# Generalizable equivariant diffusion models for non-Abelian lattice gauge theory

Diaa Eddin Habibi

Nordic Lattice Meeting,
18th May 2026,
Edinburgh

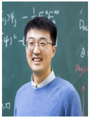
Swansea University
Prifysgol Abertawe

UKRI CENTRE FOR DOCTORAL TRAINING IN ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND ADVANCED COMPUTING

Based on arXiv:2601.19552

理化学研究所 RIKEN

東京大学 THE UNIVERSITY OF TOKYO



Lingxiao Wang

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Qianteng Zhu

Wei Wang

Zheyuan Peng

Renzo Kapust

Jan M. Pawlowski

Yang-Yang Tan

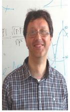
Swansea University
Prifysgol Abertawe

Gert Aarts
Swansea -> Bielefeld

Diaa E. Habibi

Shiyang Chen

Queen Mary University of London

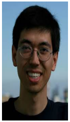


Biagio Lucini

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

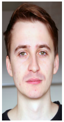
Kai Zhou

TU WIEN



Andreas Ipp

Thomas Richard Ranner



David Müller

# References

Diffusion models applications on LFT in recent years

- 2d scalar fields
  - L. Wang, G. Aarts, K. Zhou, JHEP **05** (2024) 060 [2309.17082],
  - L. Wang, G. Aarts, K. Zhou, ML4PS NeurIPS 2023 [2311.03578]
  - G. Aarts, D.E. Habibi, L. Wang, K. Zhou, Mach. Learn. Sci. Tech. 6 (2025) [2410.21212]
- 2d U(1) and non-Abelian gauge theories
  - Q. Zhu, G. Aarts, W. Wang, K. Zhou, L. Wang, JHEP **03** (2026) [2502.05504] + [2410.19602]
  - O. Vega, J. Komijani, A. El-Khadra, M. Marinkovic, [2510.26081]
  - G. Aarts, D. E. Habibi, A. Ipp, T. Ranner, D. Müller, L. Wang, K. Zhou, Q. Zhu, W. Wang, [2601.19552]
  - H. Alharazin, J. Yu Panteleeva, B.D. Sun [2602.09045]
  - J. Komijani, M. Marinkovic, L. Turgut [2605.06134] (+ in 4d!)

# Diffusion models

DMs are a class of generative AI which learn distributions from data.

- Relies on a stochastic process to learn the *score function*
- Related to Stochastic Quantization Wang, Aarts, Zhou 2309.17082

Forward process (Data to noise) ⟶

⟵ Backward process (Noise to data)

$$\dot{\mathbf{x}}(t) = \frac{1}{2}\underbrace{\mathbf{K}[\mathbf{x}(t), t]}_{\text{drift}} + \underbrace{g(t)\boldsymbol{\eta}(t)}_{\text{noise}}, \quad 0 \le t \le T$$

- Initial conditions determined by $\mathbf{x}_0 \sim p_0(\mathbf{x}_0)$

# Diffusion models

DMs are a class of generative AI which learn distributions from data.

- Relies on a stochastic process to learn the *score function*
- Related to Stochastic Quantization Wang, Aarts, Zhou 2309.17082

Forward process (Data to noise) $\longrightarrow$

$\longleftarrow$ Backward process (Noise to data)

$$\dot{\mathbf{x}}(t) = \frac{1}{2}\mathbf{K}[\mathbf{x}(t), t] - g^2(t)\underbrace{\nabla \log p_t(\mathbf{x})}_{\text{score function}} + g(t)\boldsymbol{\eta}(t),$$

- Initial conditions sampled from a simple prior, eg Gaussian
- Process starts at $t = T$ and flows backwards to $t = 0$

# Score matching

- Processes described by a time-dependent distribution

$$p_t(x) = \int \mathrm{d}x_0 \, p_t(x|x_0) p_0(x_0)$$

- The score is not known; approximate it with a time-dependent model $s_\theta(x, t) \approx \nabla \log p_t(x)$
- Train using a Fisher objective Hyvärinen 2005

$$\mathcal{L}_\theta = \frac{1}{2} \int_0^T \mathrm{d}t \, \mathbb{E}_{p_t(x)} \left[ \sigma^2(t) \| s_\theta(x, t) - \nabla \log p_t(x) \|^2 \right]$$

