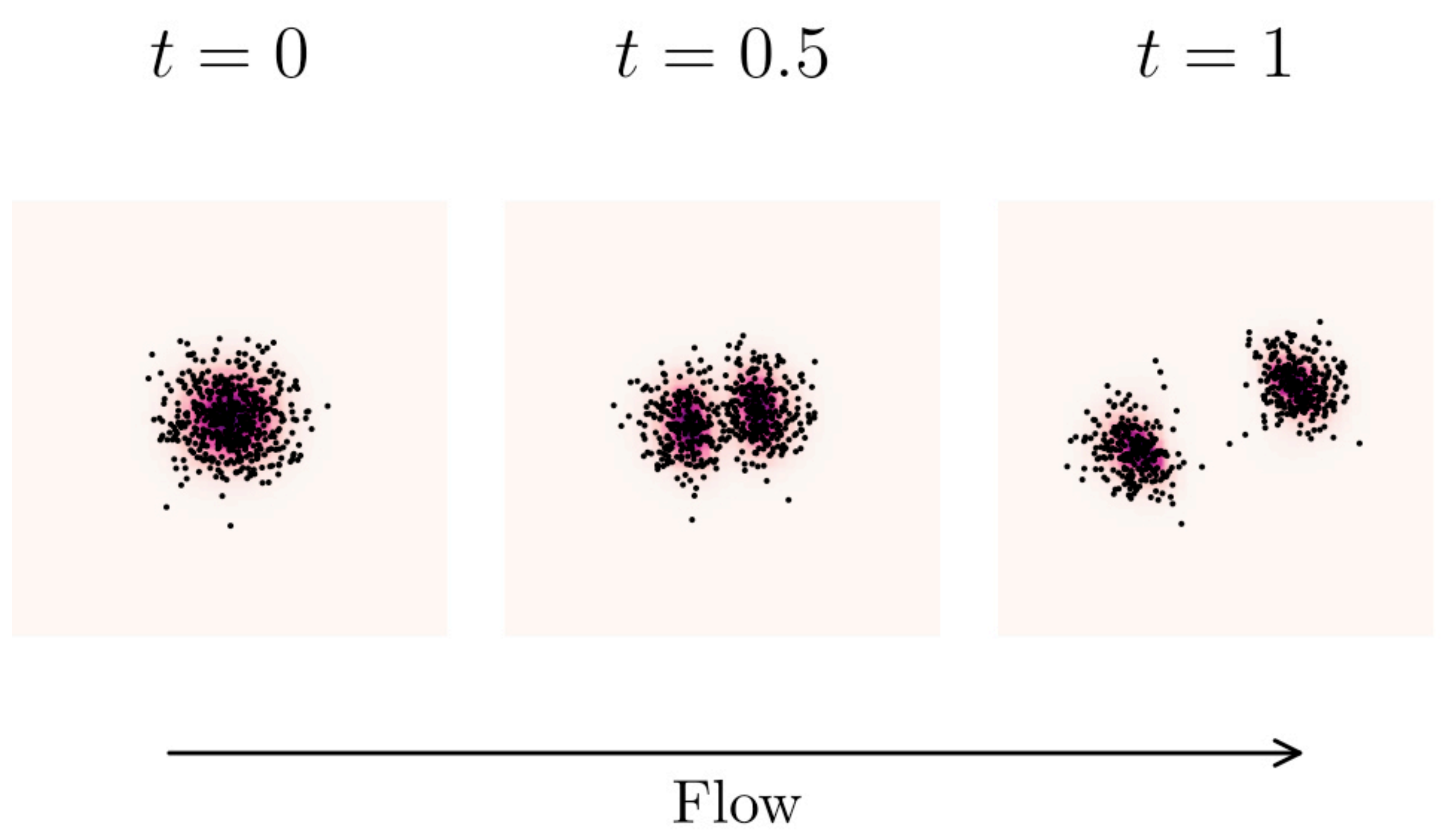
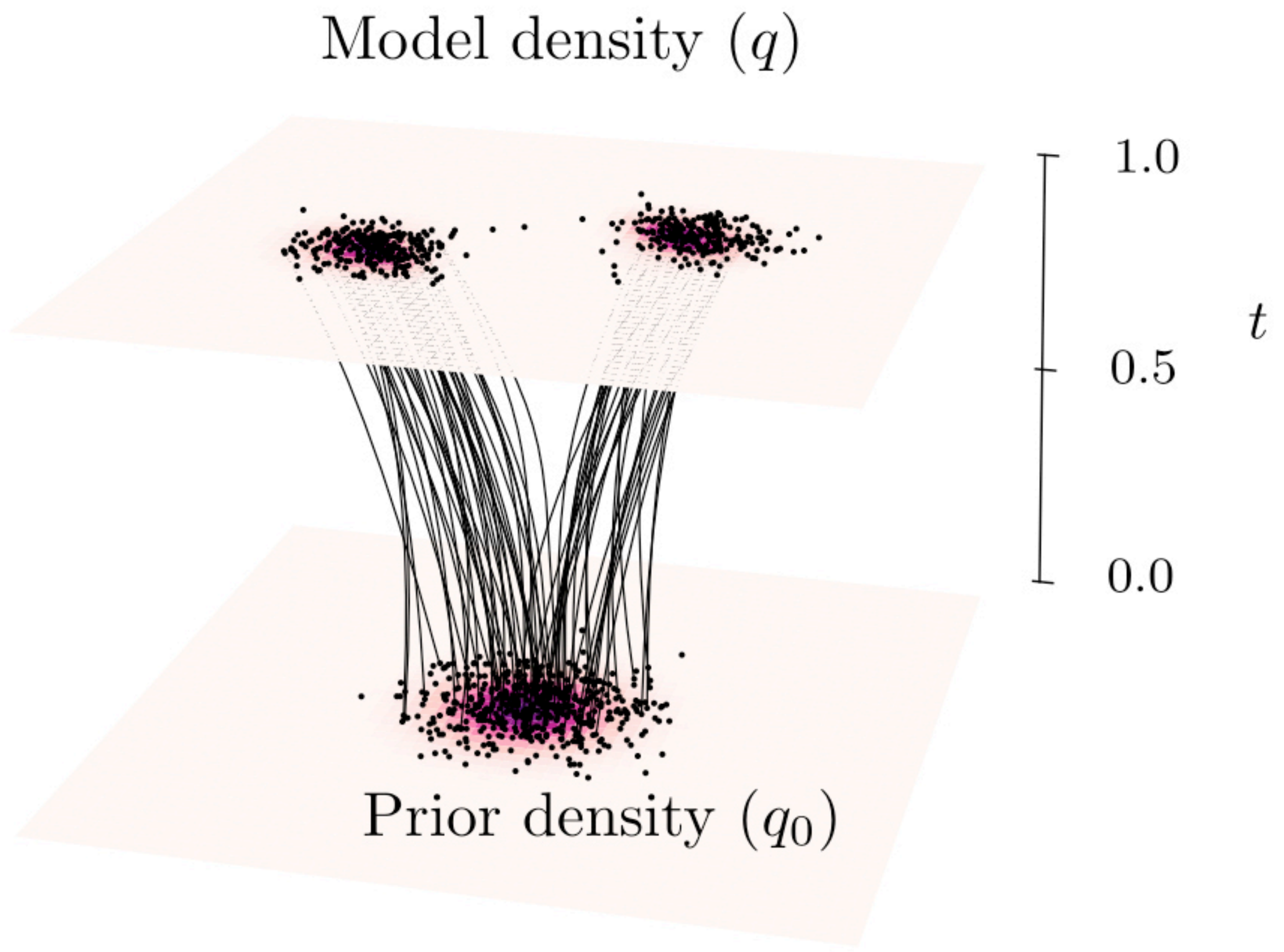
Normalizing flows for Lattice gauge theory





Gurtej Kanwar
Chancellor's Fellow in AI & Datascience
University of Edinburgh

Jul 28 - Aug 1, 2025
MITP Summer School

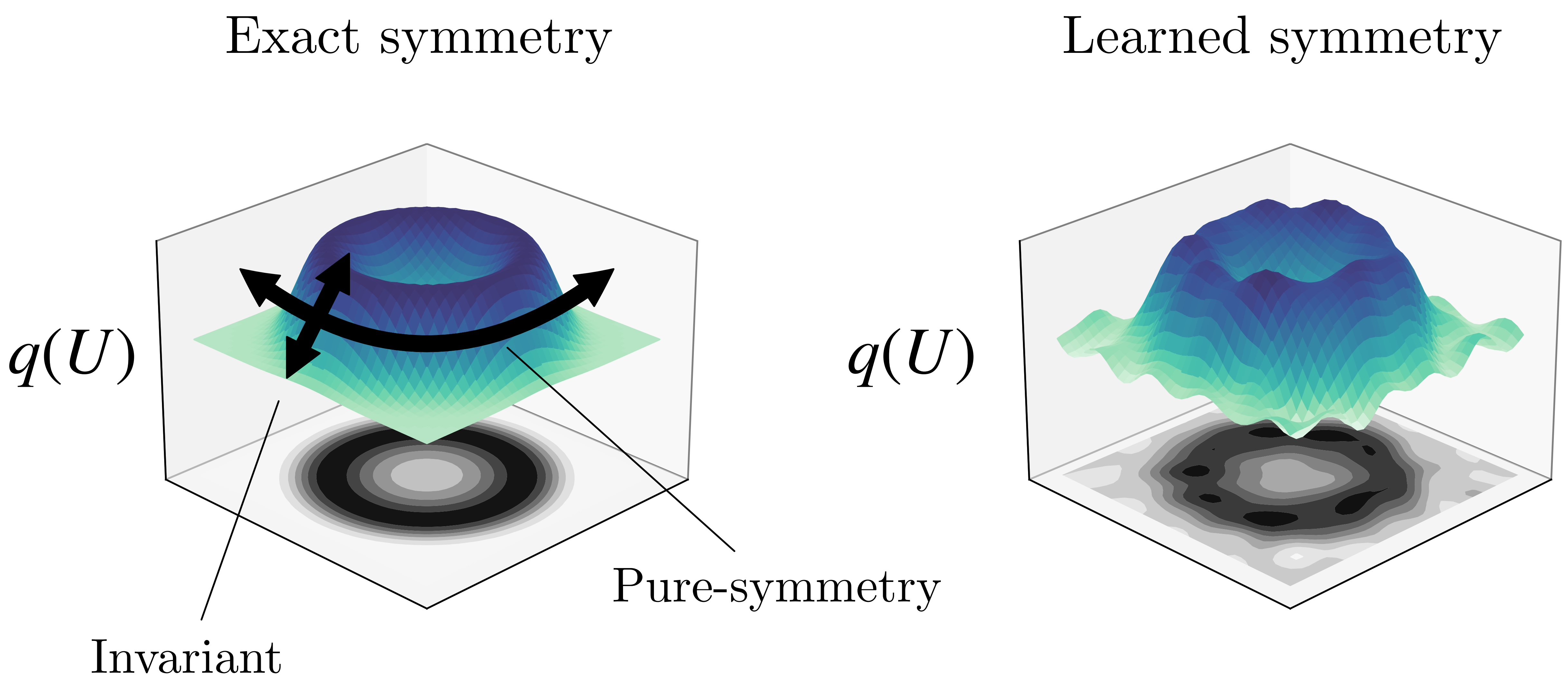
Lecture 3

Symmetries in flows

Motivation: Since target $p(U)$ is invariant under symmetries, natural to also make $q(U)$ invariant.

