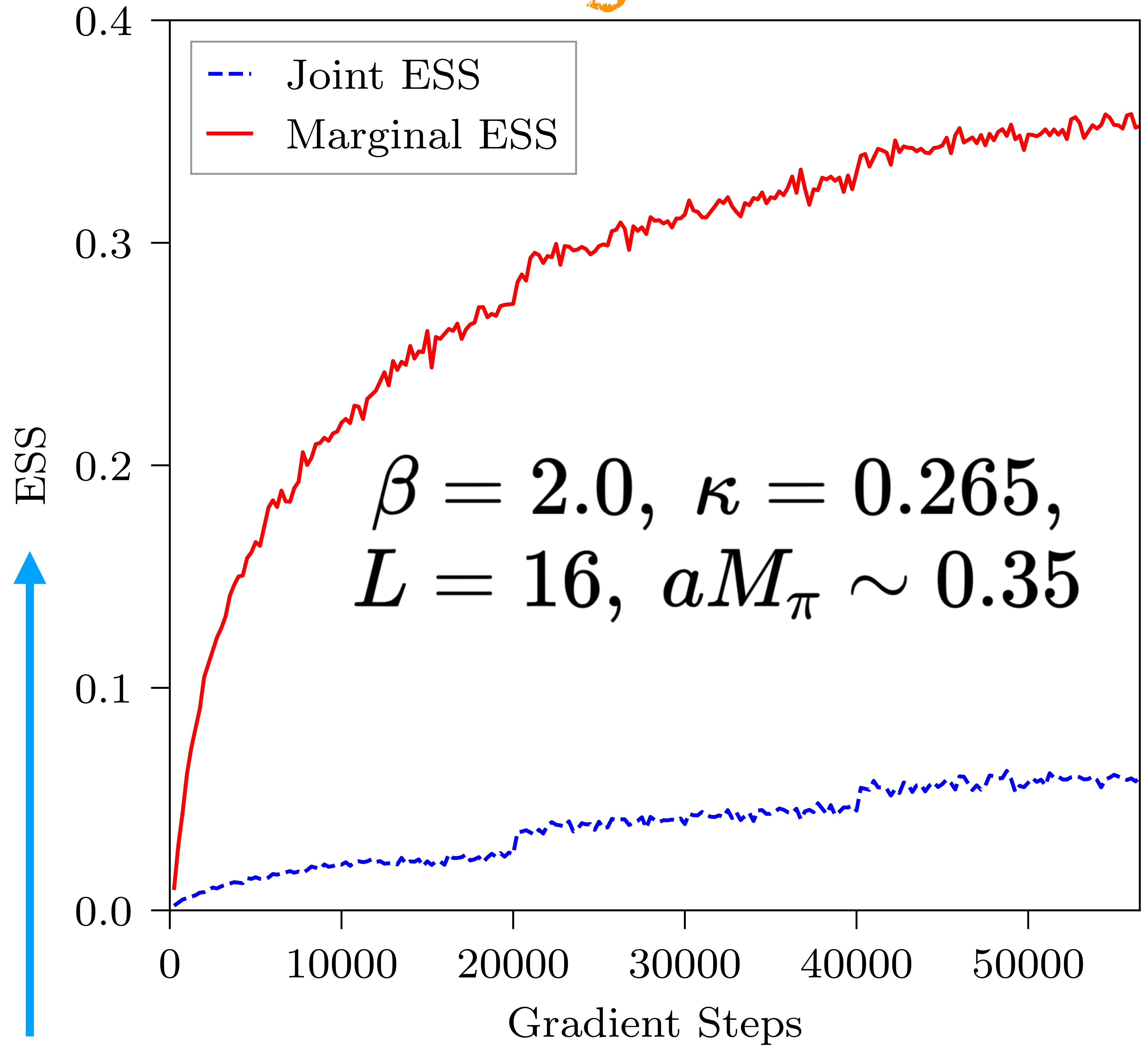


variable partitioning
ensures tractible Jacobian

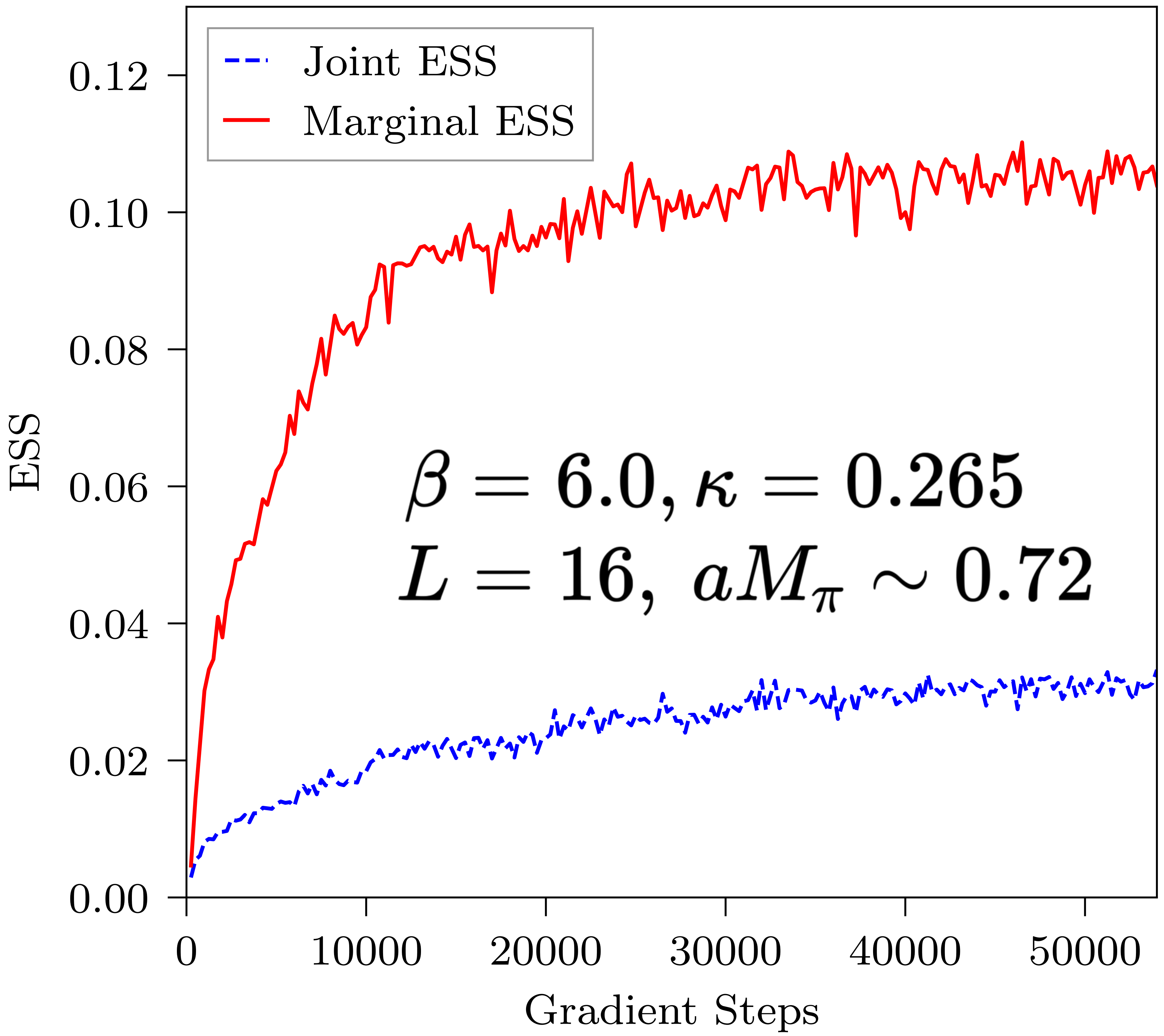
Fernando Romero-López, MIT

Examples of flow models

Schwinger Model



2D SU(3) with Nf=2



"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]

Use of preconditioners

- Use joint models in combination with preconditioners
  [Abbott, Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Tian, Urban 2207.08945]
  - Even/odd (EO) preconditioning

$$D = \begin{pmatrix} \mathbb{1} & D_{eo} \\ D_{oe} & \mathbb{1} \end{pmatrix} \qquad \det D = \det\left(\mathbb{1} - D_{eo} D_{oe}\right) \equiv \det D_{\rm sc}$$

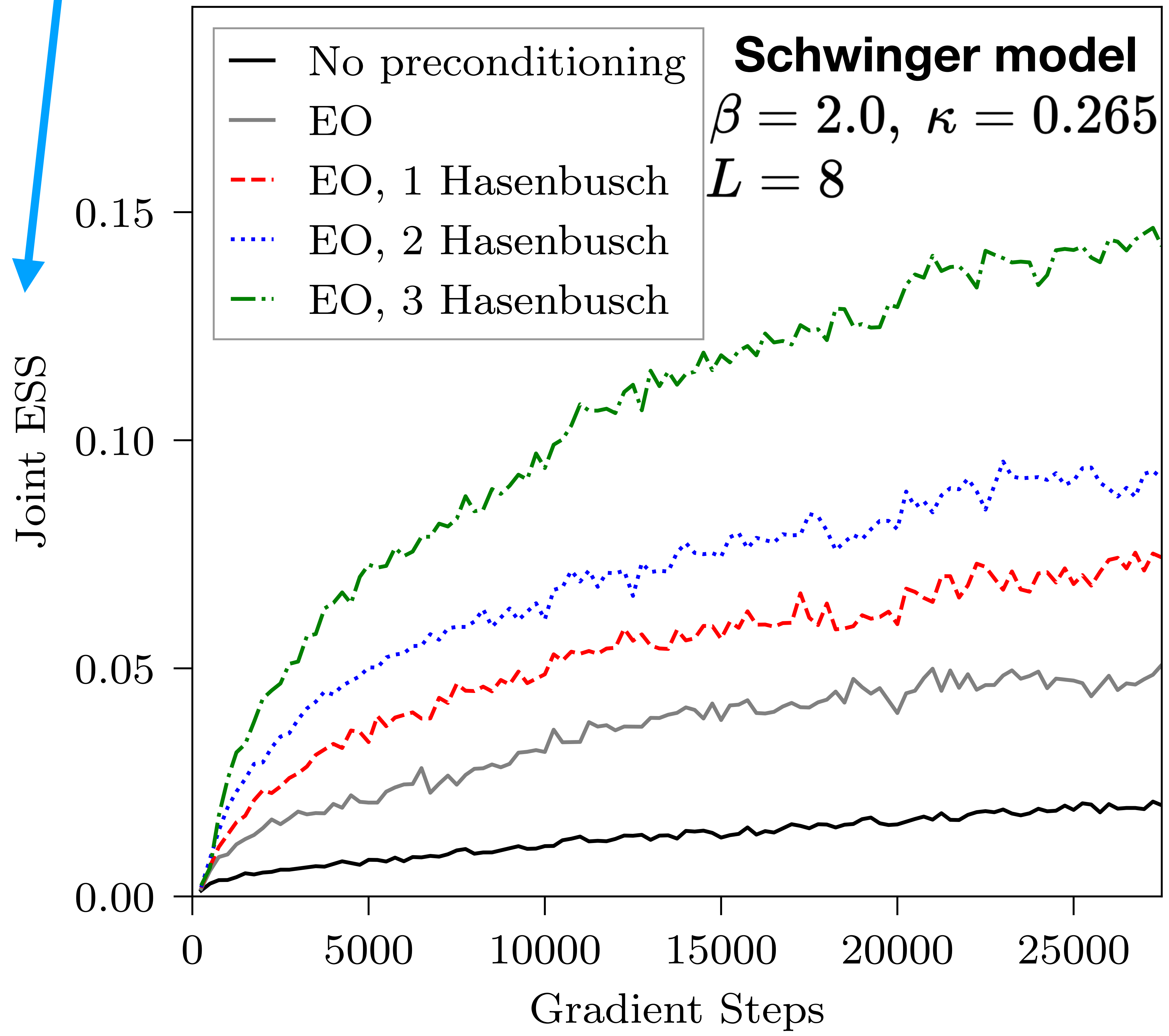
    - Less PF variables in target
  - Hasenbusch factorization

$$\det M = \left[\frac{\det M}{\det\left(M+\mu\right)}\right] \det\left(M+\mu\right)$$

    - Reduces the condition number of target

"Effective sample size" $\equiv$ model quality

Significant improvement by using preconditioners!



Fernando Romero-López, MIT

Improving determinant estimators

- Use joint models more efficiently by drawing multiple pseudofermion samples at fixed gauge field

$$w_{N_{\rm pf}}(U) = \frac{1}{N_{\rm pf}} \sum_{i=1}^{N_{\rm pf}} \frac{p\left(\phi^{(i)}, U\right)}{q\left(\phi^{(i)}, U\right)}$$

weight of each config for reweighting or Markov Chain

- This does not require reevaluating the observables

Fernando Romero-López, MIT

Improving determinant estimators

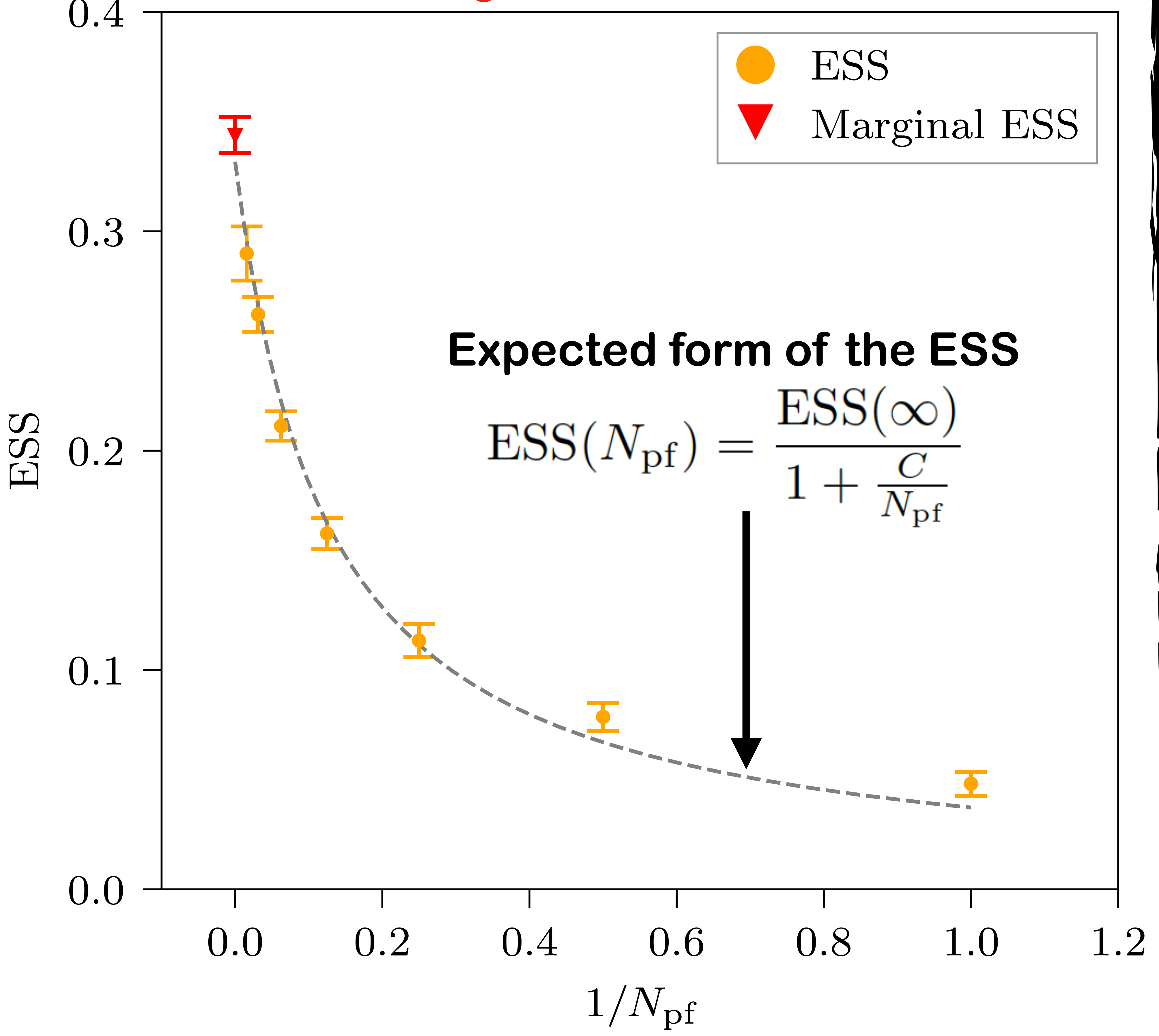
- Use joint models more efficiently by drawing multiple pseudofermion samples at fixed gauge field

$$w_{N_{\rm pf}}(U) = \frac{1}{N_{\rm pf}} \sum_{i=1}^{N_{\rm pf}} \frac{p\left(\phi^{(i)}, U\right)}{q\left(\phi^{(i)}, U\right)}$$

weight of each config for reweighting or Markov Chain

- This does not require reevaluating the observables

Numerical demonstration in the Schwinger model



Fernando Romero-López, MIT

Summary & Outlook

- Pseudofermions can be used in flows as a scalable approach to the fermion determinant

- We have developed architectures for flow models using pseudofermions in gauge theories
  - Parallel Transport Convolutional Network is the central element
    [Abbott, Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Tian, Urban 2207.08945]

- Successful proof of concept applications to 2D theories: Schwinger model and SU(3) in 2D
  - Improve model quality with preconditioners
  - Can use multiple pseudofermion draws to better estimate $\det DD^\dagger[U]$

- No fundamental limitation to apply this technology to QCD in 4D
  - Further engineering still needed for high quality models

  [See talk by Phiala Shanahan later in this session!]

Summary & Outlook

- Pseudofermions can be used in flows as a scalable approach to the fermion determinant

- We have developed architectures for flow models using pseudofermions in gauge theories
  - Parallel Transport Convolutional Network is the central element
    [Abbott, Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Tian, Urban 2207.08945]

- Successful proof of concept applications to 2D theories: Schwinger model and SU(3) in 2D
  - Improve model quality with preconditioners
  - Can use multiple pseudofermion draws to better estimate $\det DD^{\dagger}[U]$

- No fundamental limitation to apply this technology to QCD in 4D
  - Further engineering still needed for high quality models

  [See talk by Phiala Shanahan later in this session!]

Thanks!

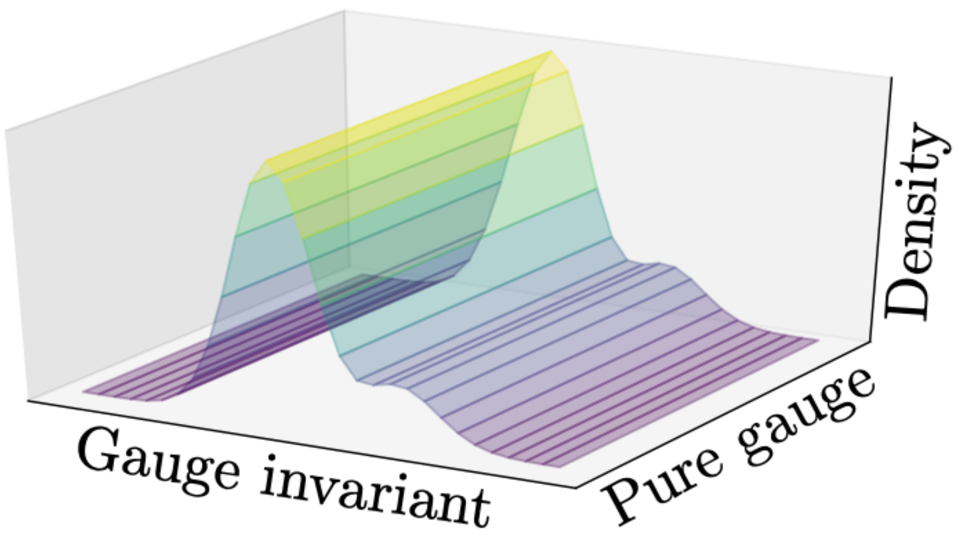
Fernando Romero-López, MIT

Back-up slides

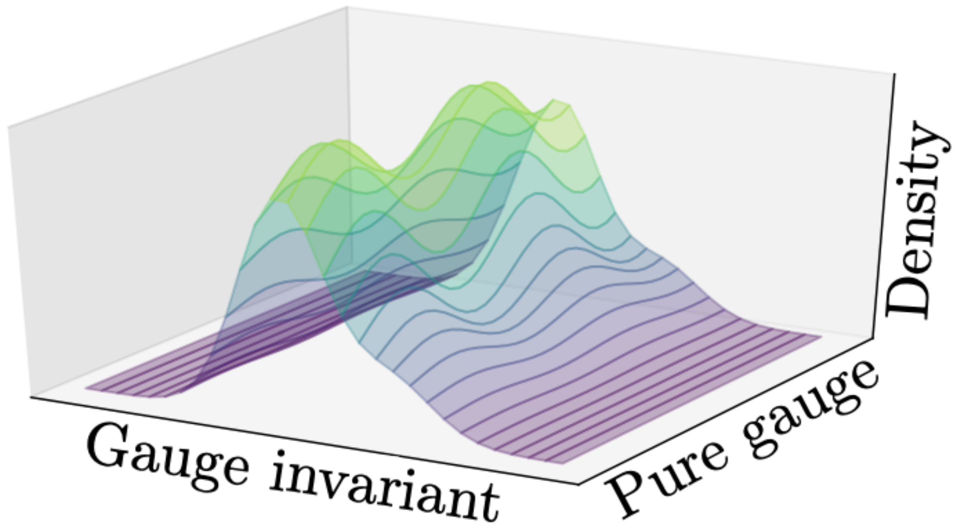
Incorporating symmetries

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

Gauge symmetries:



