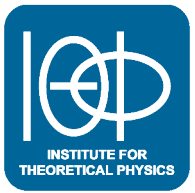
Diffusion models for lattice field theories

Andreas Ipp
Institute for Theoretical Physics, TU Wien, Austria

in collaboration with Gert Aarts, Diaa E. Habibi, David I. Müller, Thomas R. Ranner, Lingxiao Wang, Wei Wang, and Qianteng Zhu

Digital Twins for Nuclear and Particle physics – NPTwins 2026
ECT* Trento, May 12, 2026



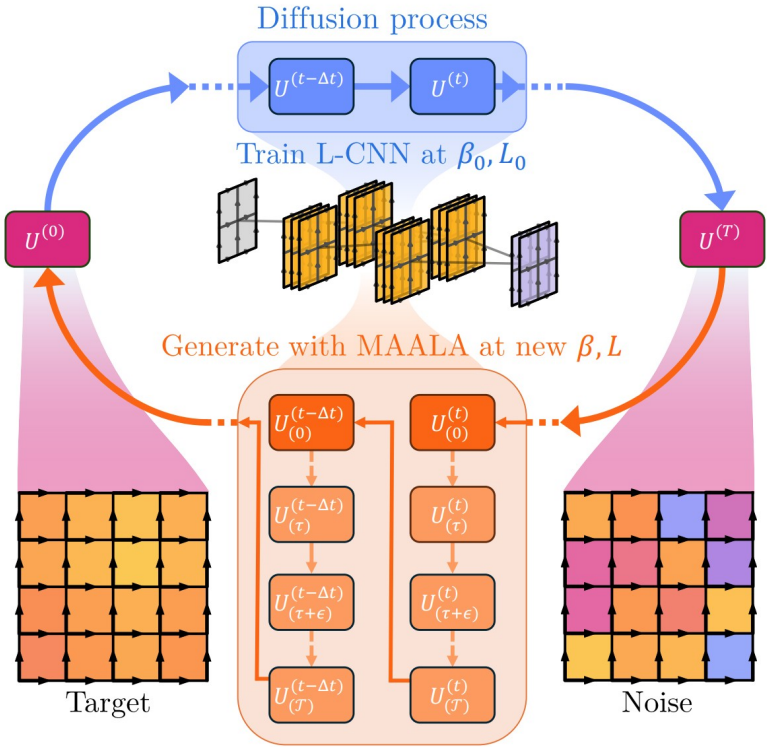


Der Wissenschaftsfonds.



Overview

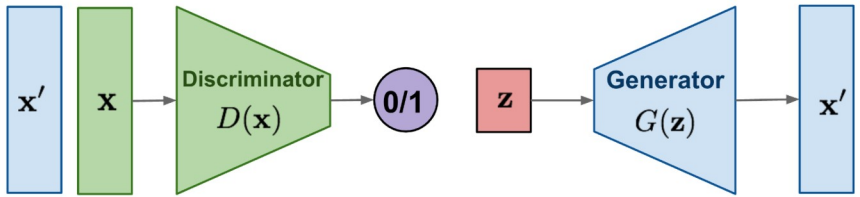
- Diffusion models
- Metropolis-adjusted annealed Langevin algorithm (MAALA)
- Lattice-gauge equivariant convolutional neural networks (L-CNNs)
- Sampling for U(2) and SU(2) gauge theories



Generalizable Equivariant Diffusion Models
for Non-Abelian Lattice Gauge Theory
Aarts, Habibi, AI, Müller, Ranner, Wang, Wang, Zhu
(2026) arXiv:2601.19552

Generative models

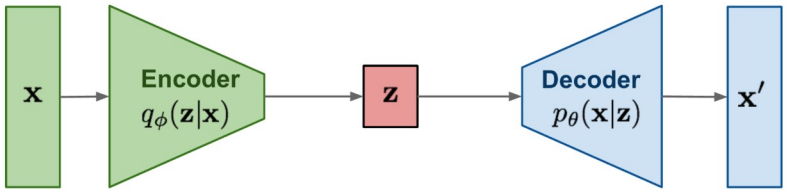
Generative Adversarial Networks:
GAN: Adversarial training (2014)



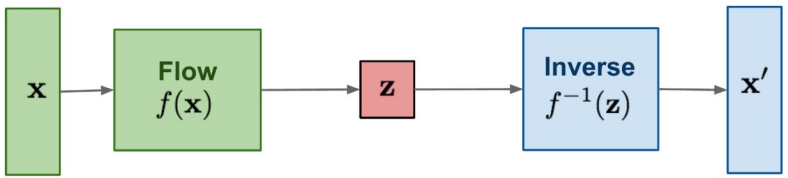


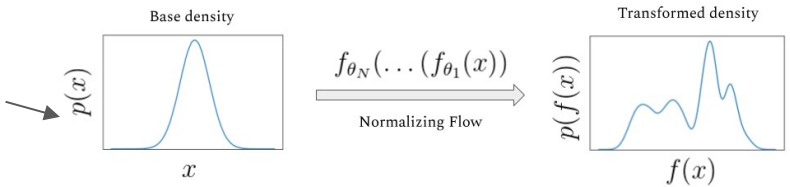
thispersondoesnotexist.com/ (2019)

Variational Autoencoders:
VAE: maximize variational lower bound



Flow-based models:
Invertible transform of distributions





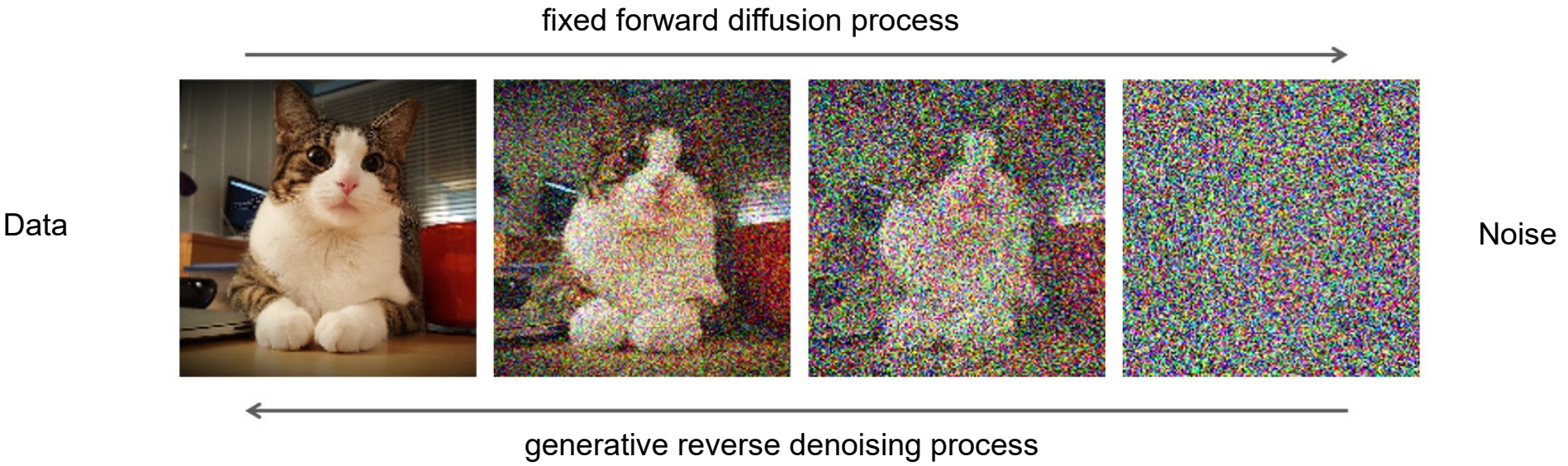
Source: The Normalizing Flow Network
(Enable gauge equivariant flows for fast sampling)

Diffusion models:
Gradually add Gaussian noise and then reverse



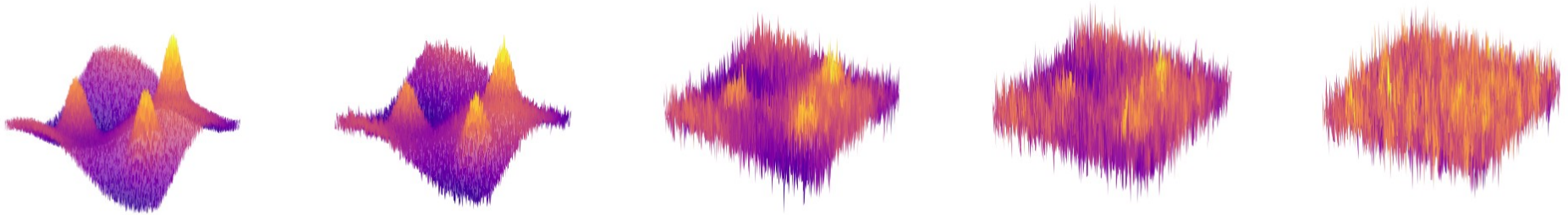
Source: Lilian Weng, Diffusion models, blog post (2021)

Diffusion model: reverse a diffusion process

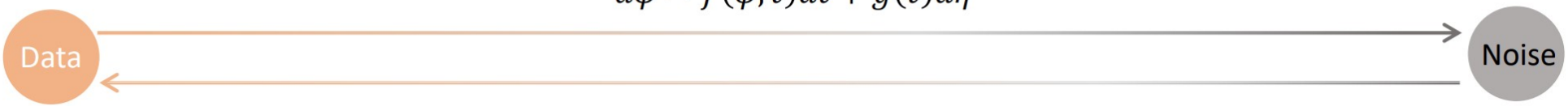


Source: Nvidia Developer blog

Diffusion models for lattice field simulations



$$d\phi = f(\phi, t)dt + g(t)d\eta$$



$$d\phi = [f(\phi, t) - g^2(t)\nabla_\phi log p_t(\phi)]dt + g(t)d\bar{\eta}$$

Physics-Conditioned Diffusion Models for Lattice Gauge Theory
Zhu, Aarts, Wang, Zhou, Wang
JHEP 03 (2026) arXiv:2502.05504

