Massachusetts
Institute of
Technology

Flow-based MCMC for SU(N) Lattice Gauge Theory

Gurtej Kanwar

Based on …

… flow-based sampling for lattice QFT:

[Albergo, GK, Shanahan **PRD100 (2019) 034515**]
[Albergo, Boyda, Hackett, GK, Cranmer, Racanière, Rezende, Shanahan **2101.08176**]

… 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**]

MITP Workshop: Machine Learning Techniques in Lattice QCD (May 24-28, 2021)

Motivations

Critical slowing down and **topological freezing** obstruct MCMC sampling near the continuum limit. See [K. Clark, Mon].

- **Problem:** Local/diffusive Markov chain updates
- Generative ML models can directly sample, may be used to propose global updates

ML models provide flexible “variational ansatz” distribution $q(U)$.

After optimizing the model “ansatz”:

$$q(U) = e^{-S_{\text{eff}}(U)} \approx p(U) = e^{-S(U) - \log Z}$$

$$\updownarrow$$

$$S_{\text{eff}}(U) \approx S(U) + \log Z$$

Efficiently sampled

Desired target

Generative modeling for LQFT

[L. Funcke, Tue]

Estimating thermodynamic observables:

- Flow-based models precisely estimate $\log Z$
- Asymptotic exactness $N \to \infty$

[S. Foreman, Thur]

Improved HMC updates:

- Flows/INNs describing modified HMC updates to (π, U)
- Topological freezing can be avoided
- Detailed balance for exactness

[D. Rezende, Mon] & this talk

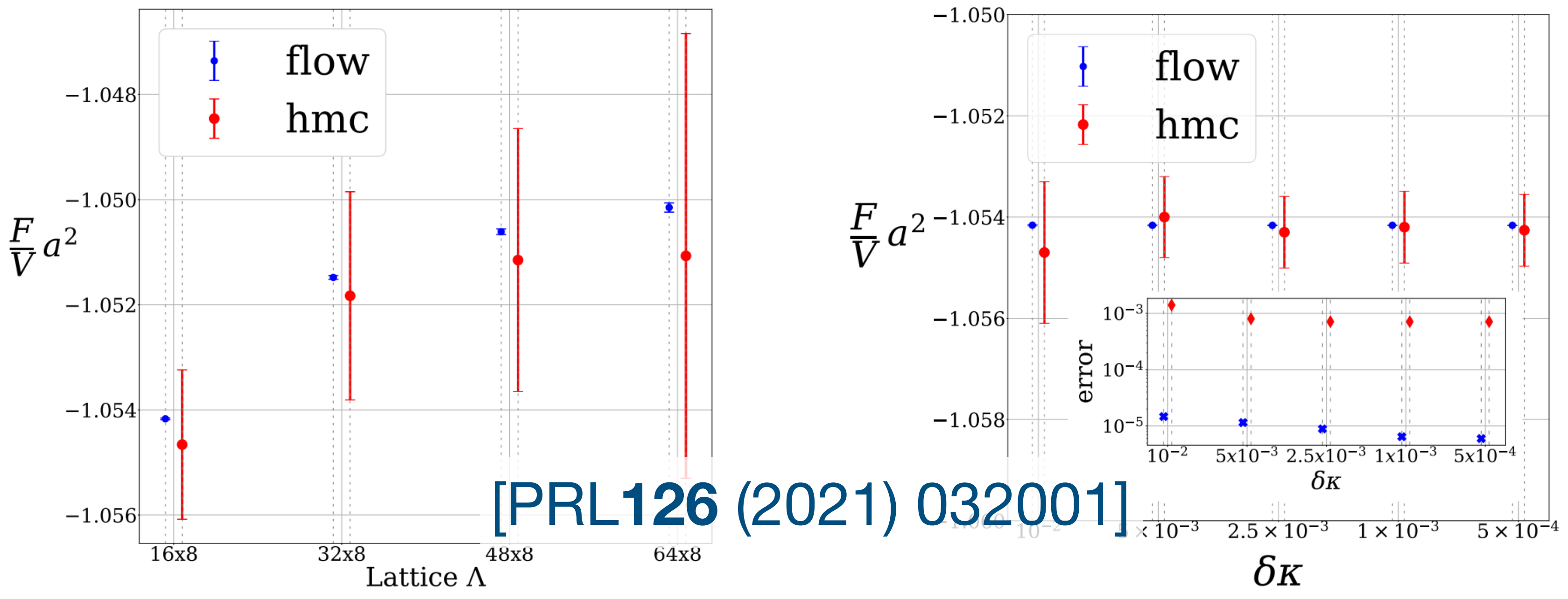
Flow-based MCMC:

- Flows directly propose new configs
- Metropolis step (satisfying balance) for exactness

Generative modeling for LQFT

[L. Funcke, Tue]

Estimating thermodynamic observables:



[PRL**126** (2021) 032001]

[S. Foreman, Thur]

Improved HMC updates:

- Flows/INNs describing modified HMC updates to (π, U)
- Topological freezing can be avoided
- Detailed balance for exactness

[D. Rezende, Mon] & this talk

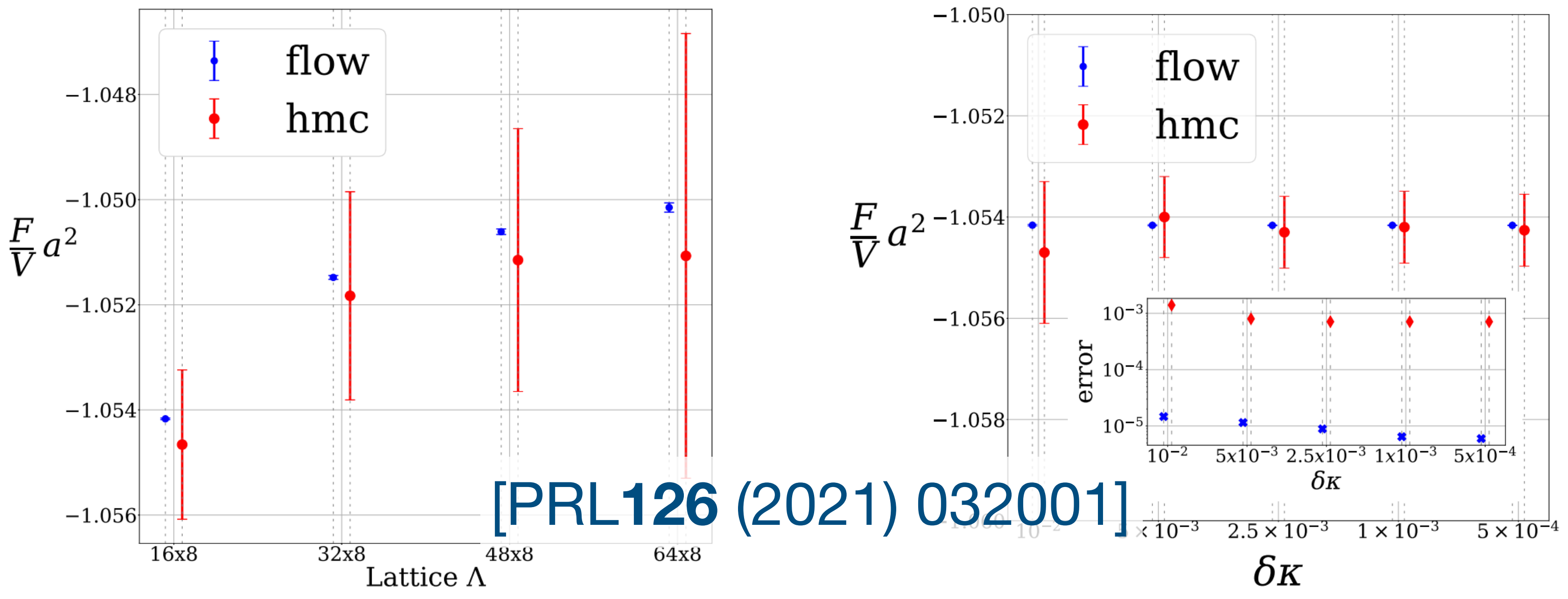
Flow-based MCMC:

- Flows directly propose new configs
- Metropolis step (satisfying balance) for exactness

Generative modeling for LQFT

[L. Funcke, Tue]

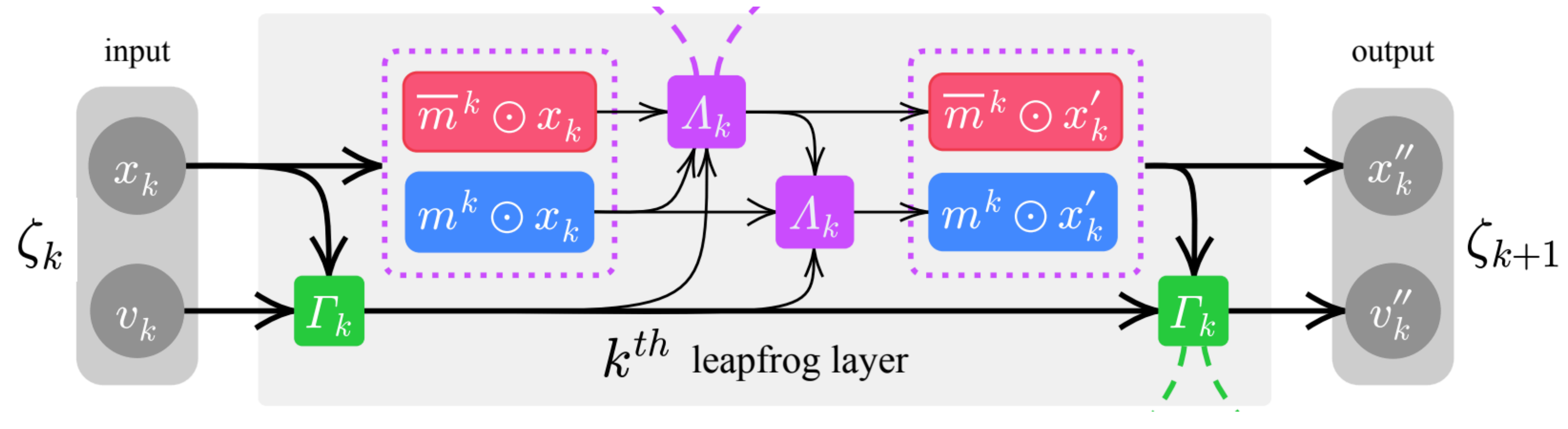
Estimating thermodynamic observables:



[PRL**126** (2021) 032001]

[S. Foreman, Thur]

Improved HMC updates:



[2105.03418]

[D. Rezende, Mon] & this talk

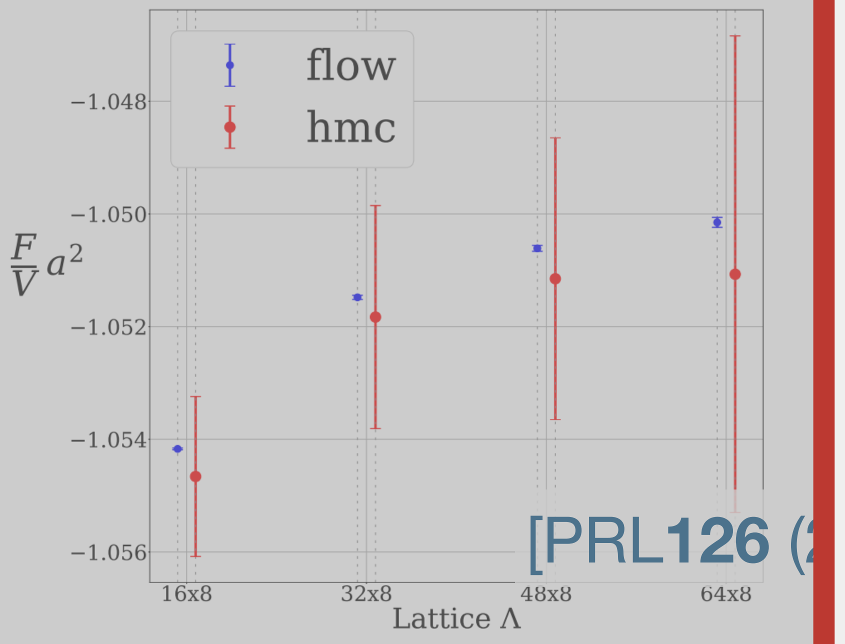
Flow-based MCMC:

- Flows directly propose new configs
- Metropolis step (satisfying balance) for exactness

Generative modeling for LQFT

[L. Funcke, Tue]

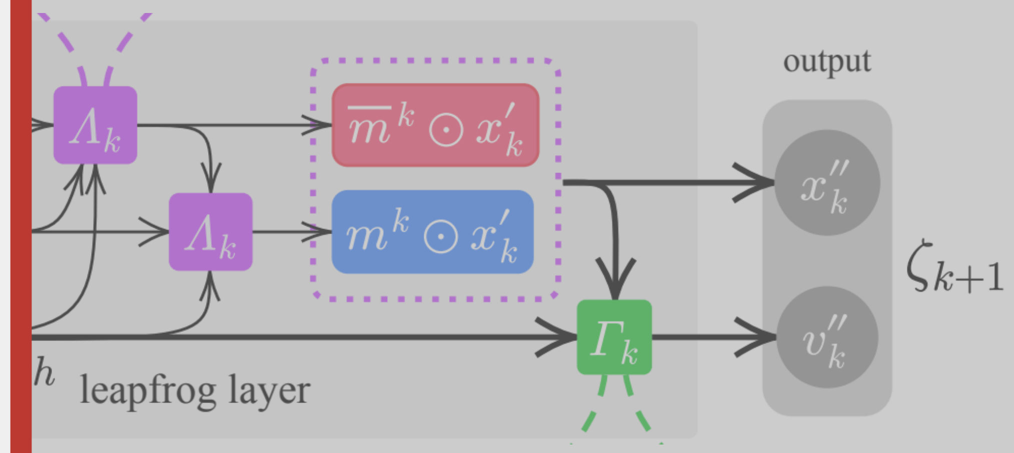
Estimating thermodynamic obser



[PRL**126** (

[S. Foreman, Thur]

Improved HMC updates:



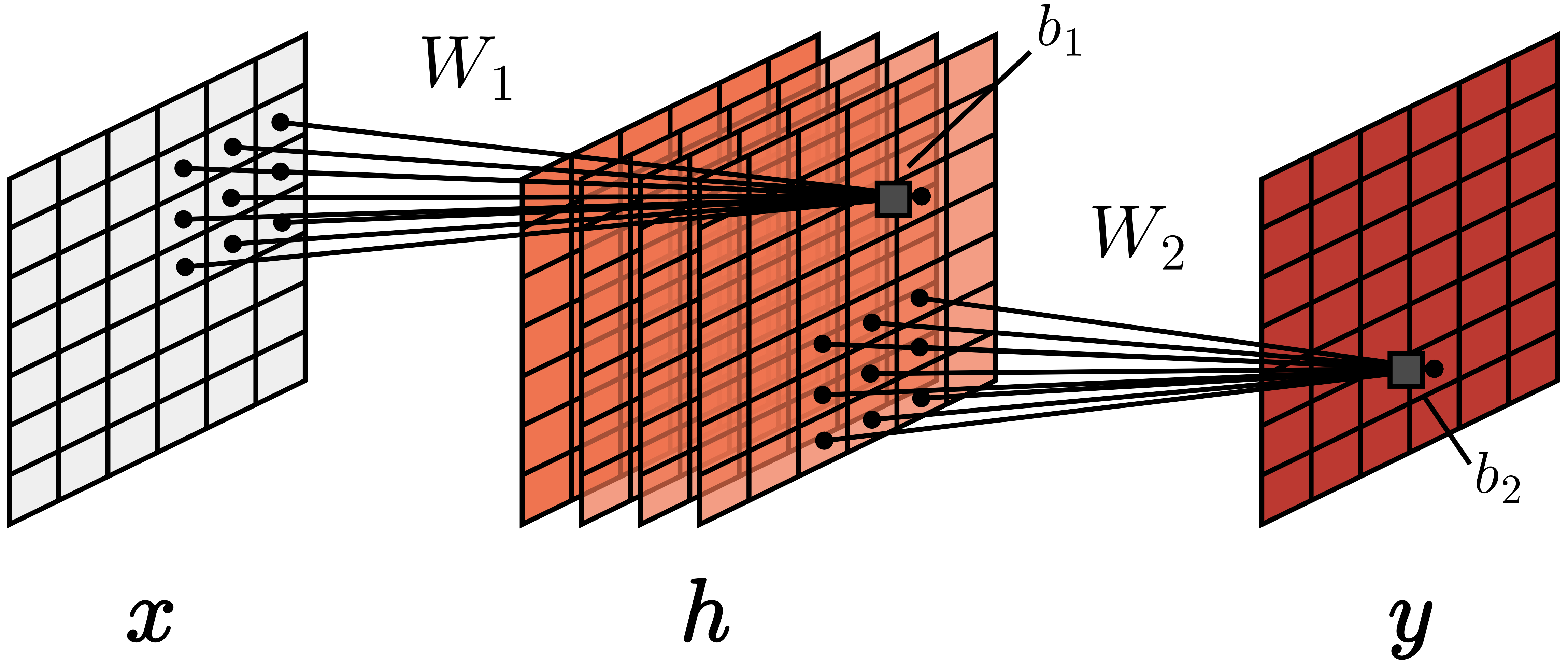
[2105.03418]

Common theme:

Black-box ML components wrapped inside exact schemes

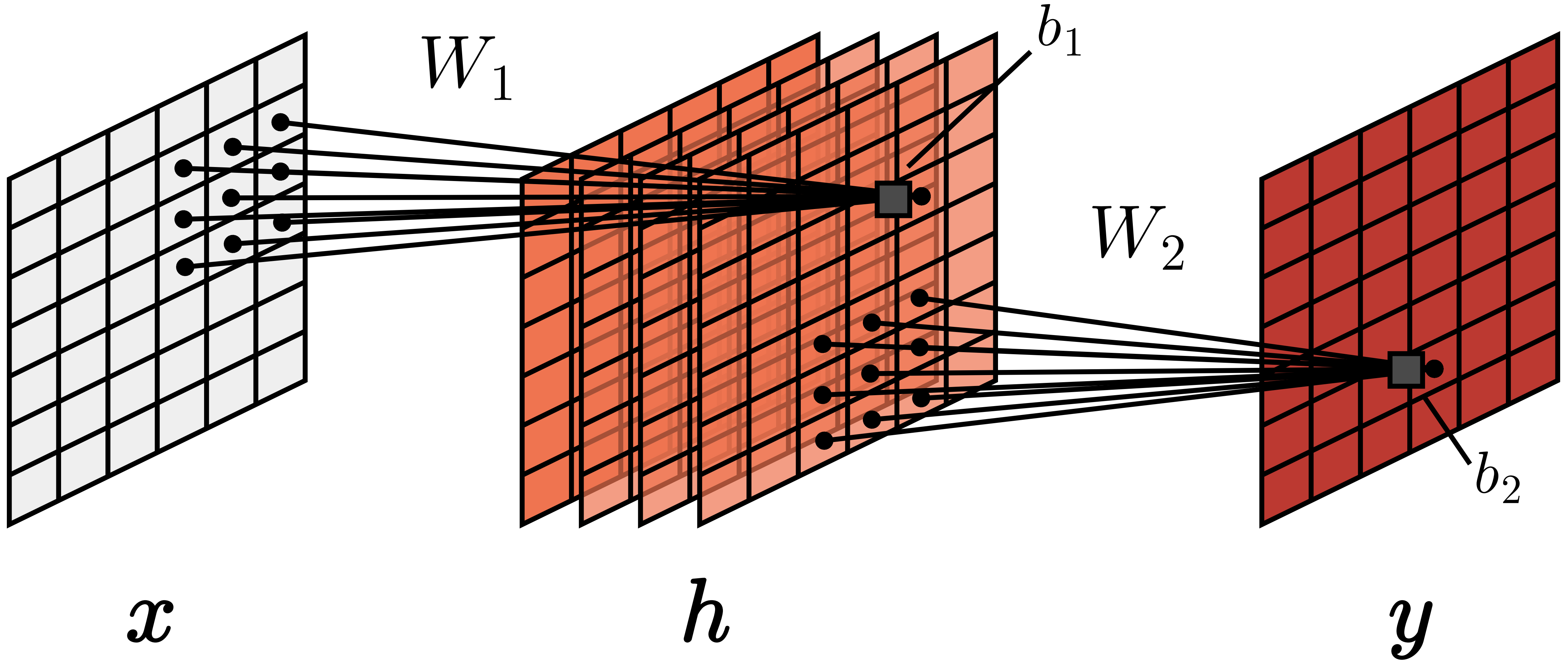
- Flows directly propose new configs
- Metropolis step (satisfying balance) for exactness