True distribution

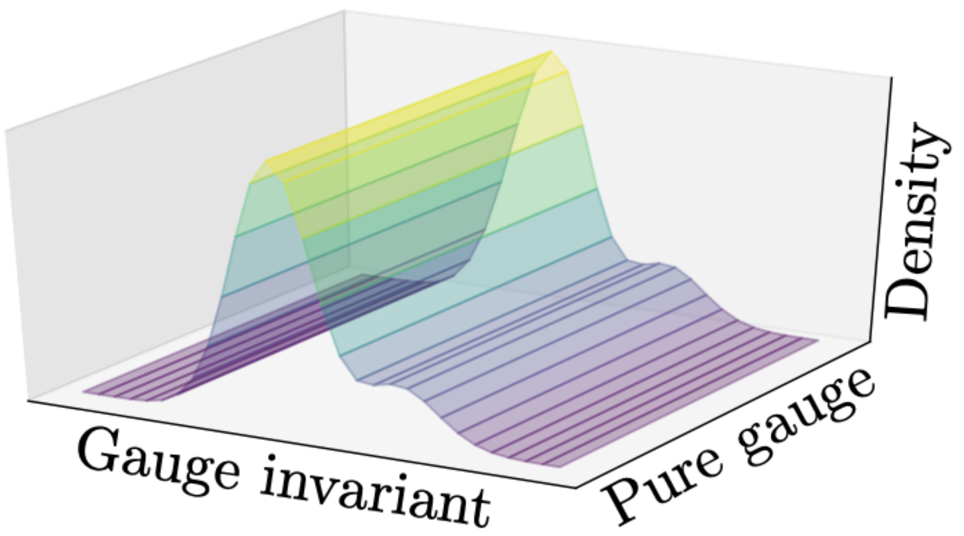


Learned by naive ML

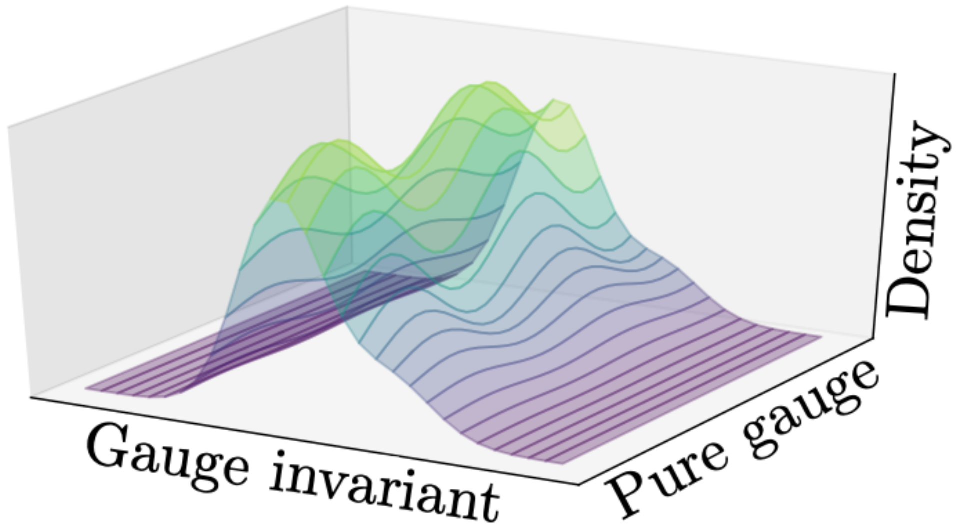
Incorporating symmetries

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

Gauge symmetries:



True distribution

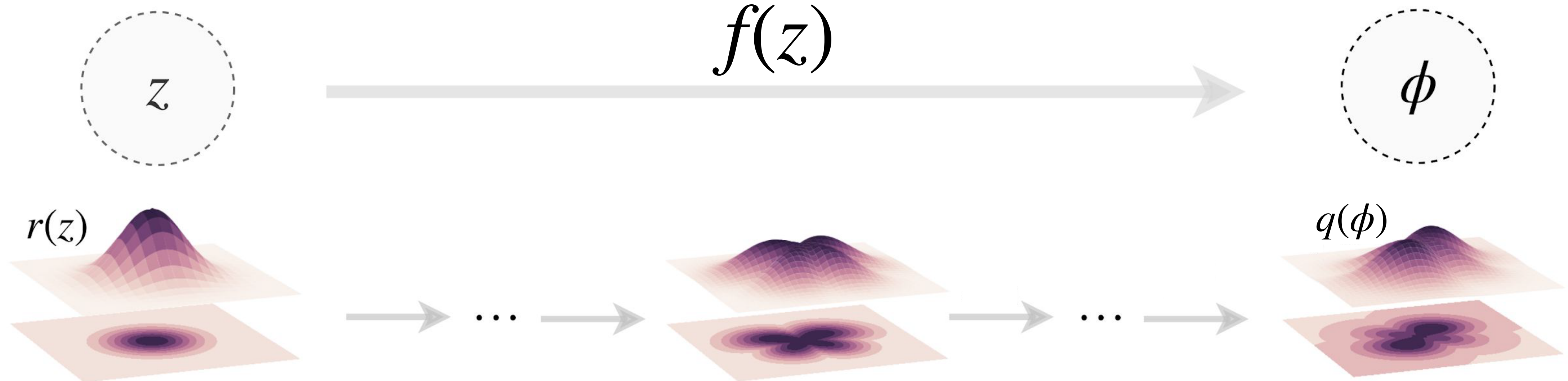


Learned by naive ML

General approach:

1. Invariant base distribution

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

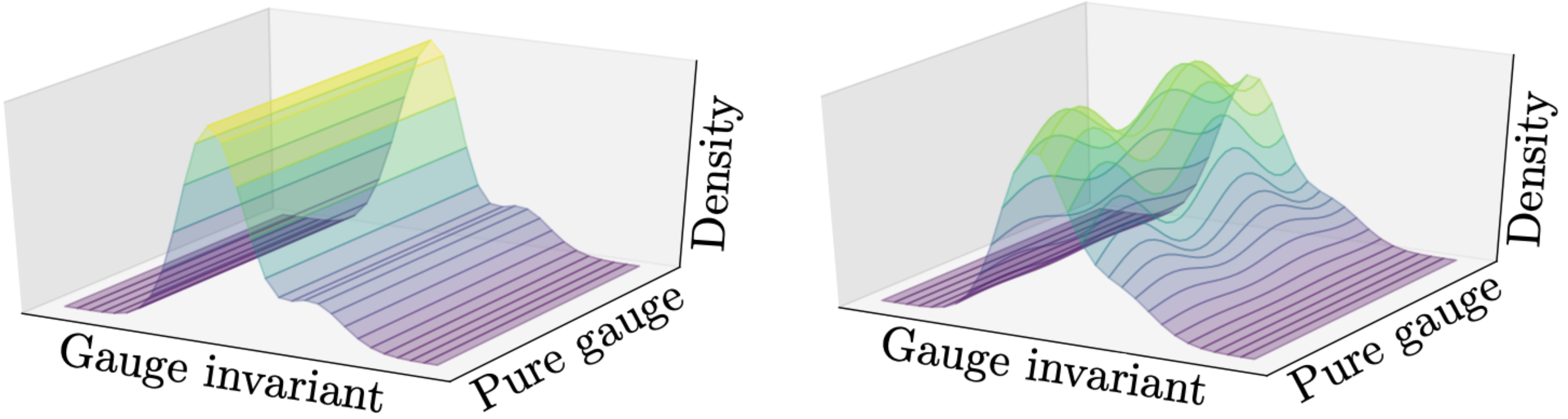


Fernando Romero-López, MIT

Incorporating symmetries

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

Gauge symmetries:



True distribution **Learned by naive ML**

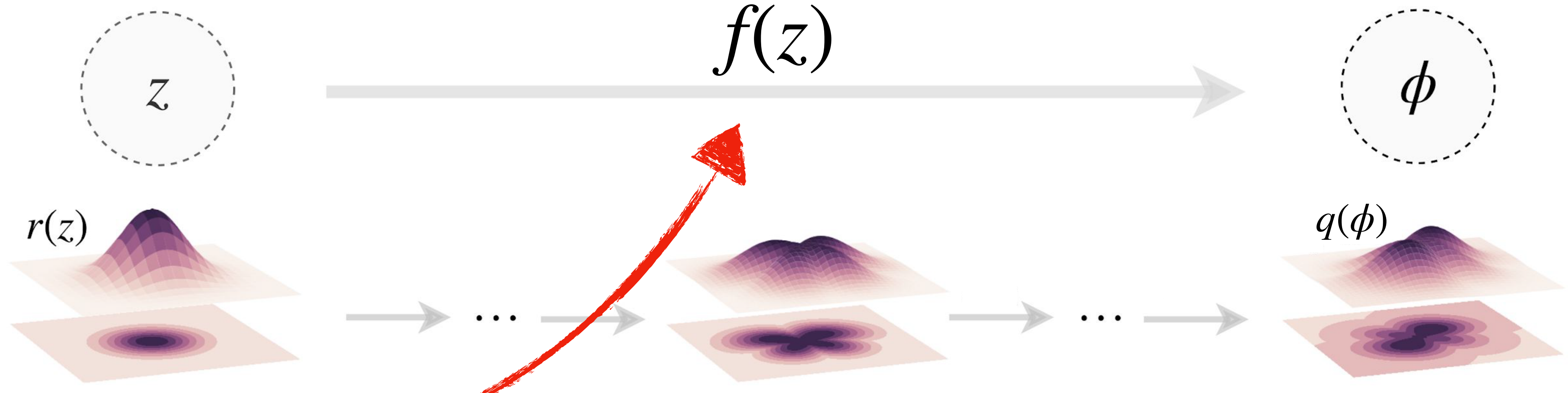
General approach:

