# Multiscale Models for Gauge Theories

Ryan Abbott

MIT

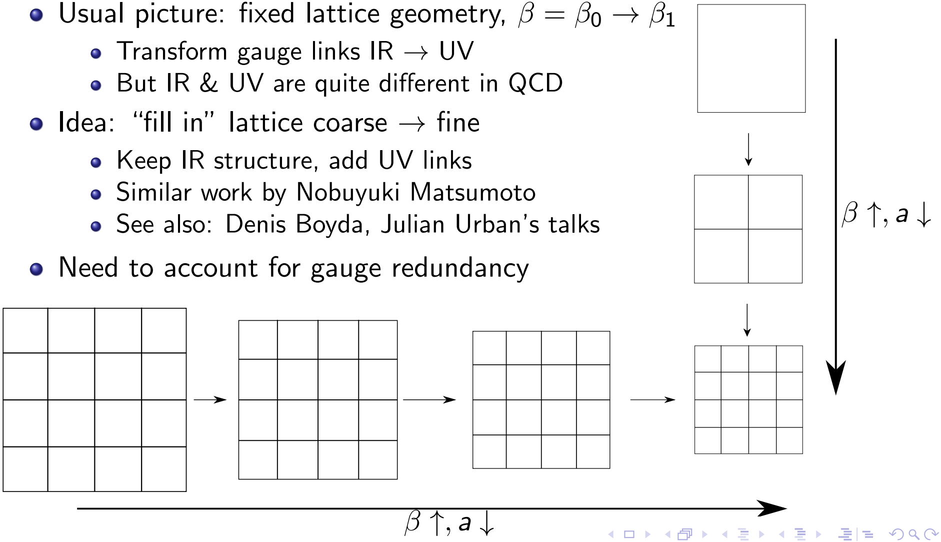
July 31, 2023

# Collaborators

- Phiala Shanahan
- Gurtej Kanwar
  - Mon. 10:00am – "Flow based sampling for lattice field theories"
- Julian Urban
  - Thursday 2:10pm – "Constructing approximate semi-analytic and machine-learned trivializing maps for lattice gauge theory"
- Denis Boyda
  - Thursday 2:50pm – "Enhanced expressivity in machine learning: application of normalizing flows in lattice QCD simulations"
- Dan Hackett
  - Thursday 3:10pm – "Practical applications of machine-learned flows on gauge fields"
- Fernando Romero-López
- + collaborators at Deep Mind

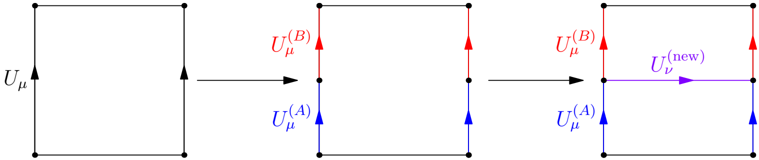
# Scale Separation and Normalizing Flows

- Usual picture: fixed lattice geometry, $\beta = \beta_0 \to \beta_1$
  - Transform gauge links IR $\to$ UV
  - But IR & UV are quite different in QCD
- Idea: "fill in" lattice coarse $\to$ fine
  - Keep IR structure, add UV links
  - Similar work by Nobuyuki Matsumoto
  - See also: Denis Boyda, Julian Urban's talks
- Need to account for gauge redundancy



# Doubling Layers (2d)

- Double lattice extent along $\hat{\mu}$ direction
- Two steps:
  - Splitting $U_\mu = U_\mu^{(A)} U_\mu^{(B)}$ (sample $U_\mu^{(B)} \sim$ Haar)
  - Sample new links $U_\nu^{(\text{new})}$ (from heatbath, flow, etc.)
- Repeat along different directions to form full fine lattice