Flow-based sampling



Flow-based sampling

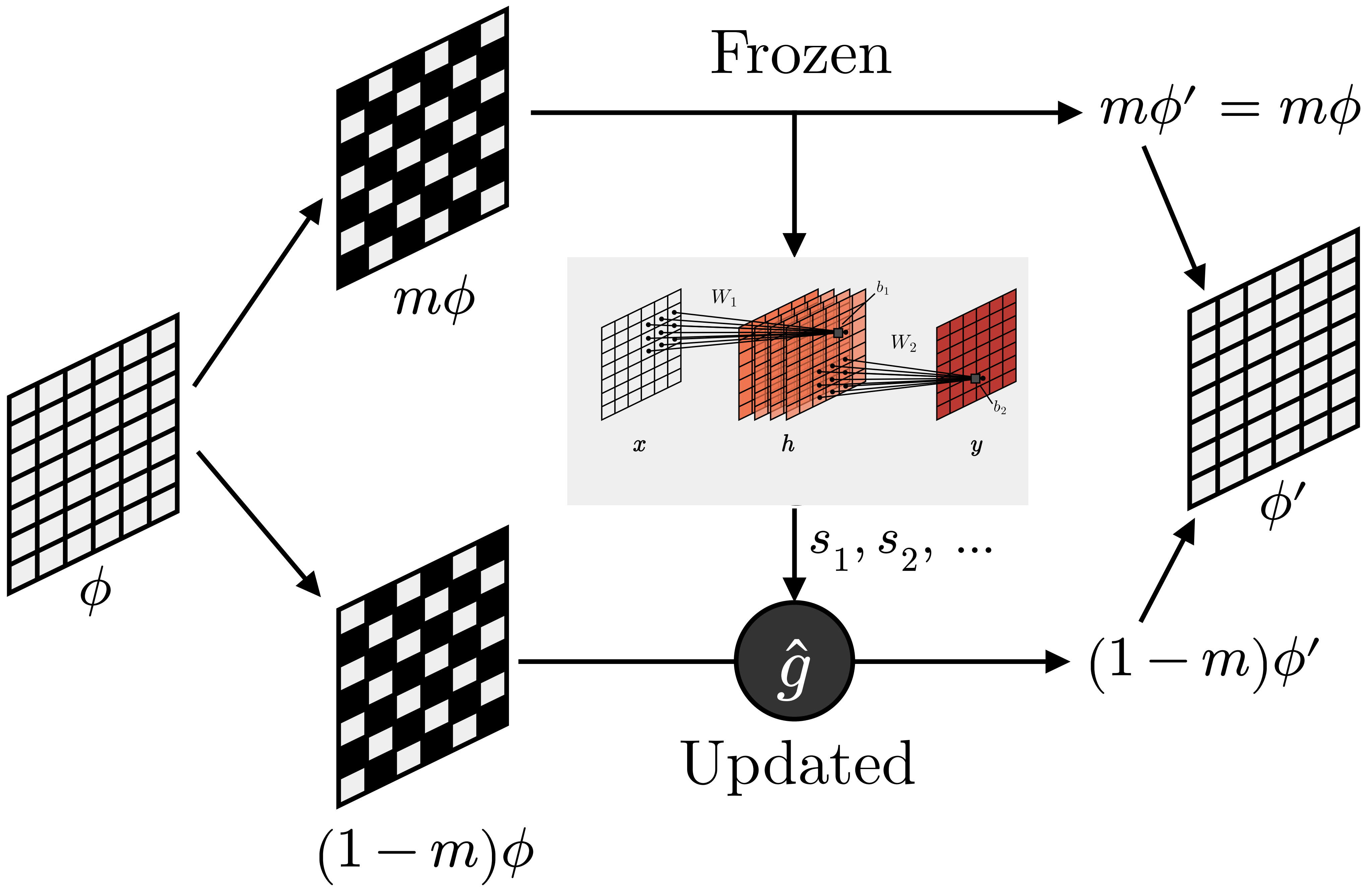
(Convolutional) neural networks: Black-box (local) function approximators



Flow-based sampling

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

Coupling layers: Invertible transformations, tractable Jacobian

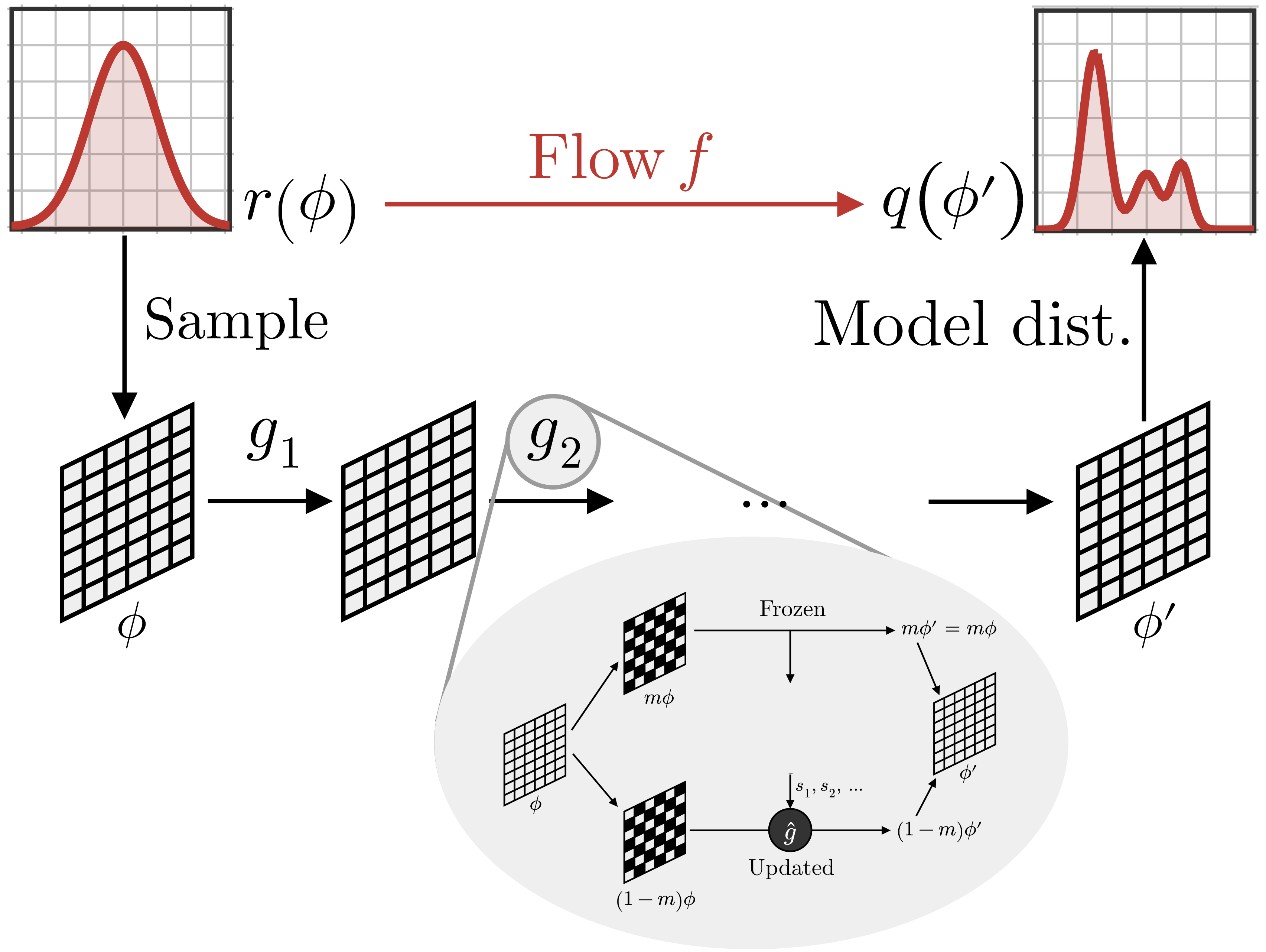


Flow-based sampling

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

Coupling layers: Invertible transformations, tractable Jacobian

Flow-based model: Transform prior density to **computable** and **sample-able** output model density



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

Flow-based sampling

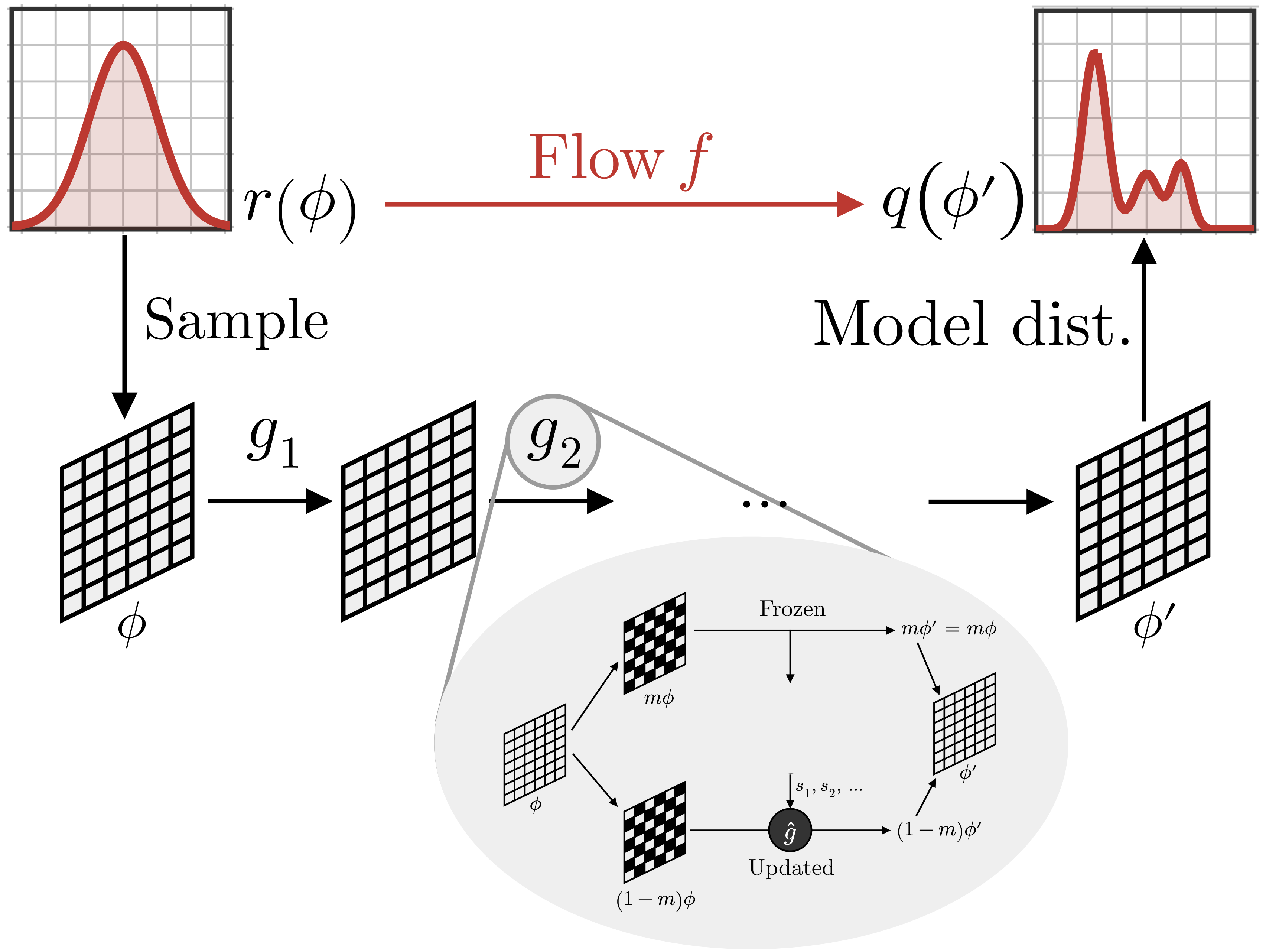
(Convolutional) neural networks: Black-box (local) function approximators

Coupling layers: Invertible transformations, tractable Jacobian

Flow-based model: Transform prior density to **computable** and **sample-able** output model 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

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

Coupling layers: Invertible transformations, tractable Jacobian

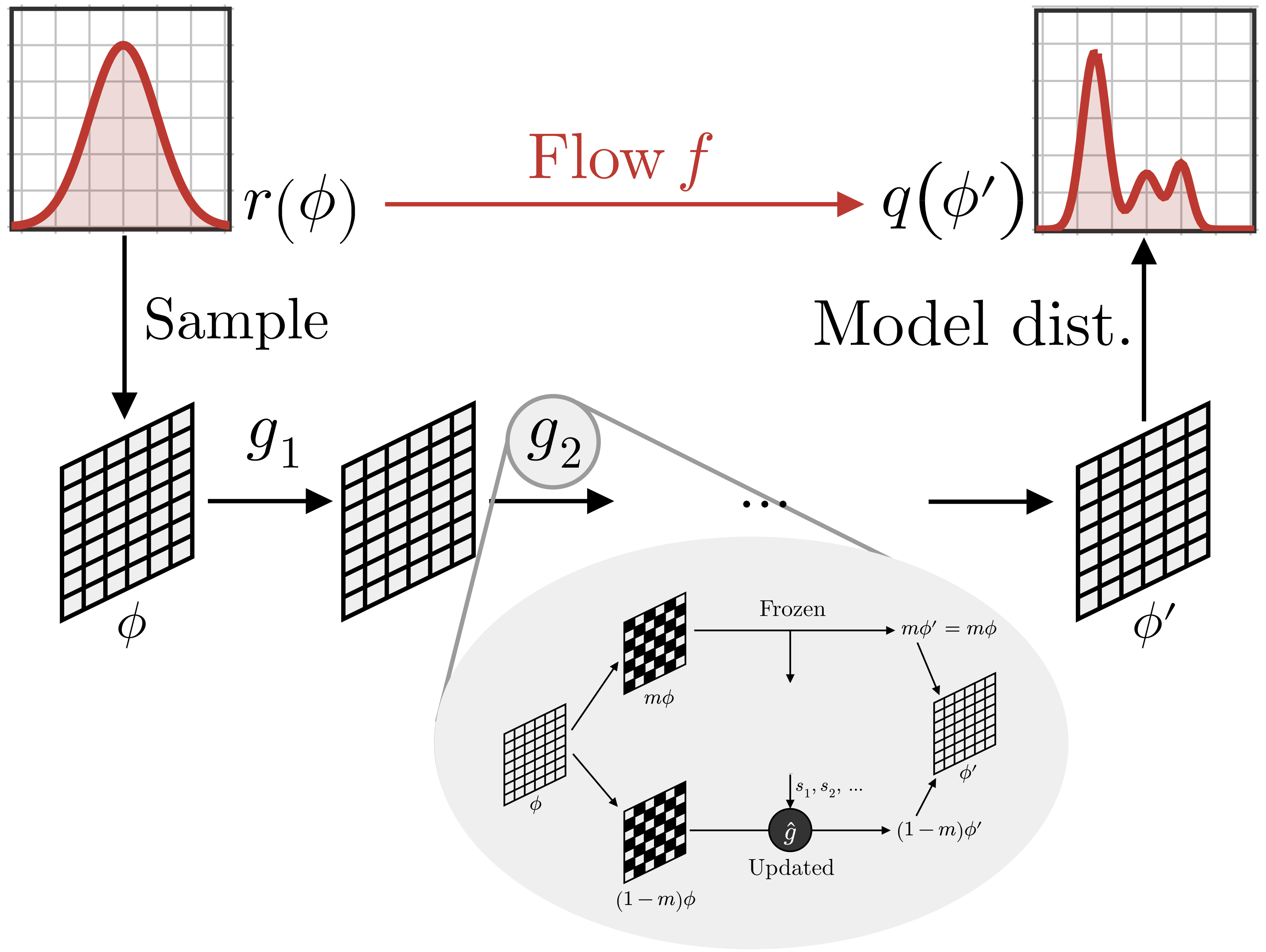
Flow-based model: Transform prior density to **computable** and **sample-able** output model density

Training:

- Measure KL divergence
- Apply gradient-based opt

Exactness:

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



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

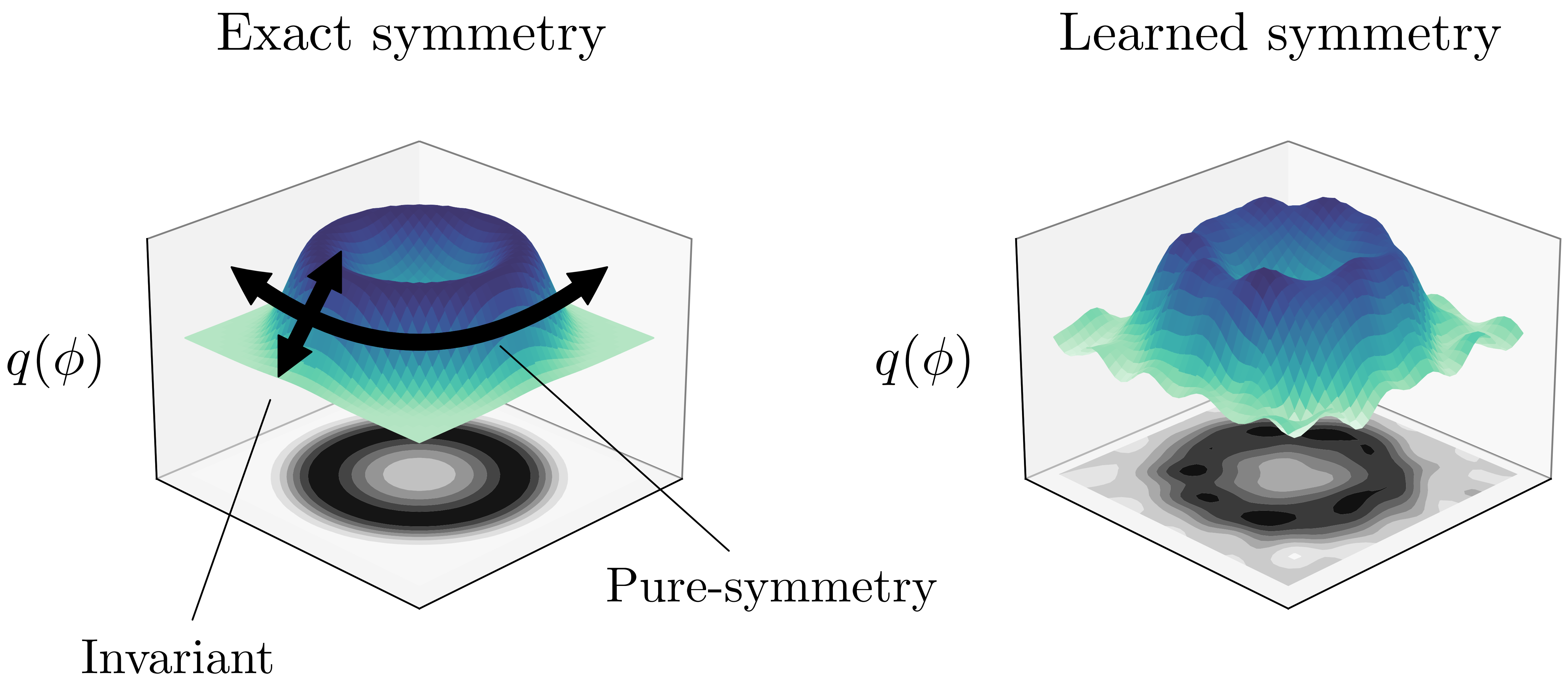
Symmetries in flows

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

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

Symmetries…

- Reduce data complexity of training
- Reduce model parameter count
- See [D. Luo, Wed] and [A. Tomiya, Fri]



Towards lattice QCD

This talk: SU(N) gauge symmetry of Lattice QCD [± small detour into U(1)]

- See [D. Rezende, Mon] and upcoming paper with

  M. S. Albergo, S. Racanière, D. J. Rezende, J. M. Urban, D. Boyda, K. Cranmer, D. C. Hackett, P. E. Shanahan

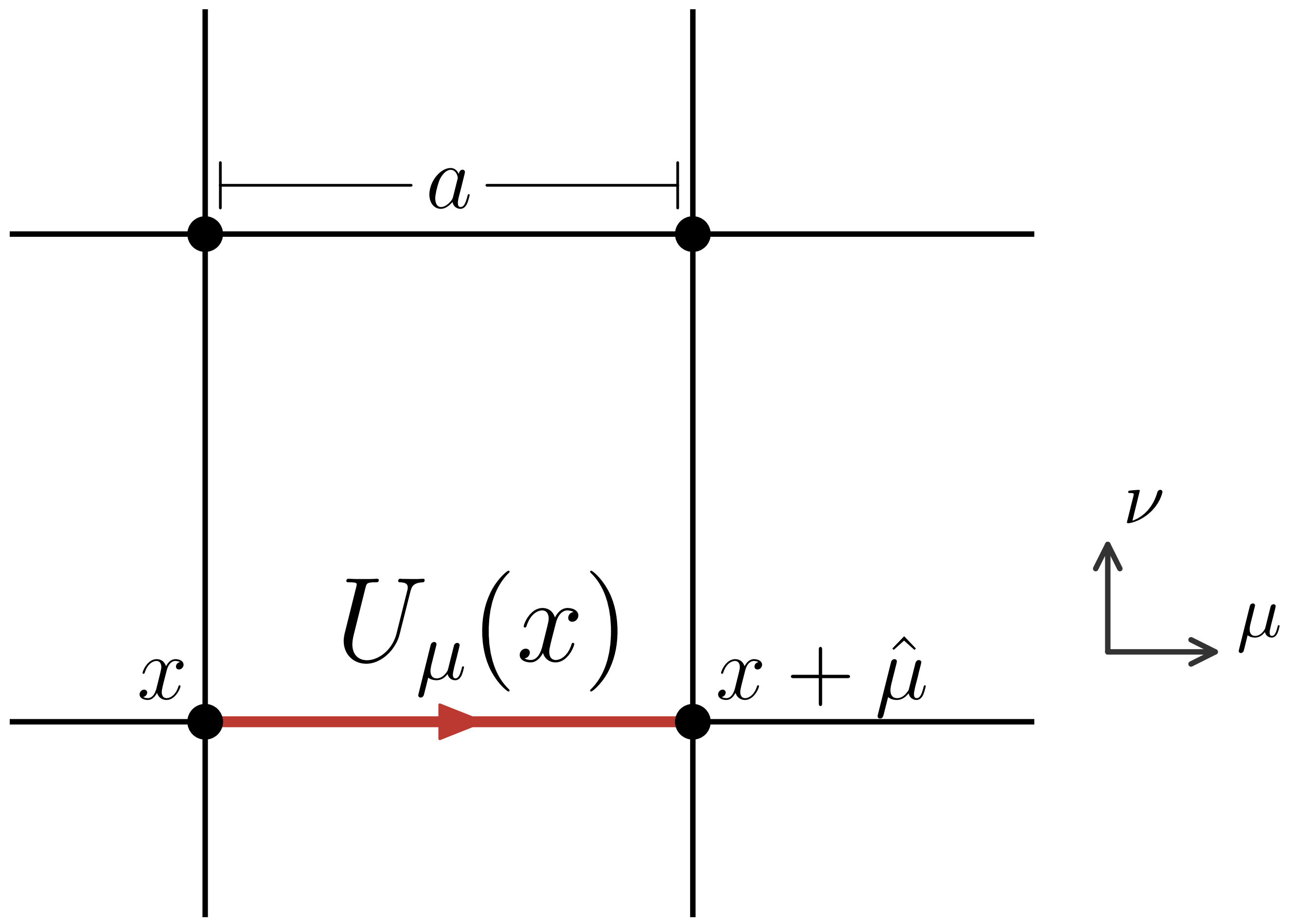
  for more on fermions

Wilson gauge action - prototypical, gauge-invariant lattice action

For U(1): $N = 1$ and no Tr

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

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



Towards lattice QCD

This talk: SU(N) gauge symmetry of Lattice QCD [± small detour into U(1)]

