Physics-Conditioned Diffusion Models for Lattice Gauge Theory

Speaker: **Qianteng Zhu**

Collaborators: Gert Aarts, Wei Wang, Kai Zhou, **Lingxiao Wang**

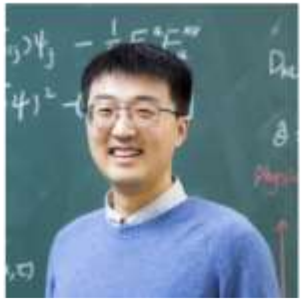
ArXiv: 2502.05504

27 February 2025

AI+HEP East Asia, Daejeon

Diffusion Models for LQFT

理化学研究所
RIKEN



Lingxiao Wang

Qianteng Zhu

SHANGHAI JIAO TONG UNIVERSITY

Swansea University
Prifysgol Abertawe

Gert Aarts

Diaa E. Habibi

Shiyang Chen

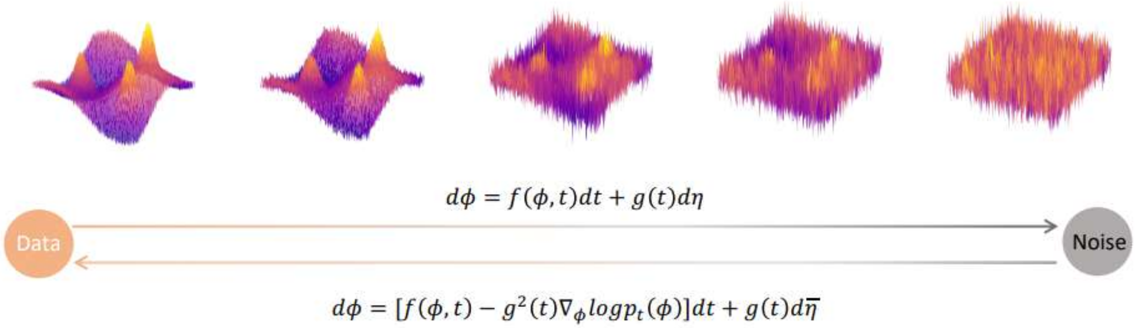
Chuanyang Li

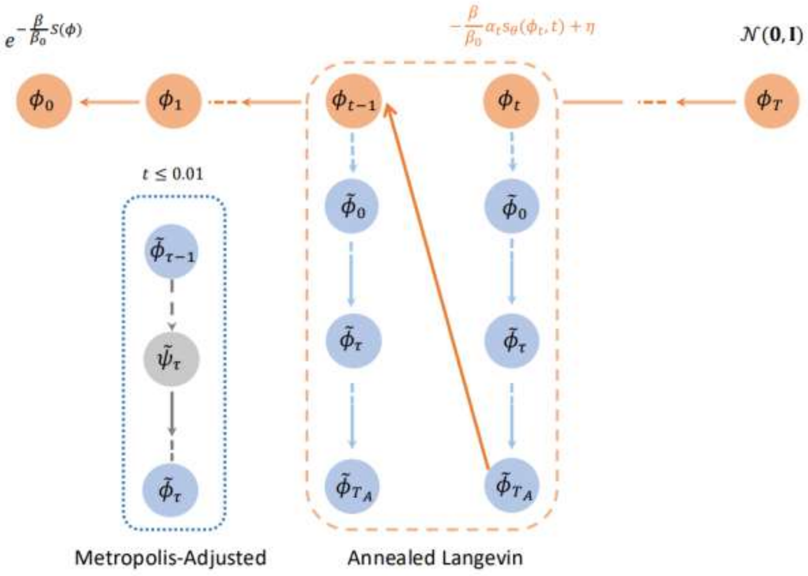
香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Kai Zhou

Outline

- Lattice Gauge theory
- Diffusion Models
- Sampler and Net Architecture
- Results
- Summary and Outlooks

