Physics-Conditioned Diffusion Models for Lattice Gauge Theory

Qianteng Zhu, Gert Aarts, Wei Wang, Kai Zhou and Lingxiao Wang

arXiv:2502.05504 [hep-lat]

Hiroshi Ohno

Journal Club
June 13, 2025

Summary

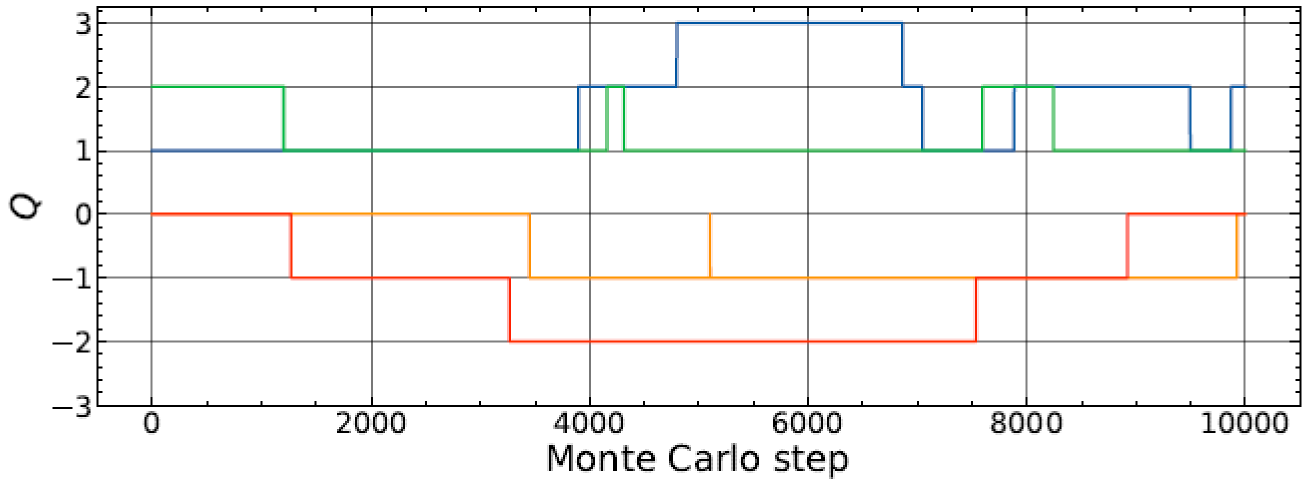
- A diffusion model for sampling in lattice gauge theory (LGT) was proposed.
  - Stochastic quantization was incorporated as physics conditioned sampling

  → Enabling to sample configurations with different couplings

- Applicability of the proposed sampler was demonstrated for 2D U(1) theory.
  - No topological freezing problem
  - Can be used for different lattice sizes and couplings without further training
  - More efficient sampling of topological quantities compared to HMC and Langevin

Motivation

- Traditional LGT sampling (e.g., HMC, Langevin) suffers from:
  - **Critical slowing down**
  - **Topological freezing**