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

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



Symmetries…

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

Aside: “equivariant” vs “covariant”

Equivariance is a property of the map $f : G \to G$ saying that it **commutes** with gauge transformations g:

$$f(g \cdot U) = g \cdot f(U)$$

Covariance is a property of the fields $U \in G$ saying that it **transforms** in a specific way under gauge transformations g:

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

Gauge symmetry: discrete flows

Gauge symmetry for $\mathrm{SU}(3)$
lattice gauge theory

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

GK, et al. PRL125 (2020) 121601

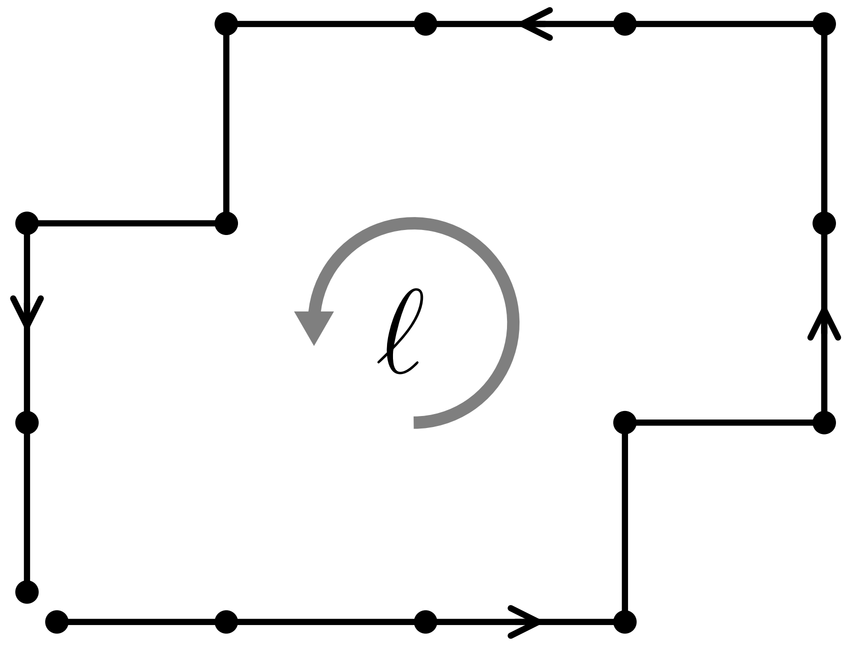
Gauge-invariant prior:

Uniform (Haar) distribution
$r(U) = 1$ works.

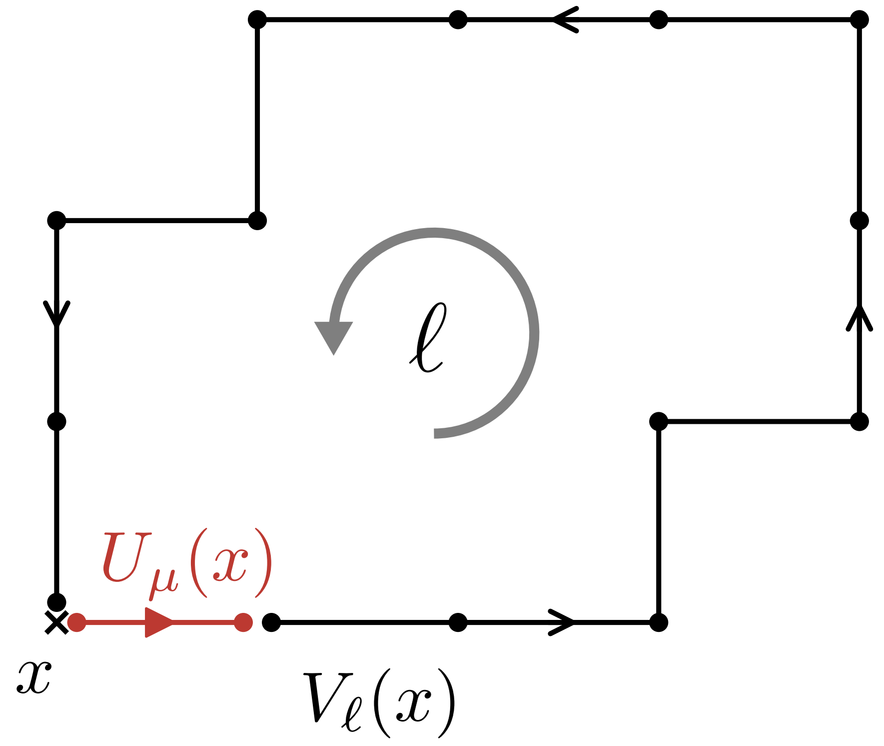
Gauge-equivariant flow:

Coupling layers acting on
(untraced) Wilson loops.

Loop transformation easier to satisfy.



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



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

Gauge symmetry: discrete flows

Gauge symmetry for $\mathrm{SU}(3)$
lattice gauge theory

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

Gauge-invariant prior:

Uniform (Haar) distribution
$r(U) = 1$ works.

Gauge-equivariant flow:

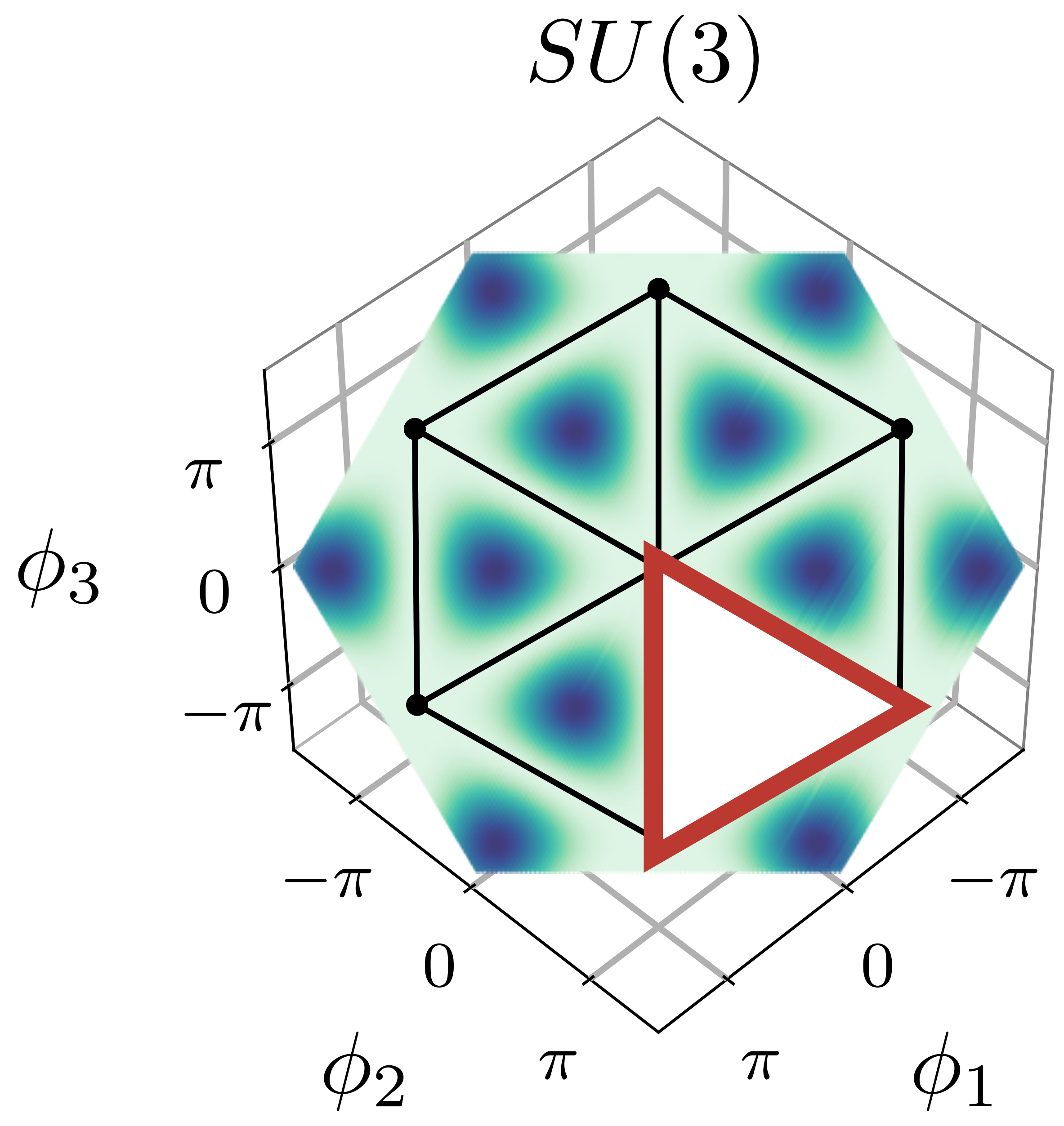
Coupling layers acting on
(untraced) Wilson loops.

Loop transformation easier to satisfy.

Custom flows designed
for $U(1)$ and $SU(N)$
gauge manifolds

GK, et al. PRL125 (2020) 121601

Rezende, et al. PMLR119 (2020) 8083
Boyda, et al. PRD103 (2021) 074504



Gauge symmetry: continuous flows

For gauge fields, continuous flows are described by velocity in the **Lie algebra:**

$$\dot{U}(t) = \nabla \varphi(U(t); t)\, U(t), \qquad \nabla \varphi(\,\cdot\,) \in \mathfrak{su}(3)^{N_l}$$

Gauge-equivariant flow:

- Make $\varphi(U)$ invariant
- Then the time-derivative transforms properly:

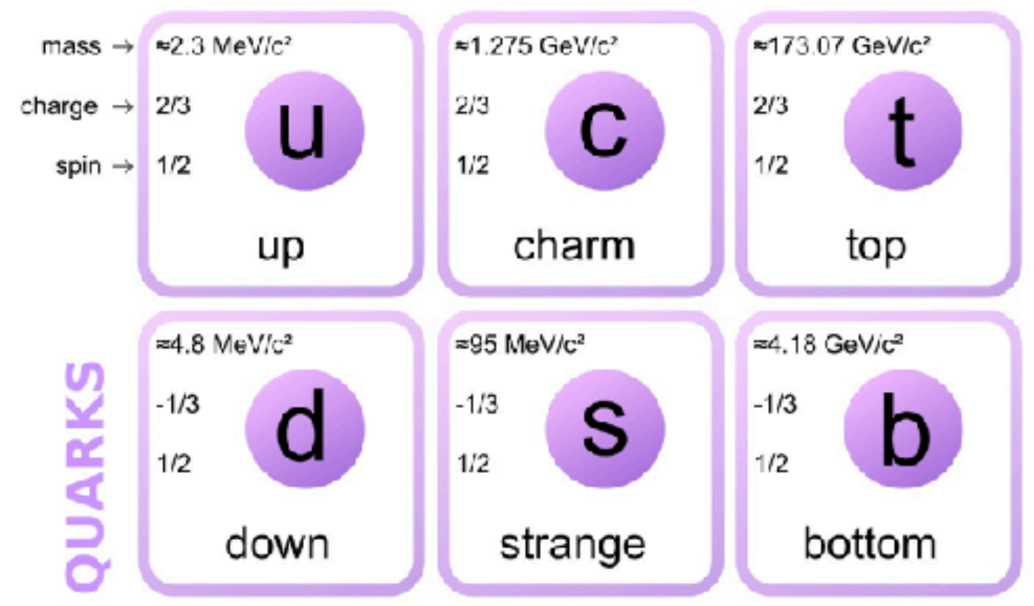
$$\underbrace{\nabla_{x,\mu}\varphi(g \cdot U)}_{\text{LEFT-ACTING DERIVATIVE LIVES AT } x}\,(g \cdot U_{\mu,x}) = (\Omega_x \nabla_{x,\mu}\varphi(U)\Omega_x^\dagger)(\Omega_x U_{\mu,x}\Omega_{x+\hat{\mu}}^\dagger) = \boxed{g \cdot \dot{U}(t)}$$

Including the quarks

Interaction between all quark flavors ($\psi_u, \psi_d, \ldots$) and gluons (U):

Action $$S_f = \sum_f \bar{\psi}_f D_f[U] \psi_f$$

Path integral $$\int \prod_f [d\bar{\psi} d\psi] e^{-S_f} = \prod_f \det(D_f[U])$$



