Flow-based models for lattice ensemble generation

Gurtej Kanwar
University of Bern

Massachusetts Institute of Technology

Based on …

… flow-based sampling for lattice QFT:

[Albergo, GK, Shanahan **PRD100 (2019) 034515**]
[Albergo, Boyda, Hackett, GK, Cranmer, Racanière, Rezende, Shanahan **2101.08176**]
[Albergo, GK, Racanière, Rezende, Urban, Boyda, Cranmer, Hackett, Shanahan **2106.05934**]
[Hackett, Hsieh, Albergo, Boyda, Chen, Chen, Cranmer, GK, Shanahan **2107.00734**]

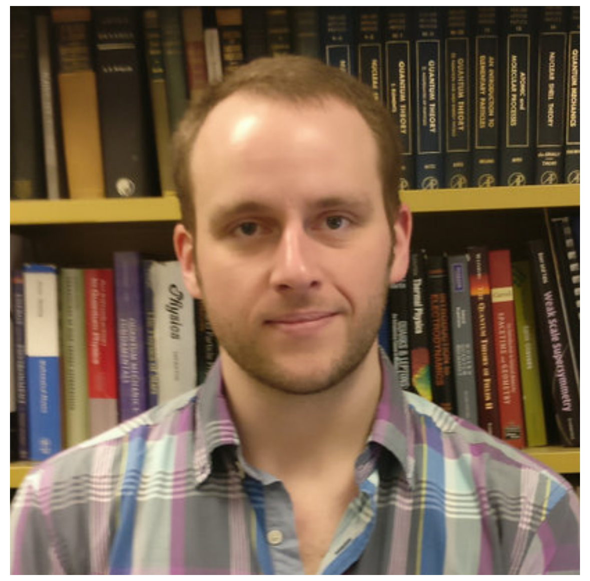
… flows for compact vars & lattice gauge theories:

[GK, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan **PRL125 (2020) 121601**]
[Rezende, Papamakarios, Racanière, Albergo, GK, Shanahan, Cranmer **ICML (2020) 2002.02428**]
[Boyda, GK, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan **PRD103 (2021) 074504**]

ML for HEP, on and off the lattice | ECT* Trento | Sept 29, 2021

Massachusetts Institute of Technology

Phiala Shanahan



Dan Hackett

Denis Boyda

DeepMind

Sébastien Racanière

Danilo Rezende

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Julian Urban

NEW YORK UNIVERSITY

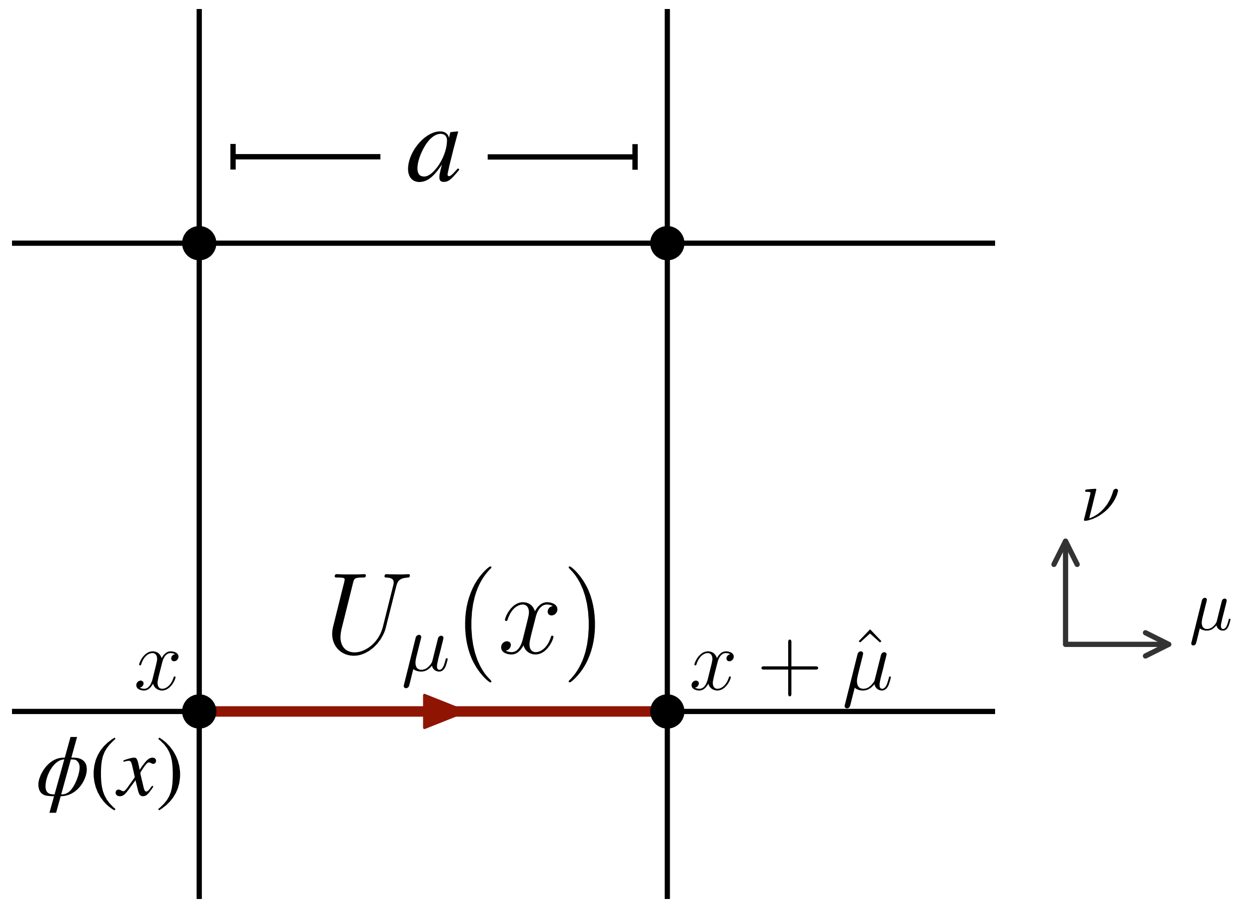
Kyle Cranmer

Michael Albergo

30 second Lattice QFT primer

Lattice discretization:

- Gauge field discretized on links $U_\mu(x) \in G$ [e.g. $\mathrm{SU}(3)$]
- Other fields $\phi(x)$ discretized to live on sites x

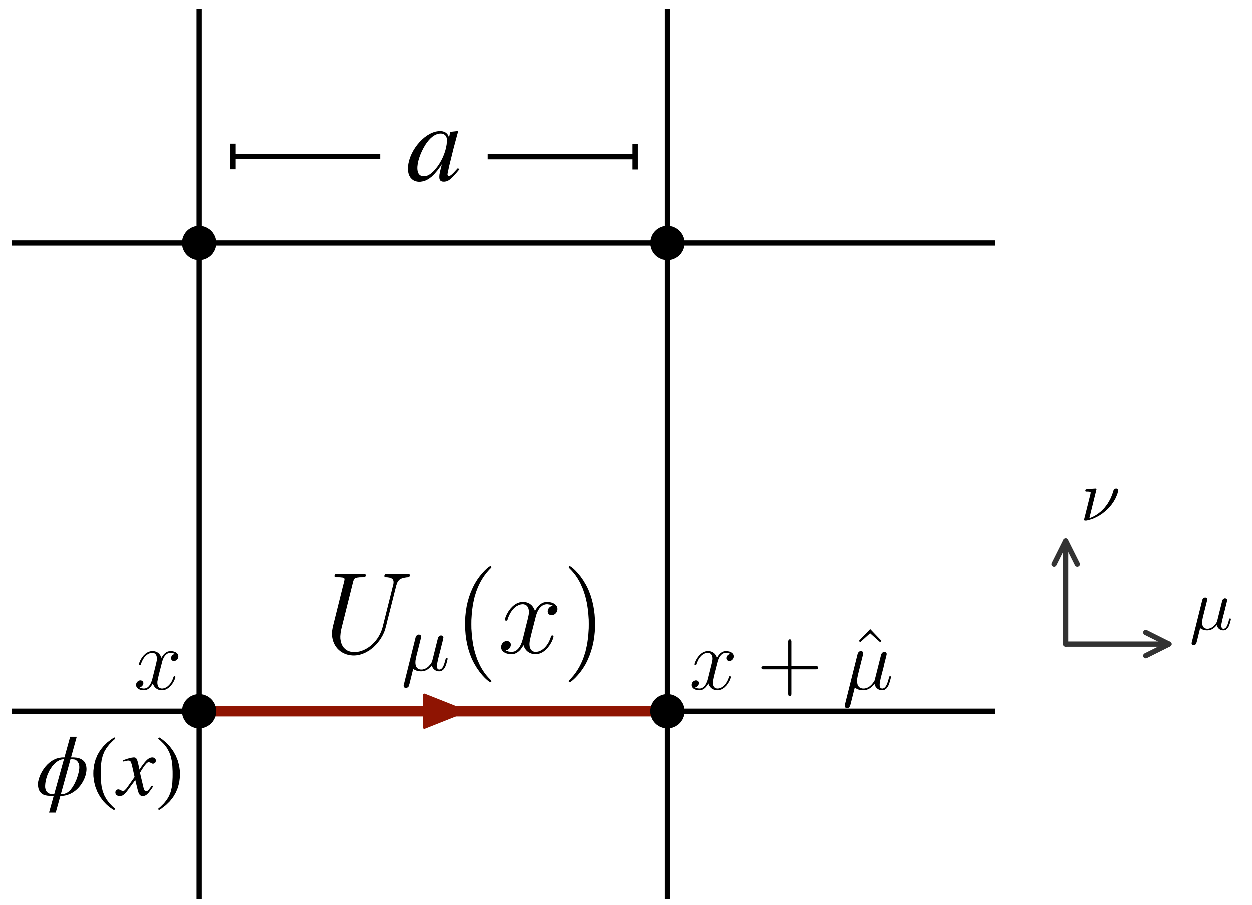


Lattice path integral → observables

Caveats:

- Euclidean spacetime $t \to i\tau$
- Discretization effects (must take $a \to 0$)

30 second Lattice QFT primer



Lattice discretization:

- Gauge field discretized on links $U_\mu(x) \in G$ [e.g. $\mathrm{SU}(3)$]
- Other fields $\phi(x)$ discretized to live on sites x

Lattice path integral → observables

Vaccum/thermal expt. value of quantum operator

✓ In principle tractable integral

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}U \, \mathcal{O}(U) \, e^{-S(U)}$$

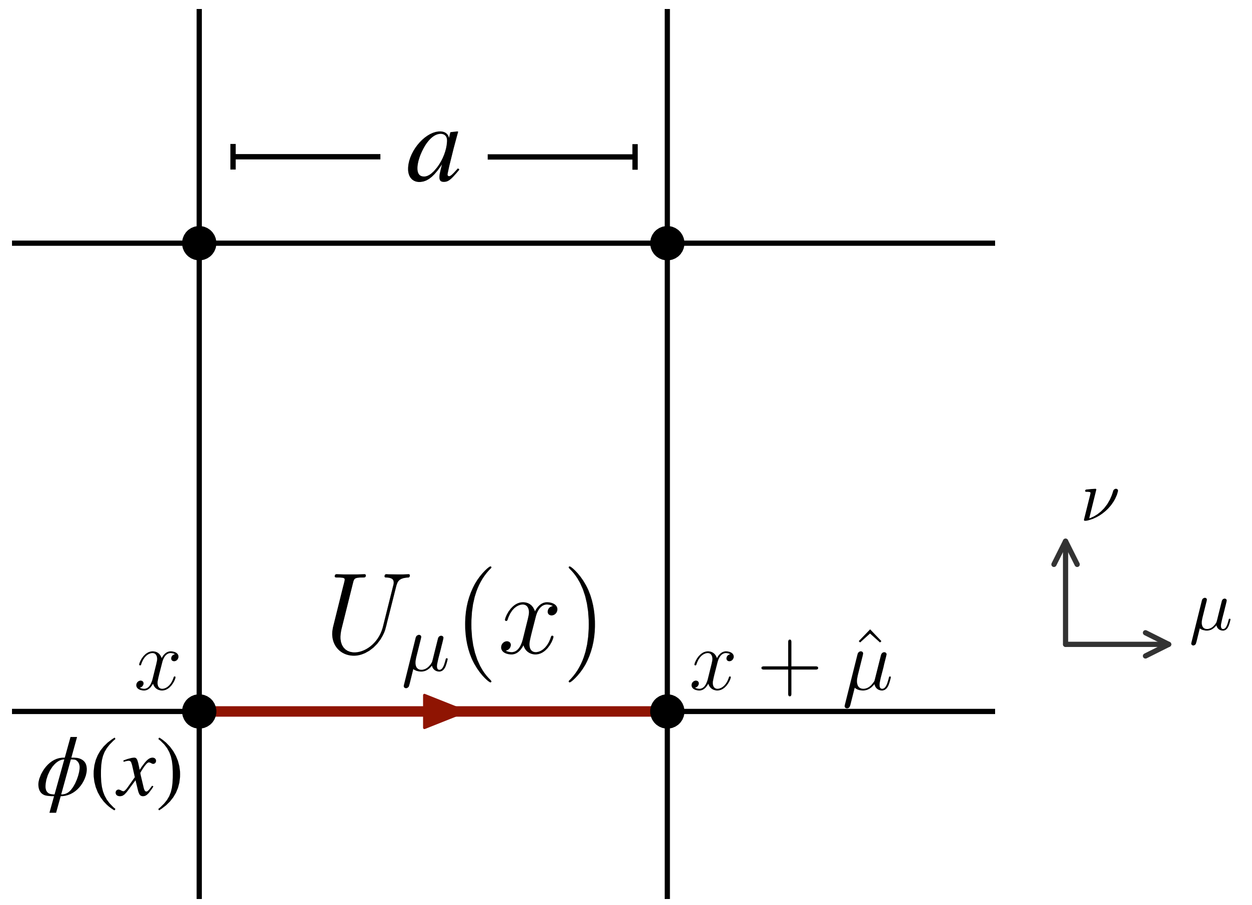
Caveats:

- Euclidean spacetime $t \to i\tau$
- Discretization effects (must take $a \to 0$)

30 second Lattice QFT primer

Lattice discretization:

- Gauge field discretized on links $U_\mu(x) \in G$ [e.g. $\mathrm{SU}(3)$]
- Other fields $\phi(x)$ discretized to live on sites x



Lattice path integral → observables

Vaccum/thermal expt. value of quantum operator

✓ In principle tractable integral

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}U \, \mathcal{O}(U) \, e^{-S(U)}$$

$$Z = \int \mathcal{D}U \, e^{-S(U)}, \qquad \int \mathcal{D}U = \prod_{x,\mu} \int dU_\mu(x)$$

Normalizing constant

Path integral measure

Caveats:

- Euclidean spacetime $t \to i\tau$
- Discretization effects (must take $a \to 0$)

Importance sampling: the workhorse of LQFT

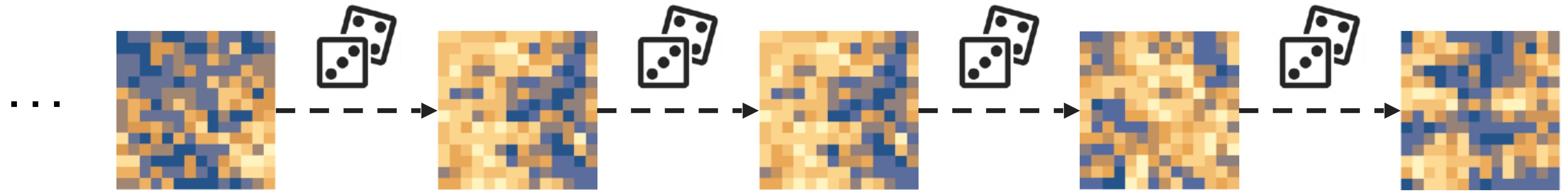
Monte Carlo sampled ensembles allow estimates of (many) QFT observables

Desired continuum QFT quantity

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}U \, \mathcal{O}(U) \, e^{-S(U)} \quad \rightarrow \quad \langle \mathcal{O} \rangle \approx \frac{1}{n} \sum_{i=1}^{n} \mathcal{O}[U_i] \quad \rightarrow \quad U_i \sim p(U) = e^{-S(U)}/Z$$

Target distribution

Markov chain Monte Carlo (MCMC)



Example: MCMC for scalar field configurations

- **Asymptotically** converges to distribution p
- However: States of the chain are “autocorrelated”
- Skip thermalization steps, ensemble “thinned” to a subset

Importance sampling: the workhorse of LQFT

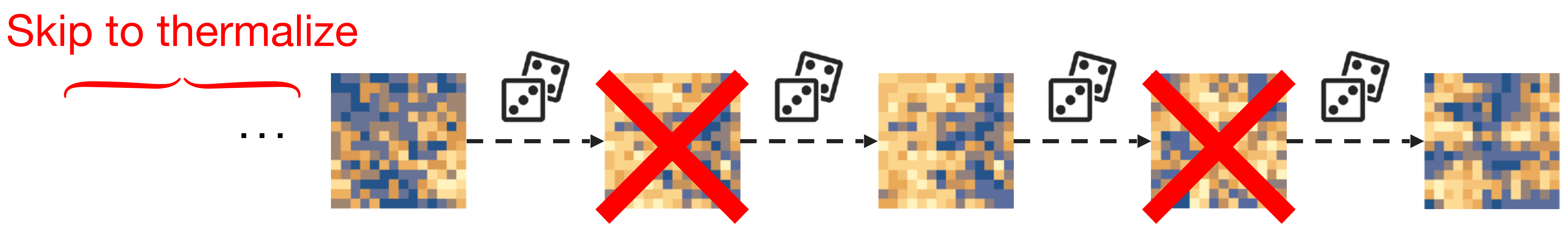
Monte Carlo sampled ensembles allow estimates of (many) QFT observables

Desired continuum QFT quantity

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}U \, \mathcal{O}(U) \, e^{-S(U)} \quad \longrightarrow \quad \langle \mathcal{O} \rangle \approx \frac{1}{n} \sum_{i=1}^{n} \mathcal{O}[U_i] \quad \longrightarrow \quad U_i \sim p(U) = e^{-S(U)}/Z$$

Target distribution

Markov chain Monte Carlo (MCMC)



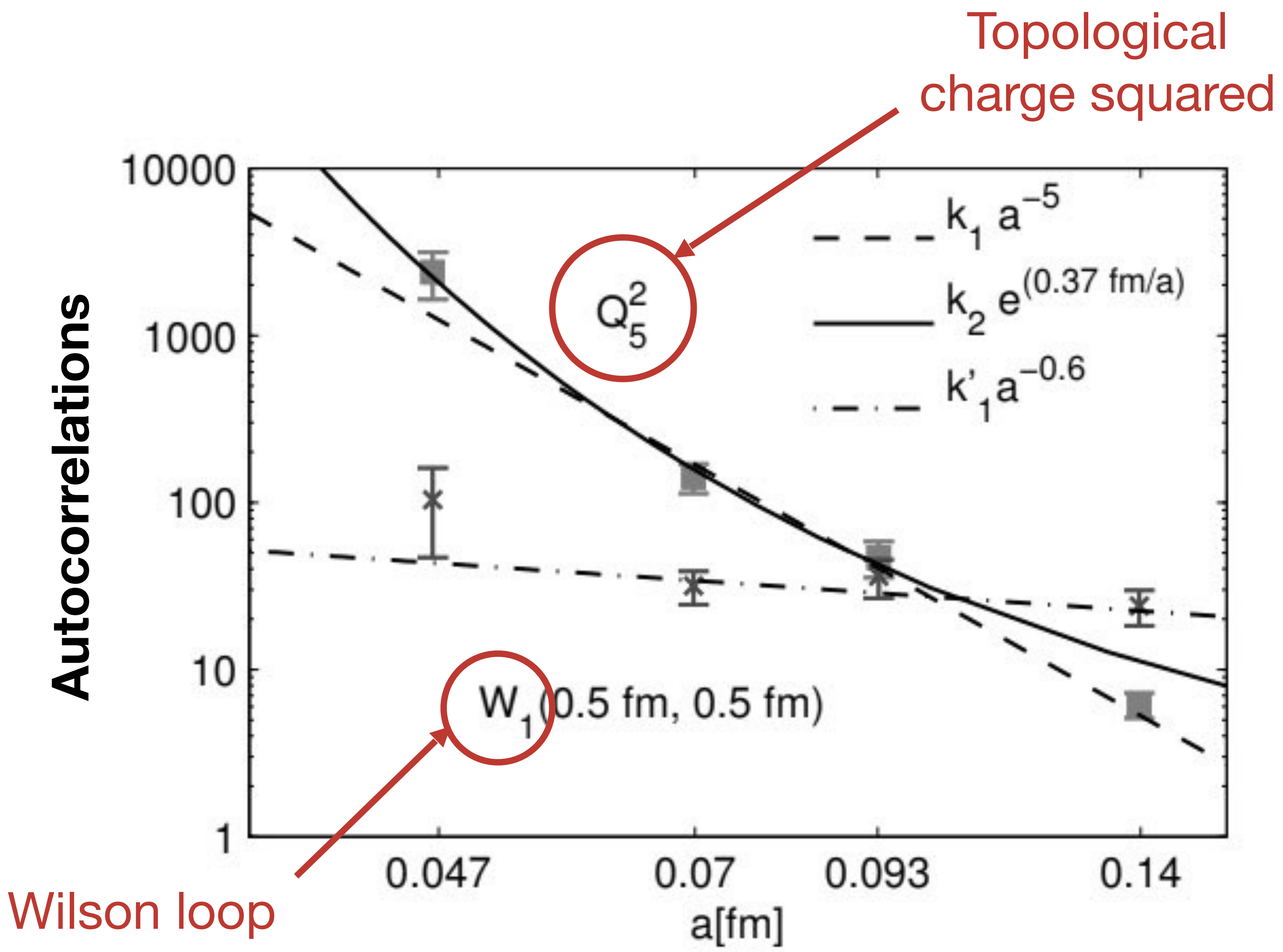
Example: MCMC for scalar field configurations

- **Asymptotically** converges to distribution p
- However: States of the chain are “autocorrelated”
- Skip thermalization steps, ensemble “thinned” to a subset

Critical slowing down (CSD)

Local/diffusive Markov chains inefficient as $a \to 0$

- Correlation length grows in lattice units, but information transfer is local
- Rare to update entire field coherently



[Schaefer et al. / ALPHA collaboration **NPB845 (2011) 93**]

Critical slowing down: diverging autocorrelations due to local mixing

Topological freezing: Markov chain gets “stuck” in topological sectors

CSD also affects a number of other models:

- CP^{N-1} [Flynn, et al. **1504.06292**]
- O(N) [Frick, et al. **PRL63 (1989) 2613**]
- ϕ^4 [Vierhaus; Thesis, **doi:10.18452/14138**]
- ...

Critical slowing down (CSD)

Local/diffusive Markov chains inefficient as $a \to 0$

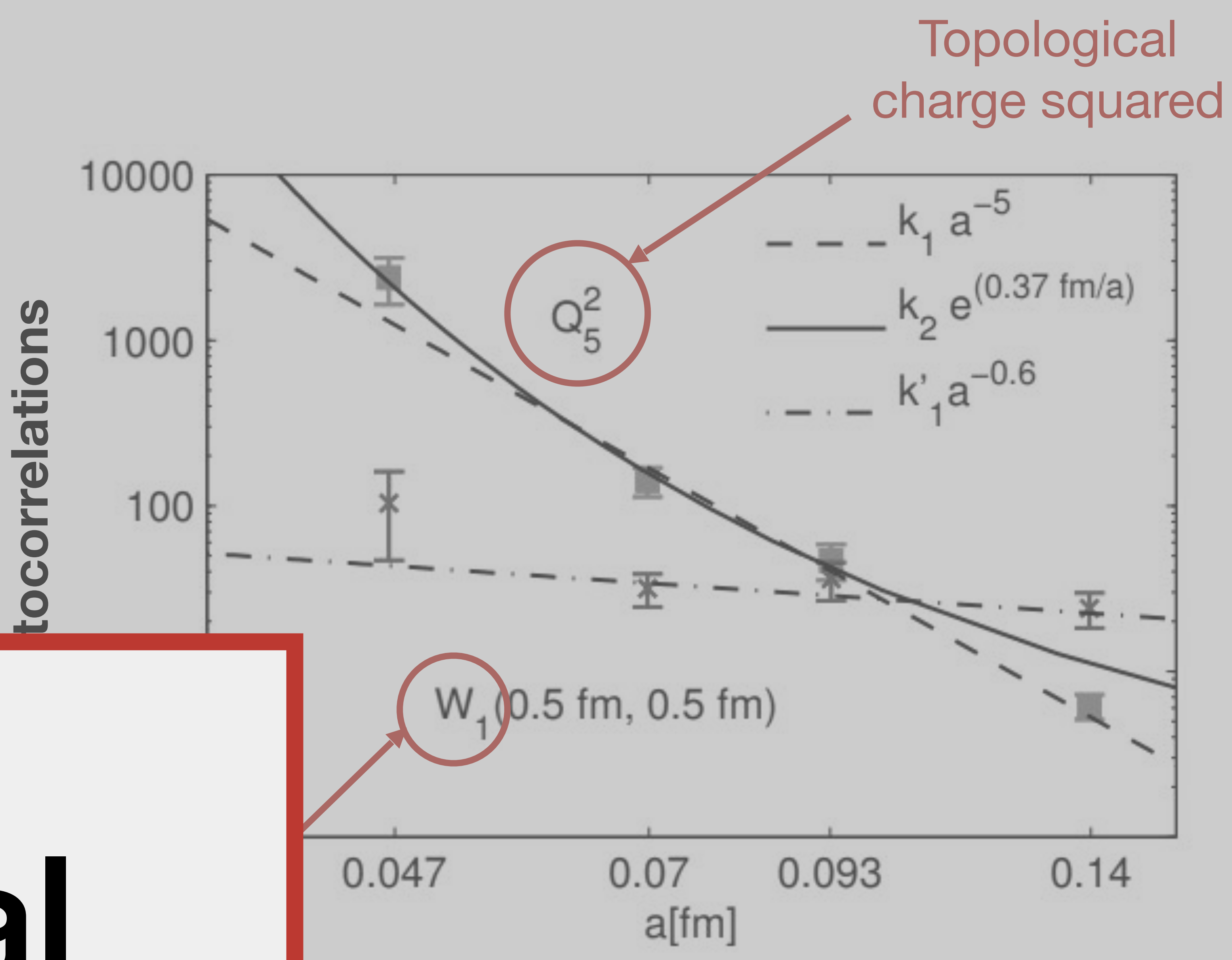
- Correlation length grows in lattice units, but information trans
- Rare to update e

Critical slowing d
due to local mixing

Topological freezing: Markov chain gets "stuck" in topological sectors

CSD & Topological freezing:

Significant obstacles to continuum limit



al. / ALPHA collaboration **NPB845 (2011) 93**]

CSD also affects a number of other models:

- CP^{N-1} [Flynn, et al. **1504.06292**]
- O(N) [Frick, et al. **PRL63 (1989) 2613**]
- ϕ^4 [Vierhaus; Thesis, **doi:10.18452/14138**]
- ...

Circumventing CSD?

Avoid diffusive/local Markov chain updates.

Proposal: Sample from generative ML models.

Caveats:

⚠️ We require **exactness**

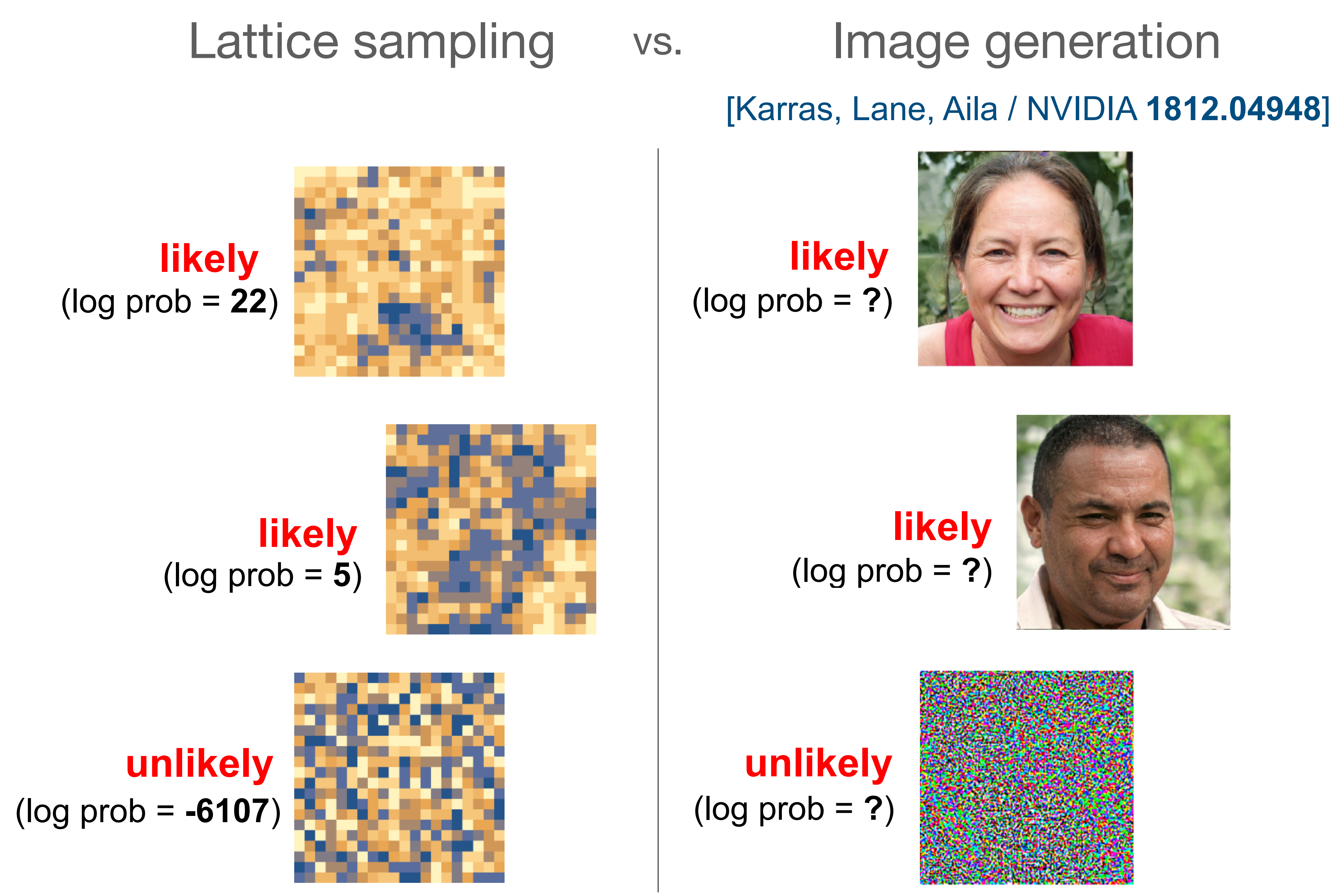
⚠️ Inverted data hierarchy
… $\sim 10^9 - 10^{10}$ DoFs in a config
… $\sim 10^3$ configs

✓ Target probability density $e^{-S(U)}/Z$

✓ Physical symmetries = “flat directions”

We choose **flow-based models.**

AKA ‘normalizing flows’



Circumventing CSD?

Avoid diffusive/local Markov chain updates.

Proposal: Sample from generative ML models.