1. Invariant base distribution

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

2. Equivariant flow

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



Fernando Romero-López, MIT

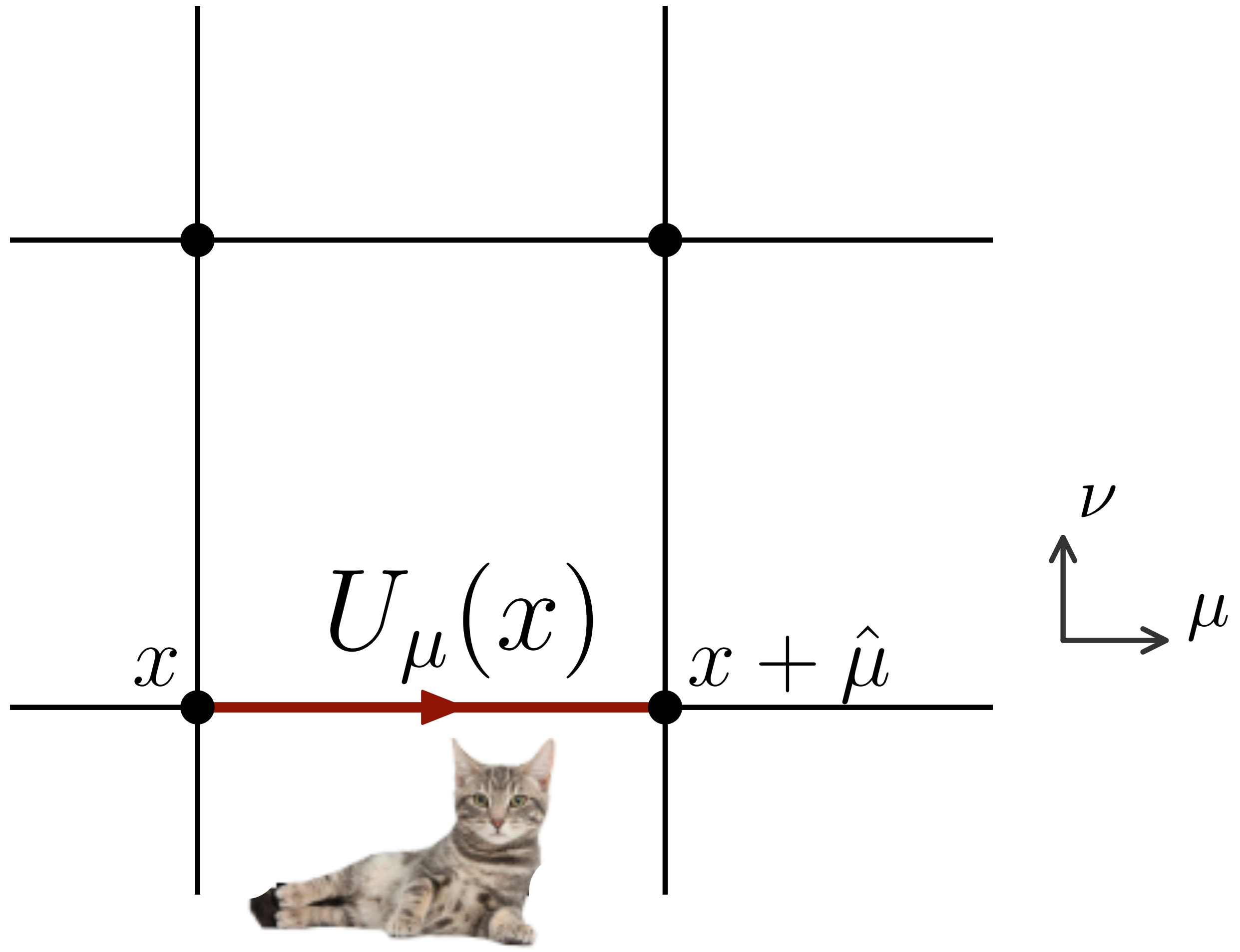
Lattice U(1) gauge symmetry

- **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

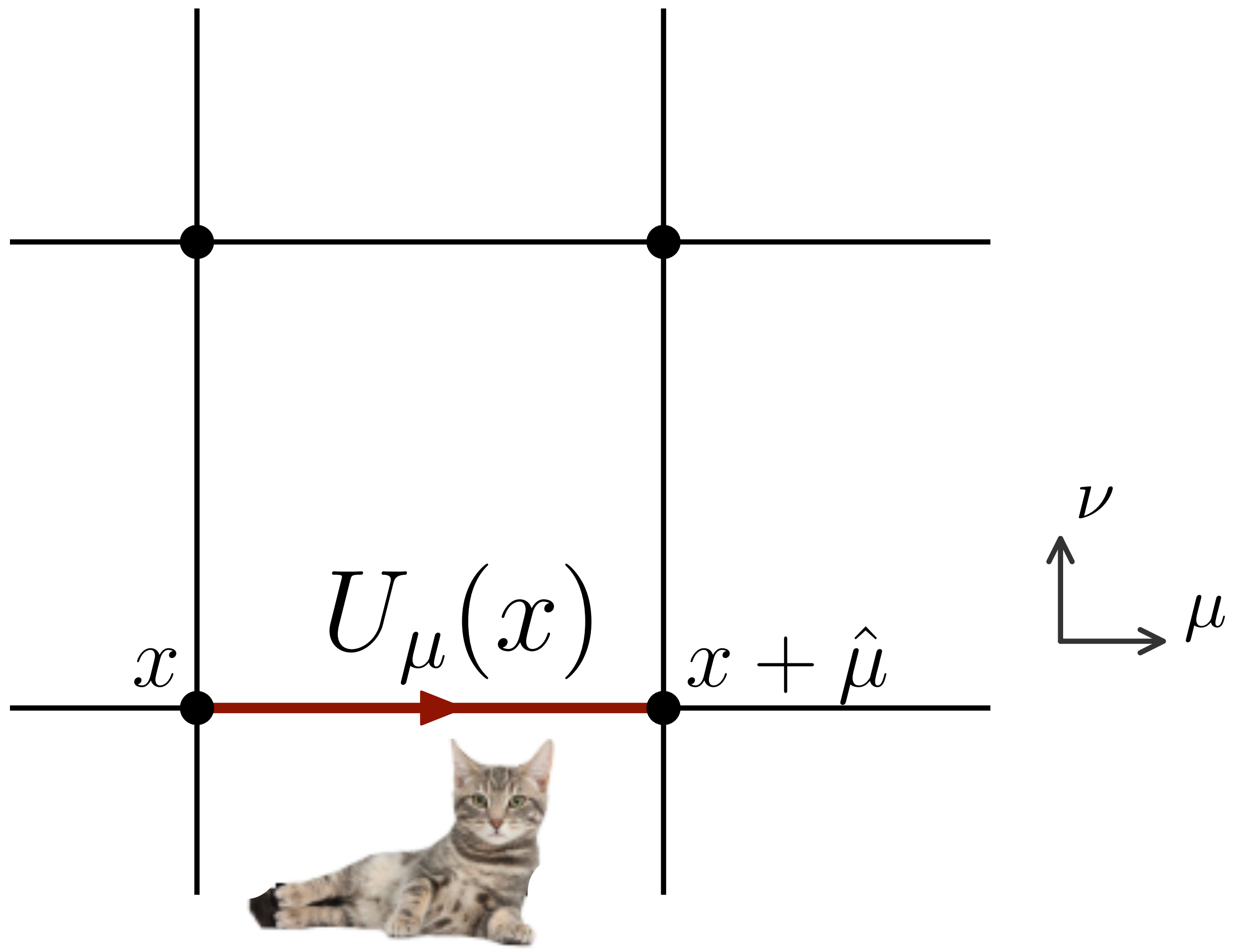
Lattice U(1) gauge symmetry

- Gauge variables are the gauge links

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

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

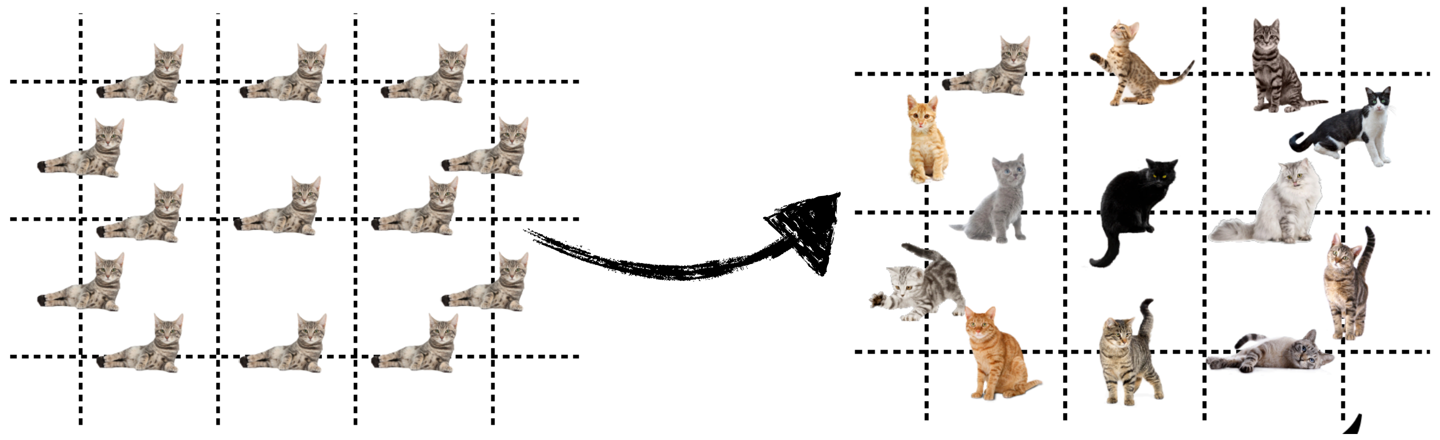
$$agA_\mu(x) \in [0, 2\pi)$$



- 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

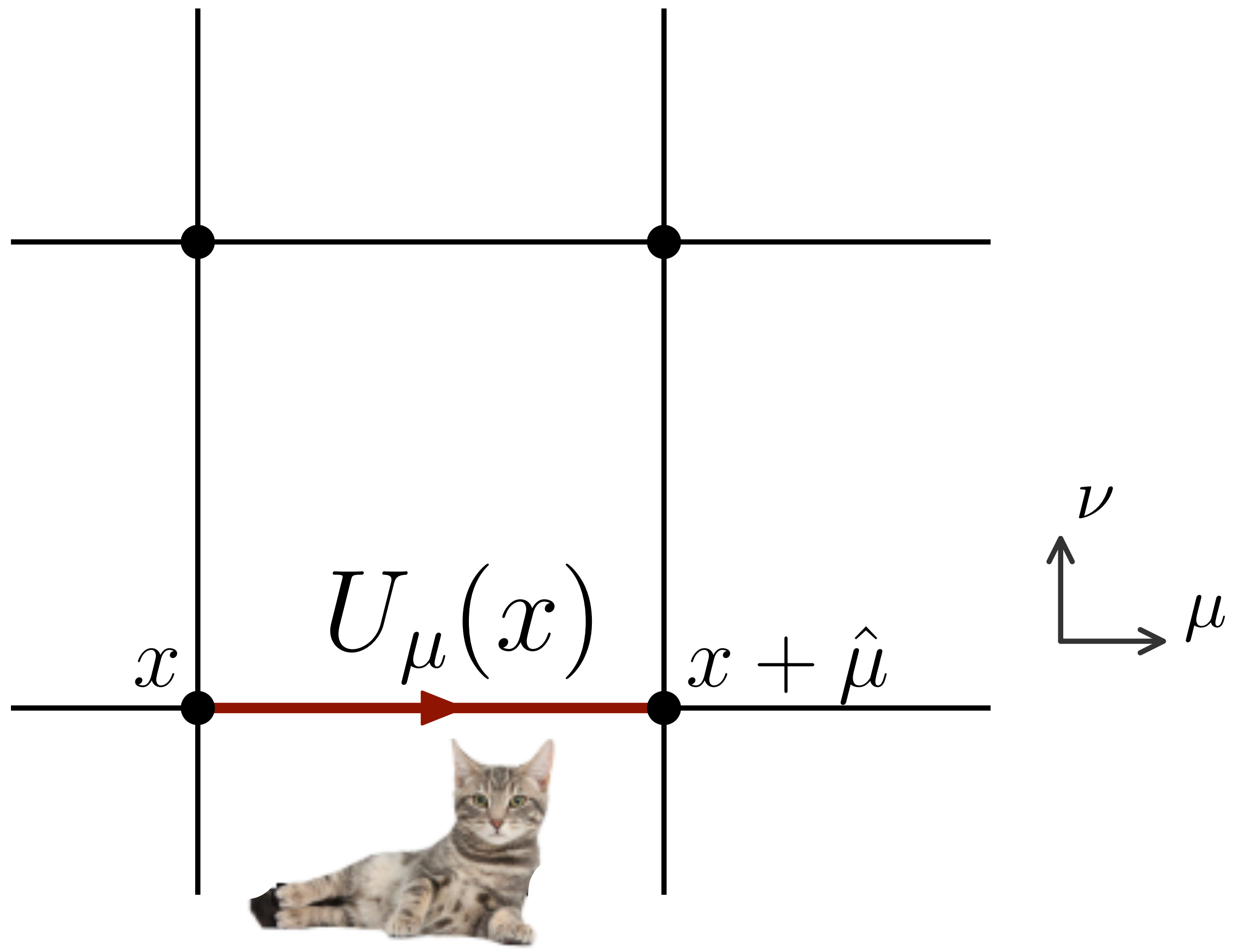
Lattice U(1) gauge symmetry

- **Gauge variables are the gauge links**

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

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

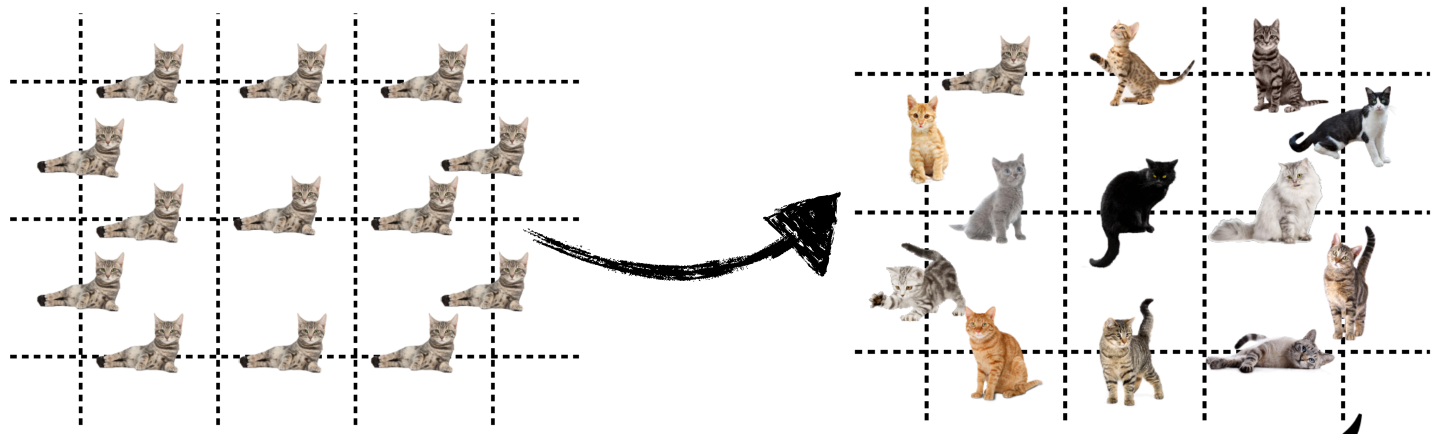
$$agA_\mu(x) \in [0, 2\pi)$$



- **A gauge transformation is:**

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

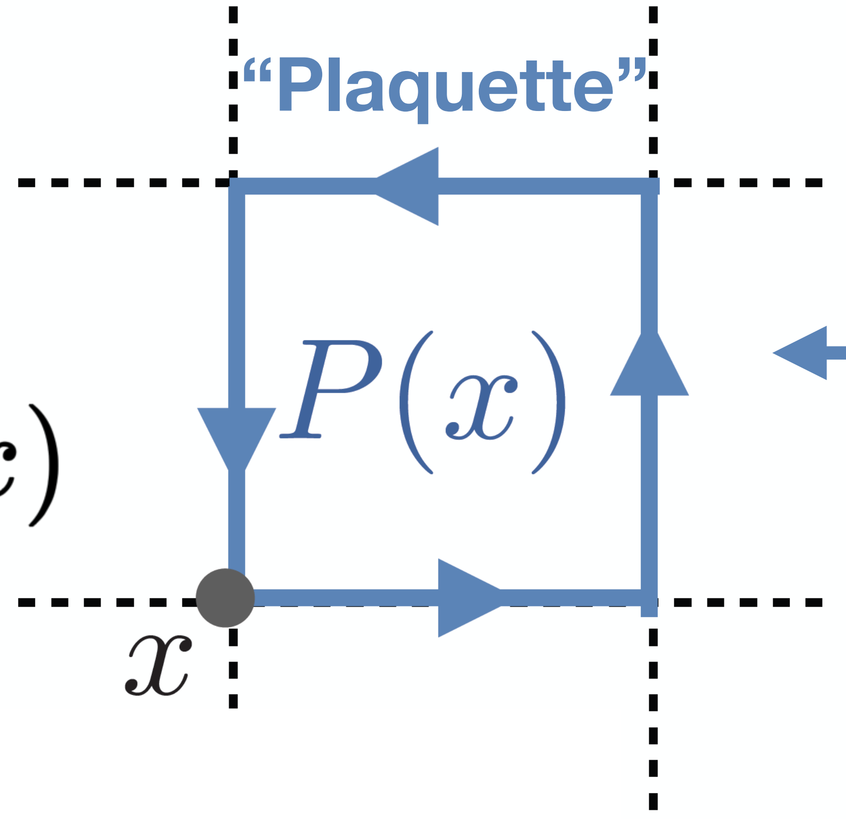
$$\text{for all } \; \Omega(x) \in \mathrm{U}(1)$$



- **Pure gauge action:**

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

inverse
gauge coupling

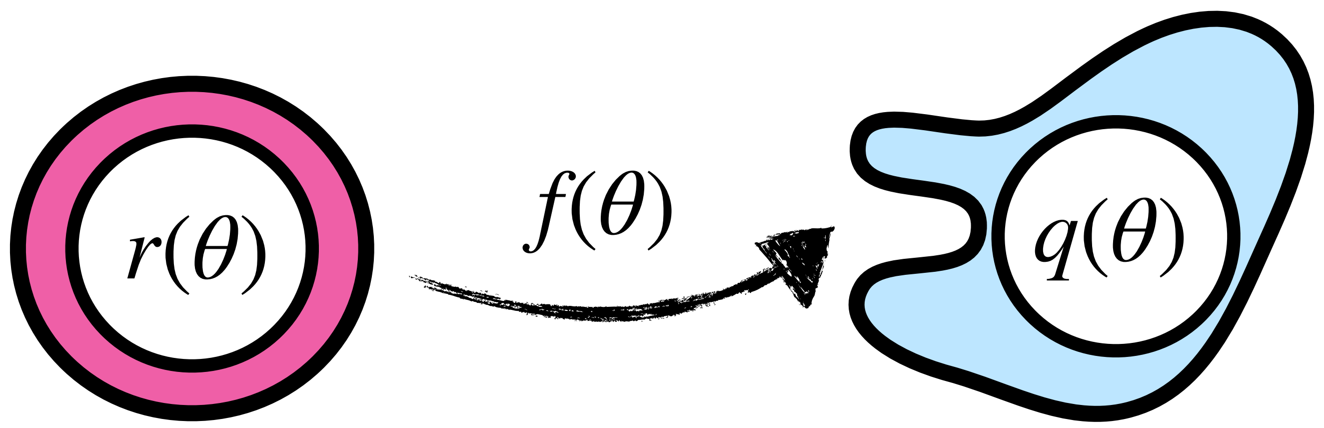


Gauge invariant
$P_{\mu\nu} \in U(1)$

Fernando Romero-López, MIT

Flows on compact variables

- **Flows on compact connected manifolds:**



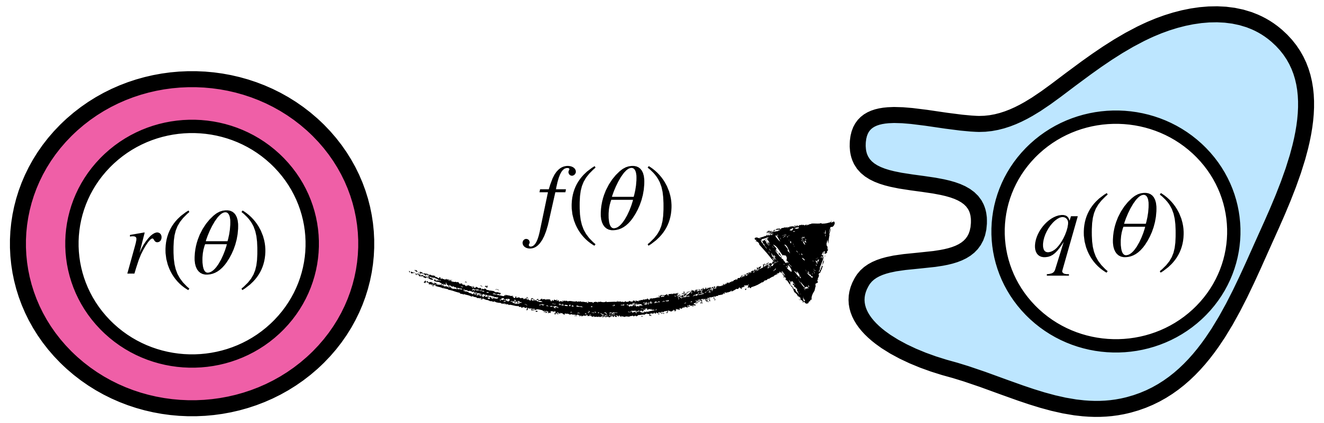
$$U_\mu(x) = \exp(i\theta) \in \mathrm{U}(1)$$

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

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:

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

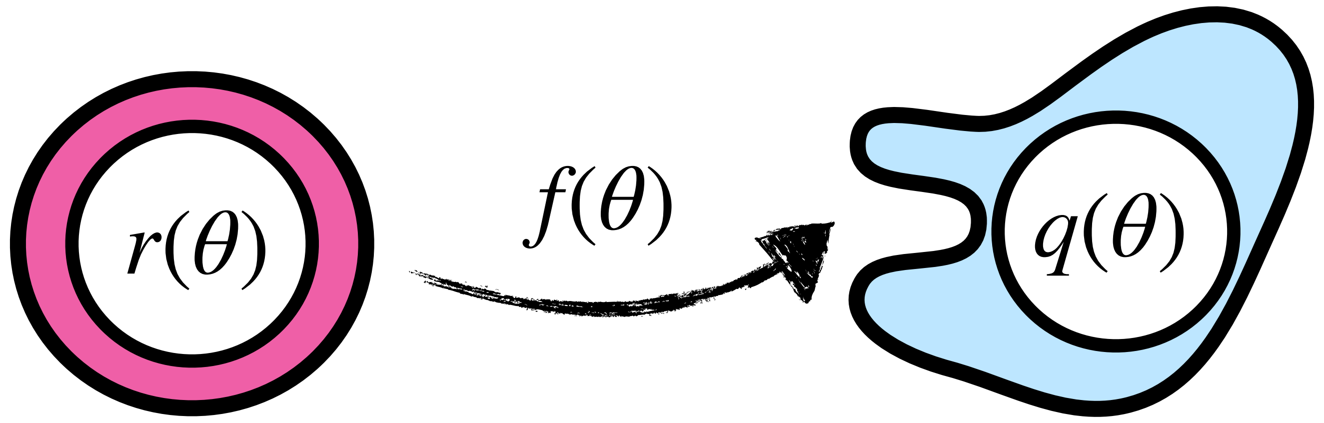
Fernando Romero-López, MIT

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:

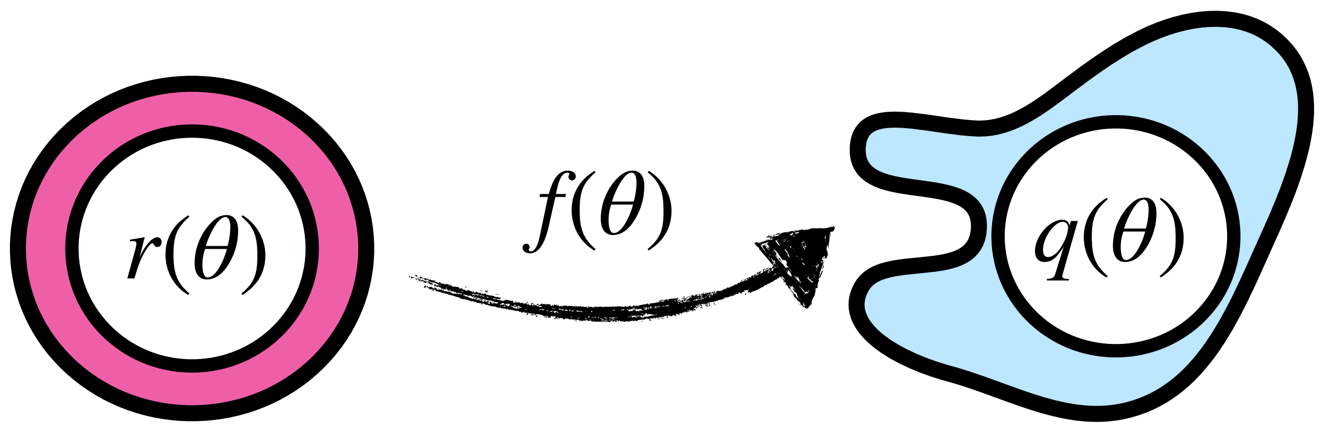
$$\begin{aligned} f(0) &= 0, \\ f(2\pi) &= 2\pi, \\ \nabla f(\theta) &> 0, \\ \nabla f(\theta)\big|_{\theta=0} &= \nabla f(\theta)\big|_{\theta=2\pi} \end{aligned}$$

monotonic, invertible

Flows on compact variables

- **Flows on compact connected manifolds:**

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



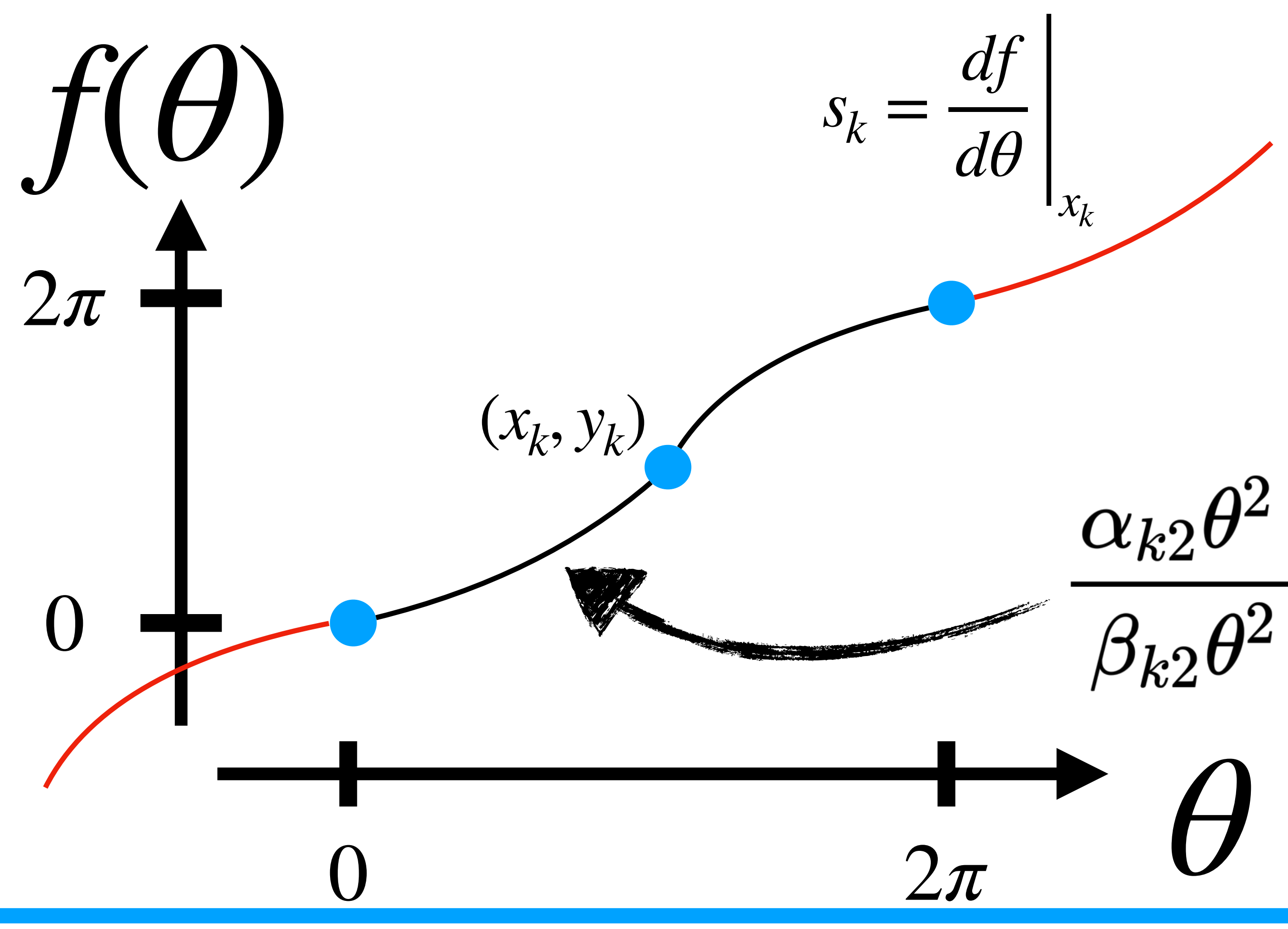
[Rezende, Papamakarios, Racanière , Albergo, Kanwar, Shanahan, Cranmer 2002.02428]

Diffeomorphism if:

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

monotonic, invertible

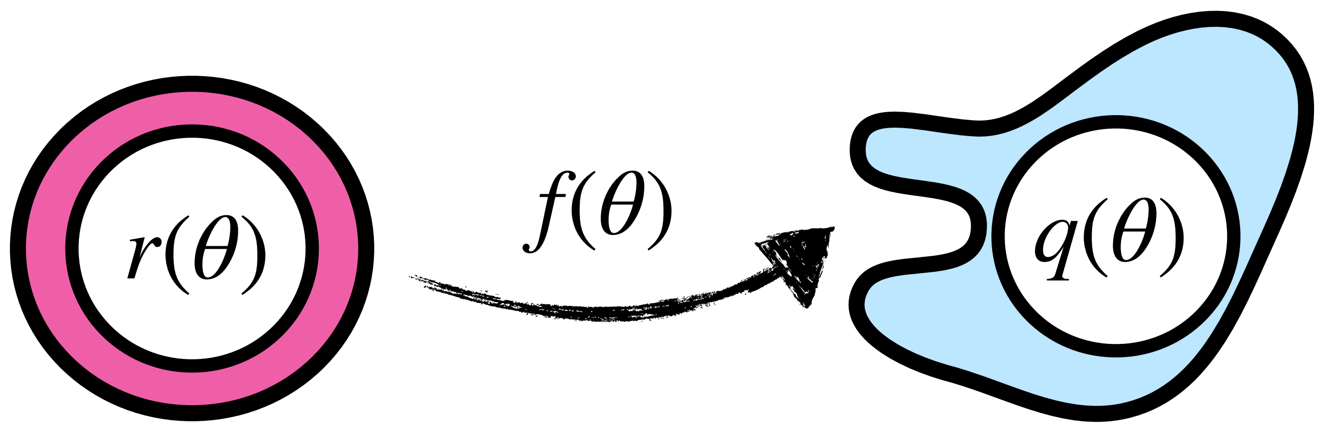
Circular splines:



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

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]

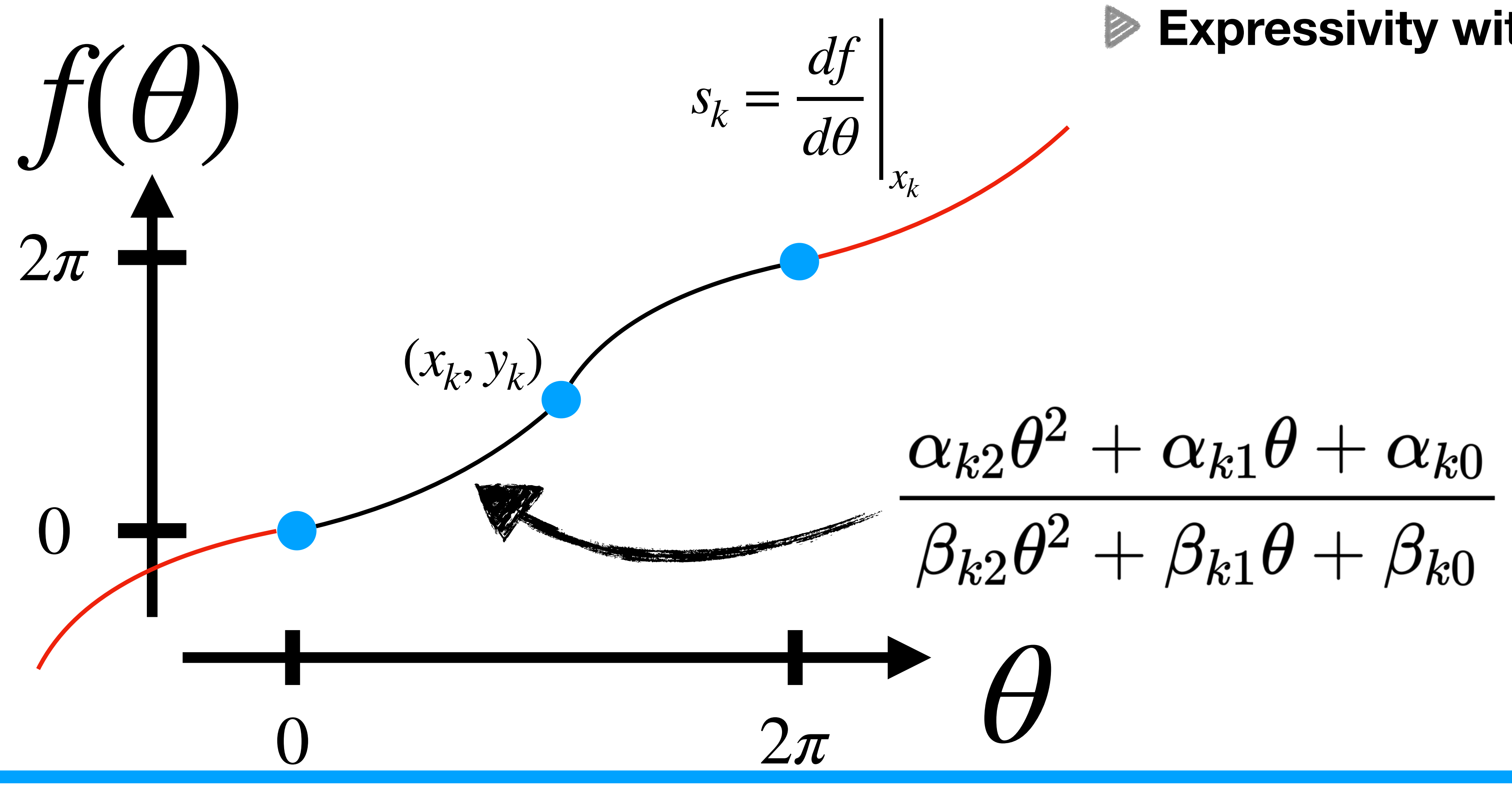
Diffeomorphism if:

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

monotonic, invertible

Circular splines:

- Trainable positions and slopes
- Expressivity with more knots



Fernando Romero-López, MIT

Flows with gauge variables

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

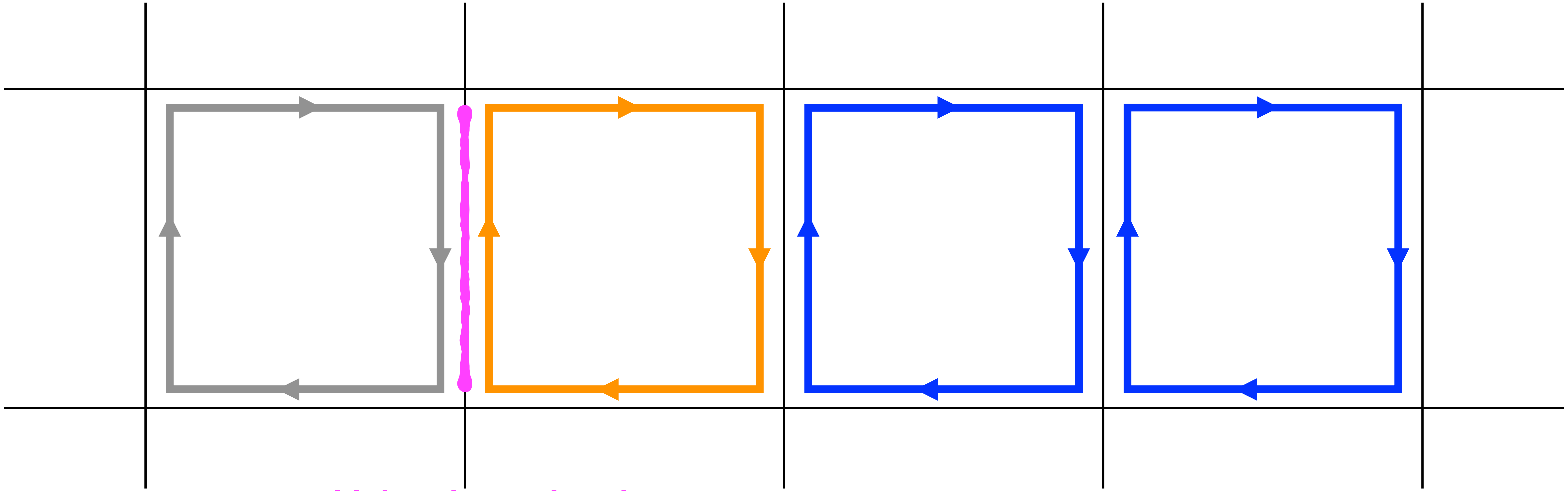
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$
$$g\big(\Omega(x)U_\mu(x)\Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$

Fernando Romero-López, MIT

Flows with gauge variables

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

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



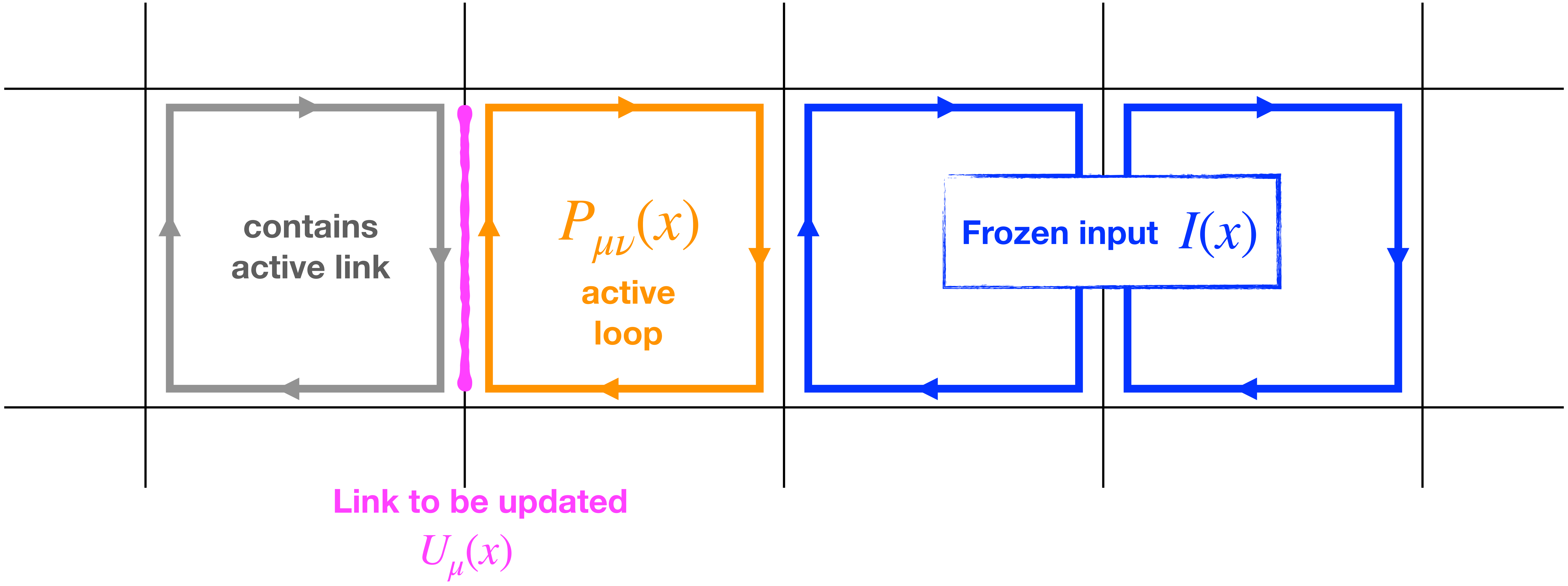
Fernando Romero-López, MIT

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

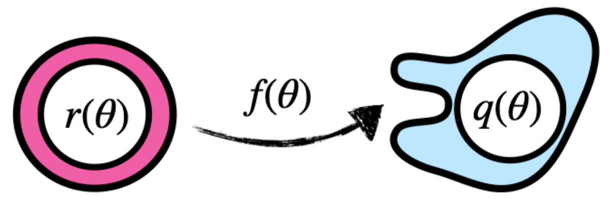
- **Gauge invariant prior + gauge-equivariant transformations:**

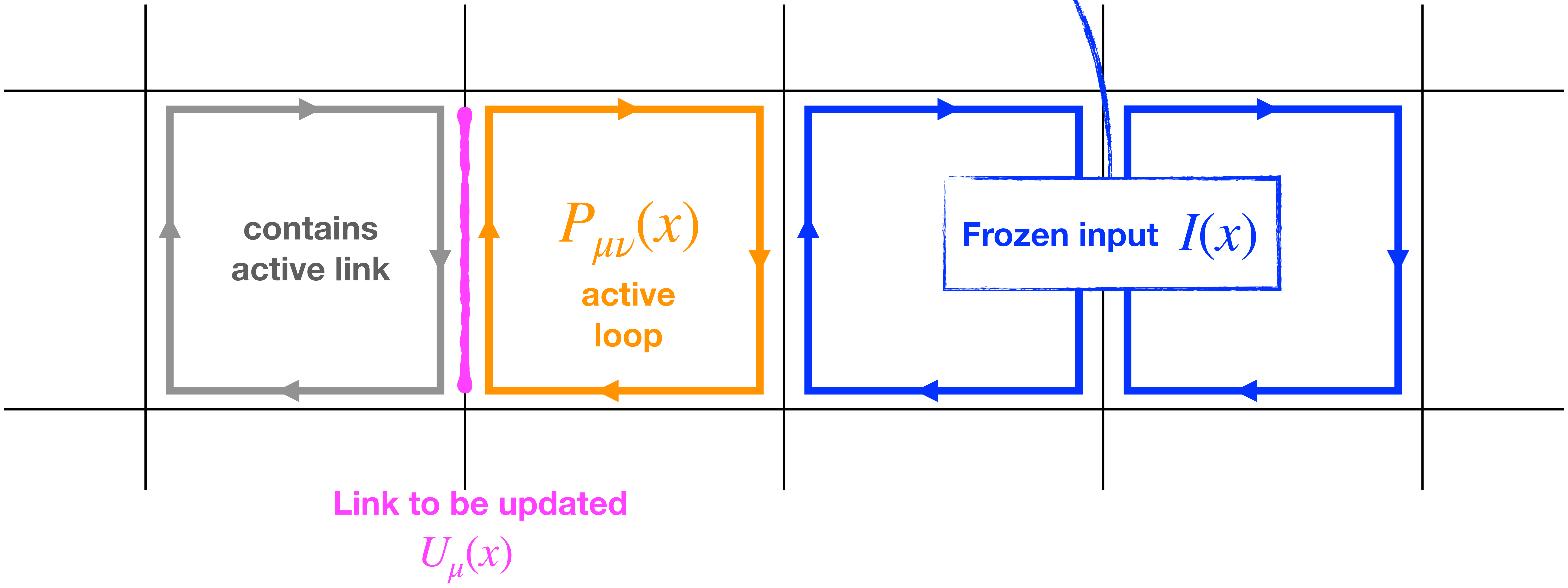
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

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

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

e.g. circular splines





Fernando Romero-López, MIT

Flows with gauge variables

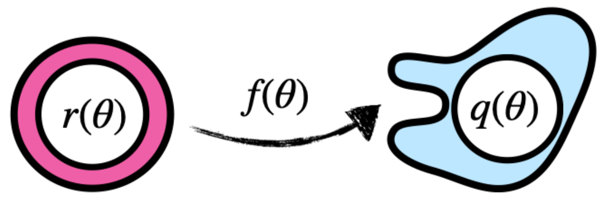
- **Gauge invariant prior + gauge-equivariant transformations:**

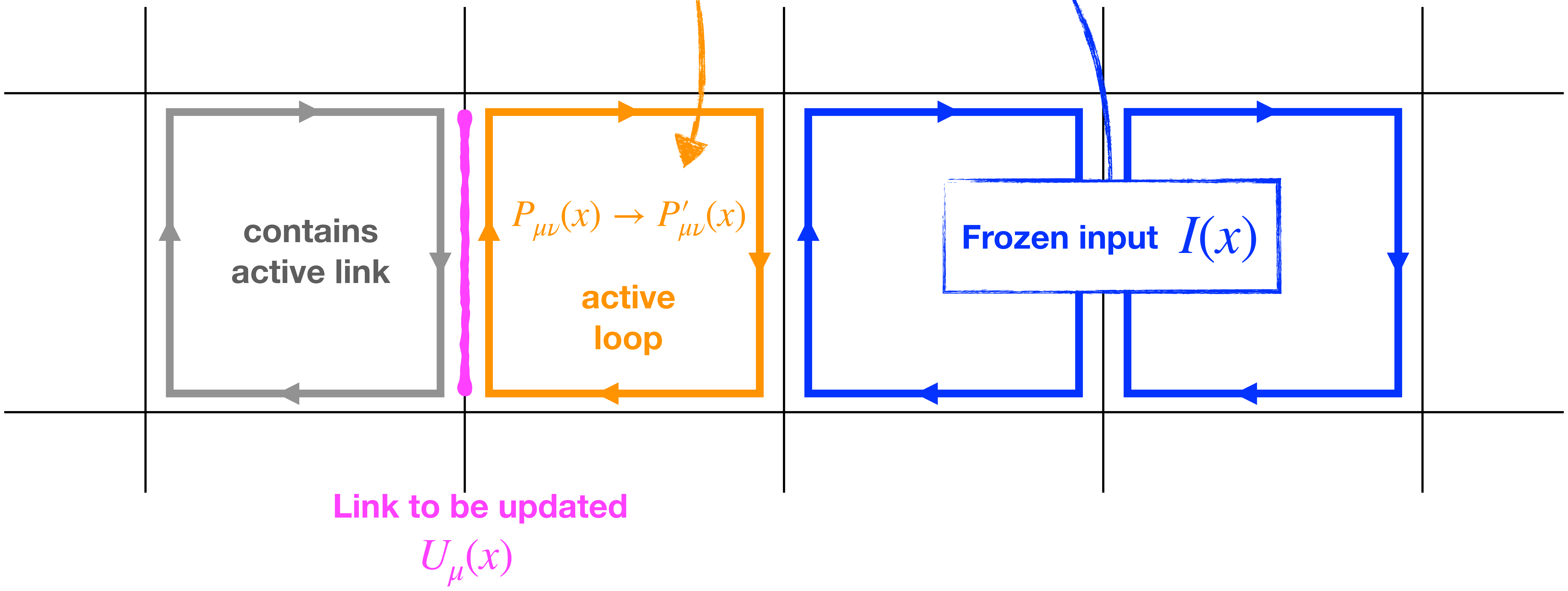
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

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

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

e.g. circular splines





Fernando Romero-López, MIT

Flows with gauge variables

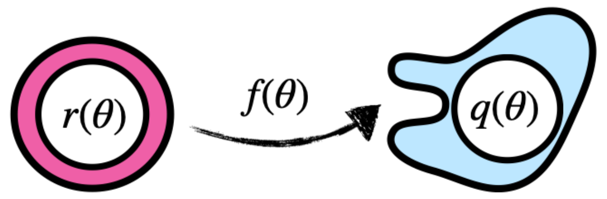
- **Gauge invariant prior + gauge-equivariant transformations:**

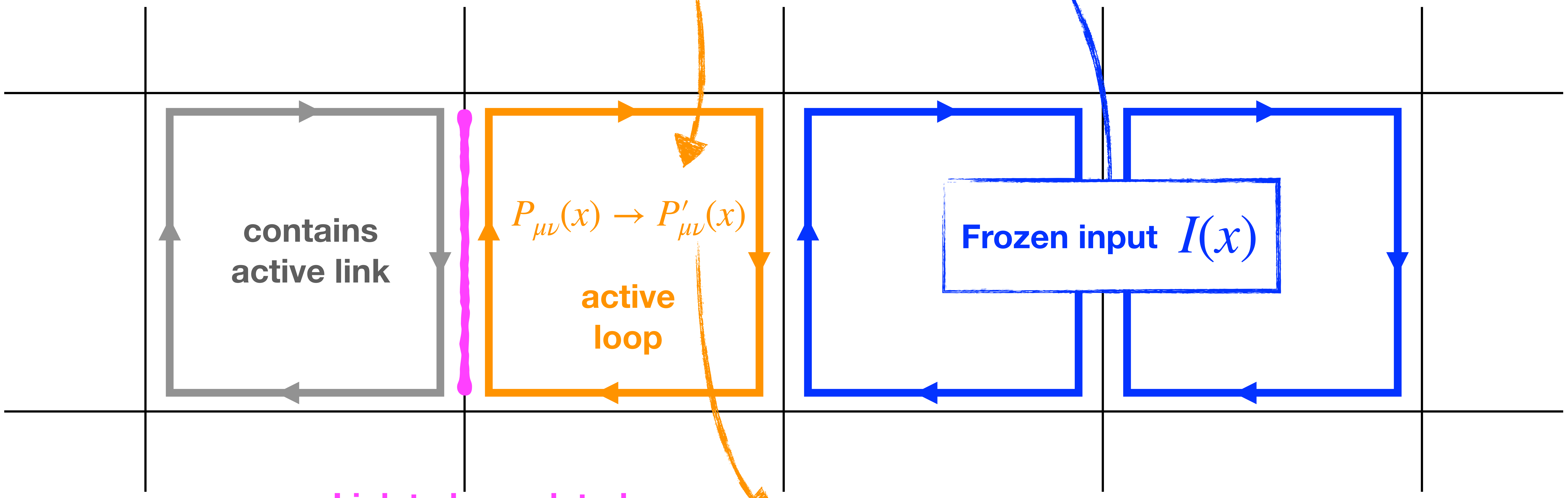
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

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

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

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

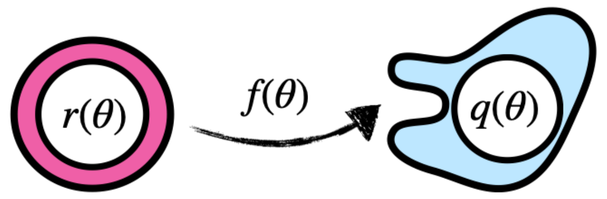
- **Gauge invariant prior + gauge-equivariant transformations:**

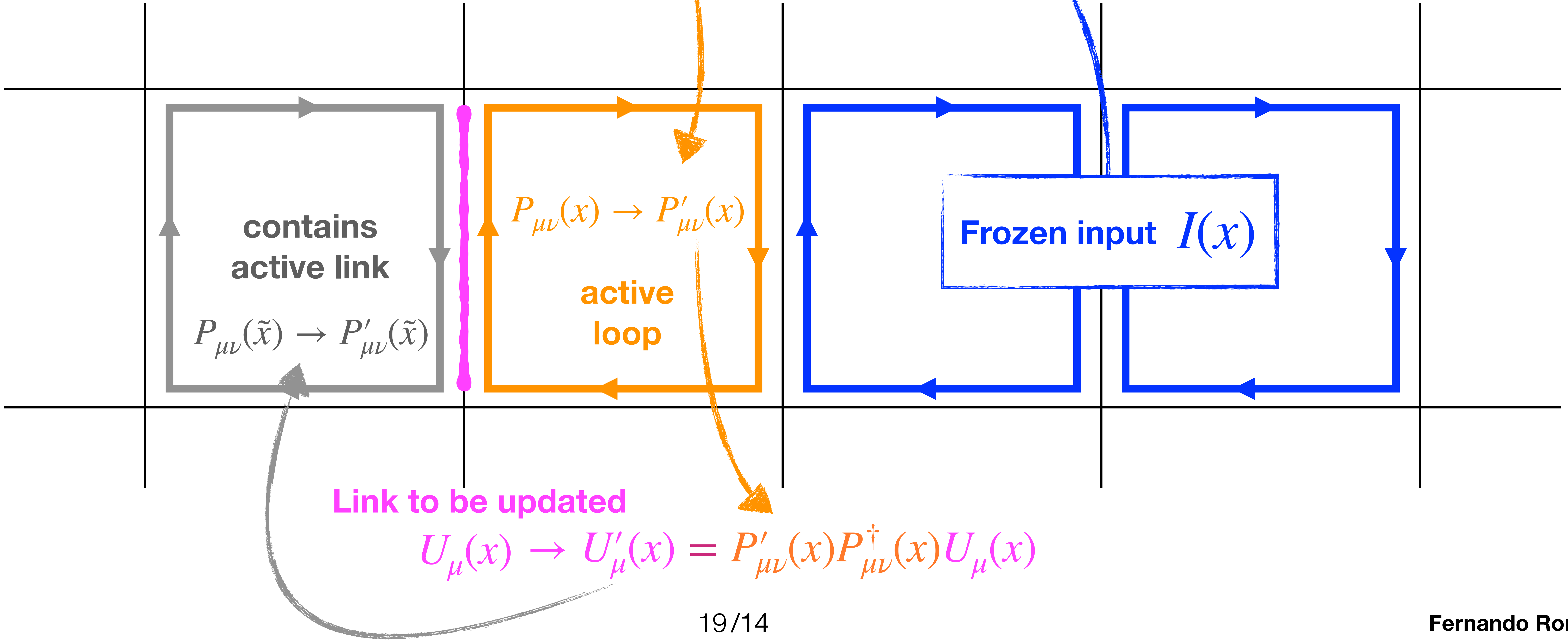
$$\theta \sim \mathrm{Uniform}(0, 2\pi)$$

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

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

e.g. circular splines





Link to be updated

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

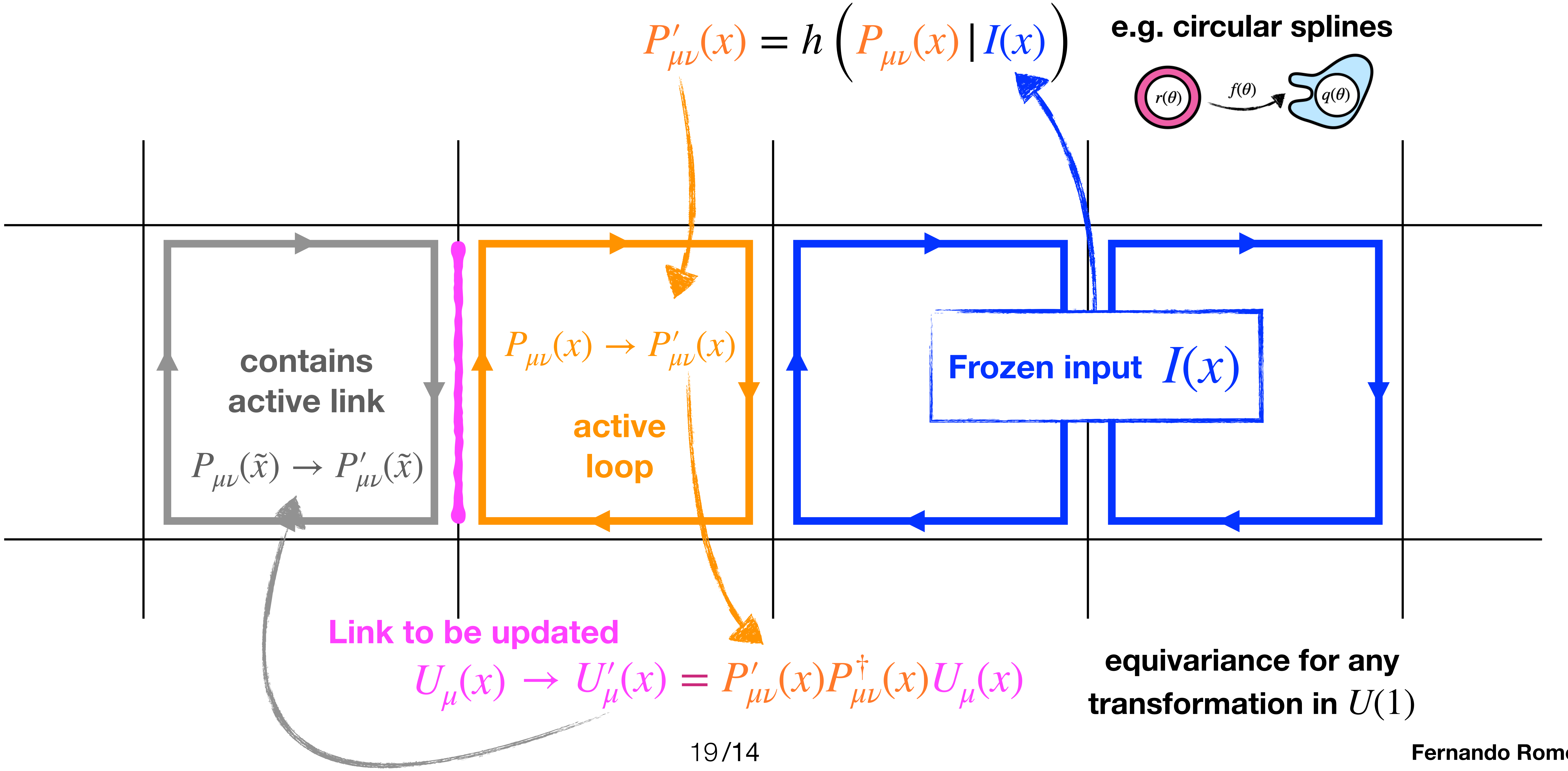
Fernando Romero-López, MIT

Flows with gauge variables

- Gauge invariant prior + gauge-equivariant transformations:

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

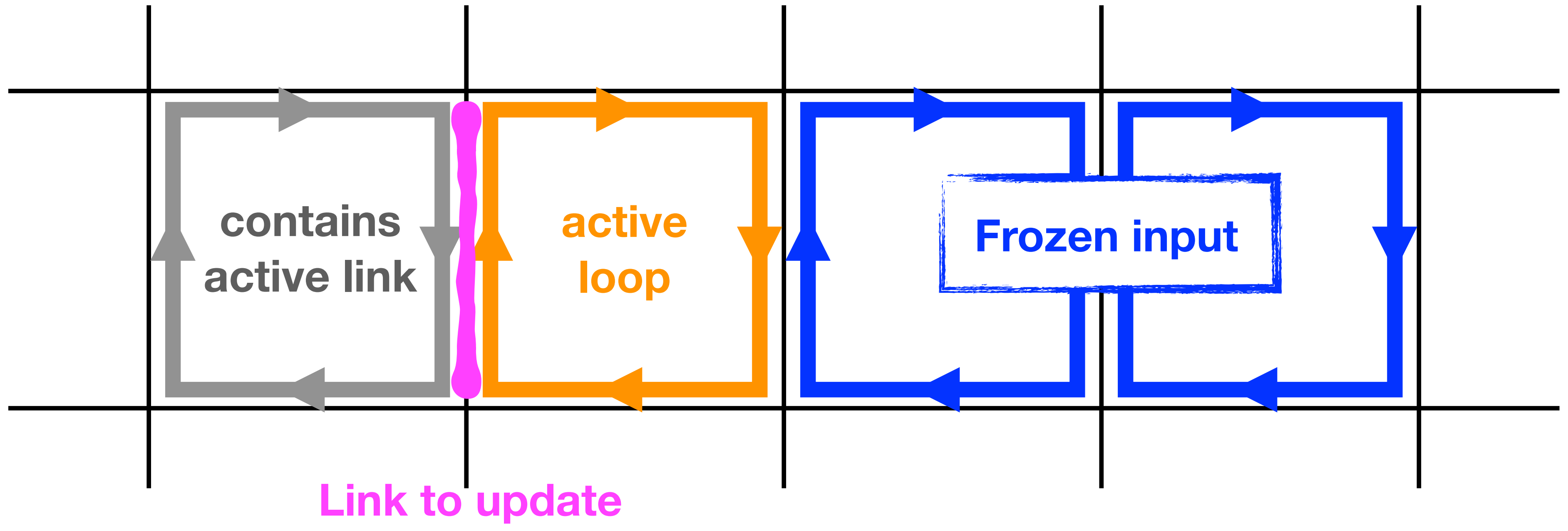
$$g\big(\Omega(x)U_\mu(x)\Omega^\dagger(x+\mu)\big) = \Omega(x)\, g(U_\mu(x))\, \Omega^\dagger(x+\mu)$$