How to describe a diffusion process

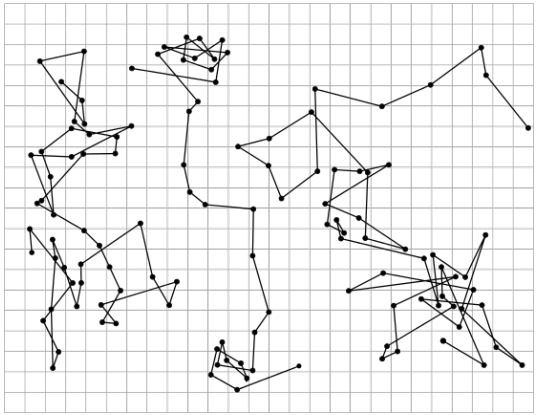


Image: Wikipedia Brownian Motion

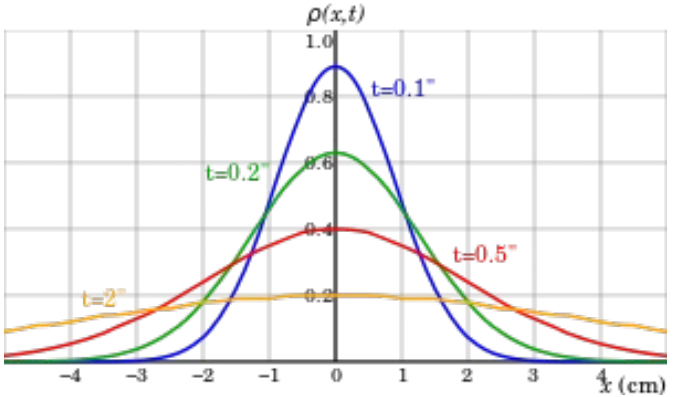


Image: Wikipedia Brownian Motion

Brownian motion:
random motion of particles
suspended in a medium

Robert Brown 1827 (microscopic observation)

Albert Einstein 1905 (link microscopic and macroscopic descriptions),
Jean Perrin 1908 (experiment; Nobel prize 1926)

Fick's (second) law
(diffusion equation)
D.. diffusion coefficient
φ.. concentration

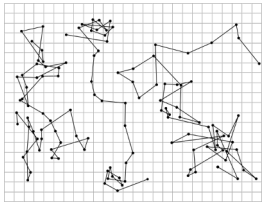
$$\frac{\partial \varphi}{\partial t} = D \frac{\partial^2 \varphi}{\partial x^2}$$

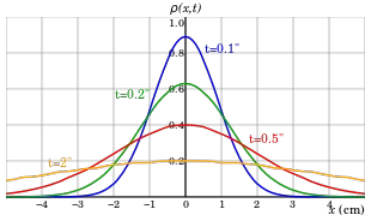
Adolf Fick 1855 (continuous equation for fluids)

Solution:

$$\varphi(x,t) = \frac{1}{\sqrt{4\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right)$$

Description of non-equilibrium systems





Langevin equation
1908
Stochastic differential equation

$$m\frac{d\mathbf{v}}{dt} = -\lambda\mathbf{v} + \boldsymbol{\eta}(t)$$

noise term η(t): Gaussian probability distribution with correlation

$$\langle \eta_i(t)\,\eta_j(t')\rangle = 2\lambda k_{\mathrm{B}} T \delta_{i,j}\delta(t-t')$$

with damping coefficient λ, and temperature *T*

Fokker-Planck equation
1914, 1917

Kolmogorov forward equation
1931

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}\left[\mu(x,t)p(x,t)\right] + \frac{\partial^2}{\partial x^2}\left[D(x,t)p(x,t)\right]$$

evolution of probability density $p(x,t)$
under drift $\mu(x,t)$ and diffusion coefficient $D(x,t)$

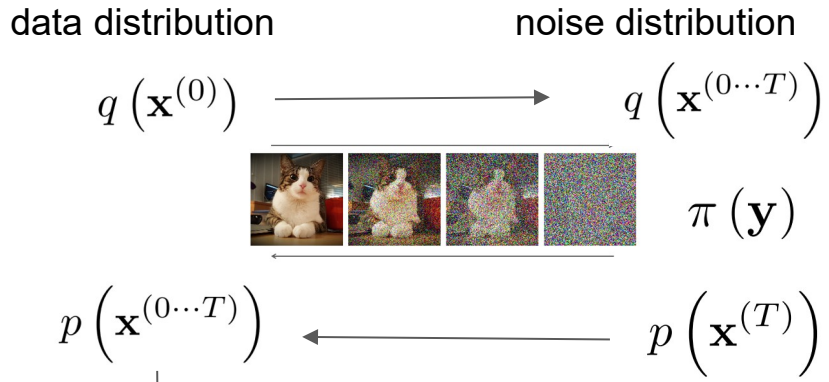
Forward and reverse trajectories

Markov diffusion kernel

diffusion rate β

$$\pi(\mathbf{y}) = \int d\mathbf{y}' T_\pi(\mathbf{y}|\mathbf{y}';\beta)\,\pi(\mathbf{y}') \qquad (1)$$

$$q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right) = T_\pi\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)};\beta_t\right). \qquad (2)$$



forward trajectory:

$$q\left(\mathbf{x}^{(0\cdots T)}\right) = q\left(\mathbf{x}^{(0)}\right)\prod_{t=1}^{T} q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right) \qquad (3)$$

backward trajectory:

$$p\left(\mathbf{x}^{(T)}\right) = \pi\left(\mathbf{x}^{(T)}\right) \qquad (4)$$

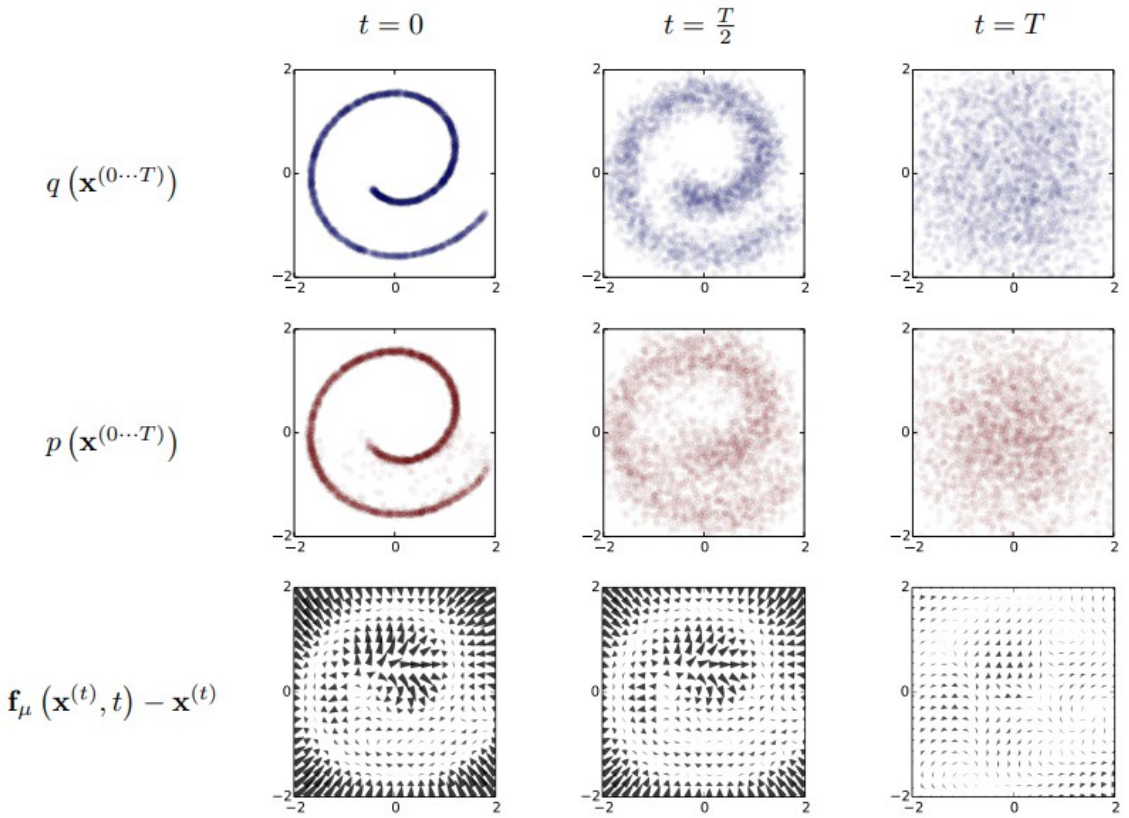
$$p\left(\mathbf{x}^{(0\cdots T)}\right) = p\left(\mathbf{x}^{(T)}\right)\prod_{t=1}^{T} p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right). \qquad (5)$$

$$p\left(\mathbf{x}^{(0)}\right) = \int d\mathbf{x}^{(1\cdots T)} p\left(\mathbf{x}^{(0\cdots T)}\right).$$

For Gaussian diffusion and small time steps, reverse diffusion process has same functional form as forward diffusion process [Feller, 1949]

Deep Unsupervised Learning using Nonequilibrium Thermodynamics
Sohl-Dickstein, Weiss, Maheswaranathan, Ganguli
arXiv:1503.03585