Caveats:

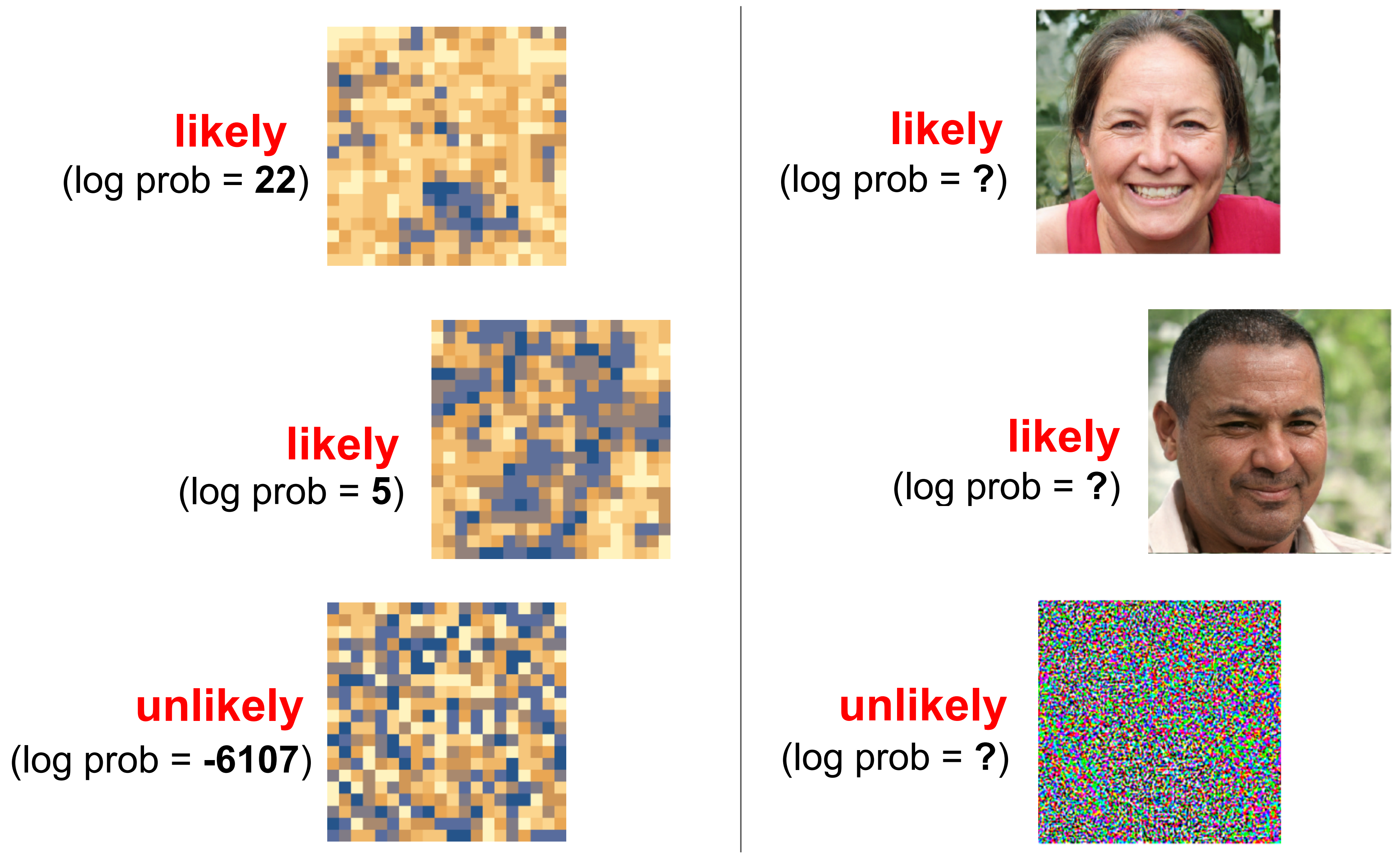
- ⚠️ We require **exactness**
- ⚠️ Inverted data hierarchy
  … ~10^9 – 10^{10} DoFs in a config
  … ~10^3 configs
- ✓ Target probability density $e^{-S(U)}/Z$
- ✓ Physical symmetries = "flat directions"

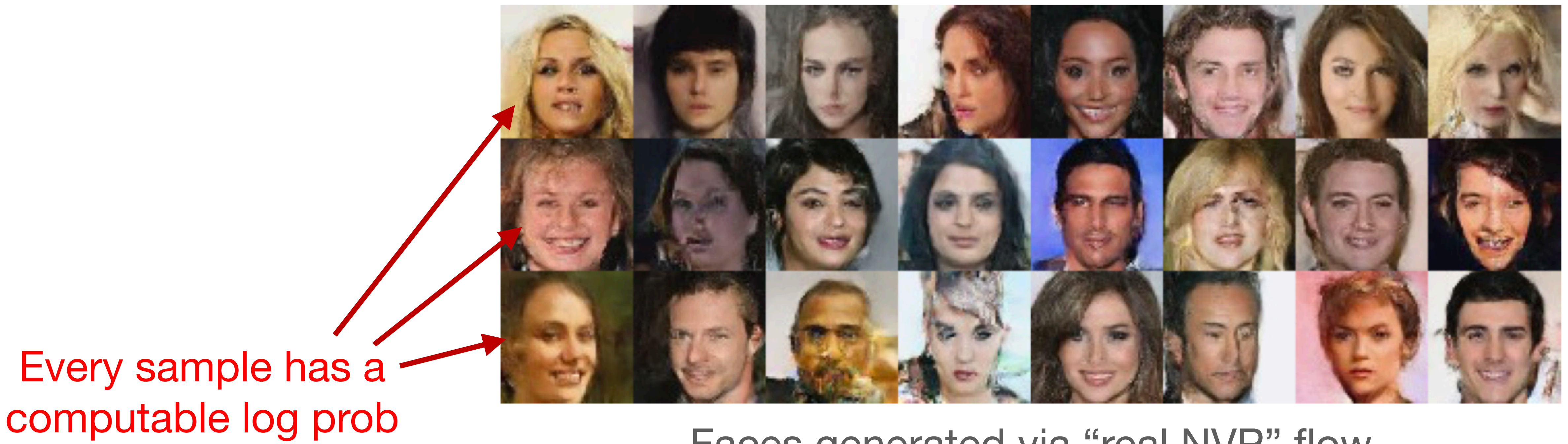
We choose **flow-based models.**

AKA 'normalizing flows'

Lattice sampling vs. Image generation

[Karras, Lane, Aila / NVIDIA **1812.04948**]





Faces generated via "real NVP" flow
[Dinh, Sohl-Dickstein, Bengio **1605.08803**]

A story of symmetries & generative models

(In three parts)

1. Flow-based generative models

2. Gauge symmetry & translational symmetry

3. Fermions & translational symmetry

A story of symmetries & generative models

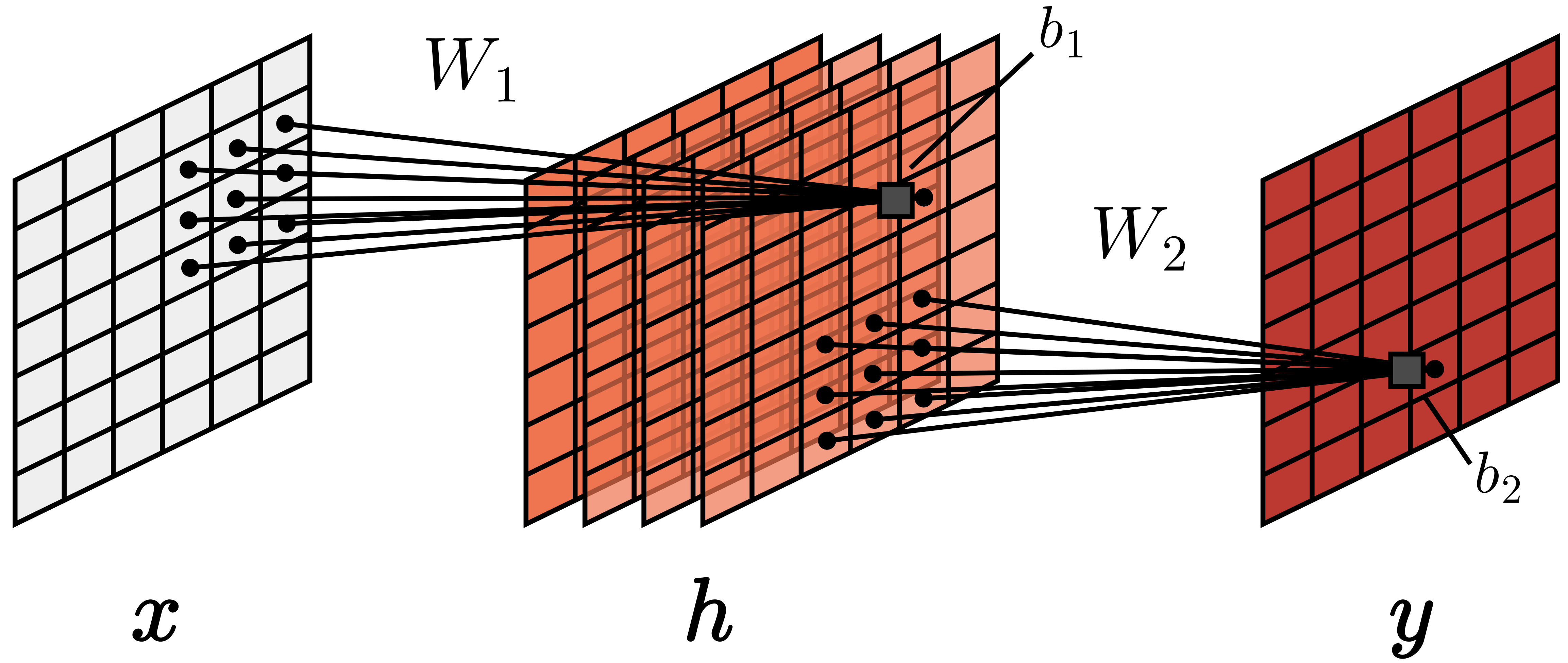
(In three parts)

1. **Flow-based generative models**

2. Gauge symmetry & translational symmetry

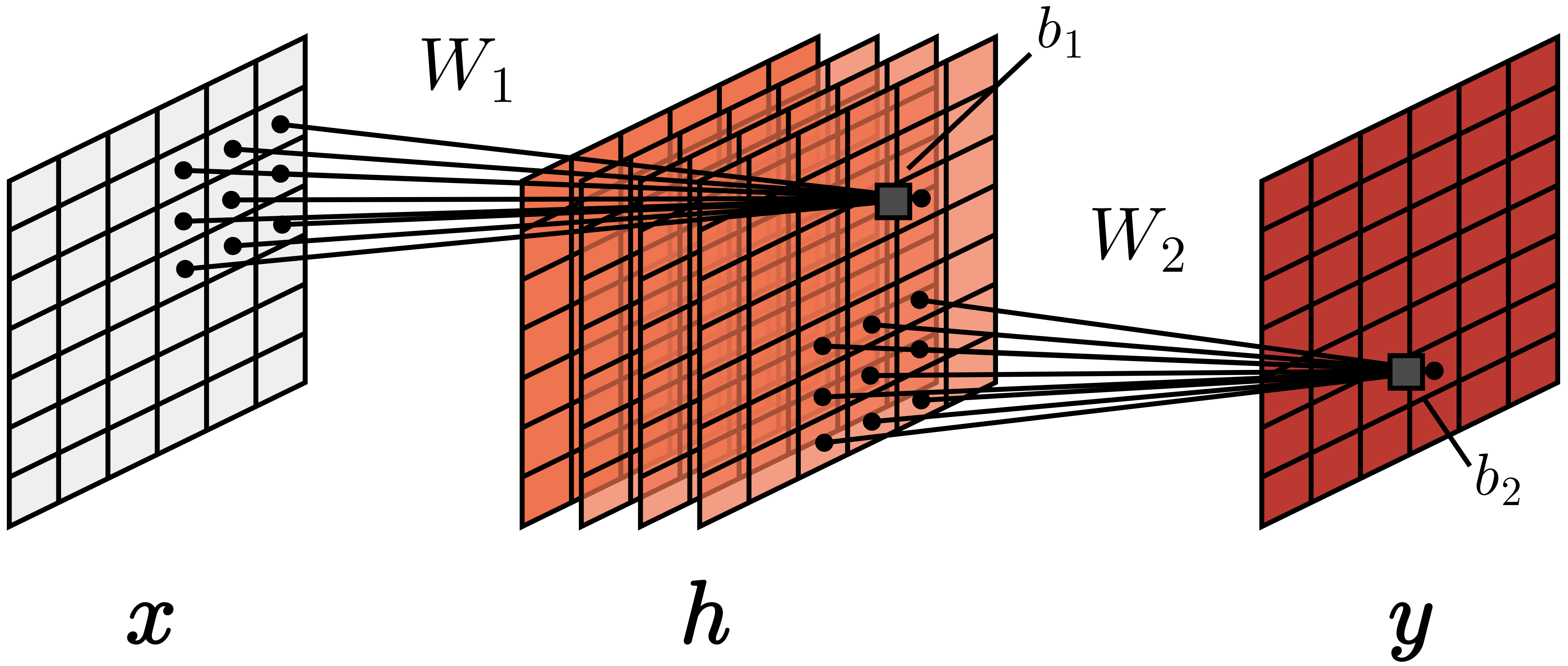
3. Fermions & translational symmetry

Flow-based sampling: Overview



Flow-based sampling: Overview

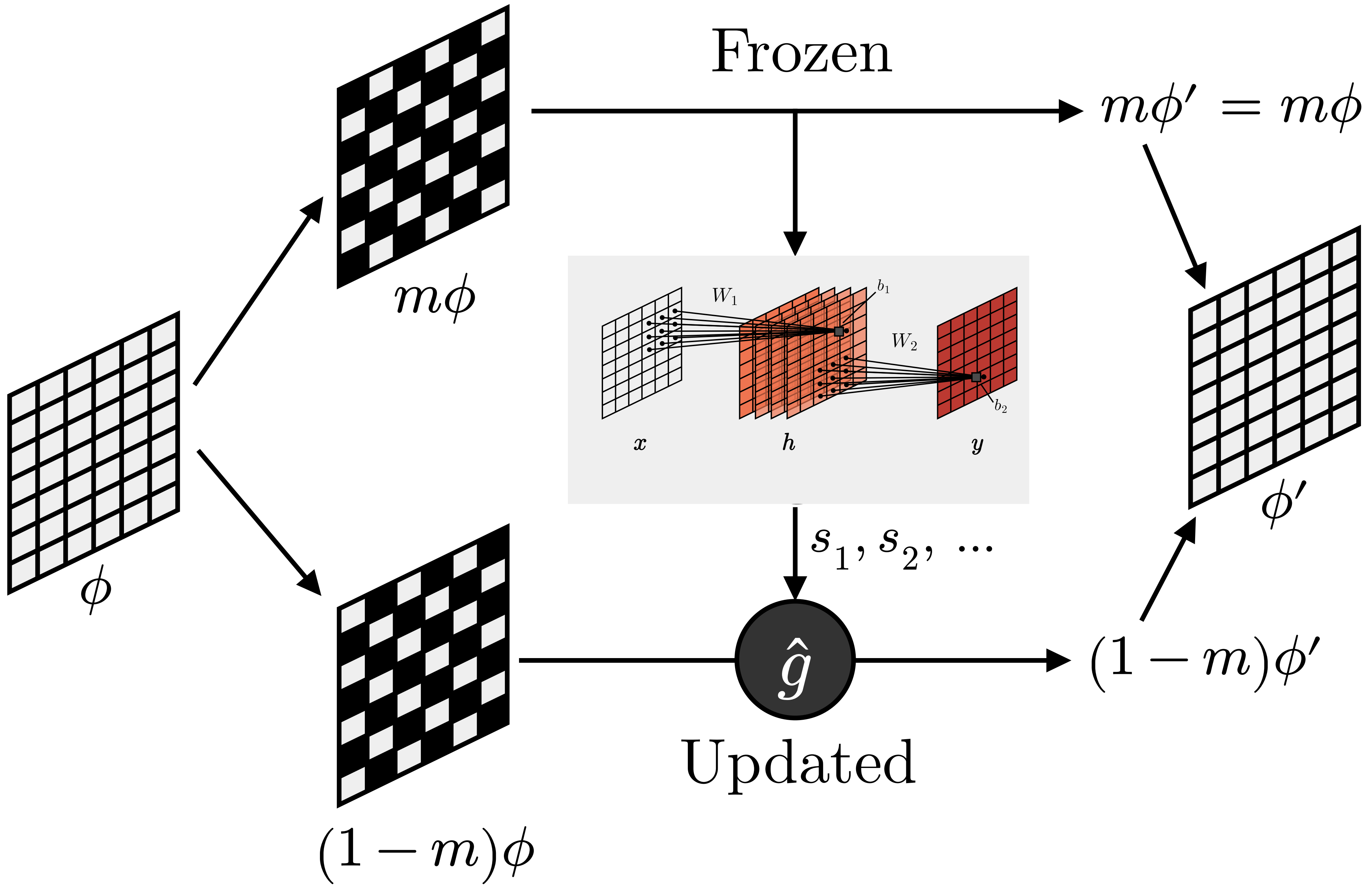
(Convolutional) neural networks: Black-box (local) function approximators



Flow-based sampling: Overview

(Convolutional) neural networks: Black-box (local) function approximators

Coupling layers: Invertible transformations, tractable Jacobian

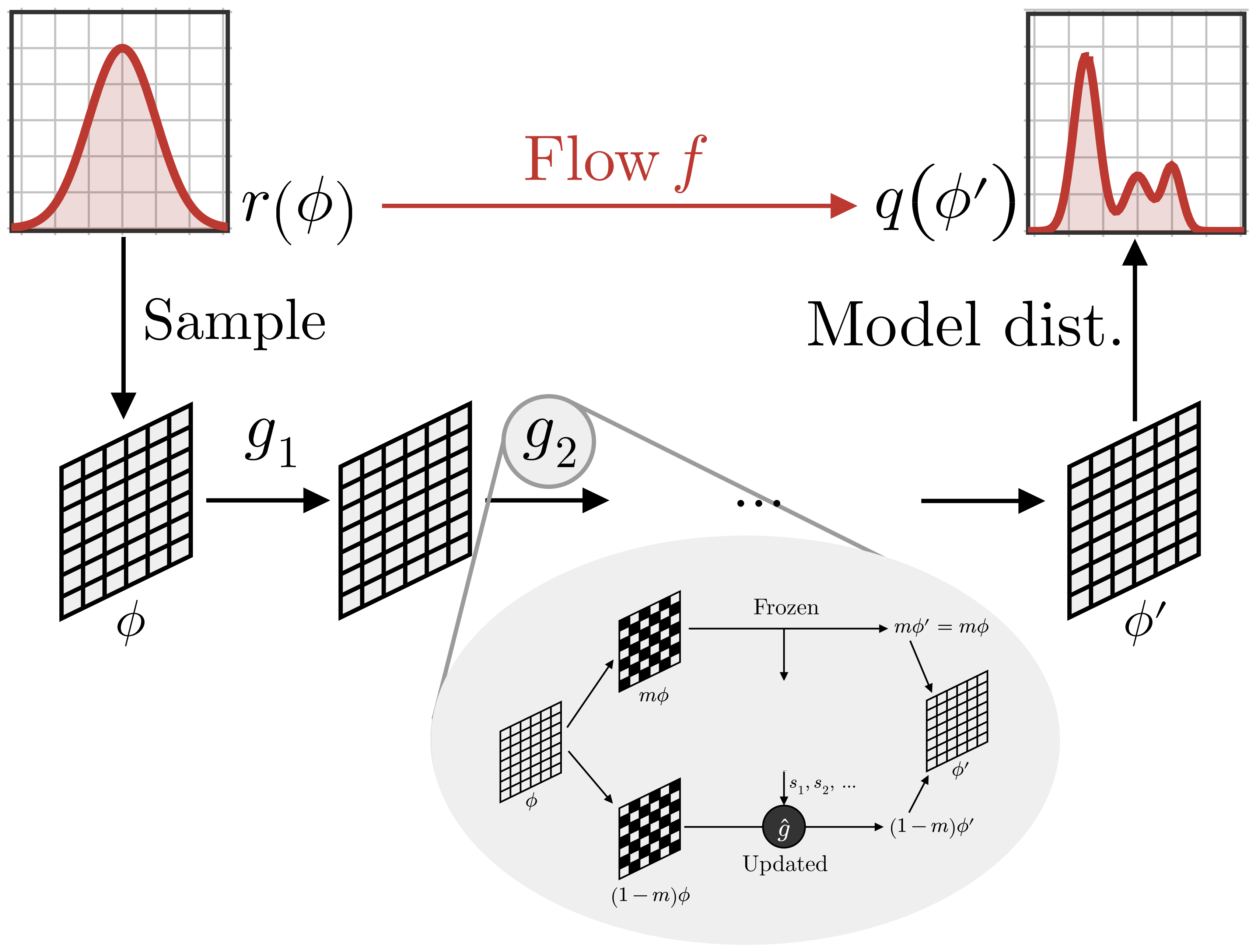


Flow-based sampling: Overview

(Convolutional) neural networks: Black-box (local) function approximators

Coupling layers: Invertible transformations, tractable Jacobian

Flow model: Prior density + flow = **sampleable + computable** output density



$$q(\phi') = r(\phi) \left| \det_{ij} \frac{\partial [f(\phi)]_i}{\partial \phi_j} \right|^{-1}$$

Flow-based sampling: Overview

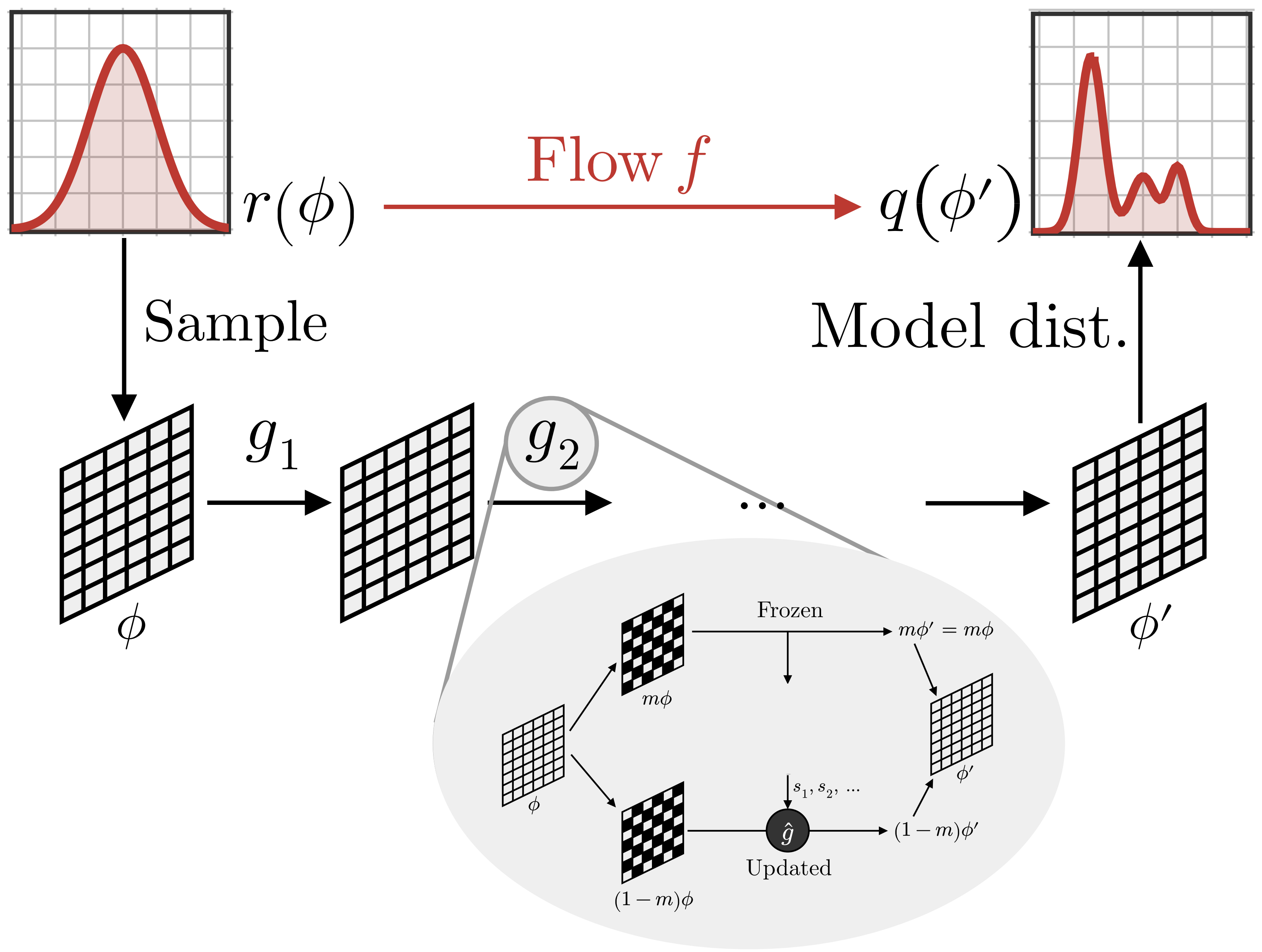
(Convolutional) neural networks: Black-box (local) function approximators

Coupling layers: Invertible transformations, tractable Jacobian

Flow model: Prior density + flow = **sampleable + computable** output density

Training:

- Measure KL divergence
- Apply gradient-based opt



$$q(\phi') = r(\phi) \left| \det_{ij} \frac{\partial [f(\phi)]_i}{\partial \phi_j} \right|^{-1}$$

Flow-based sampling: Overview

(Convolutional) neural networks: Black-box (local) function approximators

Coupling layers: Invertible transformations, tractable Jacobian

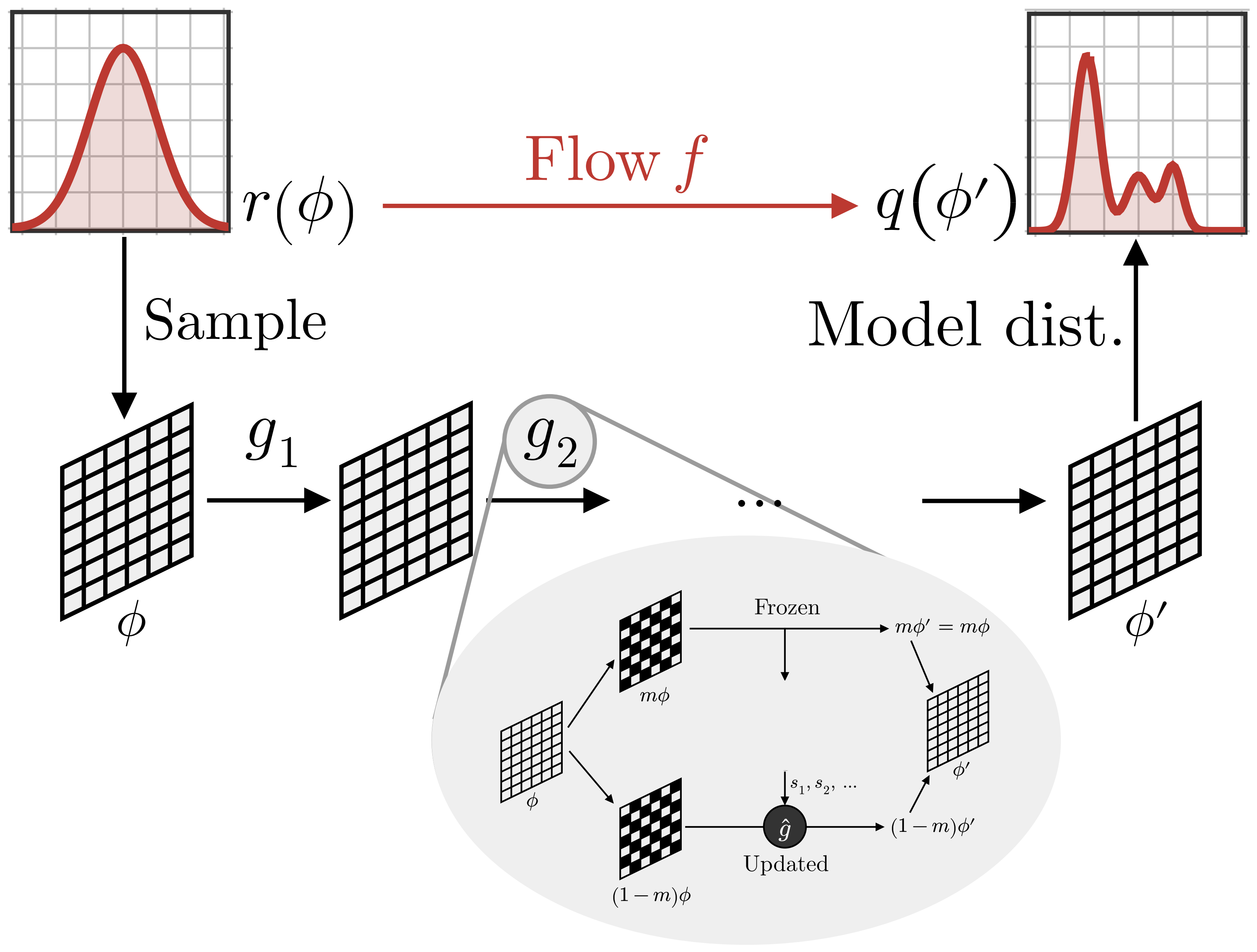
Flow model: Prior density + flow = **sampleable + computable** output density

Training:

- Measure KL divergence
- Apply gradient-based opt

Exactness:

- Use $q(\phi')$ and $p(\phi')$to correct approximation



$$q(\phi') = r(\phi) \left| \det_{ij} \frac{\partial [f(\phi)]_i}{\partial \phi_j} \right|^{-1}$$

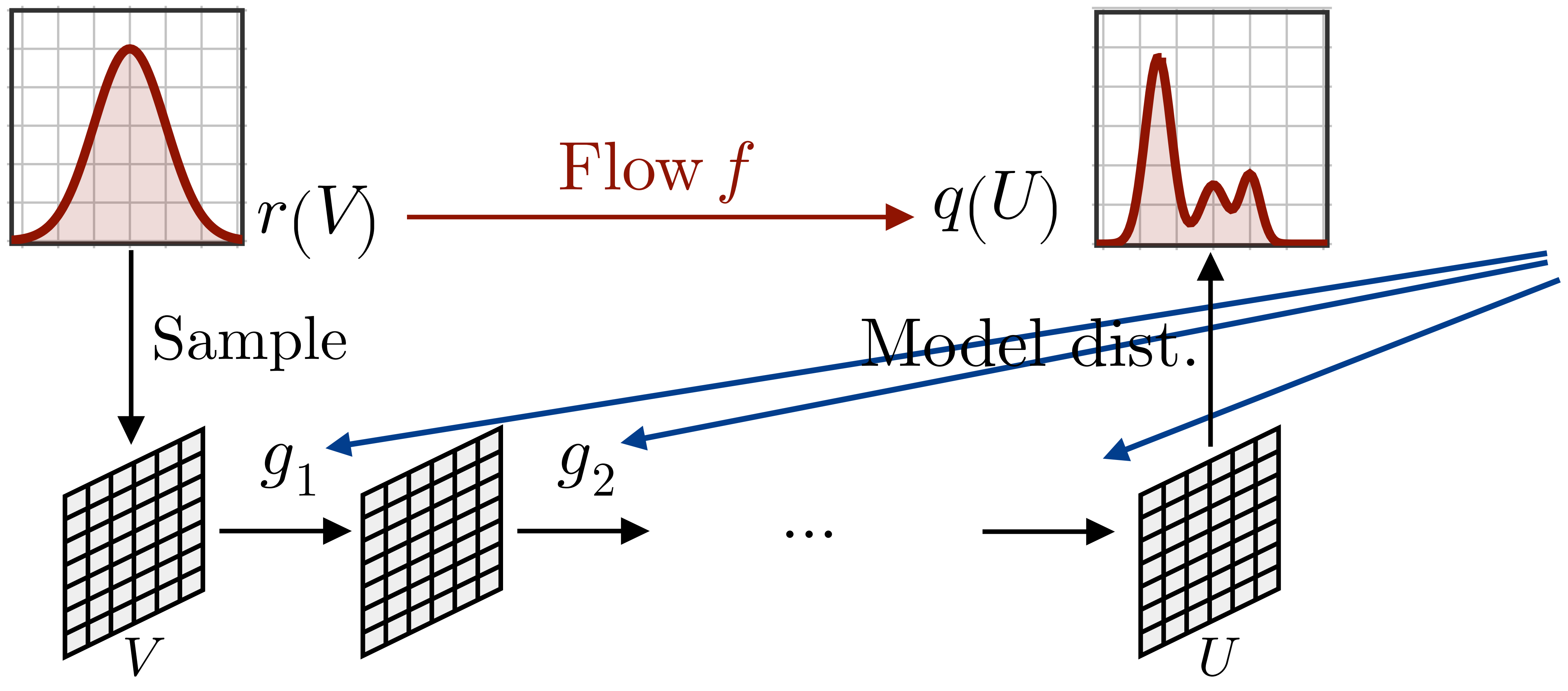
$$q(U) = r(V)\left|\det_{ij}\frac{\partial[f(V)]_i}{\partial V_j}\right|^{-1}$$

Defining the flow function

The "flow" f must be **invertible** and have **tractable Jacobian determinant**

- For LQFT, don't know what f needs to be *a priori*
- Construct expressive parameterized ansatz and optimize it

Key to expressivity — **Use composition.**



Each layer is **invertible**, has **tractable Jac.**

Simple individual layers combine to give complex transformations.

Coupling layers

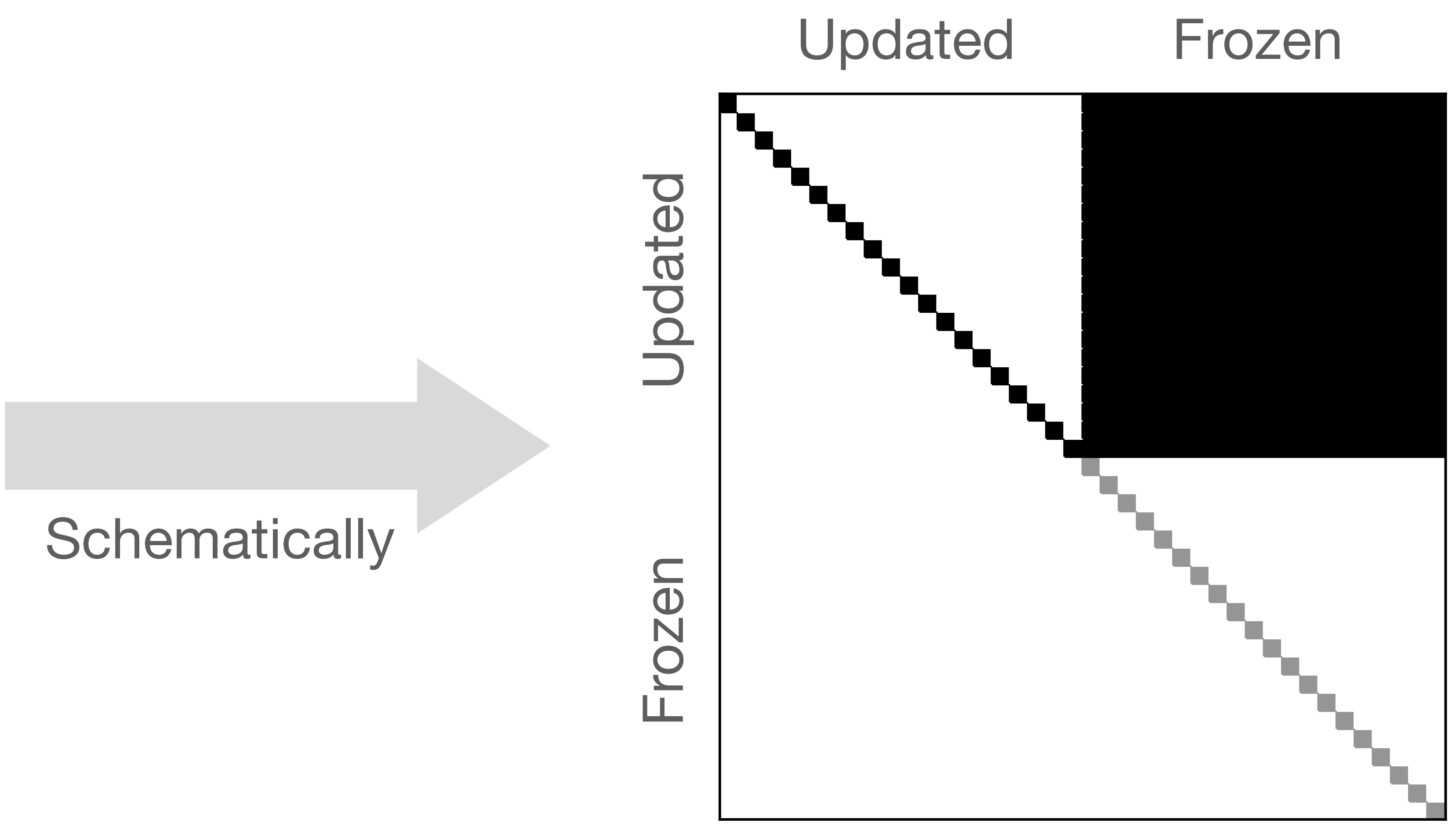
Similar to leapfrog integrator

Idea: Construct each g to act on a **subset** of components, conditioned only on the complimentary subset. “Masking pattern” m defines subsets.