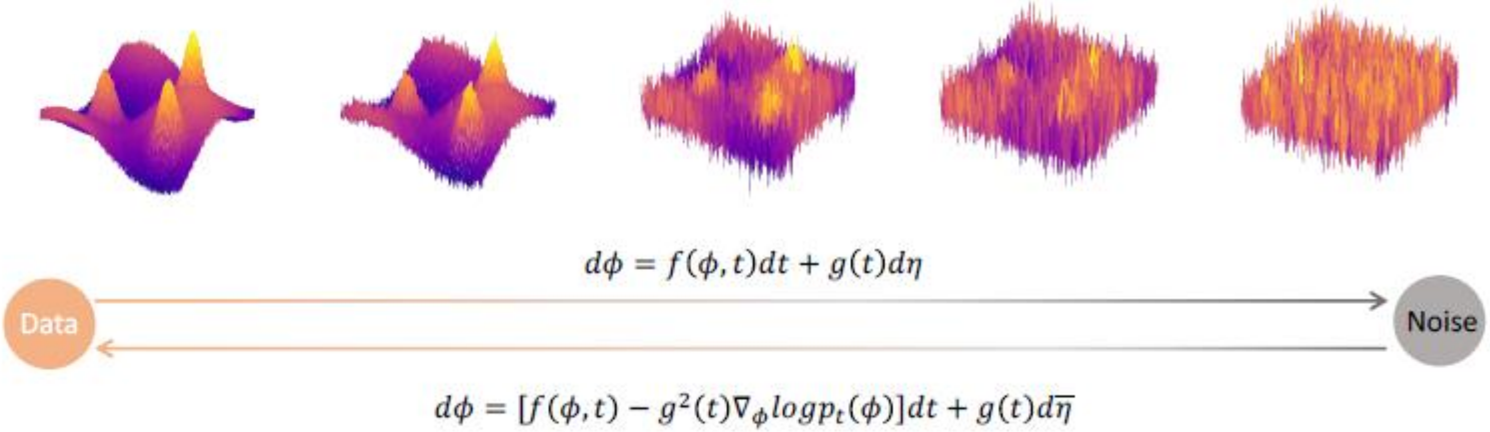


Topological charge in 2D U(1) theory

$$Q = \sum_x q(x) = \frac{1}{2\pi} \sum_x \arg U_{12}(x)$$

- Diffusion models:
  - random samples -- (denoising) --> samples following a target probability distribution
- Develop a diffusion model to overcome the problem above

Diffusion models



- Forward process:
  - Add noise through stochastic differential equation (SDE): $\frac{d\phi}{d\xi} = f(\phi, \xi) + g(\xi)\eta(\xi)$
  - f: drift term → set to zero in this study
  - η: noise term following $\langle \eta(\xi)\eta(\xi') \rangle = \delta(\xi - \xi')$
  - g: diffusion coefficient → choose $g(\xi) = \sigma^{\xi}$ in this study → transition probability: Gaussian
- Reverse process:
  - Denoise through reverse SDE: $\frac{d\phi}{dt} = \left[f(\phi, t) - g^2(t)\nabla_\phi \log p_t(\phi)\right] + g(t)\bar{\eta}(t)$
  - $\nabla_\phi \log p_t(\phi)$: unknown → model a score function $s(\phi, \xi)$ with deep neural networks

Denoising score matching objective

- Minimize the loss function θ: learnable parameters

$$\ell(\theta;\xi) \equiv \frac{1}{2}\mathbb{E}_{p_{data}(\phi)}\mathbb{E}_{\tilde{\phi}\sim\mathcal{N}(\phi,\sigma_\xi^2 I)}\left[\left\|s_\theta(\tilde{\phi},\xi) + \nabla_\phi \log p_\xi(\phi_\xi|\phi_0)\right\|_2^2\right]$$

$$= \frac{1}{2}\mathbb{E}_{p_{data}(\phi)}\mathbb{E}_{\tilde{\phi}\sim\mathcal{N}(\phi,\sigma_\xi^2 I)}\left[\left\|s_\theta(\tilde{\phi},\xi) + \frac{\tilde{\phi}-\phi}{\sigma_\xi^2}\right\|_2^2\right]$$

← p_ξ: Gaussian

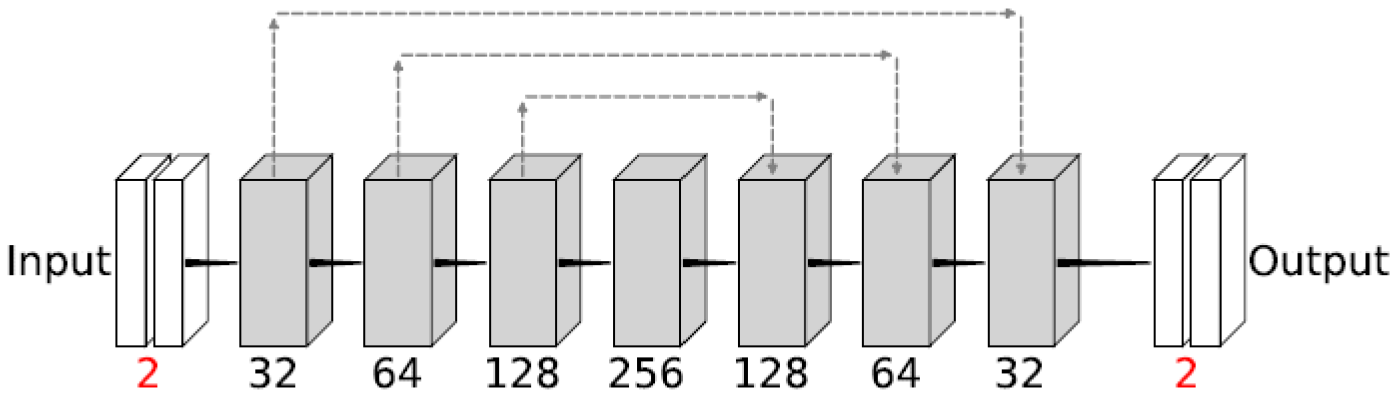
$$\mathcal{L}(\theta;\{\xi_i\}_{i=1}^{N_T}) \equiv \frac{1}{N_T}\sum_{i=1}^{N_T}\lambda(\xi_i)\ell(\theta;\xi_i), \qquad \lambda(\xi) = \sigma_\xi^2$$