Diffusion model trained on swiss roll data



forward trajectory into isotropic Gaussian distribution

trained reverse trajectory

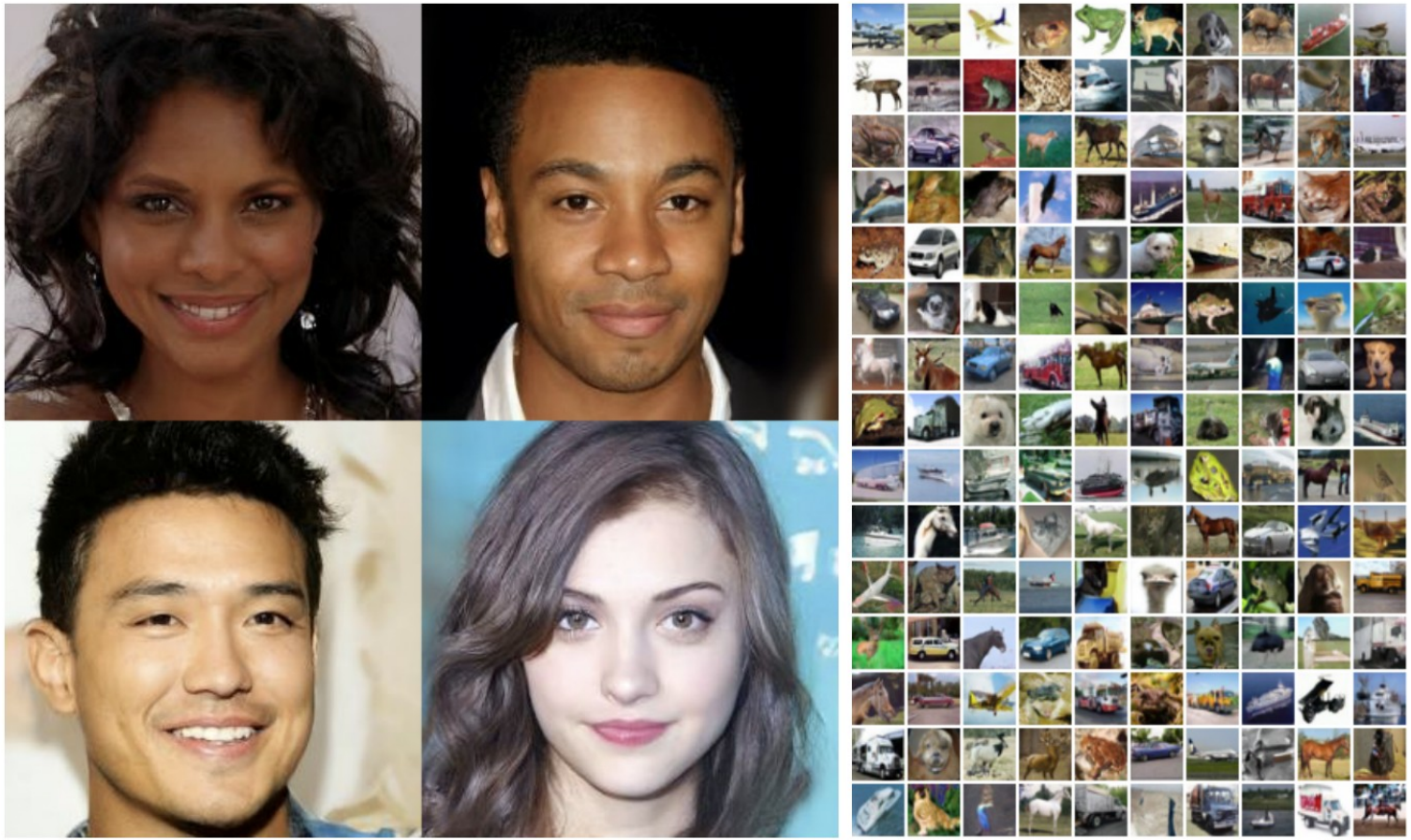
drift term for same reverse diffusion process

$$\text{Forward diffusion kernel} \quad q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right) = \mathcal{N}\left(\mathbf{x}^{(t)}; \mathbf{x}^{(t-1)}\sqrt{1-\beta_t}, \mathbf{I}\beta_t\right)$$

$$\text{Reverse diffusion kernel} \quad p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right) = \mathcal{N}\left(\mathbf{x}^{(t-1)}; \mathbf{f}_\mu\left(\mathbf{x}^{(t)}, t\right), \mathbf{f}_\Sigma\left(\mathbf{x}^{(t)}, t\right)\right)$$

Source: Fig. 1 from 1503.03585

Improvements since the 2015 paper



2006.11239

Improvements since the 2015 paper

Simplified training objective
Reparametrization to predict noise ε instead of mean μ simplifies variational bound:

$$L_{\text{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\right\|^2\right]$$

Denoising Diffusion Probabilistic Models (DDPM)
2006.11239

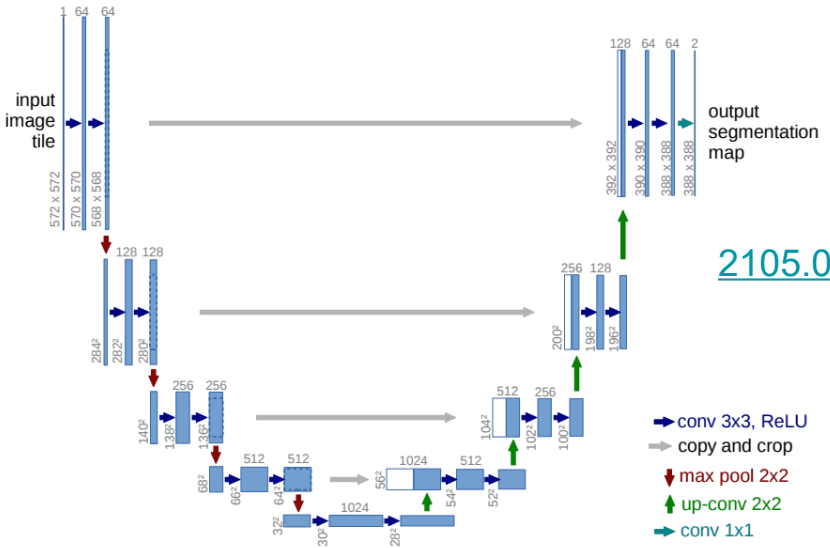
Improved diffusion schedule
Instead of linear (top) schedule for the diffusion coefficient β_t, one can use cosine (bottom) schedule:



Improved DDPM
2102.09672

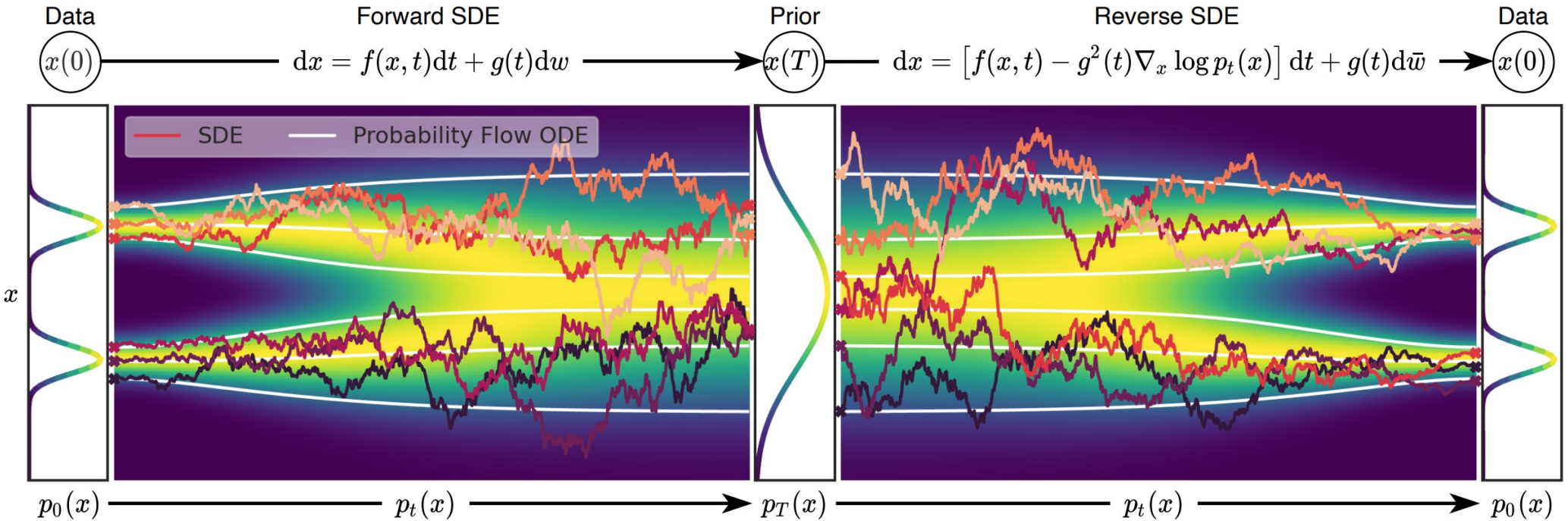
Improved network architecture
Detail improvements to the U-Net architecture (introduced in 2006.11239) include additional attention heads.

Systematic study shows that diffusion models can outperform state-of-the-art GAN-models.



2105.05233

Stochastic differential equations (SDE)



Score-Based Generative Modeling through Stochastic Differential Equations
Song et al. (2020) arXiv:2011.13456

Score function and QFT action

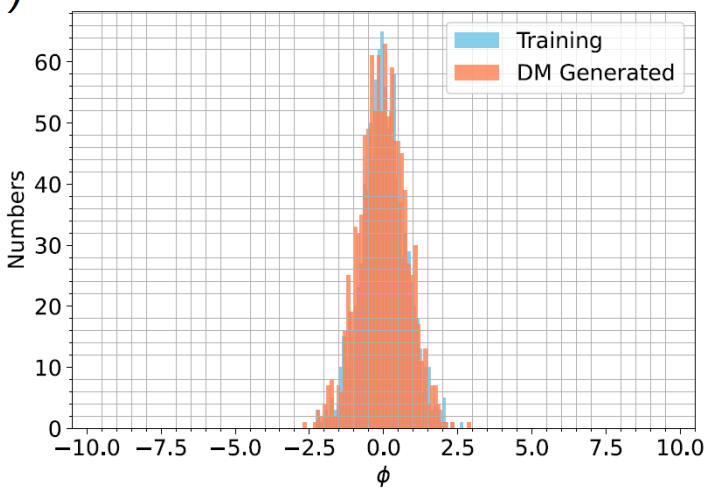
Probability of a field configuration x in quantum field theory:

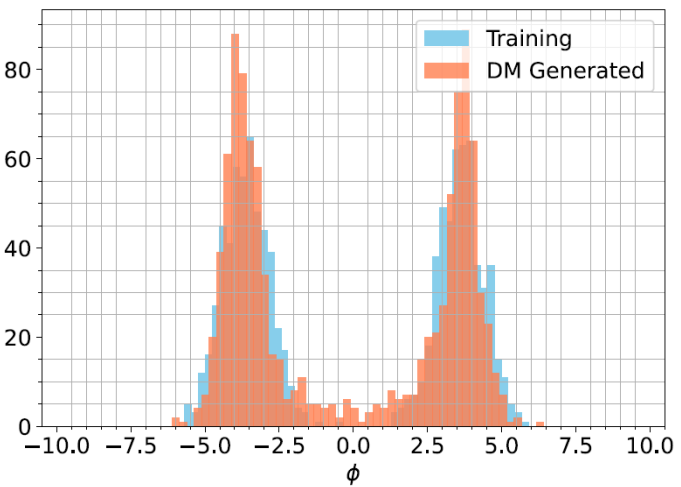
$$\langle O \rangle = \int \mathcal{D}[x]\, O(x) p(x), \quad p(x) = \frac{1}{Z} e^{-S(x)}, \quad Z = \int \mathcal{D}[x] e^{-S(x)}$$