- D_f is a sparse $O(V) \times O(V)$ matrix
- Typically use the **pseudofermion** representation for pairs of quark flavors

$$\det(D^2) \propto \int d\phi^\dagger d\phi \, e^{-\phi^\dagger D^{-1} \phi}$$

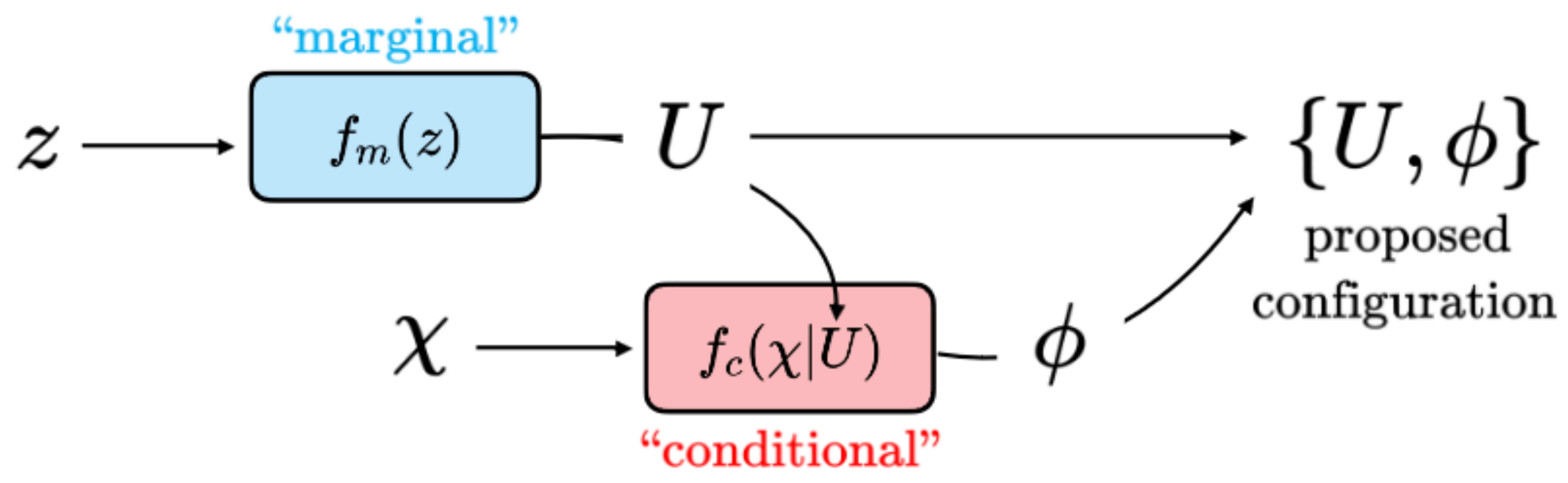
PRD104 (2021) 114507
PRD106 (2022) 014514
PRD106 (2022) 074506

Flows with pseudofermions

Pseudofermions highly effective in HMC, logical to use for flows also.

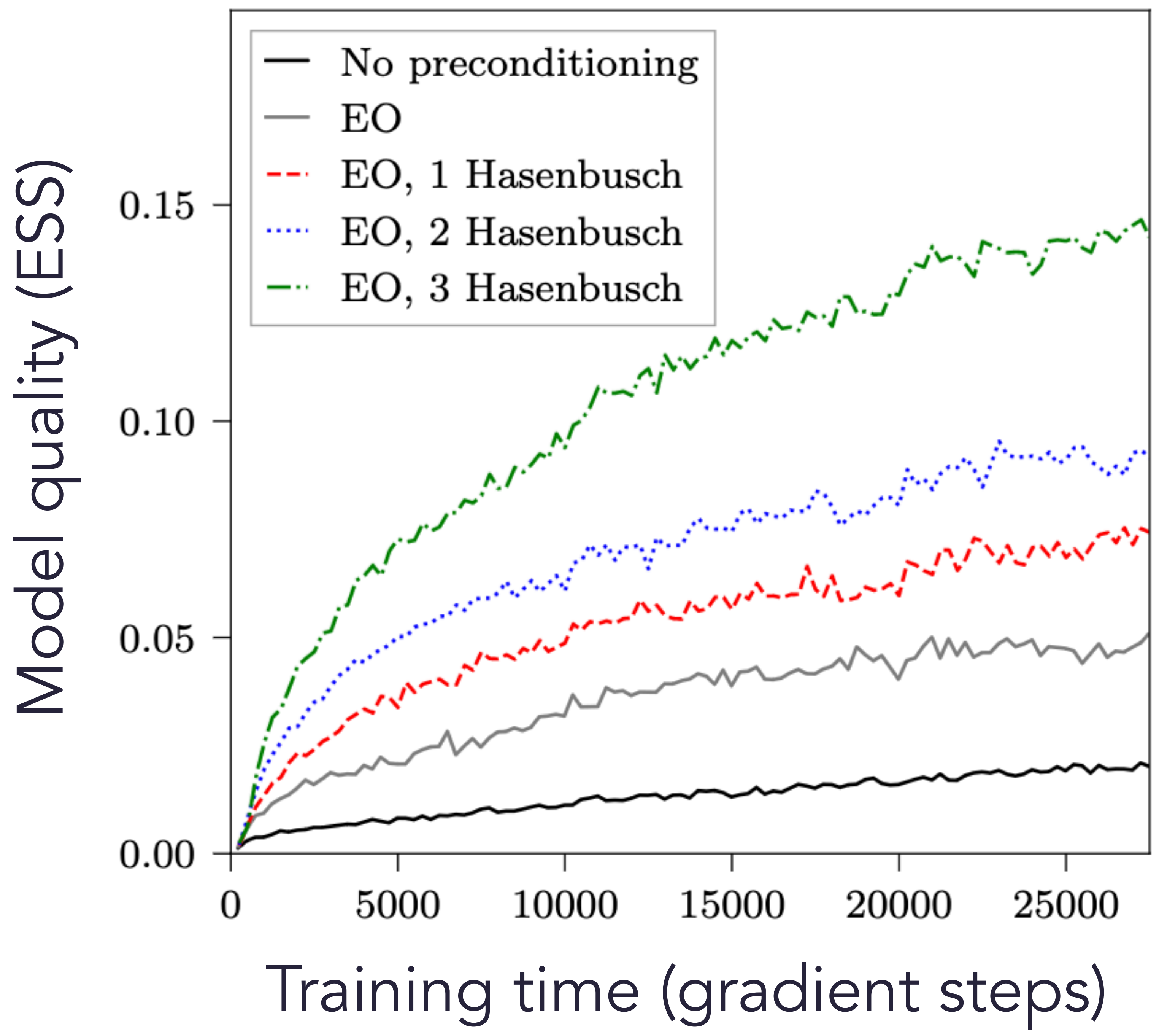
Separate coupling layers for gauge field and PFs can be composed arbitrarily

- **Simplest case:** marginal + conditional model



- **Preconditioning** works equally well for flows

- Modified Metropolis allows averaging away noise in the conditional flow



Hands-on coding

github.com/gkanwar/flow-lectures

- Set up your env and follow along (if you want!)
- Implement flow for ϕ^4 theory
- Train to map from symm to broken phase
- Explore this code further in the exercises

Outlook: important directions

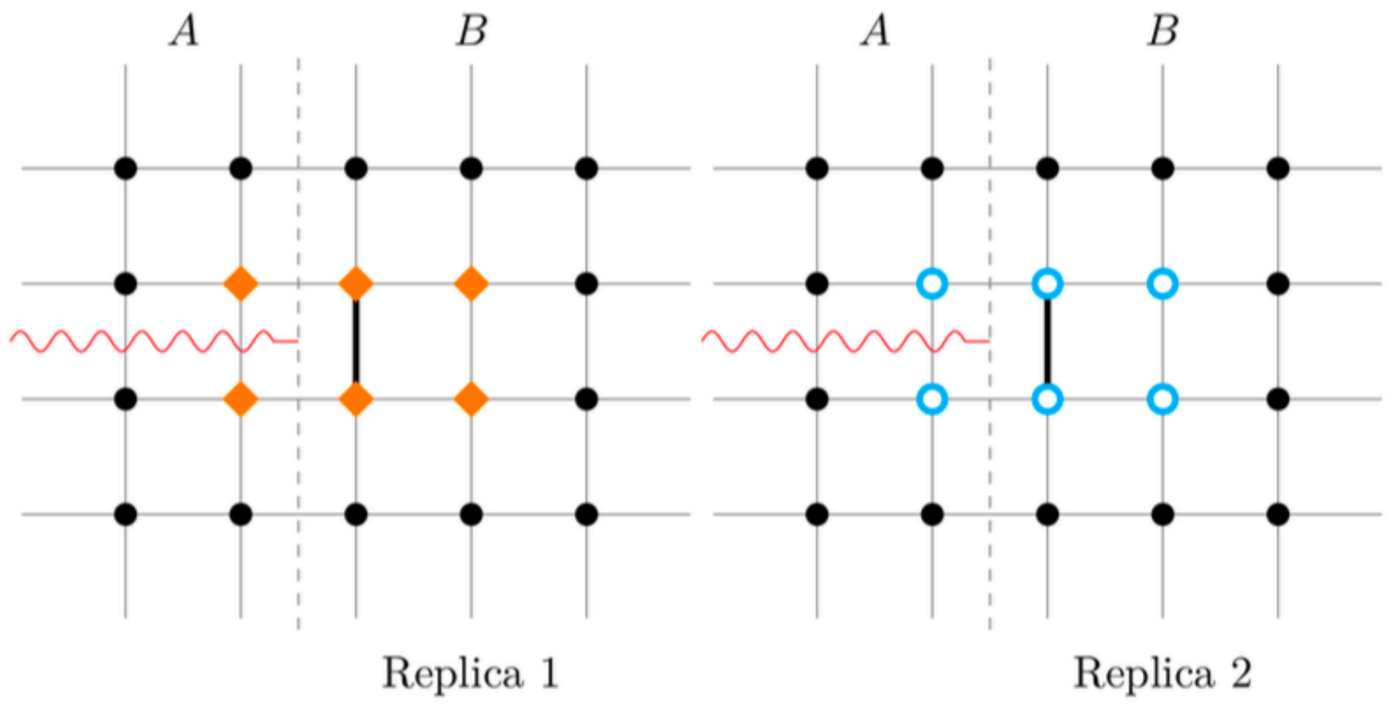
Disclaimer: personal perspective / opinions!

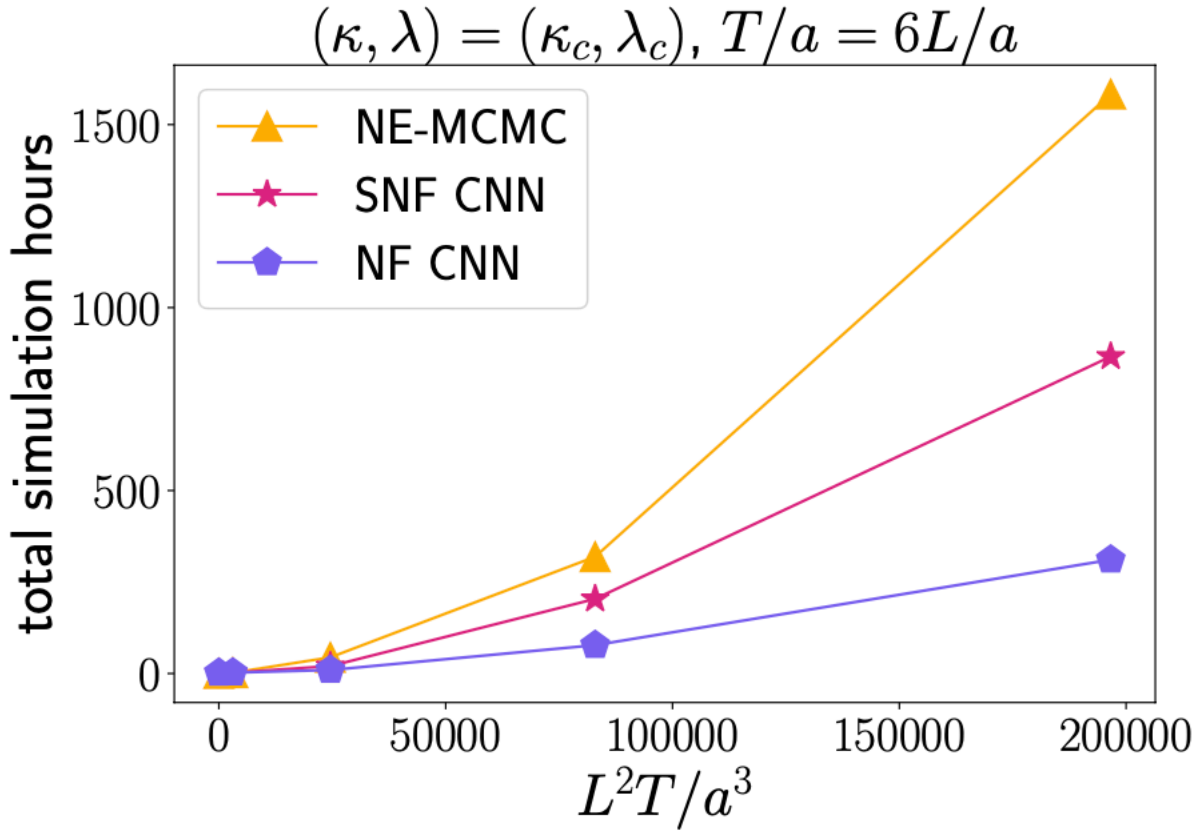
- Finding tailored applications
- Identifying good universal flows
- Connecting to RG
- Measuring continuum scaling