→ Jacobian is explicitly upper-triangular (get LDJ from diag elts)

“Updated” i ($m_i = 0$) “Frozen” i ($m_i = 1$)

$$\frac{\partial[g(V)]_i}{\partial V_j} = \left(\begin{array}{ccc|ccc} \frac{\partial[g(V)]_1}{\partial V_1} & & & & & \\ & \frac{\partial[g(V)]_2}{\partial V_2} & & & \text{(nonzero)} & \\ & & \ddots & & & \\ \hline & & & 1 & & \\ & 0 & & & 1 & \\ & & & & & \ddots \end{array}\right)$$



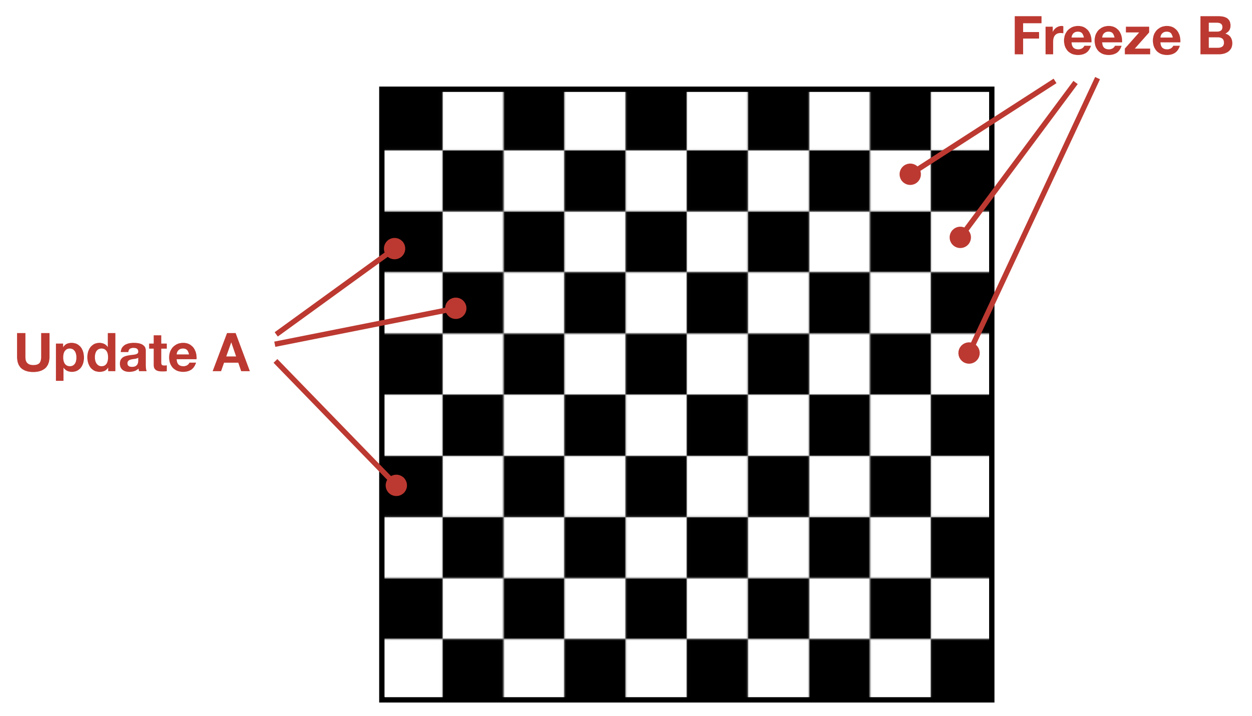
→ Invertible if each diag component invertible, $\partial[g(V)]_i/\partial V_i \neq 0$.

Example: RNVP for scalar fields

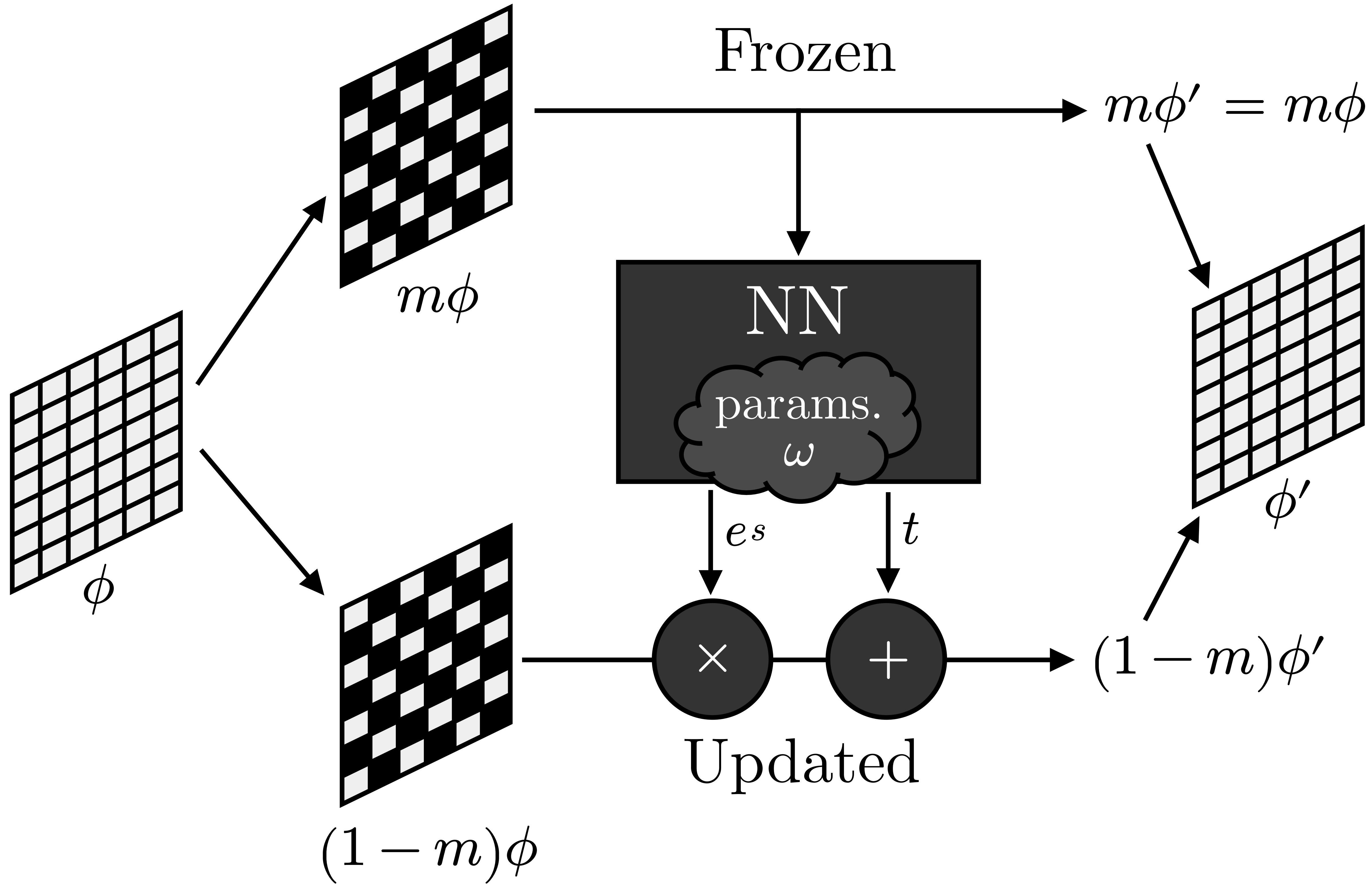
Scalar field $\phi(x) \in \mathbb{R}$ $\approx$ grayscale image

Real NVP coupling layer:

[Dinh, Sohl-Dickstein, Bengio **1605.08803**]



Checkerboard masking pattern m

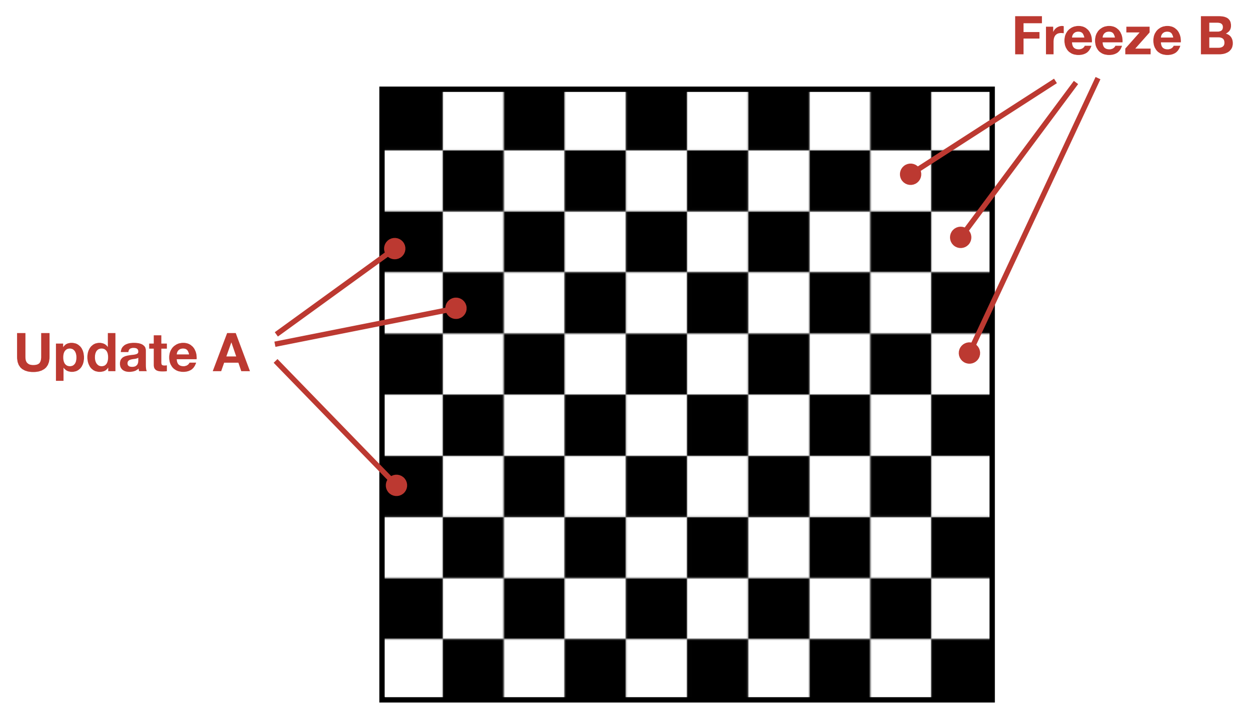


Example: RNVP for scalar fields

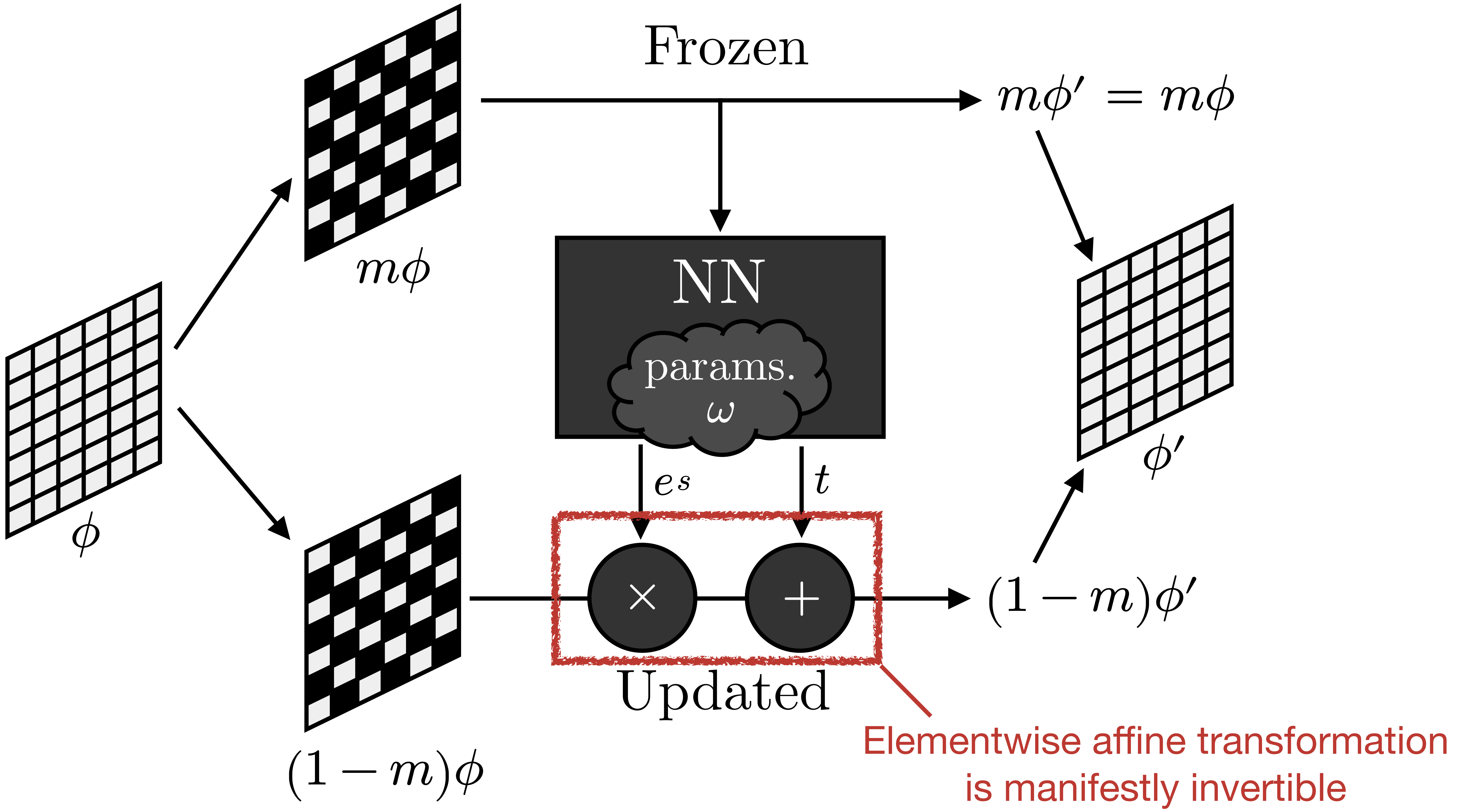
Scalar field $\phi(x) \in \mathbb{R} \approx$ grayscale image

Real NVP coupling layer:

[Dinh, Sohl-Dickstein, Bengio **1605.08803**]



Checkerboard masking pattern m

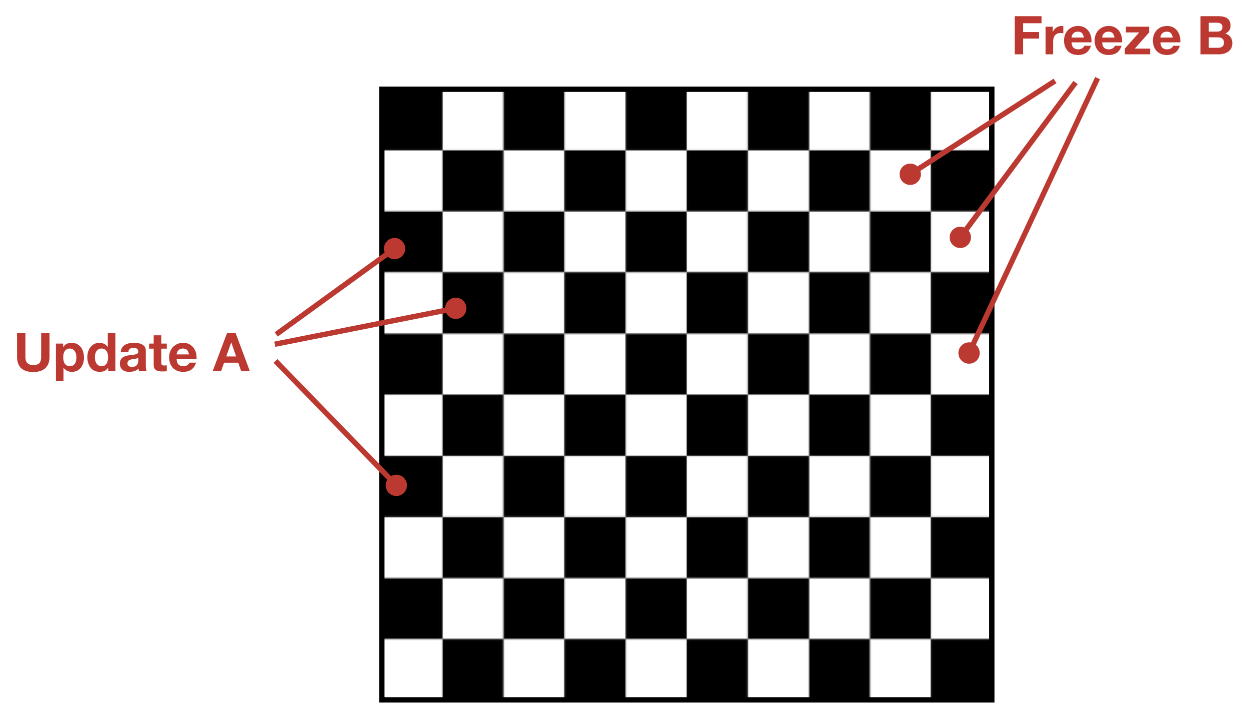


Example: RNVP for scalar fields

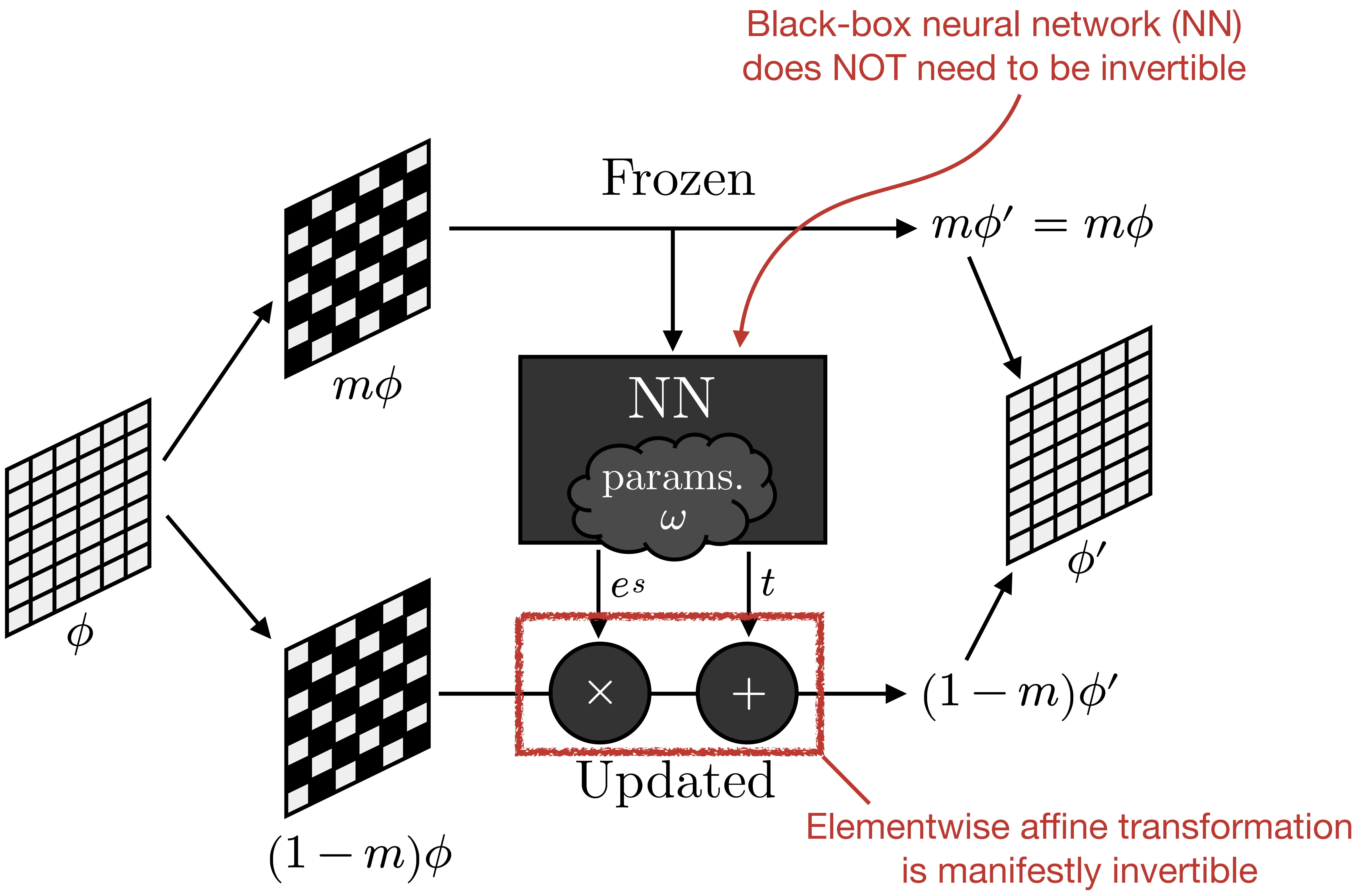
Scalar field $\phi(x) \in \mathbb{R} \;\approx\;$ grayscale image

Real NVP coupling layer:

[Dinh, Sohl-Dickstein, Bengio **1605.08803**]



Checkerboard masking pattern m



Symmetries in flows

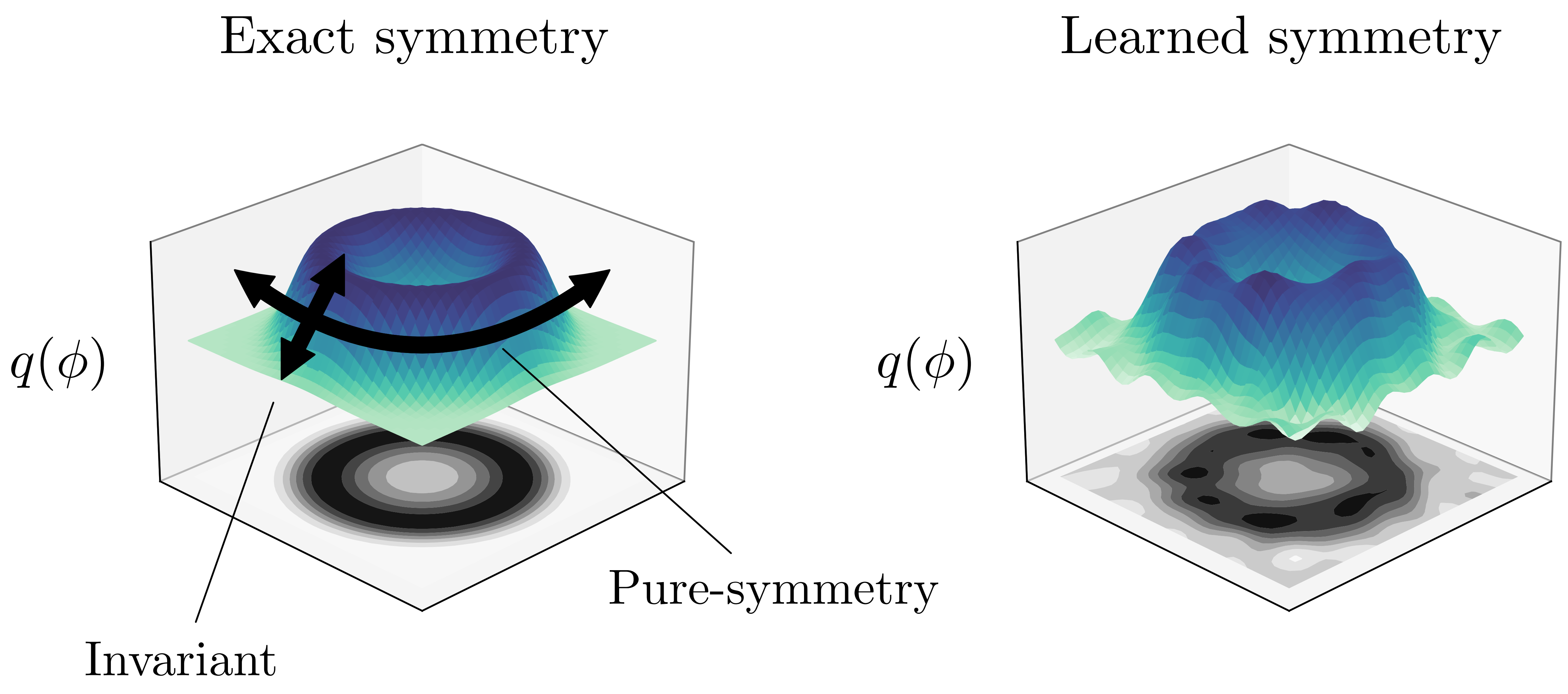
Motivation: Target $p(\phi)$ is often invariant under symmetries. Make $q(\phi)$ automatically invariant too?

Invariant prior + **equivariant** flow = symmetric model [Cohen, Welling 1602.07576]

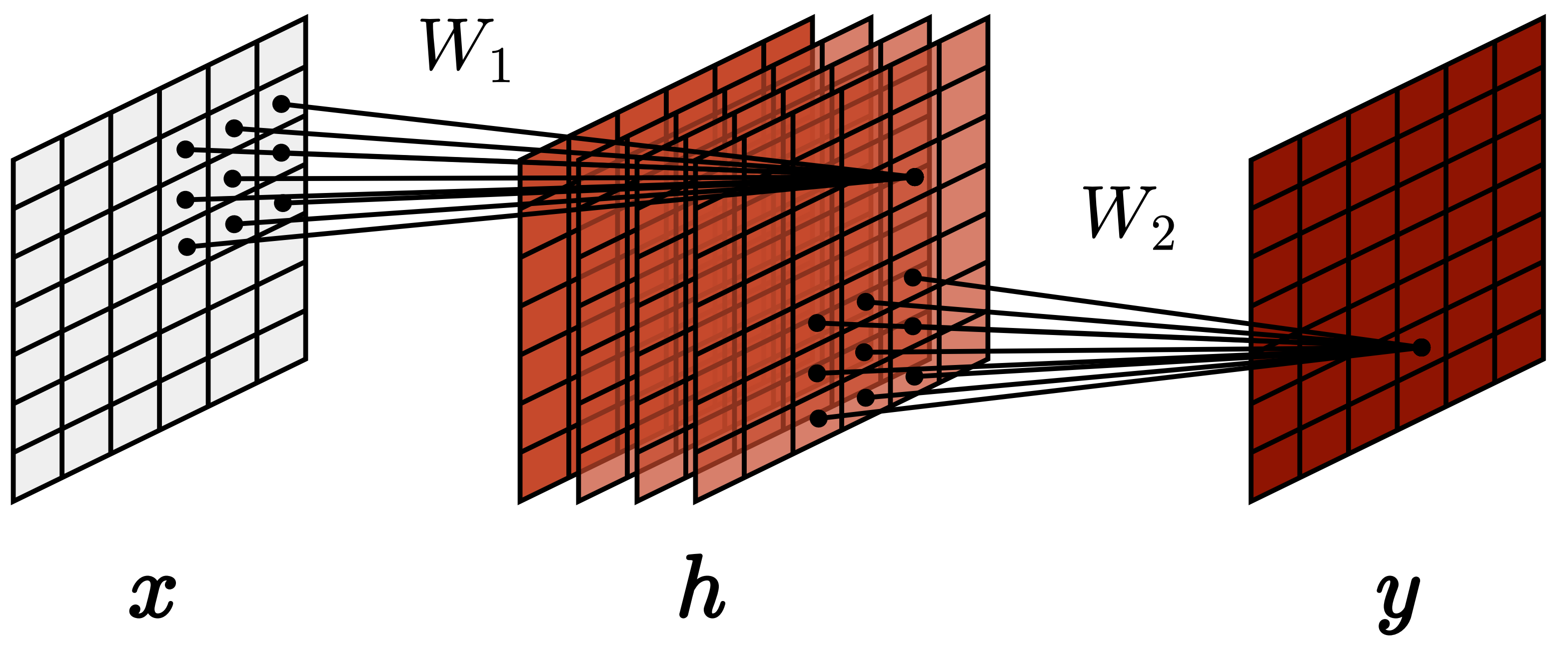
$r(t \cdot U) = r(U)$ $f(t \cdot U) = t \cdot f(U)$

Symmetries…

- ✓ Reduce data complexity of training
- ✓ Reduce model parameter count
- ✓ May make "loss landscape" easier



Translational equivariance



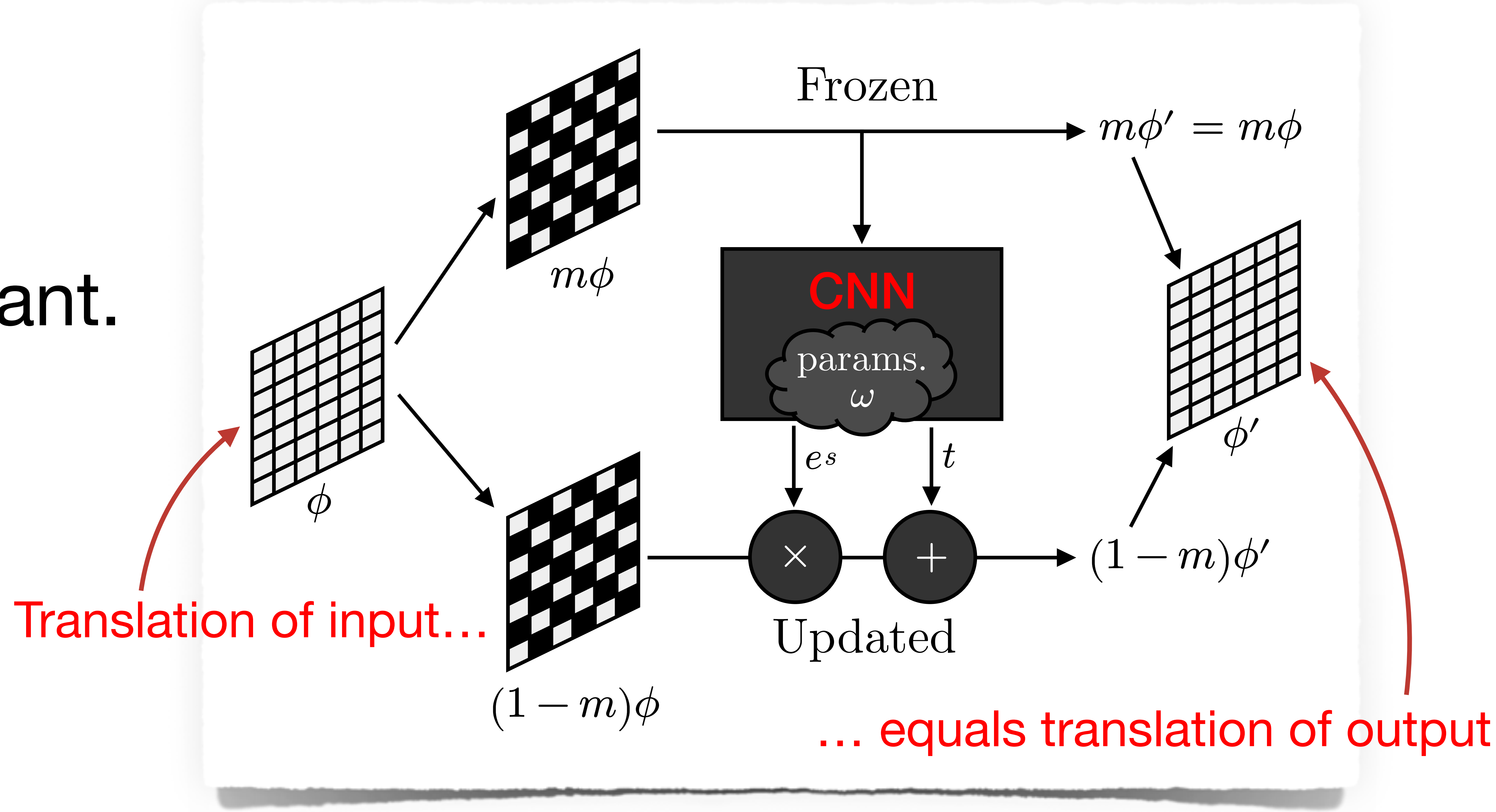
1. Use **Convolutional Neural Nets (CNNs).**

- Output values (e.g. $e^{s(x)}$ and $t(x)$) for each site are local functions of frozen DoFs

- CNNs are equivariant under translations

2. Make masking pattern (mostly) invariant.

- E.g. checkerboard



See also Self-Learning Monte Carlo (SLMC) methods:
[Huang, Wang **PRB95 (2017) 035105**;
Liu, et al. **PRB95 (2017) 041101**;
… and many more …]

Optimizing the model

Must not require a large number of samples from real distribution to optimize!

Self-training:

1. Loss function = modified **Kullback-Leibler (KL)** divergence

$$D_{KL}(q\,||\,p) := \int \mathcal{D}U\, q(U)\left[\log q(U) - \log p(U)\right] \geq 0$$

Constant shift removes unknown normalization

$$D'_{KL}(q\,||\,p) := \int \mathcal{D}U\, q(U)\left[\log q(U) + S(U)\right] \geq -\log Z$$

(Using $p(U) = e^{-S(U)}/Z$)

2. Stochastic estimate: draw samples U_i **from the model**, then measure

$$\frac{1}{M}\sum_{i=1}^{M}\left[\log q(U_i) + S(U_i)\right]$$

Exactness

Samples from **model** are from biased distribution $q(\phi) \neq p(\phi)$, but…

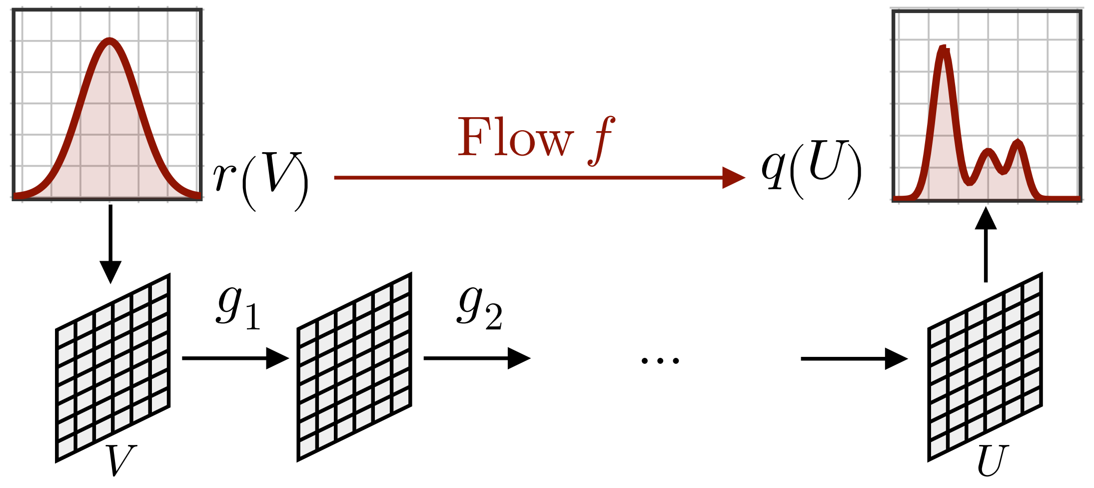
For each ϕ_i drawn from the model, we know $q(\phi_i)$ and $p(\phi_i)$

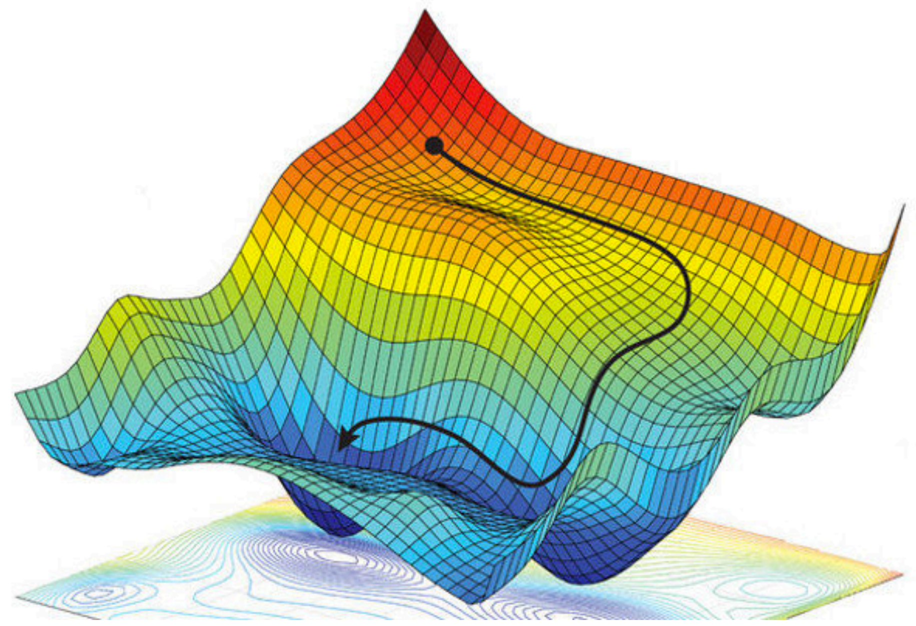
→

Exact bias correction possible
(e.g. "flow-based MCMC" or reweighting)

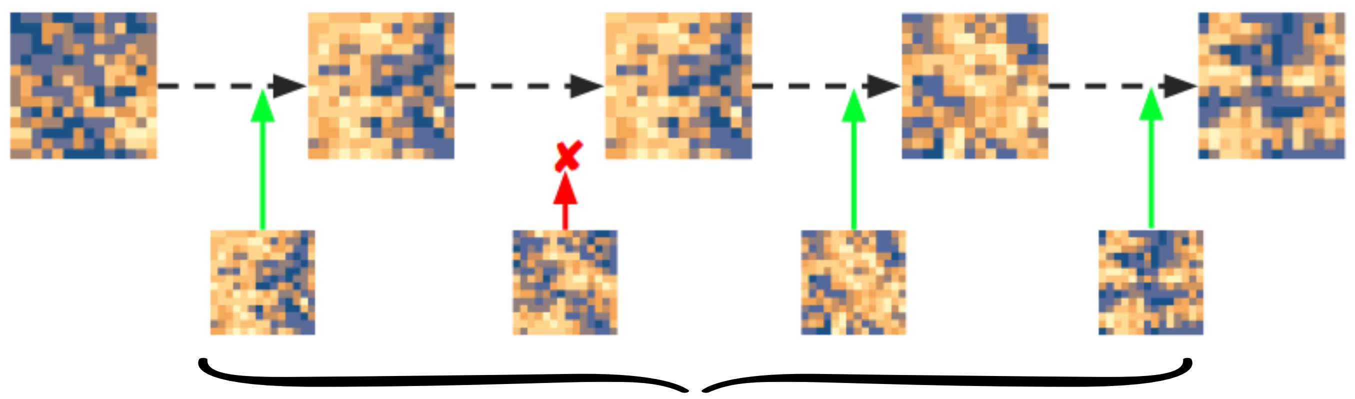
Note: Efficiency of bias correction depends on how close q and p are.