→ $\frac{(\tilde{\phi}-\phi)}{\sigma_\xi} \sim \mathcal{N}(0,1)$ and $\left\|\sigma_\xi s_\theta(\phi,\xi)\right\|_2 \propto 1$

→ order of magnitude of $\lambda(\xi)\ell(\theta;\xi)$: independent of ξ

Neural network architecture

- U-Net Architecture



  - Encoder: convolutional layers

$$h^{(l+1)} = \sigma\left(W^{(l)} * h^{(l)} + b^{(l)}\right)$$

convolution

  - Decoder: transposed convolutional layers

$$h^{(l-1)} = \sigma\left(W_T^{(l-1)} \circledast h^{(l)} + b^{(l-1)}\right)$$

deconvolution

Lattice size can be changed.

Metropolis-Adjusted Annealed Langevin Algorithm (1)

- Stochastic quantization (Parisi-Wu)

$$\frac{\partial\phi(x,\tau)}{\partial\tau} = -\frac{\delta S[\phi]}{\delta\phi(x,\tau)} + \sqrt{2}\eta(x,\tau)$$

Cf. reverse denoising process

$$\frac{\partial\phi(x,t)}{\partial t} = -g^2(t)s_{\hat{\theta}}(\phi(x,t),t) + g(t)\eta(x,t)$$

→ Well trained model: $s_{\hat{\theta}}(\phi(x,0),0;\beta) \simeq -\frac{\delta S(\phi;\beta)}{\delta\phi}$

- Physics-conditioned sampler
  - In case $S(\phi;\beta) = \beta\tilde{S}(\phi)$, e.g., pure gauge, reference β_0 can be replaced to target β

    → Fix the drift term to $-\frac{\beta}{\beta_0}g^2(t)s_{\hat{\theta}}(\phi_t,t)$

Metropolis-Adjusted Annealed Langevin Algorithm (2)

- To ensure that a Markov chain has the correct stationary measure, introduce Metropolis-Hastings tests during the final steps

$$\psi_{\tau+1} = \phi_\tau + \alpha_i f(\phi_\tau, \tau) + \sqrt{2\alpha_i}\eta$$