- See [D. Rezende, Mon] and upcoming paper with

  M. S. Albergo, S. Rac Hackett, P. E. Shanahan

  for more on ferm

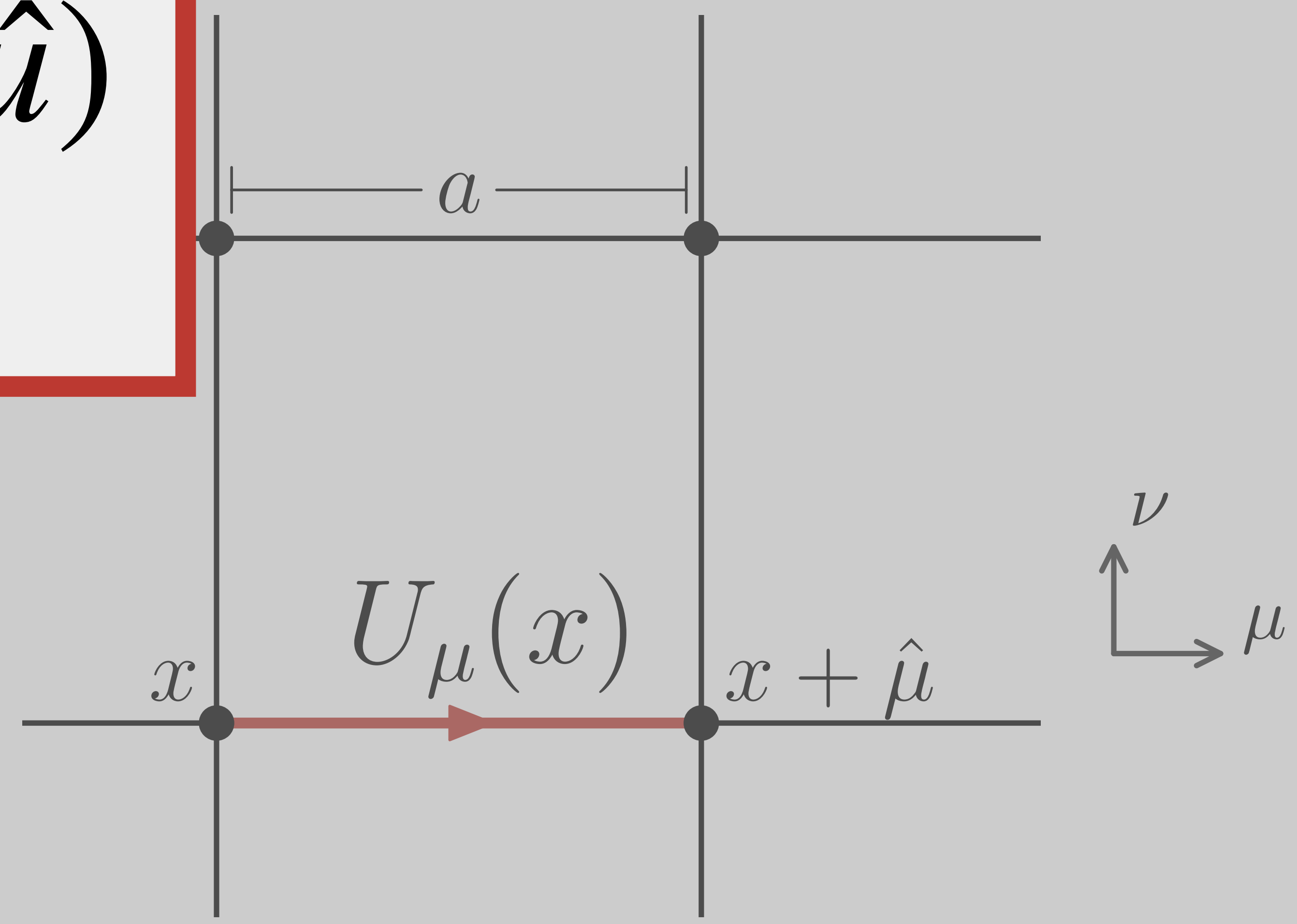
Wilson gauge act tion

Invariant under gauge transforms:

$$U_\mu(x) \;\rightarrow\; \Omega(x)\, U_\mu(x)\, \Omega^\dagger(x+\hat{\mu})$$

$$S(U) = -\frac{\beta}{N} \sum_{x} \sum_{\mu<\nu}$$

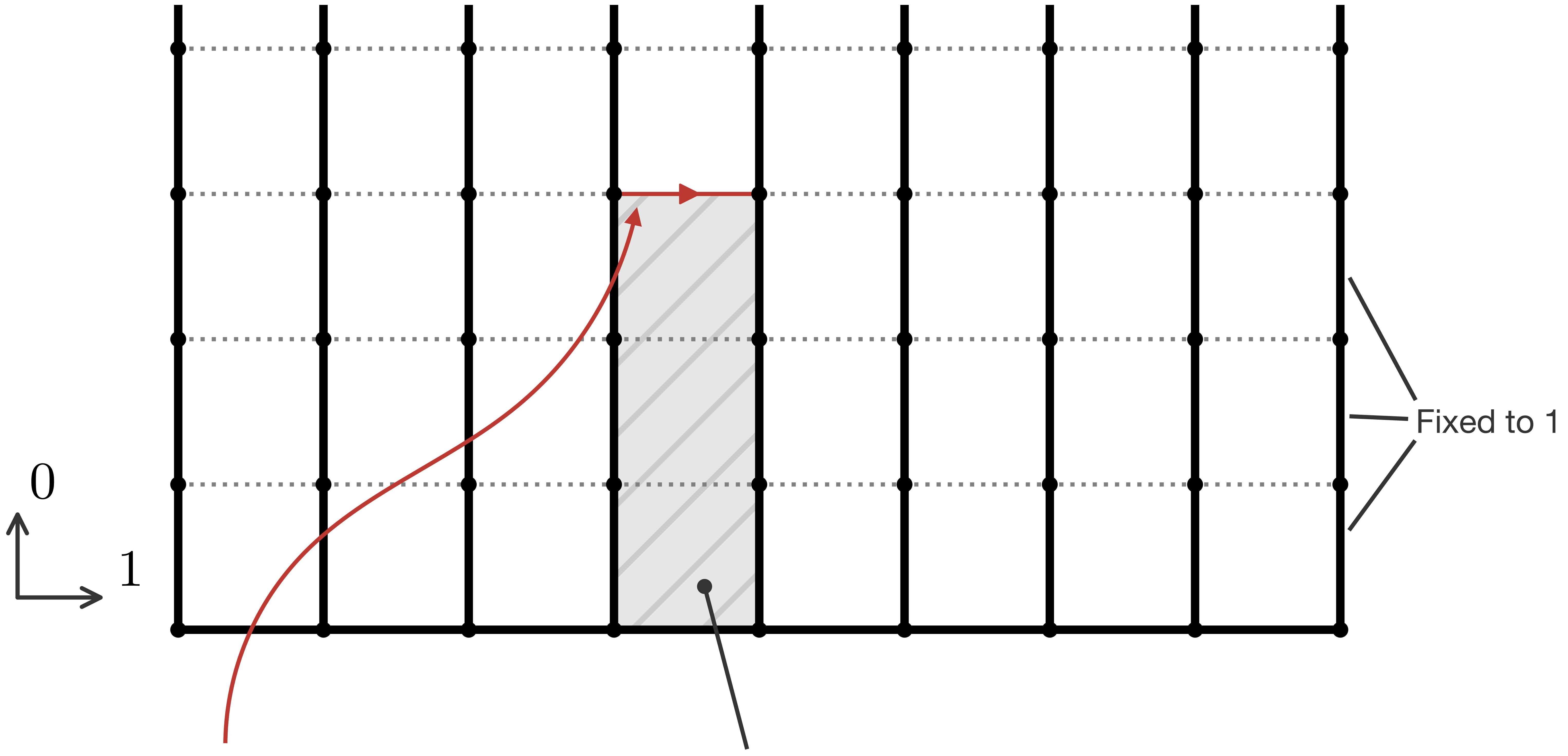
$$P_{\mu\nu}(x) = U_\mu(x) U_\nu(x+\hat{\mu}) U_\mu^\dagger(x+\hat{\nu}) U_\nu^\dagger(x)$$



Gauge fixing?

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

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



Link physically encodes **Wilson loop** around shaded region

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**

$$\left.\begin{array}{ll}\textbf{Landau gauge:} & U_\mu^{\text{fix}}(x) = \text{argmin}_{U^\Omega} \sum_x \sum_{\mu=1}^{N_d} \text{ReTr}[U_\mu^\Omega(x)] \\ \\ \textbf{Coulomb gauge:} & U_\mu^{\text{fix}}(x) = \text{argmin}_{U^\Omega} \sum_x \sum_{\mu=1}^{N_d-1} \text{ReTr}[U_\mu^\Omega(x)] \end{array}\right\}$$

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

Gauge symmetries in un-fixed 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:

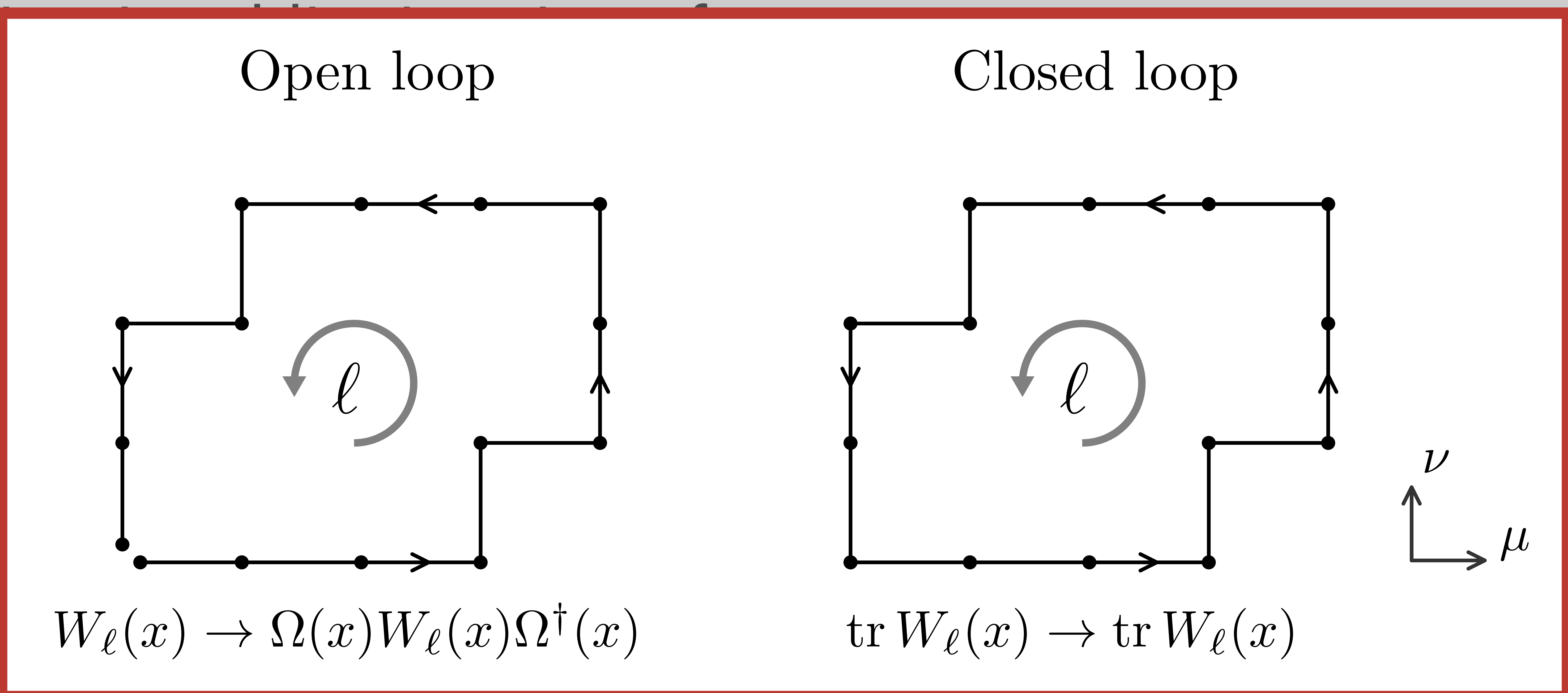
Coupling layers acting on (untraced) Wilson loops.

Loop transformation easier to satisfy.

Gauge symmetries in un-fixed flows

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

Carefully cons



$$W_\ell(x) \to \Omega(x) W_\ell(x) \Omega^\dagger(x)$$

$$\mathrm{tr}\, W_\ell(x) \to \mathrm{tr}\, W_\ell(x)$$

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** $W_\ell(x)$.

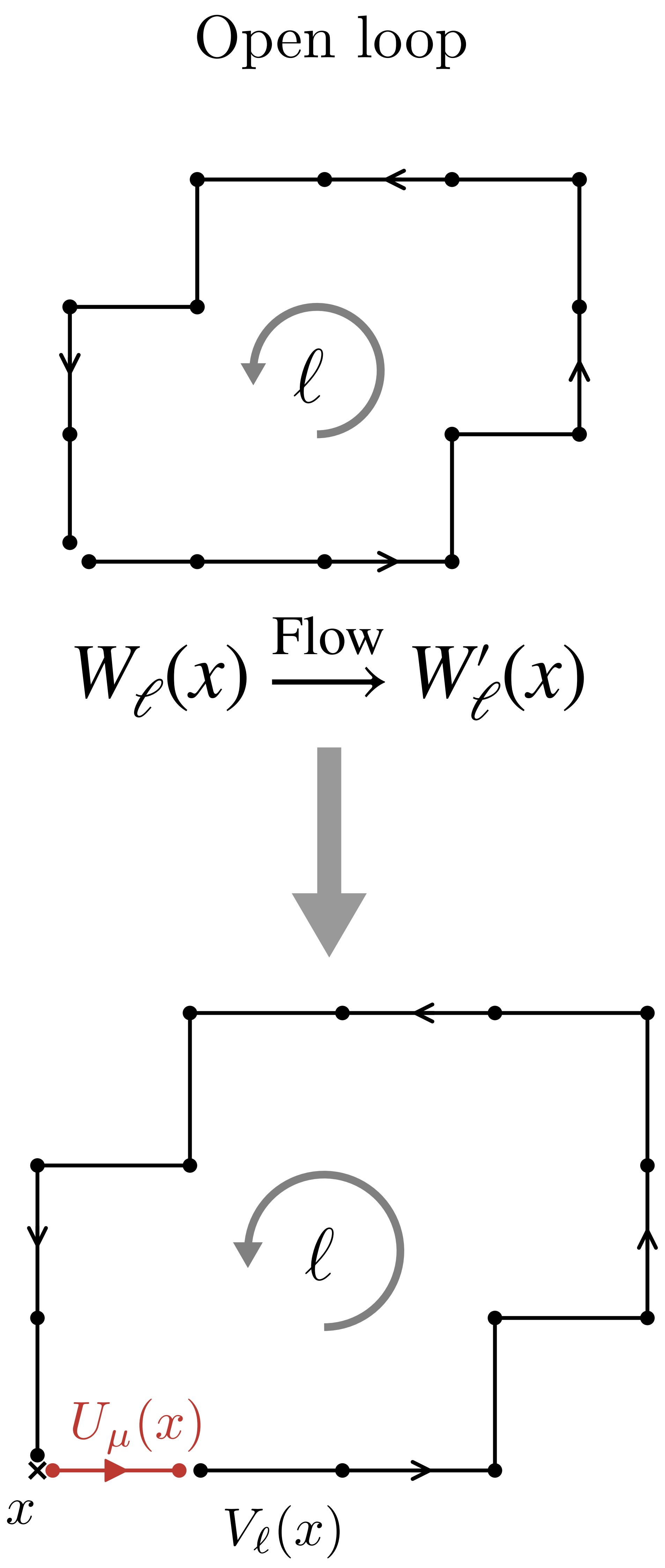
Inner coupling layer [function of $W_\ell(x)$]

- "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.



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

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

[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.

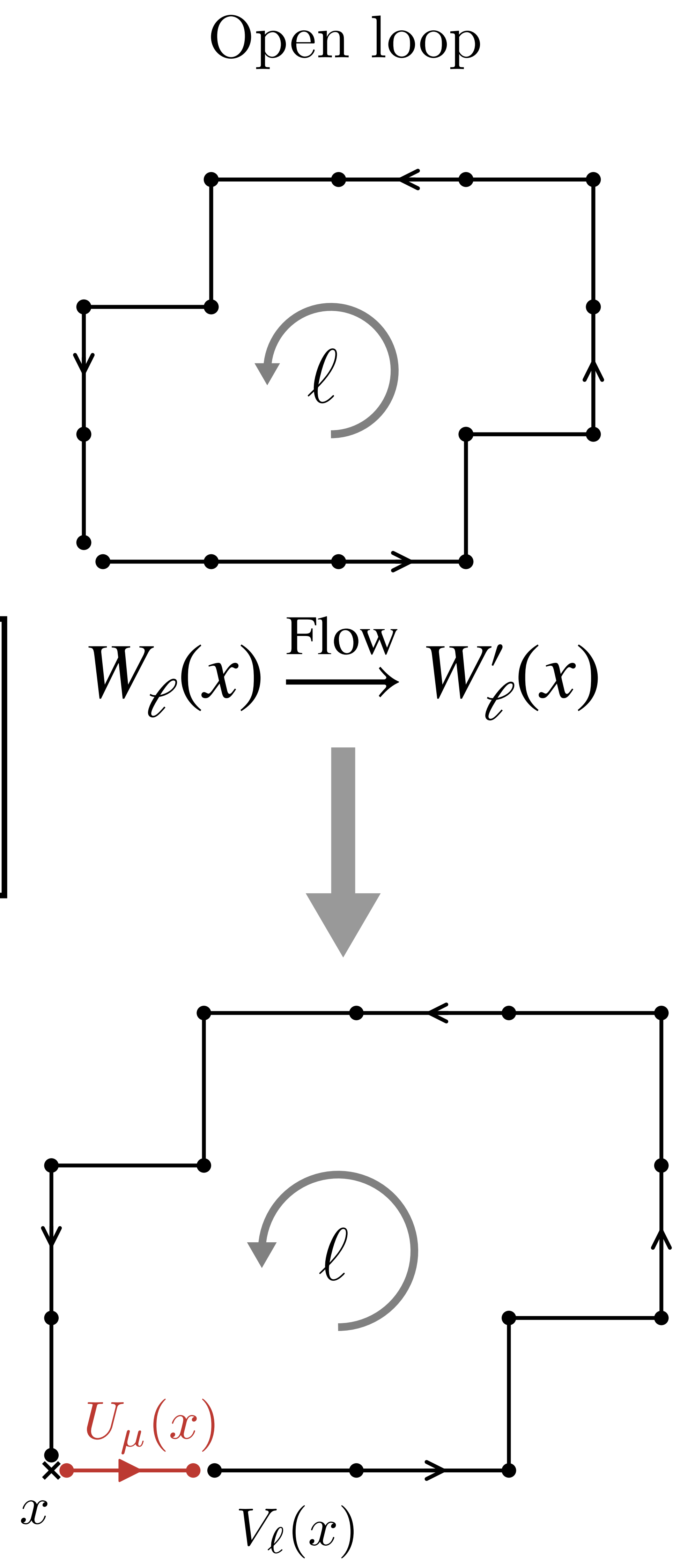
Gauge invariant!

* 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.