- Skipping some step, the above relation can be made tractable

$$\begin{aligned} \mathcal{L}_\theta &= \frac{1}{2} \int_0^T \mathrm{d}t \, \mathbb{E}_{p_t(x)} \left[ \| \sigma(t) s_\theta(x, t) + \frac{x_t - x_0}{\sigma(t)} \|^2 \right] \\ &= \frac{1}{2} \int_0^T \mathrm{d}t \, \mathbb{E}_{p_t(x)} \left[ \| \sigma(t) s_\theta(x, t) + \eta(t) \|^2 \right] \end{aligned}$$

# Stochastic Quantization

- In SQ, evolve random field configurations according to Langevin equation in fictitious time $t$ Parisi, Wu 1980

$$\frac{\mathrm{d}\phi(x,\tau)}{\mathrm{d}\tau} = -\frac{\partial S[\phi(x,\tau)]}{\partial \phi(x,\tau)} + \sqrt{2}\,\eta(x,\tau), \quad \eta \sim \mathcal{N}(0,1)$$

- At $\tau \to \infty$, recover probability distribution, $p(\phi) \propto e^{-S[\phi]}$
- Compare with the reverse process of a diffusion model (no drift term)

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = -g^2(t)\nabla_\phi \log p_t(\phi) + g(t)\eta(x,t)$$

- At $t = 0$, the model score learns the drift in SQ,

$$\nabla_\phi \log p_0(\phi) = -\nabla_\phi S[\phi]$$

# Non-Abelian stochastic processes

- For applications to LGT, diffusion process needs to be adapted to group manifold geometry

- Stochastic dynamics known from stochastic quantization

$$U'_{x,\mu} = \exp\left[ i \sum_{a=1} T^a \left( \epsilon K^a_{x,\mu} + \sqrt{2\epsilon}\eta^a_{x,\mu} \right) \right] U_{x,\mu}$$

- Drift term $K^a_{x,\mu}(U) = -D^a_{x,\mu} S_E(U)$ and $\eta^a_{x,\mu} \sim \mathcal{N}(0,1)$

- Process will equilibrate to action, eg, Wilson for $SU(N)$

$$Z = \int DU \, e^{-S_E(U)}, \qquad S_E(U) = \frac{\beta}{N_c} \sum_{x,\mu<\nu} \mathrm{Re}\,\mathrm{Tr}\left[\mathbb{I} - P_{\mu\nu}\right]$$

# Non-Abelian diffusion models

- Define group-preserving forward process as solution to Stratonovich SDE

$$dU_{x,\mu} = i \sum_a T^a \left[K^a_{x,\mu}(U)\, dt + g(t) dW^a_{x,\mu}\right] \circ U_{x,\mu}$$

- Equivalently to flat space case

$$K^a_{x,\mu}(U) \to K^a_{x,\mu}(U) - g^2(t) D^a_{x,\mu} \log p_t(U)$$

  with time running backward $t = T \to 0$.
- The initial conditions of the reverse process are sampled from the Haar uniform.

# Loss function - Score matching

- Approximate score function by neural network $s^a_{\theta;x,\mu}(U,t)$, minimizing

$$\int_0^T \mathrm{d}t \underset{\substack{p_0(U^{(0)}) \\ p_{0t}(U^{(t)}|U^{(0)})}}{\mathbb{E}} \left[ \sum_{x,\mu,a} \left( s^a_{\theta;x,\mu}(U^{(t)},t) - D^a_{x,\mu} \ln p_{0t}(U^{(t)}|U^{(0)}) \right)^2 \right]$$

- Consider forward process with zero drift $K^a_{x,\mu} = 0$,

$$U_{x,\mu} = \exp\left[ i\sigma(t) \sum_a T^a \eta^a_{x,\mu} \right] U_{x,\mu}, \quad \sigma^2(t) = \int_0^t ds\, g^2(s)$$

- Disregard contributions from multiple wrappings

$$p_{0t}(U^{(t)}|U^{(0)}) \approx p(\eta) \propto \exp\left[ -\frac{1}{2} \sum_{a,x,\mu} (\eta^a_{x,\mu})^2 \right]$$