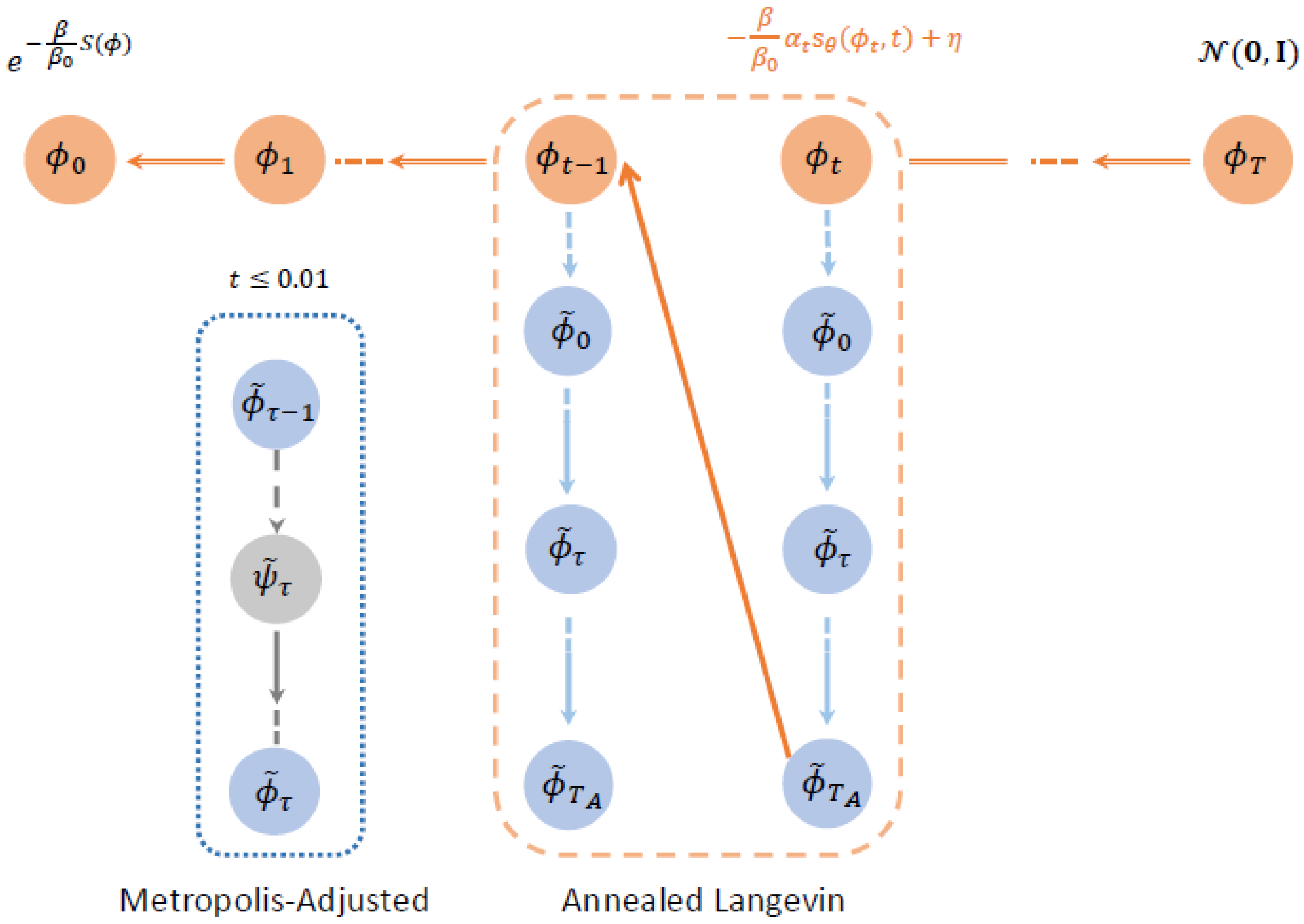
$$\phi_{\tau+1} = \begin{cases} \psi_{\tau+1} & \text{with probability } \min\left\{1, \frac{p(\psi_{\tau+1})q(\phi_\tau|\psi_{\tau+1})}{p(\phi_\tau)q(\psi_{\tau+1}|\phi_\tau)}\right\}, \\ \phi_\tau & \text{with the remaining probability,} \end{cases}$$

$$q(\phi_\tau|\psi_{\tau+1}) = \frac{1}{(4\pi\alpha_i)^{n/2}} \exp\left(-\frac{1}{4\alpha_i}\|\phi_\tau - (\psi_{\tau+1} + \alpha_i f(\psi_{\tau+1}, \tau+1))\|_2^2\right)$$

➡

$$\frac{p(\psi_{\tau+1})q(\phi_\tau|\psi_{\tau+1})}{p(\phi_\tau)q(\psi_{\tau+1}|\phi_\tau)} = \exp\Bigg(-S(\psi_{\tau+1}) - \frac{1}{4\alpha_i}\|\phi_\tau - (\psi_{\tau+1} + \alpha_i f(\psi_{\tau+1}, \tau+1))\|_2^2 + S(\phi_\tau) + \frac{1}{4\alpha_i}\|\psi_{\tau+1} - (\phi_\tau + \alpha_i f(\phi_\tau, \tau))\|_2^2\Bigg).$$

Metropolis-Adjusted Annealed Langevin Algorithm (3)