Score function:

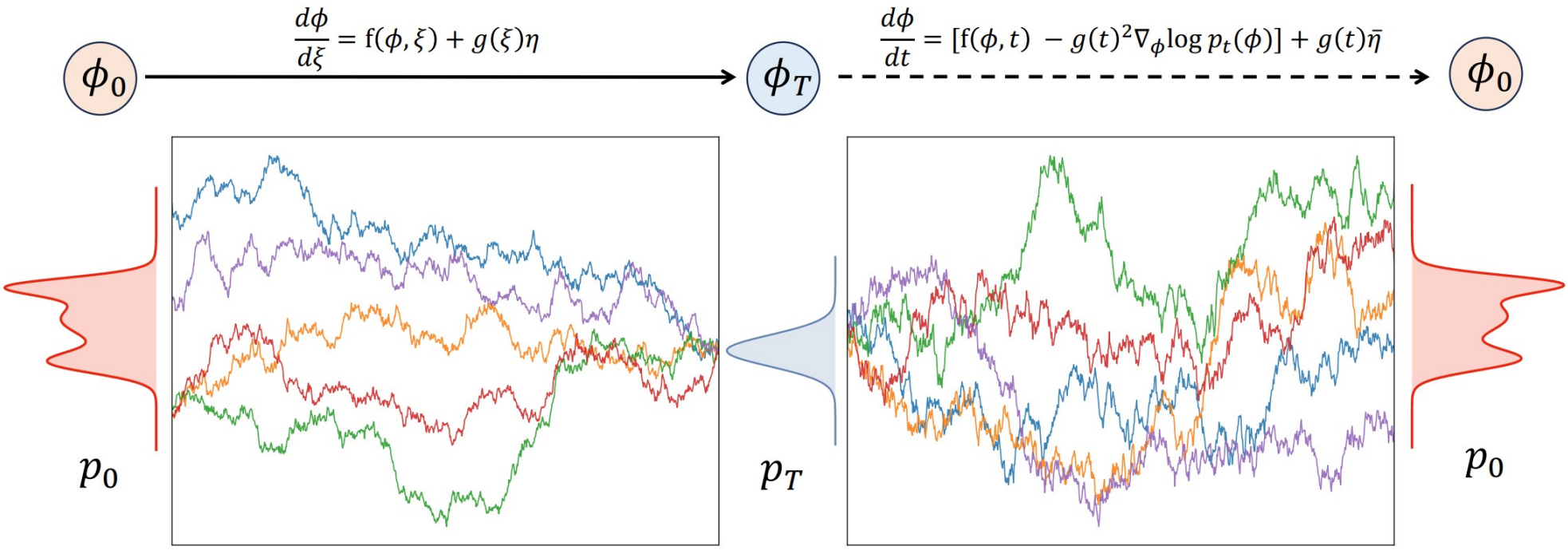
$$\nabla_x \log p(x) = -\nabla_x S(x)$$

Diffusion Models as Stochastic Quantization in Lattice Field Theory
Wang, Aarts, Zhou, JHEP 05 (2020)
arXiv:2309.17082





Forward diffusion and reverse denoising



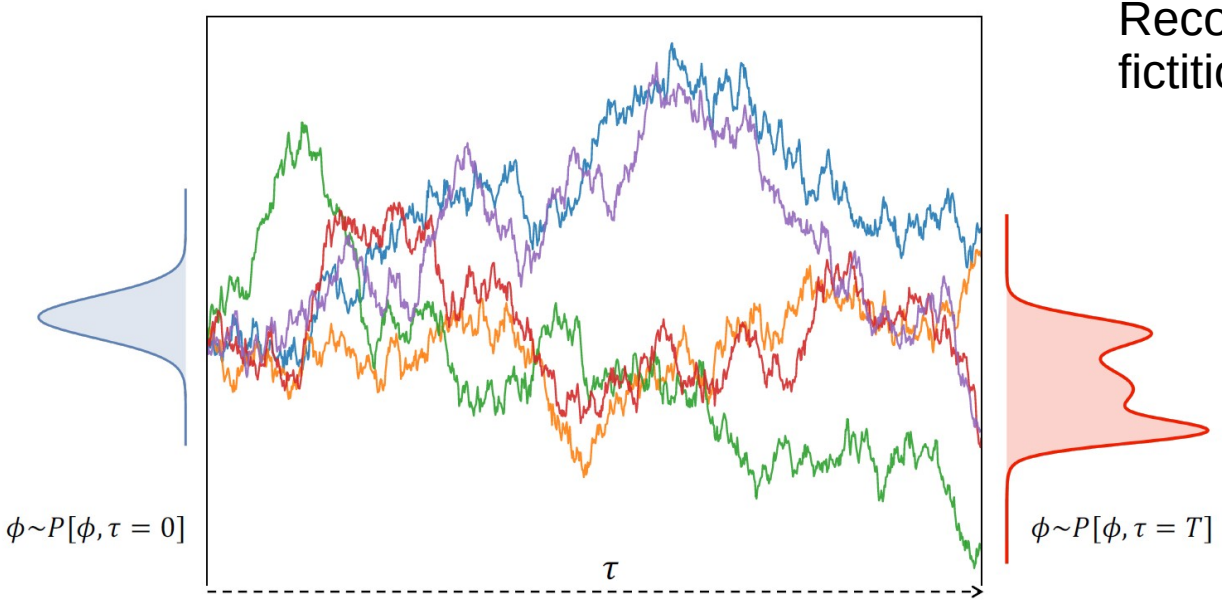
Diffusion Models as Stochastic Quantization in Lattice Field Theory
Wang, Aarts, Zhou, JHEP 05 (2020) arXiv:2309.17082

Stochastic quantization (SQ)

Evolve random field configuration according to Langevin equation:

Parisi, Wu (1980)

$$\frac{\partial x_\tau}{\partial \tau} = -\frac{\delta S(x_\tau)}{\delta x_\tau} + \sqrt{2}\eta, \quad \eta \sim \mathcal{N}(0,1)$$



Recover probability distribution for large fictitious time $\tau \to \infty$

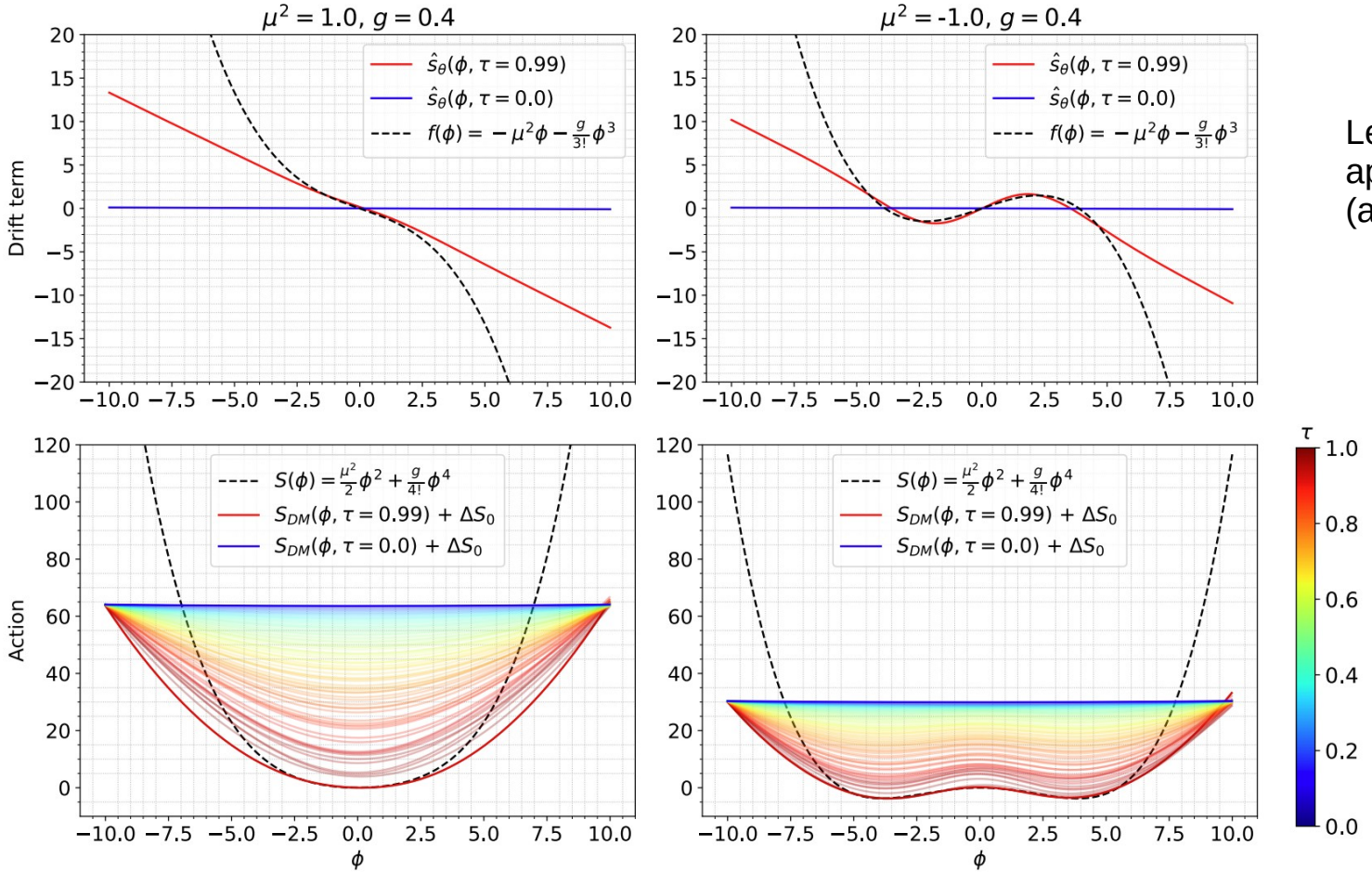
$$p(x) \propto e^{-S(x)}$$

Drift term in SQ is the score we learn in the diffusion model

$$-\nabla_x S(x) = \nabla_x \log p(x)$$

Diffusion Models as Stochastic Quantization in Lattice Field Theory
Wang, Aarts, Zhou, JHEP 05 (2020)
arXiv:2309.17082