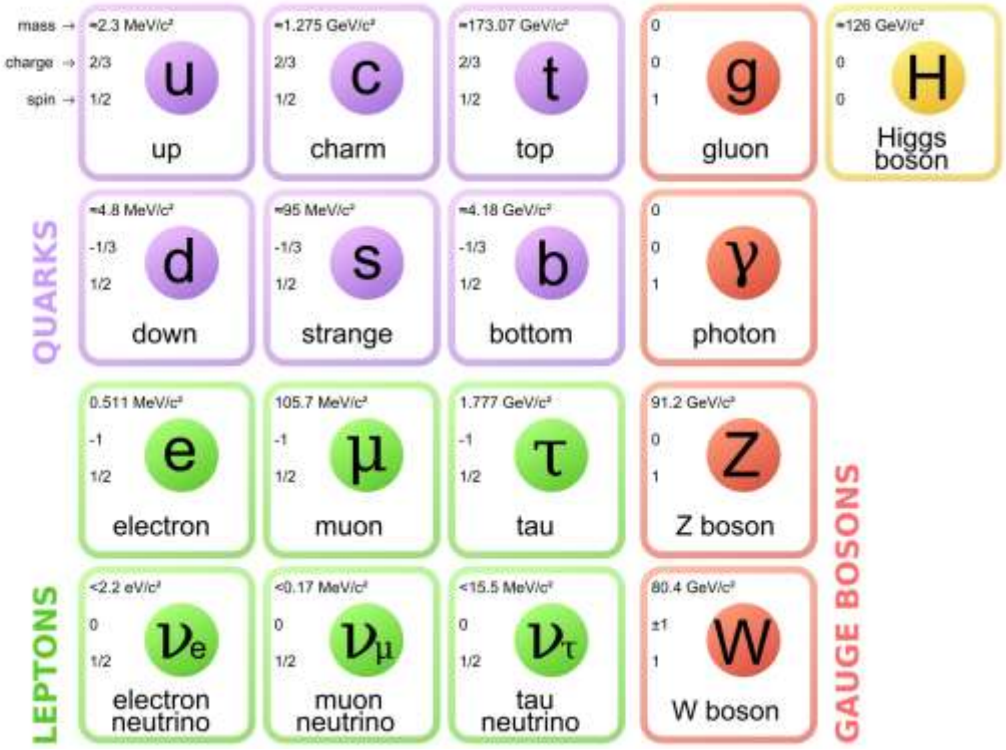




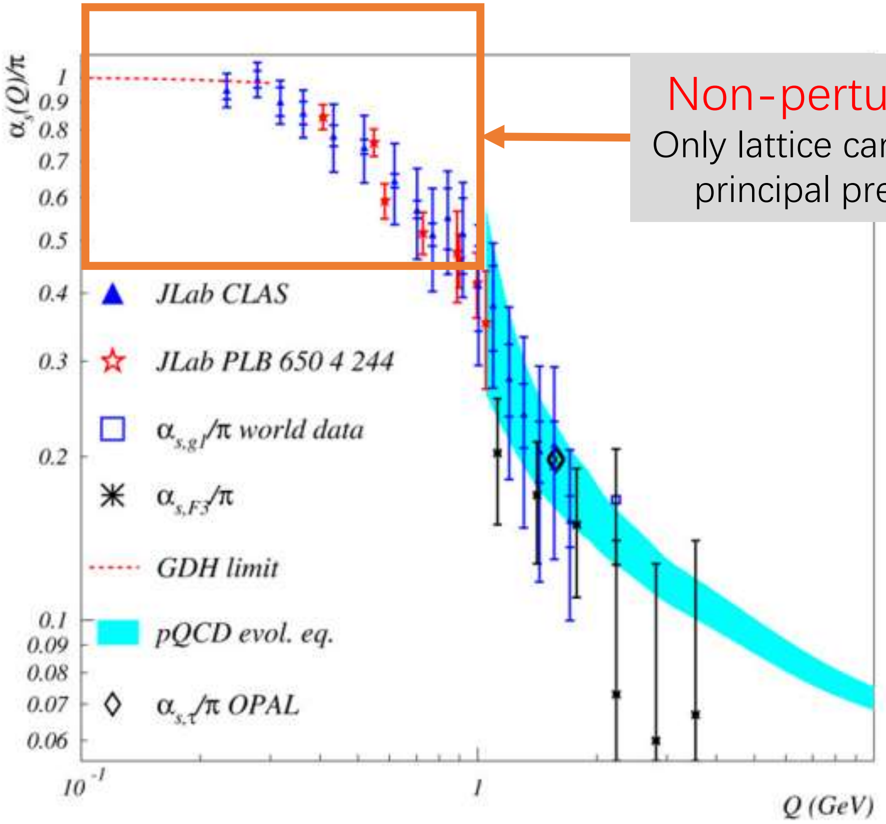
Lattice Gauge Theory: Gauge And Lattice

Standard model of particle physics

$$Standard\ model = SU(3) \otimes SU(2) \otimes U(1)$$



QCD coupling with energy scale



Non-perturbative
Only lattice can give first principal prediction

Physics Letters B 665 (2008) 349–351

Lattice Gauge Theory: Action

Gauge theory on lattice

Gauge field as links

$$U_\mu(n) = \exp(iaA_\mu(n))$$

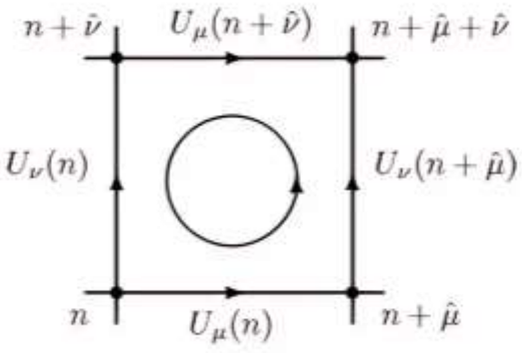
Gauge invariant Wilson loop

$$U_{\mu\nu}(n) = U_\mu(n)\, U_\nu(n+\hat{\mu})\, U_\mu(n+\hat{\nu})^\dagger U_\mu(n)^\dagger$$

Wilson Gauge action

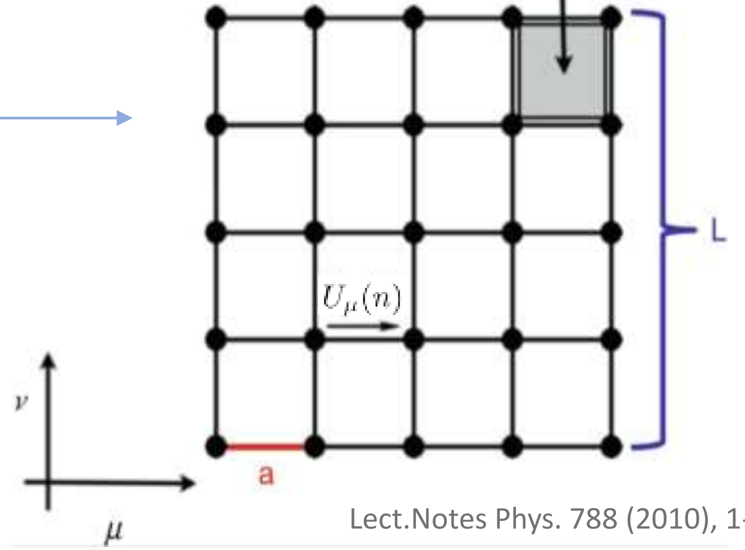
$$S[U] = \frac{\beta}{N}\sum_n \sum_{\mu<\nu} \mathrm{Re}\, Tr(1 - U_{\mu\nu}(n))$$

Gauge invariant plaquette



SU(N) Matrix

$U_{\mu\nu}(n)$



Lect.Notes Phys. 788 (2010), 1-343

Lattice Gauge Theory: Generating Configurations

Observables and configurations

Observable (configuration average):

$$\langle O\rangle = \frac{1}{Z}\int \mathcal{D}[U]e^{-S_G[U]}O[U] \;\; with \;\; Z = \int \mathcal{D}[U]e^{-S_G[U]}$$

First thing:

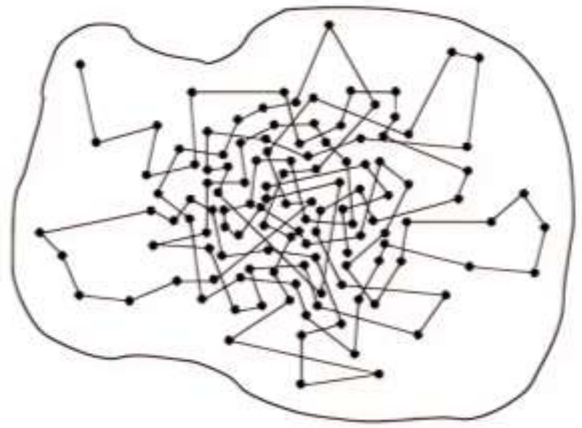
Generate configurations

$$dP[U] = \frac{1}{Z}e^{-S[U]}dU$$

$$Z = \int dU e^{-S[U]}$$

Traditional Algorithm

Markov chain Monte Carlo



$$P(i \to j) = \min\left(1, \frac{T(j \to i)e^{-S[j]}}{T(i \to j)e^{-S[i]}}\right)$$

Lattice Gauge Theory: Critical Slowing Down (CSD)

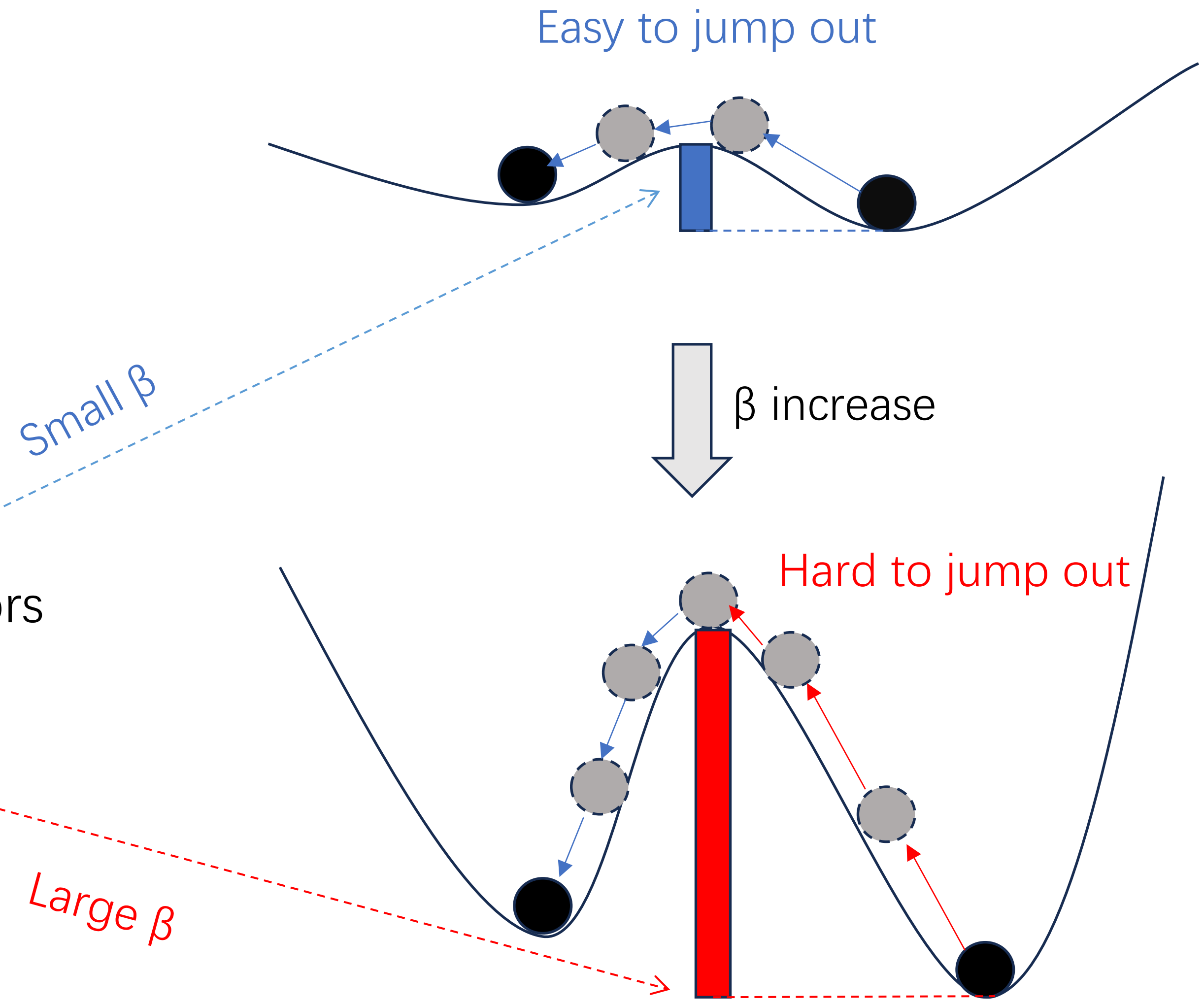
Transition between modes

$$P(i \to j) = \min\left(1, \frac{T(j \to i)e^{-S[j]}}{T(i \to j)e^{-S[i]}}\right)$$

Action wall
prevents transition between separated sectors

$$S[U] = \frac{\beta}{N}\sum_{n}\sum_{\mu<\nu} \mathrm{Re}\,Tr(1 - U_{\mu\nu}(n)) \propto \beta$$

CSD: Nucl.Phys.B Proc.Suppl. 17 (1990) 93-102,
Phys.Rev.Lett. 58 (1987) 86-88,
Phys.Lett.B 594 (2004) 315-323



Lattice Gauge Theory: Topological Freezing

Lattice definition and properties

Topological charge on Lattice for 2d U(1) :