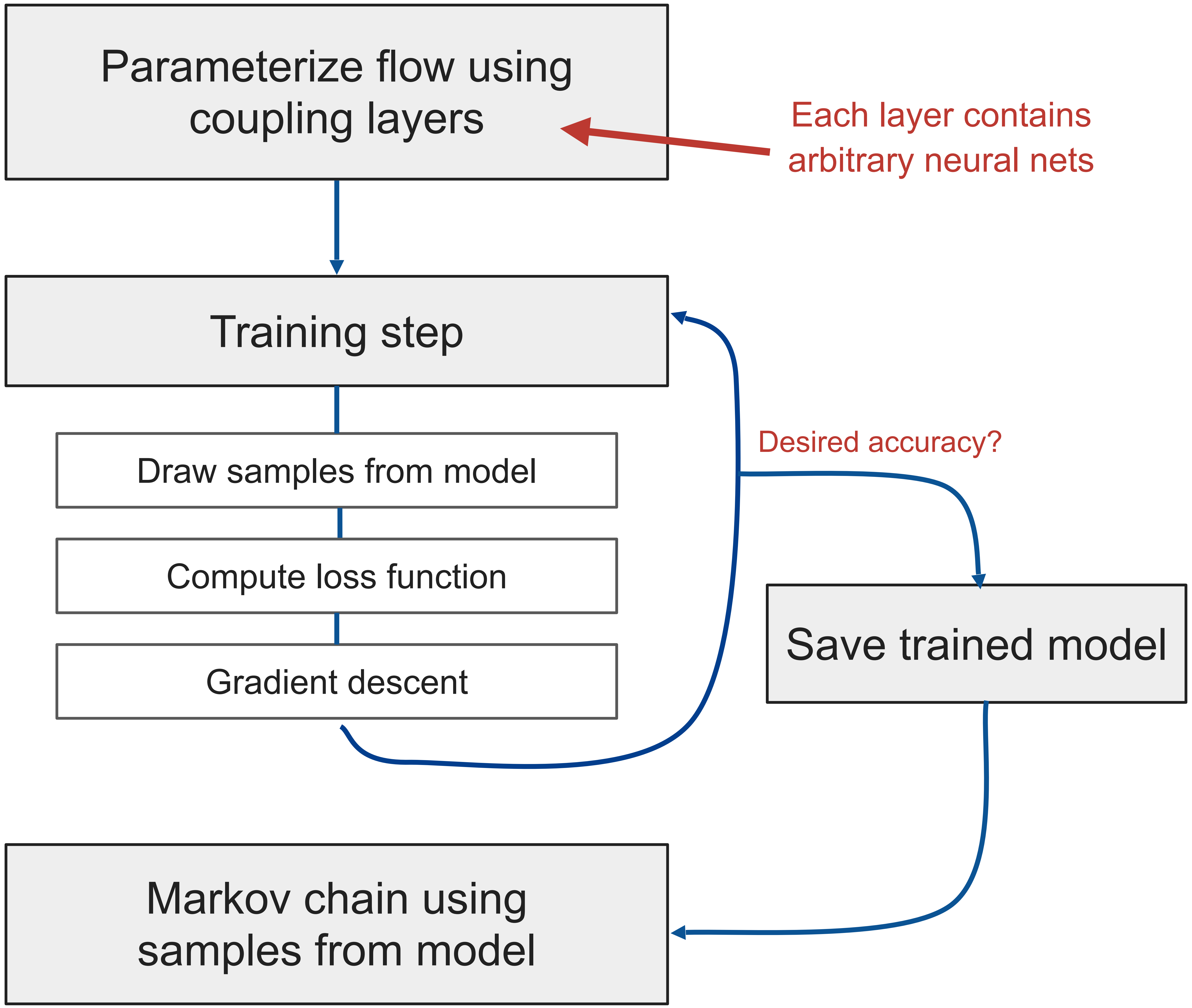
Birds-eye view





[Image credit: 1805.04829]





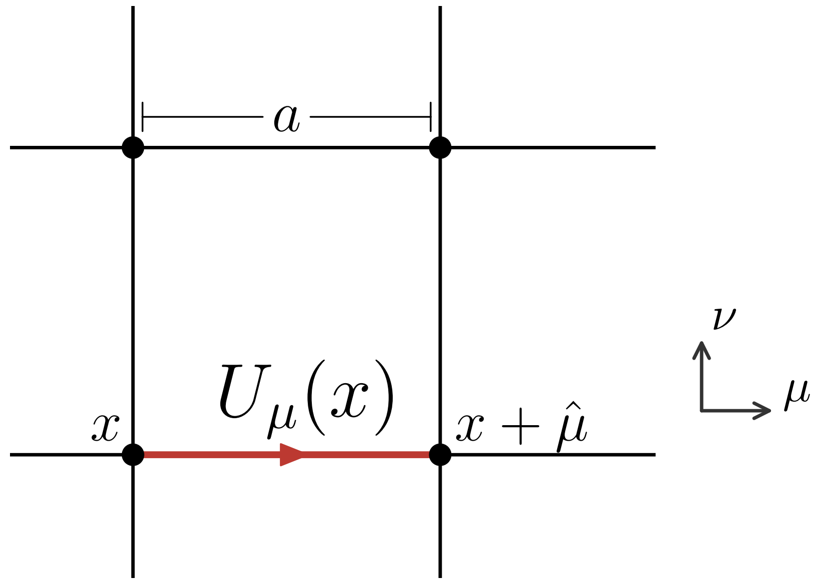
A story of symmetries & generative models

(In three parts)

1. Flow-based generative models
2. **Gauge symmetry & translational symmetry**
3. Fermions & translational symmetry

Lattice gauge theory & Symmetries

Gauge field **discretized** in terms of parallel transporters (links) $U_\mu(x)$.

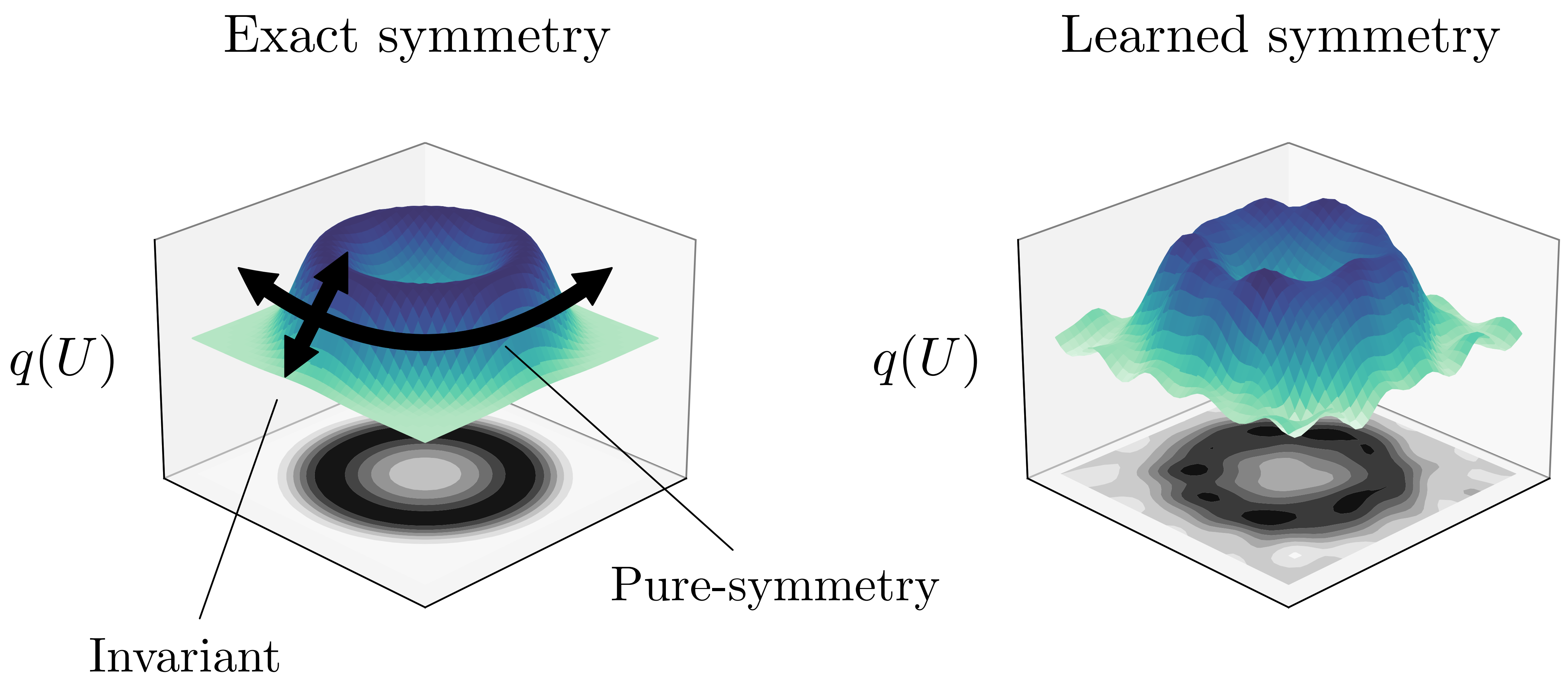


Lattice gauge theory actions (typically) **satisfy symmetries**:

1. (Discrete) translational symmetries
2. Hypercubic symmetries
3. Gauge symmetries

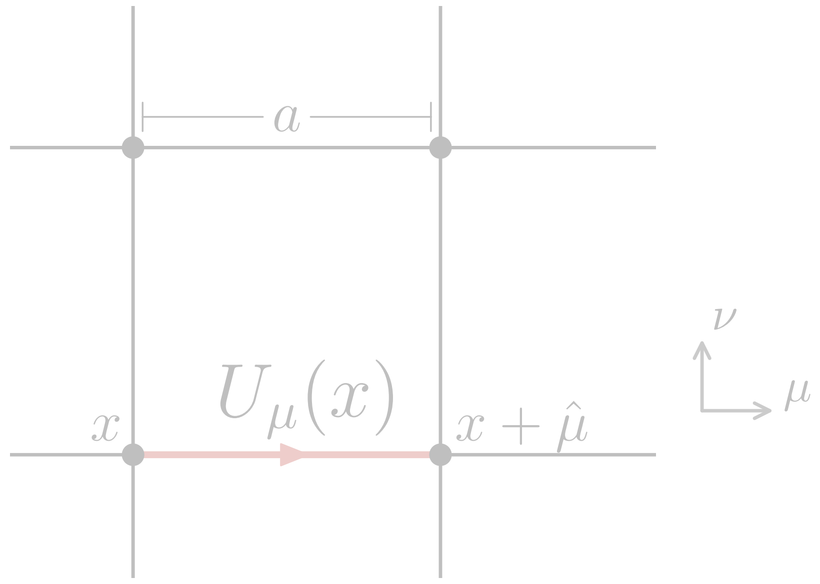
$$(\Omega \cdot U)_\mu(x) = \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu})$$

Symmetries **factor** distribution into uniform component along symmetry direction, and non-uniform component along invariant direction. Schematically:

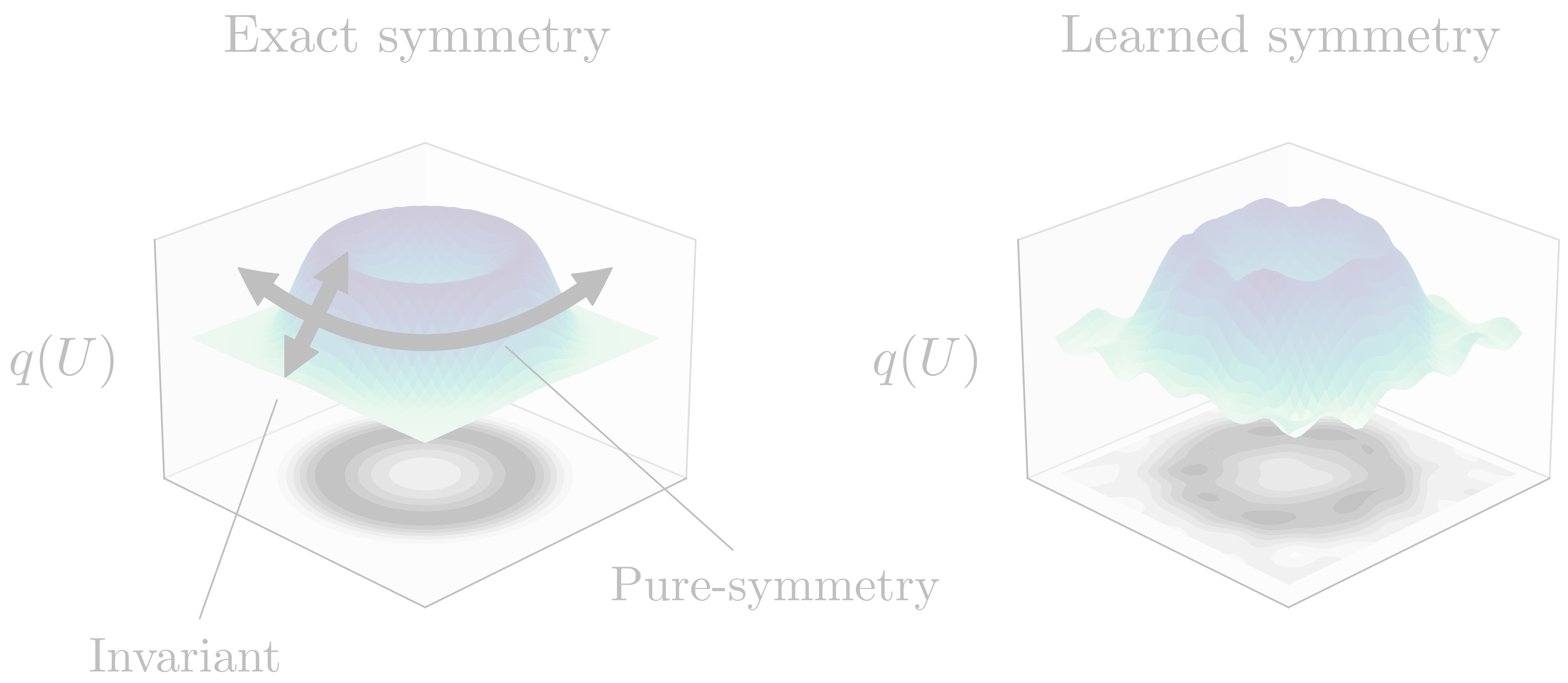


Lattice gauge theory & Symmetries

Gauge field **discretized** in terms of parallel transporters (links) $U_\mu(x)$.



Symmetries **factor** distribution into uniform component along symmetry direction, and non-uniform component along invariant direction. Schematically:



Lattice gauge theory actions (typically) **satisfy symmetries**:

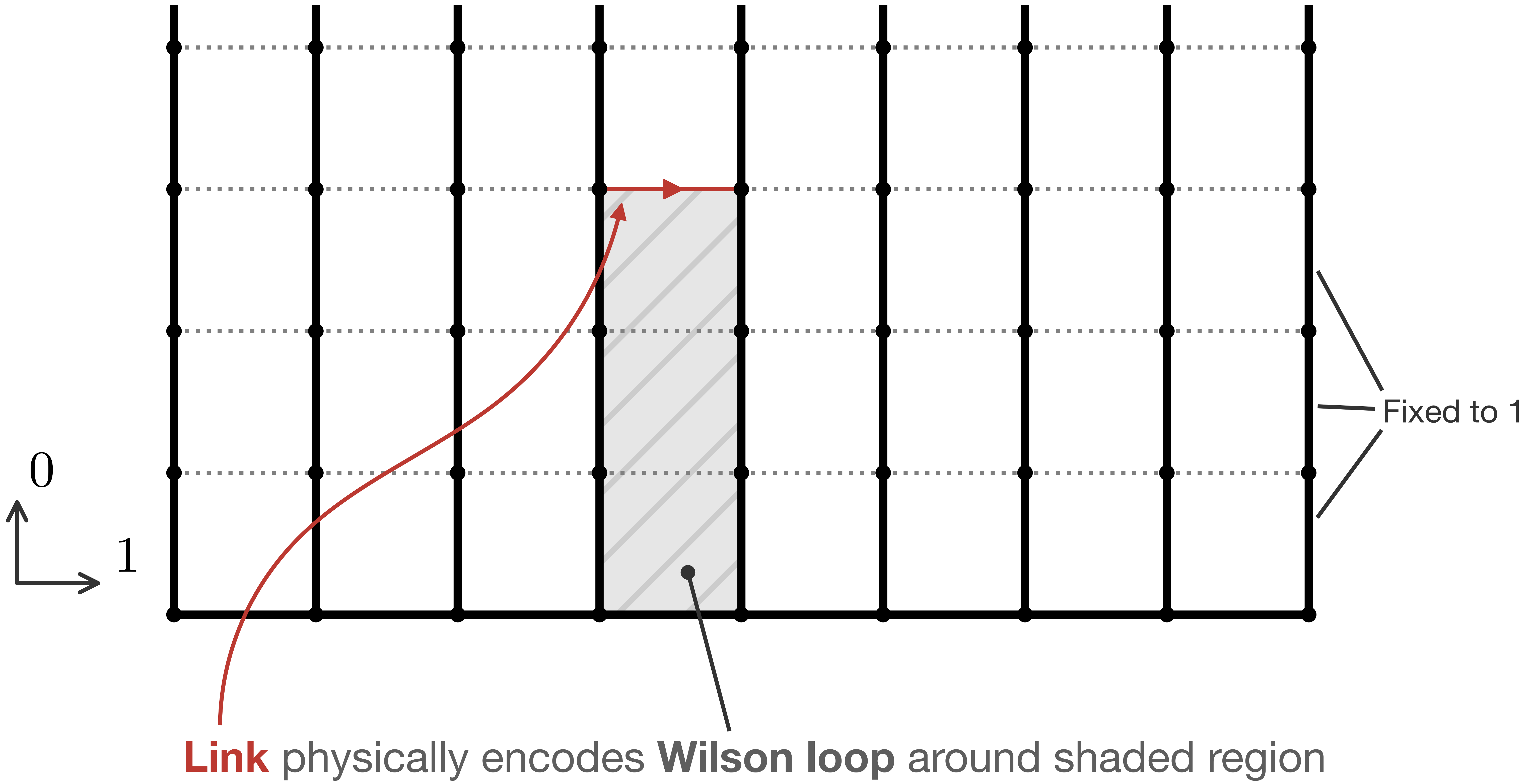
1. (Discrete) translational symmetries
2. Hypercubic symmetries
3. Gauge symmetries

$$(\Omega \cdot U)_\mu(x) = \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu})$$

Gauge symmetry via gauge fixing?

Where gauge DoFs are explicitly factored out, e.g. maximal tree

Explicit gauge fixing is at odds with **translational symmetry** + **locality**



Gauge symmetry via gauge fixing?

Where gauge DoFs are fixed by solving a constraint, e.g. Landau gauge

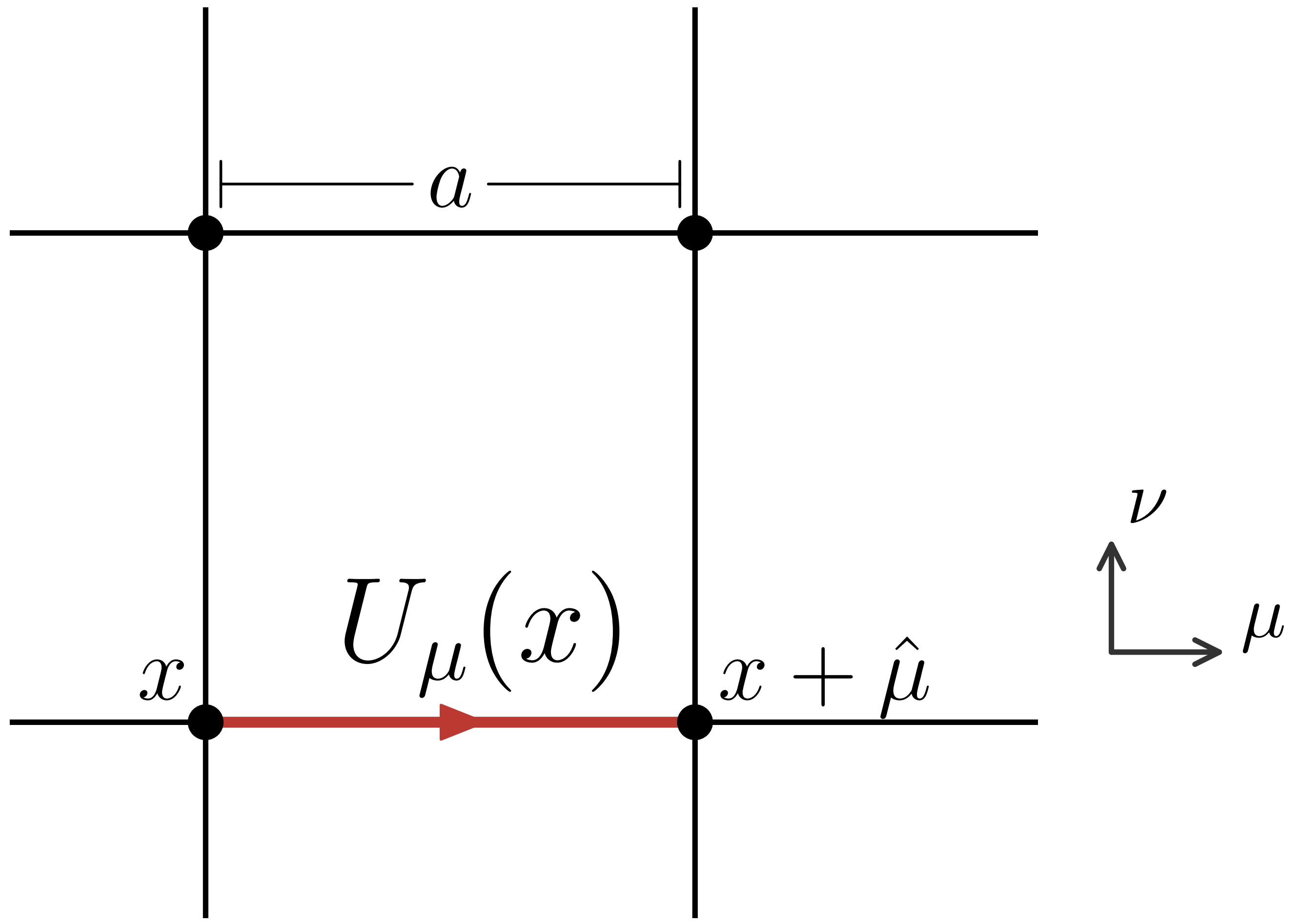
Implicit gauge fixing difficult to act on via **flow-based models**

Landau gauge: $$U_\mu^{\mathrm{fix}}(x) = \mathrm{argmin}_{U^\Omega} \sum_x \sum_{\mu=1}^{N_d} \mathrm{ReTr}[U_\mu^\Omega(x)]$$

Coulomb gauge: $$U_\mu^{\mathrm{fix}}(x) = \mathrm{argmin}_{U^\Omega} \sum_x \sum_{\mu=1}^{N_d-1} \mathrm{ReTr}[U_\mu^\Omega(x)]$$

Unclear how to invertibly transform $U_\mu^{\mathrm{fix}}(x)$.

Gauge symmetries in flows



Choose to act on the un-fixed link representation $U_\mu(x)$.

Carefully construct architecture to enforce…

Gauge-invariant prior:

Not very difficult!
Uniform distribution works.

With respect to Haar measure

$$r(U) = 1$$

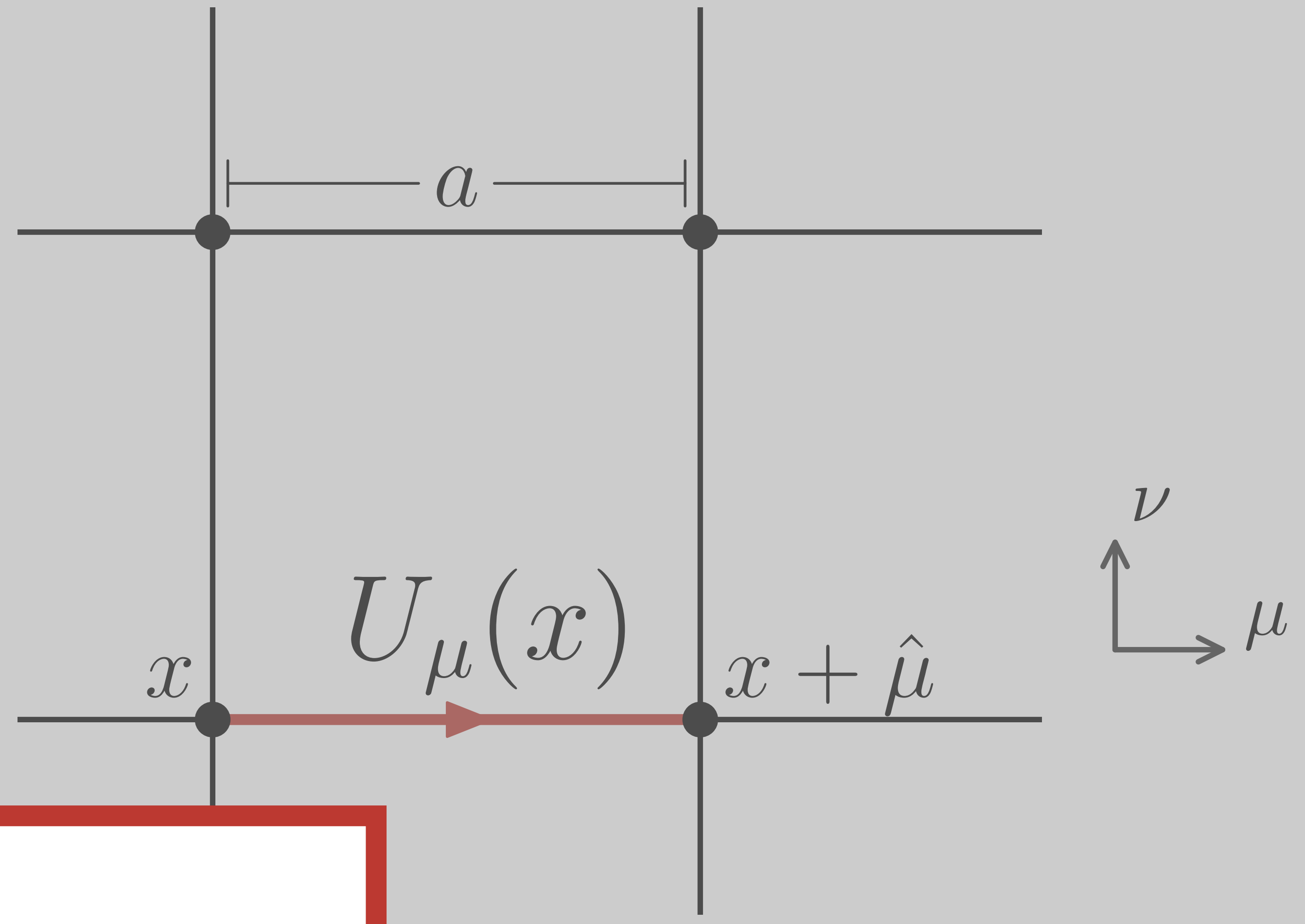
Gauge-equivariant flow:

Coupling layers acting on (untraced) Wilson loops.

Loop transformation easier to satisfy.

Gauge symmetries in flows

Choose to act on the un-fixed link representation $U_\mu(x)$.

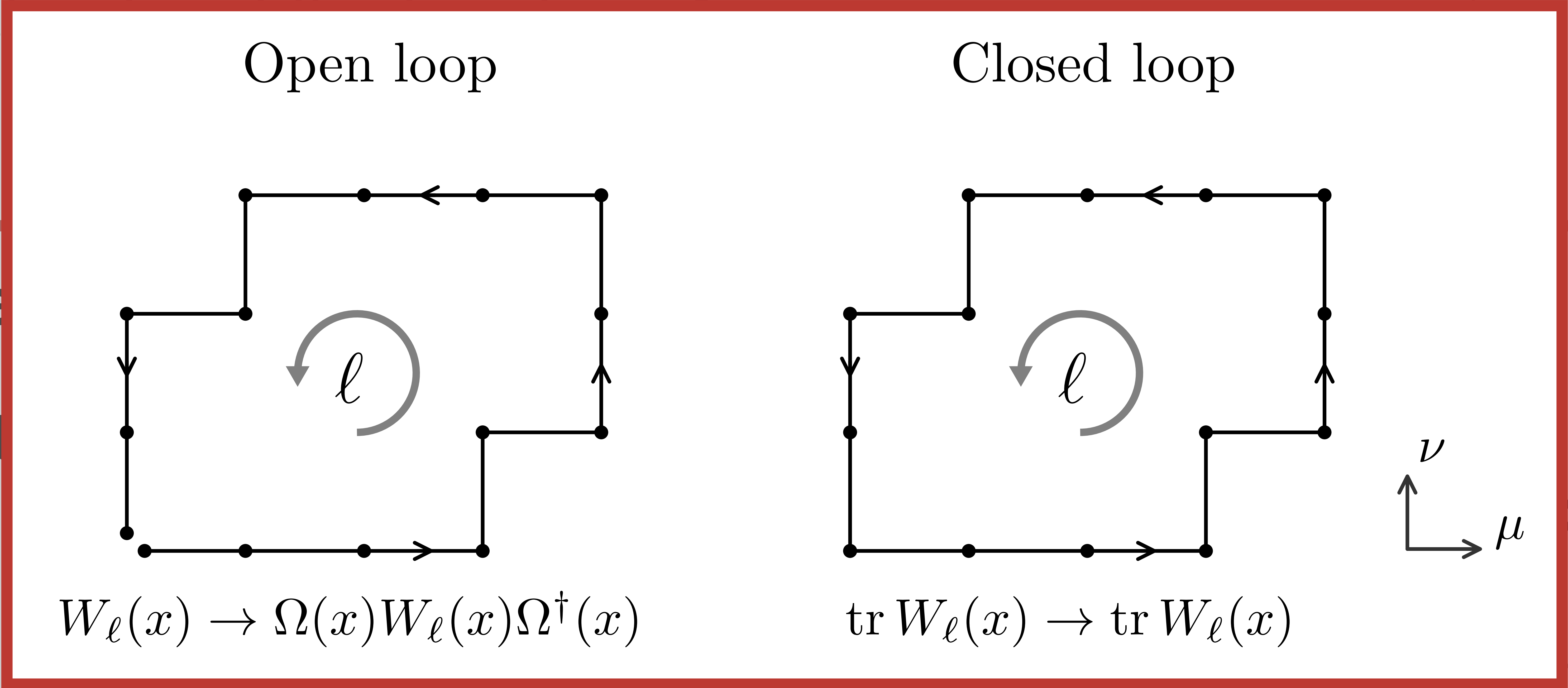


Carefully cons

Gauge

Not
Uniform

With respect to Haar measure



iant flow:

acting on
on loops.

tion easier
to satisfy.

Gauge-equivariant coupling layer

Compute a **field of Wilson loops** $W_\ell(x)$.

Inner coupling layer [function of $W_\ell(x)$]

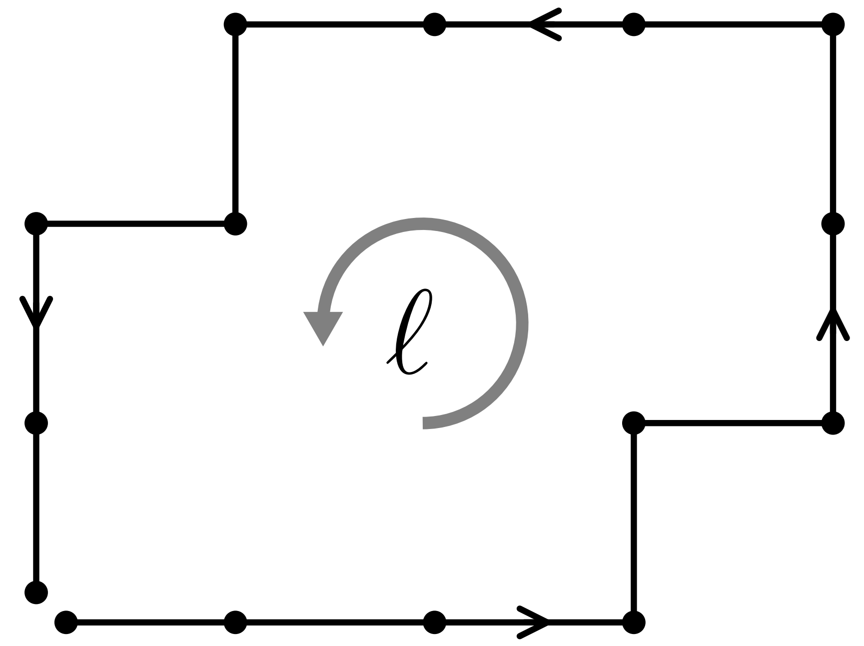
- "Actively" update a subset of loops.*
- Condition on "frozen" closed loops.