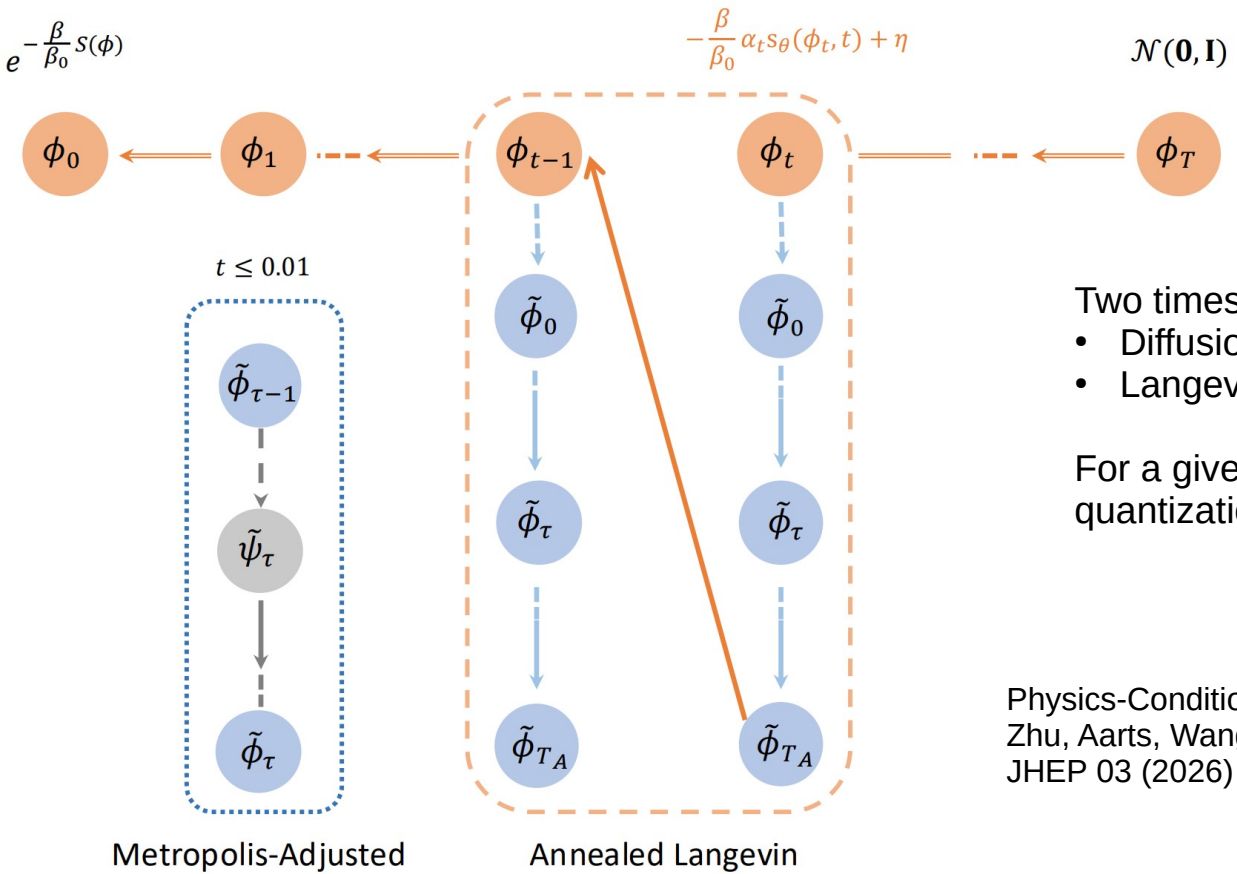
Drift term and effective action for toy model



Learned score function $\hat{s}$ approximates drift term f (after 500 iterations)

Diffusion Models as Stochastic Quantization in Lattice Field Theory
Wang, Aarts, Zhou, JHEP 05 (2020) arXiv:2309.17082

Metropolis-adjusted annealed Langevin algorithm (MAALA)



Two times:
- Diffusion time t
- Langevin time τ

For a given diffusion time t run stochastic quantization along τ.

Physics-Conditioned Diffusion Models for Lattice Gauge Theory
Zhu, Aarts, Wang, Zhou, Wang
JHEP 03 (2026) arXiv:2502.05504

QCD on the lattice

Lattice Gauge Theory allows for precision determination of non-perturbative quantities in quantum chromodynamics (QCD).

Calculate hadron masses, decay constants, thermodynamic quantities at finite temperature (QCD phase transition), topological properties, or glueball spectra.

Solve Feynman path integral numerically using Monte Carlo methods

Gauge transformation:
Rotation in internal color space at each space-time point

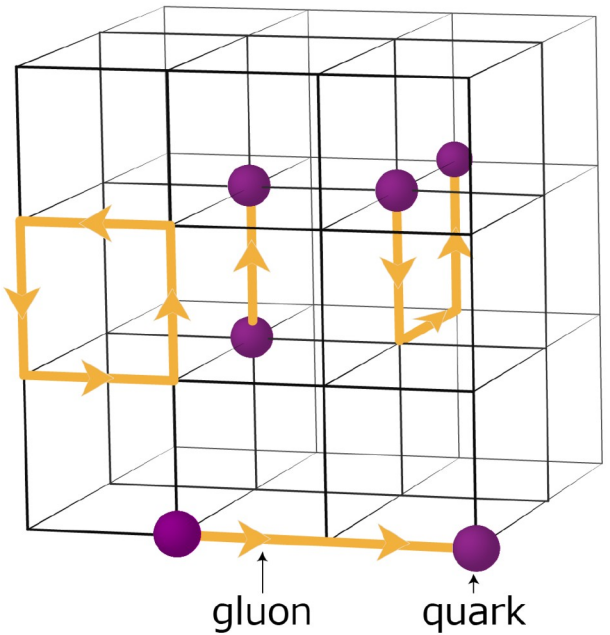


Image from Bi et al. EPJ Web Conf. 245 (2020) 09008

Machine learning approaches:

- Normalizing flow
  Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan (2020), ...
- Continuous normalizing flow
  Gerdes, de Haan, Bondesan, Cheng (2024)
- Diffusion models
  Zhu, Aarts, Wang, Zhou, Wang (2024)
- Fixed point action
  Holland, AI, Müller, Wenger (2024)

Wilson action for gluons:

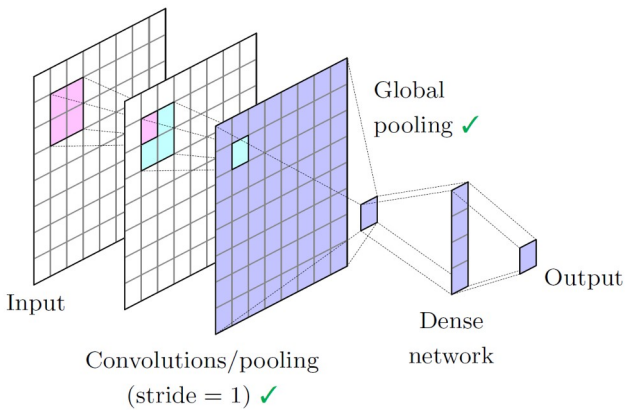
$$S_W[U] = \frac{2}{g^2} \sum_{x \in \Lambda} \sum_{\mu < \nu} \mathrm{Tr}\left[\mathbb{1} - U_{x,\mu\nu}\right]$$

Gauge transformation of link variables:

$$U_\mu(n) \;\rightarrow\; U'_\mu(n) = \Omega(n)\, U_\mu(n)\, \Omega(n+\hat{\mu})^\dagger$$

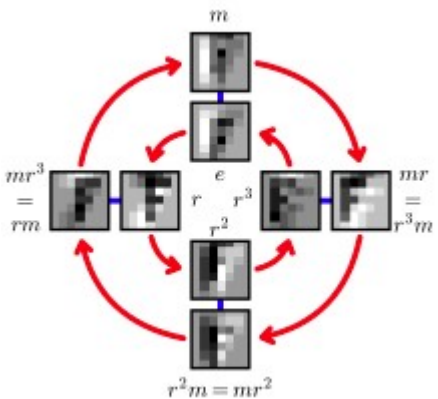
Symmetries on the lattice

Translational symmetry
→ Convolutional neural networks (CNNs)



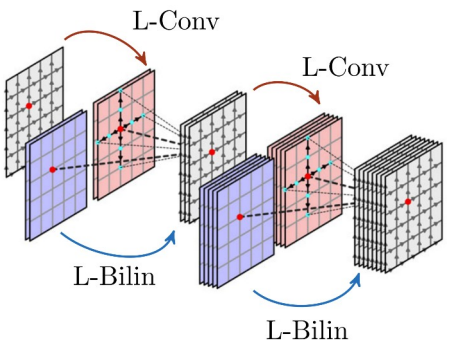
Bulusu, Favoni, AI, Müller, Schuh,
Phys. Rev. D 104 (2021) 074504

Rotation, mirror symmetry
→ Group equivariant CNNs (G-CNNs)



Cohen, Welling, ICML 2016

Lattice gauge symmetry
→ Lattice gauge equivariant CNNs (L-CNNs)