equivariance for any transformation in $U(1)$

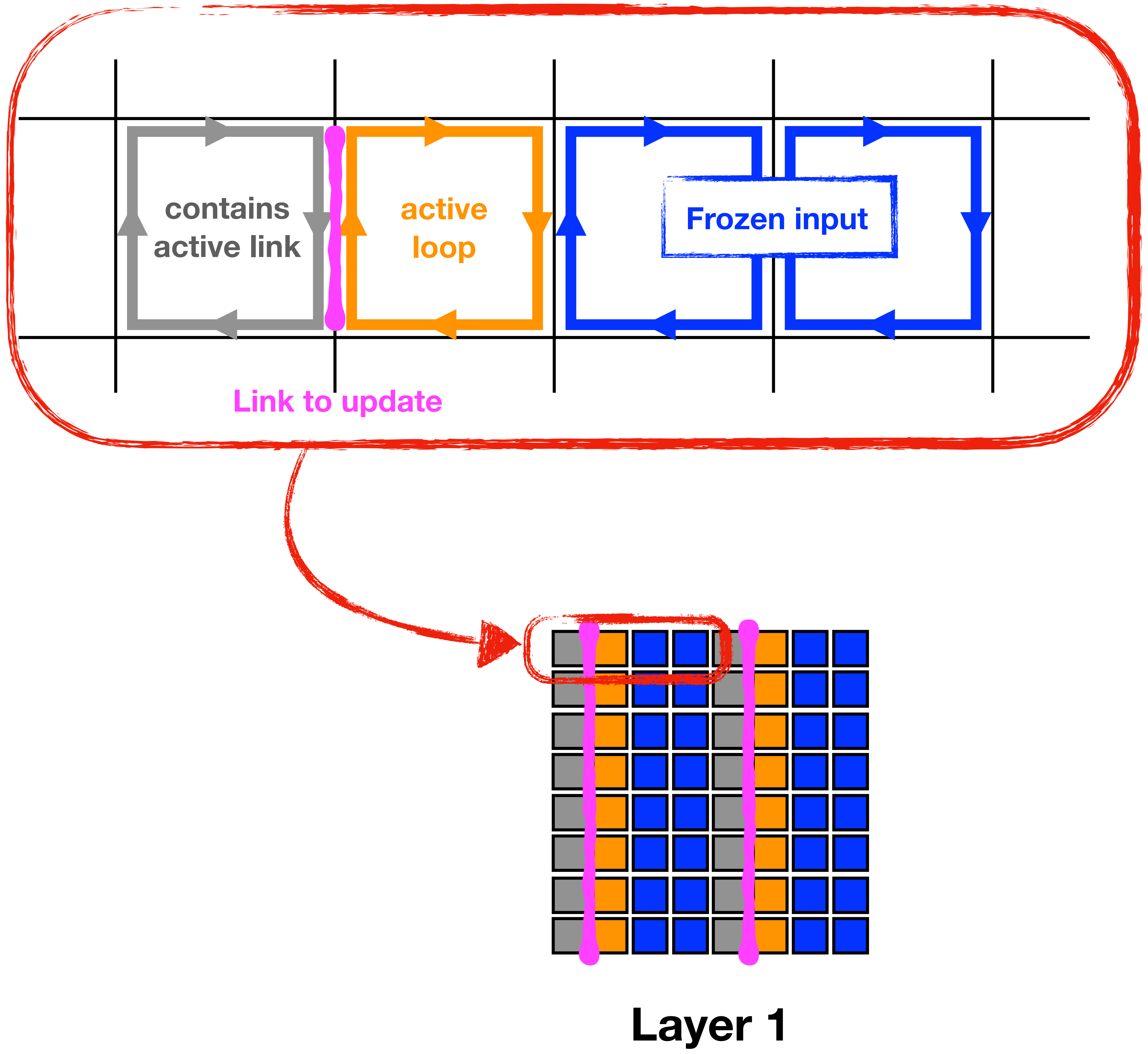
Fernando Romero-López, MIT

Gauge-equivariant models



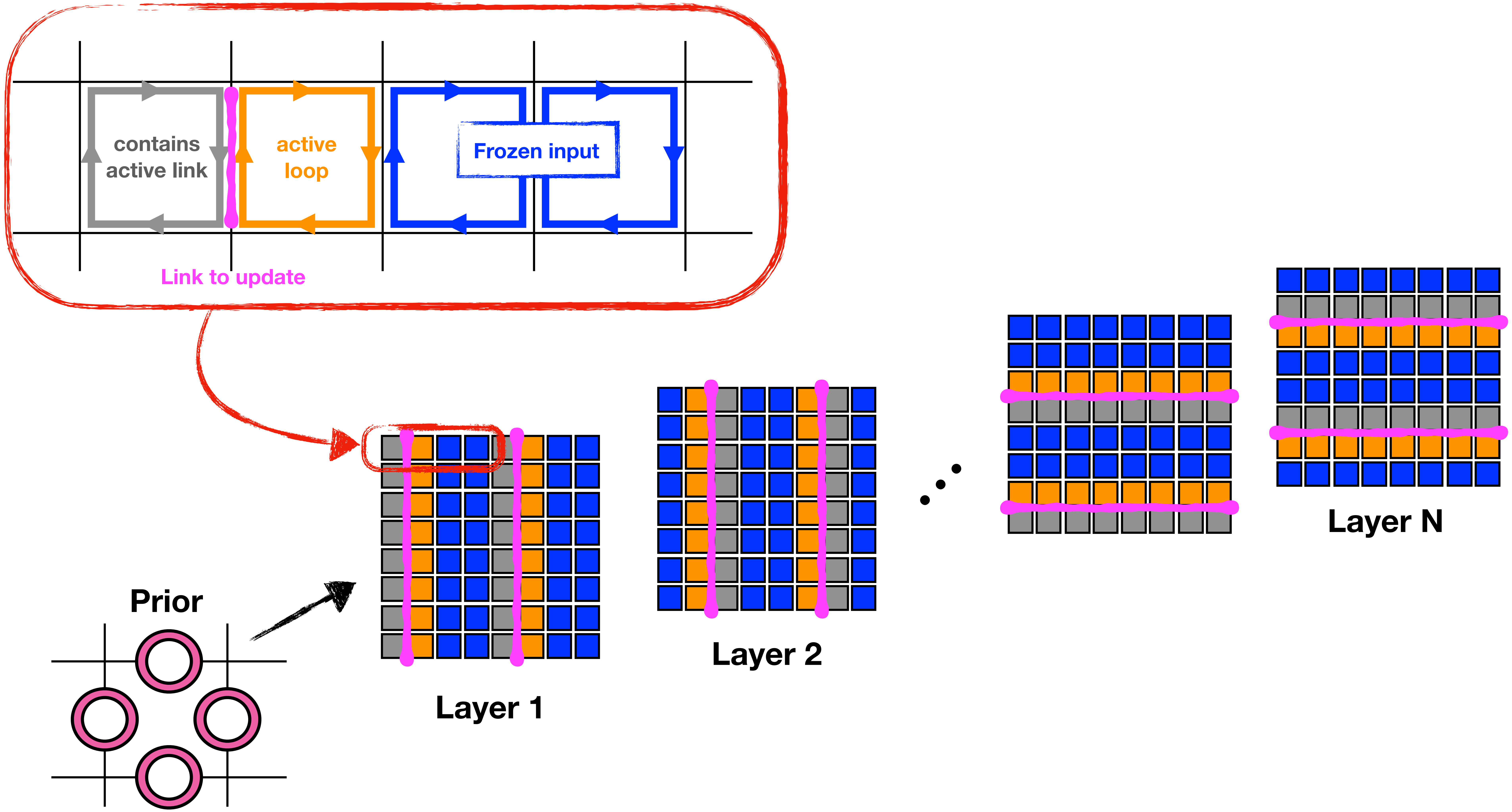
Fernando Romero-López, MIT

Gauge-equivariant models



Fernando Romero-López, MIT

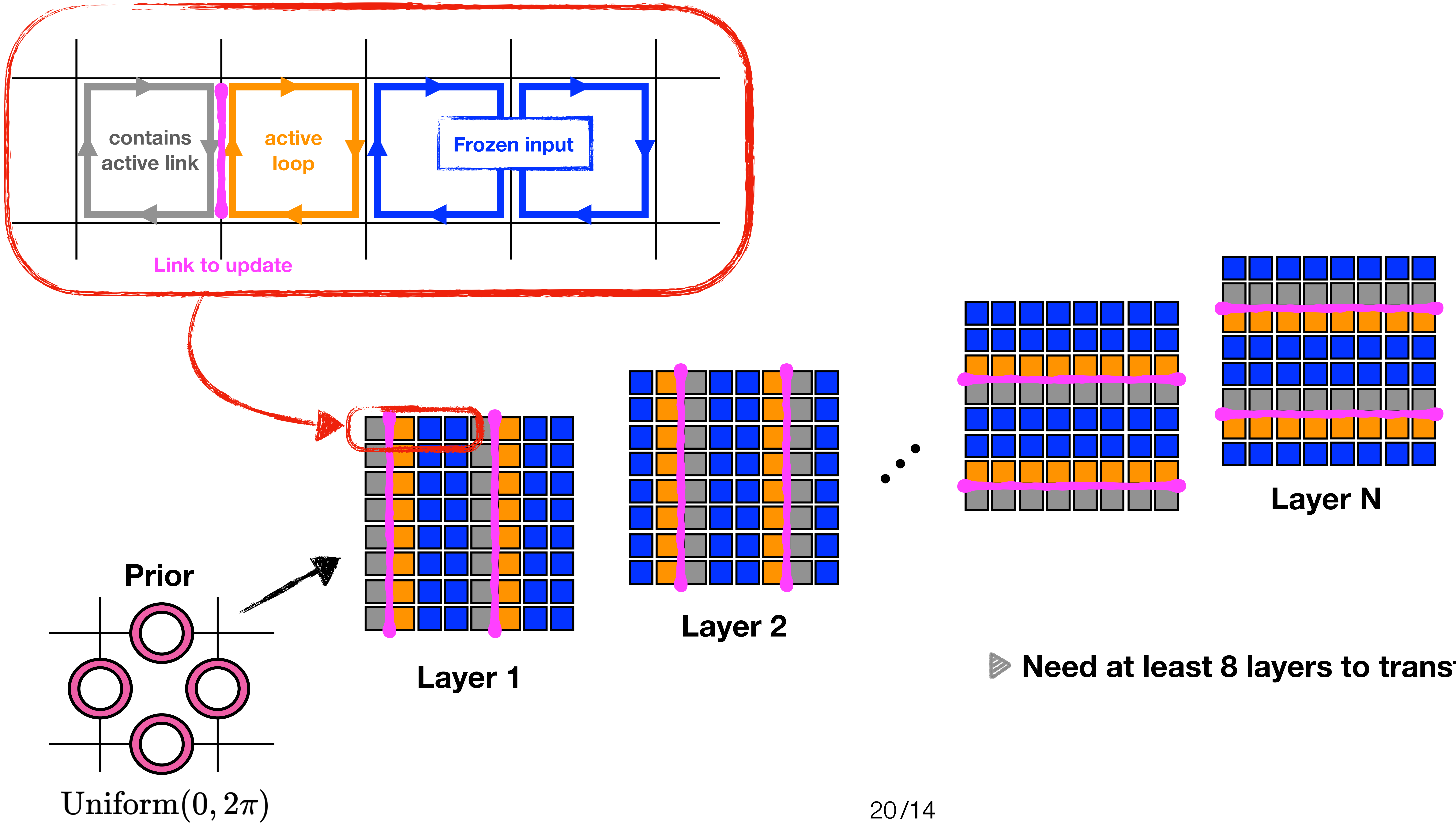
Gauge-equivariant models



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

Fernando Romero-López, MIT

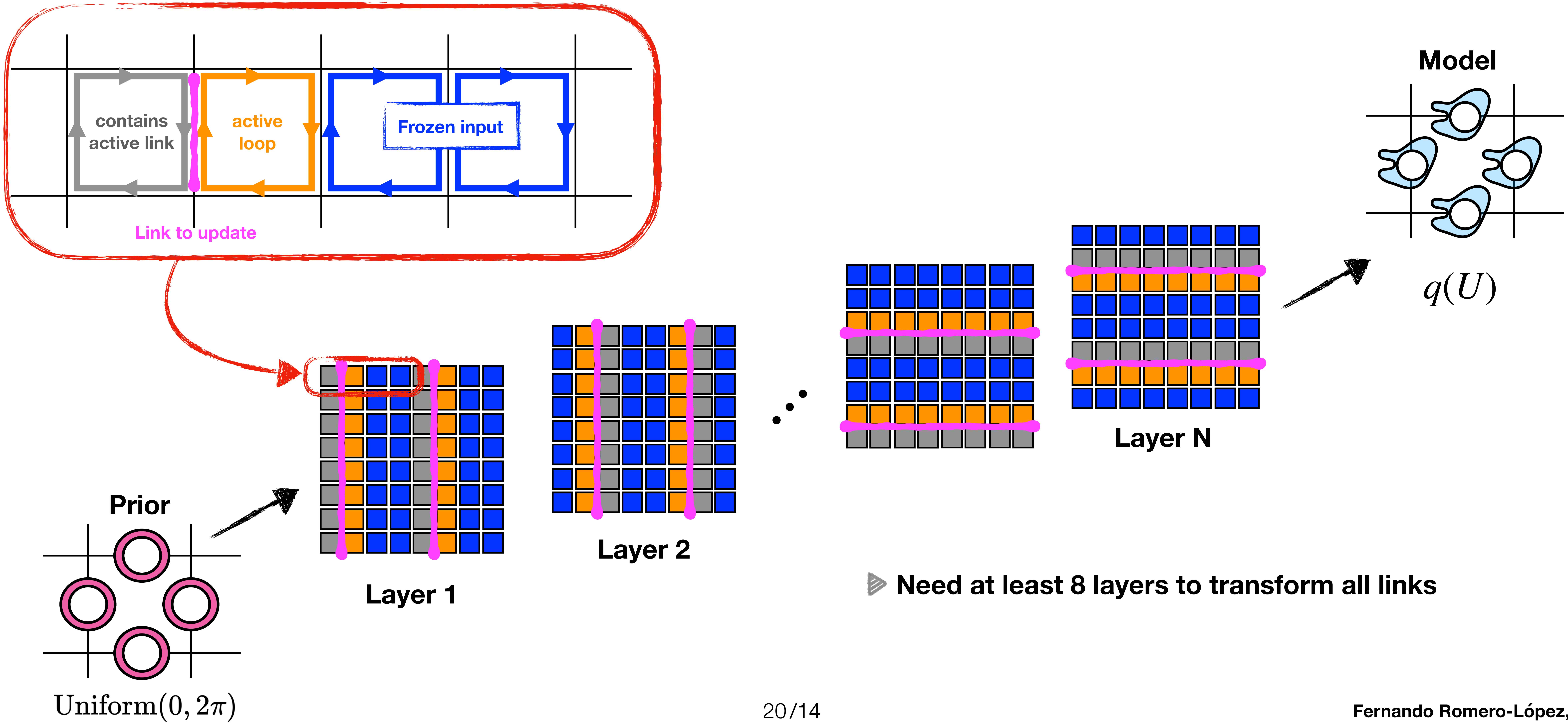
Gauge-equivariant models



- Need at least 8 layers to transform all links

Fernando Romero-López, MIT

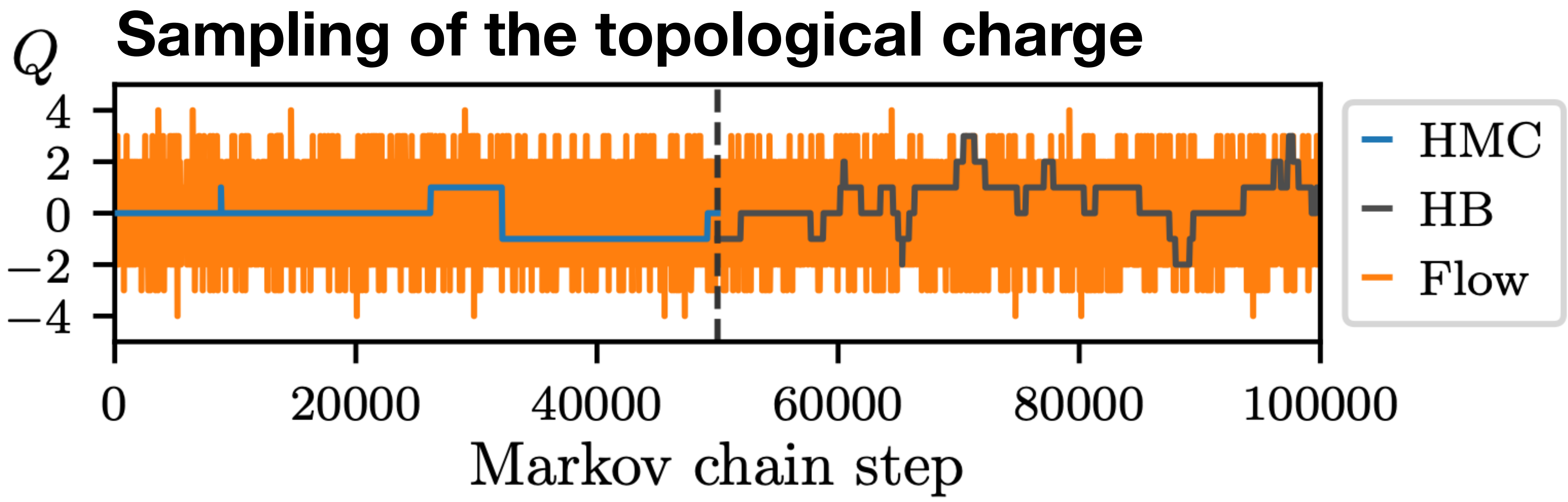
Gauge-equivariant models



- Need at least 8 layers to transform all links

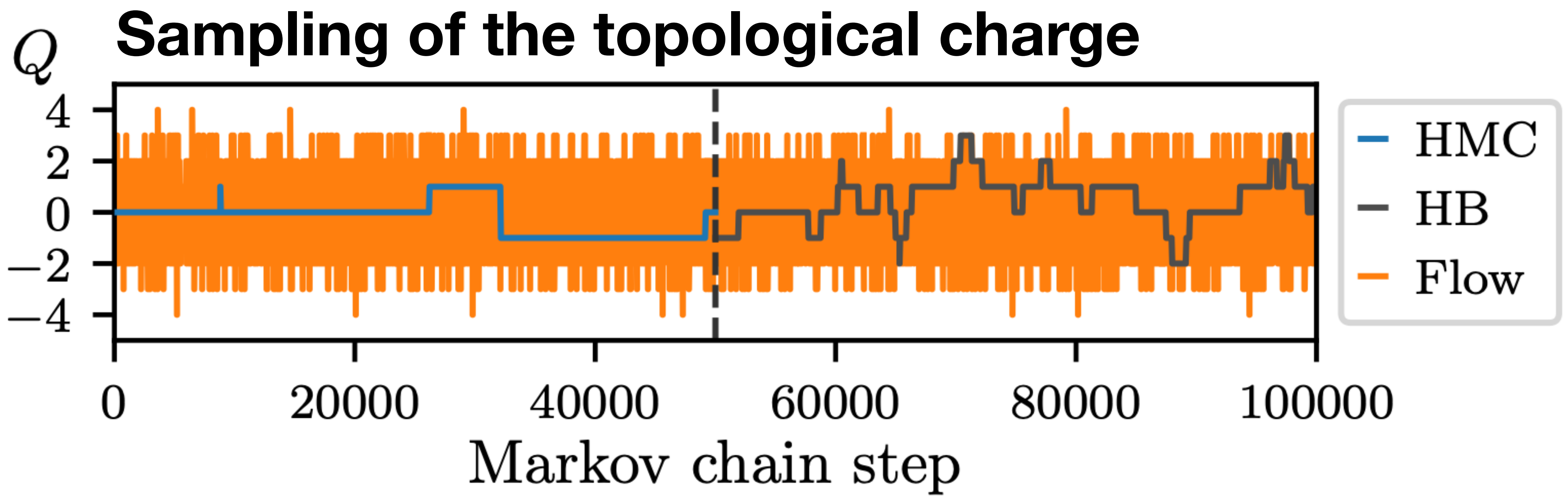
Fernando Romero-López, MIT

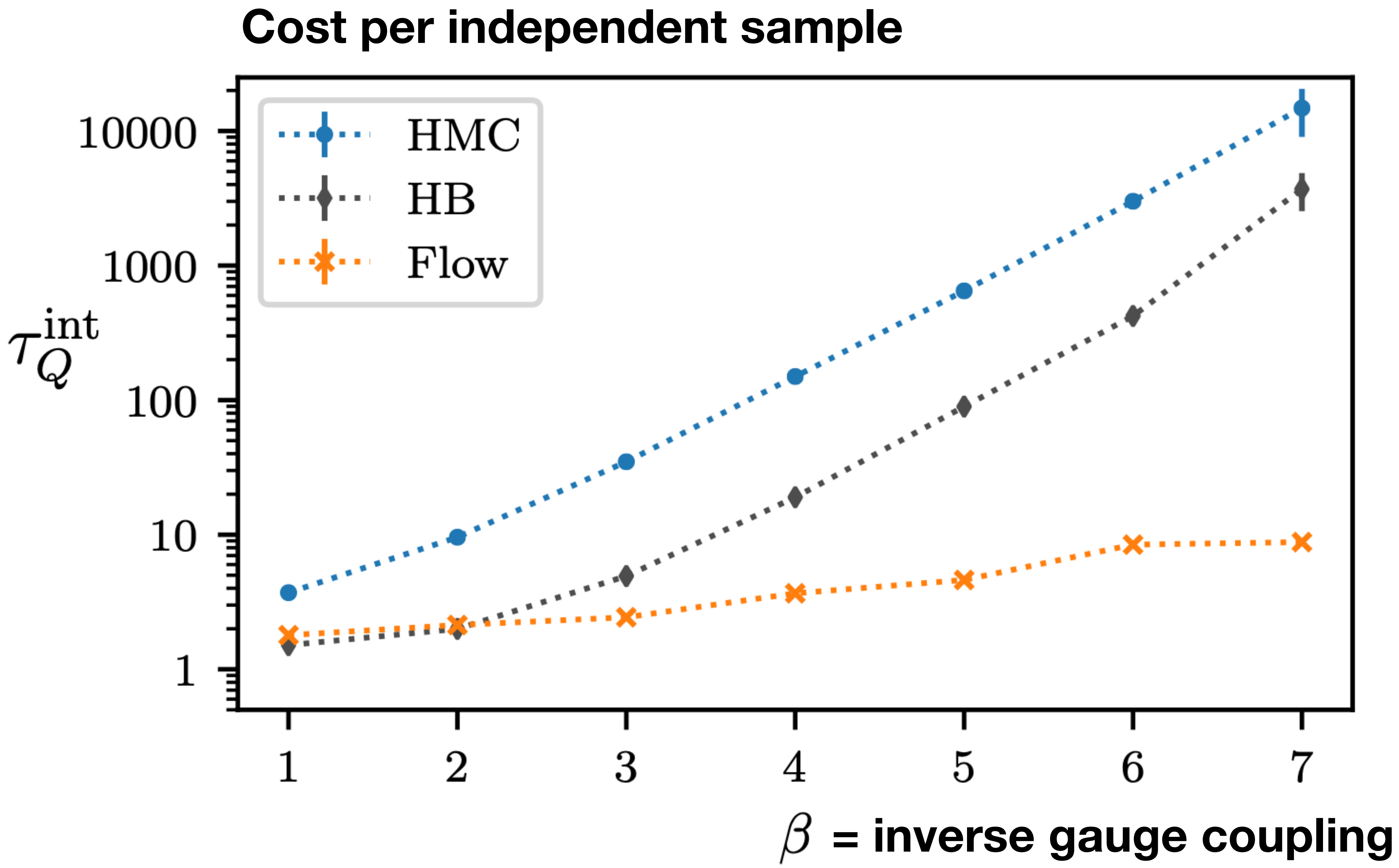
Results for pure U(1) theory



Fernando Romero-López, MIT

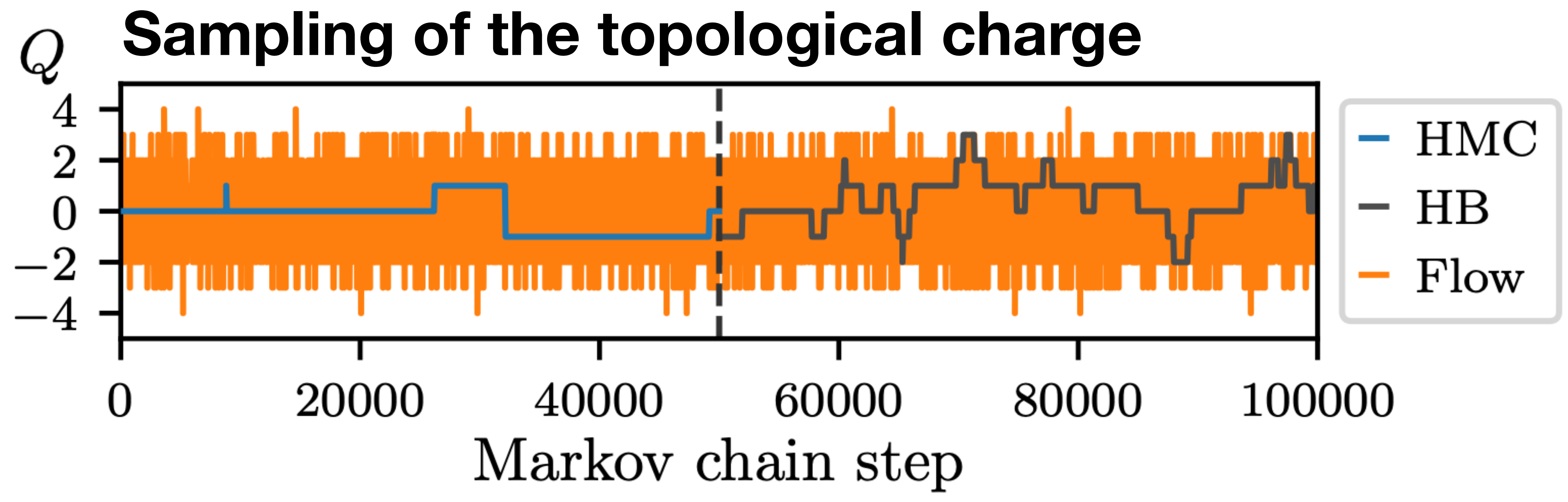
Results for pure U(1) theory

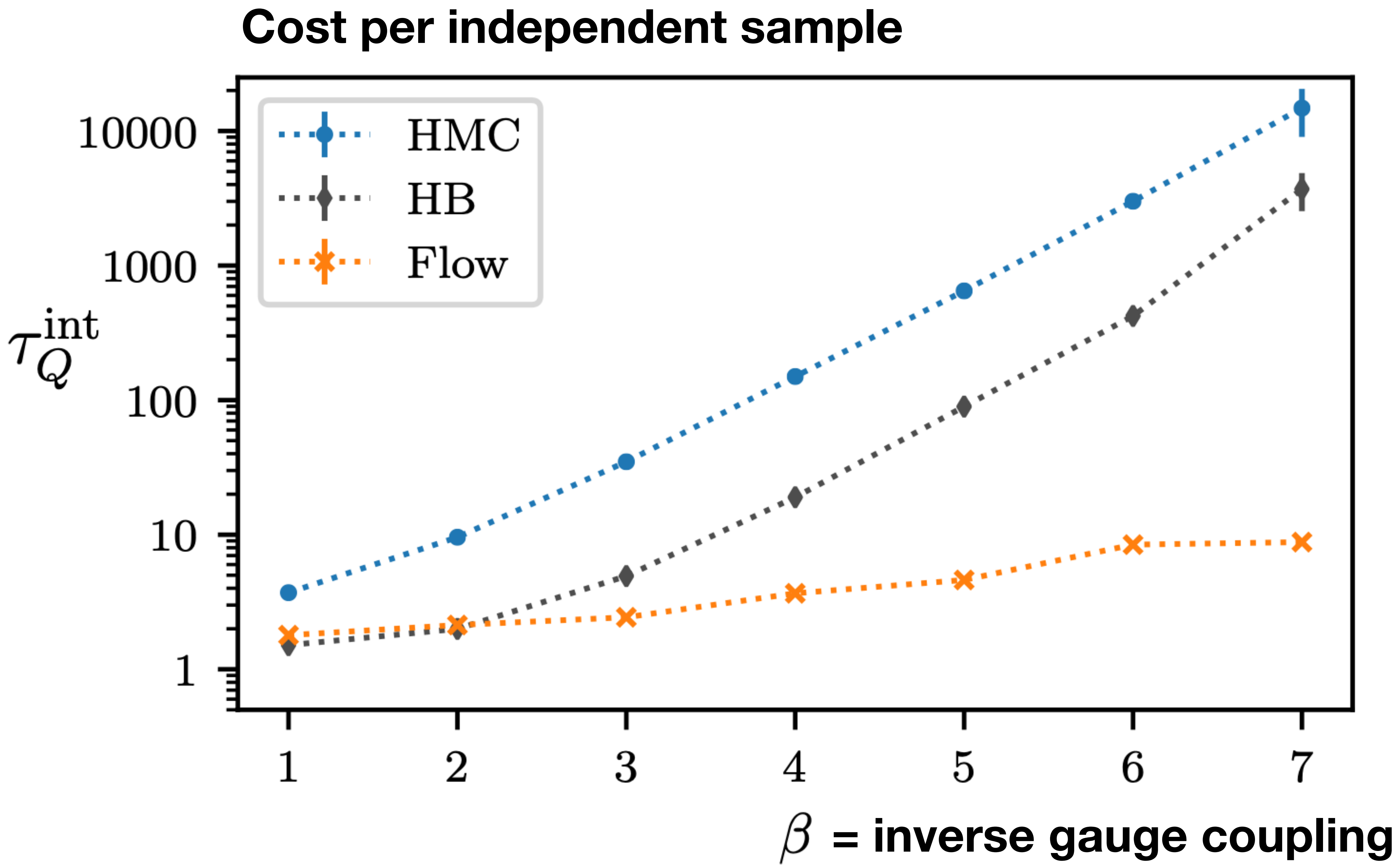




Fernando Romero-López, MIT

Results for pure U(1) theory



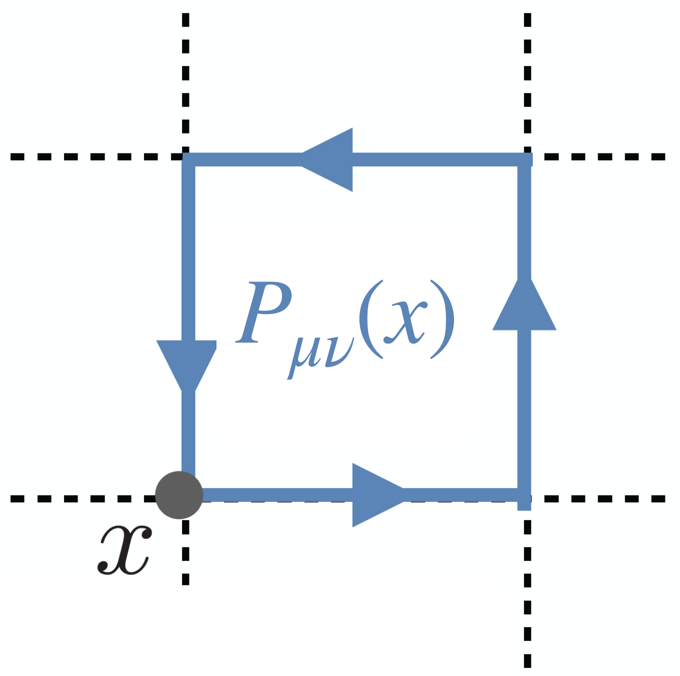


✓ **Flow-based sampling has no topological freezing**

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

Flows on SU(3)

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

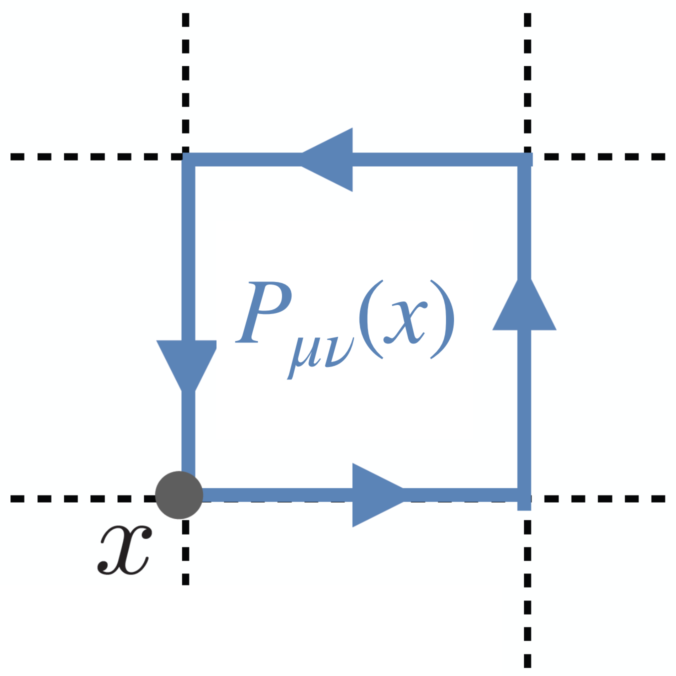


$$P_{\mu\nu} = X \begin{pmatrix} e^{i\theta_1} & 0 & 0 \\ 0 & e^{i\theta_2} & 0 \\ 0 & 0 & e^{i\theta_3} \end{pmatrix} X^\dagger$$

Fernando Romero-López, MIT

Flows on SU(3)

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

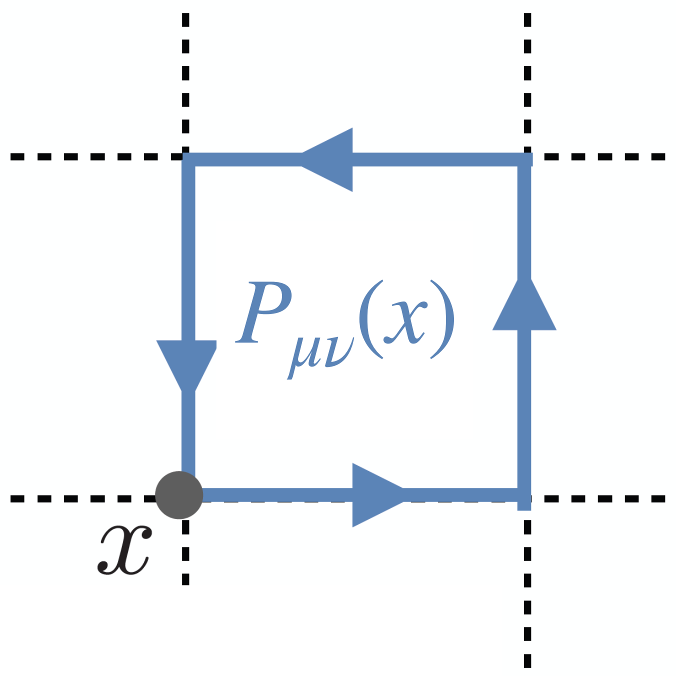


$$P_{\mu\nu} = X \begin{pmatrix} e^{i\theta_1} & 0 & 0 \\ 0 & e^{i\theta_2} & 0 \\ 0 & 0 & e^{i\theta_3} \end{pmatrix} X^\dagger$$

Use transformation on eigenvalues

Flows on SU(3)

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



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

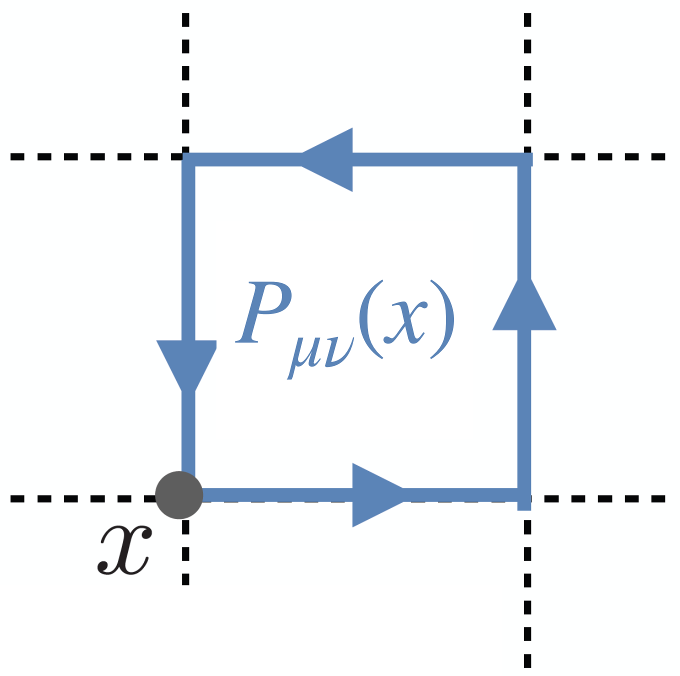
Use transformation on eigenvalues

- 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>

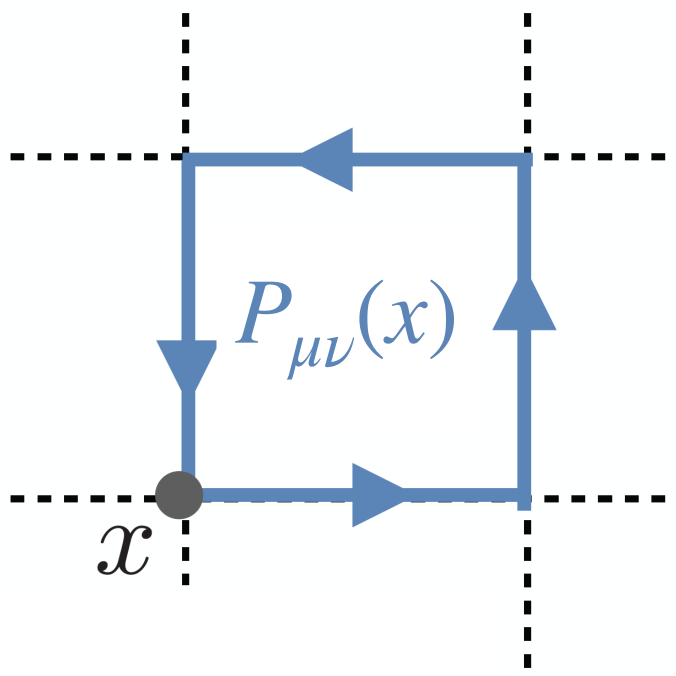
  - Place eigenvalues in canonical ordering
  - Transform eigenvalues
  - Undo canonical ordering

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

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]

The road to at-scale 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

[Boyda, Kanwar, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan 2008.05456]

(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]

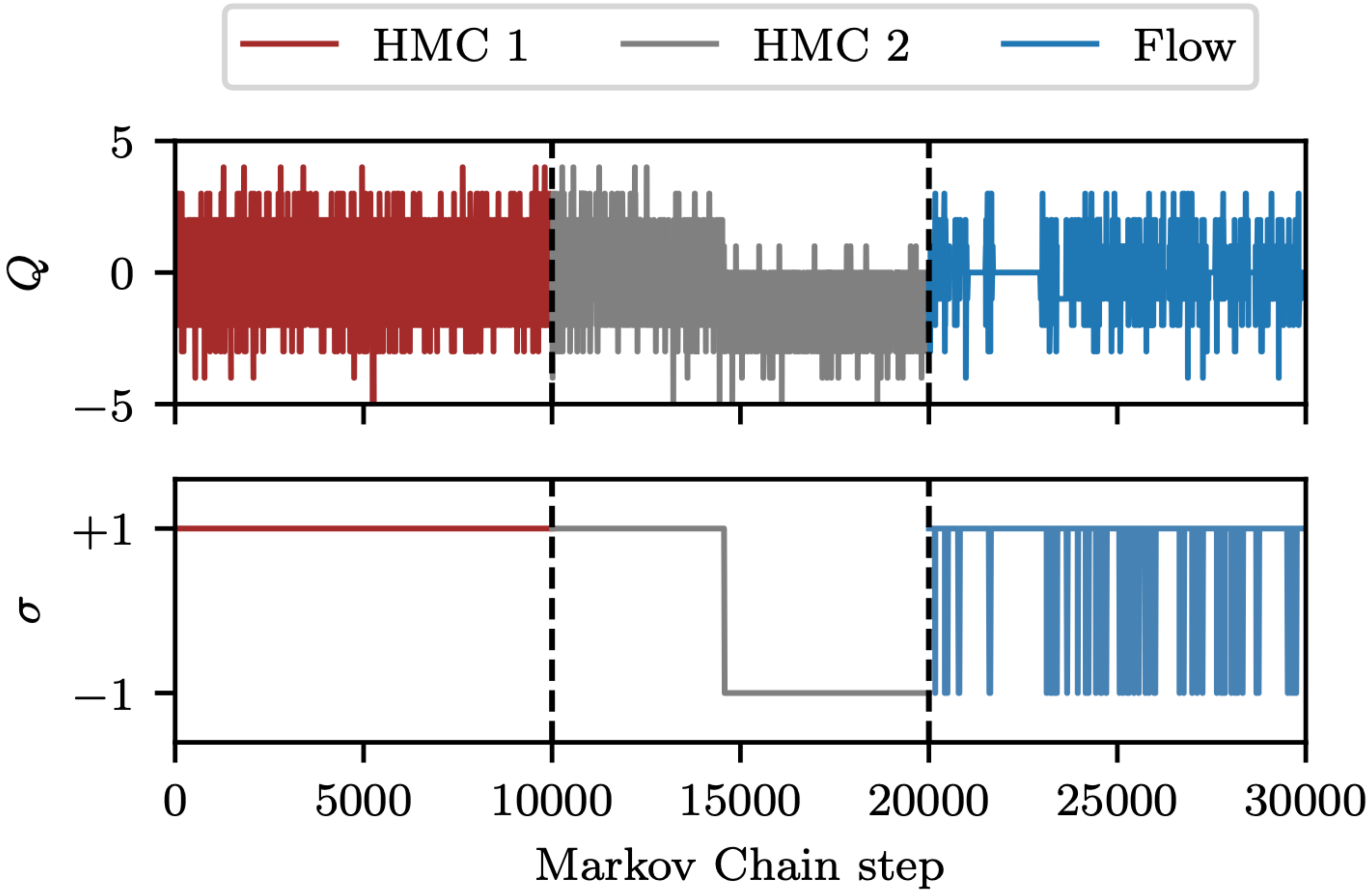
Schwinger and 2D QCD with Pseudofermions