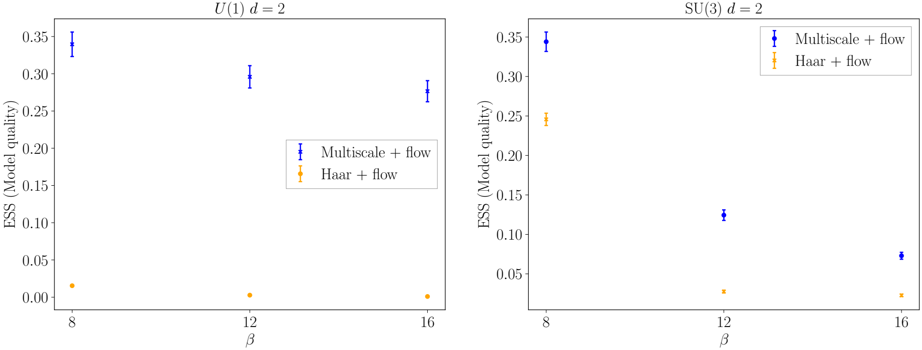


# Results: 2d

- Multiscale model = Haar $2^2$ + doubling layers + fine-lattice flow
- Compare against Haar $8^2$ + fine-lattice flow

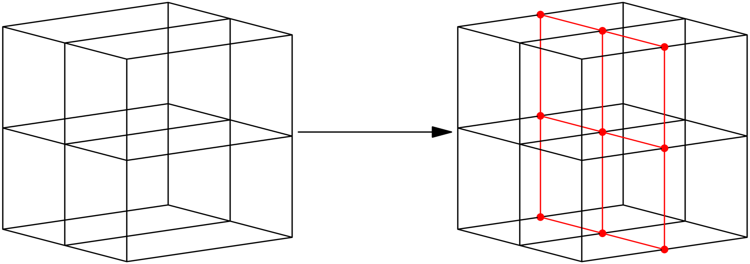
Small flow for comparison purposes

- ESS $\sim$ (independent) acceptance rate



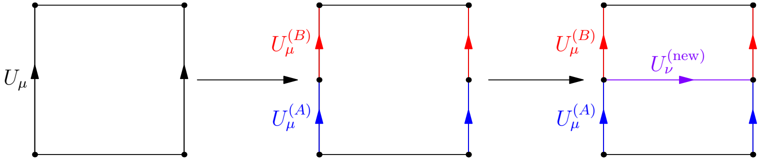
# Higher-dimensional doubling layers

- Need to generate lower-dimensional *slice*
- Slice contains both UV and IR links
  - $\Longrightarrow$ use (recursive) multiscale model



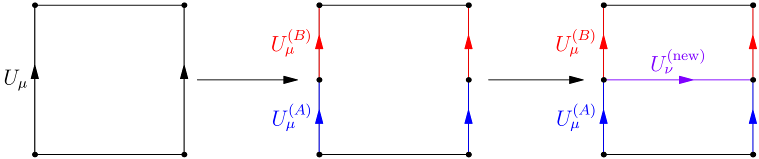
# Doubling Layers (General Dimension)

- Double lattice extent along $\hat{\mu}$ direction
- Two steps:
  - Splitting $U_\mu = U_\mu^{(A)} U_\mu^{(B)}$ (sample $U_\mu^{(B)} \sim$ Haar)
  - Sample new links $U_\nu^{(\text{new})}$ (from heatbath, flow, etc.)
- Repeat along different directions to form full fine lattice
- Also keep track of staples from higher dimensions



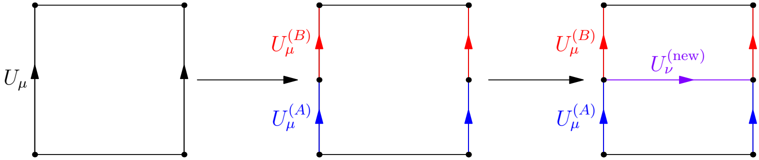
# Doubling Layers (General Dimension)