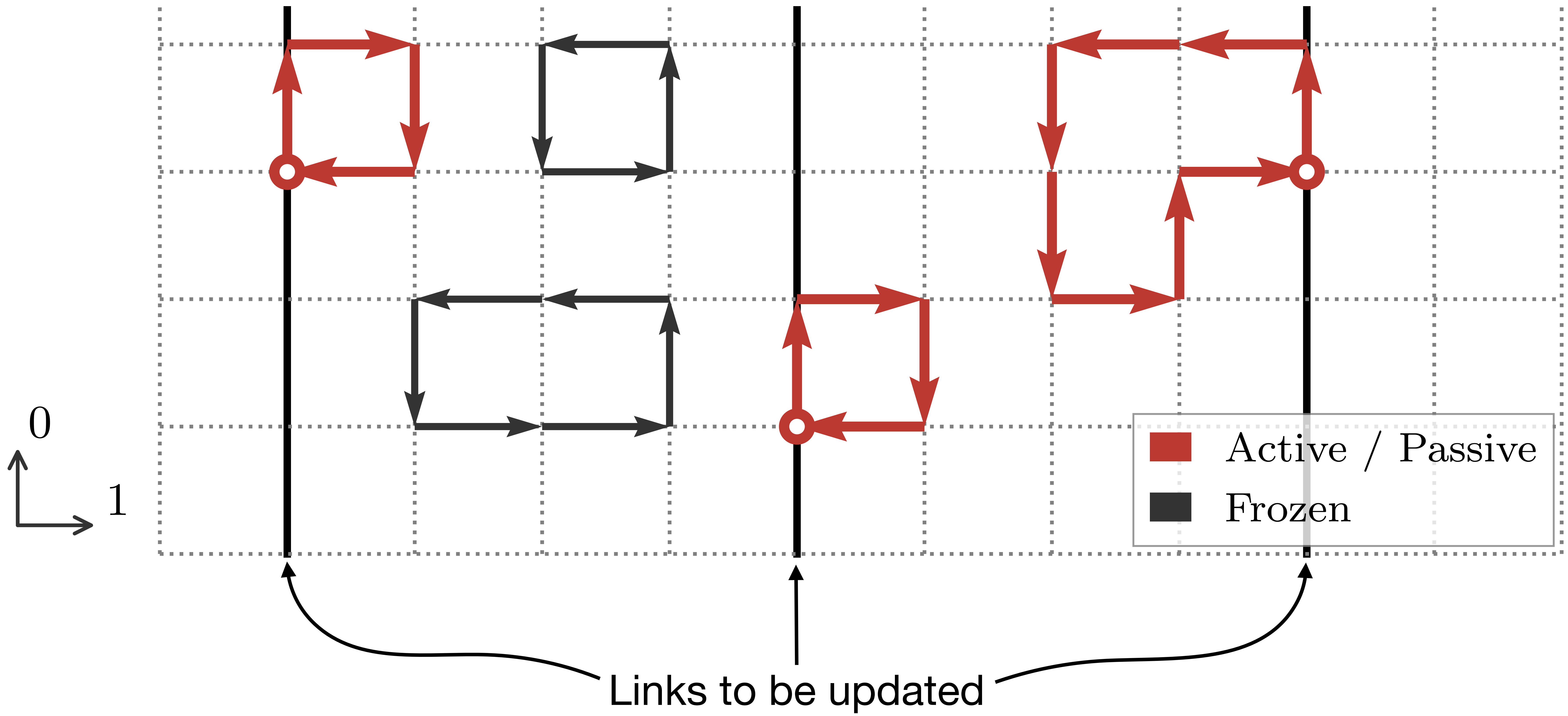
$$W_\ell(x) \xrightarrow{\text{Flow}} W'_\ell(x)$$

$$U'_\mu(x) = W'_\ell(x)\, V_\ell^\dagger(x)$$

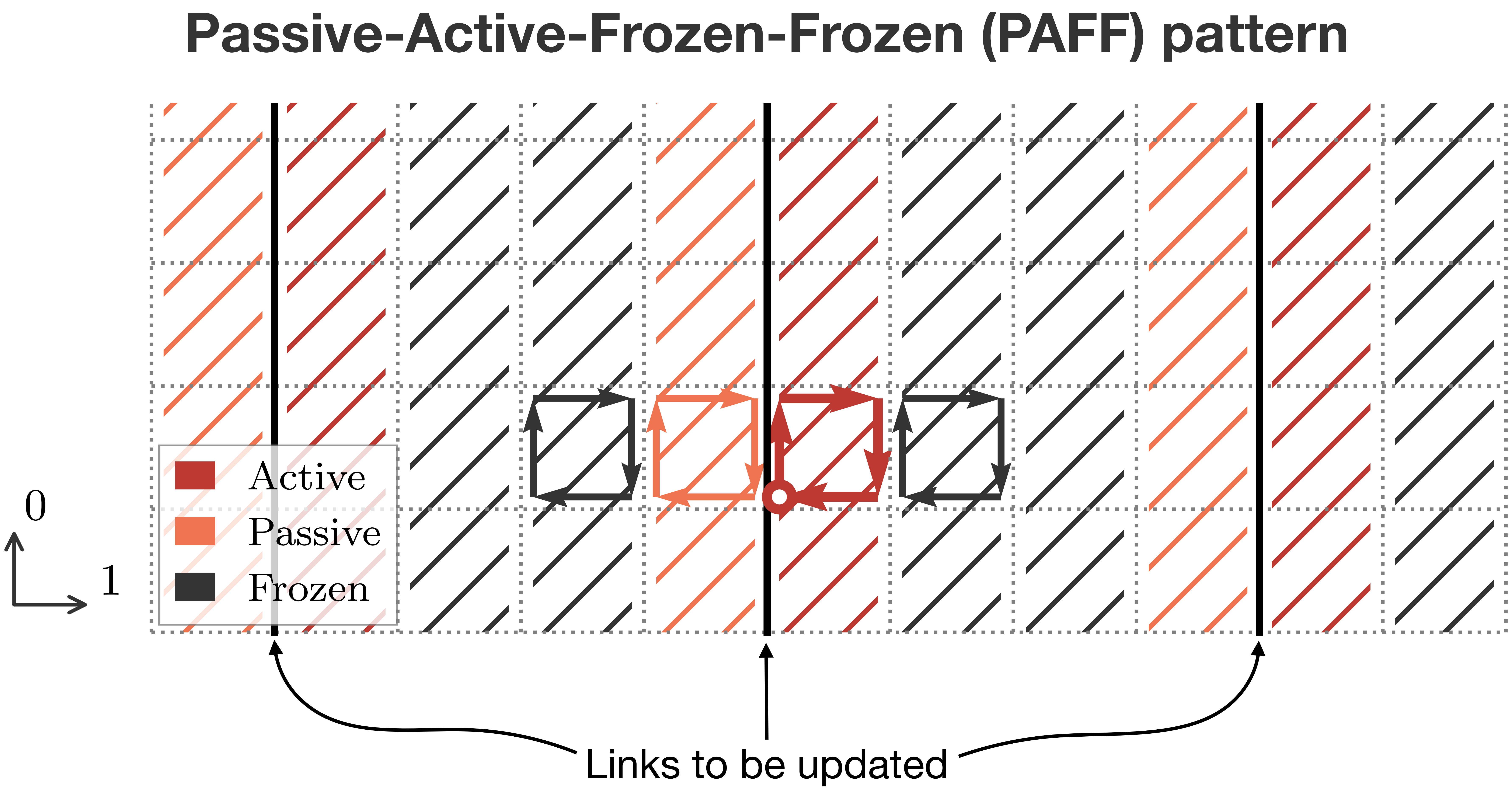
[GK, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan **PRL125 (2020) 121601**]
[Boyda, GK, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan **PRD103 (2021) 074504**]

Active, passive, and frozen loops

Examples of active/passive/frozen loops



Active, passive, and frozen loops


Passive-Active-Frozen-Frozen (PAFF) pattern

Links to be updated

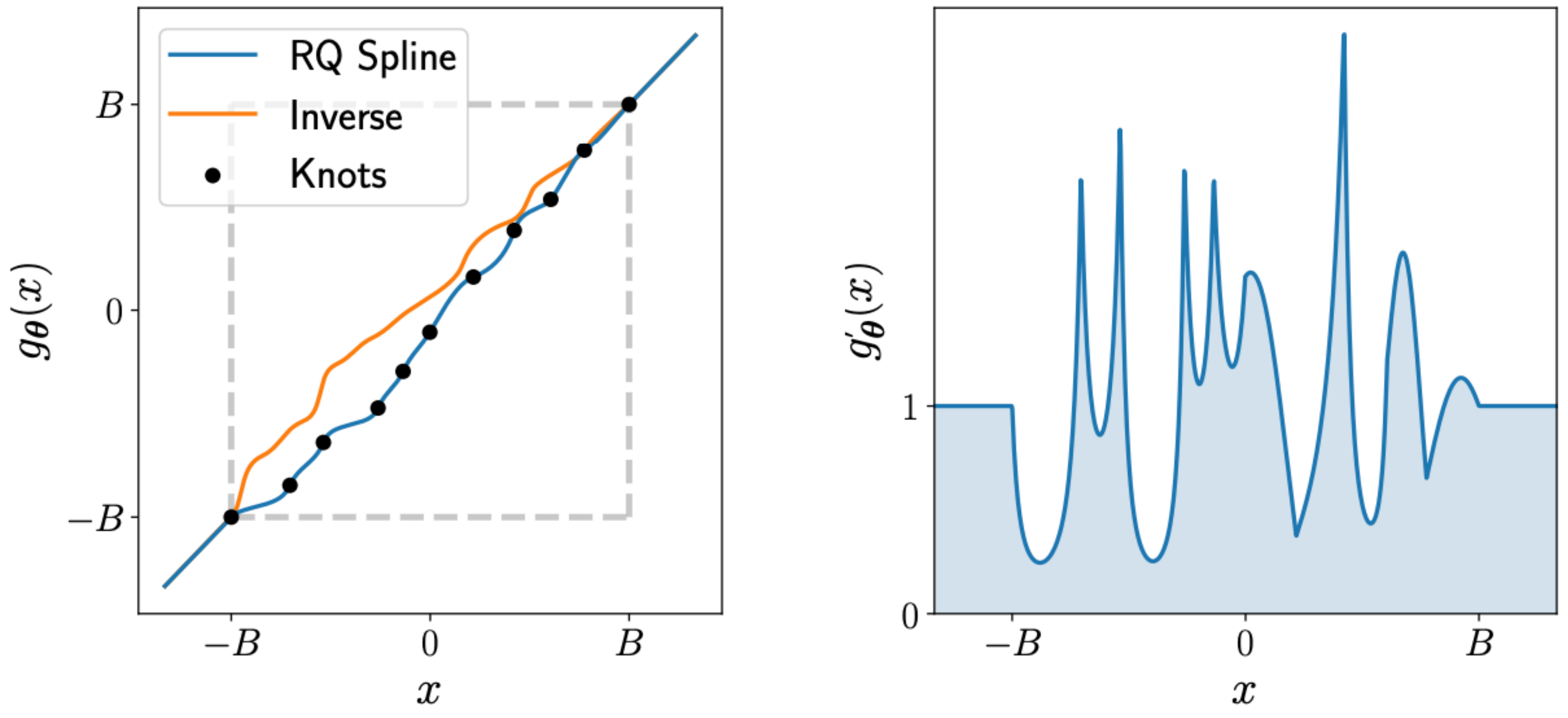
U(1) kernels

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 $x' \to \theta' \in [-\pi, \pi]$

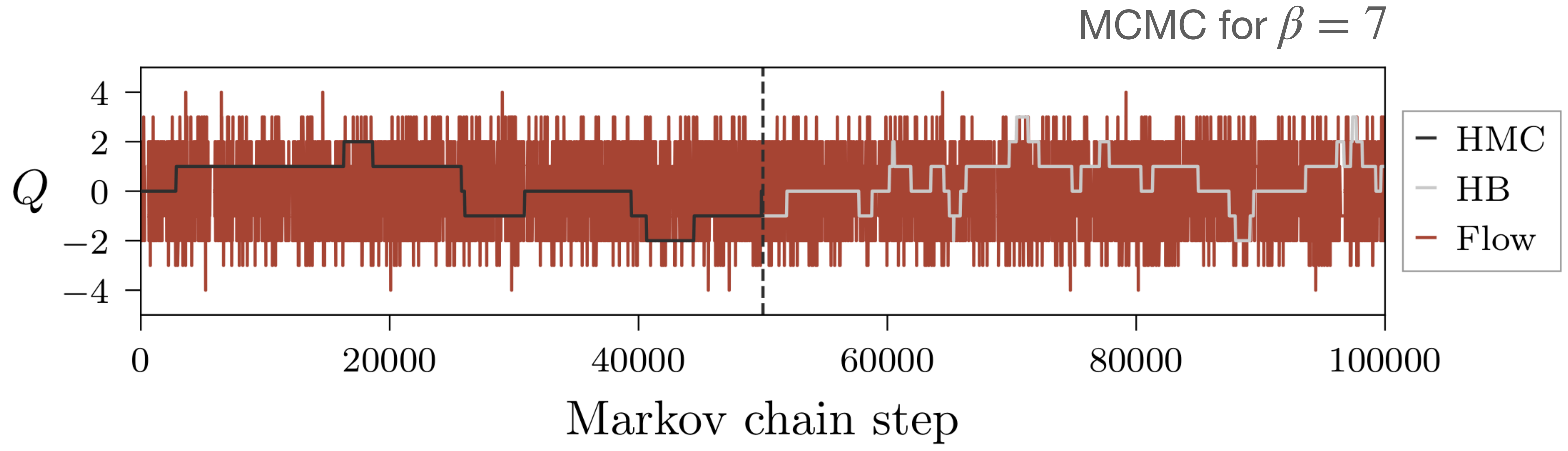
Circular invertible splines:

- Spline “knots” trainable fns
- Identify endpoints π and $-\pi$
- Number of knots ↔ expressivity

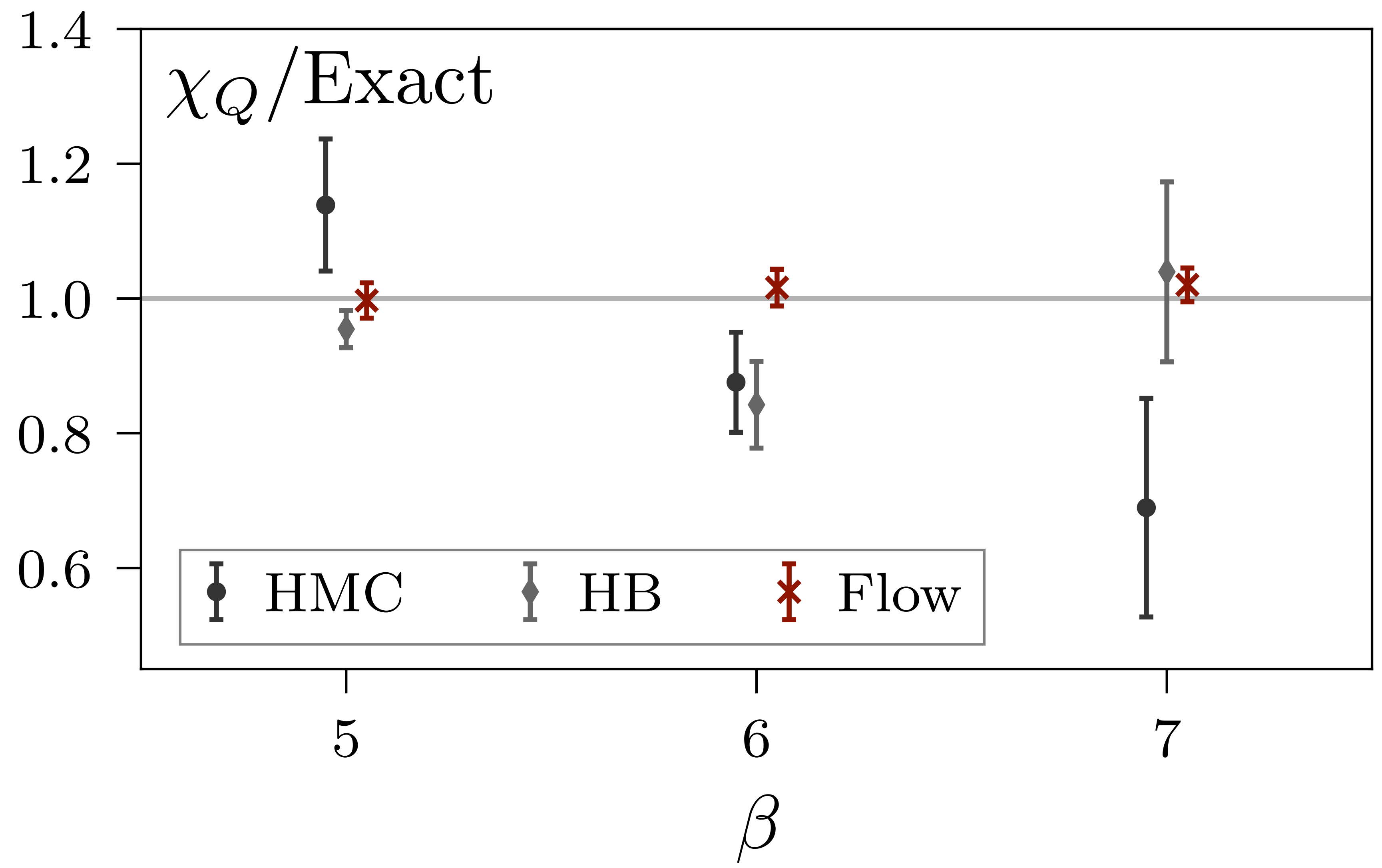
Learning U(1) gauge theory

There is exact lattice topology in 2D.

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



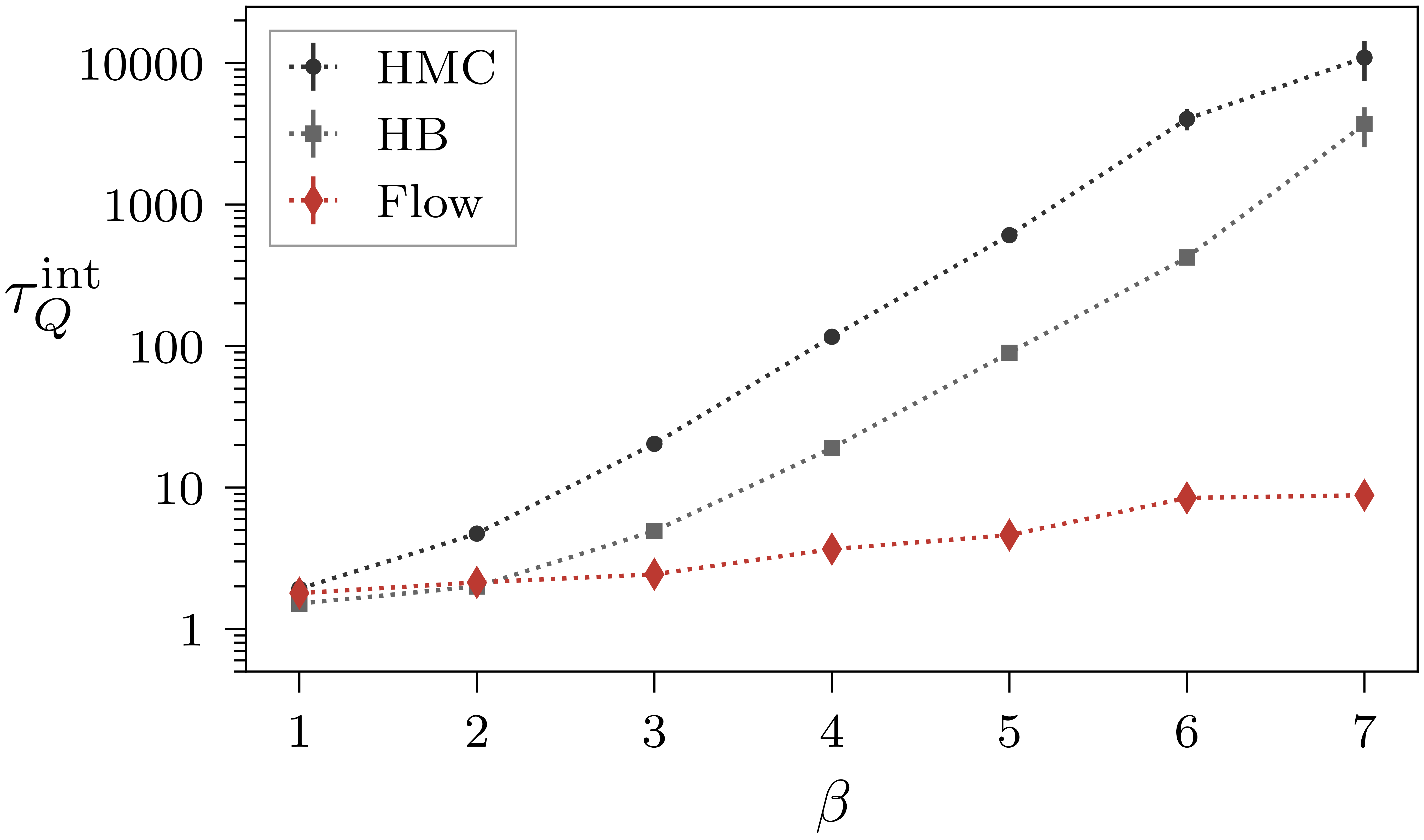
- Compared **flow, analytical, HMC**, and **heat bath** on 16×16 lattices for $\beta = \{1,\ldots,7\}$
- Topo freezing in HMC and heat bath
- Gauge-equiv flow-based model at each β
- Flow-based MCMC observables agree



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

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

Topological freezing mitigated



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**.