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

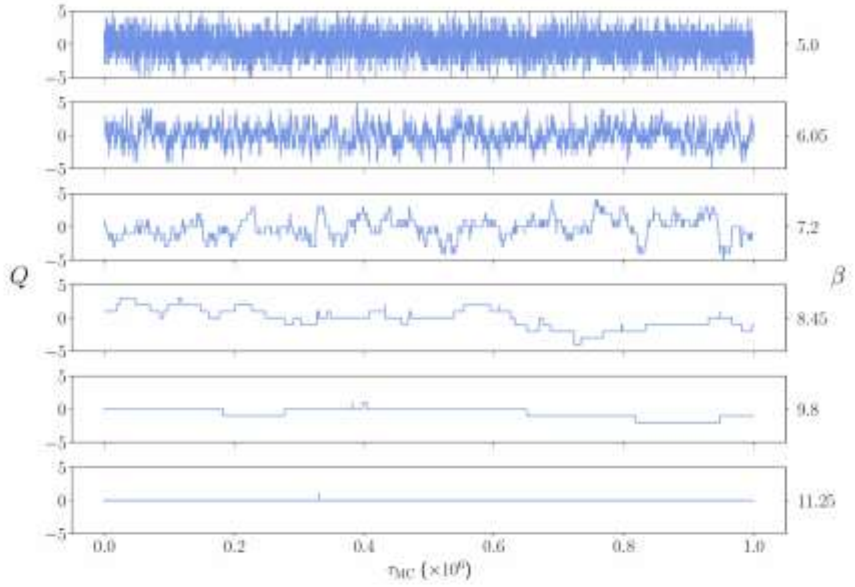
Atiyah-Singer index theorem:

- Q must be integer
- Related to left- and right-handed zero modes

$$Q = n_- - n_+$$

Separated topological sectors

Traditional HMC is low in efficiency



Eur.Phys.J.C 81 (2021) 10, 873

Improvements are needed

Lattice Gauge Theory: Mitigating CSD

Our new proposal

Only train model once, we can generate at

- Different couplings
- Different sizes

Our previous attempts

- **Diffusion for phi4 lattice theory**
  - JHEP 05 (2024) 060 [2309.17082 [hep-lat]]
  - NeurIPS workshop 2023 “ML and the Physical Sciences” [2311.03578 [hep-lat]]
- **Gauge Field**
  - NeurIPS workshop 2024 “ML and the Physical Sciences” [2410.21212 [hep-lat]]
- **Complex action**
  - Lattice 2024 [2412.01919 [hep-lat]]

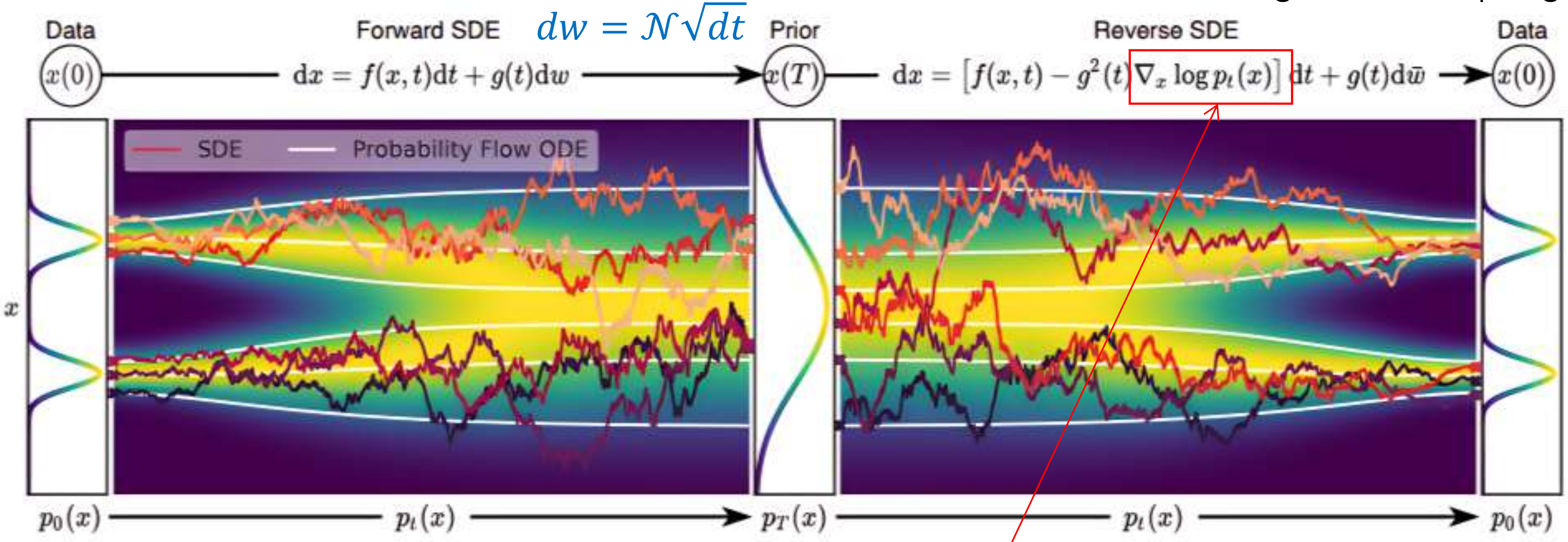
Other attempts

- **Normalizing flow for phi4 and U(1) lattice theory**
  - Phys. Rev. Lett. 125, 121601 (2020)[2003.06413 [hep-lat]]
- **Normalizing flow for lattice Schwinger model**
  - Phys. Rev. D. 106. 014514 (2022)[2202.11712 [hep-lat]]
- **Stochastic normalizing flow**
  - JHEP 07 (2022) 015 [2210.03139[hep-lat]]
- **Stochastic normalizing flow for SU(3)**
  - LATTICE 2024 (2025) 040 [2409.18861 [hep-lat]]
- **Winding HMC**
  - Eur. Phys. J. C 81, 873 (2021) [2106.14234 [hep-lat]]

Diffusion Model: Score Matching

(This term is missing when sampling)