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

The Schwinger model at criticality

Topological charge

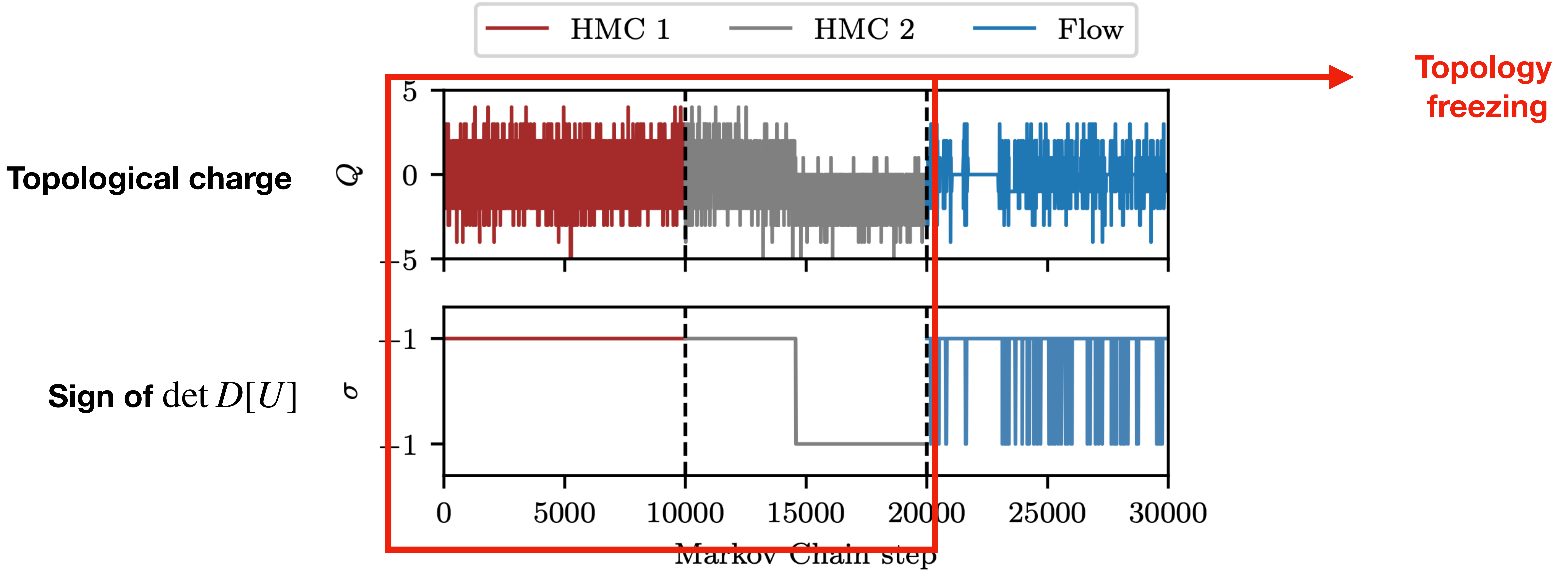
Sign of $\det D[U]$



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

Fernando Romero-López, MIT

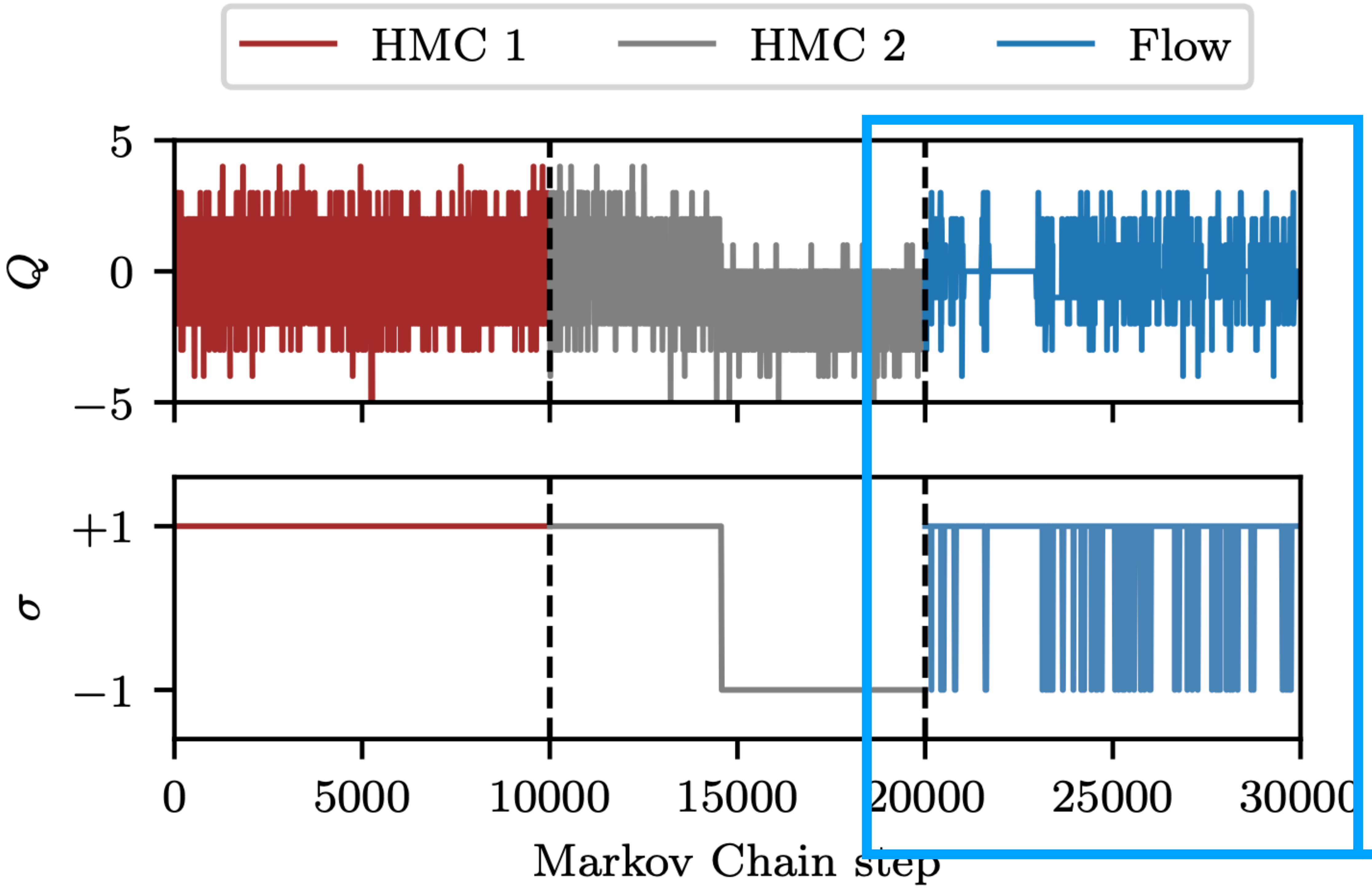
The Schwinger model at criticality



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

Fernando Romero-López, MIT

The Schwinger model at criticality



Topological charge

Sign of $\det D[U]$

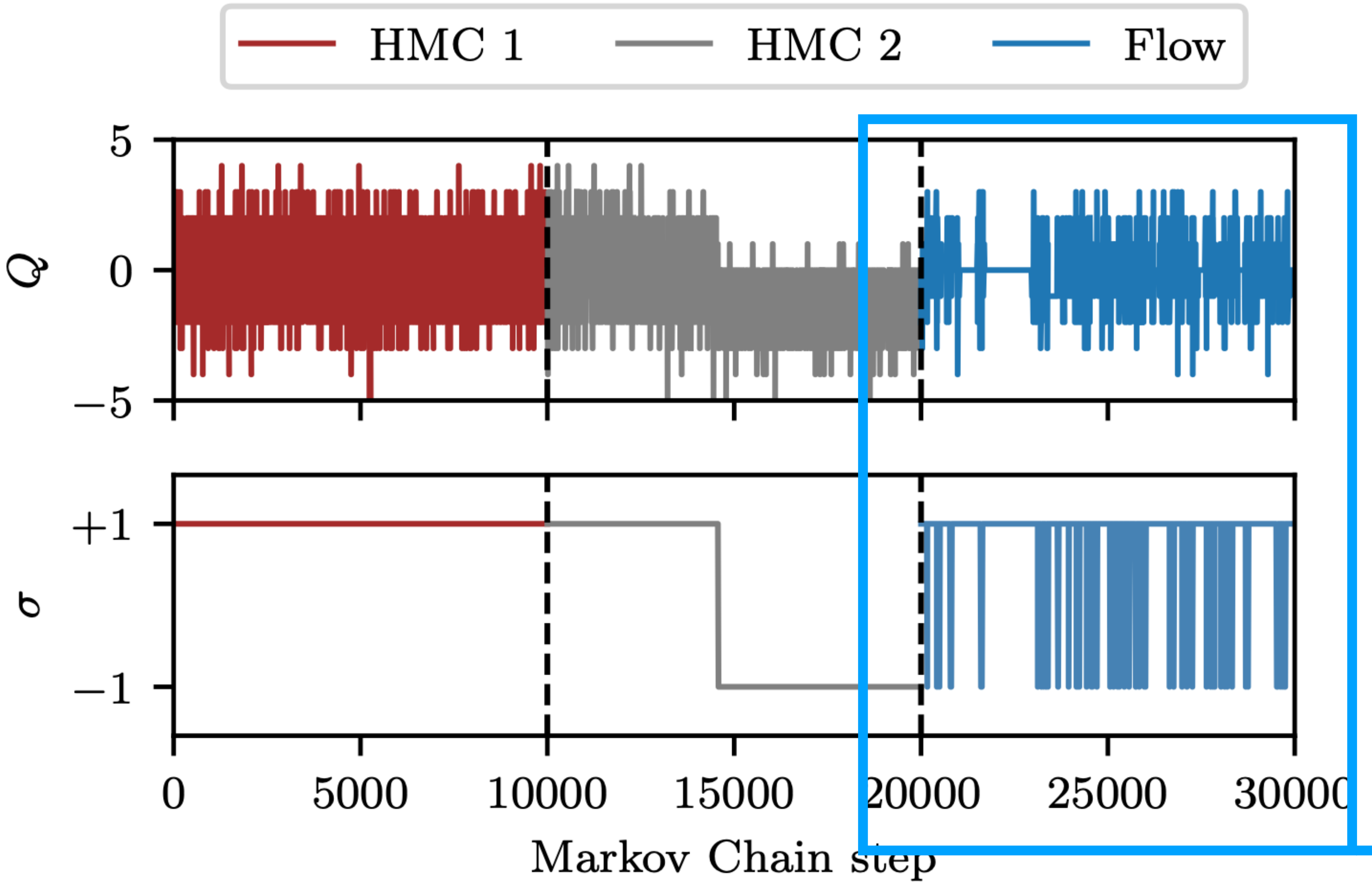
Correct sampling

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

The Schwinger model at criticality

Topological charge

Sign of $\det D[U]$



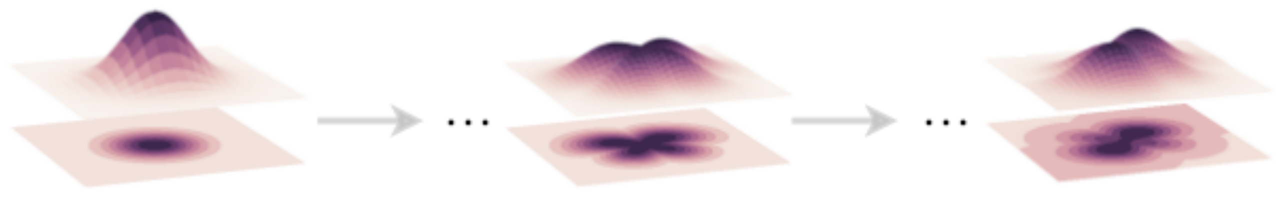
Correct sampling

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

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

Fernando Romero-López, MIT

Lattice QFT via flow models

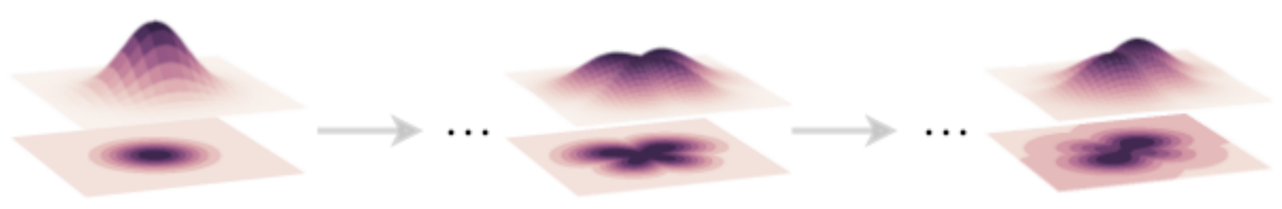


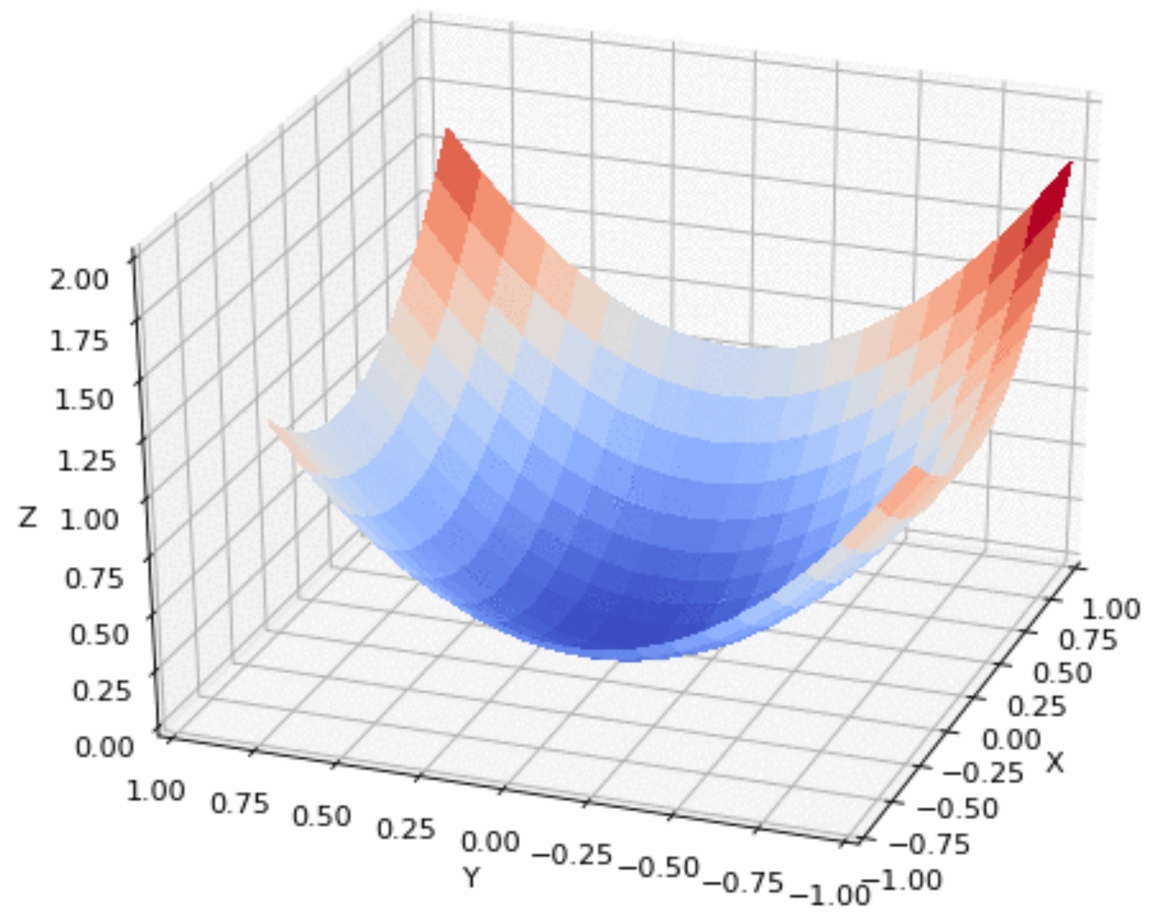
Parameterize flow using coupling layers

Each layer contains arbitrary neural nets

[credits: G. Kanwar]

Lattice QFT via flow models





Parameterize flow using coupling layers

Each layer contains arbitrary neural nets

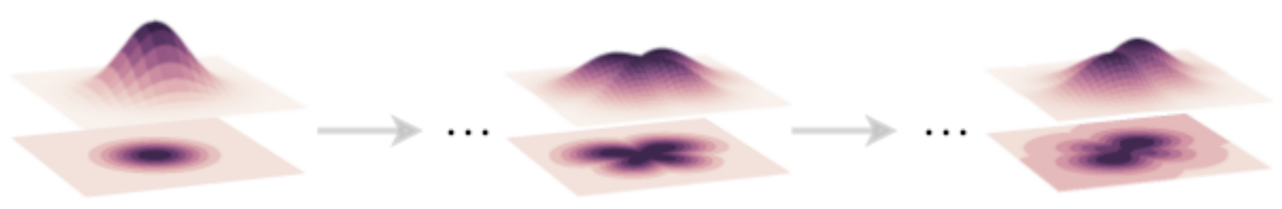
Training step

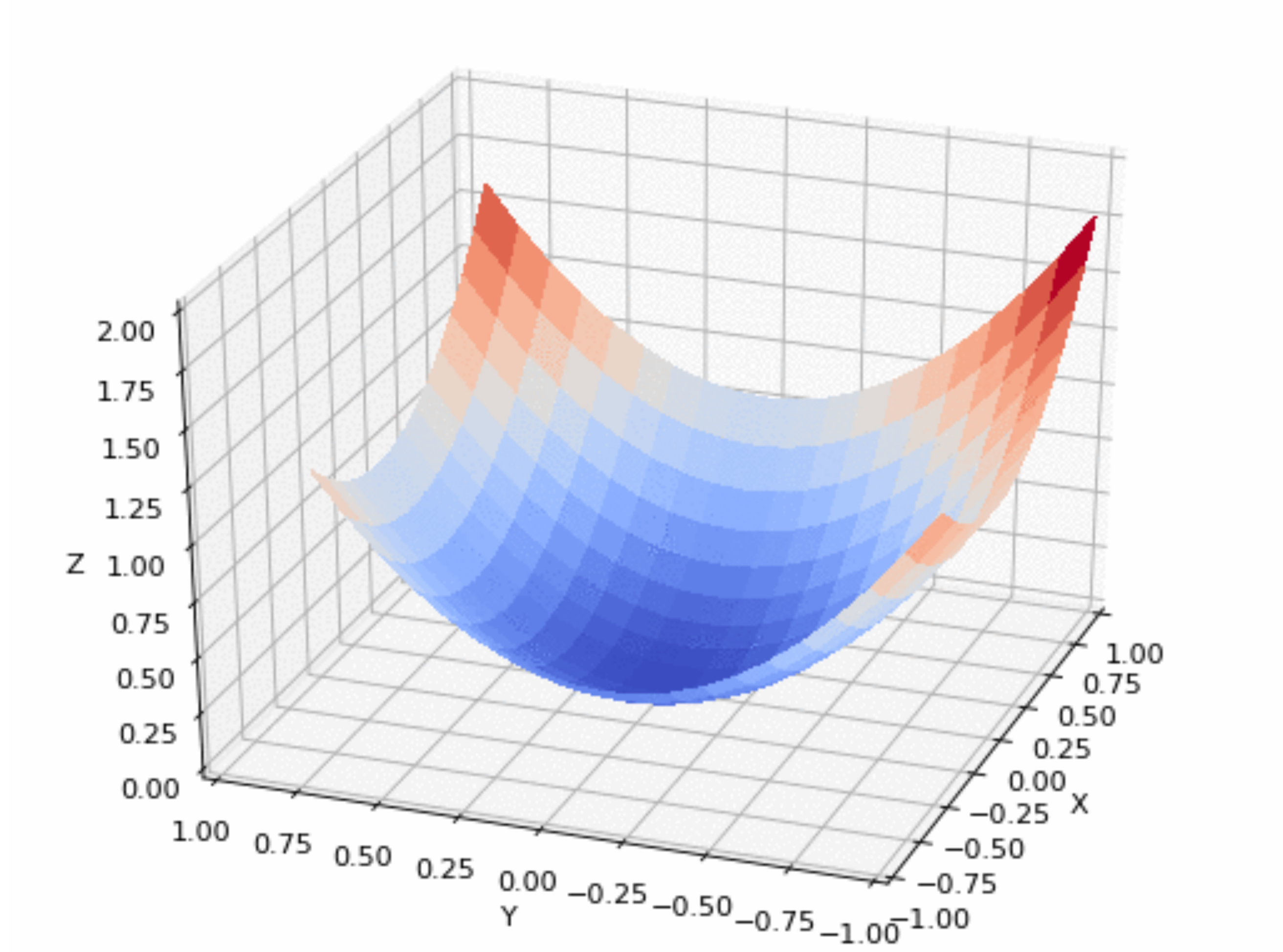
- Draw samples from model
- Compute loss function
- Gradient descent