Finding tailored applications (1)

Flow-based Sampling for Entanglement Entropy and the Machine Learning of Defects

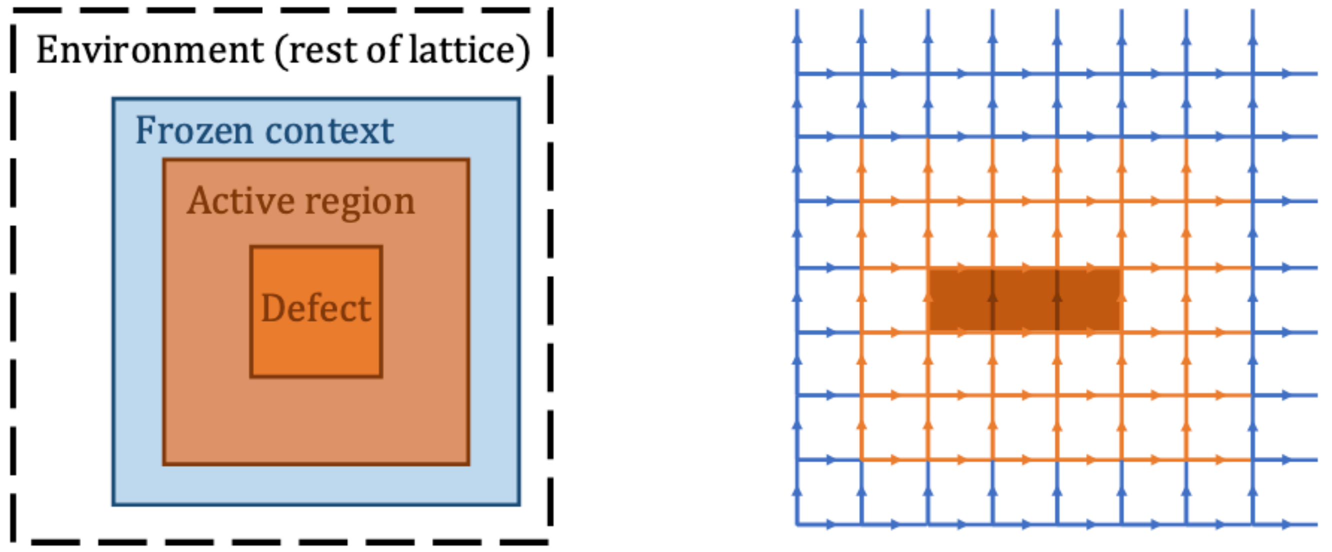
Andrea Bulgarelli[1,*] Elia Cellini[1] Karl Jansen[2,3] Stefan Kühn[3] Alessandro Nada[1] Shinichi Nakajima[4,5,6] Kim A. Nicoli[7,8] and Marco Panero[1]

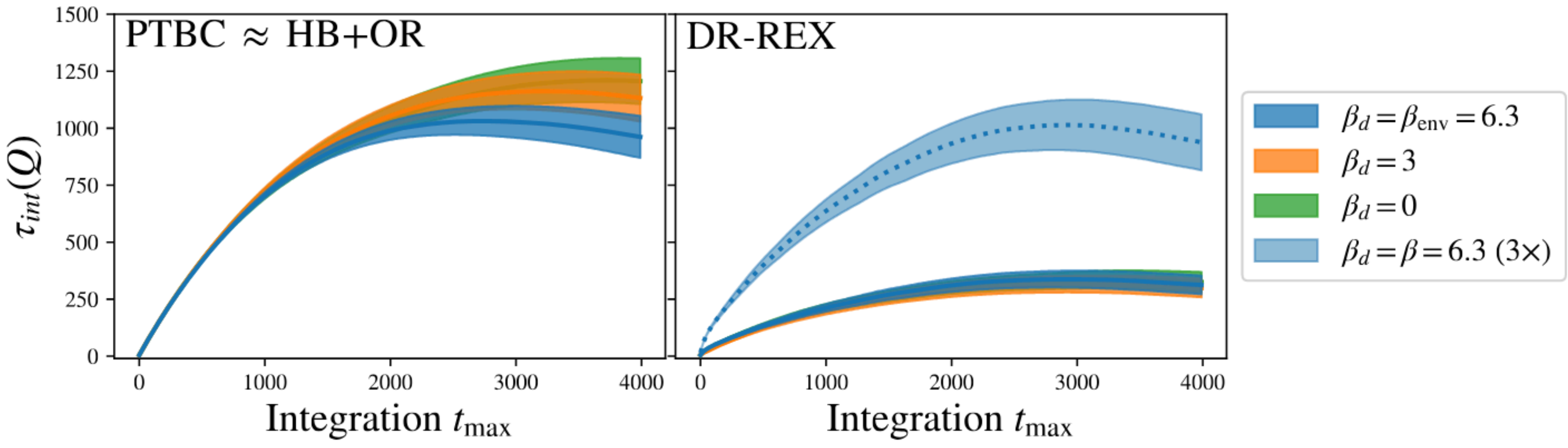




Practical applications of machine-learned flows on gauge fields

Ryan Abbott,[b,c] Michael S. Albergo,[d] Denis Boyda,[b,c] Daniel C. Hackett,[a,b,c,*] Gurtej Kanwar,[e] Fernando Romero-López,[b,c] Phiala E. Shanahan[b,c] and Julian M. Urban[b,c]





Finding tailored applications (2)

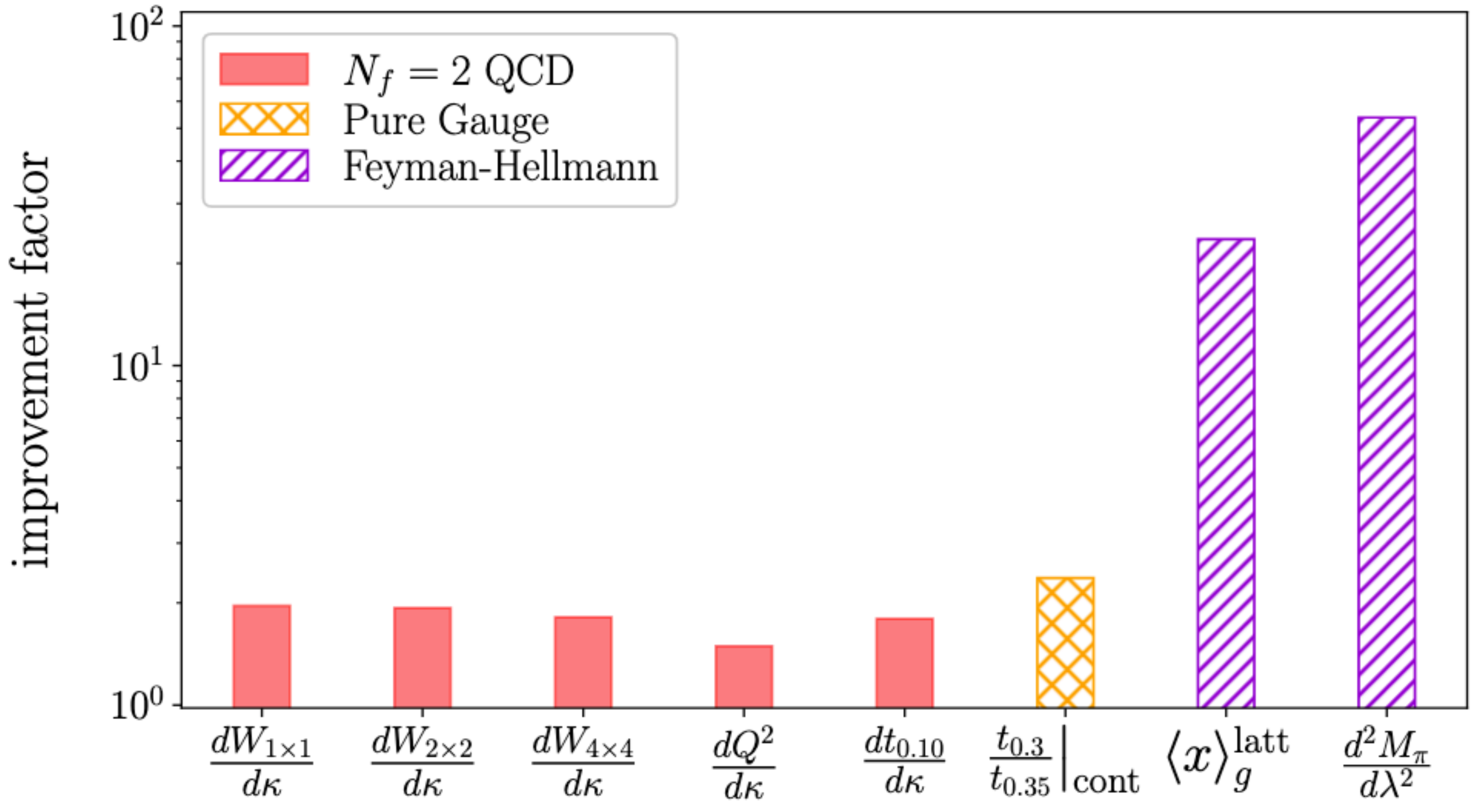
Applications of flow models to the generation of correlated lattice QCD ensembles

Ryan Abbott,[1,2] Aleksandar Botev,[3] Denis Boyda,[1,2] Daniel C. Hackett,[4,1,2] Gurtej Kanwar,[5] Sébastien Racanière,[3] Danilo J. Rezende,[3] Fernando Romero-López,[1,2] Phiala E. Shanahan,[1,2] and Julian M. Urban[1,2]

Use a trained **flow model** to map configurations between the distributions given by α_1 and α_2. Including flow reweighting factors, correlated differences can be calculated as:

$$\langle \mathcal{O}(U) - w(f(U))\, \mathcal{O}(f(U)) \rangle_{\alpha_1} \ , \qquad (7)$$

where $w(f(U)) = p_{\alpha_2}(f(U))/q(f(U))$, such that a perfect flow would remove the reweighting factors entirely.