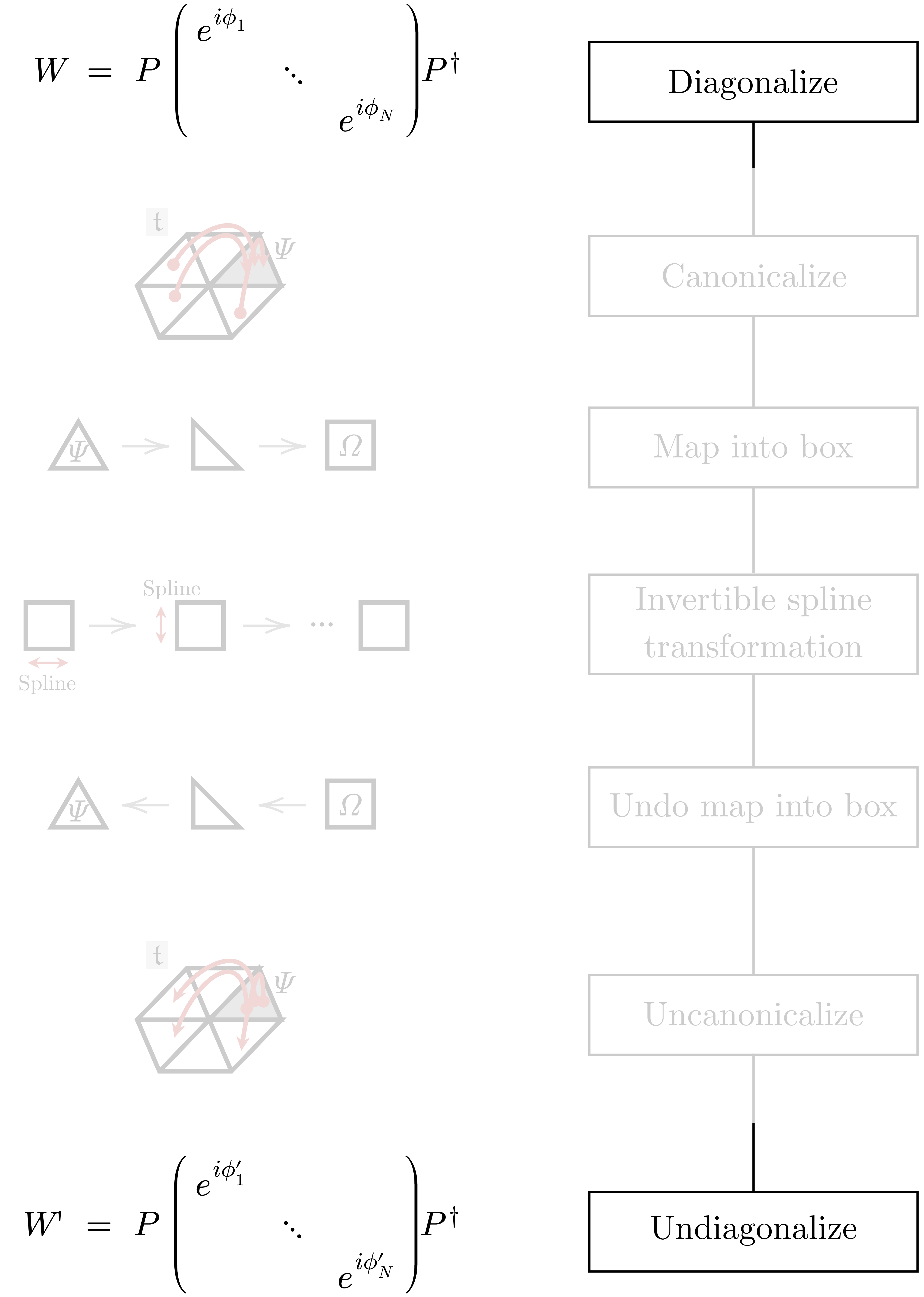
$$h(\Omega W \Omega^\dagger) = \Omega h(W) \Omega^\dagger$$

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".



$$W = P \begin{pmatrix} e^{i\phi_1} & & \\ & \ddots & \\ & & e^{i\phi_N} \end{pmatrix} P^\dagger$$

Diagonalize

Canonicalize

Map into box

Invertible spline transformation

Undo map into box

Uncanonicalize

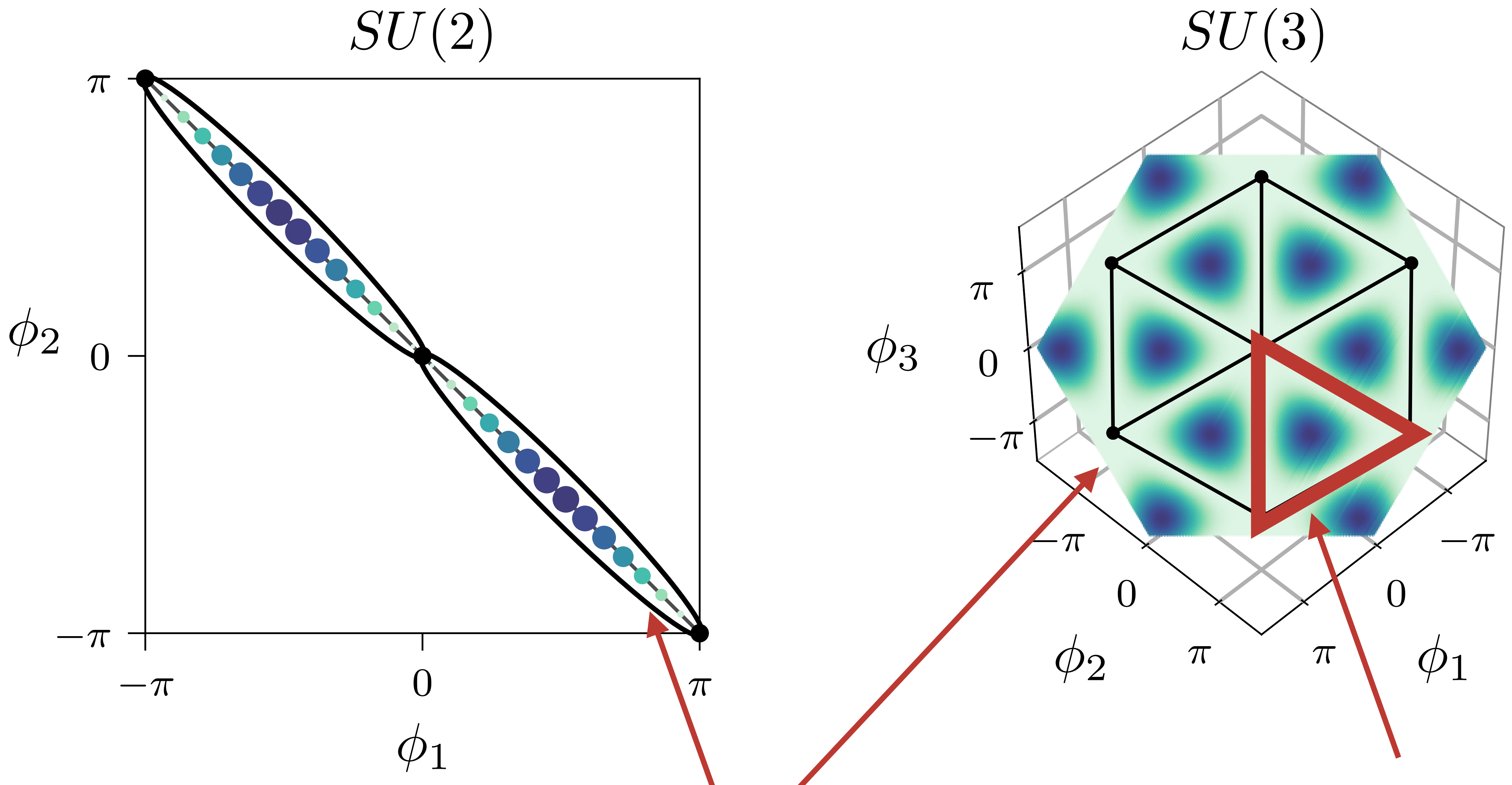
Undiagonalize

$$W' = P \begin{pmatrix} e^{i\phi'_1} & & \\ & \ddots & \\ & & e^{i\phi'_N} \end{pmatrix} P^\dagger$$

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

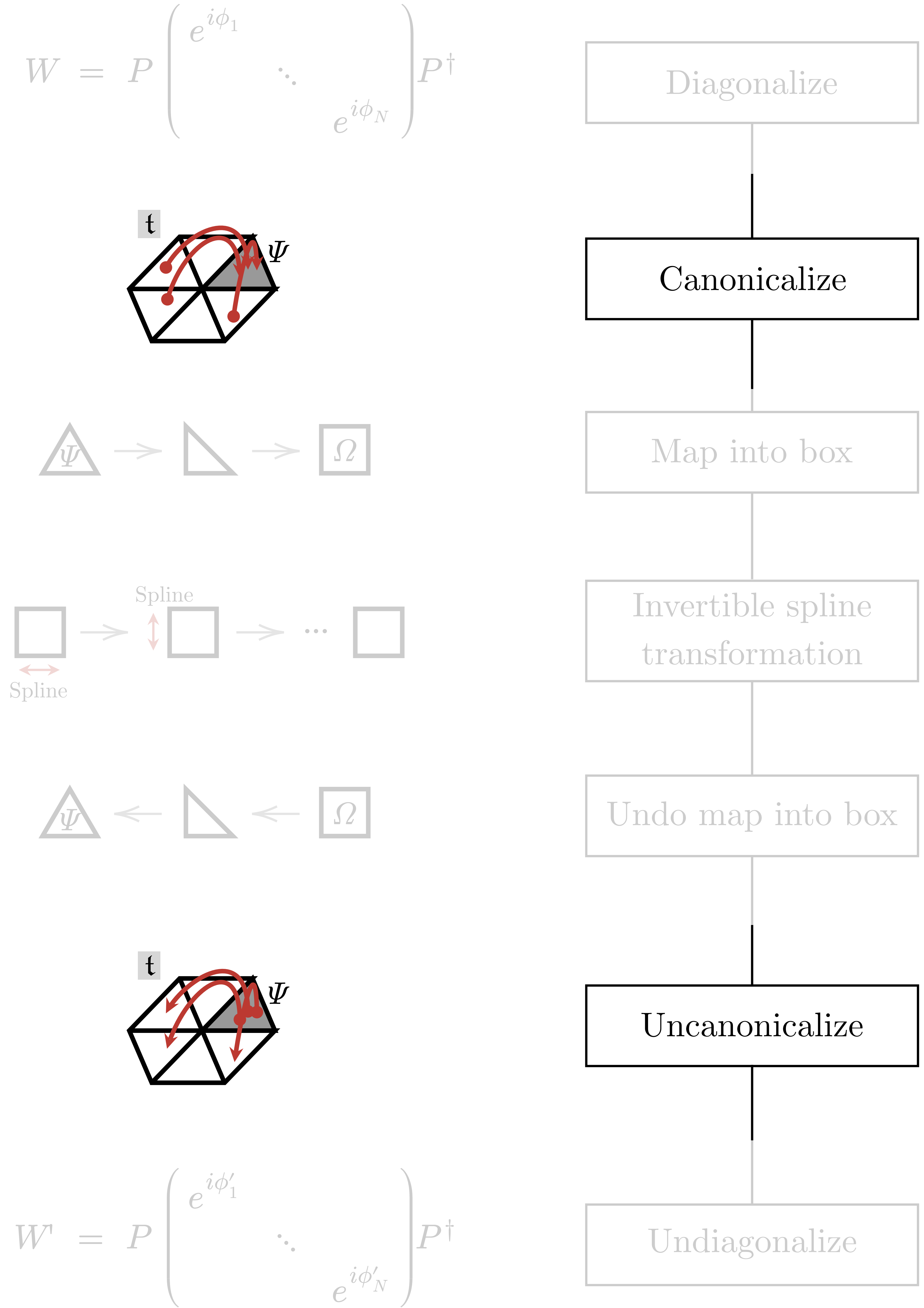
SU(N) kernels:
Permutation equivariance

See also [J. Thaler, Wed]
for perm-inv NNs



Sub-manifold of
$\det = 1$ eigenvalues

“Cell”, related by
perms of eigenvalues
to other cells.



$$W = P \begin{pmatrix} e^{i\phi_1} & & \\ & \ddots & \\ & & e^{i\phi_N} \end{pmatrix} P^\dagger$$

Diagonalize

Canonicalize

Map into box

Invertible spline transformation

Undo map into box

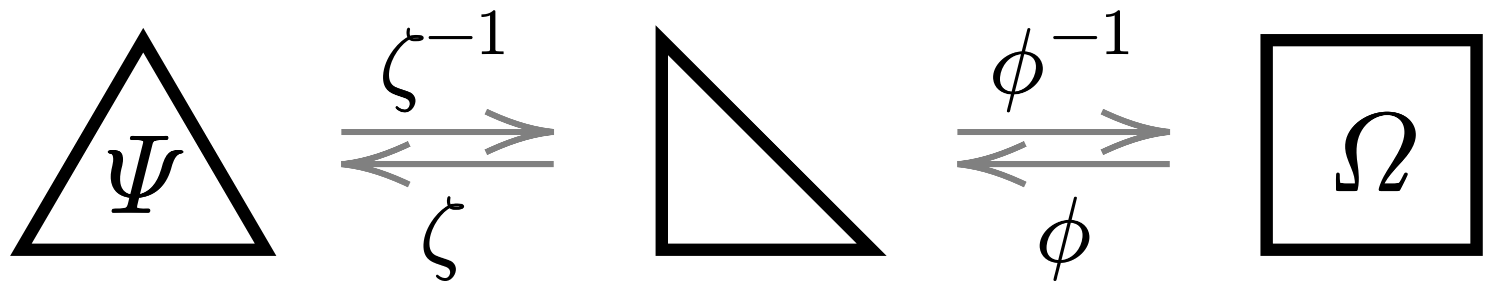
Uncanonicalize

$$W' = P \begin{pmatrix} e^{i\phi'_1} & & \\ & \ddots & \\ & & e^{i\phi'_N} \end{pmatrix} P^\dagger$$

Undiagonalize

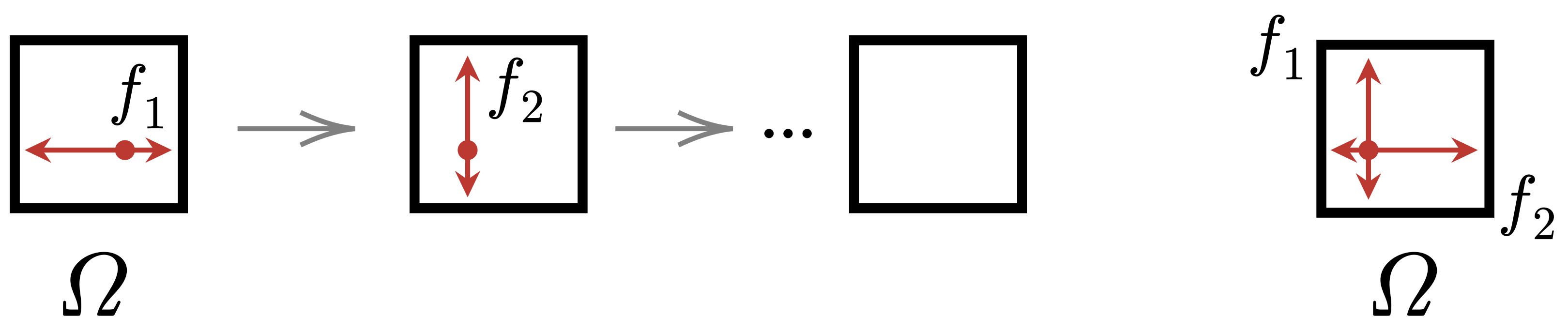
SU(N) kernels:
Transform the canonical cell

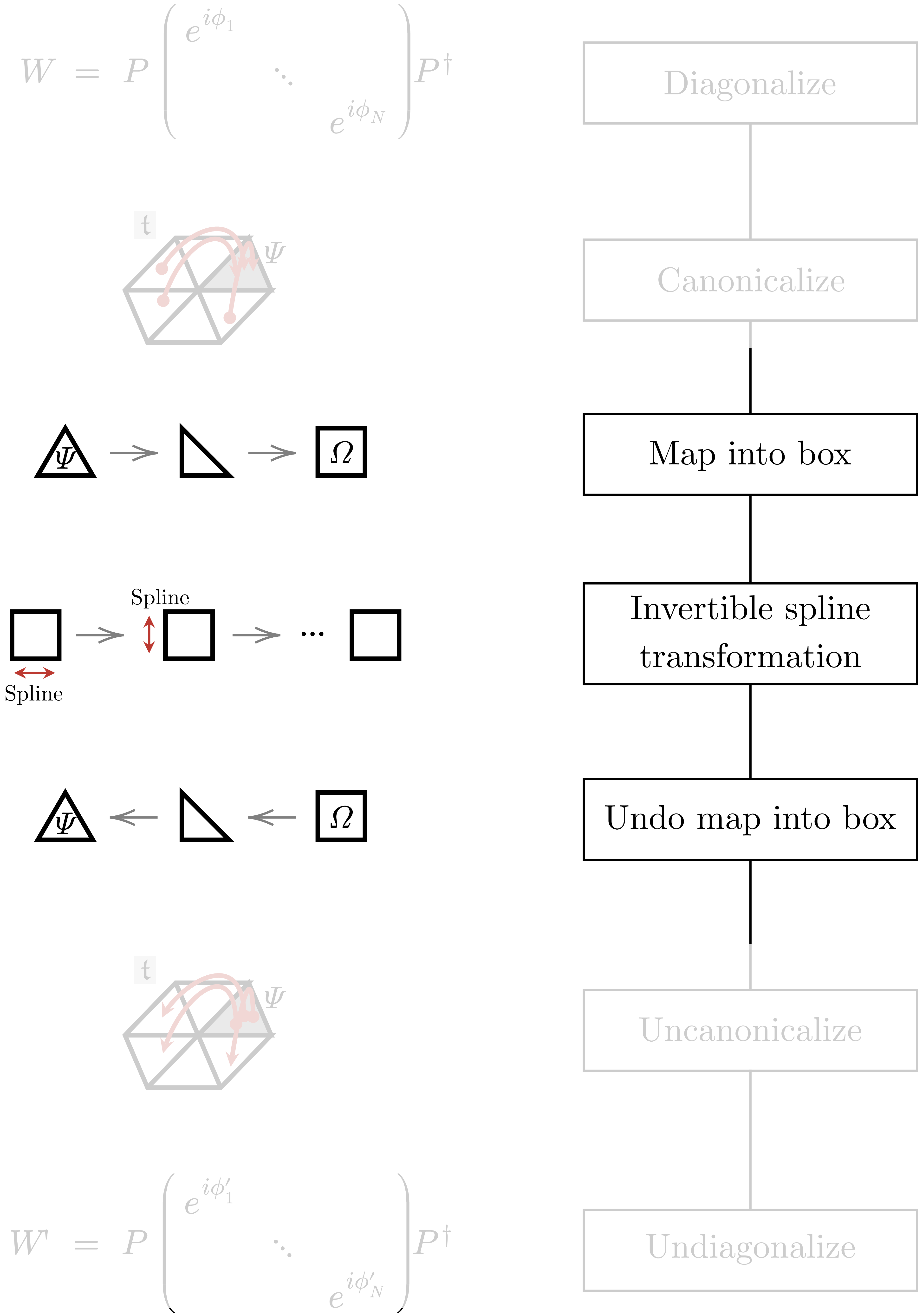
Change variables to rectilinear box Ω



Transform by acting on coords of box Ω, either…

Autoregressive … or … Independent





$$W = P \begin{pmatrix} e^{i\phi_1} & & \\ & \ddots & \\ & & e^{i\phi_N} \end{pmatrix} P^\dagger$$

Diagonalize

Canonicalize

Map into box

Invertible spline transformation

Undo map into box

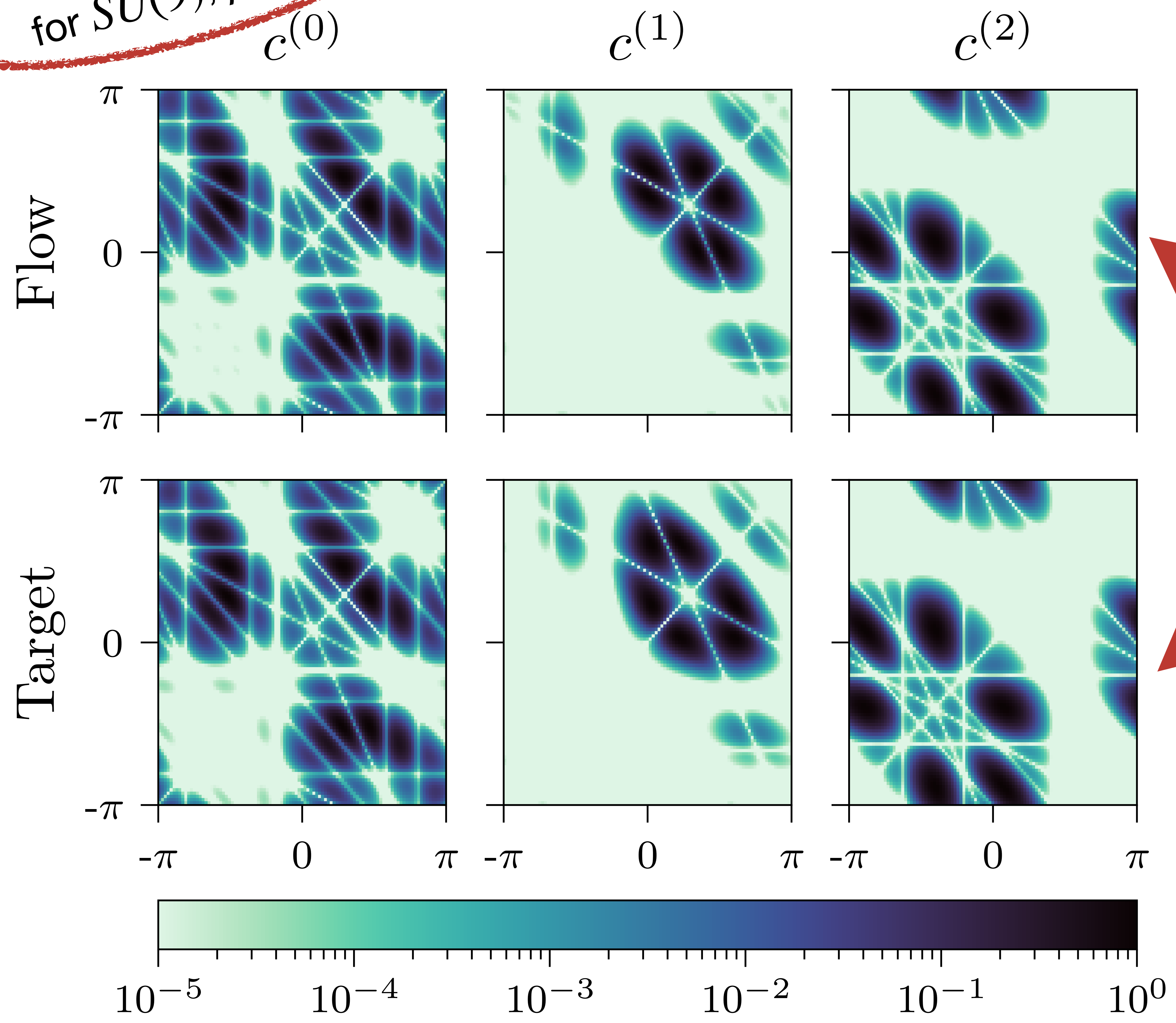
Uncanonicalize

$$W' = P \begin{pmatrix} e^{i\phi'_1} & & \\ & \ddots & \\ & & e^{i\phi'_N} \end{pmatrix} P^\dagger$$