Favoni, AI, Müller, Schuh,
Phys.Rev.Lett. 128 (2022) 032003

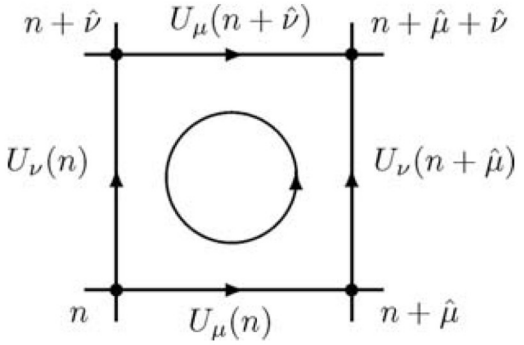
L-CNN data

Combine lattice links *U*
and locally transforming objects *W*

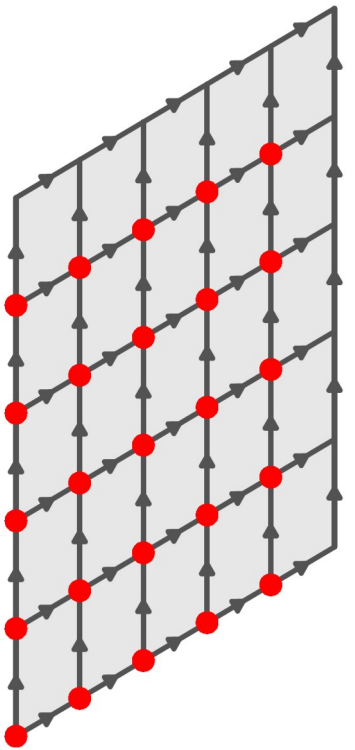
$$\text{tuple } (\mathcal{U}, \mathcal{W})$$

$$\mathcal{U} = \{U_{x,\mu}\} \text{ SU}(N_c) \text{ matrices}$$

$$\mathcal{W} = \{W_{x,i}\} \text{ with } W_{x,i} \in \mathbb{C}^{N_c \times N_c}$$



from: Gattringer, Lang (2010)



$$\mathcal{U} = \{U_{\mathbf{x},\mu}\}$$

$$\mathcal{W} = \{W_{\mathbf{x},\mu\nu}\}$$

Gauge transformation

$$T_\Omega U_{x,\mu} = \Omega_x U_{x,\mu} \Omega^\dagger_{x+\mu}$$

$$T_\Omega W_{x,i} = \Omega_x W_{x,i} \Omega^\dagger_x$$

Gauge equivariant (gauge covariant) function

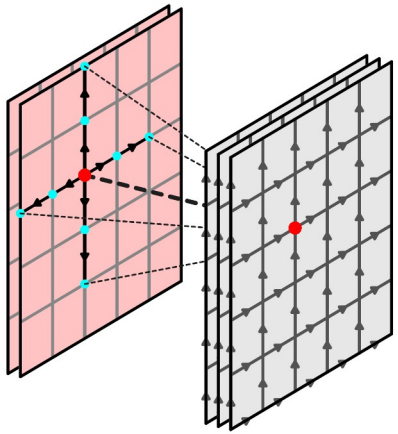
$$f(T_\Omega \mathcal{U}, T_\Omega \mathcal{W}) = T'_\Omega f(\mathcal{U}, \mathcal{W})$$

Gauge invariant function

$$f(T_\Omega \mathcal{U}, T_\Omega \mathcal{W}) = f(\mathcal{U}, \mathcal{W})$$

Lattice gauge equivariant layers

Convolution (L-Conv)

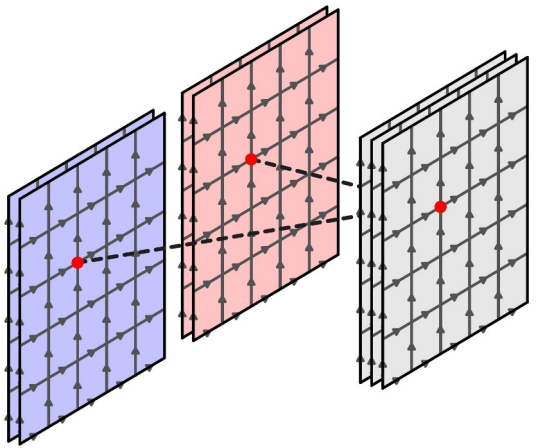


Convolution with shared weights and proper parallel transport along coordinate axes

$$(\mathcal{U},\mathcal{W}) \to (\mathcal{U},\mathcal{W}')$$

$$W'_{\mathbf{x},i} = \sum_{j,\mu,k} \omega_{i,j,\mu,k} U_{\mathbf{x},k\cdot\mu} W_{\mathbf{x}+k\cdot\mu,j} U^{\dagger}_{\mathbf{x},k\cdot\mu}$$

Bilinear layer (L-Bilin)

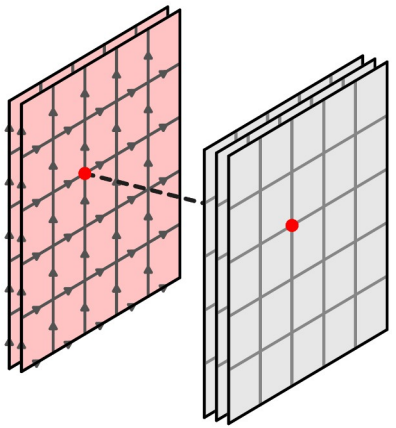


Multiply *W* at each lattice point

$$(\mathcal{U},\mathcal{W}) \times (\mathcal{U},\mathcal{W}') \to (\mathcal{U},\mathcal{W}'')$$

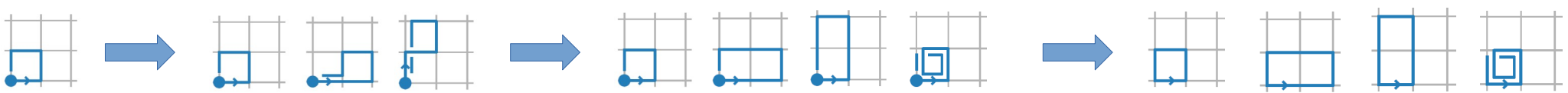
$$W''_{\mathbf{x},i} = \sum_{j,k} \alpha_{ijk} W_{\mathbf{x},j} W'_{\mathbf{x},k}$$

Trace layer

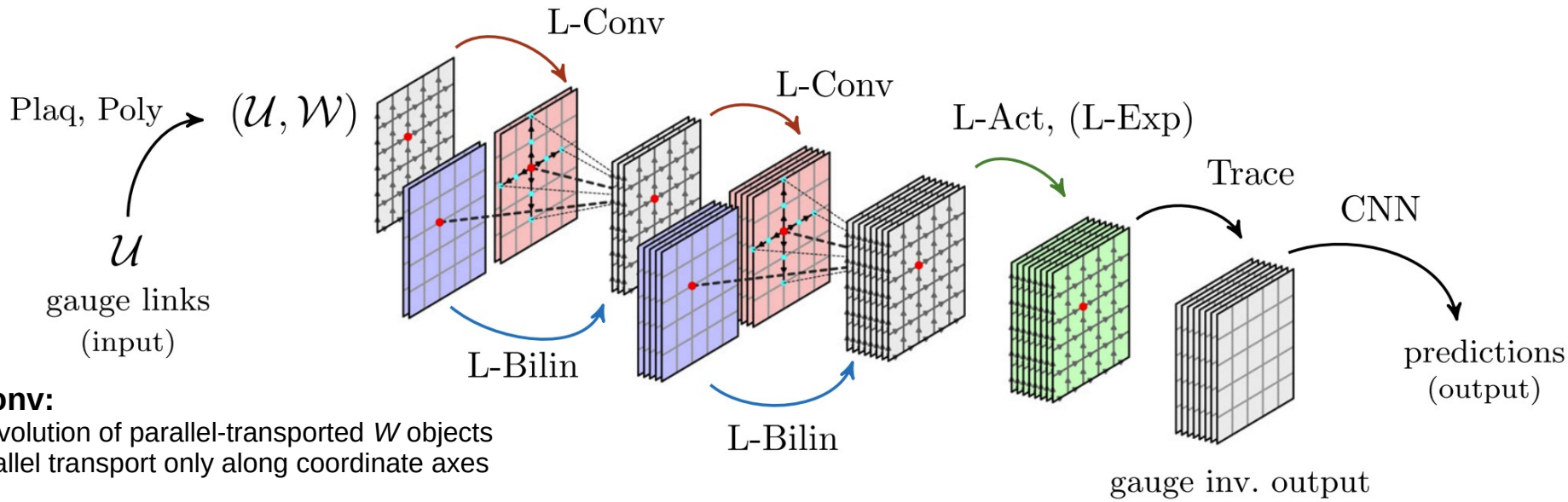


Generate gauge invariant output

$$w_{\mathbf{x},i} = \mathrm{Tr}\, W_{\mathbf{x},i} \in \mathbb{C}$$



Generic L-CNN



L-Conv:
* convolution of parallel-transported *W* objects
* parallel transport only along coordinate axes

L-Bilin:
* bilinear layer, product of locally transforming objects

L-Act:
* activation functions multiply *W* objects by scalar, gauge-invariant functions

L-Exp:
* update link variables using exponential map

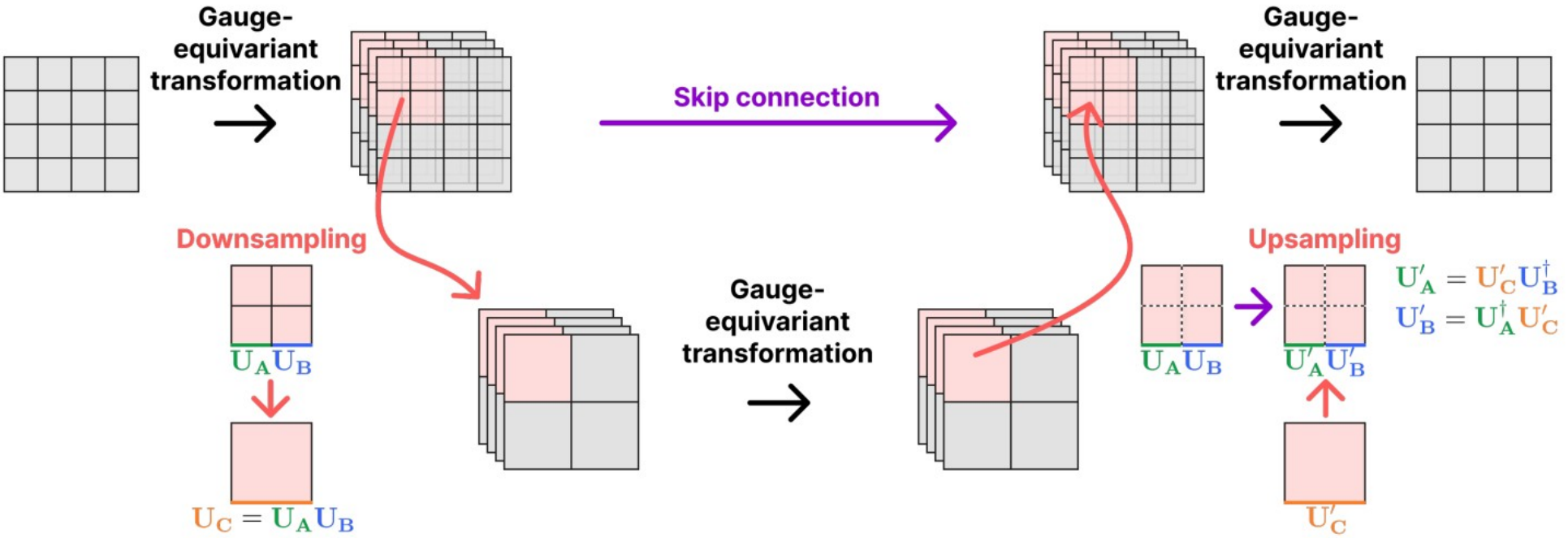
Trace:
* calculate gauge invariant trace