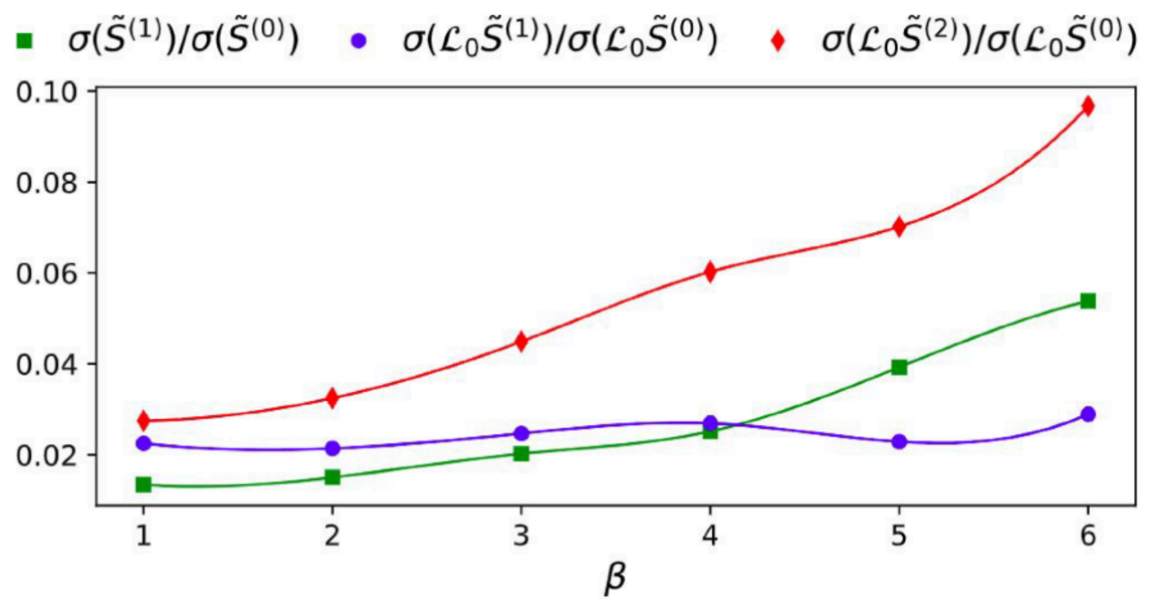


Identifying good universal flows (1)

No clear proof of universality for **discrete flows**

Questions about a Luscher-style loop expansion for **continuous flows**

$$\tilde{S}(U_t, t) = \frac{\beta}{6} \sum_{k=0}^{\infty} \left(\frac{t\beta}{6}\right)^k \tilde{S}^{(k)}(U_t)$$



- $t \leq 1$ ✔ Limit of the integral
- $\beta < 6$ ✗ Region of physical interest at $\beta \gtrsim 6$
- Order of magnitude of $\tilde{S}^{(k)}$??
- **Our conclusion:** <u>it does not converge!</u>

$$\mathcal{L}_0 \tilde{S}^{(2)} = \left(\partial W_0, \partial \tilde{S}^{(1)}\right)$$

Simone Bacchio, ML4PhysChem, Bonn (2024)

Identifying good universal flows (2)

So what can we do about it?

Option 1: Be practical

- Continue to find effective ad-hoc extensions of our architectures
- Measure model quality vs. parameter complexity

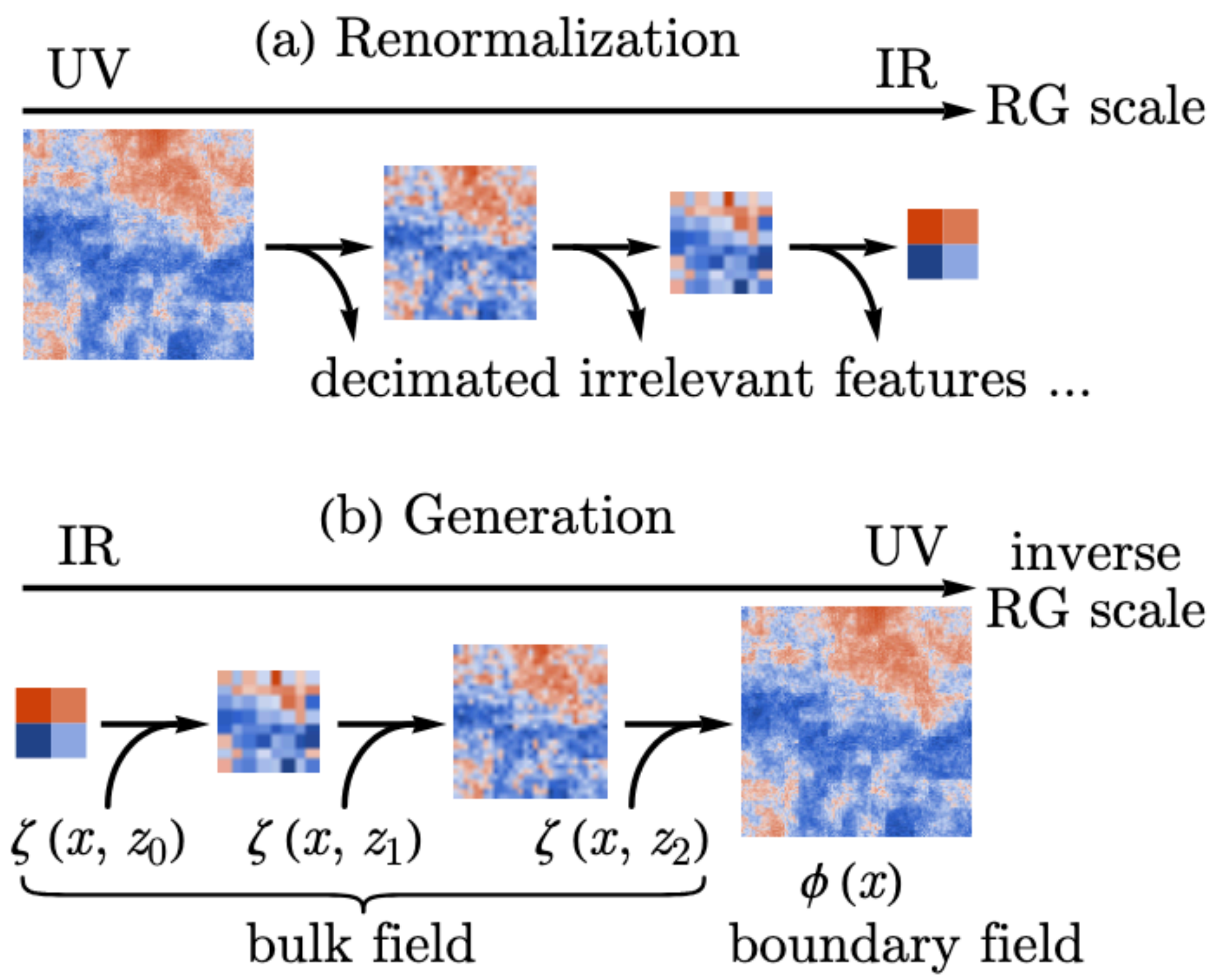
Option 2: Be analytical

- Attempt to construct coupling layers/potentials with universality **guarantees**
- Measure model quality vs. parameter complexity

Connecting to RG (1)

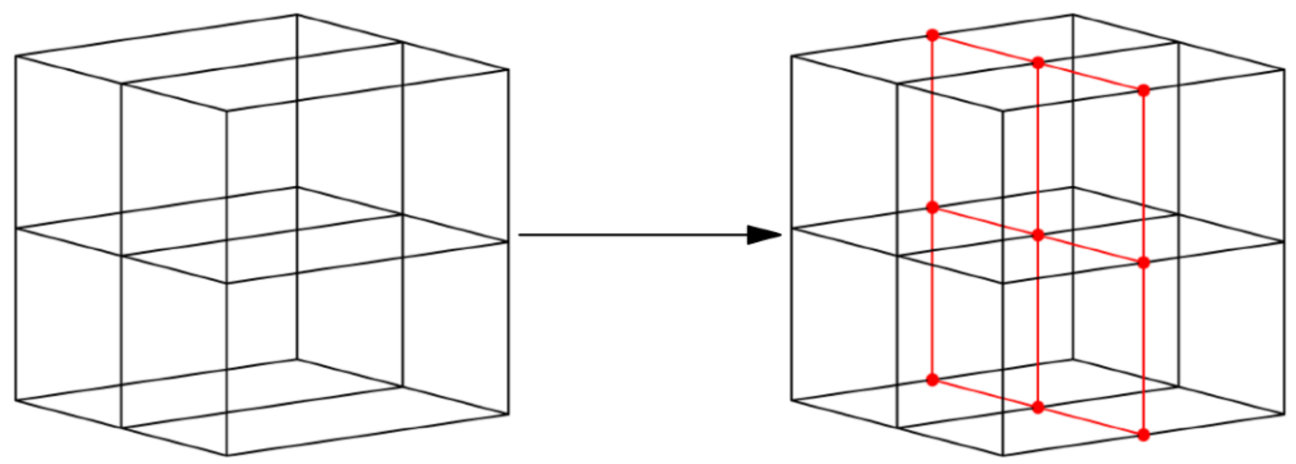
Machine Learning Holographic Mapping by Neural Network Renormalization Group

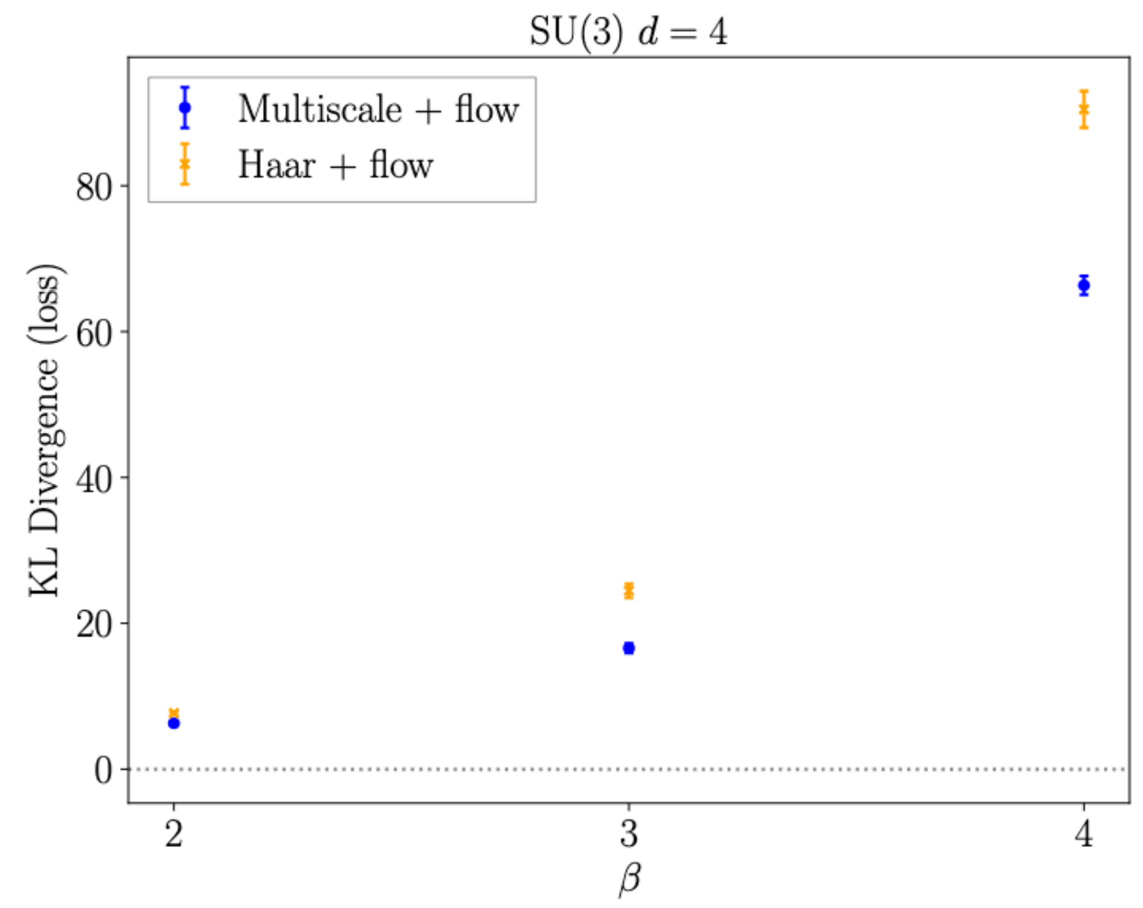
Hong-Ye Hu,[1] Shuo-Hui Li,[2,3] Lei Wang,[2,4,5] and Yi-Zhuang You[1,*]



Multiscale Normalizing Flows for Gauge Theories

Ryan Abbott,[a,b,*] Michael S. Albergo,[c] Denis Boyda,[a,b] Daniel C. Hackett,[f,a,b] Gurtej Kanwar,[a,b,d] Fernando Romero-López,[a,b] Phiala E. Shanahan[a,b] and Julian M. Urban[a,b,e]