Gauge invariant!

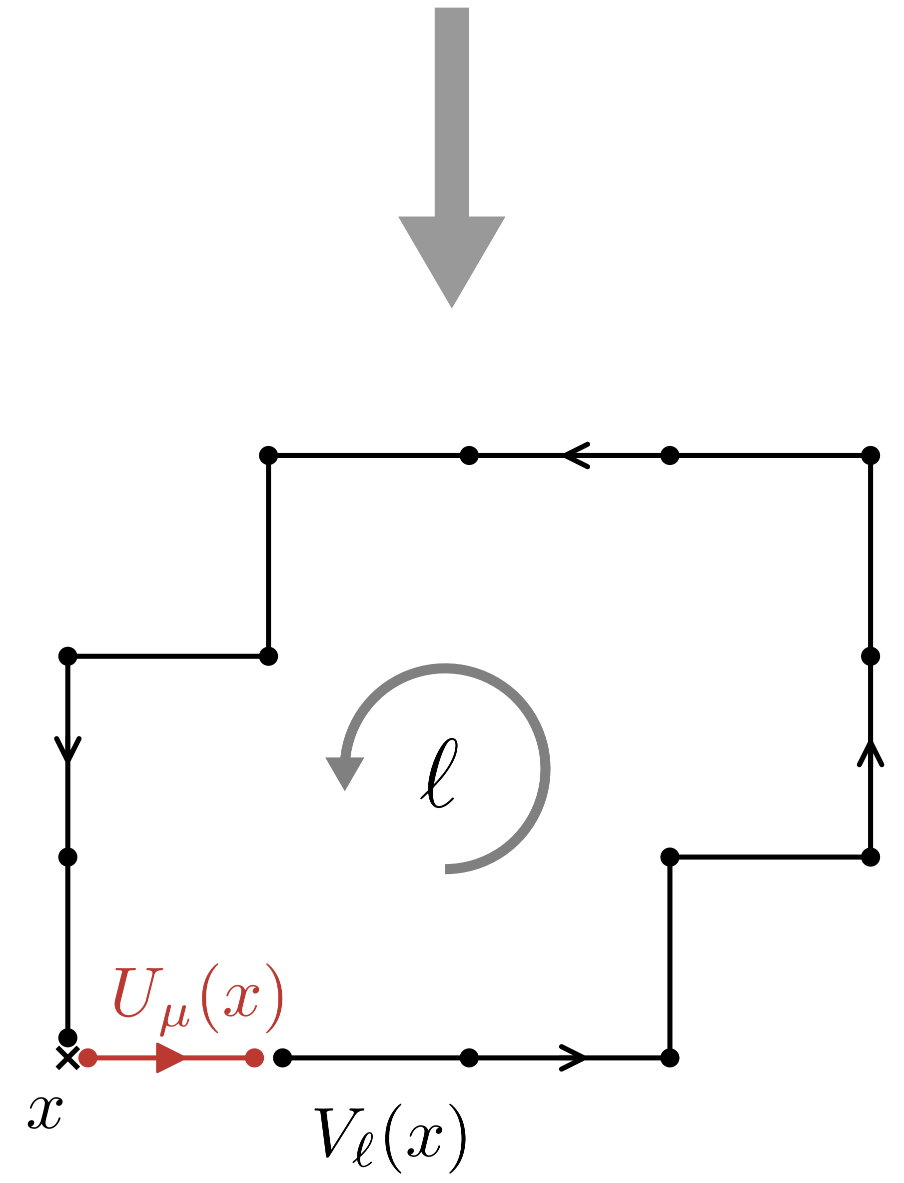
Outer coupling layer [function of $U_\mu(x)$]

- Solve for link update to satisfy actively updated loops.
- Other loops in $W_\ell(x)$ may "passively" update.

Open loop



$$W_\ell(x) \xrightarrow{\text{Flow}} W'_\ell(x)$$



$$U'_\mu(x) = W'_\ell(x)\, V^\dagger_\ell(x)$$

[GK, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan **PRL125 (2020) 121601**]
[Boyda, GK, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan **PRD103 (2021) 074504**]

Gauge-equivariant coupling layer

Compute a **field of Wilson loops** $W_\ell(x)$.

Inner coupling layer [function of $W_\ell(x)$]

- "Actively" update a subset of loops.*
- Condition on "frozen" closed loops.

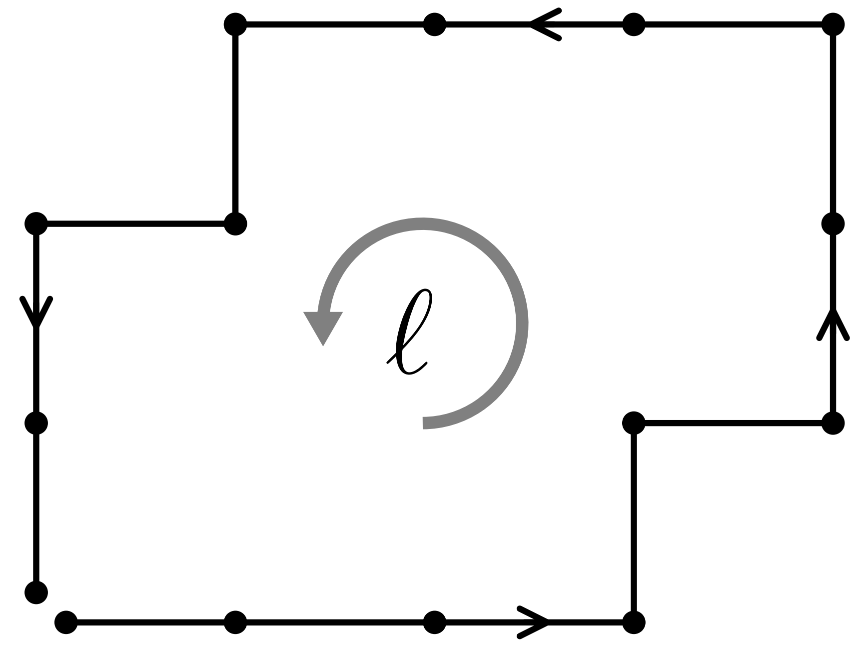
Gauge invariant!

> * This **"kernel"** must satisfy:
>
> $$h(W_\ell^\Omega(x)) = h^\Omega(W_\ell(x))$$

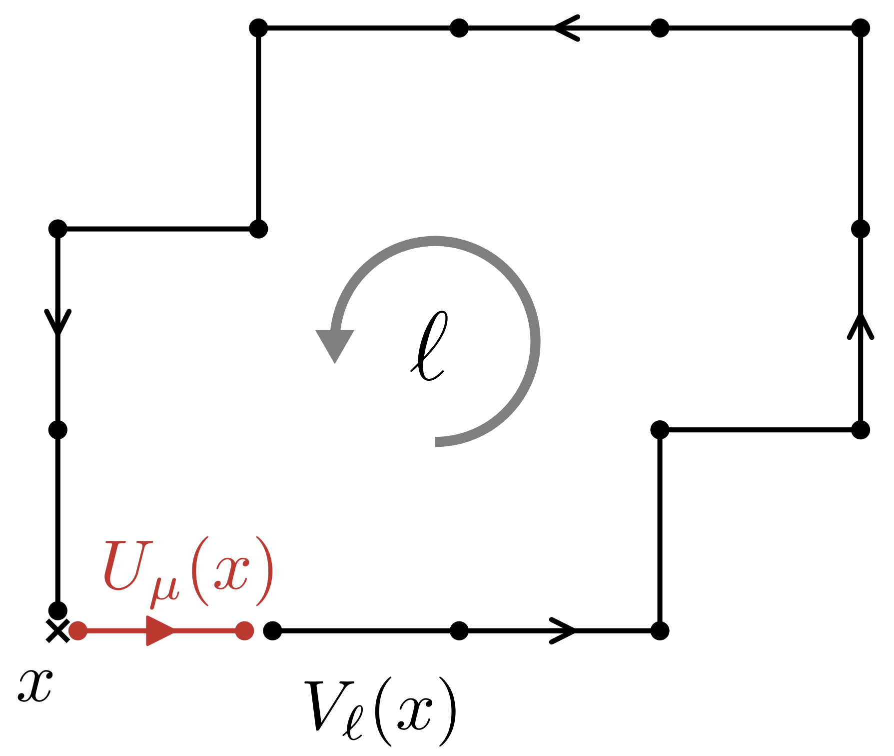
Outer coupling layer [function of $U_\mu(x)$]

- Solve for link update to satisfy actively updated loops.
- Other loops in $W_\ell(x)$ may "passively" update.

Open loop



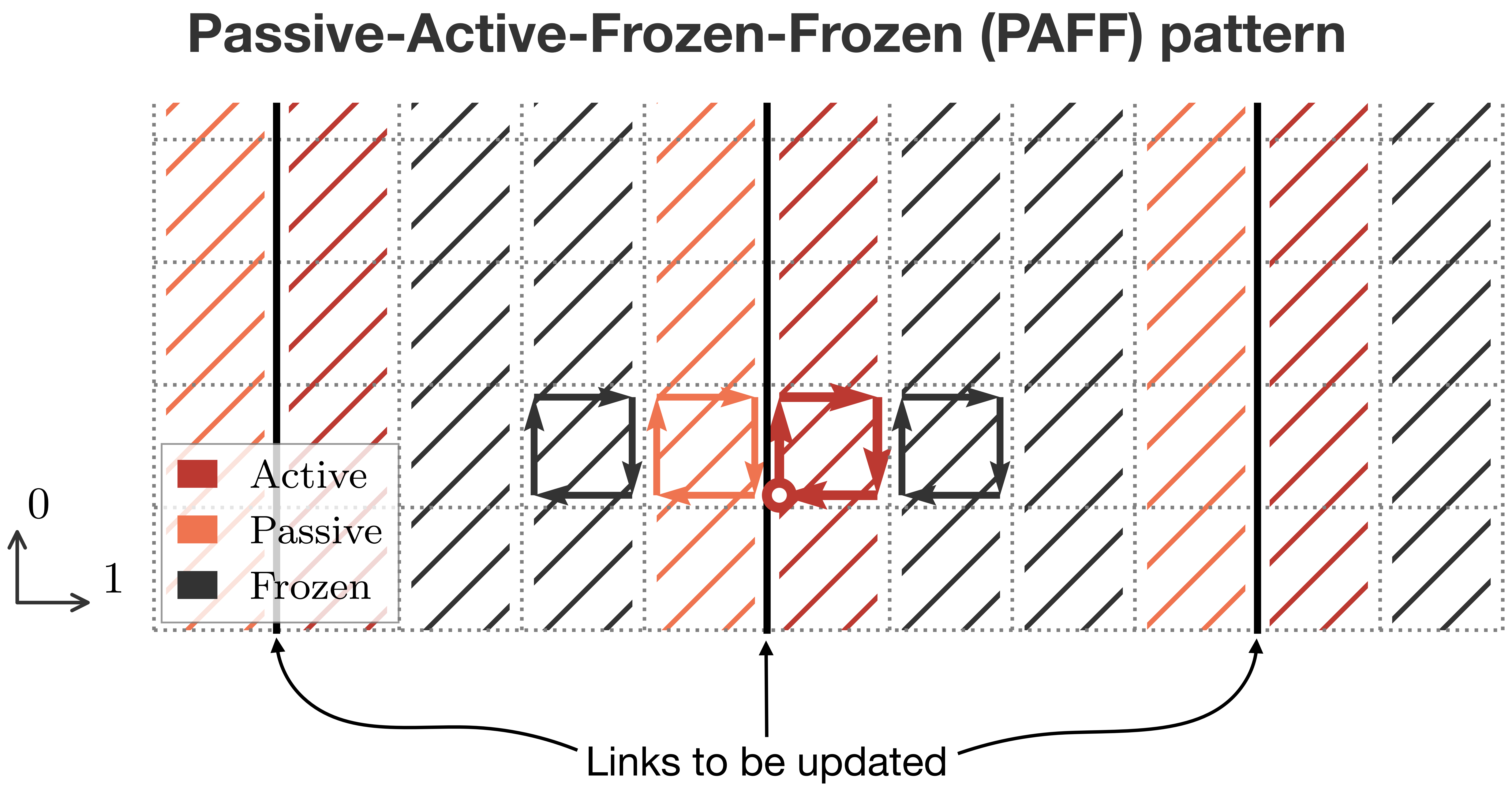
$$W_\ell(x) \xrightarrow{\text{Flow}} W'_\ell(x)$$



$$U'_\mu(x) = W'_\ell(x)\, V^\dagger_\ell(x)$$

[GK, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan **PRL125 (2020) 121601**]
[Boyda, GK, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan **PRD103 (2021) 074504**]

Active, passive, and frozen loops



Kernels

Coupling layers required **kernels** satisfying **conjugation equivariance**:

$$h(\Omega W \Omega^\dagger) = \Omega h(W) \Omega^\dagger$$

U(1): Trivially satisfied because $h(\Omega W \Omega^\dagger) = h(W) = \Omega h(W) \Omega^\dagger$.

However, invertible transforms on the compact domain required.

[Rezende, Papamakarios, Racanière, Albergo, GK, Shanahan, Cranmer; **ICML (2020) 2002.02428**]

SU(N): Non-trivial constraint requiring some **fun mathematical engineering**…

SU(N) kernels: strategy

SU(N) matrix-conj. equivariance is **non-trivial**.

$$h(\Omega W \Omega^\dagger) = \Omega h(W) \Omega^\dagger$$

Useful observations:

- Conjugation only rotates eigenvectors.
- Spectrum is invariant.
- Wilson loop spectrum encodes gauge-invariant physics → **This is what we want to transform.**

Strategy: Invertibly transform only the spectrum of W via a "spectral map".

Or, "spectral flow".

[Boyda, GK, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan **PRD103 (2021) 074504**]

SU(N) kernels: strategy

SU(N) matrix-conj. equivariance is **non-trivial**.

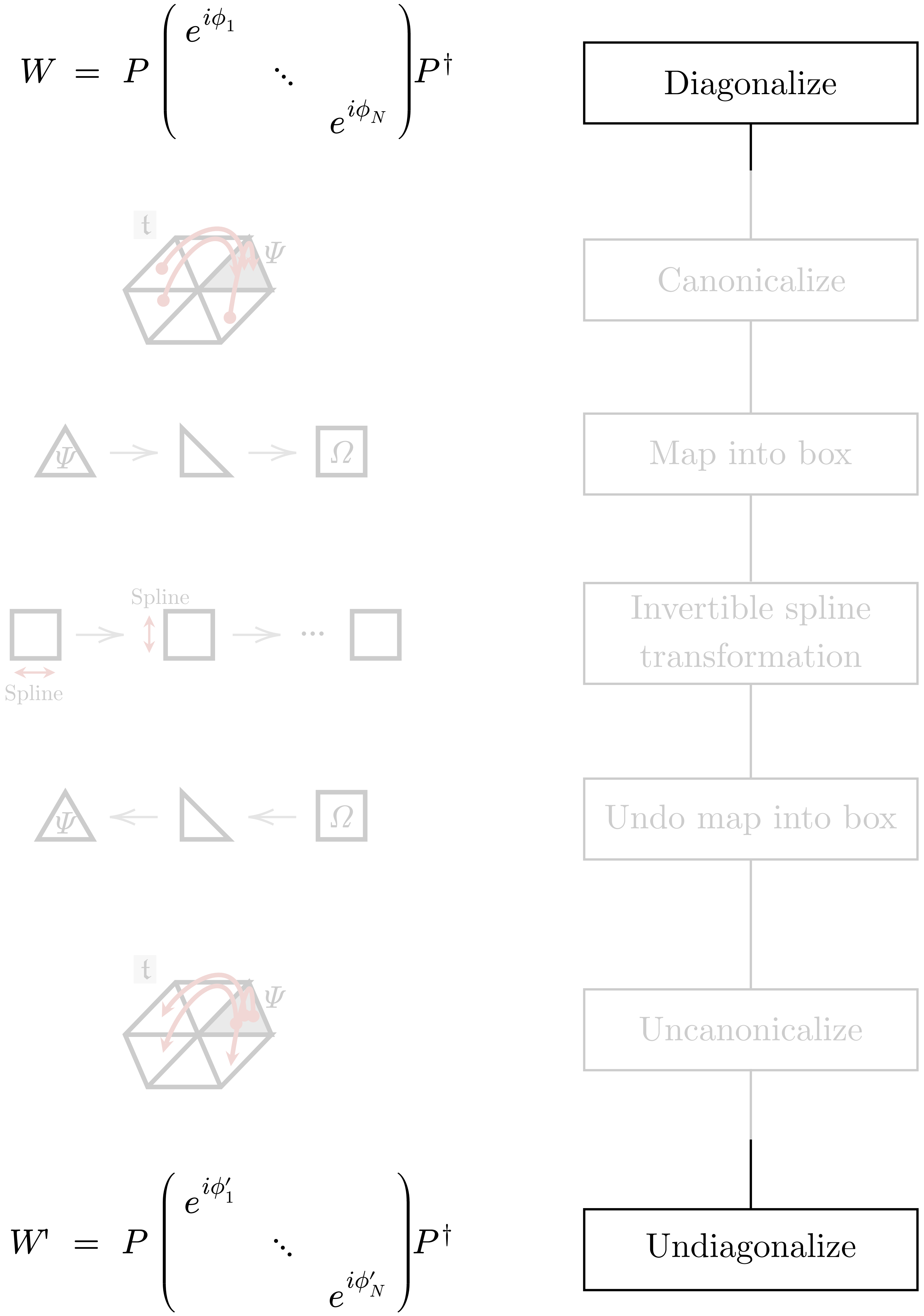
$$h(\Omega W \Omega^\dagger) = \Omega h(W) \Omega^\dagger$$

Useful observations:

- Conjugation only rotates eigenvectors.
- Spectrum is invariant.
- Wilson loop spectrum encodes gauge-invariant physics → **This is what we want to transform.**

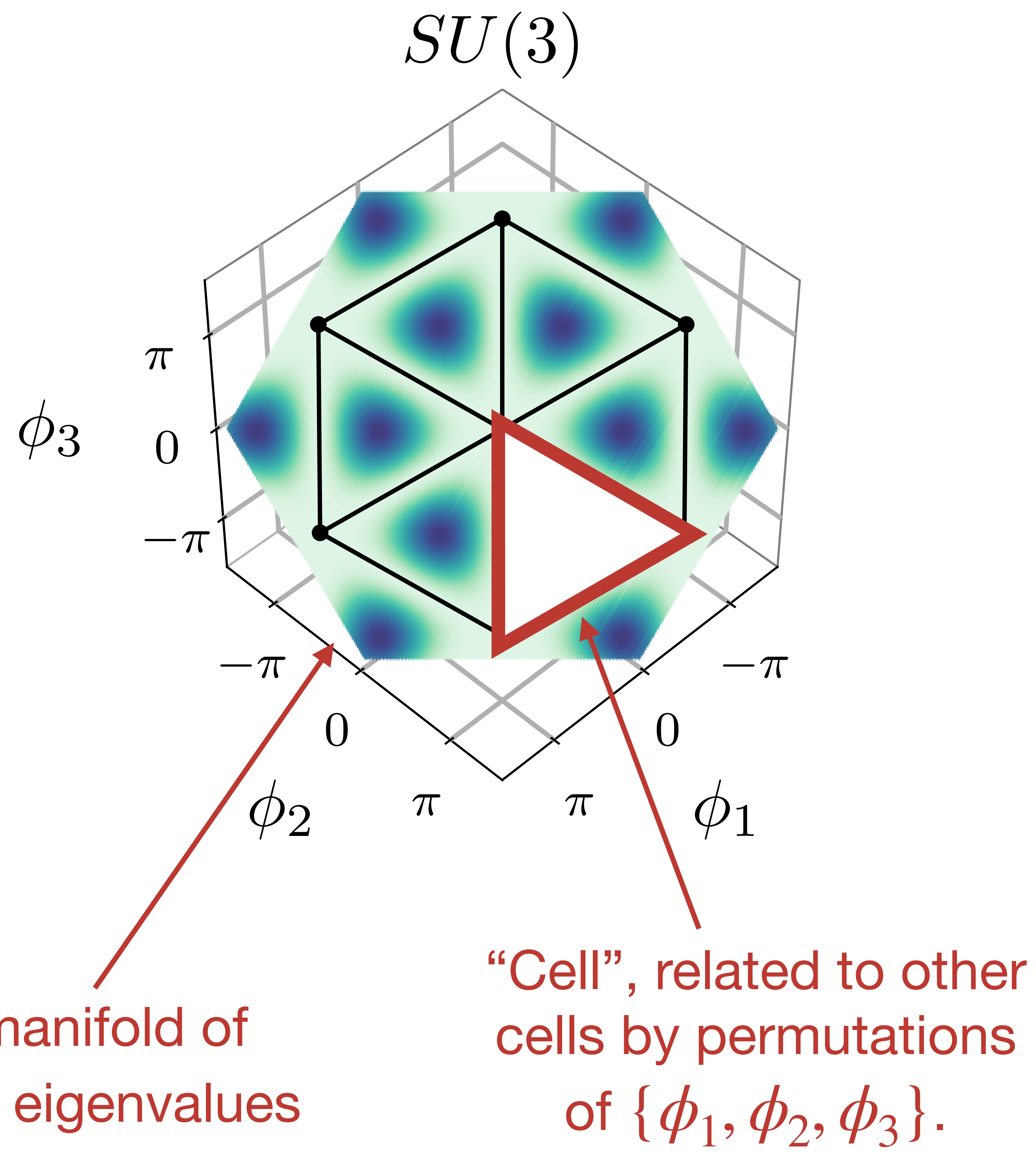
Strategy: Invertibly transform only the spectrum of W via a "spectral map".

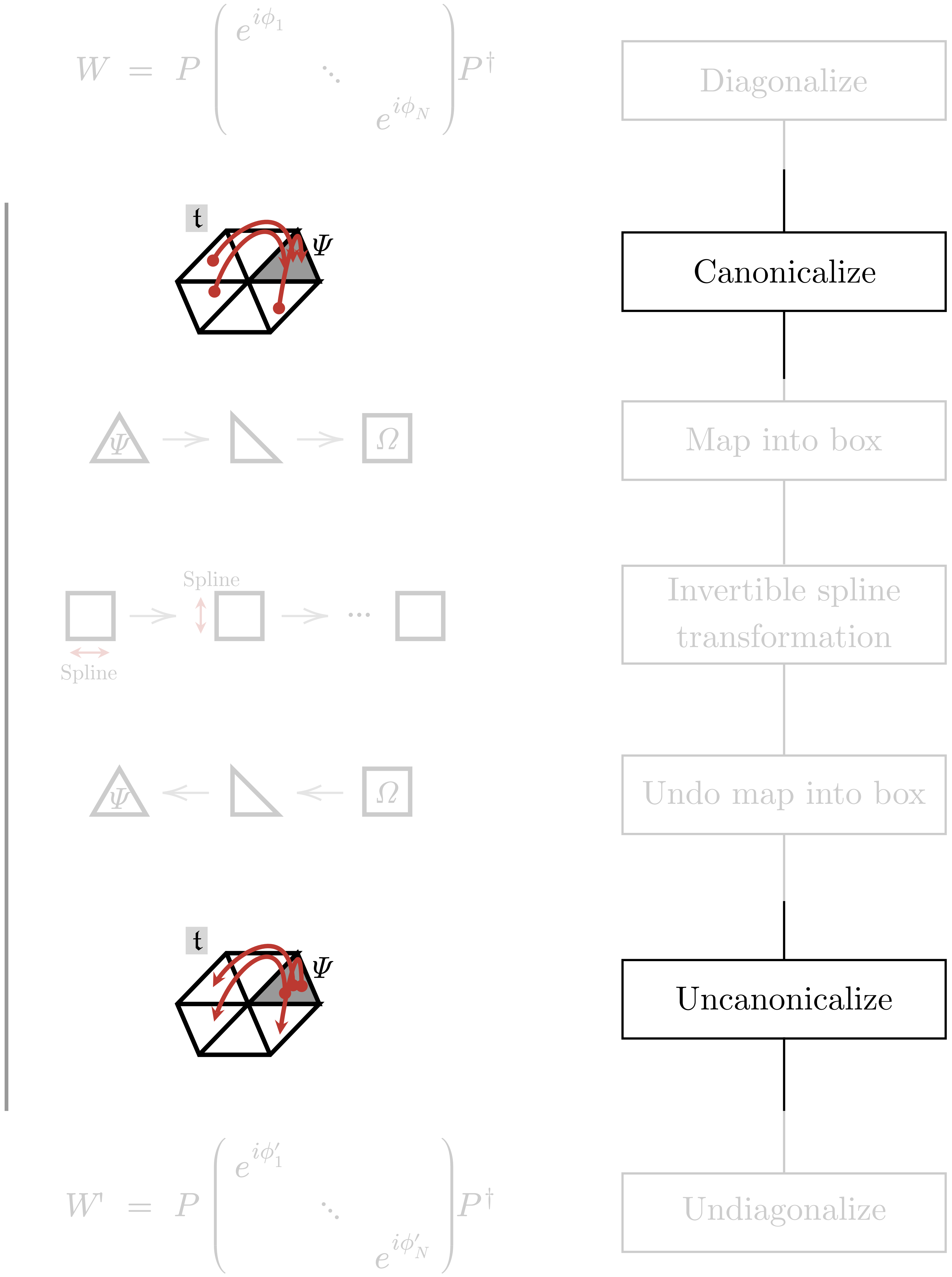
Or, "spectral flow".



[Boyda, GK, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan **PRD103 (2021) 074504**]

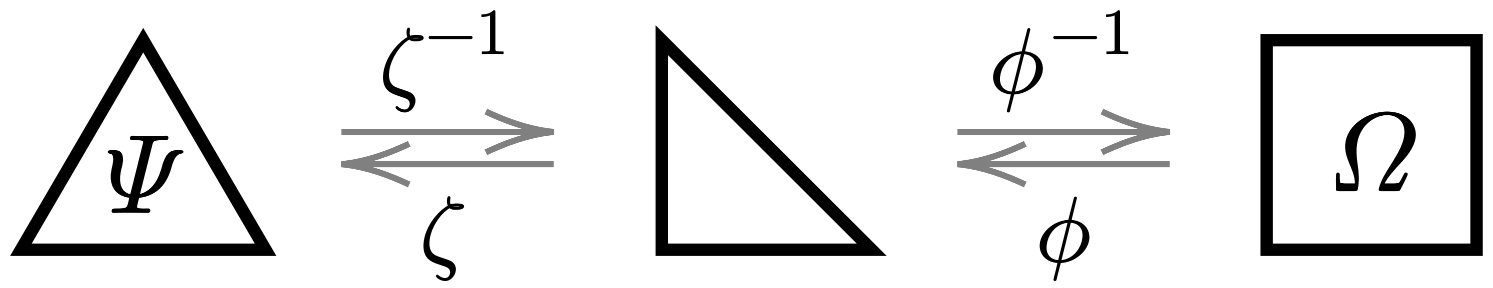
SU(N) kernels:
Permutation equivariance





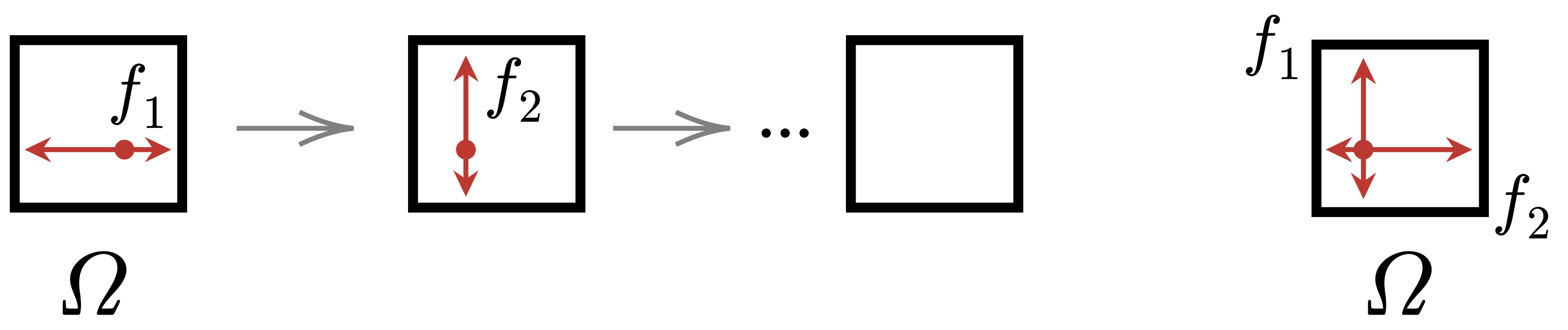
SU(N) kernels:
Transform the canonical cell

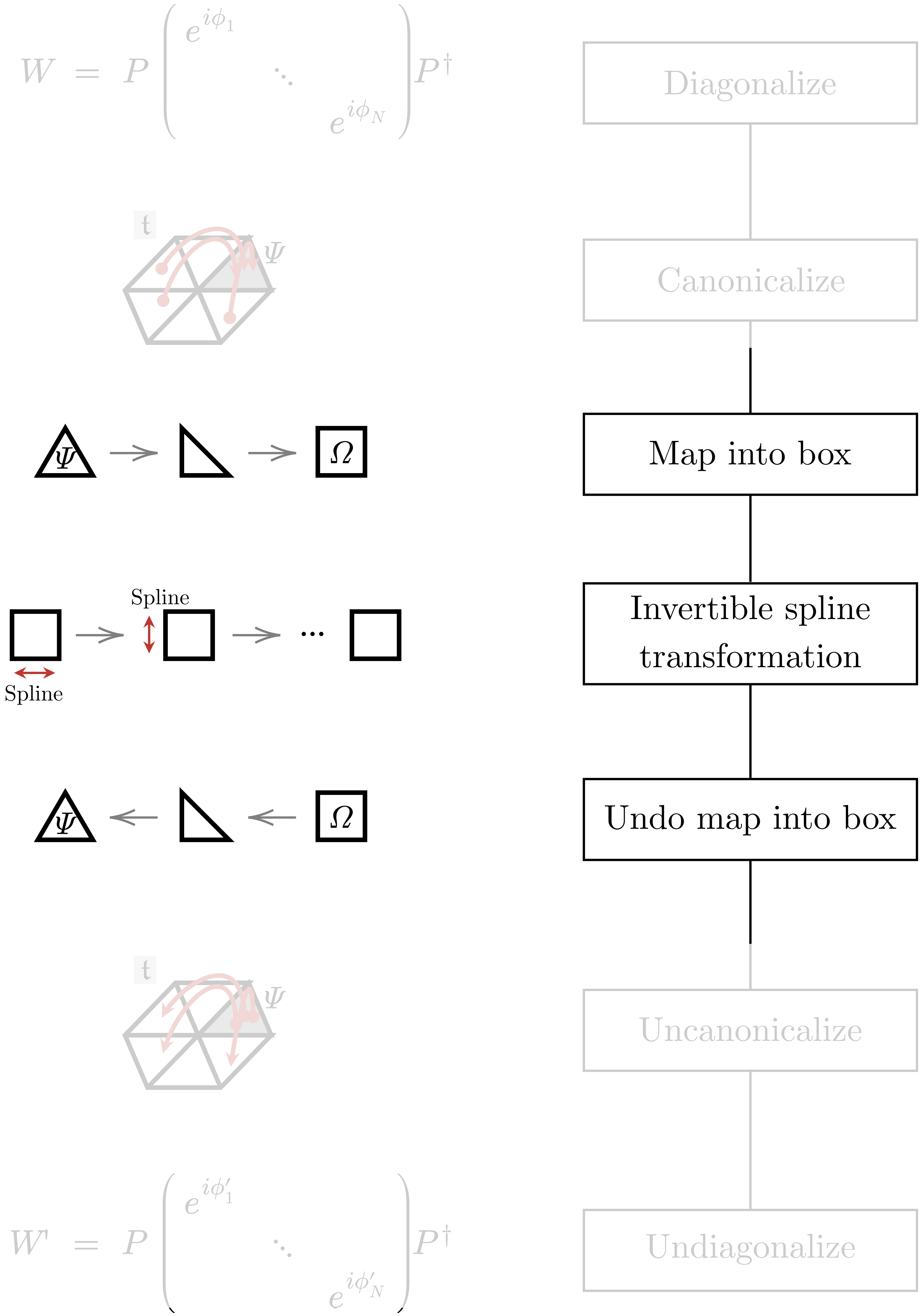
Change variables to rectilinear box Ω



Transform by acting on coords of box Ω, either…

Autoregressive … or … Independent



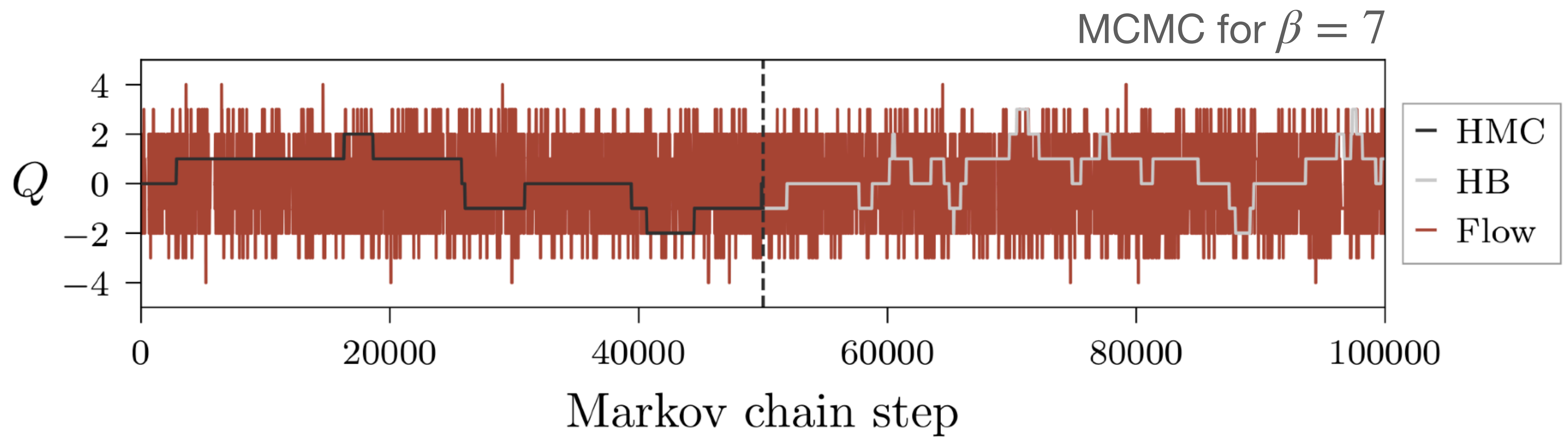


U(1) gauge theory in 1+1D

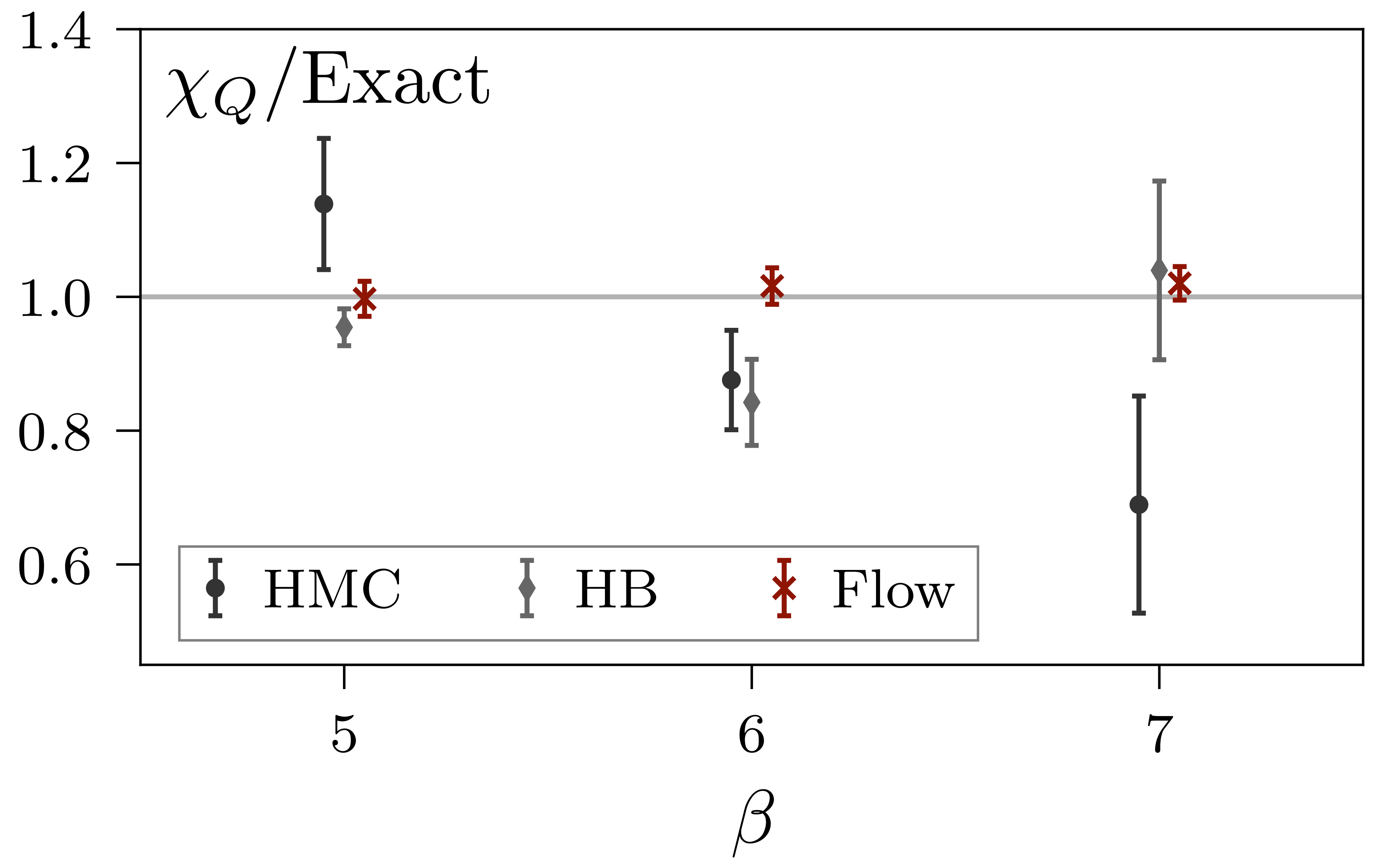
$$S(U) = -\beta \sum_x \sum_{\mu<\nu} \mathrm{Re}\, P_{\mu\nu}(x)$$
$$P_{\mu\nu}(x) = U_\mu(x) U_\nu(x+\hat{\mu}) U_\mu^\dagger(x+\hat{\nu}) U_\nu^\dagger(x)$$