Numerical experiments

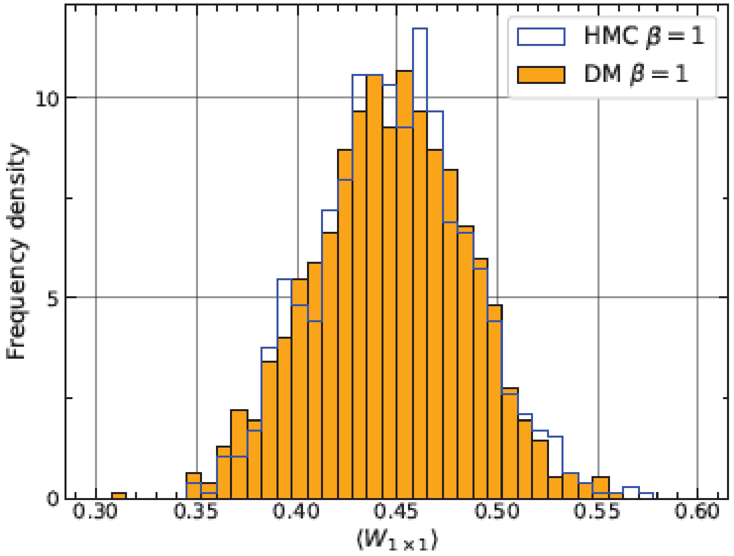
- 2D U(1) theory

$$S = -\beta \sum_x \mathrm{Re}(U_{x,\Box}) \qquad U_{x,\Box} \equiv U_{x,\mu} U_{x+\hat{\mu},\nu} U^{\dagger}_{x+\hat{\nu},\mu} U^{\dagger}_{x,\nu} \qquad U_{x,\mu} = e^{i\phi_{x,\mu}}$$

- Inverse coupling
  - $\beta = 1.0, 3.0, 5.0, 7.0, 9.0, 11.0$

- Lattice size
  - $L \times L = 8 \times 8, 16 \times 16, 32 \times 32$

Generation at different lattice sizes ($\beta = 1$)



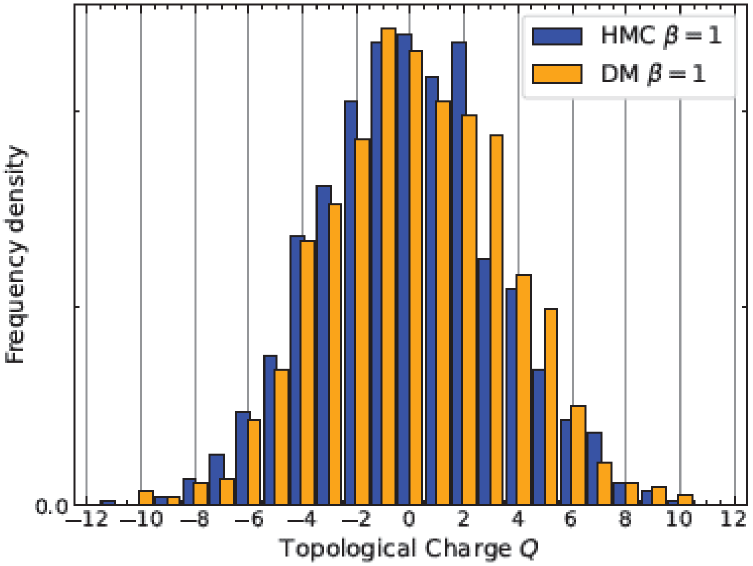
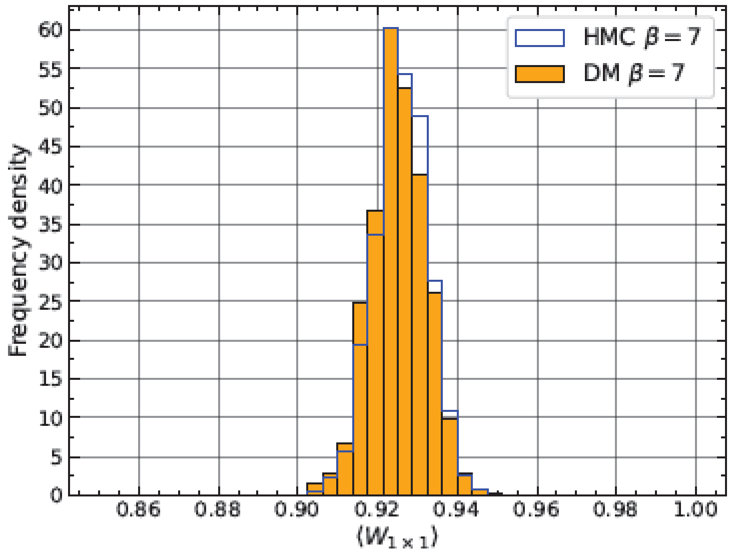