Undiagonalize

Testing SU(N) kernels

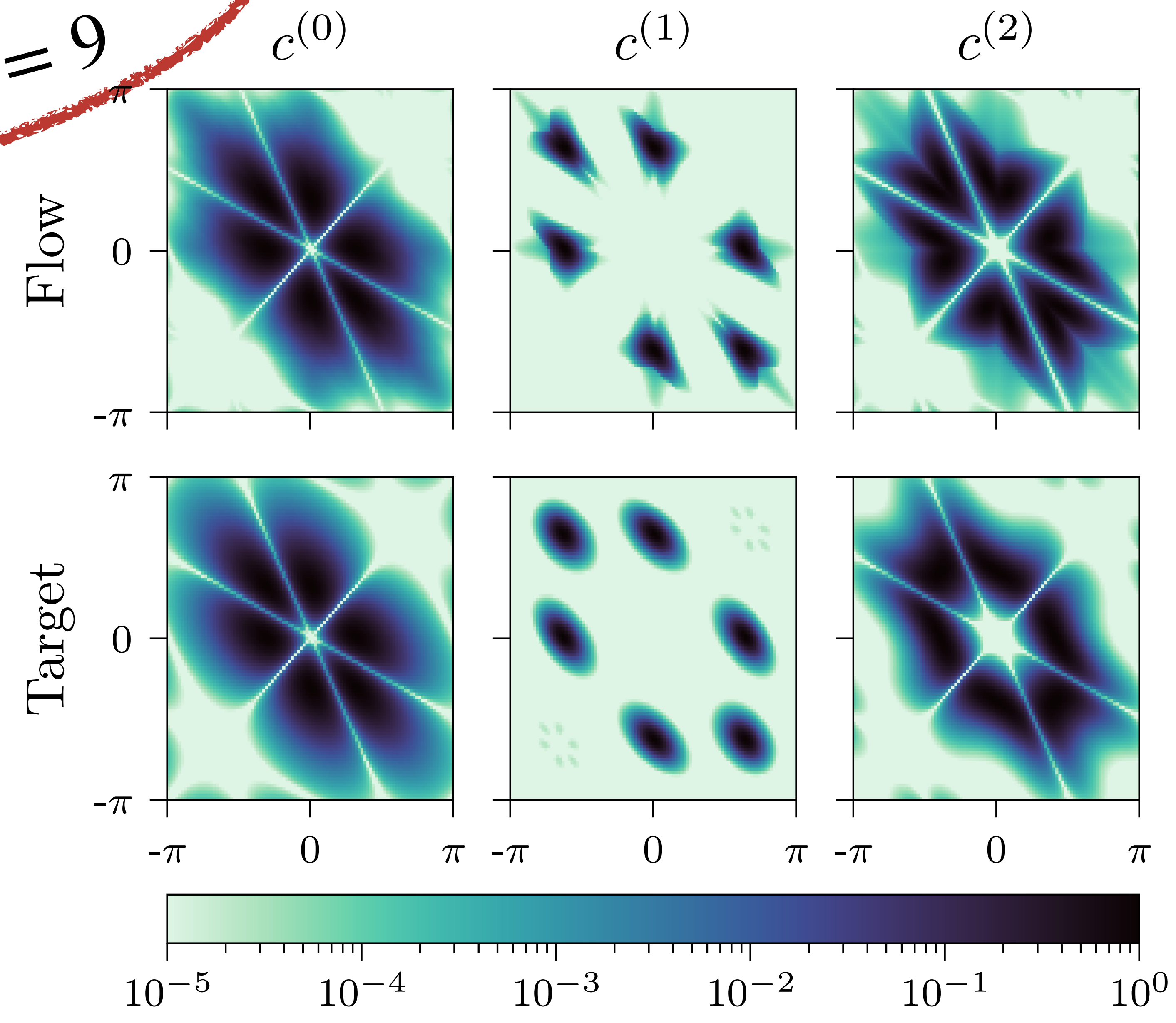
Plaquette distributions for $SU(9)$, $\beta = 9$



Agree!

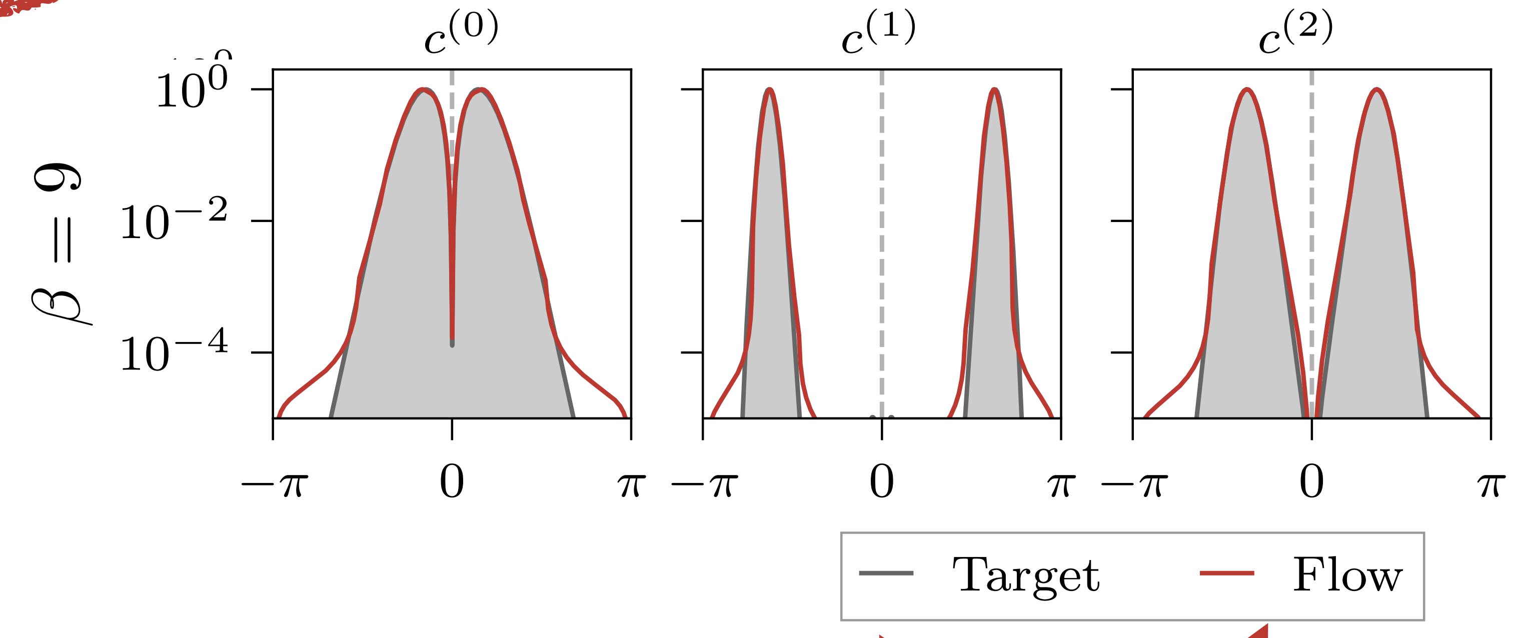
Density has zeros on vertical, horizontal, and diagonal lines where the slice crosses walls of cells

Plaquette distributions for $SU(3)$, $\beta = 9$



Agree!

Plaquette distributions for $SU(2)$, $\beta = 9$

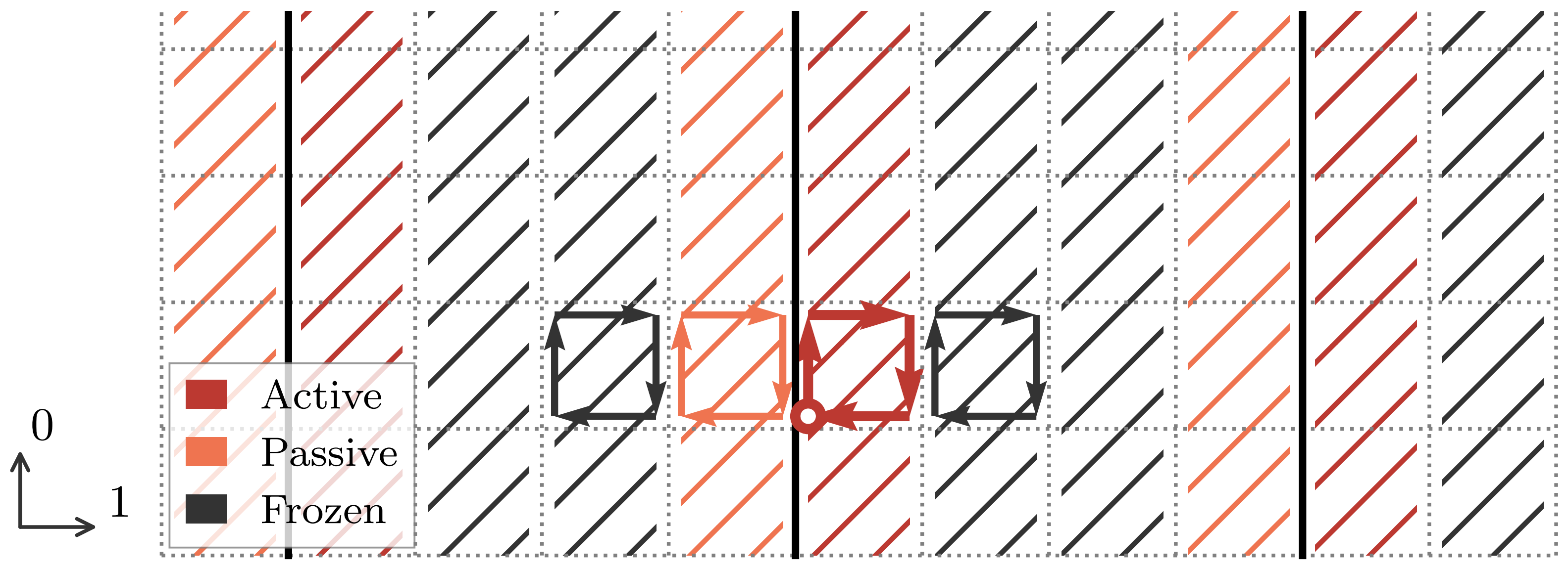


Agree!

Learning SU(2) and SU(3) gauge theory

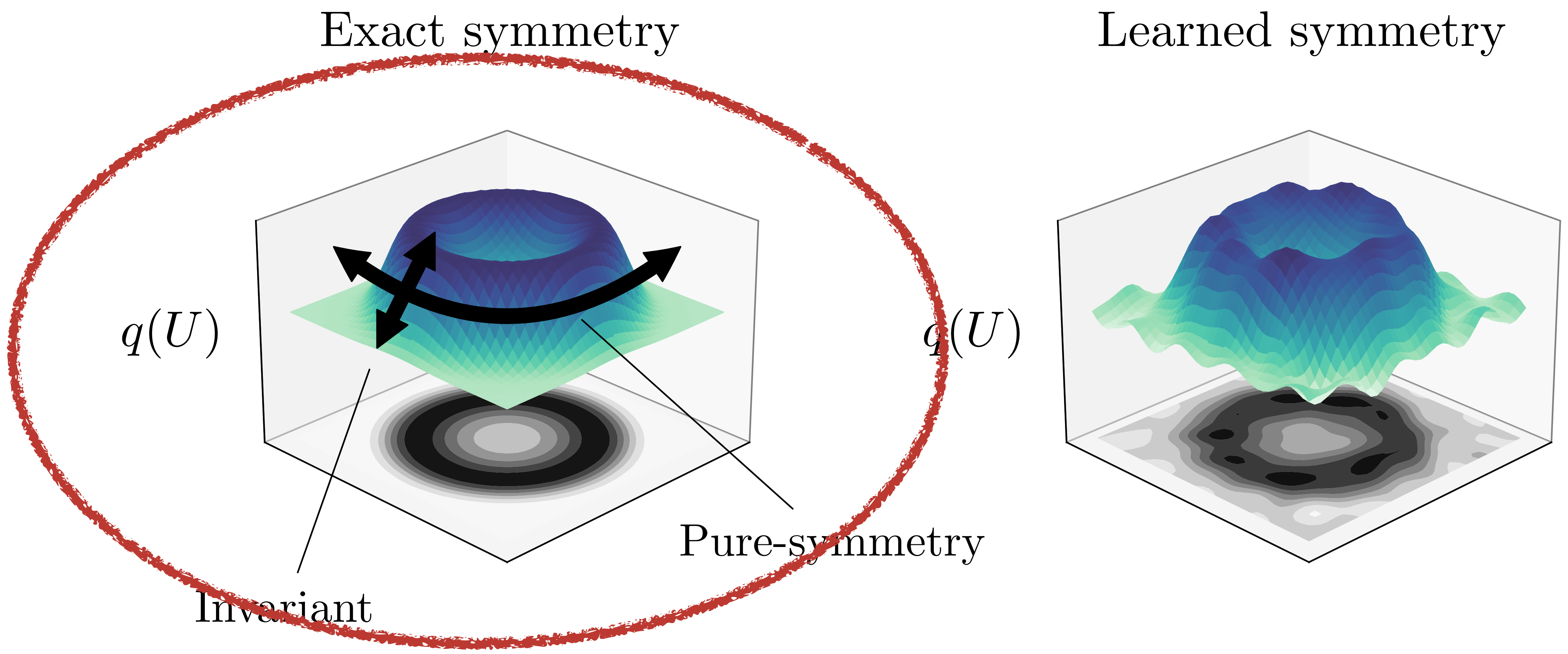
Normalizing flows trained for 2D lattice gauge theory on 16×16 lattices.

- Approx matched ’t Hooft couplings, giving $\beta = \{1.8, 2.2, 2.7\}$ for $SU(2)$ and $\beta = \{4.0, 5.0, 6.0\}$ for $SU(3)$



- 48 **PAFF coupling layers**, update all links 6 times
- No equivalent topo freezing, studied **absolute model quality** instead

All flow-based models exactly gauge-equiv by construction

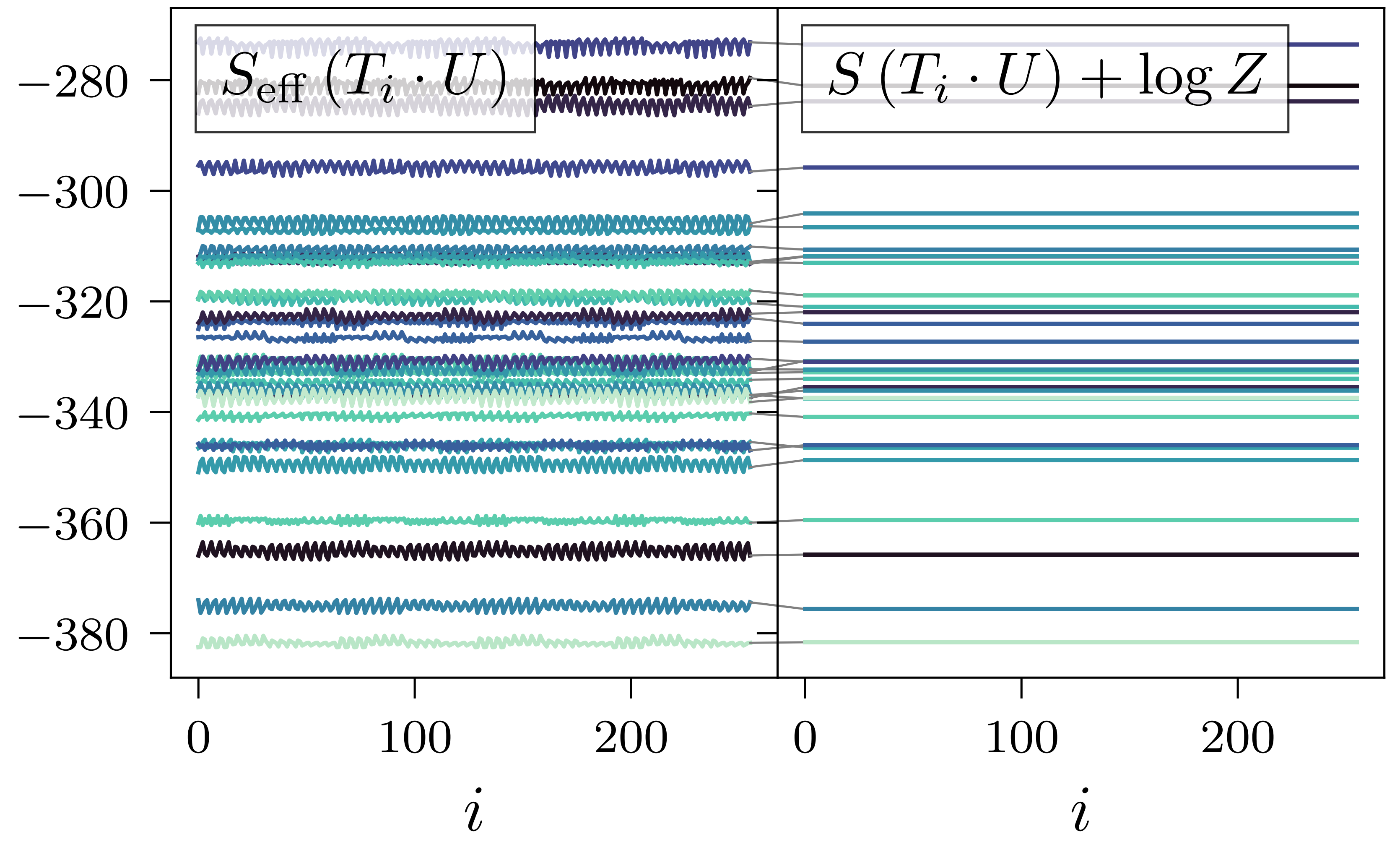


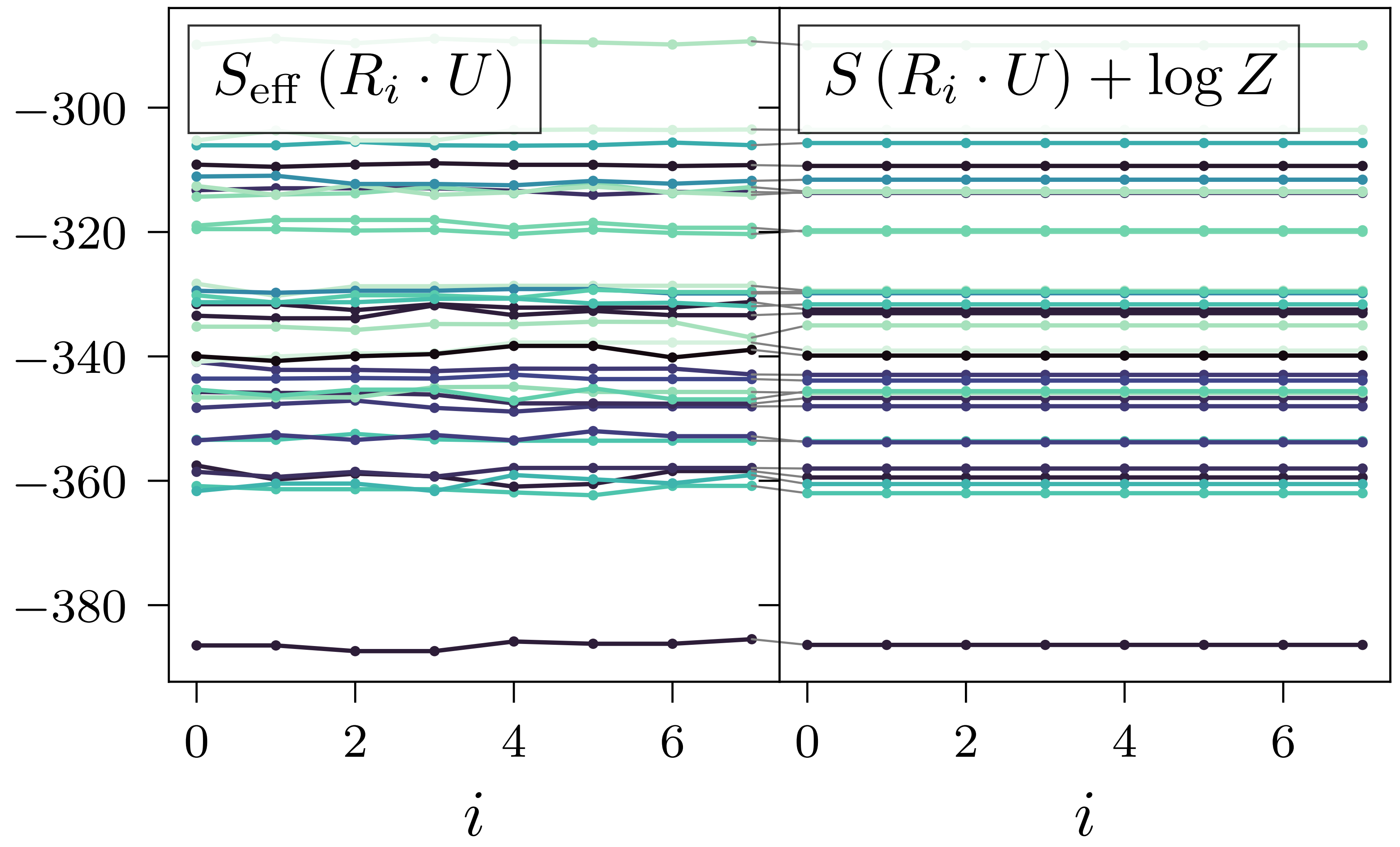
Results for SU(2) and SU(3) gauge theory

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

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

Measure of "effective" # samples from target dist for each sample drawn from model (100% = perfect model)


Exact translational subgroup, residual learned


Rotation and reflection symmetry learned

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

Summary and Outlook

Gauge symmetry encoded in flow models while preserving (most of) translational symmetry

- Gauge equivariant coupling layers
- Kernels for $U(1)$ and $SU(N)$

High-quality models produced for $U(1)$, $SU(2)$, and $SU(3)$ lattice gauge theory in 1+1D

Future work:

1. Higher spacetime dims. Choices of untraced loops to transform, gauge-inv loops as input?
2. Training for multimodal distributions

   [Hsieh, Chen, Chen, Albergo, Boyda, Cranmer, Hackett, GK, Saito, Shanahan; **In preparation**]
3. Training hyperparameter tuning, different model arch for inner flows
4. Incorporation of dynamical fermions

   [Albergo, GK, Racanière, Rezende, Urban, Boyda, Cranmer, Hackett, Shanahan; **In preparation**]

Open Questions

1. Universality?

- Limits in which (gauge-equiv) flow-based models capture arbitrary (gauge theory) distributions?
- How to avoid / diagnose silent “expressivity failure”?