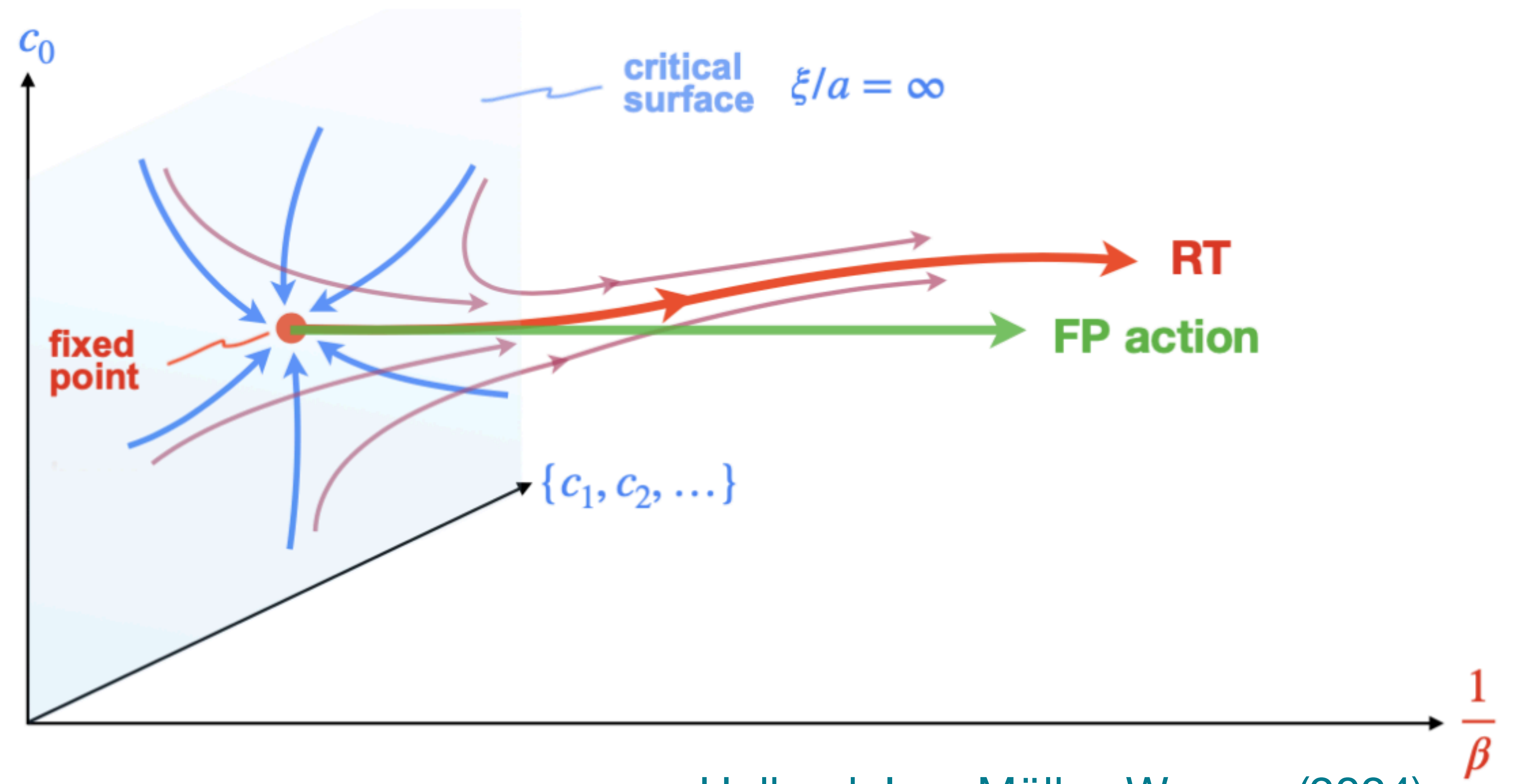




Connecting to RG (2)

Lift restriction to reproduce specific discretized actions

- Only need a sequence of theories on a line of constant physics to the continuum
- Theories related by blocking RG transformations would be a natural choice (related to search for "quantum perfect" actions)

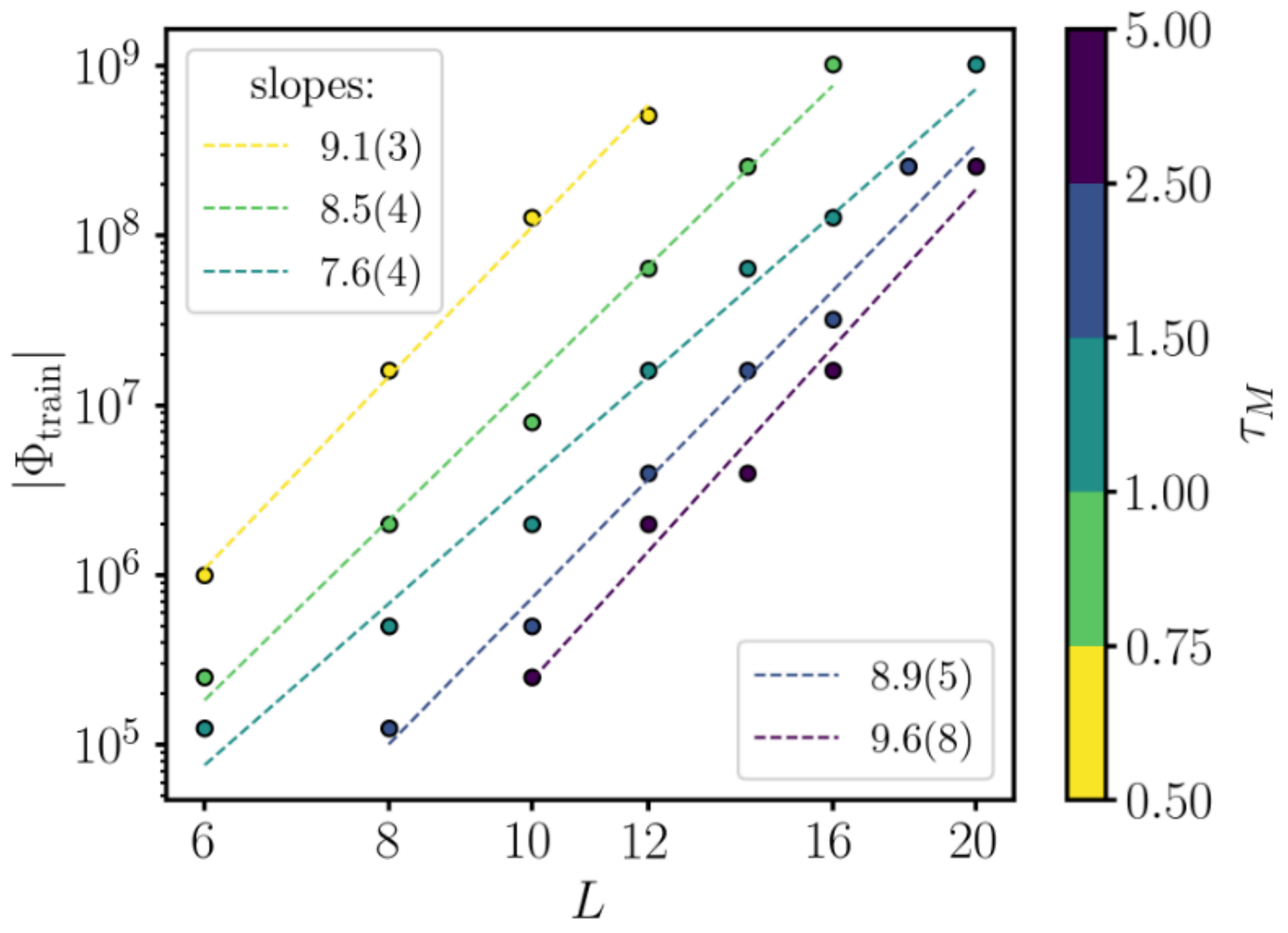


Holland, Ipp, Müller, Wenger (2024)

Measuring continuum scaling (1)

Efficient Modelling of Trivializing Maps for Lattice ϕ^4 Theory Using Normalizing Flows: A First Look at Scalability

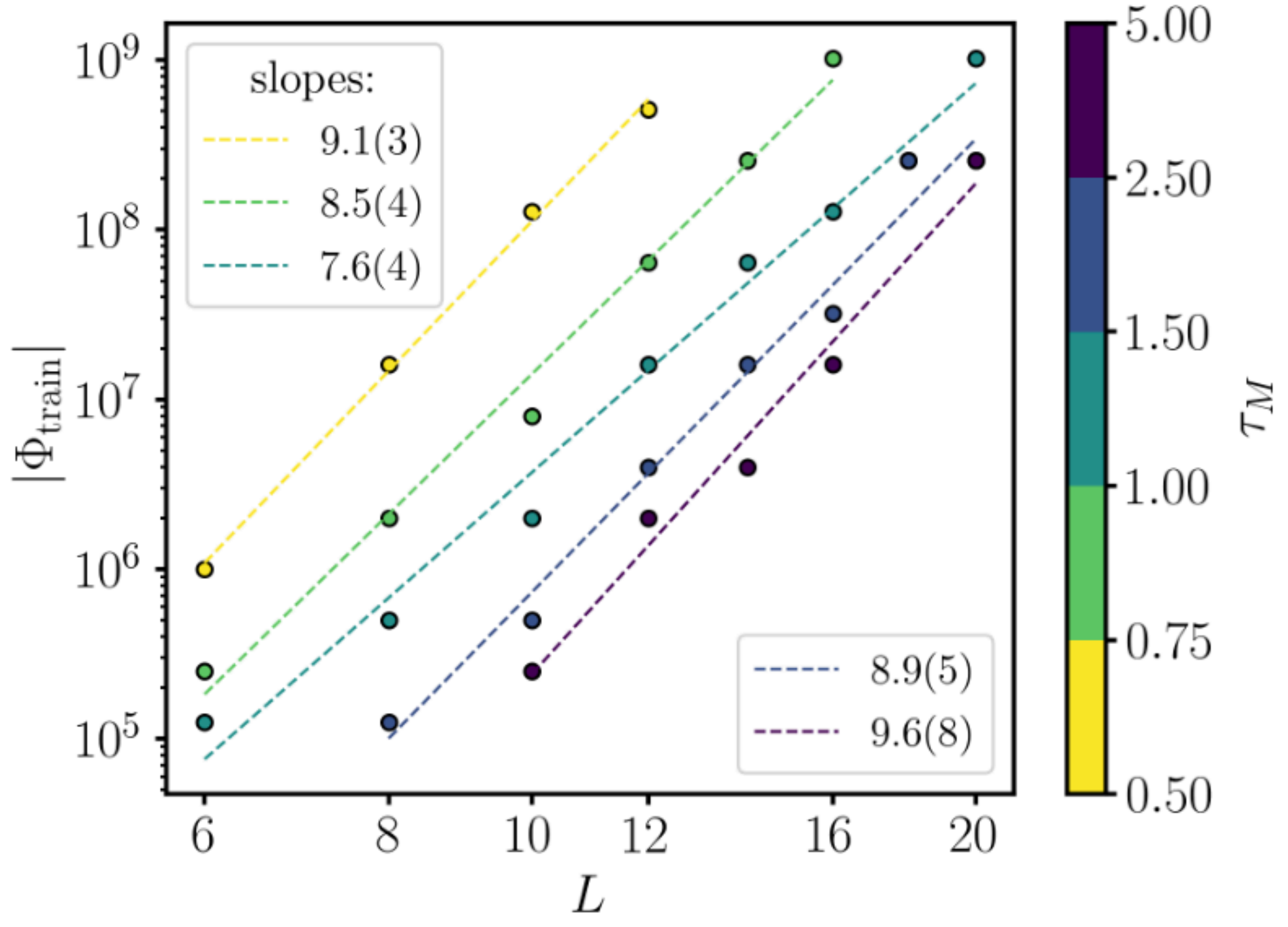
Luigi Del Debbio, Joe Marsh Rossney,* and Michael Wilson