Plaq:
* generate all possible plaquettes

Poly:
* generate all possible Polyakov loops

Favoni, AI, Müller, Schuh,
Phys.Rev.Lett. 128 (2022) 032003

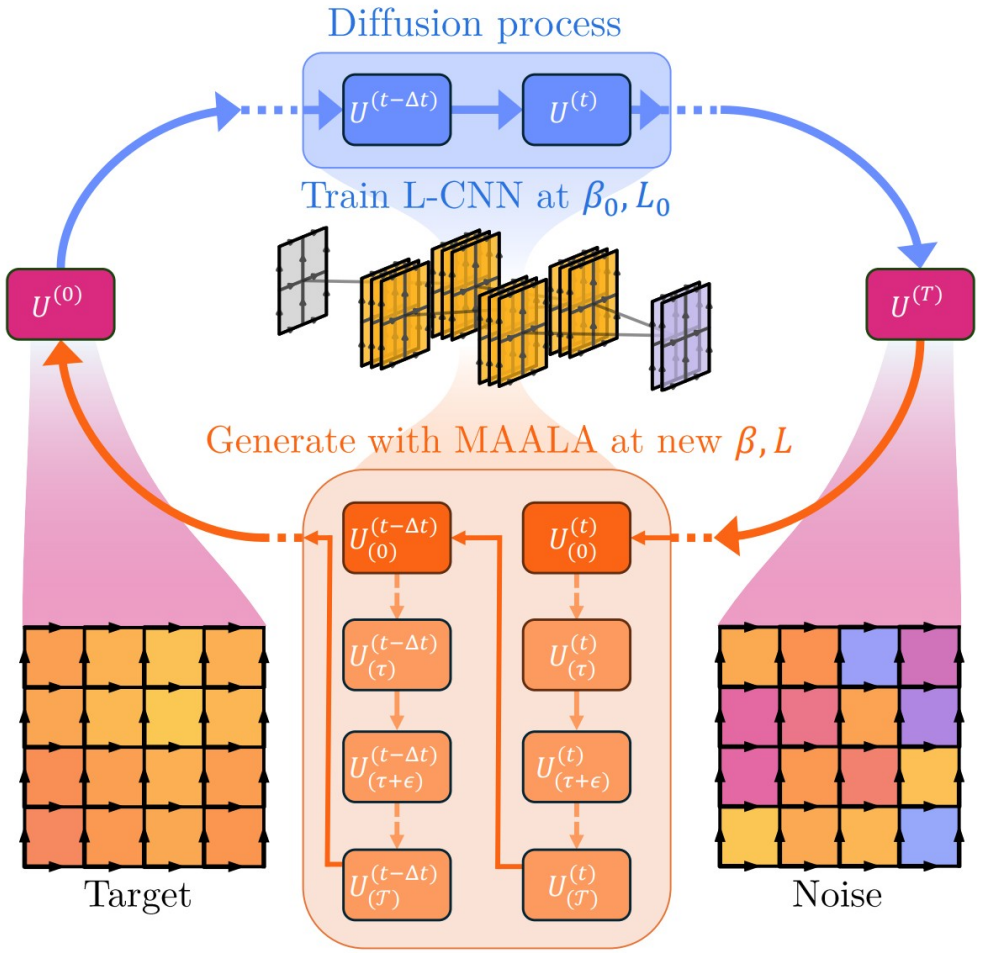
Gauge equivariant U-Net



Gauge equivariant U-Net architecture with downsampling, upsampling, and skip connections.

Diffusion model for SU(N) gauge theories
Komijani, Marinkovic, Turgut (2026)
arXiv:2605.06134

Metropolis-adjusted annealed Langevin algorithm (MAALA)



Train an L-CNN with diffused samples at a given inverse coupling β_0 and lattice size L_0.

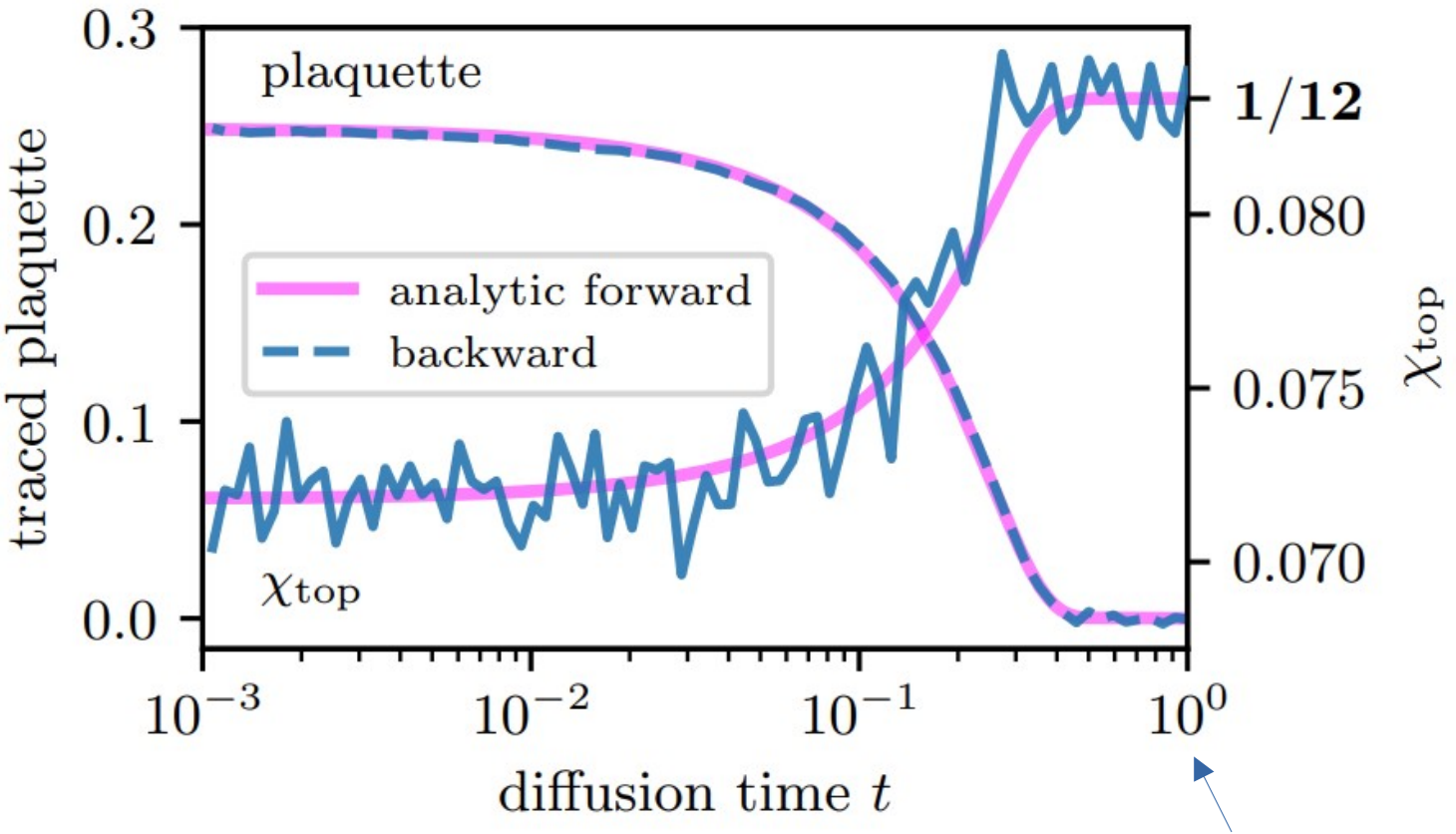
Solve the reversed diffusion process using MAALA to generate new independent samples at arbitrary β and L.

If action depends on overall factor, scaling with β'/β is possible:

$$\frac{\partial x_{t,\tau}}{\partial \tau} = \frac{\beta'}{\beta} s_\theta(x_{t,\tau}, t) + \sqrt{2}\eta.$$

Generalizable Equivariant Diffusion Models for Non-Abelian Lattice Gauge Theory
Aarts, Habibi, AI, Müller, Ranner, Wang, Wang, Zhu (2026) arXiv:2601.19552