2. Hierarchical structures?

- Asymptotic freedom, RG information
- Hierarchical models to efficiently capture many scales?

E.g. Hierarchical based on “disentanglers” and “decimators”: [Li, Wang **PRL121 (2018) 260601**]

3. Continuum scaling?

- How to scale model expressivity as $a \to 0$?
- Computational cost?
- How to scale model training?

Posted last night: [Del Debbio, Marsh-Rossney, Wilson **2105.12481**]

Backup Slides

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.

Translational equivariance

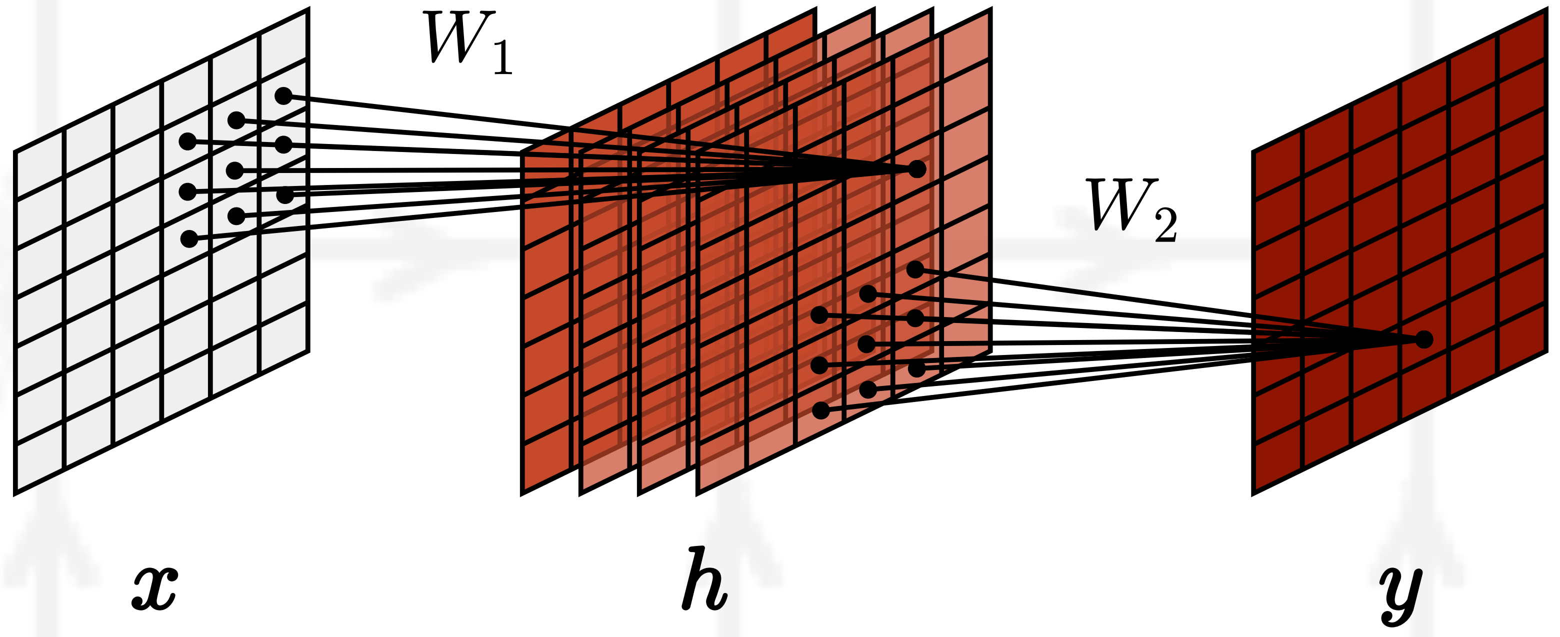
1. Make context functions Convolutional Neural Nets:

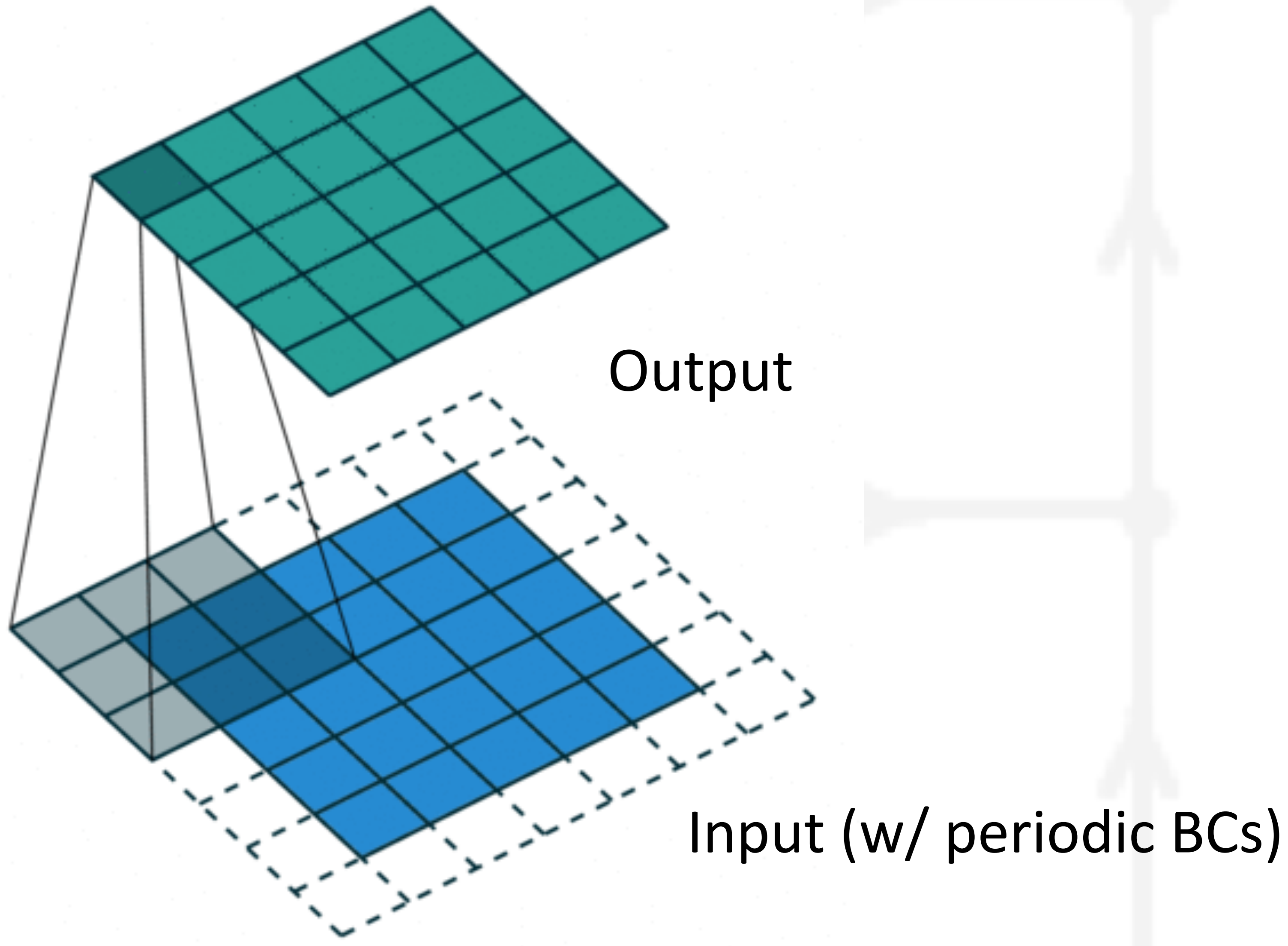
   - Compute output value for each site from linear transform of nearby DOF only
   - Reuse same weights, scanning kernel across the lattice

CNNs are equivariant under translations.

2. Make masking pattern (mostly) translationally invariant.

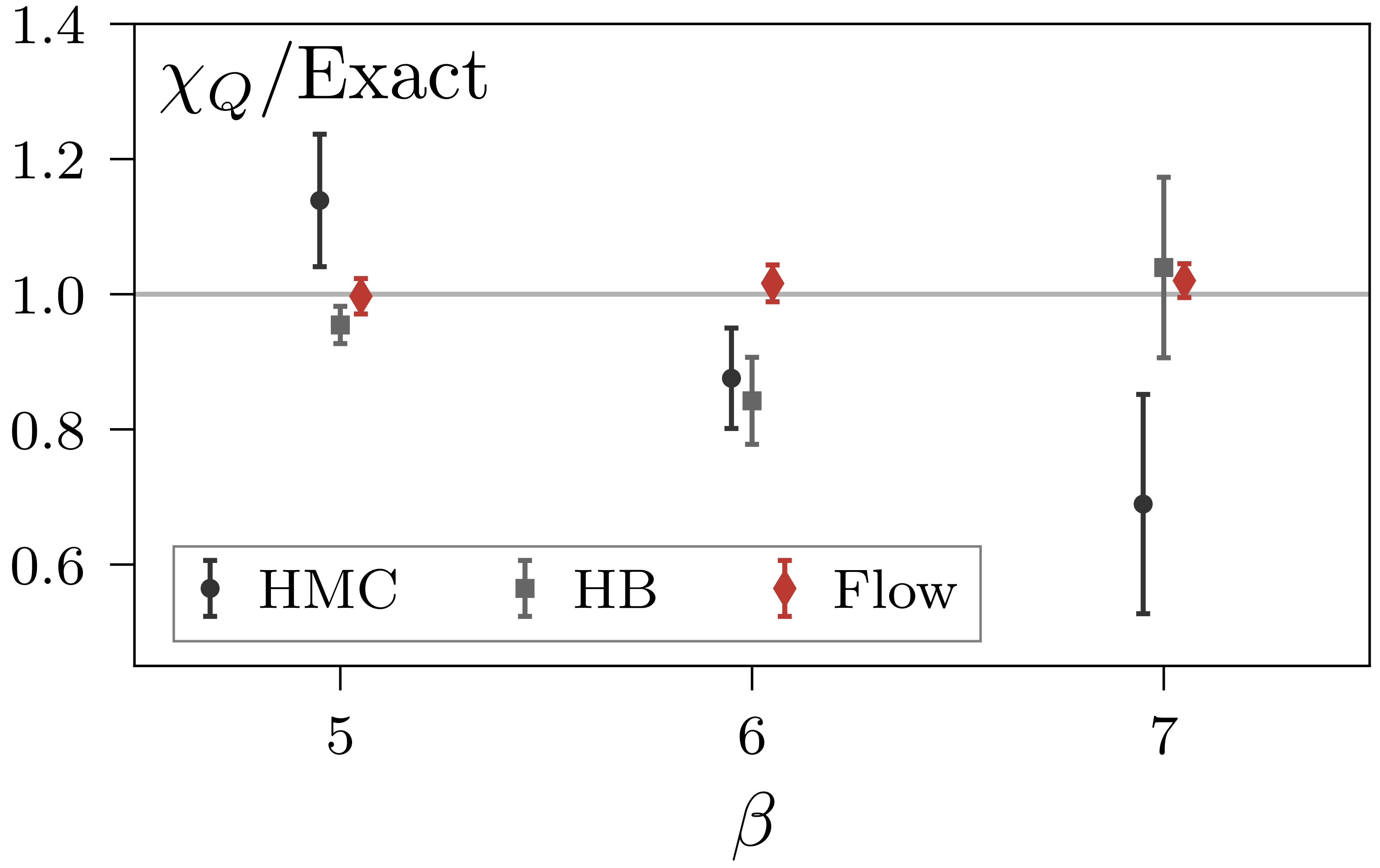
   - E.g. checkerboard is symmetric modulo $\mathbb{Z}_2$ even/odd
   - Gauge theory: translational equiv modulo $\mathbb{Z}_4 \times \mathbb{Z}_4$

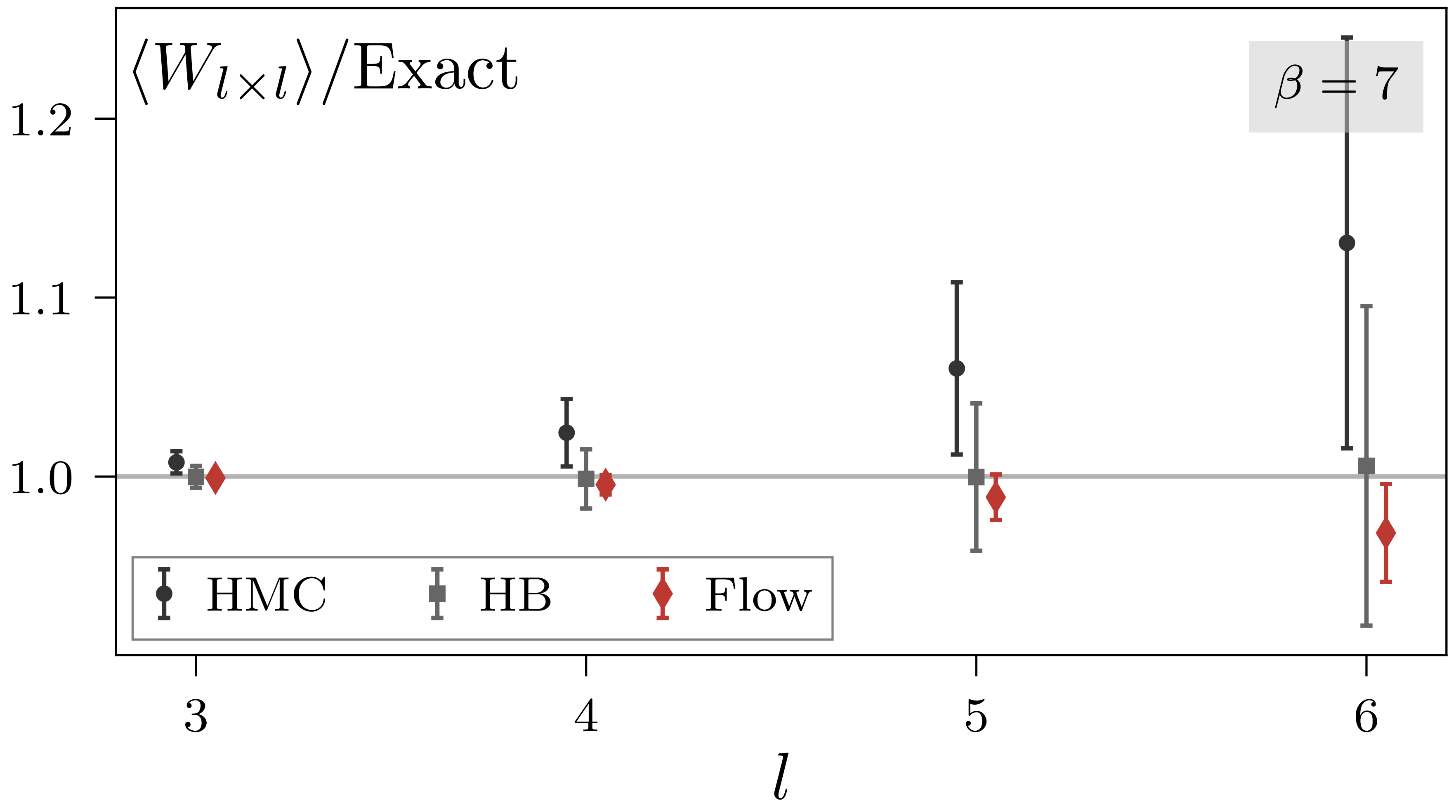


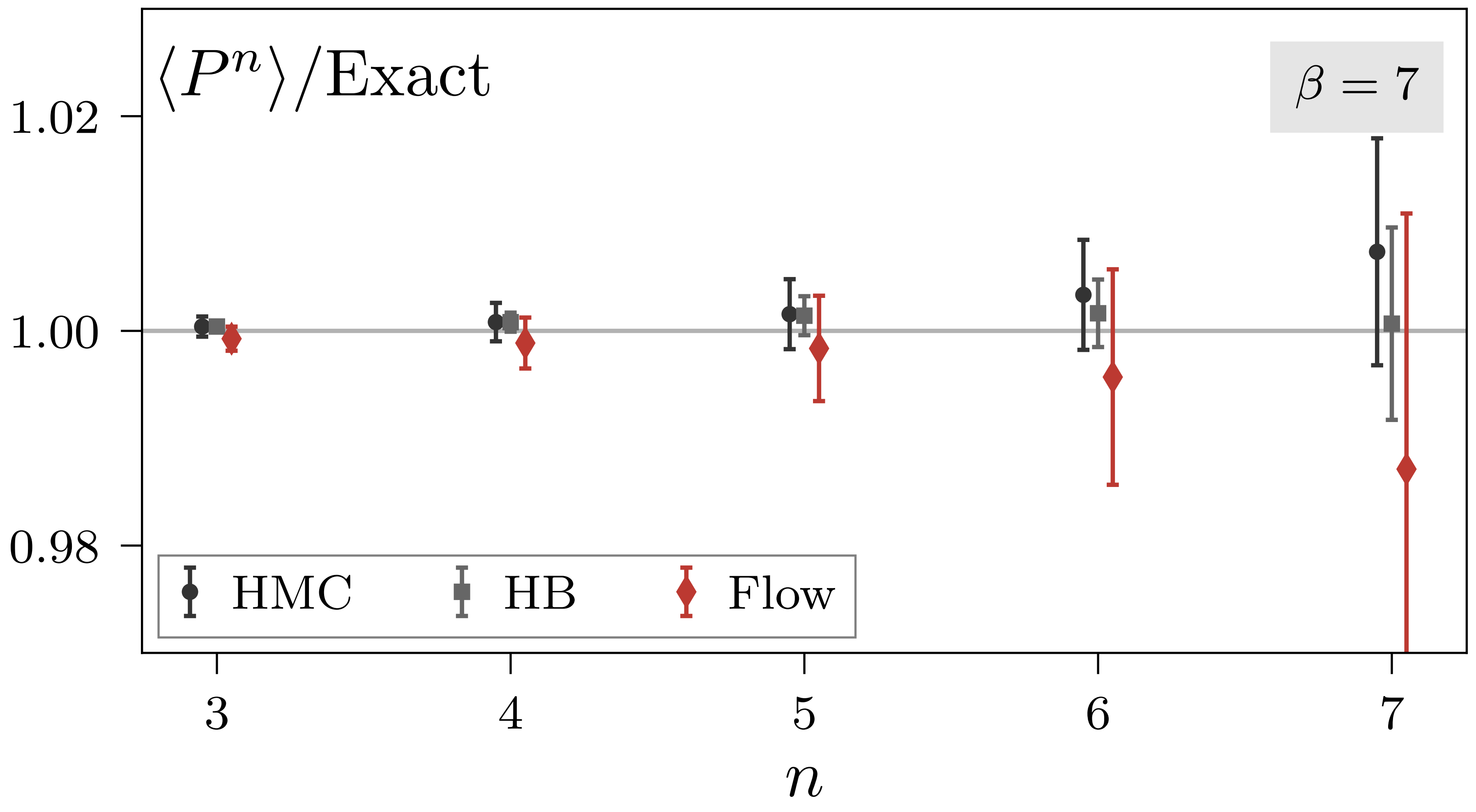


[freecodecamp.org]