Measuring continuum scaling (1)

Efficient Modelling of Trivializing Maps for Lattice ϕ^4 Theory Using Normalizing Flows: A First Look at Scalability

Luigi Del Debbio, Joe Marsh Rossney,* and Michael Wilson



A First Look at

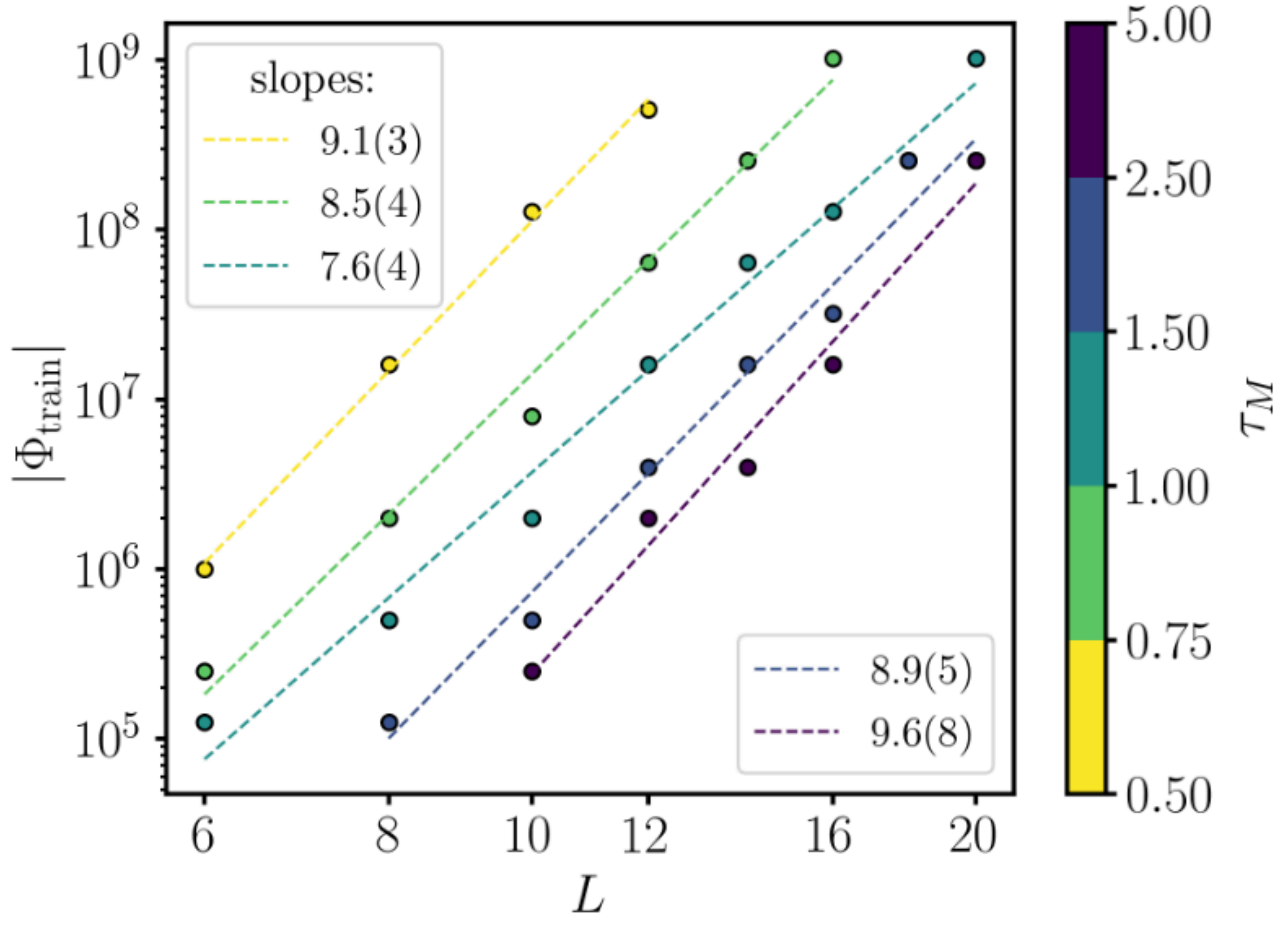
When investigating the scaling of training costs (Section VI D), we used normalizing flows that are a specific hybrid of affine coupling layers and rational quadratic splines, with the parameters of the transformations generated by fully-connected feed-forward neural networks containing a single hidden layer of size $H = |\Lambda|$. In Section VI C we report on the observations that led us to converge on this particular design.

Conclusion: This question needs to be revisited *in concert with* the previous points.

Measuring continuum scaling (1)

Efficient Modelling of Trivializing Maps for Lattice ϕ^4 Theory Using Normalizing Flows: A First Look at Scalability

Luigi Del Debbio, Joe Marsh Rossney,* and Michael Wilson



So should we despair?

A First Look at

When investigating the scaling of training costs (Section VI D), we used normalizing flows that are a specific hybrid of affine coupling layers and rational quadratic splines, with the parameters of the transformations generated by fully-connected feed-forward neural networks containing a single hidden layer of size $H = |\Lambda|$. In Section VI C we report on the observations that led us to converge on this particular design.

Conclusion: This question needs to be revisited *in concert with* the previous points.

Measuring continuum scaling (2)

Several useful target theories for this kind of study:

- Scalar field theory with quartic coupling
- O(N) non-linear sigma models
- **CP(N-1) models in 2D**
  - Sweet spot of interesting features, yet low cost!
- U(1) lattice gauge theory in 2+1D
- SU(N) lattice gauge theory in 3+1D

Extra slides

What about volume scaling?

Abbott, et al. 2211.07541

Fixed models will always* scale exponentially poorly with the **physical volume**.

* in a direct sampling scheme

- Expect variance of log reweighting factors to scale as $(L/\xi)^d$
  Scaling relation $\mathrm{ESS}(V) = \mathrm{ESS}(V_0)^{V/V_0}$, where $V_0 \sim \xi^d$
- This says nothing about scaling towards the continuum limit!

We should be thinking about targeting boxes of size $\approx \xi^d$.

- For larger volumes, hybrid/multilevel sampling schemes should be used

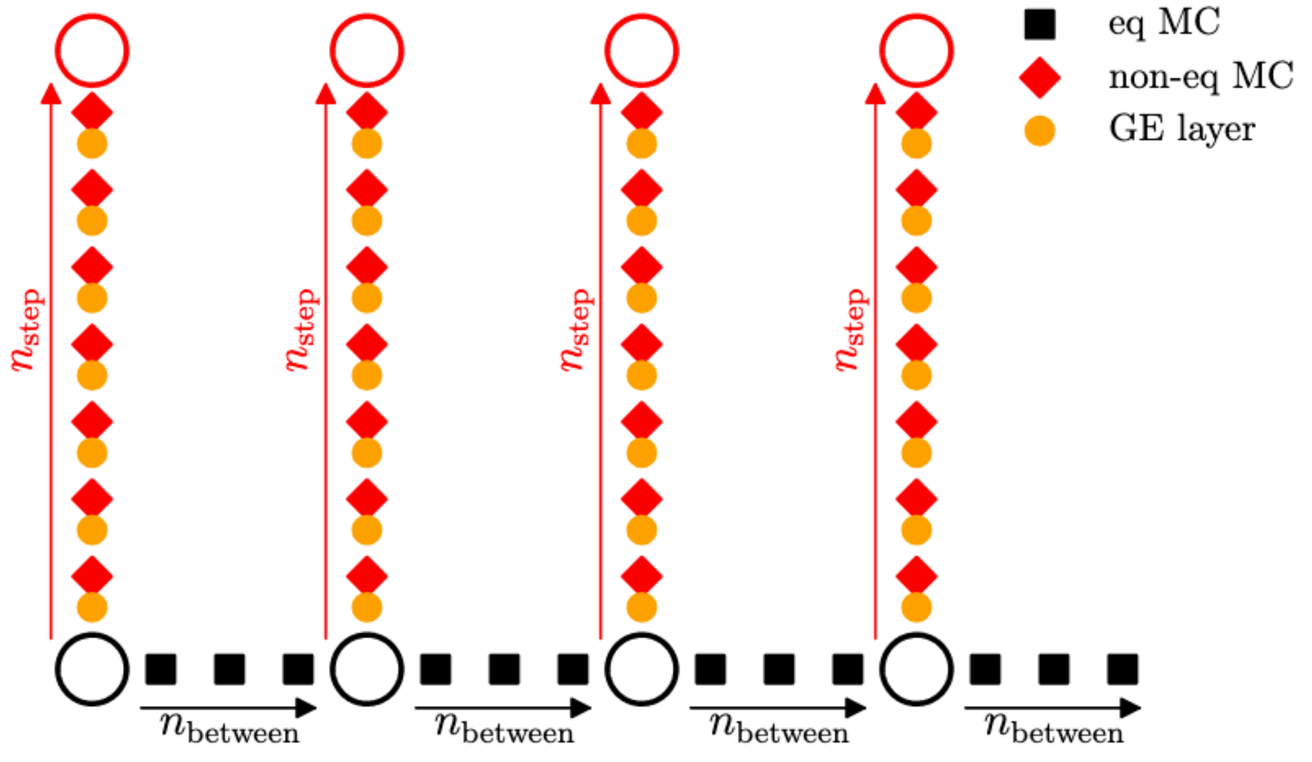
Identifying good universal flows (3)

Methods exist to arbitrarily improve sample quality even with limited flows

Stochastic normalizing flows as non-equilibrium transformations

Michele Caselle[1,2,*] Elia Cellini[1,2,†] Alessandro Nada[1,‡] and Marco Panero[1,2,§]

$$U_0 \xrightarrow{g_1} g_1(U_0) \xrightarrow{P_{c(1)}} U_1 \xrightarrow{g_2} g_2(U_1) \xrightarrow{P_{c(2)}} U_2 \xrightarrow{g_3} \ldots \xrightarrow{P_{c(n_{\rm step})}} U_{n_{\rm step}}$$

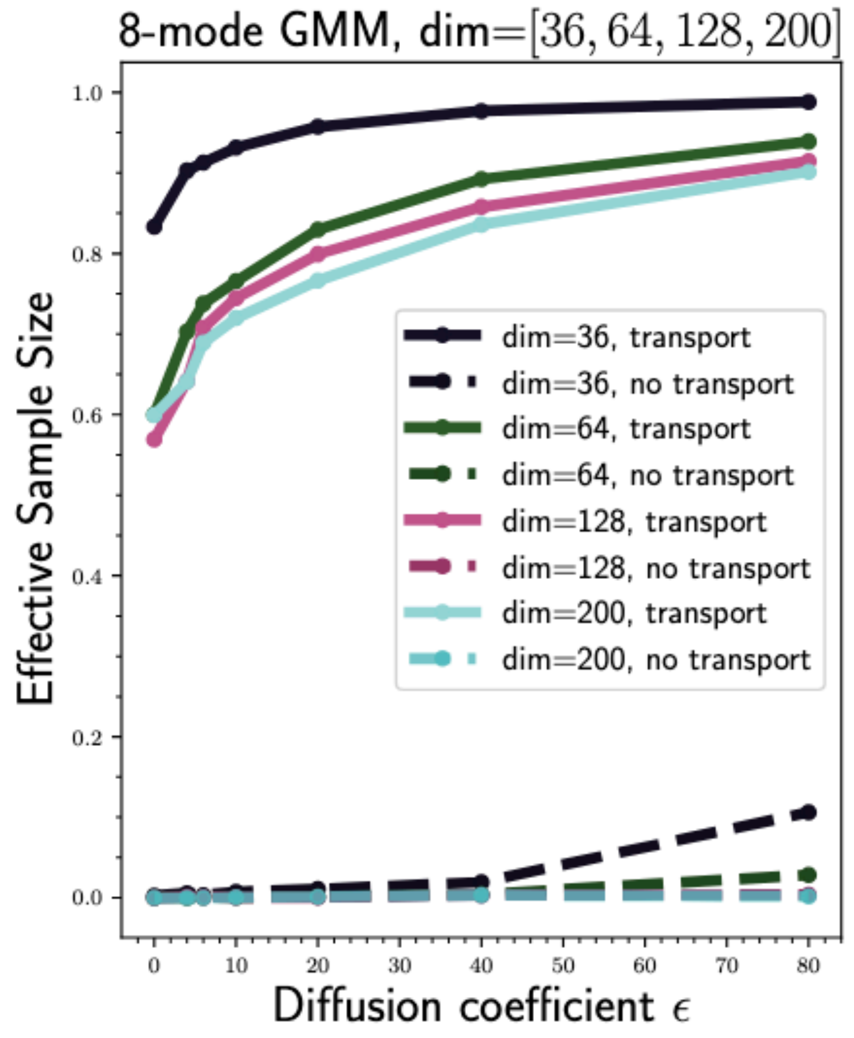


NETS: A NON-EQUILIBRIUM TRANSPORT SAMPLER

Michael S. Albergo
Society of Fellows
Harvard University
Cambridge, MA 02138, USA
malbergo@fas.harvard.edu

Eric Vanden-Eijnden
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012, USA
eve2@cims.nyu.edu

- The resultant samplers can be adapted *after training* by tuning the integration time-step as well as the diffusivity to improve the sample quality, which we demonstrate on high-dimensional numerical experiments below.