- Double lattice extent along $\hat{\mu}$ direction
- Two steps:
  - Splitting $U_\mu = U_\mu^{(A)} U_\mu^{(B)}$ (sample $U_\mu^{(B)} \sim$ Haar)
  - Sample new links $U_\nu^{(\text{new})}$ (from heatbath, flow, etc.) Multiscale model
- Repeat along different directions to form full fine lattice
- Also keep track of staples from higher dimensions



# Doubling Layers (General Dimension)

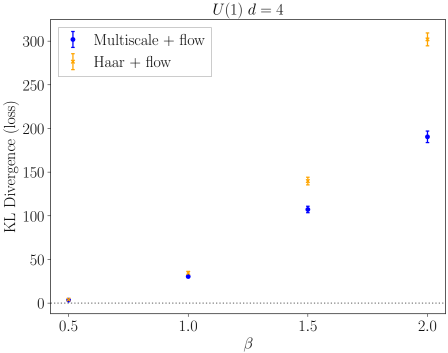
- Double lattice extent along $\hat{\mu}$ direction
- Two steps:
  - Splitting $U_\mu = U_\mu^{(A)} U_\mu^{(B)}$ (sample $U_\mu^{(B)} \sim$ Haar) flow, heatbath, etc.
  - Sample new links $U_\nu^{(\text{new})}$ (from heatbath, flow, etc.) Multiscale model
- Repeat along different directions to form full fine lattice
- Also keep track of staples from higher dimensions

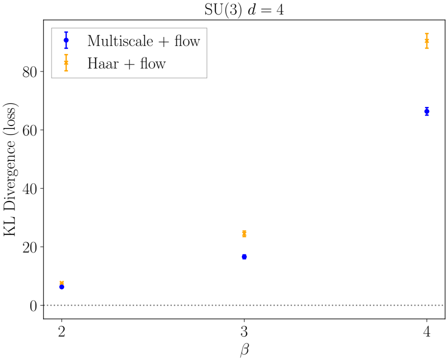


# Results: SU(3) and $U(1)$ in 4d

- Multiscale model = Haar $2^4$ + doubling layers + fine-lattice flow
  - (+ additional layer of recursion)
- Compare against Haar $4^4$ + fine-lattice flow

Small flow for comparison purposes



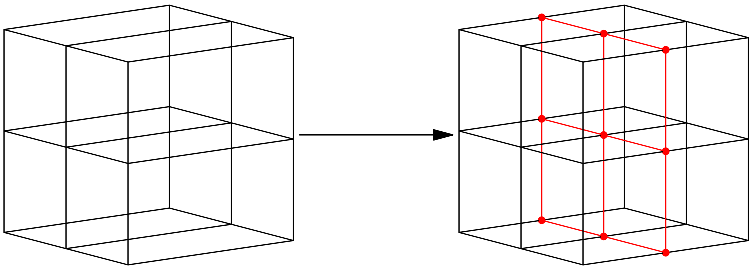

- Note: normalized with rough estimate of $\log Z$ at each $\beta$

# Future work

- Model improvements
  - More equivariant information/correlations
  - Better link-level conditional flows
  - Fermions (multigrid?)
- Combine with approaches beyond direct sampling
  - e.g. CRAFT, Parallel Tempering, Feynman-Hellman, . . .
    - see Dan Hackett's talk (Thursday 3:10pm)
  - HMC on trivialized distribution
    - see previous talk by Nobuyuki Matsumoto

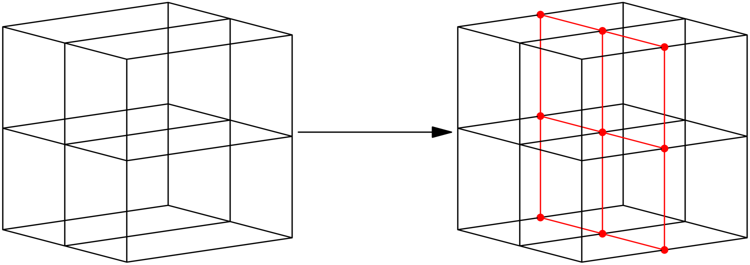
# Conclusions

- Scale separation enables new types of normalizing-flow models
- Implemented in arbitrary dimension $U(1)$ and SU(3)
- Promising early results, but more work needed



# Conclusions

- Scale separation enables new types of normalizing-flow models
- Implemented in arbitrary dimension $U(1)$ and SU(3)
- Promising early results, but more work needed
- Thanks! Questions?