$dw = \mathcal{N}\sqrt{dt}$

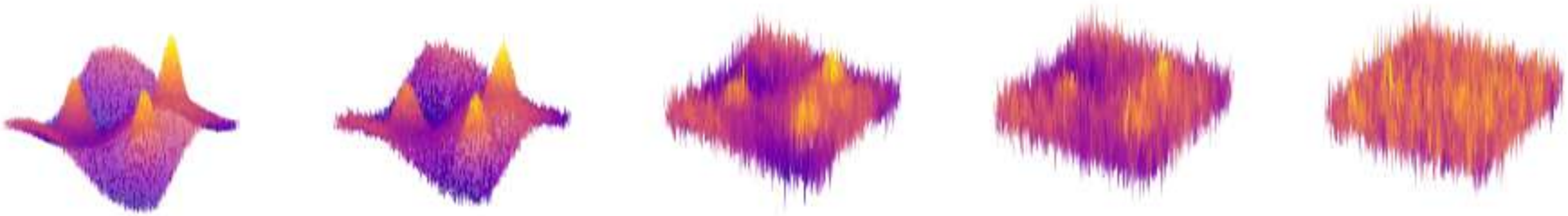


A neural network is trained to fit the missing term

$$\mathcal{L}_\theta = \sum_{i=1}^{N} \sigma_i^2 \mathbb{E}_{p_0(\phi_0)} \mathbb{E}_{p_i(\phi_i|\phi_0)} \left[\left\| \boxed{s_\theta(\phi_i, t)} - \nabla_{\phi_i} \log p_i(\phi_i \mid \phi_0) \right\|_2^2 \right]$$

Diffusion Model: Forward/Reverse SDE

$p_0[\phi] \longleftrightarrow p_{0.1}[\phi] \longleftrightarrow p_{0.2}[\phi] \longleftrightarrow p_t[\phi] \cdots \longleftrightarrow p_T[\phi]$



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

$f = 0$

Data

t is noise scale

$d\eta = \mathcal{N}\sqrt{dt}$

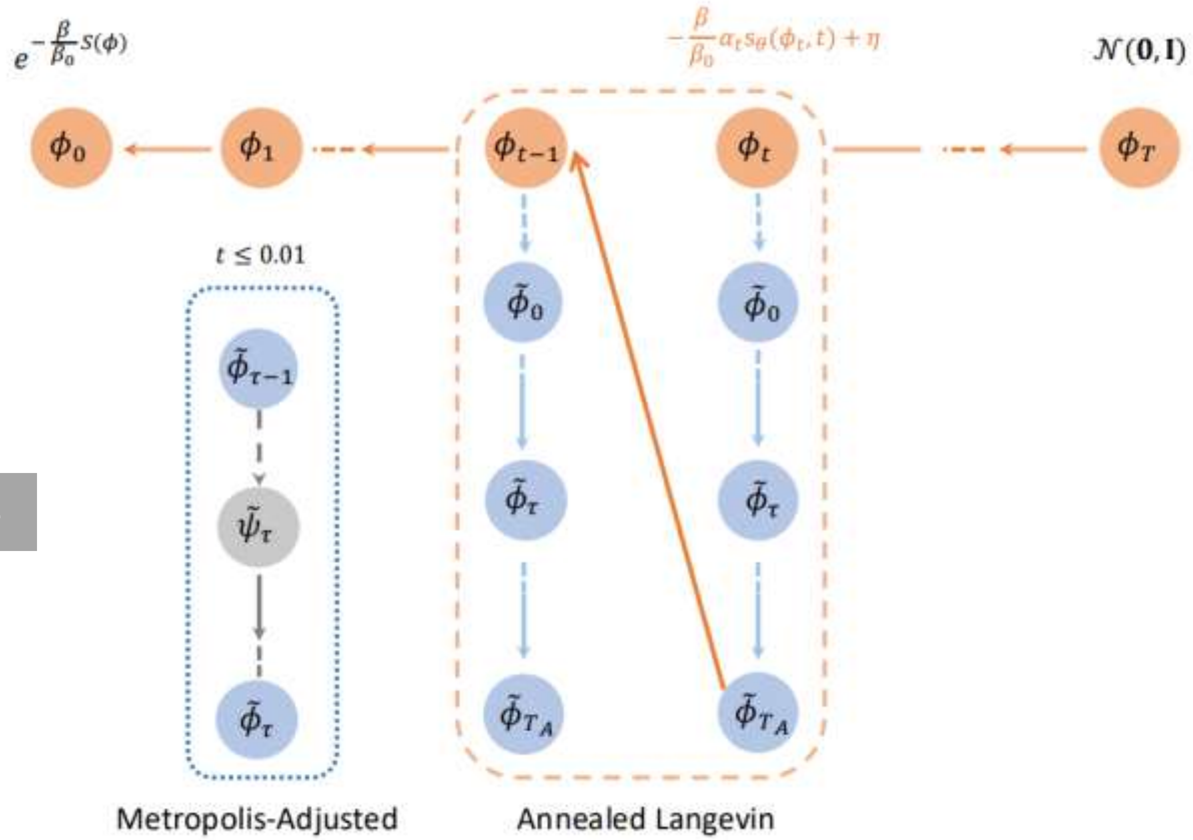
Noise

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

Approximated by Score net

Sampler: Metropolis-Adjusted Annealed Langevin

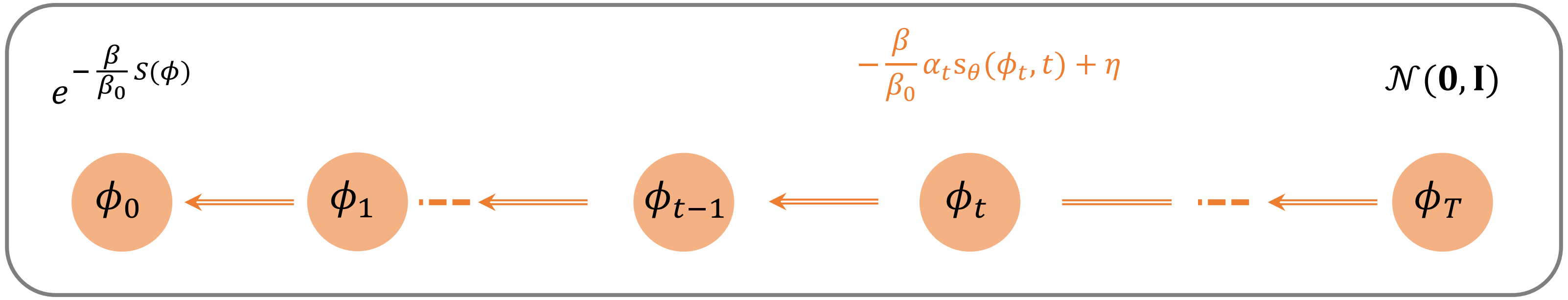
1. Horizontal : Euler-Maruyama scheme gradually decrease noise