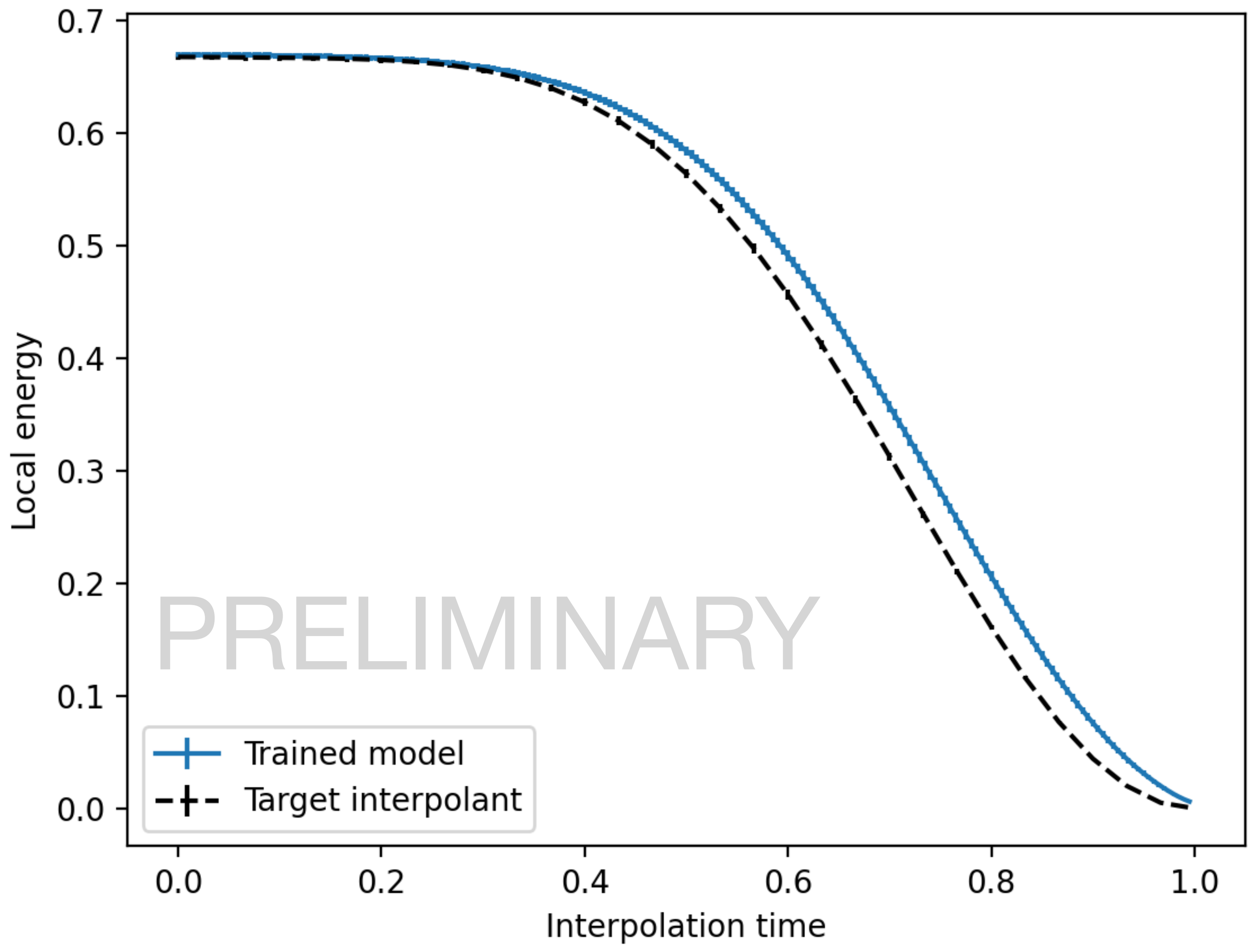


Investigating good universal flows (4)

Is it best to use flows for **crude measure transport** (corrected by other methods) or do we gain by **improving flow quality**?

Currently looking to answer this question in the context of CP^{N-1} models

- Training flows to **prescribed path** of distributions is useful to isolate this question
- Possible paths:
  1. Annealing path (triv map)
  2. Path defined by diffusion
  3. Stochastic interpolants
  4. Others?

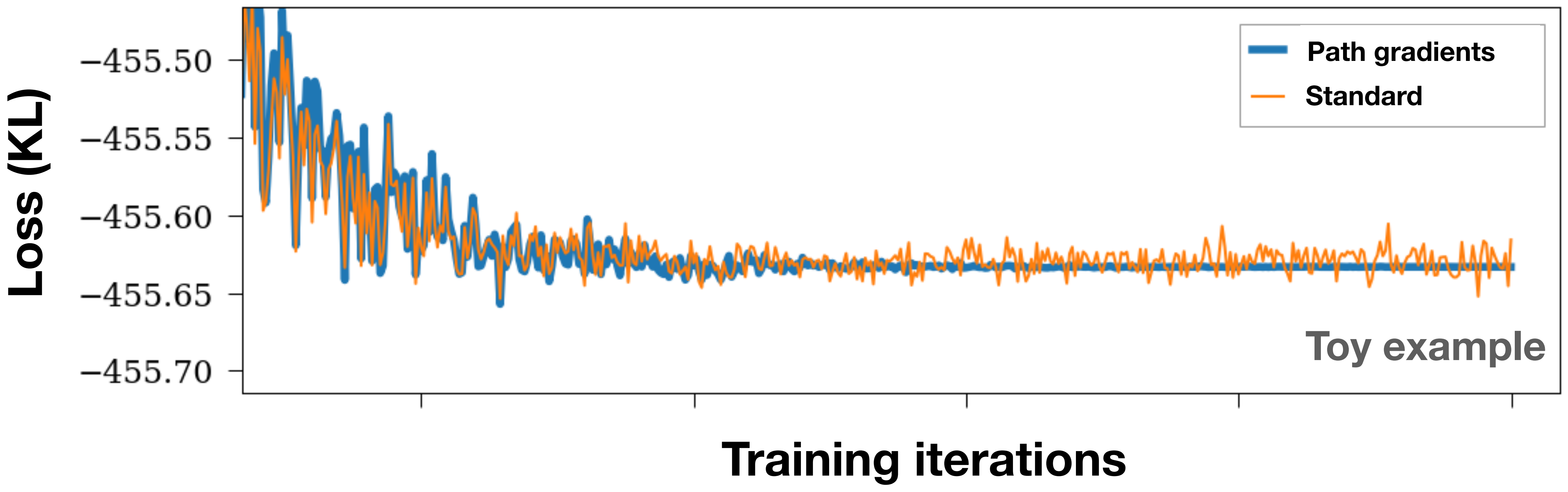


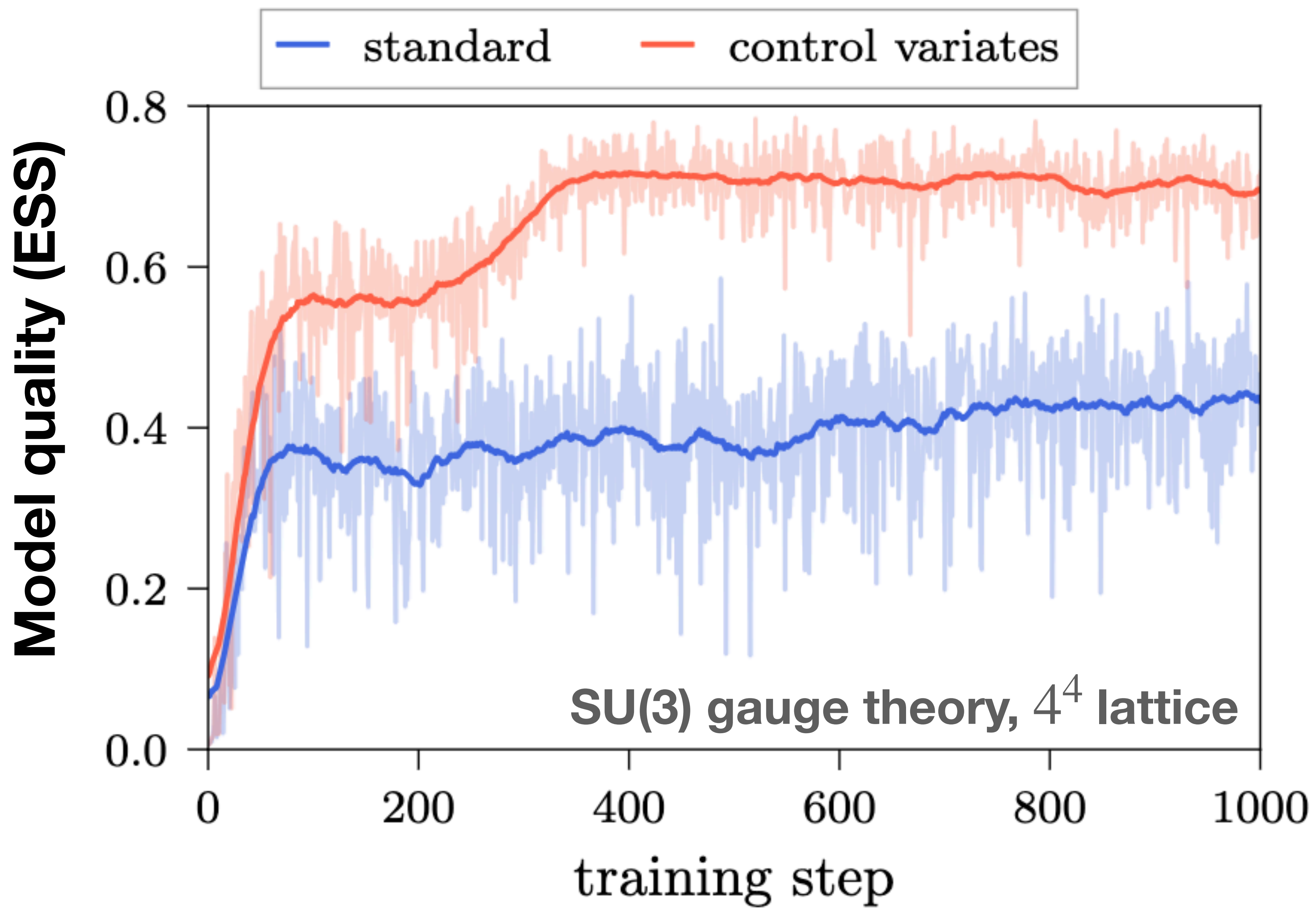
Better training procedures

- Minimize gradient noise with control variates or path gradients

  Ridder, Botev (2017) “Variance Reduction” doi:10.1002/9781118445112.stat07975
  Vaitl, Nicoli, Nakajima, Kessel (2022) 2207.08219
  Białas, Korcyl, Stebel (2022) 2202.01314

- Particularly helpful with convergence near minimum of loss function





Abbott, et al. (2023) 2305.02402