# Backup

# Staple-Conditional Flows

- Fundamental building block for multiscale models
- Inputs: gauge link $U \in G$, "staples" $S_1, \ldots, S_n \in G$
  - Gauge group $G = U(1)$ or $\mathrm{SU}(N)$
  - Staples $\sim$ inverse gauge links
- Start from either Haar-uniform or other tractable distribution
  - e.g. heatbath $U \sim e^{-\sum_i \beta_i \mathrm{Re}\,\mathrm{tr}(US_i)}$ for $G = U(1)$
- Build flow transforms from previous components
  - Spectral flows [Boyda et al., 2008.05456]
  - Residual flows [Abbott et al., 2305.02402]
  - Continuous flows [Bacchio et al. 2212.08469]

# Comments on Scaling

- Reference: [Abbott et al., 2211.07541]
- Scaling depends strongly every aspect of the model
  - E.g. use of flow, architecture choices, training choices
  - Makes extrapolating beyond any particular choice difficult

## Use of Flow

- Direct Sampling (Independence Metropolis)
- HMC on trivialized distribution [Lüscher 0907.5491]
- Generalize proposal distribution [Foreman et al., 2112.01582]
- Subdomain updates [Finkenrath, 2201.02216]
- Stochastic Normalizing Flows [Wu et al. 2002.0670]
- CRAFT [Matthews et al. 2201.13117]

# Comments on Scaling

- Reference: [Abbott et al., 2211.07541]
- Scaling depends strongly every aspect of the model
  - E.g. use of flow, architecture choices, training choices
  - Makes extrapolating beyond any particular choice difficult

## Architecture Choices

- Choice of coupling layers (spectral, residual, continuous)
- Choice of Neural networks (CNN, fully-connected, gauge-equivariant)
  - Gauge-equivariant networks [Favoni et al., 2012.12901]
- Choice of invariant context passed to networks
- Size of model (# layers, NN sizes)

# Comments on Scaling

- Reference: [Abbott et al., 2211.07541]
- Scaling depends strongly every aspect of the model
    - E.g. use of flow, architecture choices, training choices
    - Makes extrapolating beyond any particular choice difficult

## Training Choices

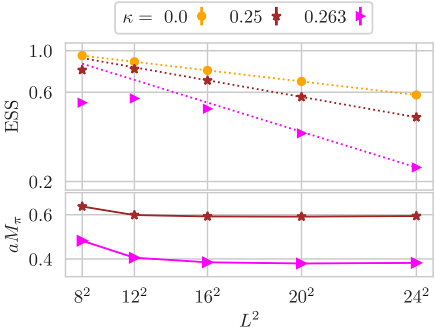
- Optimizer (Adam, SGD, higher-order optimizers)
- Choice of Loss (reverse/forward KL, MSE, ...)
- Computation of gradients (path gradients/control variates)
- Hyperparameter choices (batch size, learning rate)
    - Hyperparameter scheduling
- Volume chosen for training

# Comments on Scaling - Exponential Volume Scaling

- For $L/\xi \gg 1$, $\xi =$ correlation length, direct volume transfer

$$ESS(V) = ESS(V_0)^{V/V_0}$$

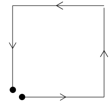
- Prevents *direct sampling* in thermodynamic limit $L/\xi \to \infty$
  - Does not apply to continuum limit $L/\xi \sim m_\pi L$ fixed, $\xi/a \to \infty$
  - Typically $4 \lesssim m_\pi L \lesssim 10 \implies$ no in principle issue