3. Detailed balance

2. Vertical: ensure equilibrium at each noise scale

Sampler(horizontal): Score And Action



Objective to minimize:

$$\frac{1}{2}\mathbb{E}_{p_{data}}\left[\left\|s_\theta(\phi) - \nabla_\phi \log p_{data}(\phi)\right\|_2^2\right]$$

$$q_\sigma(\tilde{\phi}) \triangleq \int q_\sigma(\,\tilde{\phi} \mid \phi\,)p_{data}(\phi)d\phi$$

Equivalent to minimize:

$$\frac{1}{2}\mathbb{E}_{q_\sigma(\tilde{\phi}|\phi)p_{data}}\left[\left\|s_\theta(\tilde{\phi}, t) - \nabla_{\tilde{\phi}} \log q_\sigma(\,\tilde{\phi} \mid \phi\,)\right\|_2^2\right]$$

Critical behavior of trained score function:

$$s_\theta(\phi, t \to 0) = \nabla_x \log q_\sigma(\phi) = \nabla_\phi \log p_{data}(\phi) = -\frac{\partial S}{\partial \phi} \propto \beta$$

Critical relation of score functions:

$$s_{\theta^*}(\phi, t \to 0, \beta = \beta_{target}) = \frac{\beta_{target}}{\beta_0} \cdot s_\theta(\phi, t \to 0, \beta = \beta_0)$$

Sampler(vertical): Stochastic Quantization

Stochastic Quantization

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

$$\langle \eta(x,\tau) \rangle = 0, \quad \langle \eta(x,\tau)\eta(x',\tau') \rangle = 2\alpha\delta(x-x')\delta(\tau-\tau')$$

τ: fictitious time, α: diffusion constant

L. Wang, et al. JHEP 05, 060 (2024)

Fokker-Plank Equation

$$\frac{\partial P[\phi,\tau]}{\partial \tau} = \alpha \int d^n x \left\{ \frac{\delta}{\delta \phi} \left(\frac{\delta}{\delta \phi} + \frac{\delta S_E}{\delta \phi} \right) \right\} P[\phi,\tau]$$

Equilibrium solution (long-time limit),

$$P_{\text{eq}}[\phi] \propto e^{-\frac{1}{\alpha} S_E[\phi]}$$

$$\frac{\partial \phi(x,\tau)}{\partial \tau} = (\beta_{target}/\beta_0) \cdot s_{\hat{\theta}}(\phi(x,\tau), t_i) + \eta(x,t)$$

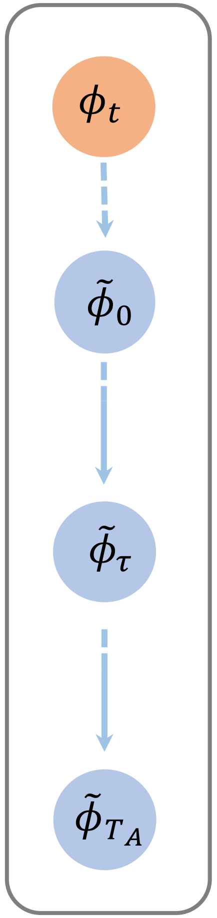
Fix t_i

$$p(\phi, t_i) \propto e^{-S_{DM}(\phi, t_i)}$$

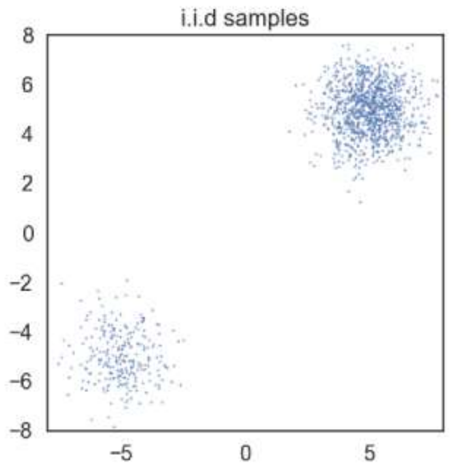
$$\partial S_{DM}(\phi, t_i)/\partial \phi = -\frac{\beta}{\beta_0} \hat{s}_\theta(\phi, t_i)$$