There is exact lattice topology in 2D.

$$Q = \frac{1}{2\pi} \sum_x \arg(P_{01}(x))$$



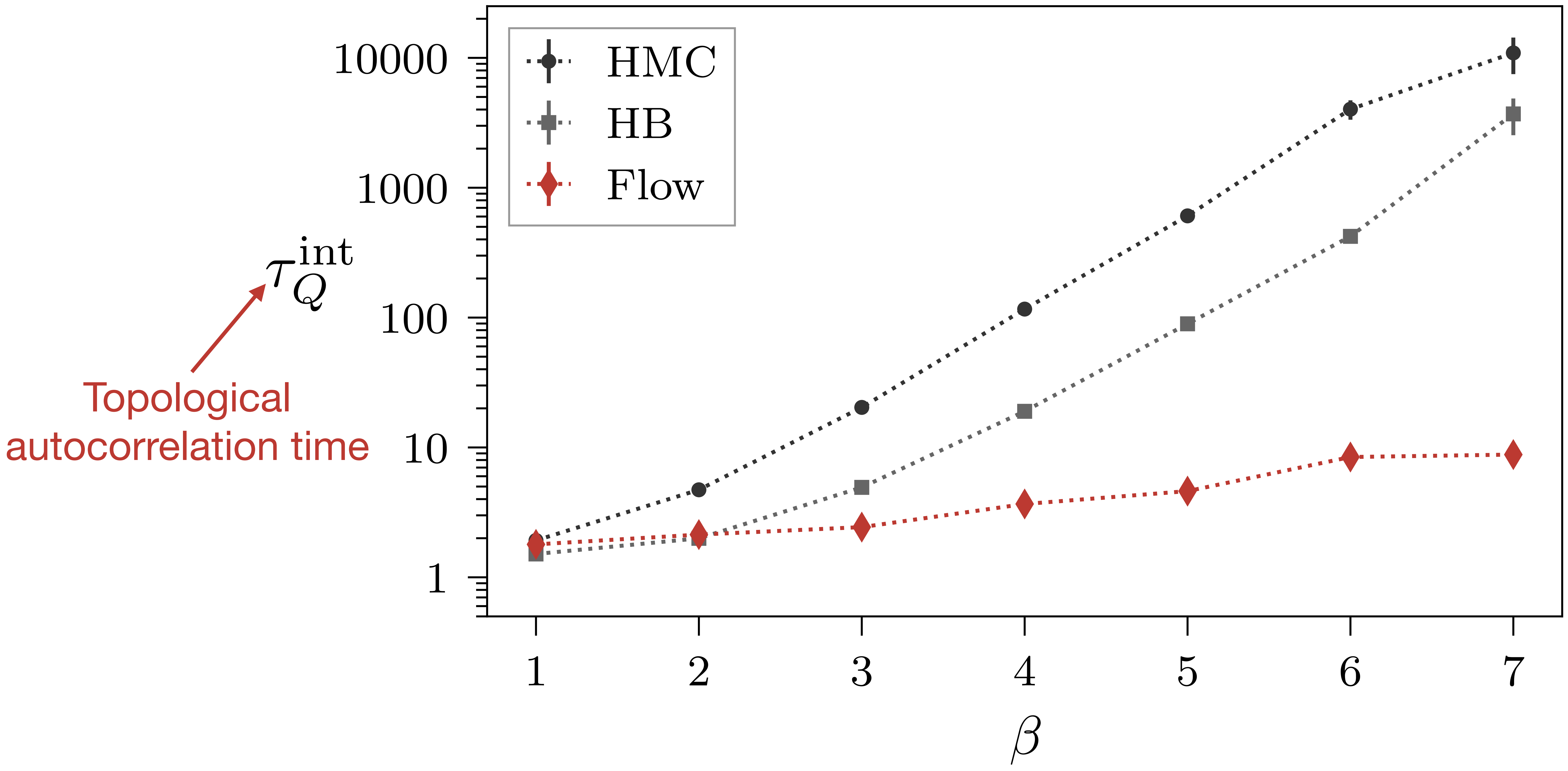
- Compared **flow, analytical, HMC**, and **heat bath** on 16×16 lattices for $\beta = \{1,\ldots,7\}$
- Topo freezing in HMC and heat bath
- Gauge-equiv flow-based model at each β
- Flow-based MCMC observables agree



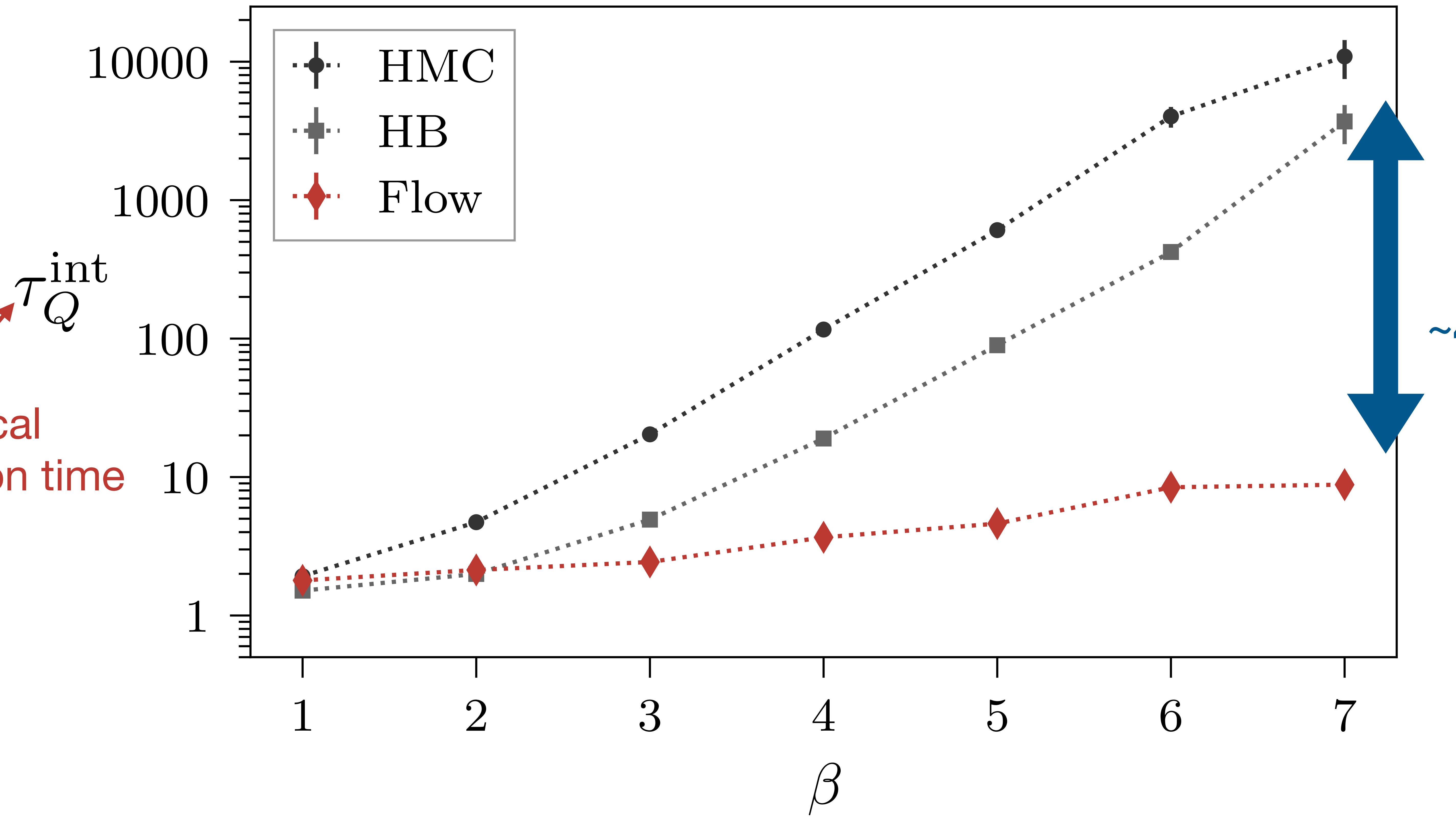
Topological susceptibility $\chi_Q = \langle Q^2/V \rangle$

[GK, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan **PRL125 (2020) 121601**]

U(1) topological freezing mitigated



U(1) topological freezing mitigated



Topological autocorrelation time

Autocorrelation time reduced by **~4 orders of magnitude!**

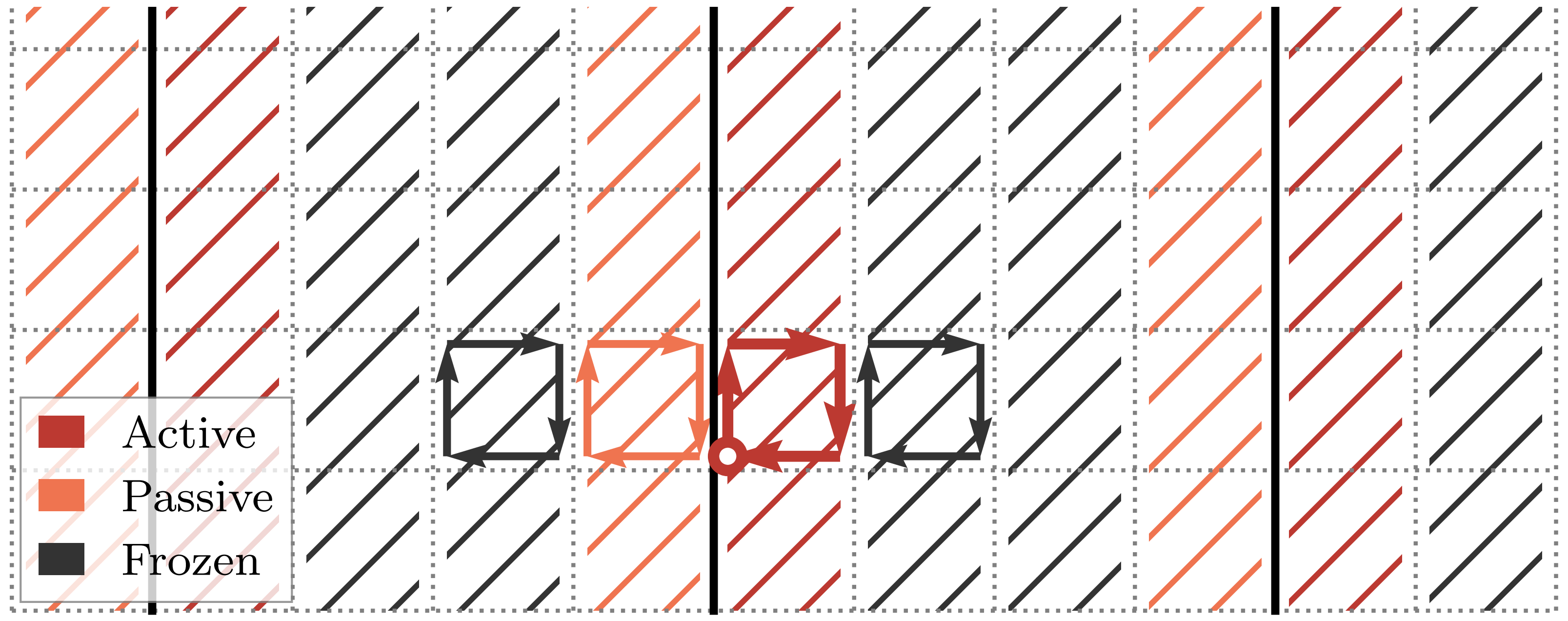
SU(N) gauge theory in 1+1D

$$S(U) = -\frac{\beta}{N}\sum_x \sum_{\mu<\nu} \mathrm{ReTr}\, P_{\mu\nu}(x)$$

$$P_{\mu\nu}(x) = U_\mu(x)U_\nu(x+\hat{\mu})U_\mu^\dagger(x+\hat{\nu})U_\nu^\dagger(x)$$

Gauge-equiv flow models for 2D lattice gauge theory on 16×16 lattices.

- Matched 't Hooft couplings:
  $SU(2) \iff \beta = \{1.8, 2.2, 2.7\}$
  $SU(3) \iff \beta = \{4.0, 5.0, 6.0\}$



- 48 **PAFF coupling layers**, links updated 6 times each

- No equivalent to $U(1)$ topological freezing, studied **absolute model quality**

[Boyda, GK, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan **PRD103 (2021) 074504**]

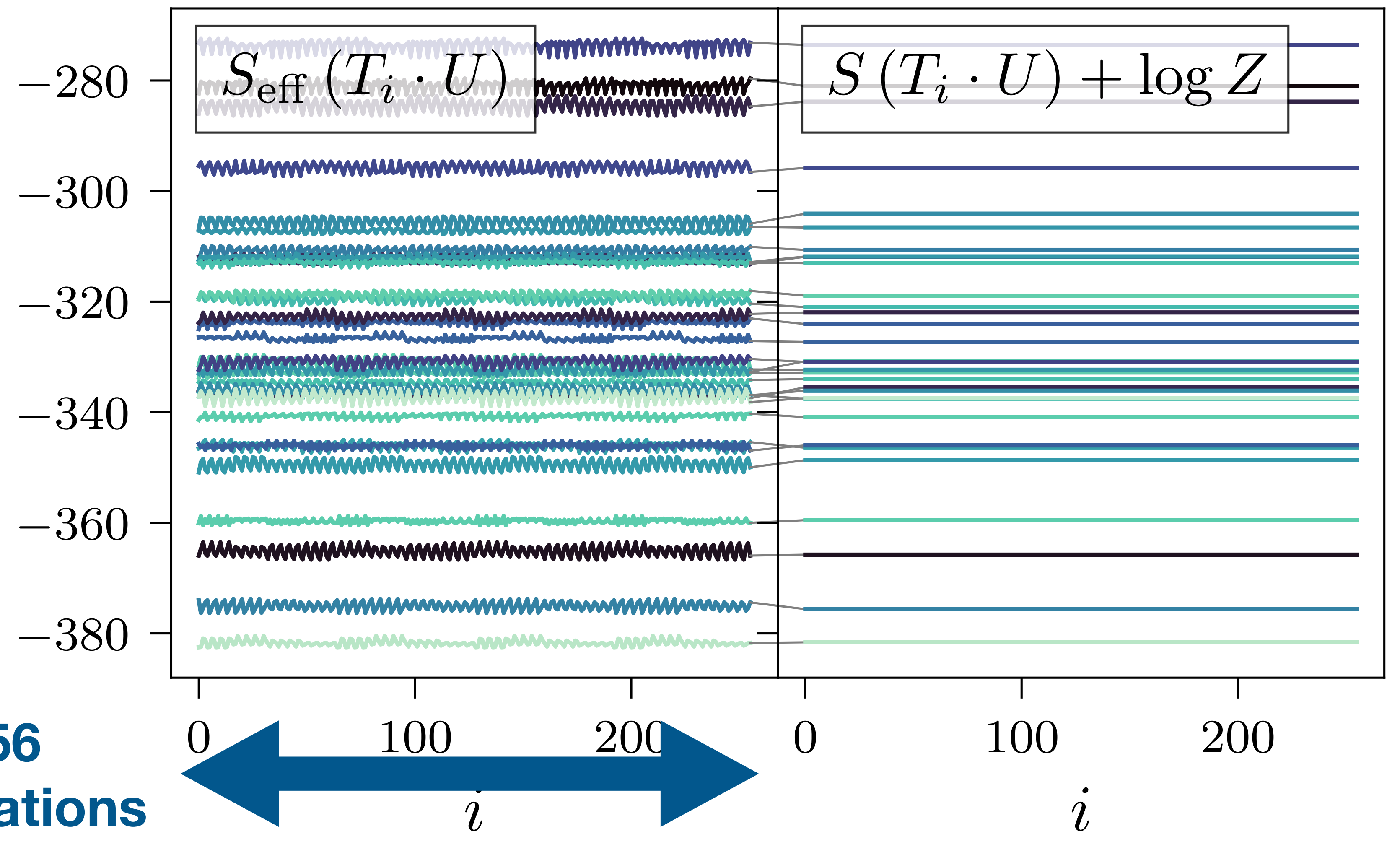
Results for SU(2) and SU(3) gauge theory

- Flow-based MCMC observables agree with analytical
- **High-quality models:** autocorrelation time in flow-based Markov chain τ_{int} = 1—4

	SU(2)			SU(3)		
β	1.8	2.2	2.7	4.0	5.0	6.0
ESS(%)	91	80	56	88	75	48

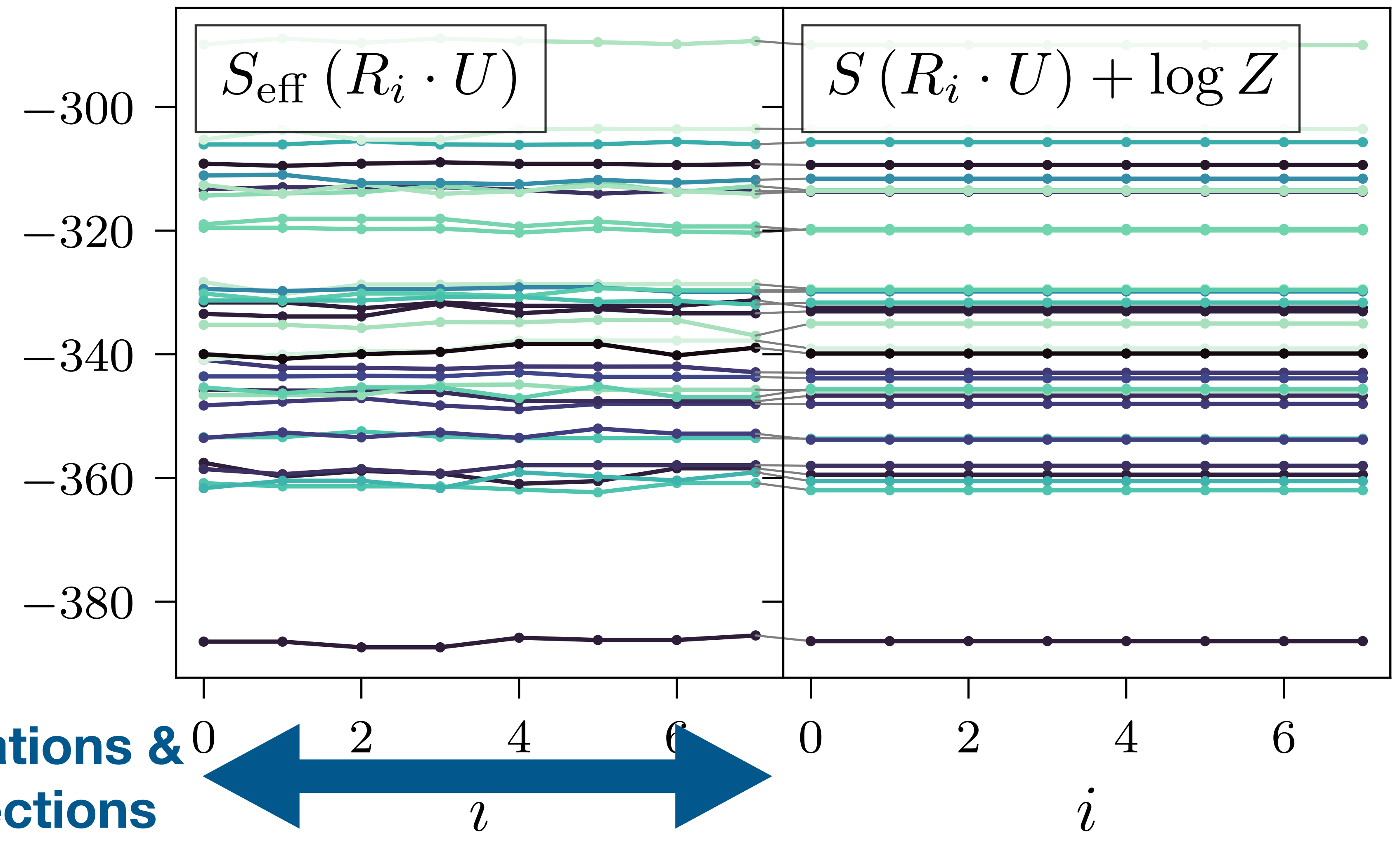
✓ “Effective Sample Sizes” indicating model overlap onto target are all larger than ~50% (100% = perfect model)

Exact translational subgroup; residual learned



256 translations

Rotation and reflection symmetry learned



8 rotations & reflections

Results for SU(2) and SU(3) gauge theory

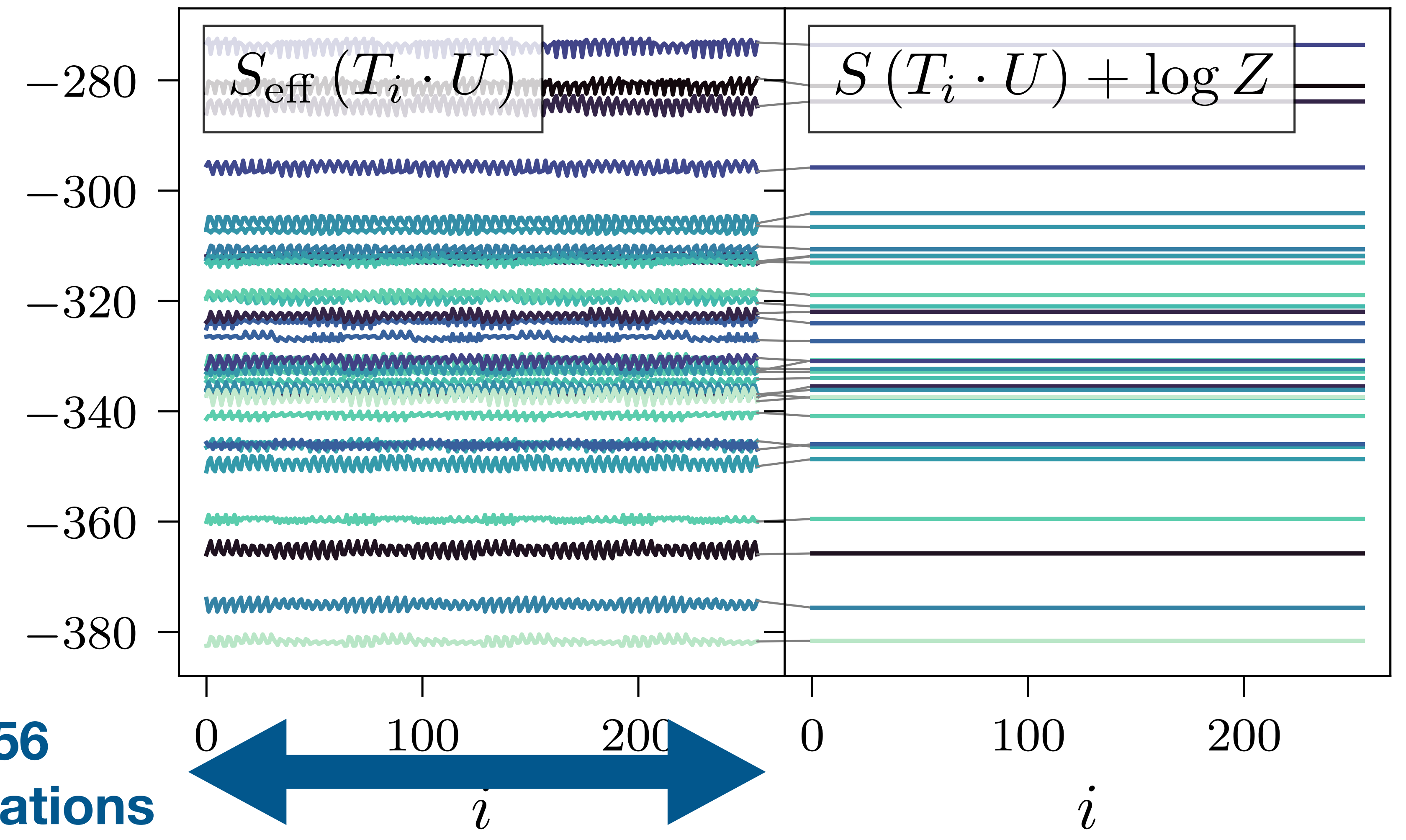
- Flow-based MCMC observables agree with analytical

- **High-quality models:** autocorrelation time in flow-based Markov chain τ_{int} = 1—4

	SU(2)			SU(3)		
β	1.8	2.2	2.7	4.0	5.0	6.0
ESS(%)	91	80	56	88	75	48

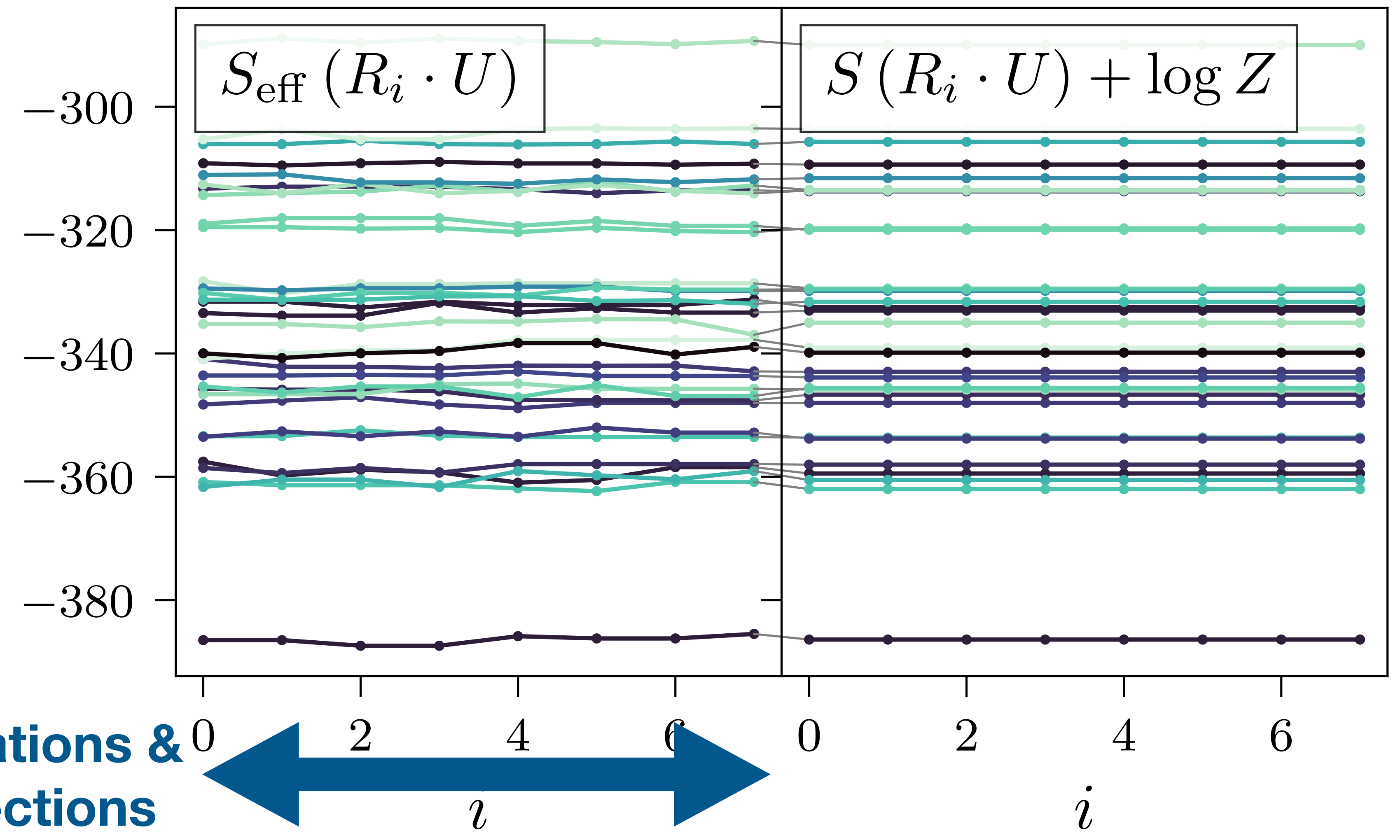
✓ "Effective Sample Sizes" indicating model overlap onto target are all larger than ~50% (100% = perfect model)

Exact translational subgroup; residual learned



256 translations

Rotation and reflection symmetry learned



8 rotations & reflections

Promising early results. No theoretical obstacle to scaling to 4D $SU(N)$ lattice gauge theory.

A story of symmetries & generative models

(In three parts)

1. Flow-based generative models

2. Gauge symmetry & translational symmetry

3. **Fermions & translational symmetry**

Fermions in field theory

Grassmann representation in path integral means…

… we cannot sample fermion fields

… integrating out fermions results in costly fermion determinants

$$\int \mathcal{D}\psi \mathcal{D}\bar{\psi} \prod_f e^{-\bar{\psi}_f D_f \psi_f} = \prod_f \det D_f$$

Pseudofermions used in standard MCMC for theories with dynamical fermions.

$$\int \mathcal{D}\psi \mathcal{D}\bar{\psi} \prod_f e^{-\bar{\psi}_f D_f \psi_f} \propto \int \mathcal{D}\varphi \mathcal{D}\varphi^\dagger \prod_k e^{-\varphi_k^\dagger \mathcal{M}_k^{-1} \varphi_k}$$

Starting point for flow-based sampling

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

5 ways to marginalize

Any could in principle be learned by flow-based models.

Below: Bosonic part of action written generically as $S_B(\phi)$

Name	Probability density
Joint[A]	$p(\phi, \varphi) = \frac{1}{Z} \exp(-S_B(\phi) - \varphi^\dagger [\mathcal{M}(\phi)]^{-1} \varphi)$
ϕ-marginal	$p(\phi) = \frac{Z_\mathcal{N}}{Z} \exp(-S_B(\phi)) \det \mathcal{M}(\phi)$
φ-conditional[A,B]	$p(\varphi \vert \phi) = \frac{1}{Z_\mathcal{N} \det \mathcal{M}(\phi)} \exp(-\varphi^\dagger [\mathcal{M}(\phi)]^{-1} \varphi)$
φ-marginal[C]	$p(\varphi) = \frac{1}{Z} \int d\phi \, \exp(-S_B(\phi) - \varphi^\dagger [\mathcal{M}(\phi)]^{-1} \varphi)$
ϕ-conditional[A]	$p(\phi \vert \varphi) = \frac{\exp(-S_B(\phi) - \varphi^\dagger [\mathcal{M}(\phi)]^{-1} \varphi)}{\int d\phi \, \exp(-S_B(\phi) - \varphi^\dagger [\mathcal{M}(\phi)]^{-1} \varphi)}$

Expensive to evaluate det exactly

Can actually be sampled directly (e.g. pseudofermion refresh in HMC)

Intractable density (even unnormalized)

Translational invariance

Pseudofermion fields $\varphi(x)$ satisfy **antiperiodic BCs** in the time direction.

Marginalizations with PFs should address this for translational equivariance.