- Tractable loss function

$$\int_0^T \mathrm{d}t \underset{\substack{p_0(U^{(0)}) \\ p_{0t}(U^{(t)}|U^{(0)})}}{\mathbb{E}} \left[ \sum_{x,\mu,a} \left( s^a_{\theta;x,\mu}(U^{(t)},t) + \frac{\eta^a_{x,\mu}}{\sigma(t)} \right)^2 \right]$$

- Process interpolates gauge invariant objects, prior and target distributions
- The score model itself has to respect gauge equivariance
- *Lattice Convolutional Neural Networks* (L-CNNs) can express arbitrary gauge invariant and equivariant functions

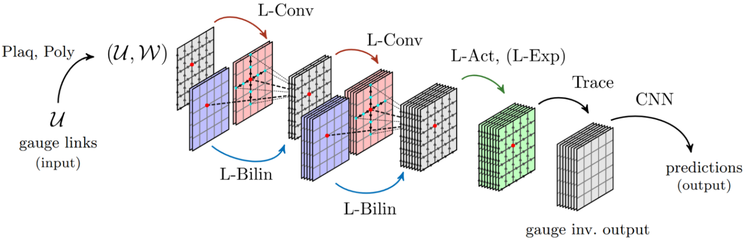



Figure: M. Favoni, A. Ipp, D.I. Müller, D. Schuh, Phys. Rev. Lett. **128**, 3 (2022)

# DMs for LFT

- Diffusion models are intended to generate field configurations.
- To do so, they need to be trained on a high-quality dataset
- What's the advantage?
- Extrapolate to bigger lattices
- Extrapolate to different couplings $\rightarrow$ Metropolis-adjusted annealed Langevin

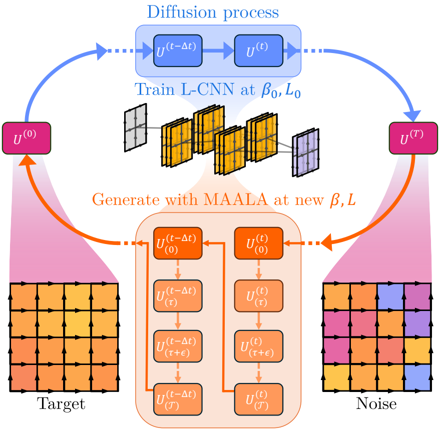
# Metropolis-adjusted Annealed Langevin

- After training at $\beta_0$, the model can be employed at different $\beta$
- Solve SDE using *score rescaling* to generalize in $\beta$

$$U^t_{\tau+\epsilon} = \exp\left[i\left(\epsilon\frac{\beta}{\beta_0}\hat{s}_\theta(U,t)+\sqrt{2\epsilon}\hat{\eta}\right)\right]U^t_\tau$$

- Include accept reject w.r.t. target action

$$p_{\text{acc}} = \min\left\{1, \frac{p(\widetilde{U}^{(t)}_{(\tau)})}{p(U^{(t)}_{(\tau)})}\frac{q(U^{(t)}_{(\tau)}|\widetilde{U}^{(t)}_{(\tau)})}{q(\widetilde{U}^{(t)}_{(\tau)}|U^{(t)}_{(\tau)})}\right\}$$