$$S_{DM}(\phi, t_i \to 0) \approx S(\phi, \beta)$$

Sampler(vertical): Annealing To Separated Modes

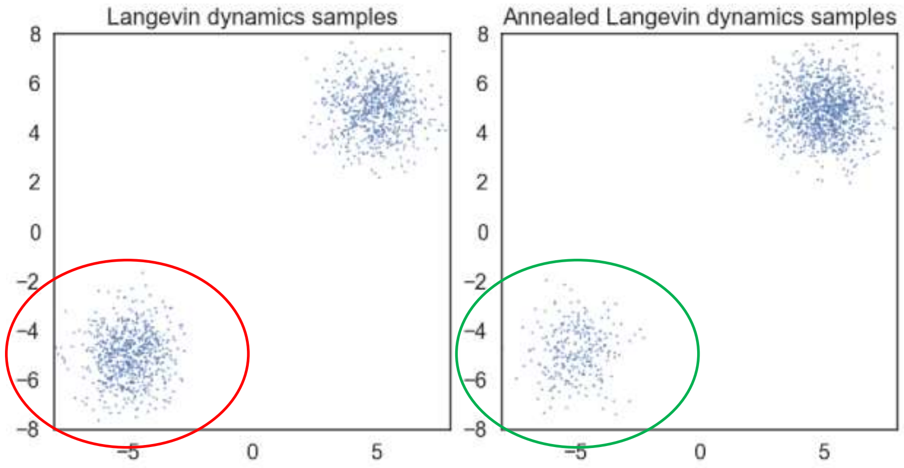


Data has separated modes



Separated + Different weights

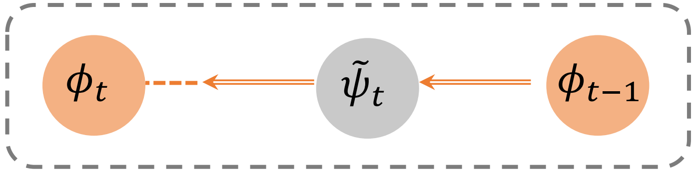
Sampler with/without Annealing



Annealed Langevin catches distribution better

1907.05600

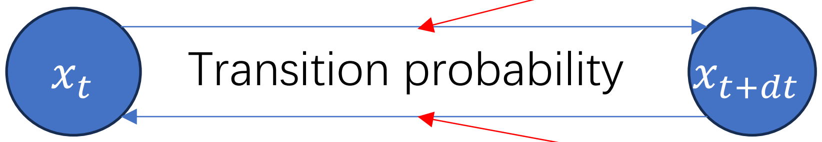
Sampler: Metropolis-Hasting Procedures



Equilibrium on Markov chain

x_{t+dt} and x_t are close to equilibrium

$$x_{t+dt} = x_t + drift(x_t, t) \cdot dt + \boxed{\sqrt{2dt} \cdot dw}$$



$$x_t = x_{t+dt} + drift(x_{t+dt}, t + dt) \cdot dt + \boxed{\sqrt{2dt} \cdot dw}$$

Metropolis-Hasting for diffusion

Known target distribution $p(\phi) = \frac{1}{Z} e^{-S[\phi]}$

proposal: $x_{t+dt} = x_t + drift(x_t, t) \cdot dt + \sqrt{2dt} \cdot dw$

Accept according to: $\left\{ \min \left\{ 1, \frac{p(x_{t+1}) q(x_t | x_{t+1})}{p(x_t) q(x_{t+1} | x_t)} \right\} \right.$

$$q(x' \mid x) \propto \exp\left(-\frac{1}{4dt} \| x' - x - drift(x, t) \cdot dt \|_2^2 \right)$$

Net Architecture : What Does U-Net Need?

Locality !

Local definition of Action

$$S[U] = \frac{\beta}{N}\sum_{n}\sum_{\mu<\nu} \mathrm{Re}\,Tr(1 - U_{\mu\nu}(n))$$

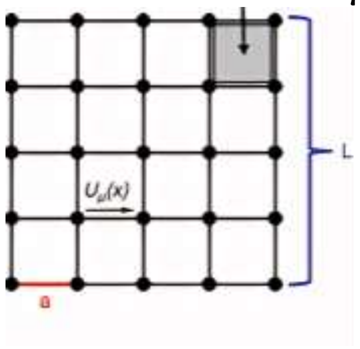
*fully convolutional network:
size independent

$$\mathrm{d}\mathbf{x}_t = [\boldsymbol{f}(\mathbf{x}_t, t) - g(t)^2 \boxed{\mathbf{s}_\theta(\mathbf{x}_t, t)}]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}_t$$

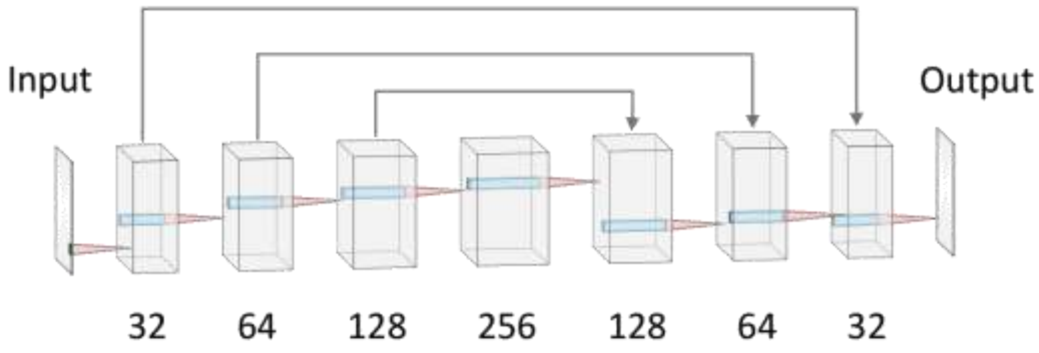
At least including one surrounding grid

$$\frac{\partial \phi(x,\tau)}{\partial \tau} = -\boxed{\frac{\delta S_E[\phi]}{\delta \phi(x,\tau)}} + \eta(x,\tau)$$

$$U_{\mu\nu}(n) = U_\mu(n)\, U_\nu(n+\hat{\mu})\, U_\mu(n+\hat{\nu})^\dagger U_\mu(n)^\dagger$$



3*3 convolutional Kernel



Circular padding

Results: Training Performance

Diffusion model and HMC setups

Training data:

- Inverse coupling **β =1**
- Lattice size **L=16** in 2-dim
- 30720 configurations from HMC
    - 200 pre-equilibrium steps
    - 128 updates
    - Data augmentation by gauge transformation
- Diffusion coefficient $\sigma_t = 25^t$
- Observables:
    - 1×1 Wilson loops

$U_{\mu\nu}(n) = U_\mu(n)\, U_\nu(n+\hat{\mu})\, U_\mu(n+\hat{\nu})^\dagger U_\mu(n)^\dagger$