Building blocks:

Explicit averaging:

- CNN outputs averaged over time translations with correct BCs

Restricted CNNs:

- Channels either **antiperiodic (AP)** or **periodic (P)** in time
- Operations restricted for well-defined outputs (either P or AP)
- AP activations only odd fns

Invertible linear layers:

- Flow = composed linear operators $\mathscr{W}_n \circ \ldots \circ \mathscr{W}_1$
- Each $\mathscr{W}_i$ is a conv with fixed direction (and correct BCs)

$$\begin{bmatrix} a_1 & & & & \pm b_1 \\ b_2 & a_2 & & 0 & \\ & \cdots & \cdots & & \\ & 0 & & & \\ & & & b_L & a_L \end{bmatrix}$$

$$\det \mathscr{W}_i = \prod_k a_k \pm \prod_k b_k$$

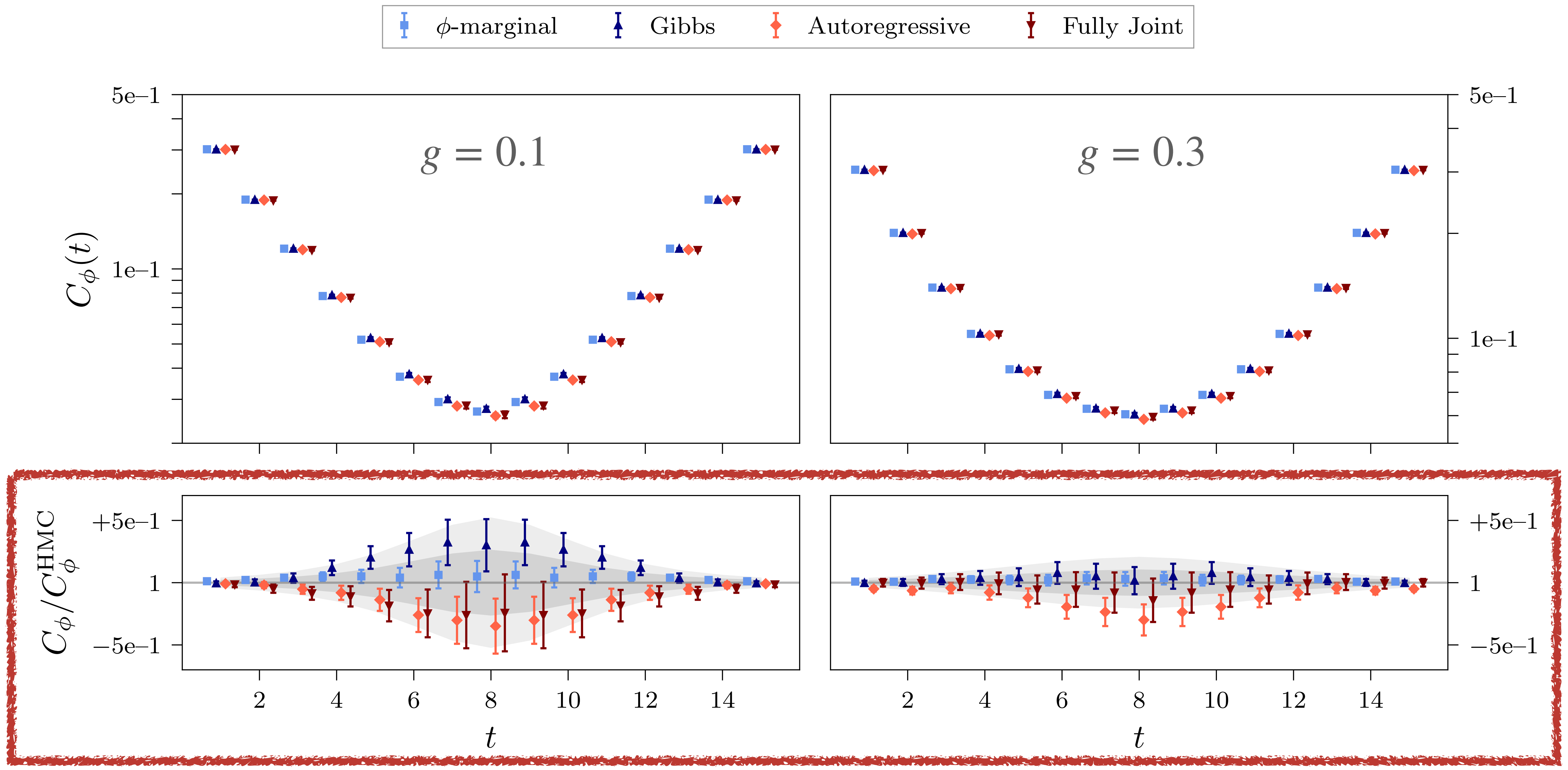
Results for Yukawa model

Staggered Dirac op with Yukawa coupling $g\phi\bar{\psi}\psi$ and mass term $M\bar{\psi}\psi$

Studied 2D ϕ^4 model coupled via Yukawa interaction to staggered ψ

$$S(\phi,\psi) = \sum_{x\in\Lambda}[-2\sum_{\mu=1}^{d}\phi(x)\phi(x+\hat{\mu}) + (m^2+2d)\phi(x)^2 + \lambda\phi(x)^4] + \sum_{f=1}^{N_f}\bar{\psi}_f D_f[\phi]\psi_f$$

- 16×16 lattices
- Two degenerate fermions ($N_f = 2$)
- Massless ($M = 0$)
- Variety of models, all 4 sampling schemes



✓ Correlation functions effectively reproduced

Summary and Outlook

Symmetries allow efficient & consistent training of flow-based models.

Gauge symmetry + **translational** symmetry addressed throughout.

Effective models produced for $U(1)$, $SU(2)$, $SU(3)$ lattice gauge theory and a ϕ^4 Yukawa model in 1+1D.

Future directions:

1. Higher spacetime dims
2. Tuning of training hyperparameters
3. Efficient model architectures at scale?

Summary and Outlook

Symmetries allow efficient & consistent training of flow-based models.

Gauge symmetry + **translational** symmetry addressed throughout.

Effective models produced for $U(1)$, $SU(2)$, $SU(3)$ lattice gauge theory and a ϕ^4 Yukawa model in 1+1D.

Future directions:

1. Higher spacetime dims
2. Tuning of training hyperparameters
3. Efficient model architectures at scale?

See also:

Approaches to multimodal sampling and mixed HMC + flow-based sampling:

[Hackett, Hsieh, Albergo, Boyda, Chen, Chen, Cranmer, GK, Shanahan; **2107.00734**]

Jupyter notebook tutorial:

[Albergo, Boyda, Hackett, GK, Cranmer, Racanière, Rezende, Shanahan; **2101.08176**]

Backup Slides

Exactness: Flow-based MCMC

Markov chain constructed using Independence Metropolis accept/reject on model proposals.

"Embarrassingly parallel" step!

- **Independent** proposals U' from model distribution q
- **Accept** proposal U', making it next elt of Markov chain, with probability

$$p_{\mathrm{acc}}(U \to U') = \min\left(1, \frac{p(U')}{q(U')}\frac{q(U)}{p(U)}\right).$$

- If **rejected**, duplicate previous elt of Markov chain
    - Only need to compute observables on duplicated elts once!

Exactness: Reweighting

- Also possible to reweight independently drawn samples:

$$\langle \mathcal{O} \rangle = \frac{\int \mathcal{D}U\, q(U) \left[\mathcal{O}(U) \frac{p(U)}{q(U)} \right]}{\int \mathcal{D}U\, q(U) \left[\frac{p(U)}{q(U)} \right]}$$

- May be preferable when observables $\mathcal{O}(U)$ are efficiently computed, and sampling is expensive.

- Observables $\mathcal{O}(U)$ are expensive in lattice QCD. We prefer resampling or MCMC approaches in these settings.

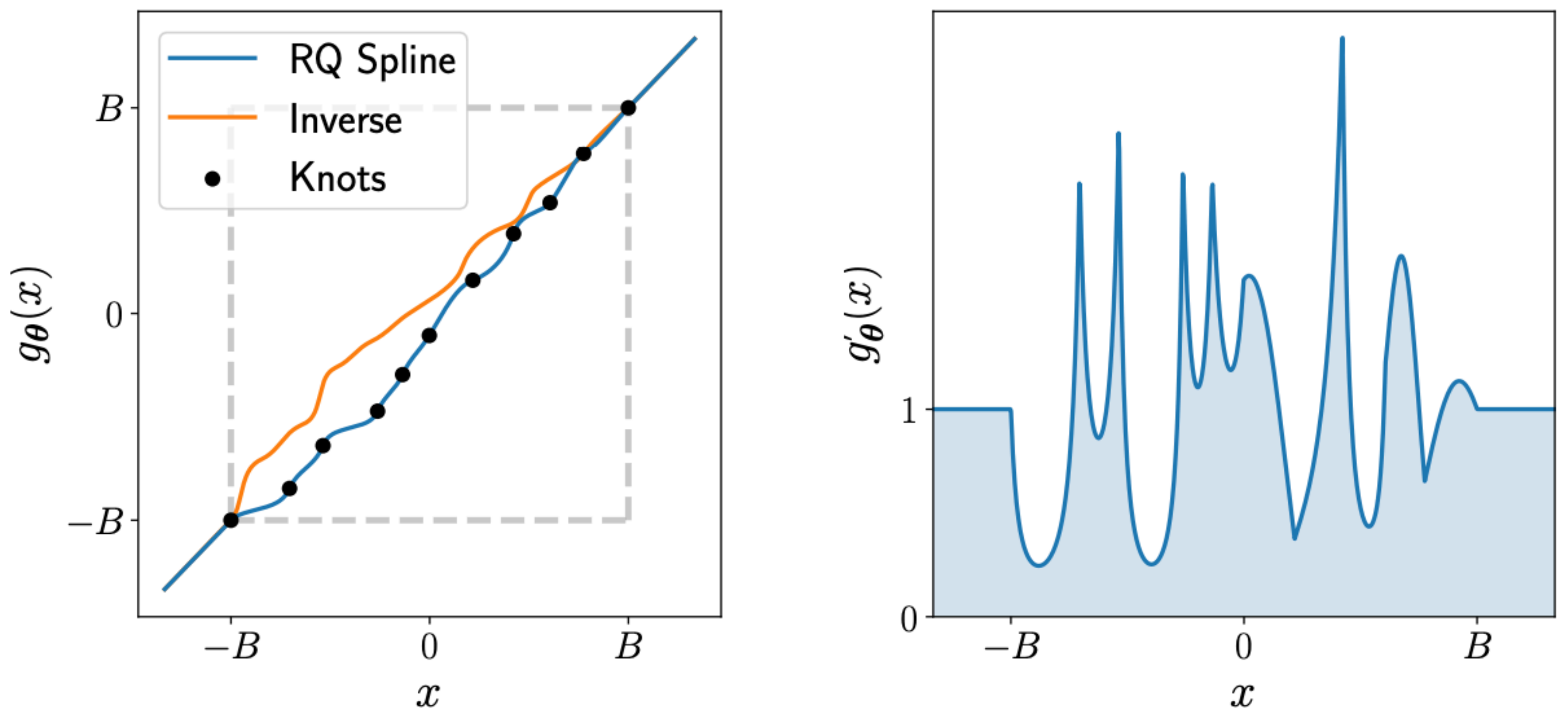
U(1) kernels

Conjugation equivariance trivially satisfied: $h(\Omega W \Omega^\dagger) = h(W) = \Omega h(W) \Omega^\dagger$.

Invertible maps on U(1) variables:

- **Periodic / compact domain** must be addressed.
- For details, see:
  [Rezende, Papamakarios, Racanière, Albergo, GK, Shanahan, Cranmer; **ICML (2020) 2002.02428**]

[Durkan, Bekasov, Murray, Papamakarios **1906.04032**]



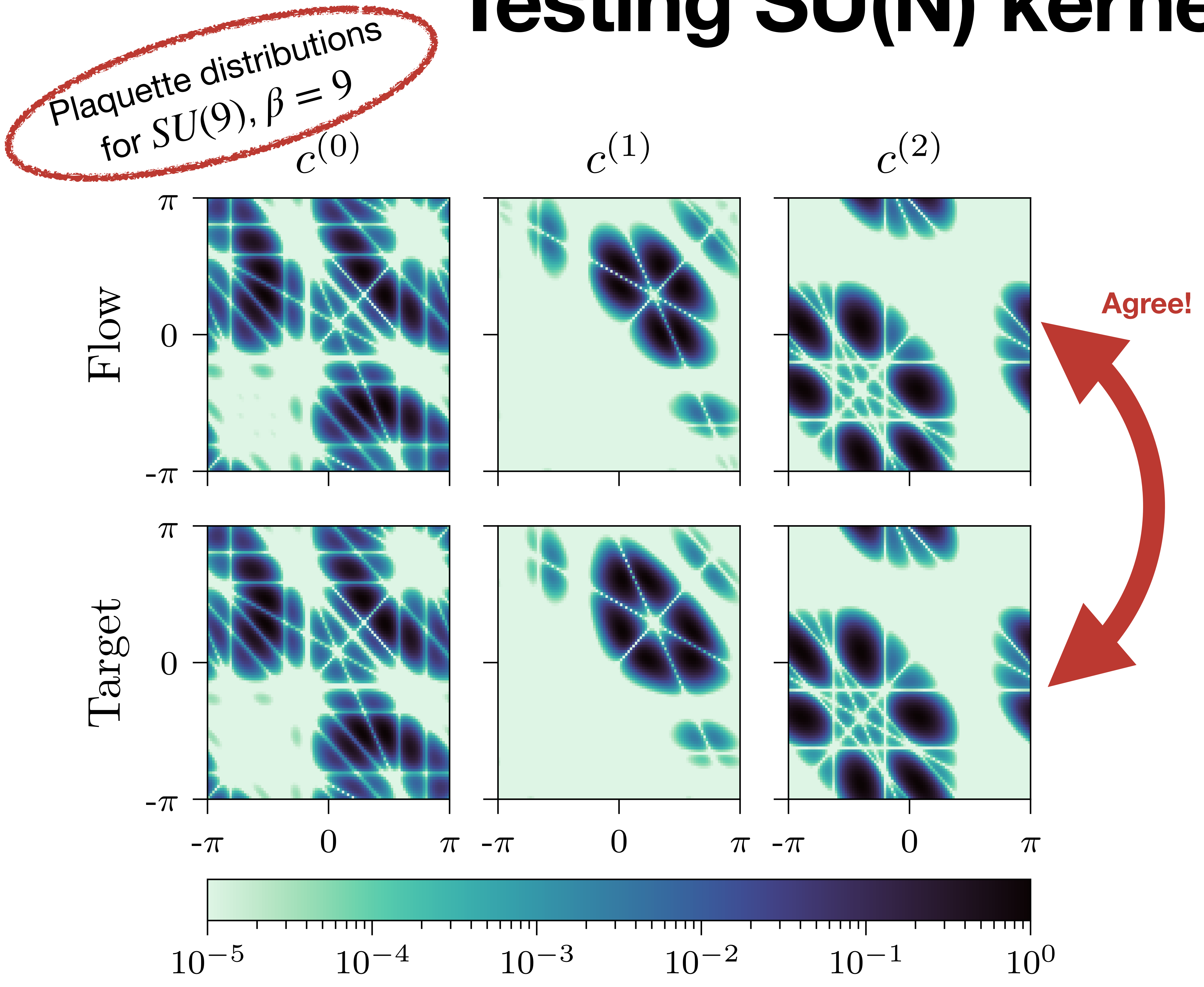
Non-compact projection:

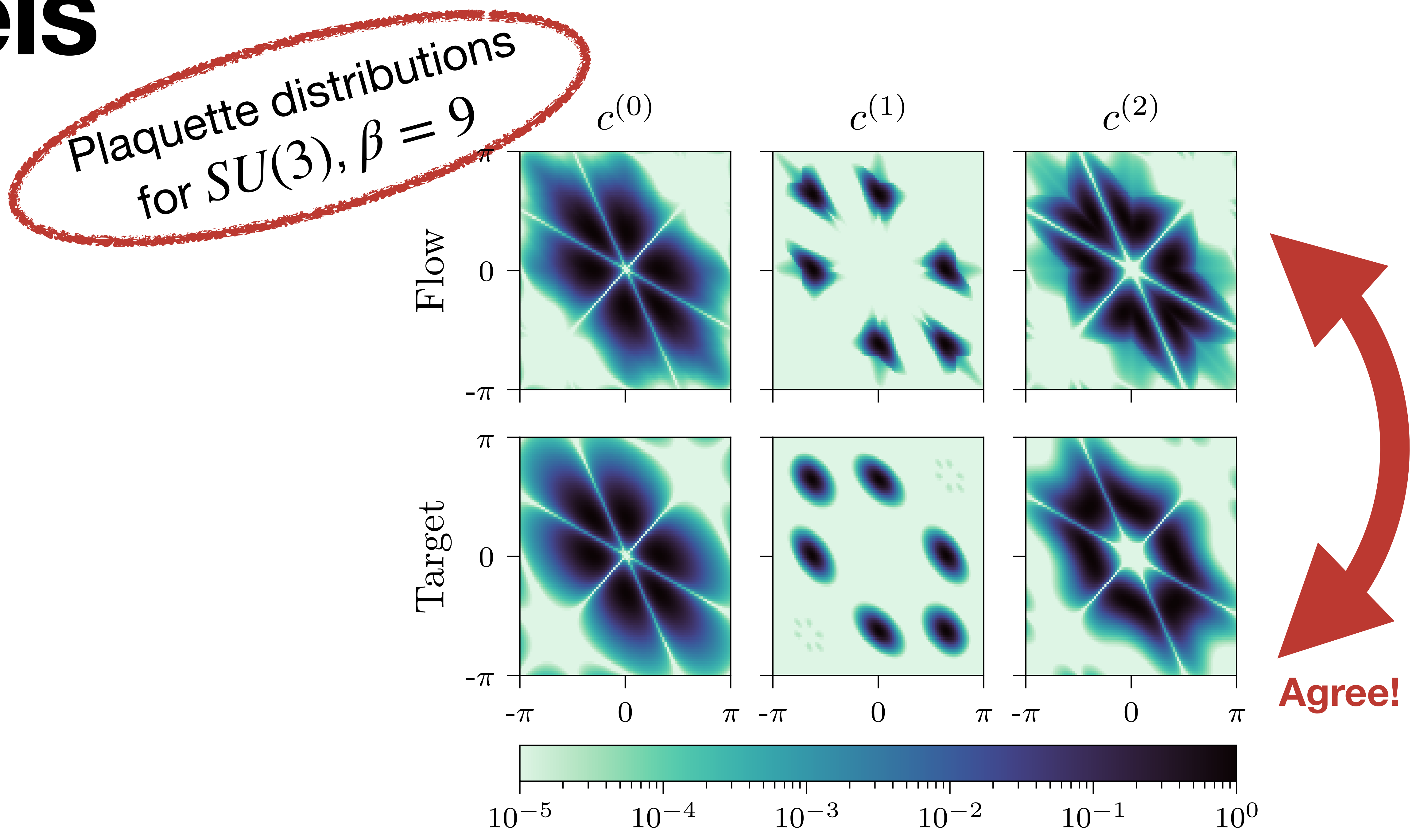
- Map $\theta \to x \in \mathbb{R}$, e.g. $\arctan(\theta/2)$
- Transform $x \to x'$ as usual
- Map $x' \to \theta' \in [-\pi, \pi]$

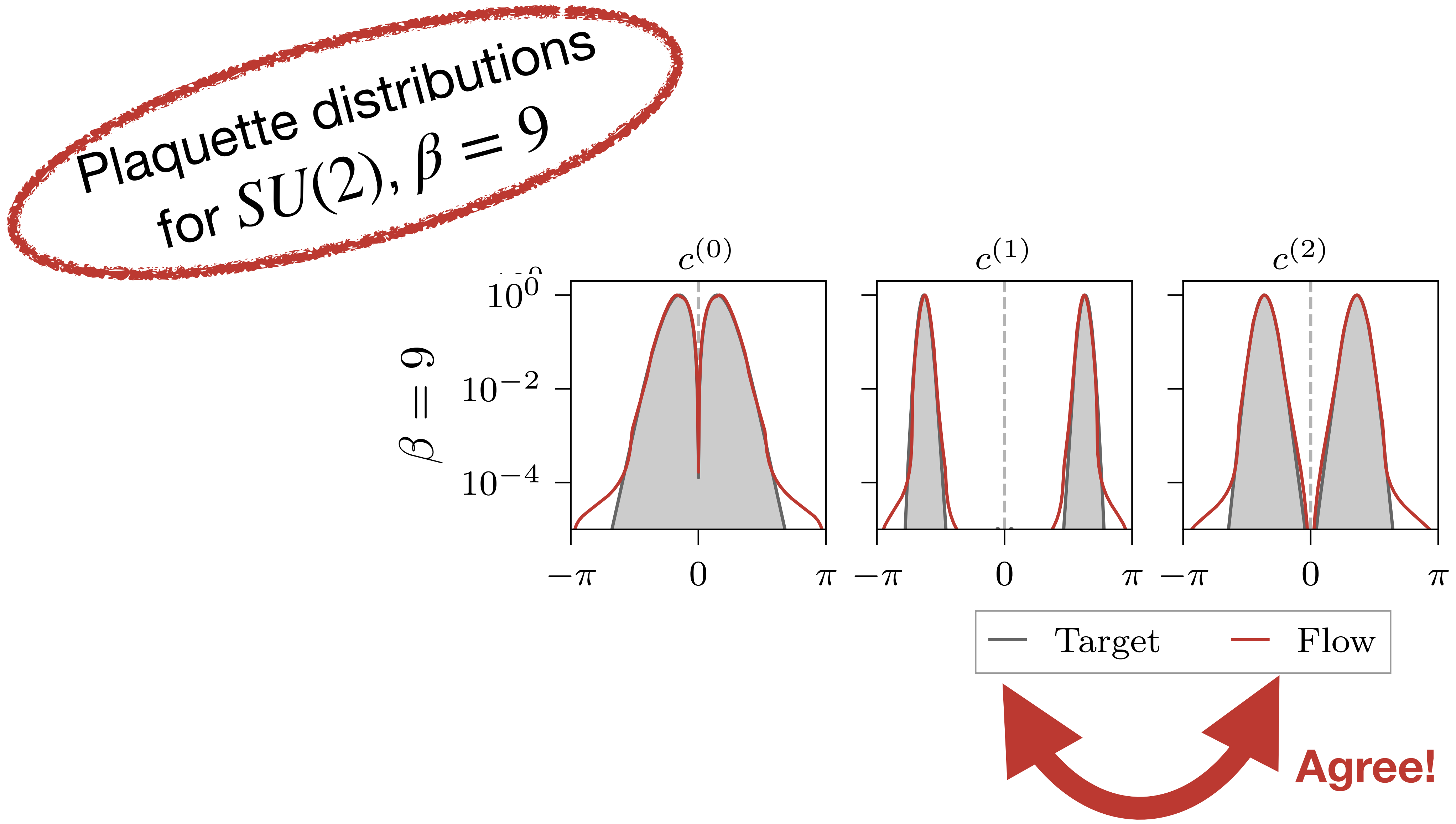
Circular invertible splines:

- Spline "knots" trainable fns
- Identify endpoints π and $-\pi$
- Number of knots $\leftrightarrow$ expressivity

Testing SU(N) kernels

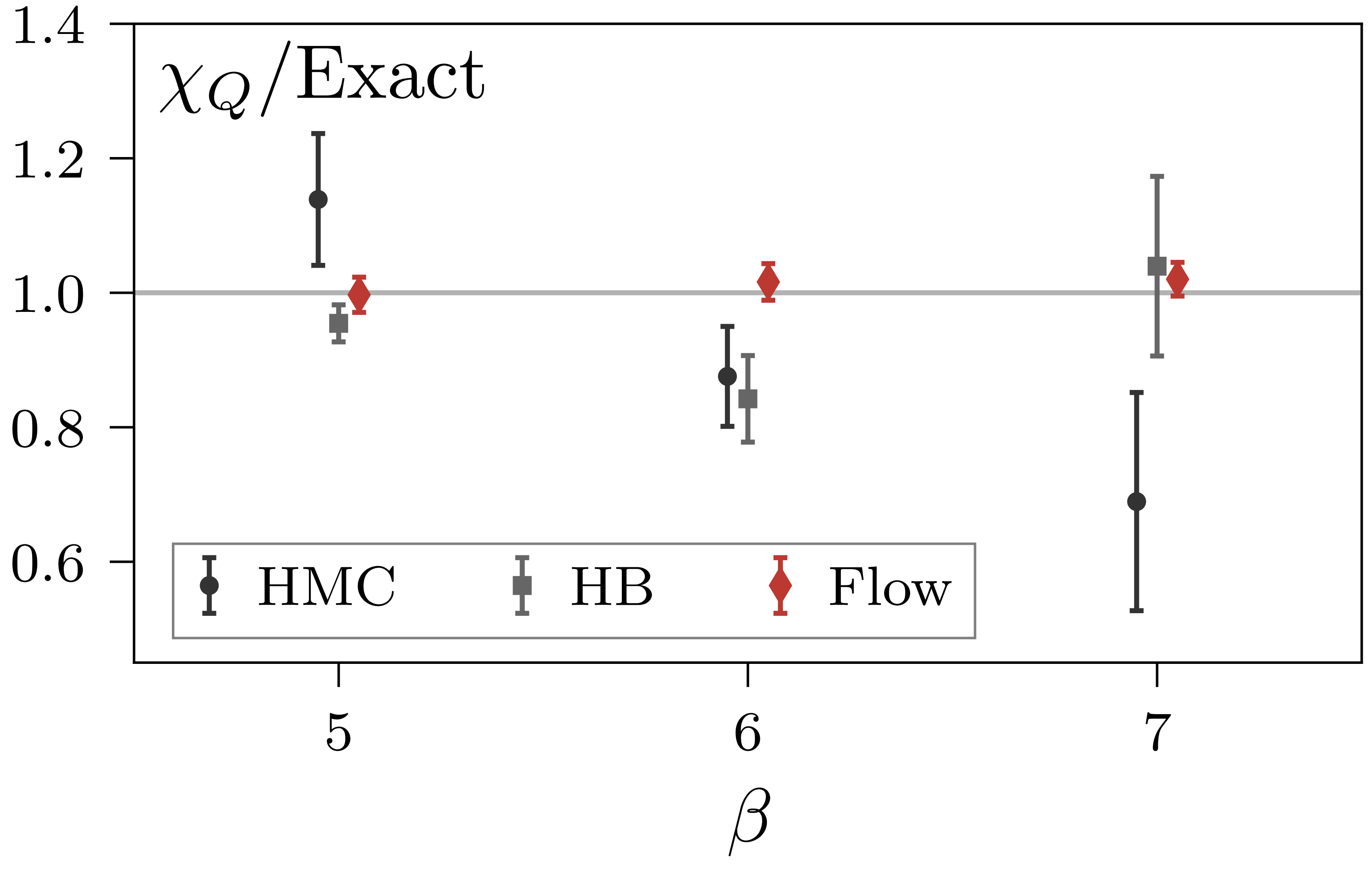


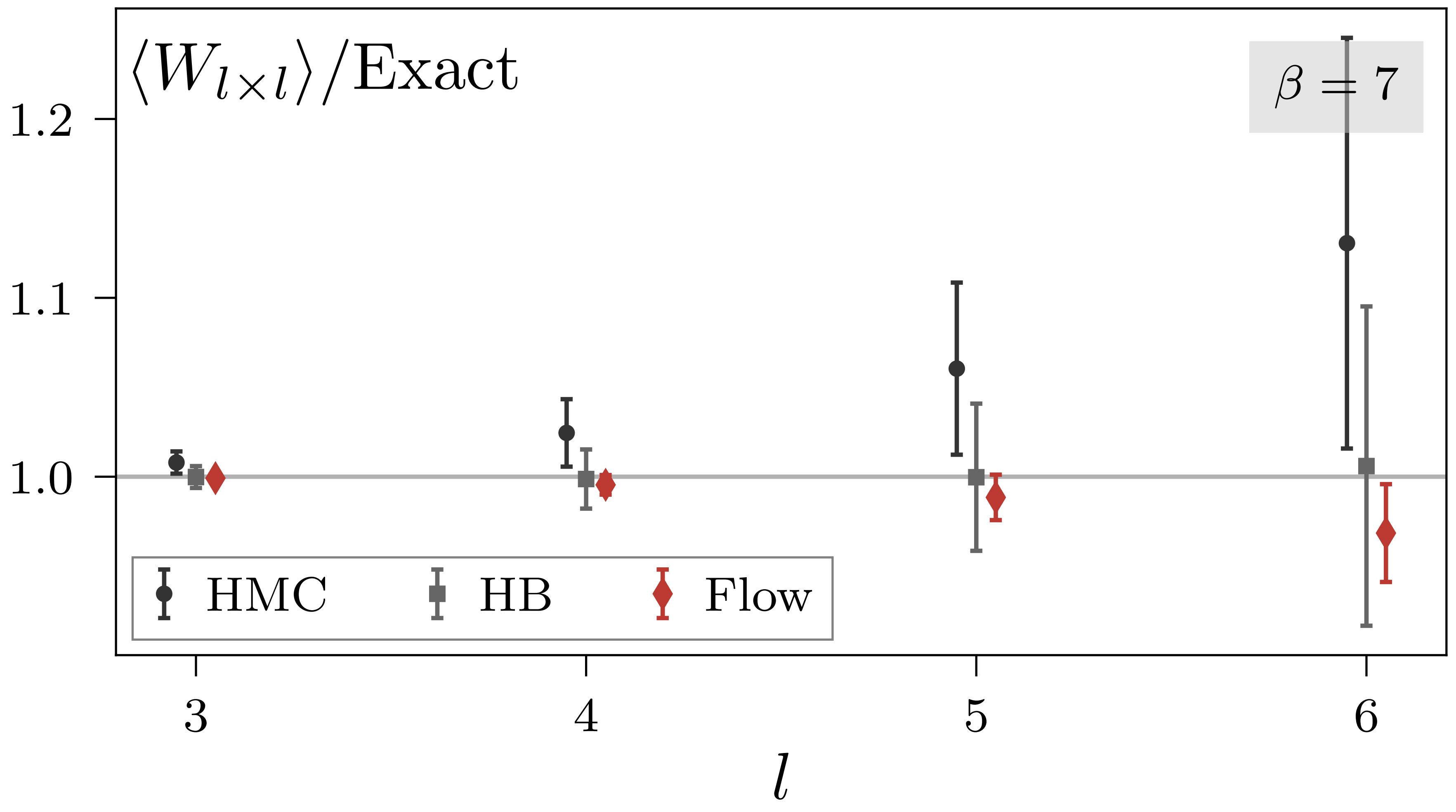


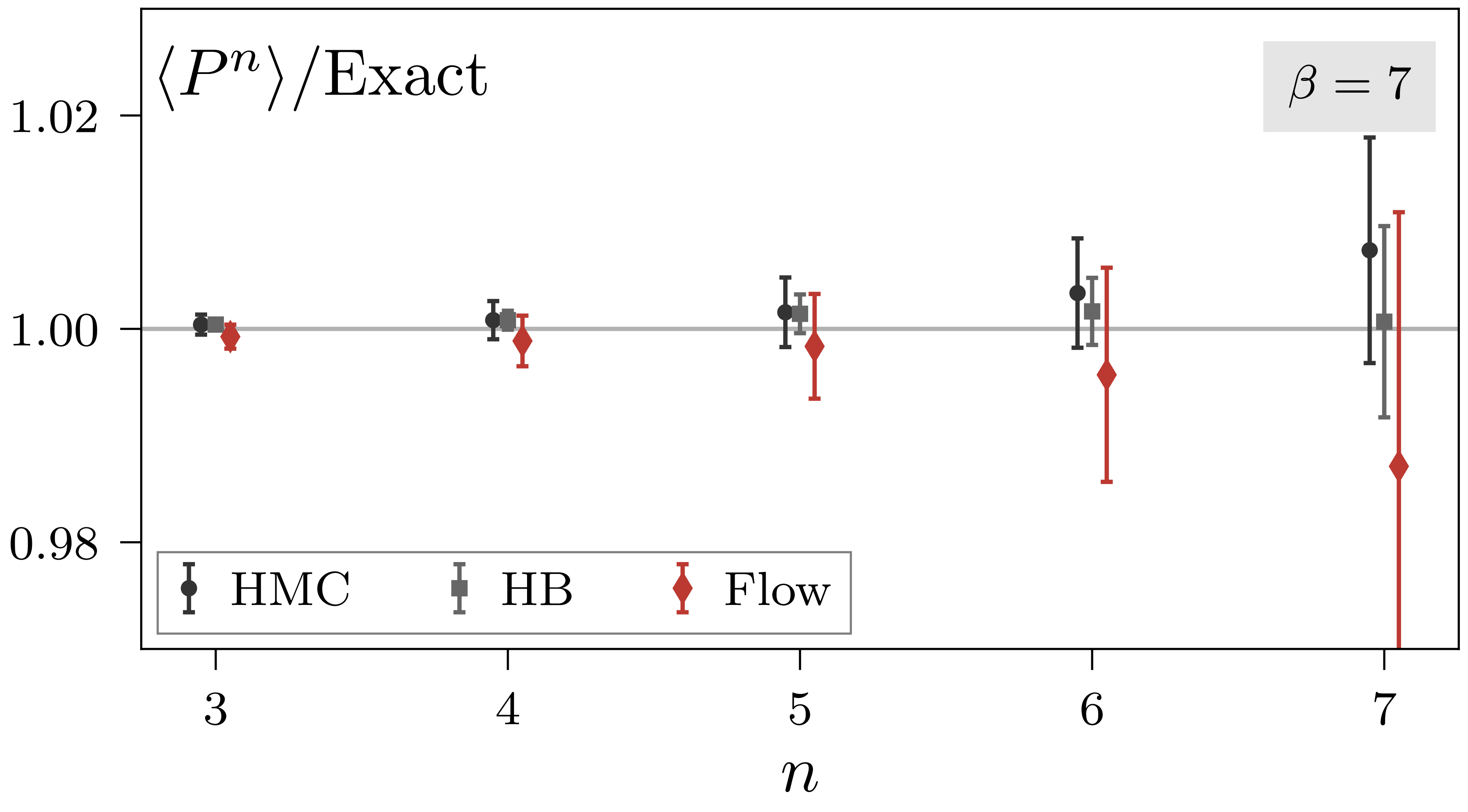


Density has zeros on vertical, horizontal, and diagonal lines where the slice crosses walls of cells