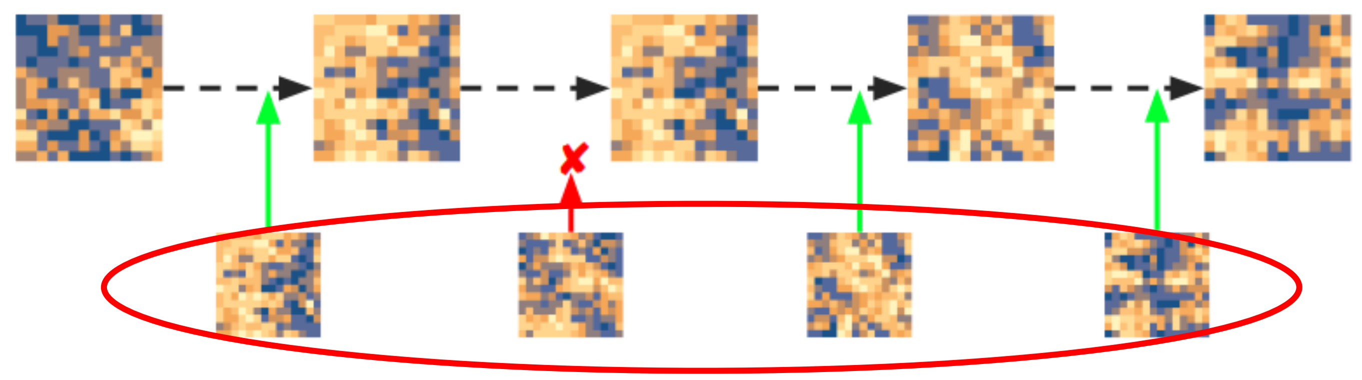
[credits: G. Kanwar]

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

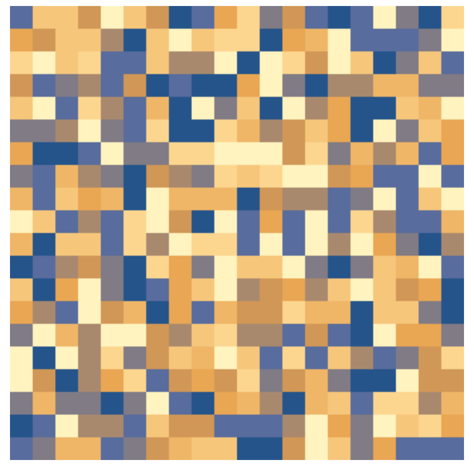
Markov chain using samples from model

[credits: G. Kanwar]

Fernando Romero-López, MIT

Implementing flows for scalar theory

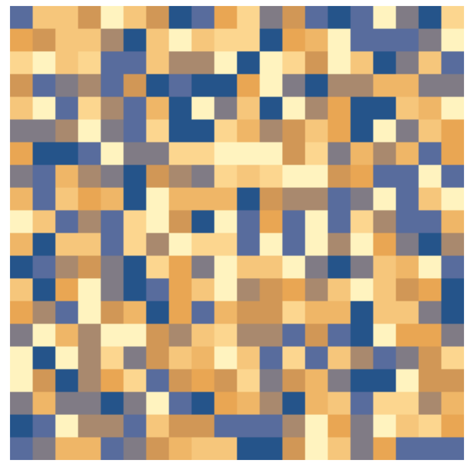
$\phi(x) \in (-\infty, +\infty)$



Fernando Romero-López, MIT

Implementing flows for scalar theory

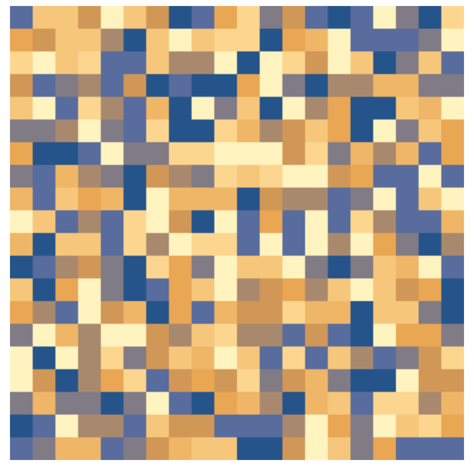
$\phi(x) \in (-\infty, +\infty)$



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

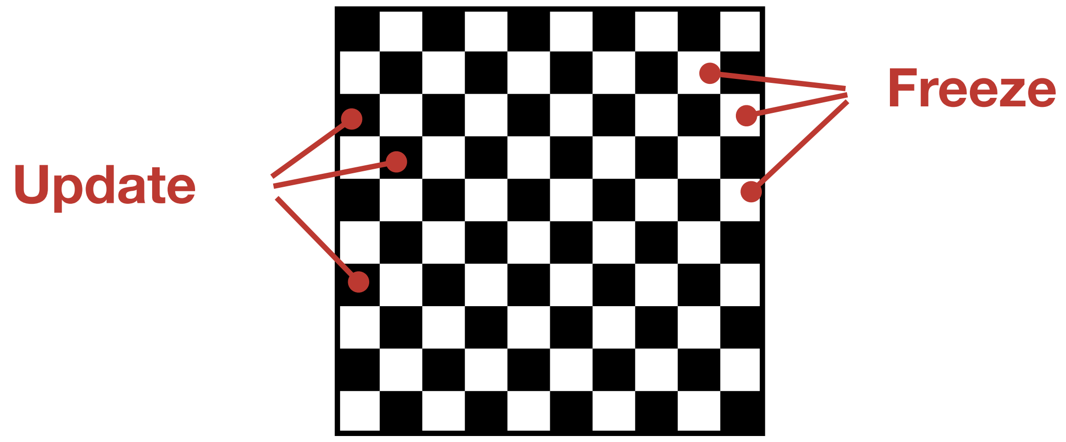
Implementing flows for scalar theory

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



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

"Masking pattern" m defines subsets.

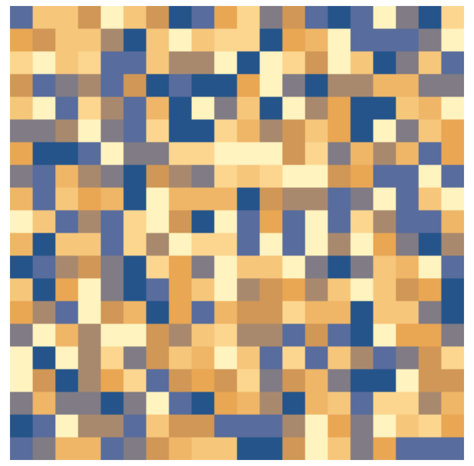


$$\phi' = \begin{cases} \phi'_{\text{frozen}} = \phi_{\text{frozen}} \\ \phi'_{\text{active}} = h(\phi_{\text{frozen}}) \times \phi_{\text{active}} + t \end{cases}$$

Fernando Romero-López, MIT

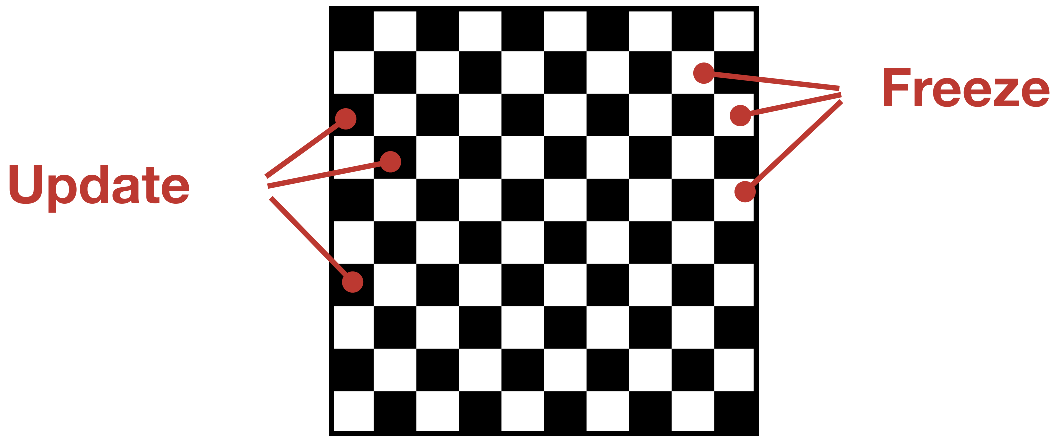
Implementing flows for scalar theory

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



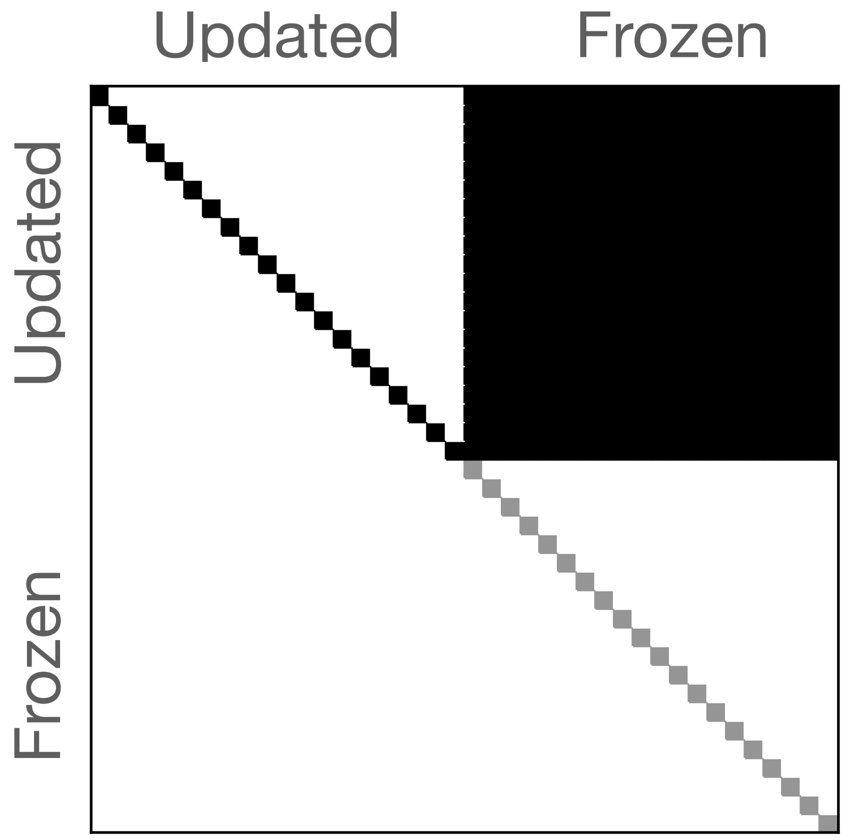
Idea: Each layer acts on a subset of components, conditioned only on the complimentary subset.

"Masking pattern" m defines subsets.



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

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

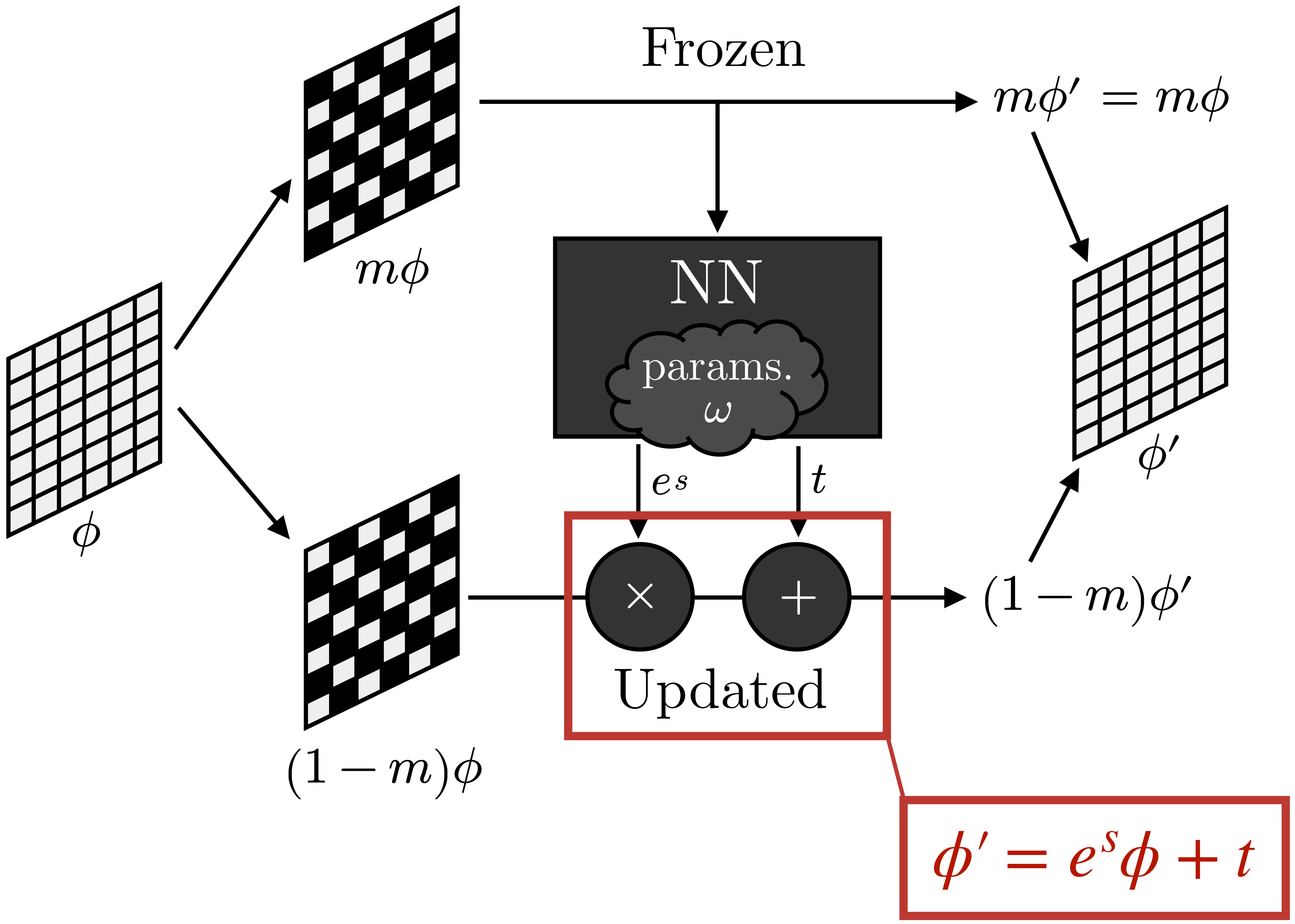


Fernando Romero-López, MIT

Implementing flows for scalar theory

Non-volume presering coupling layer:

[Dinh, Sohl-Dickstein, Bengio 1605.08803]



Fernando Romero-López, MIT

Flows for fermonic gauge theories

Flows for fermonic gauge theories

- **Straightforward approach consists on integrating out fermions:**

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

Fernando Romero-López, MIT

Flows for fermonic gauge theories

- **Straightforward approach consists on integrating out fermions:**

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

Fernando Romero-López, MIT

Flows for fermonic gauge theories

- Straightforward approach consists on integrating out fermions:

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

- Full action can be expressed only in terms of gauge variables:

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

"exact determinant"

- One can use gauge-equivariant architectures from pure gauge models:

[Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan 2003.06413]
[Boyda, Kanwar, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan 2008.05456]

Fernando Romero-López, MIT

Training and using the models

- Train to minimize e.g. Kullback-Leibler divergence:

$$D_{\mathrm{KL}}(q||p) = \int d\phi\, q(\phi)[\log q(U) + S(\phi)] + (\mathrm{const})$$

- Self-training: 1. Draw samples from the model to measure sample mean of $[\log q(U) + S(\phi)]$

  2. Gradient-based methods to optimize model parameters (e.g. Adam optimizer)
  [Kingma, Ba, arXiv:1412.6980]

Fernando Romero-López, MIT

Training and using the models

- Train to minimize e.g. Kullback-Leibler divergence:

$$D_{\mathrm{KL}}(q||p) = \int d\phi\, q(\phi)[\log q(U) + S(\phi)] + (\mathrm{const})$$

- Self-training: 1. Draw samples from the model to measure sample mean of $[\log q(U) + S(\phi)]$

  2. Gradient-based methods to optimize model parameters (e.g. Adam optimizer)
  [Kingma, Ba, arXiv:1412.6980]

! Trained models are not perfect but exactness is essential.

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

Improving determinant estimators

- For a fixed U, the fermion determinant can be estimated as

$$\det M(U) \propto \int d\phi\, e^{-S_{\rm pf}(\phi,U)} = \int d\phi\, q(\phi \mid U) \frac{e^{-S_{\rm pf}(\phi,U)}}{q(\phi \mid U)} = \left\langle \frac{e^{-S_{\rm pf}(\phi,U)}}{q(\phi \mid U)} \right\rangle_q$$

- By drawing more PF samples at fixed gauge field, one can improve the determinant estimator

$$w_{N_{\rm pf}}(U) = \frac{1}{N_{\rm pf}} \sum_{i=1}^{N_{\rm pf}} \frac{p\left(\phi^{(i)}, U\right)}{q\left(\phi^{(i)}, U\right)}$$

weight of each config for reweighting or Markov Chain

- This does not require reevaluating the observables

Improving determinant estimators

- For a fixed U, the fermion determinant can be estimated as

$$\det M(U) \propto \int d\phi\, e^{-S_{\rm pf}(\phi,U)} = \int d\phi\, q(\phi \mid U) \frac{e^{-S_{\rm pf}(\phi,U)}}{q(\phi \mid U)} = \left\langle \frac{e^{-S_{\rm pf}(\phi,U)}}{q(\phi \mid U)} \right\rangle_q$$

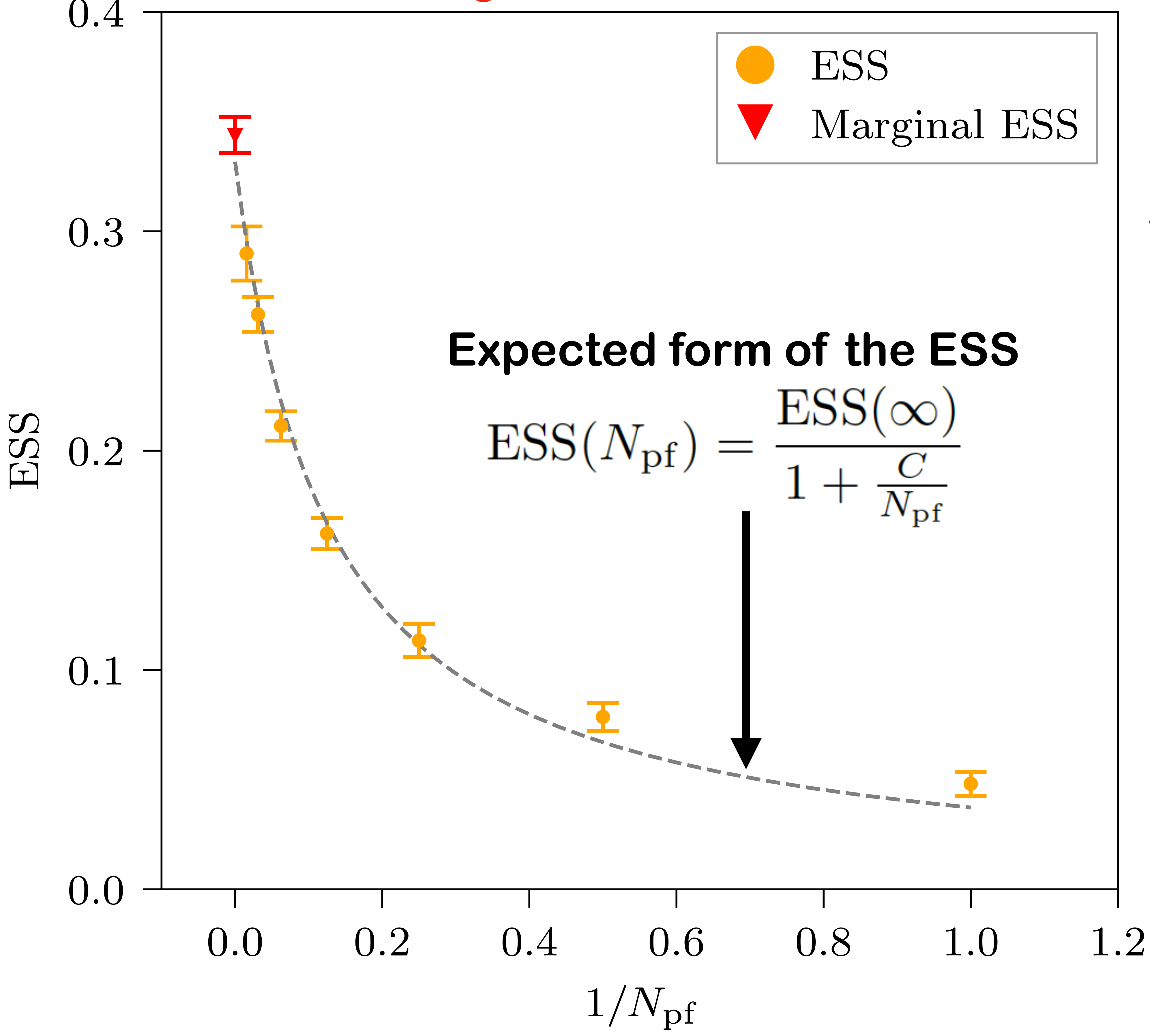
- By drawing more PF samples at fixed gauge field, one can improve the determinant estimator

$$w_{N_{\rm pf}}(U) = \frac{1}{N_{\rm pf}} \sum_{i=1}^{N_{\rm pf}} \frac{p\left(\phi^{(i)}, U\right)}{q\left(\phi^{(i)}, U\right)}$$

weight of each config for reweighting or Markov Chain

- This does not require reevaluating the observables

Numerical demonstration in the Schwinger model



Fernando Romero-López, MIT