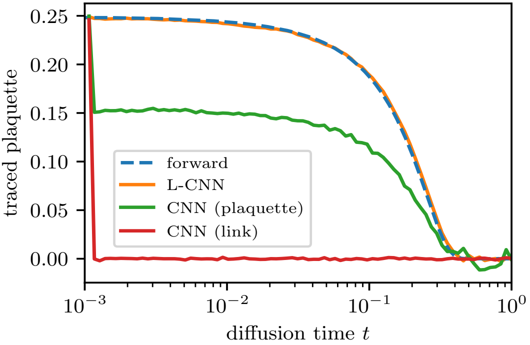


# Numerical setup

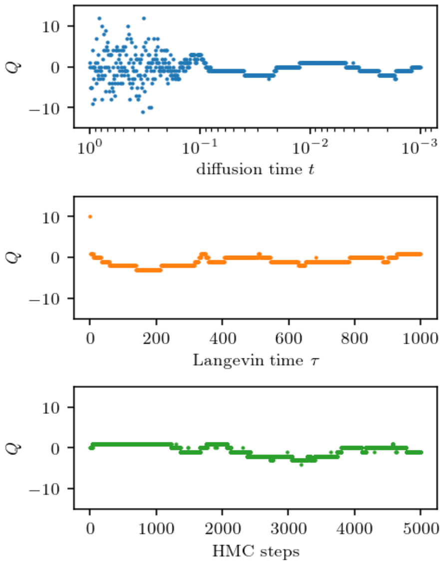
- We consider $U(2)$ and $SU(2)$ in two dimensions
- Train using 100k HMC-generated data at $\beta_0 = 2$, $L = 16$
- Employ at range of values $2 \leq \beta \leq 18$, and $L = 16, 32, 64$
- Wilson loops $W_{\ell\times\ell}$ and topological charge/susc. ($U(2)$ only)
- In strong coupling limit, $\langle W_{\ell\times\ell}\rangle \to 0$ and $\langle \chi_{\text{top}}\rangle \to \frac{1}{12}$

# Comparison with standard CNNs

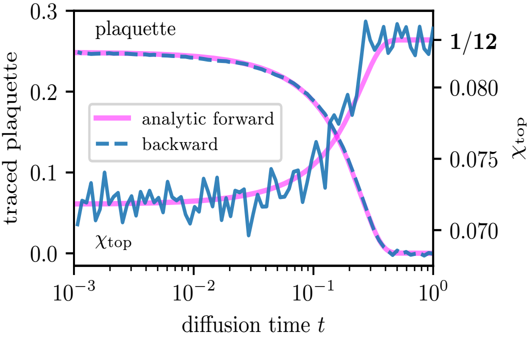
- Naive link implementation on standard CNNs fail
- CNNs with invariant representations struggle to fully capture target
- L-CNNs manage to follow analytical expectations
- Minimal corrections from the Metropolis step



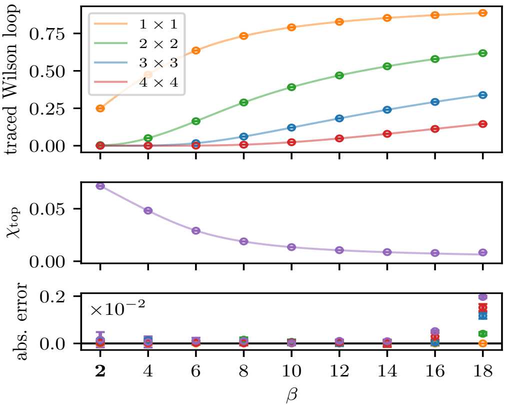
- Evolution of topological charge $Q$ for one field configuration at $\beta = 14, L = 16$
- **MAALA** explores many topological vacua in the early stages of the de-noising process
- Standard **SQ** and **HMC** might remain trapped in a few topological sectors.



- Observables interpolate between target and strong coupling values
- Forward process solved analytically using character expansions
- Time dependence $\exp(-C_2(t)\sigma(t))$
- Excellent agreement with de-noising process



- $W_{\ell\times\ell}$ and $\chi_{top}$ in $U(2)$
- Trained at $\beta_0 = 2$, generalize in $\beta$
- Extrapolation breaks down at $\beta \sim 16$
- However, it covered a large range of couplings



# Detailed comparison $U(2)$

Table: Numerical results for U(2) for larger Wilson loops and the topological susceptibility at $L = 16$, $32$ and $64$ and three separate inverse couplings $\beta$. The DM was trained on $\beta = 2$, $L = 16$ (highlighted). For $L = 32$ and $L = 64$ only $1000$ samples were generated.