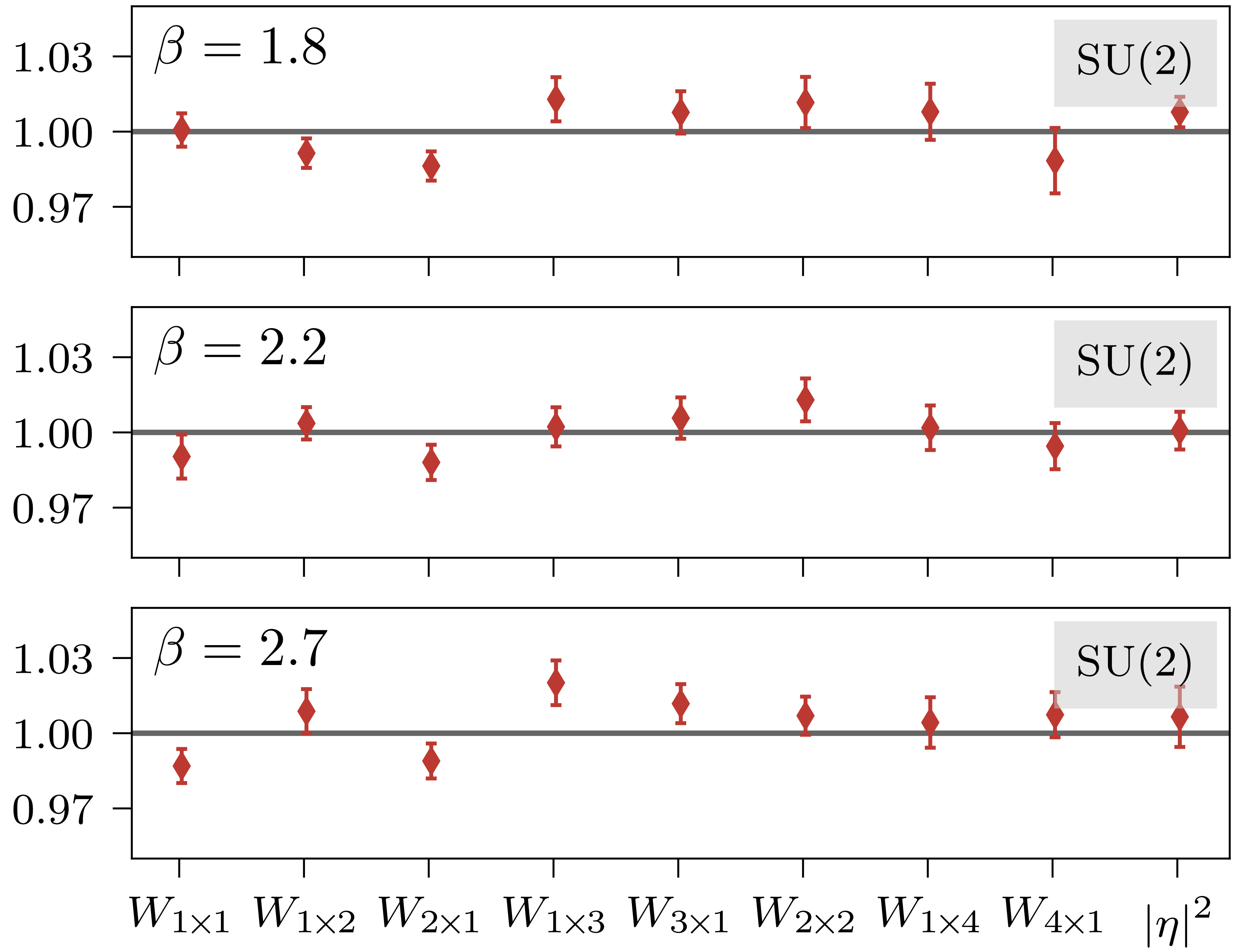
U(1) observables

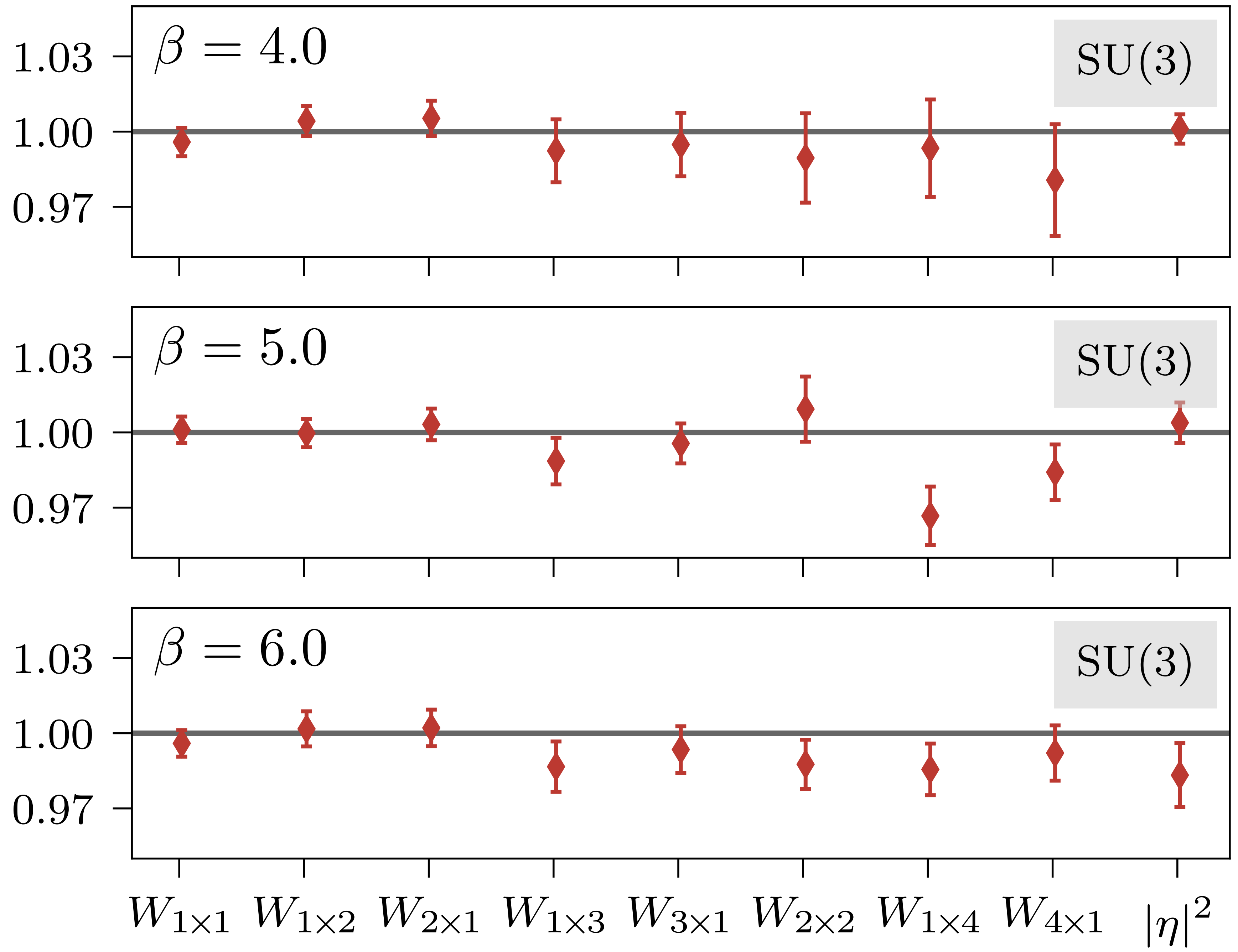






SU(N) observables



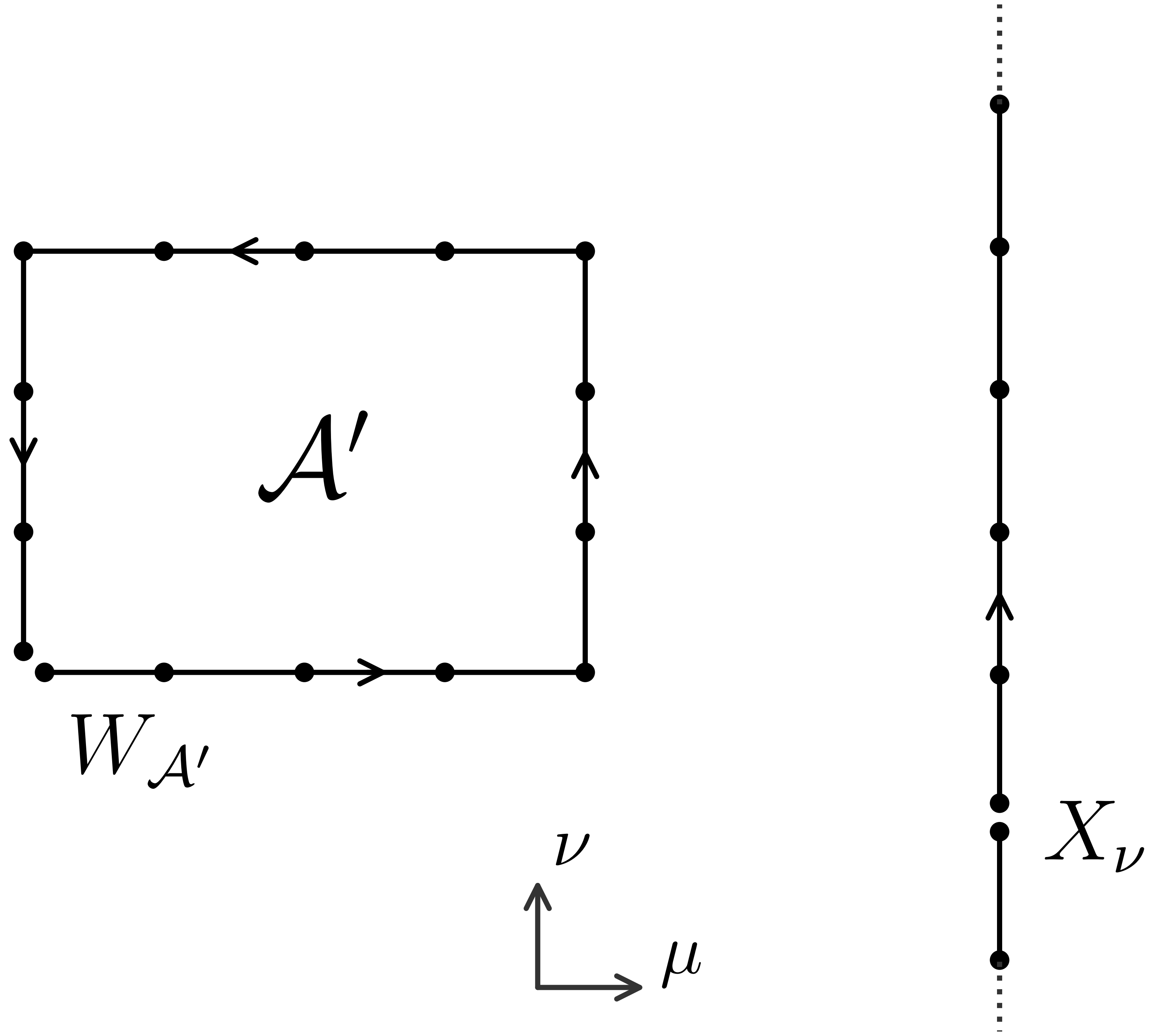


Center symmetry

Using **only contractible loops** in coupling layers enforces center symmetry.

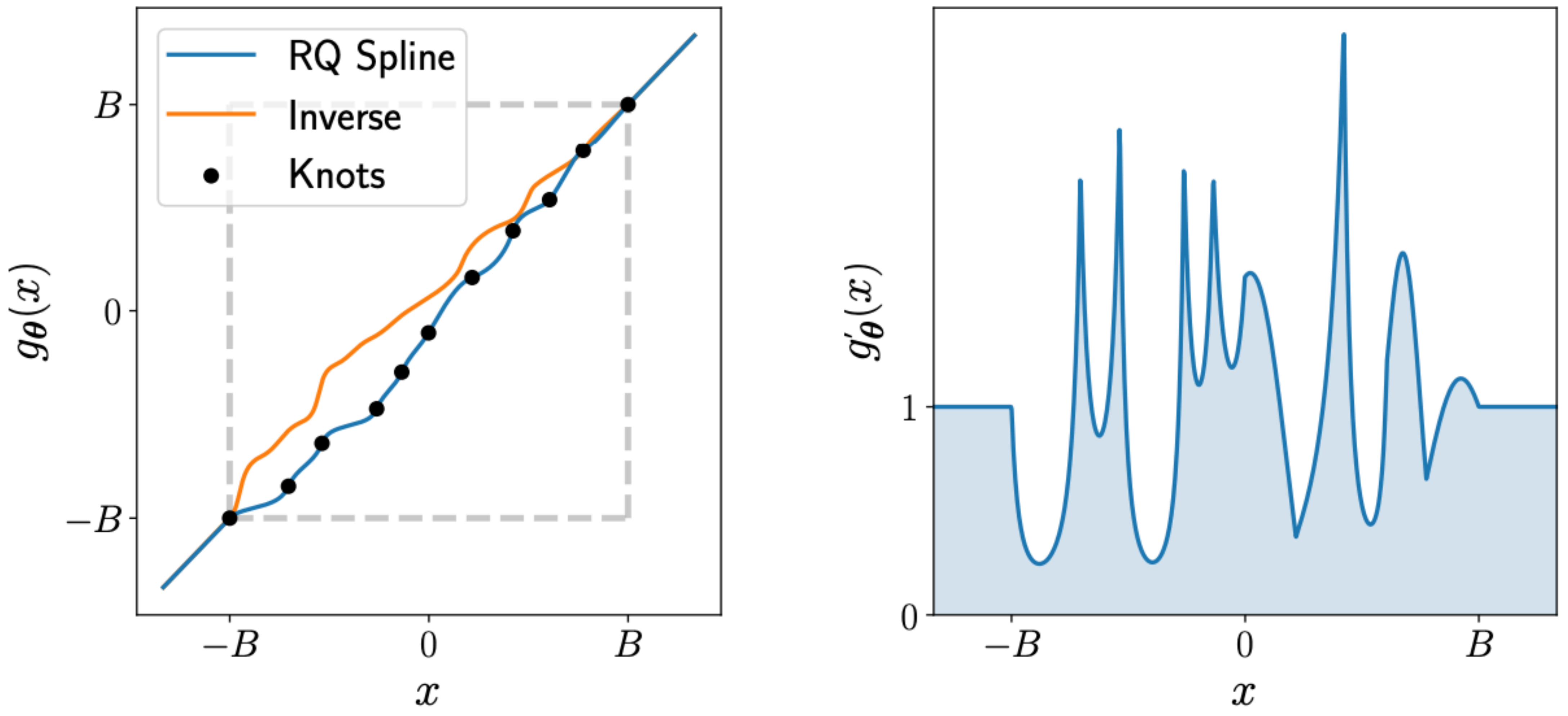
Fundamental fermions:

- Center symmetry explicitly broken
- Must include non-contractible loops (e.g. Polyakov) in the set of frozen and/or transformed loops



Details of $SU(2)$ models

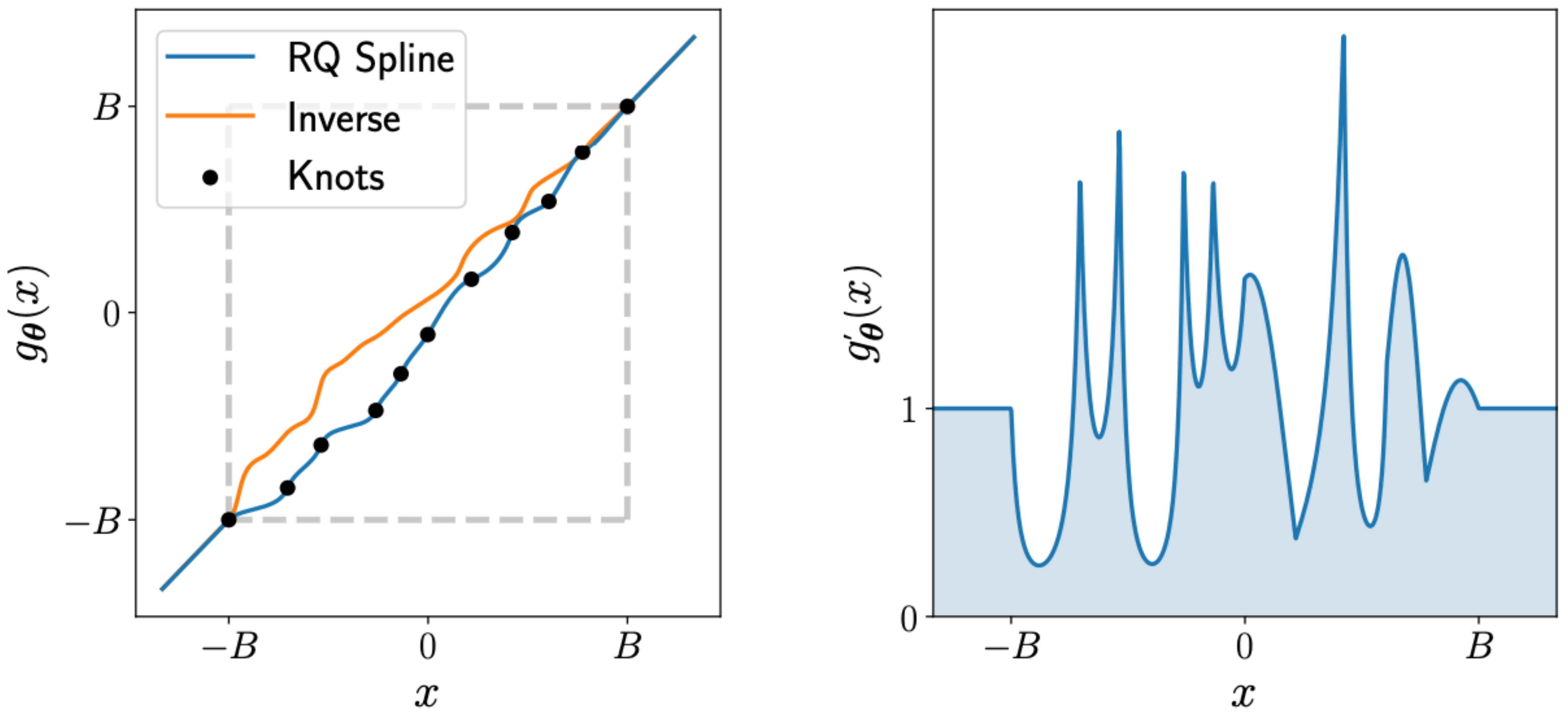
- Inner flow on open box Ω is a spline flow with **4 knots**
  - B and $-B$ boundaries align to 0 and 1 edges of the open box
- CNNs to compute the knot locations
  - 32 hidden channels
  - 2 hidden layers



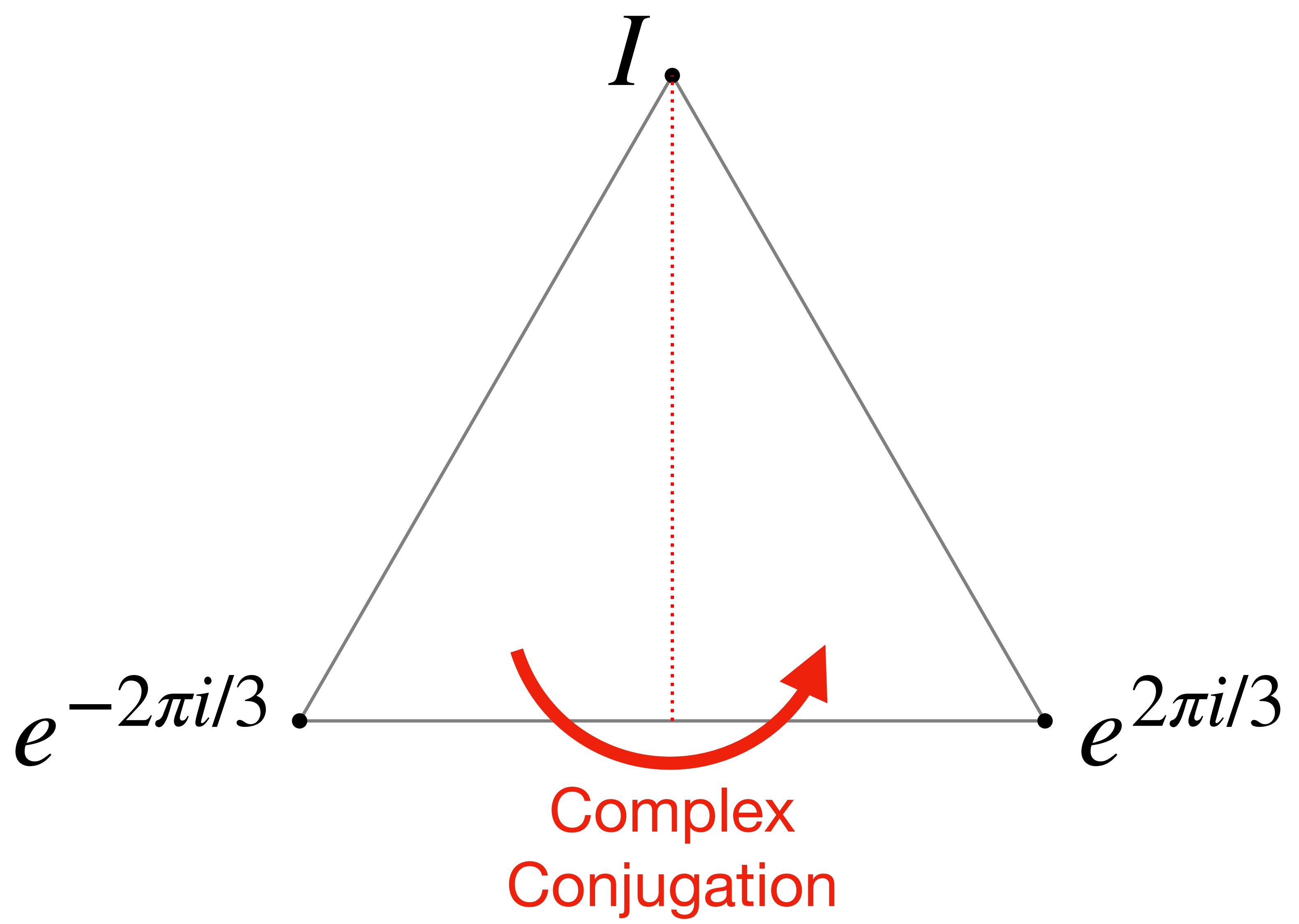
[Durkan, Bekasov, Murray, Papamakarios 1906.04032]

Details of $SU(3)$ models

- Inner flow on open box Ω is a spline flow with **16 knots**
    - B and $-B$ boundaries align to 0 and 1 edges of the open box
- CNNs to compute the knot locations
    - 32 hidden channels
    - 2 hidden layers
- Exact conjugation equivariance also imposed



[Durkan, Bekasov, Murray, Papamakarios 1906.04032]

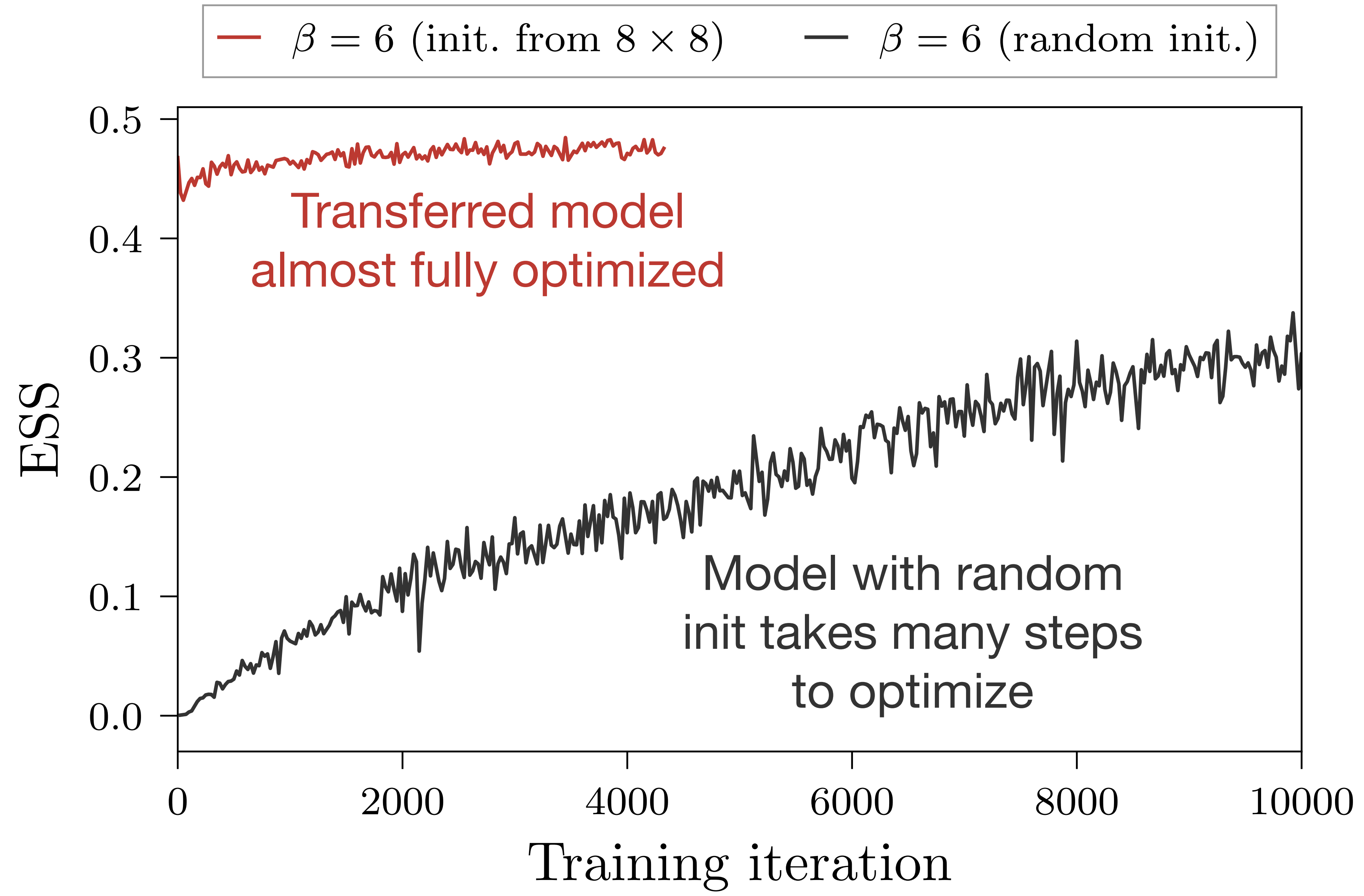


Gauge theory model training

- Adam optimizer ~ stochastic grad. descent with momentum
  - Batches of size 3072 per gradient descent step
  - Monitored value of **effective sample size** (ESS)

$$\mathrm{ESS} = \frac{\left(\frac{1}{n}\sum_i w(U_i)\right)^2}{\frac{1}{n}\sum_i w(U_i)^2}, \quad U_i \sim q(U)$$

$$w(U) = p(U)/q(U)$$ "reweighting factors"



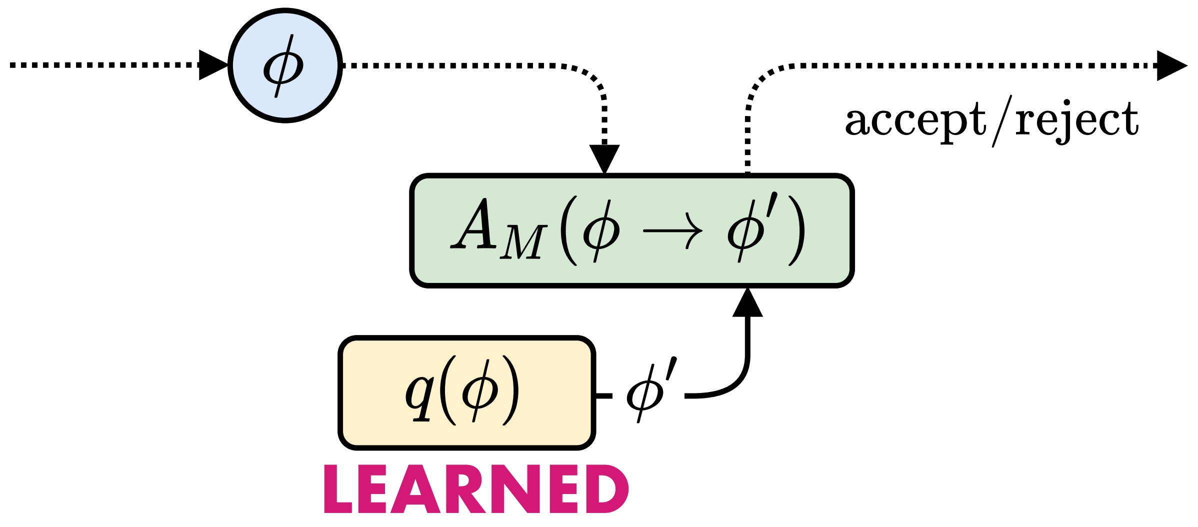
- **Transfer learning:** model trained first on 8×8 then used to initialize model for training on 16×16

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

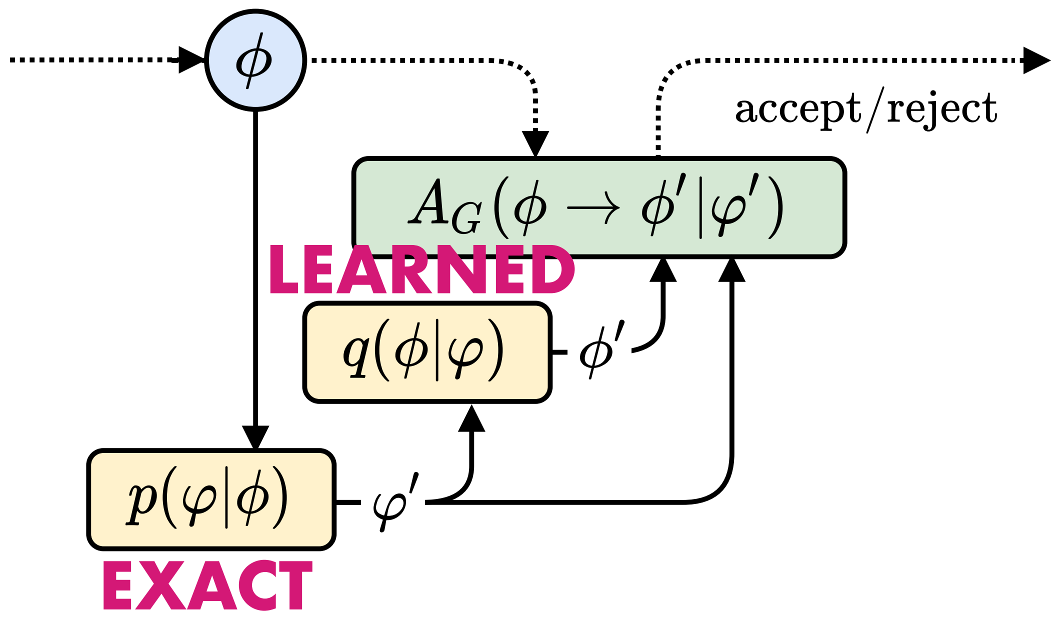
Proposed exact sampling schemes

Using a variety of learned densities $q(\ldots)$ — Best choice not yet clear!

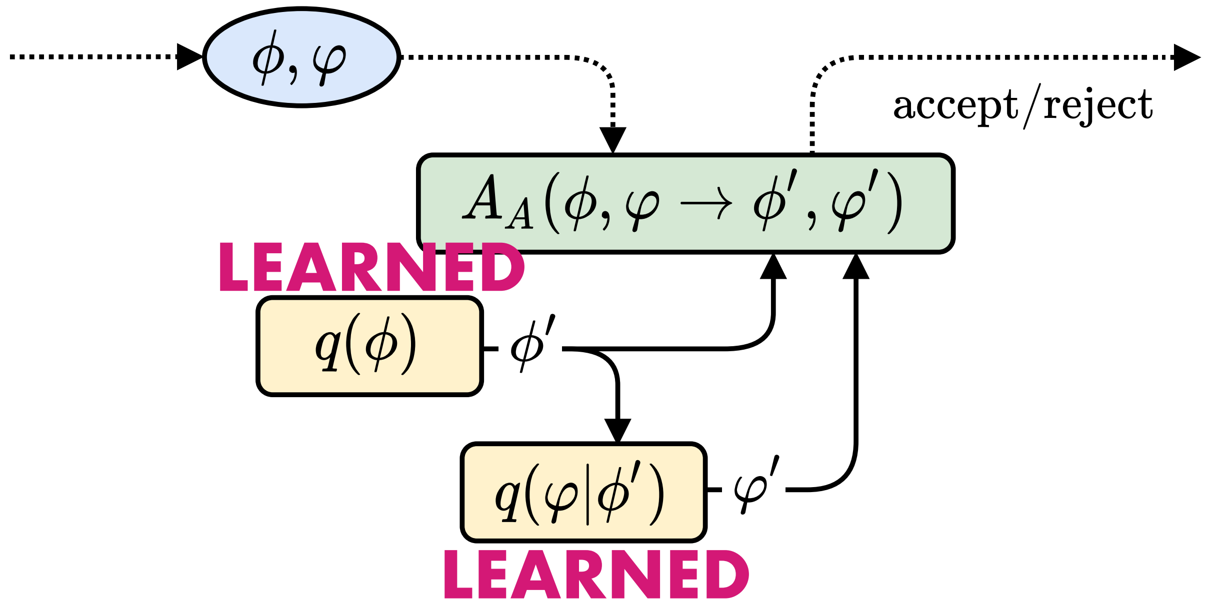
(1) ϕ-marginal



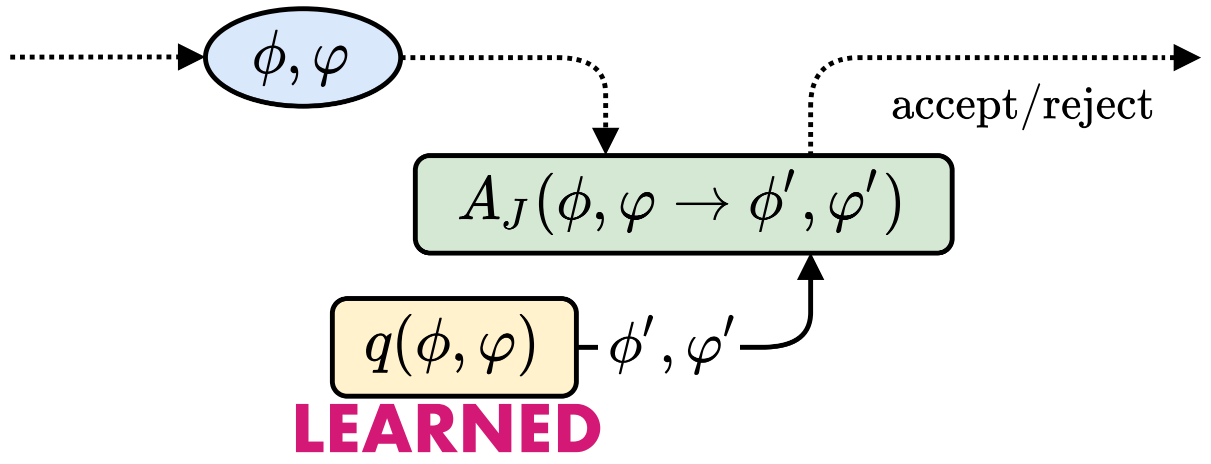
(2) Gibbs



(3) Autoregressive



(4) Joint



Key takeaways:

- **Exact** regardless of quality of modeled densities $q(\ldots)$
- Can define sampler over
  - … bosonic fields alone (ϕ) or
  - … bosonic + PF fields (ϕ, φ)
- For **Gibbs**, even a perfect model may have residual autocorrelations