U(1) study observables







Map into canonical cell

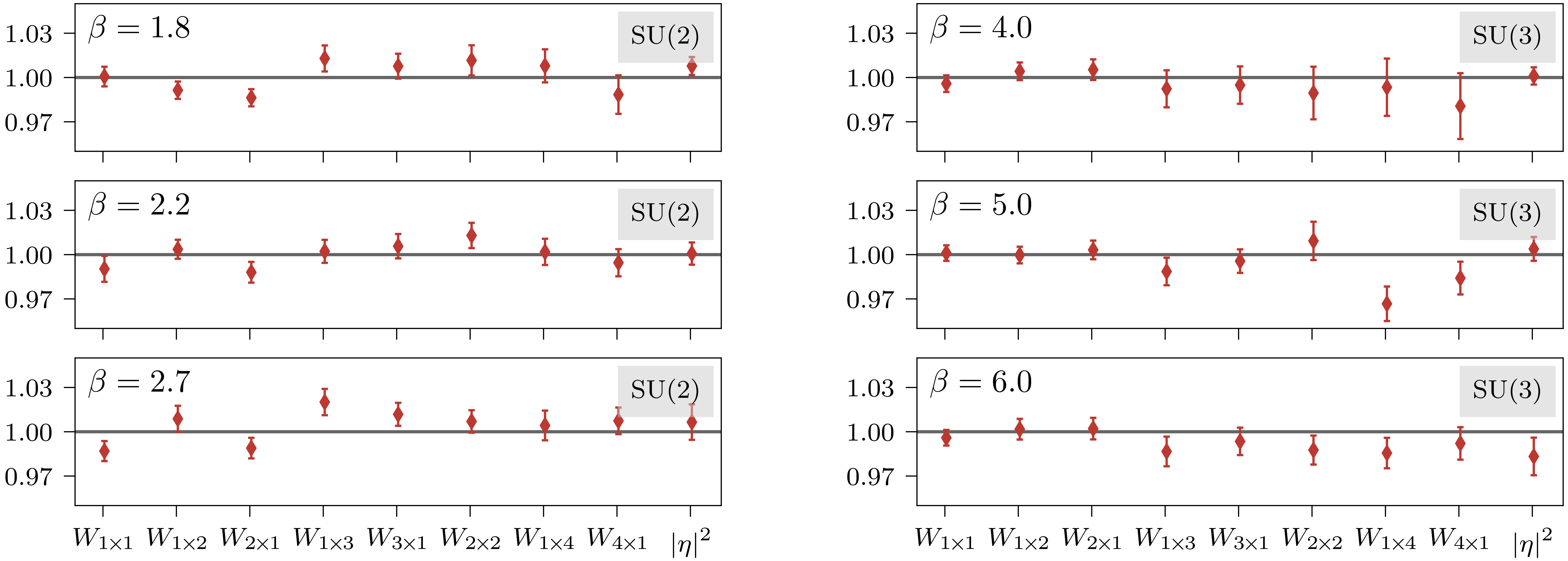
Want to permute eigenvalues into canonical order

- Sorting doesn't work directly: discontinuities when θ_k jumps across the $\pm\pi$ boundary
- Simply trying all $N!$ permutations is slow for large N
- Need to ensure permutation taking points in the same cell to canonical is the same

Short algorithm based on sorting works; Algorithm 1 of [‡]

[‡] [Boyda, **GK**, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan 2008.05456]

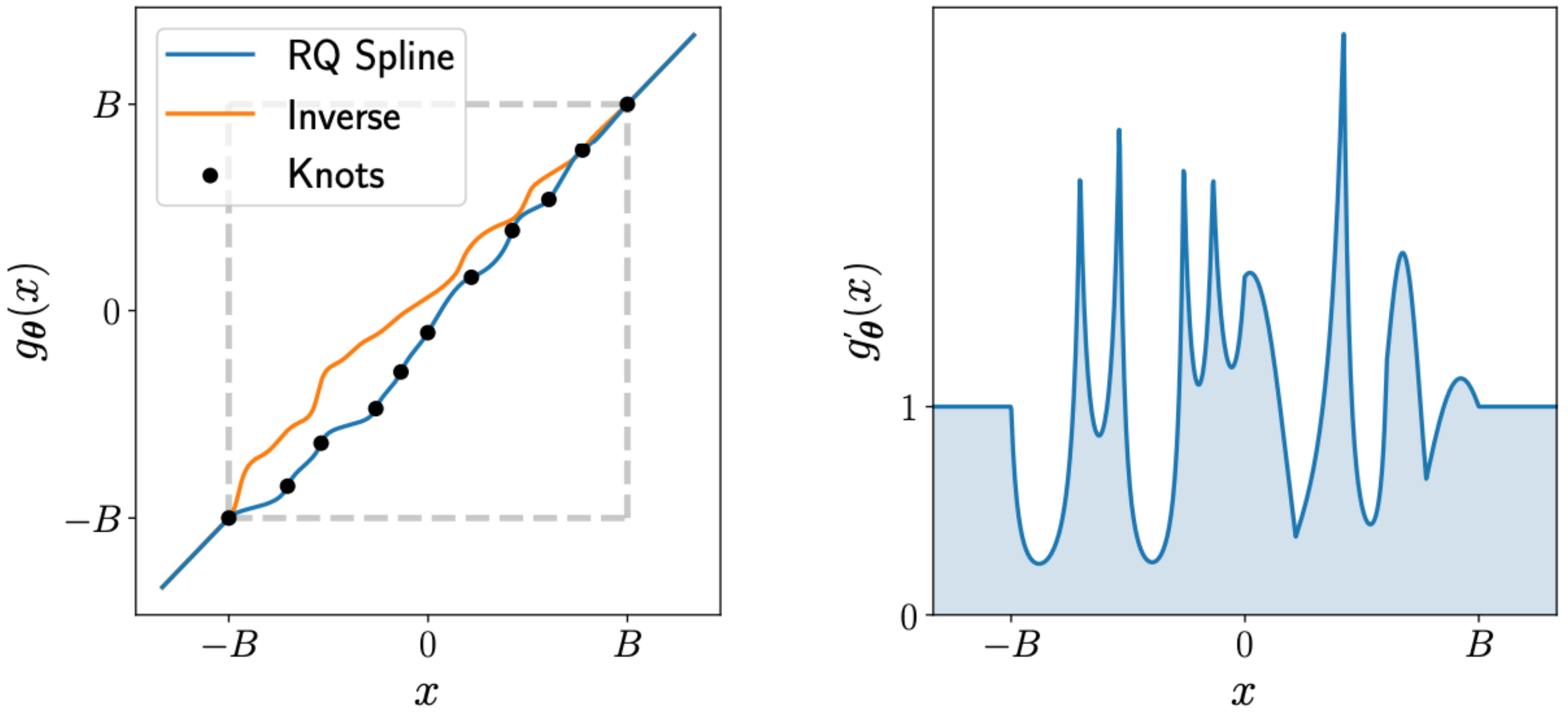
SU(N) study observables



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



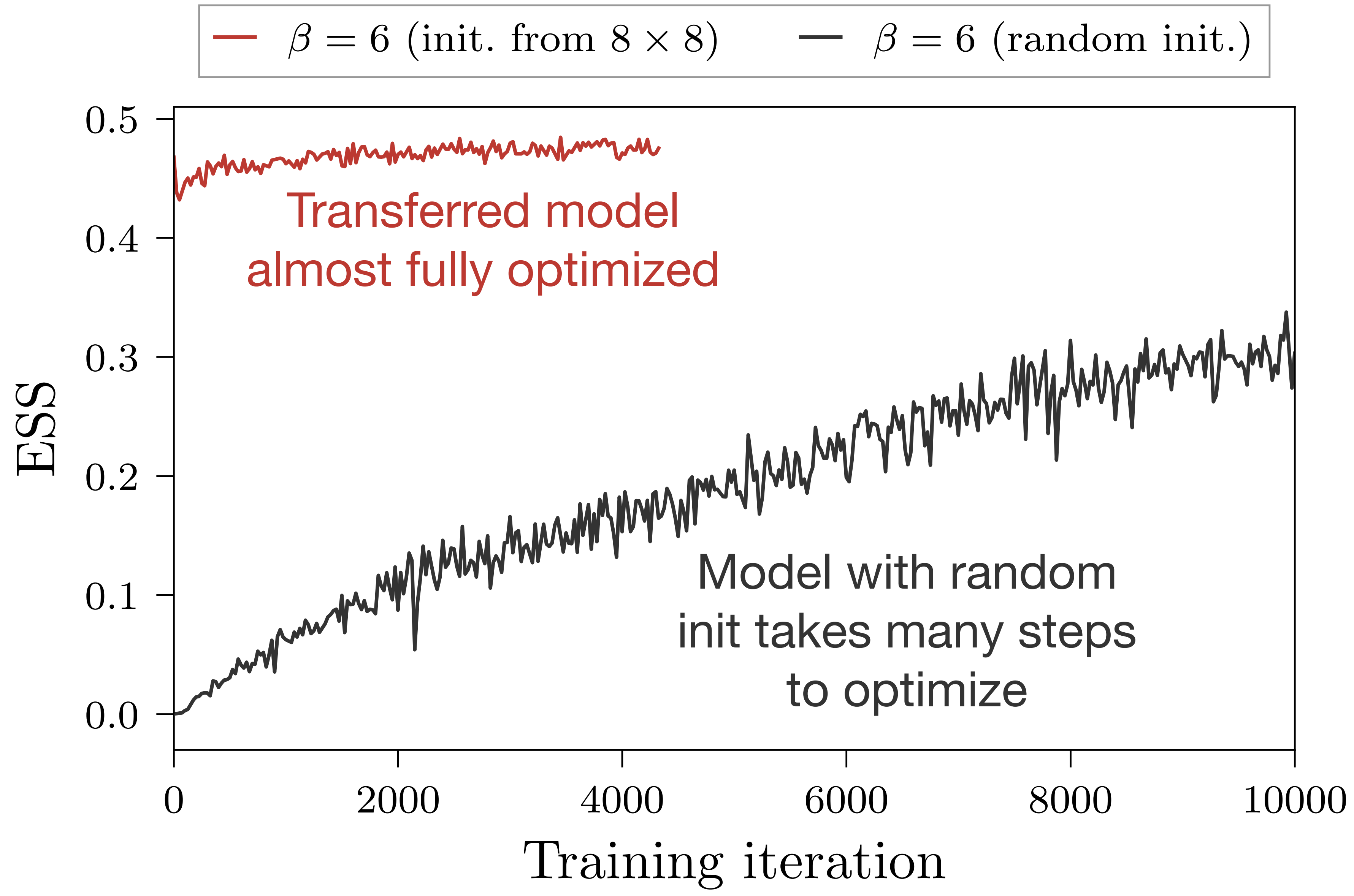
[Durkan, Bekasov, Murray, Papamakarios 1906.04032]

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)

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

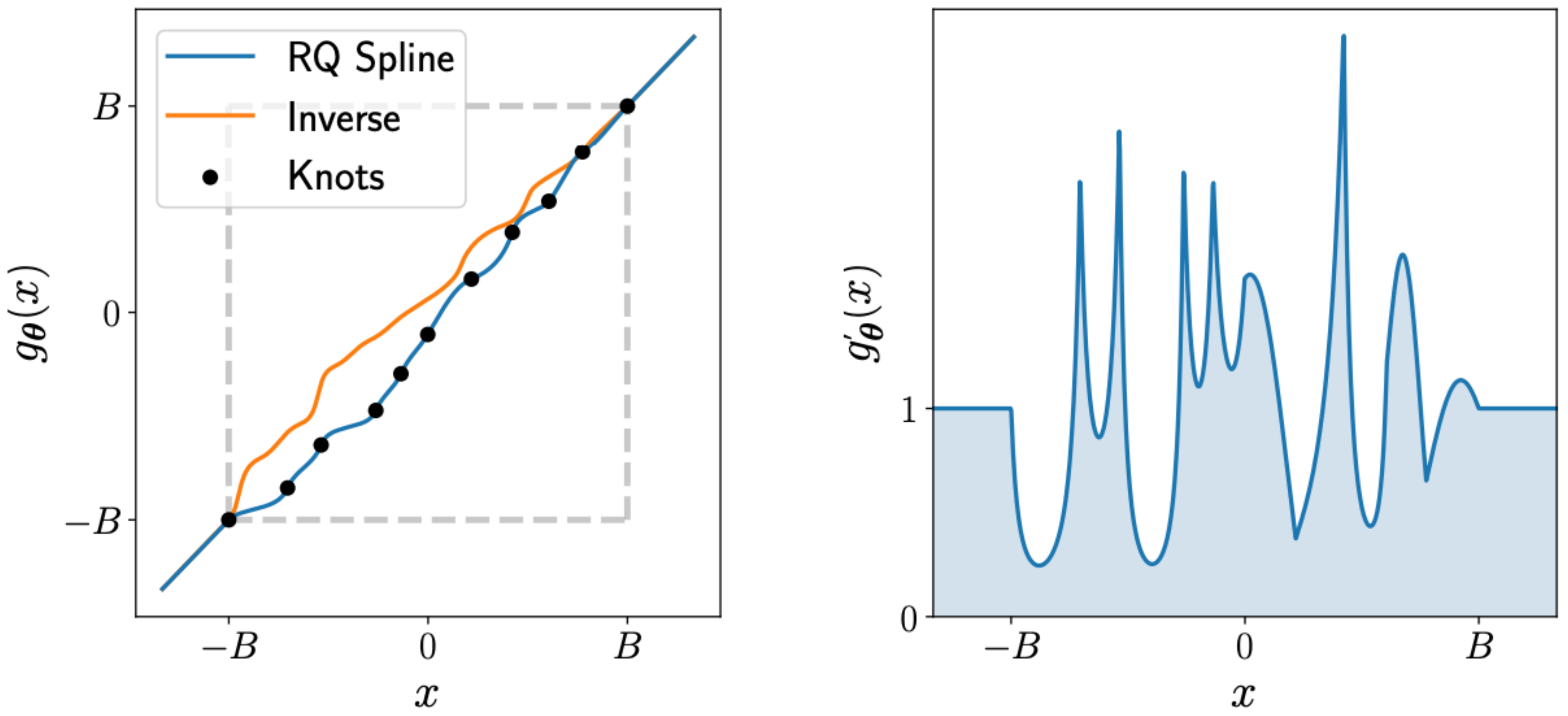
$$w(U) = p(U)/q(U)$$ “reweighting factors”



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

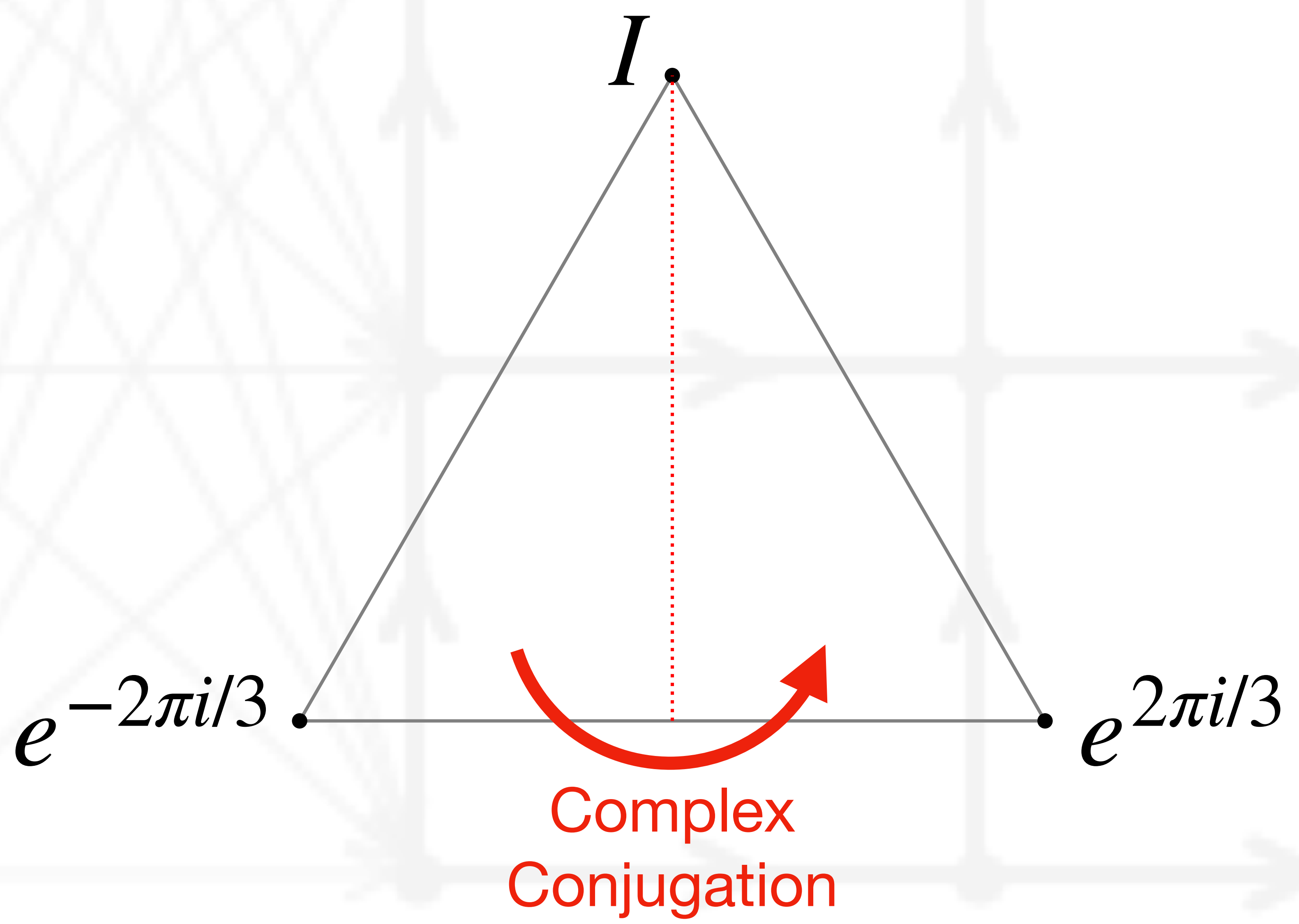
Details of $SU(3)$ models

- 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



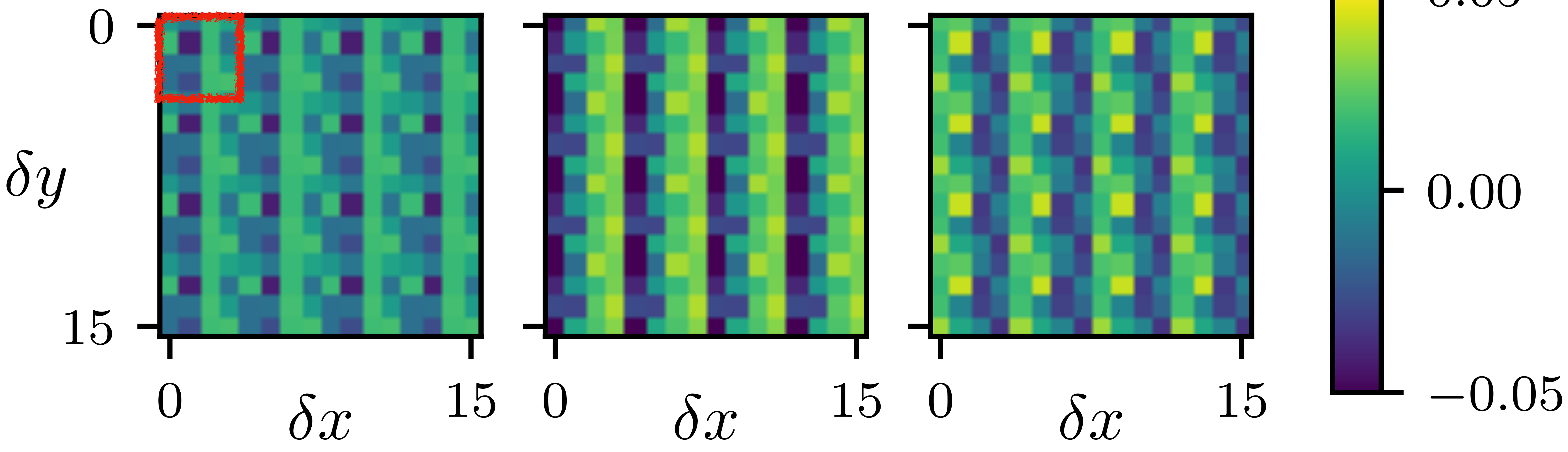
[Durkan, Bekasov, Murray, Papamakarios 1906.04032]

- CNNs to compute the knot locations
  - 32 hidden channels
  - 2 hidden layers
- Exact conjugation equivariance also imposed



Translational symmetry breaking pattern

- Masking patten = repeating tile of size 1×4
- Rotate / translate the pattern between layers
- $\mathbb{Z}_4 \times \mathbb{Z}_4$ symmetry breaking



Log density in 4x4 region extends by unbroken part of translational symmetry to rest of the lattice

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