Table 1. Comparison of observables for $\beta = 1$ at different lattice sizes

Lattice Size (L)	1×1 Wilson Loop				Topological Susceptibility			
	HMC	DM	Langevin	Exact	HMC	DM	Langevin	Exact
8	0.447(72)	0.445(74)	0.443(80)	0.446	0.0402(17)	0.0413(18)	0.0418(18)	0.0406
16	0.447(37)	0.446(37)	0.444(36)	0.446	0.0416(16)	0.0422(17)	0.0421(20)	0.0406
32	0.446(18)	0.445(19)	0.445(18)	0.446	0.0428(19)	0.0415(18)	0.0412(17)	0.0406
64	0.446(9)	0.446(11)	0.445(9)	0.446	0.0426(19)	0.0427(20)	0.0420(19)	0.0406

Generation in the topology frozen region ($\beta = 7$)



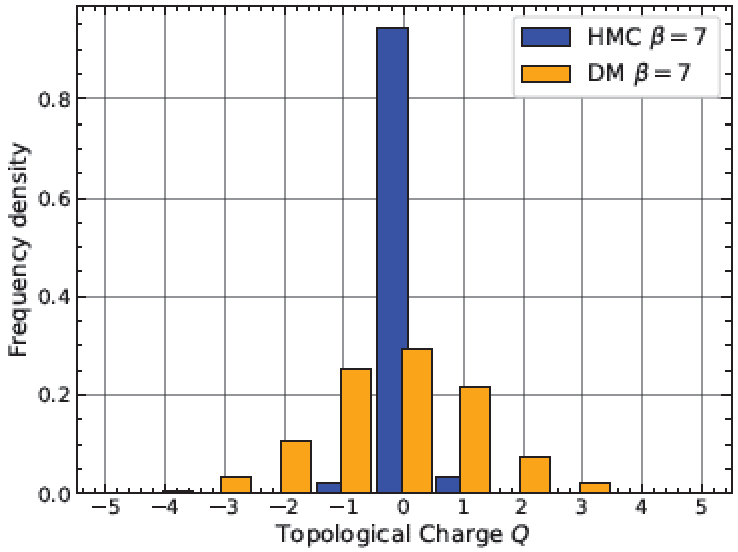


Table 2. Comparison of the $l \times l$ Wilson Loops for $L = 16, \beta = 7$

Loop size (l)	HMC	DM	Langevin	Exact
1	0.926(7)	0.926(7)	0.924(6)	0.926
2	0.737(31)	0.737(32)	0.730(34)	0.734
3	0.510(67)	0.496(72)	0.489(73)	0.498
4	0.311(97)	0.283(96)	0.283(106)	0.290

Generation at different lattice sizes and couplings

Table 3. Comparison of observables for $\beta = 7$ at different lattice sizes

Lattice Size (L)	1×1 Wilson Loop				Topological Susceptibility			
	HMC	DM	Langevin	Exact	HMC	DM	Langevin	Exact
8	0.927(13)	0.926(13)	0.921(13)	0.926	0.00006(3)	0.0040(12)	0.0143(5)	0.0040
16	0.926(7)	0.926(7)	0.924(6)	0.926	0.00013(2)	0.0043(5)	0.0131(5)	0.0039
32	0.926(3)	0.925(4)	0.924(4)	0.926	0.00013(2)	0.0040(4)	0.0137(6)	0.0039

Table 4. Comparison of observables for $L = 16$ at different couplings

coupling (β)	1×1 Wilson Loop				Topological Susceptibility			
	HMC	DM	Langevin	Exact	HMC	DM	Langevin	Exact
3	0.811(17)	0.811(17)	0.809(17)	0.810	0.0096(4)	0.0114(6)	0.0106(14)	0.0111
5	0.894(9)	0.894(9)	0.891(10)	0.894	0.0048(2)	0.0058(5)	0.0075(3)	0.0057
7	0.926(7)	0.926(7)	0.924(6)	0.926	0.00013(2)	0.0045(5)	0.0131(5)	0.0039
9	0.944(3)	0.942(4)	0.940(6)	0.942	0	0.0031(4)	0.0154(7)	0.0029
11	0.954(3)	0.953(4)	0.950(5)	0.953	0	0.0025(3)	0.0165(13)	0.0024