Learned generative process with the forward diffusion process



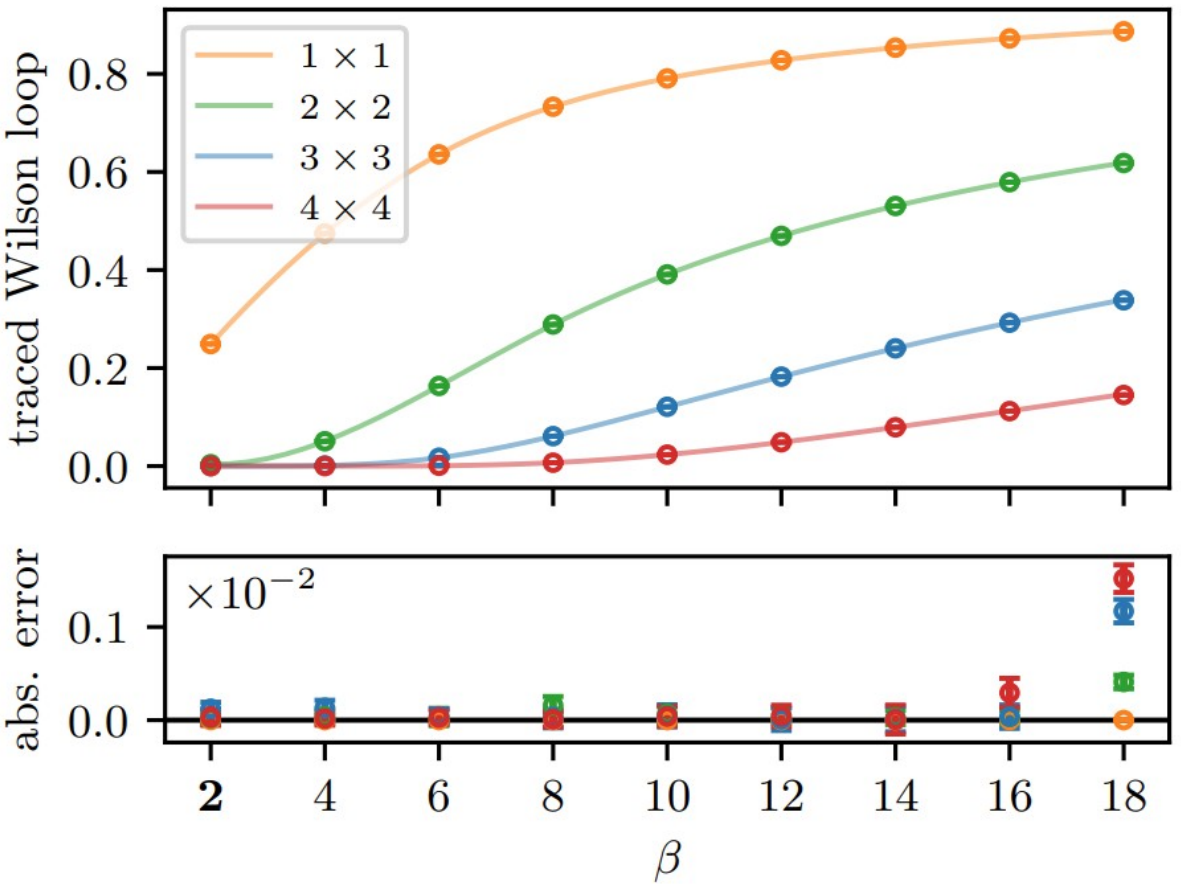
Close agreement between diffusion model and the analytical diffusion process for traced plaquette and topological susceptibility χ_{top}.

$\beta = 2$ on a $L = 16$ lattice

Strong coupling limit

Generalizable Equivariant Diffusion Models for Non-Abelian Lattice Gauge Theory
Aarts, Habibi, AI, Müller, Ranner, Wang, Wang, Zhu
(2026) arXiv:2601.19552

Predictions of Wilson loops



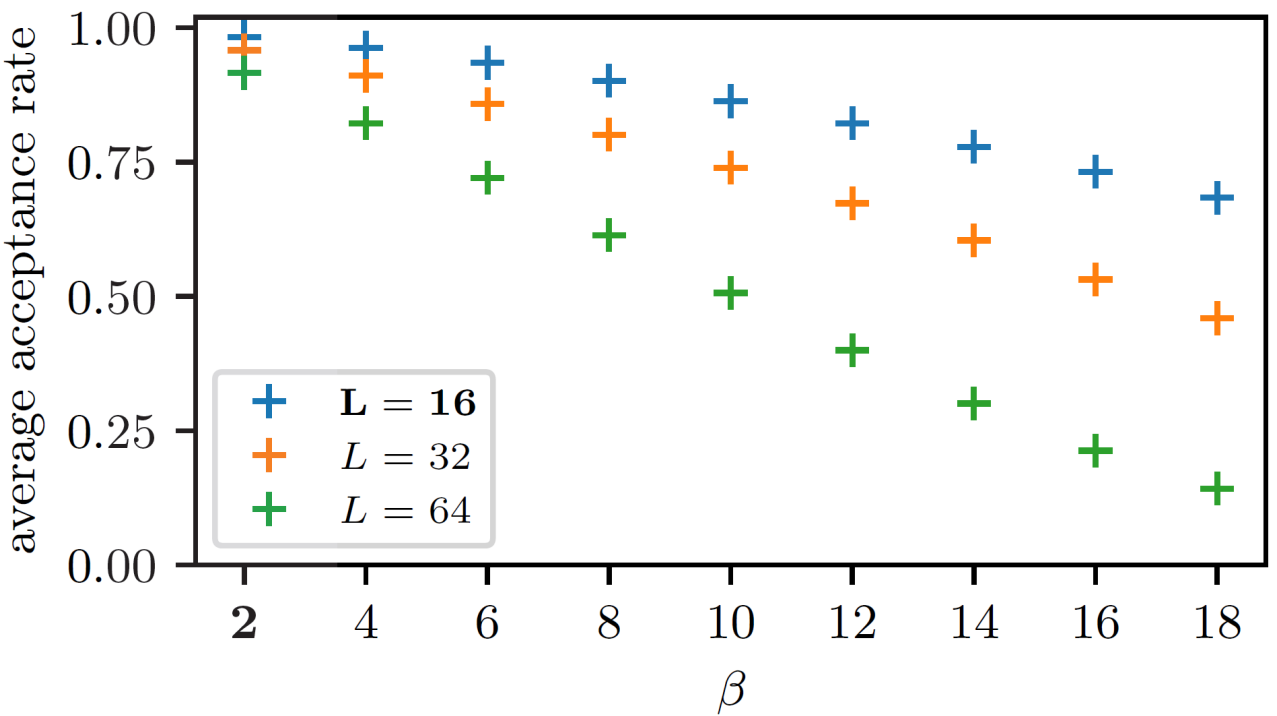
Predictions for two-dimensional U(2) gauge theory for n × n Wilson loops.

- Trained at inverse coupling β = 2, L = 16
- Evaluated at larger β

Significant deviations only occur for the largest β values.

Generalizable Equivariant Diffusion Models for Non-Abelian Lattice Gauge Theory
Aarts, Habibi, Al, Müller, Ranner, Wang, Wang, Zhu (2026) arXiv:2601.19552

Metropolis acceptance rate



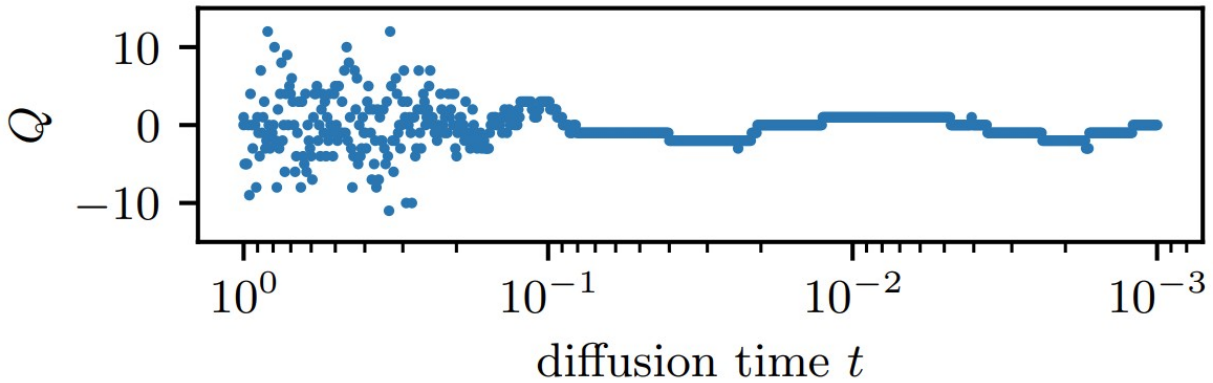
Metropolis accept / reject step for generated configuration.

- Trained at inverse coupling β = 2, L = 16
- Evaluated at larger β

Significant deviations only occur for the largest β values.

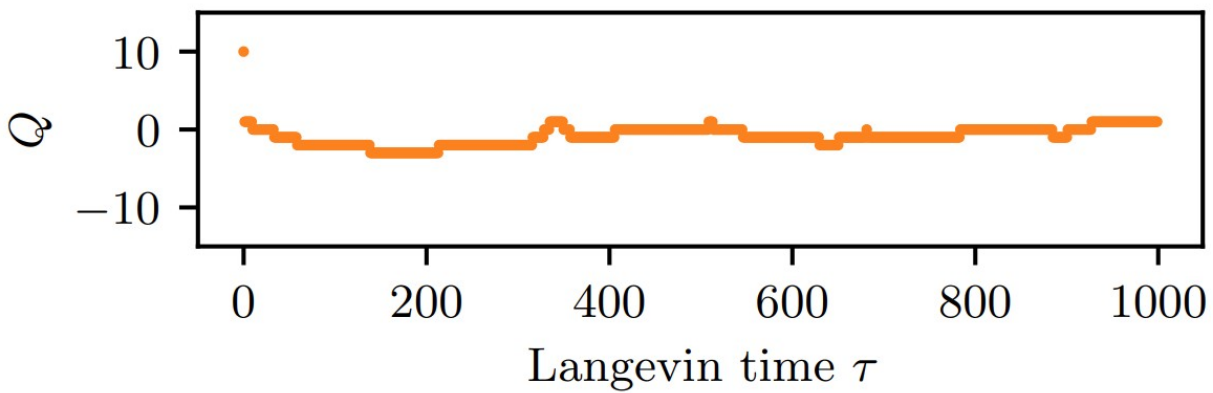
Generalizable Equivariant Diffusion Models for Non-Abelian Lattice Gauge Theory
Aarts, Habibi, AI, Müller, Ranner, Wang, Wang, Zhu (2026) arXiv:2601.19552

MAALA vs. Stochastic quantization



Evolution of the topological charge Q for one field configuration (β = 14, L = 16)

MAALA initially exhibits frequent transitions and efficiently probes several topological sectors.



Stochastic quantization with the Wilson action remains trapped in a few topological sectors for extended periods.

Generalizable Equivariant Diffusion Models for Non-Abelian Lattice Gauge Theory
Aarts, Habibi, Al, Müller, Ranner, Wang, Wang, Zhu (2026) arXiv:2601.19552

Conclusions

- Diffusion Models successfully sample non-Abelian gauge theories by utilizing gauge-equivariant L-CNN architectures.
- Independent sample generation effectively eliminates critical slowing down and topological freezing.
- Score rescaling enables a single trained model to generalize across a wide range of couplings and lattice sizes.

Outlook

- Scaling to 4D SU(3) theories with large volumes.
- Combine with Fixed-Point Action simulations to suppress lattice artifacts.
- Extend the framework to include fermionic fields.