# Spectral Flows

[Boyda et al., 2008.05456]

- Transform untraced plaquette $P_{\mu\nu}$
- Under gauge transformation $\Omega(x) \in \mathrm{SU}(N)$




$$(\Omega \cdot P)_{\mu\nu}(x) = \Omega(x) P_{\mu\nu}(x) \Omega(x)^\dagger$$

- Given $h : \mathrm{SU}(N) \to \mathrm{SU}(N)$, transform $U_\mu$ so $P_{\mu\nu} \mapsto h(P_{\mu\nu})$

$$f(U_\mu) = h(P_{\mu\nu}) P^\dagger_{\mu\nu} U_\mu$$

- Gauge equivariance $\iff$ *conjugation equivariance*:

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

- Achieve by transforming eigenvalues for fixed eigenvectors

# Residual Flows

- Inspired by Lüscher's trivializing map [Lüscher 0907.5491]
- Transform active links via

Lie-algebra-valued derivative

$$U_\mu(x) \mapsto e^{i\epsilon\partial_{x,\mu}\phi(U)} U_\mu(x)$$

- Gauge-invariant "potential" $\phi(U)$
  - Example: $\phi(U) \propto S_{\text{Wilson}}(U) \implies$ Wilson flow/stout smearing
  - More complex:

$$\phi(U) = \sum_x \sum_{\mu \neq \nu} c_{\mu\nu}(x; U_{\text{frozen}}) \mathrm{Re}\,\mathrm{Tr}(P_{\mu\nu})$$

- Small but finite $\epsilon$ for invertibility ($\epsilon \lesssim 1/8$)

# Spectral Flows

Goal: $h(\Omega X \Omega^\dagger) = \Omega h(X) \Omega^\dagger$

- Conjugation invariant data $\Leftrightarrow$ eigenvalues
- Diagonalize $X \in \mathrm{SU}(N)$ via eigenbasis $V$:

$$X = V \begin{pmatrix} e^{i\theta_1} & & \\ & \ddots & \\ & & e^{i\theta_N} \end{pmatrix} V^\dagger \mapsto V \begin{pmatrix} e^{i\theta_1'} & & \\ & \ddots & \\ & & e^{i\theta_N'} \end{pmatrix} V^\dagger$$

- Define $h : \mathrm{SU}(N) \to \mathrm{SU}(N)$ by action on $\{\theta_1, \ldots, \theta_N\}$
  - Need to be careful about order $\Rightarrow$ choose canonical order
  - Note: $\theta_k$ not independent, $\prod_k e^{i\theta_k} = \det X = 1 \Rightarrow$ remove $\theta_N$

# Training

- Model density $q(\phi)$, target $p(\phi) = \frac{1}{Z} e^{-S(\phi)}$
- Reverse Kullback Leibler (KL) loss $\mathcal{L}$:

$$\begin{aligned}\mathcal{L} &= D_{KL}(q||p) \\ &= \int \mathrm{d}\phi \, q(\phi) \log \frac{q(\phi)}{p(\phi)} \\ &= \mathbb{E}_{\phi \sim q} [\log q(\phi) + S(\phi)] + \log Z\end{aligned}$$

*Model* samples

Constant ($\Rightarrow$ can ignore)

**Key facts**

$D_{KL}(q||p) \geq 0$
$D_{KL}(q||p) = 0 \Leftrightarrow q = p$

# Unbiased sampling

- Independence Metropolis: accept $\phi \to \phi' \sim q(\phi')$ with probability

$$P_{\text{accept}}(\phi \to \phi') = \min\left(1, \frac{p(\phi')}{p(\phi)} \frac{q(\phi)}{q(\phi')}\right)$$

- Hybrid methods
  - Alternate HMC/flow updates
  - HMC on trivialized distribution [Lüscher 0907.5491]
  - Subdomain updates [Finkenrath, 2201.02216]
  - CRAFT/Annealed Importance Sampling [Matthews et al. 2201.13117]
  - . . .