| $U(2)$ | | $2 \times 2$ | $3 \times 3$ | $4 \times 4$ | $\chi_{\text{top}}$ |
|---|---|---|---|---|---|
| $\boldsymbol{\beta = 2}$ | $\boldsymbol{L = 16}$ | $0.00381(7)$ | $1.3(7) \cdot 10^{-4}$ | $-3(9) \cdot 10^{-5}$ | $0.0716(3)$ |
| | $L = 32$ | $0.0038(1)$ | $-1(1) \cdot 10^{-4}$ | $0(1) \cdot 10^{-4}$ | $0.070(1)$ |
| | $L = 64$ | $0.00378(6)$ | $2(6) \cdot 10^{-5}$ | $5(7) \cdot 10^{-5}$ | $0.072(1)$ |
| | exact | $0.003835$ | $3.660 \cdot 10^{-6}$ | $2.163 \cdot 10^{-10}$ | $0.071802$ |
| $\beta = 10$ | $L = 16$ | $0.39106(9)$ | $0.1209(1)$ | $0.0234(1)$ | $0.01345(8)$ |
| | $L = 32$ | $0.3909(2)$ | $0.1210(2)$ | $0.0236(2)$ | $0.0137(2)$ |
| | $L = 64$ | $0.39108(7)$ | $0.12102(9)$ | $0.02335(8)$ | $0.0135(2)$ |
| | exact | $0.390989$ | $0.120884$ | $0.023370$ | $0.013442$ |
| $\beta = 18$ | $L = 16$ | $0.61824(7)$ | $0.3383(1)$ | $0.1450(1)$ | $0.00837(5)$ |
| | $L = 32$ | $0.6184(1)$ | $0.3390(2)$ | $0.1459(2)$ | $0.0084(1)$ |
| | $L = 64$ | $0.6184(1)$ | $0.3390(2)$ | $0.1460(2)$ | $0.0084(1)$ |
| | exact | $0.618647$ | $0.339427$ | $0.146478$ | $0.006399$ |

# Detailed comparison $SU(2)$

Table: Numerical results for SU(2) calculations on $L = 16$, $32$ and $64$ lattices for Wilson loops at several inverse couplings $\beta$. The DM was trained on $\beta = 2$, $L = 16$.

| $SU(2)$ | | $1 \times 1$ | $2 \times 2$ | $3 \times 3$ | $4 \times 4$ |
|---|---|---|---|---|---|
| $\boldsymbol{\beta = 2}$ | $\boldsymbol{L = 16}$ | **0.4333(2)** | **0.0357(3)** | **0.0006(3)** | $-0.0003(3)$ |
| | $L = 32$ | 0.4333(4) | 0.0349(5) | 0.0014(5) | $-0.0001(5)$ |
| | $L = 64$ | 0.4333(2) | 0.0351(3) | 0.0002(3) | 0.0004(2) |
| | exact | 0.433127 | 0.035194 | 0.000536 | $1.534 \cdot 10^{-6}$ |
| $\beta = 10$ | $L = 16$ | 0.85420(7) | 0.5325(3) | 0.2421(5) | 0.0807(5) |
| | $L = 32$ | 0.8541(1) | 0.5326(4) | 0.2422(8) | 0.0808(8) |
| | $L = 64$ | 0.85420(5) | 0.5323(2) | 0.2421(4) | 0.0805(4) |
| | exact | 0.854185 | 0.532364 | 0.242086 | 0.080322 |
| $\beta = 18$ | $L = 16$ | 0.91793(4) | 0.7097(2) | 0.4621(4) | 0.2533(6) |
| | $L = 32$ | 0.91782(7) | 0.7096(3) | 0.4620(7) | 0.253(1) |
| | $L = 64$ | 0.91792(3) | 0.7099(2) | 0.4628(4) | 0.2545(5) |
| | exact | 0.917894 | 0.709854 | 0.462520 | 0.253909 |

# Summary & Outlook

- Diffusion models offer a new approach for ensemble generation in LFT
- Coupled with gauge equivariant architectures (L-CNNs), successfully sample non-Abelian LGTs
- Score rescaling enables single trained models to generalize across wide range of couplings
- Scaling to 4D SU(3) theories with large volumes (see also work by Komijani, Marinkovic, Turgut [2605.06134])
- Extend to include fermionic fields and test critical scaling

# Summary & Outlook

- Diffusion models offer a new approach for ensemble generation in LFT
- Coupled with gauge equivariant architectures (L-CNNs), successfully sample non-Abelian LGTs
- Score rescaling enables single trained models to generalize across wide range of couplings
- Scaling to 4D SU(3) theories with large volumes (see also work by Komijani, Marinkovic, Turgut [2605.06134])
- Extend to include fermionic fields and test critical scaling

Thank you for listening!