- Topological charge and susceptibility

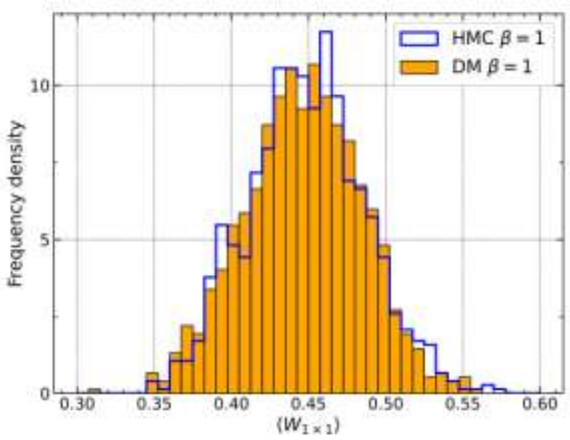
$Q = \sum_n q(n) = \frac{1}{2\pi} \sum_n \arg U_{12}(n)$

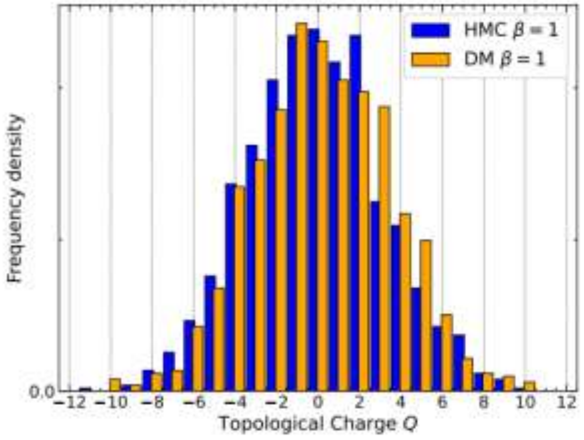
$\chi_Q = \langle Q^2/V \rangle$

Performance on training condition

1024 configurations sampled at β=1, L=16

Lattice Size (L)	1 × 1 Wilson Loop				Topological Susceptibility			
	HMC	DM	Langevin	Exact	HMC	DM	Langevin	Exact
16	0.447(37)	0.446(37)	0.444(36)	0.446	0.0416(16)	0.0422(17)	0.0421(20)	0.0406



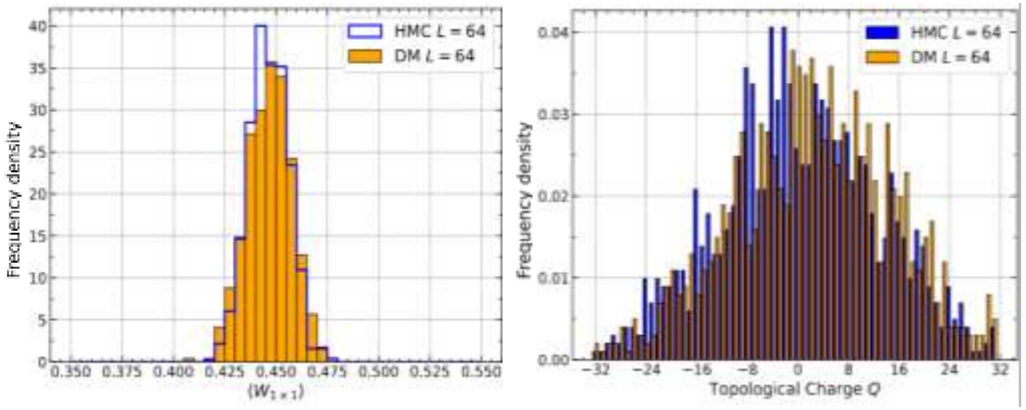


Results: Extrapolations

Extrapolation to larger size

1024 configurations sampled at β=1, L=64

Lattice Size (L)	1×1 Wilson Loop				Topological Susceptibility			
	HMC	DM	Langevin	Exact	HMC	DM	Langevin	Exact
64	0.446(9)	0.446(11)	0.445(9)	0.446	0.0426(19)	0.0427(20)	0.0420(19)	0.0406

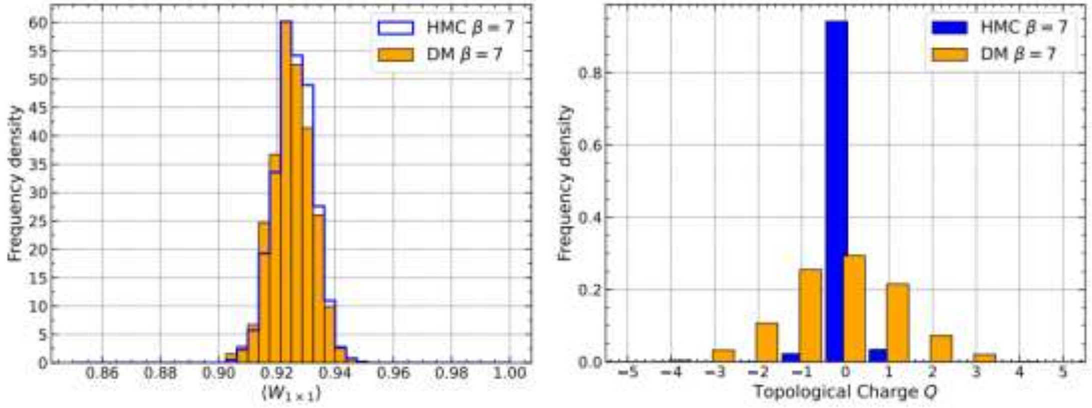


*Model is only trained at β=1, L=16

Extrapolation to larger coupling

1024 configurations sampled at β=7, L=16

coupling (β)	1×1 Wilson Loop				Topological Susceptibility			
	HMC	DM	Langevin	Exact	HMC	DM	Langevin	Exact
7	0.926(7)	0.926(7)	0.924(6)	0.926	0.00013(2)	0.0045(5)	0.0131(5)	0.0039



Results: Larger Wilson Loops

Higher-order observables

N×N Wilson loop:

- Quark potential
- Confinement

$$\langle W(C)\rangle = \left\langle Tr\ P\ \exp(\oint A_\mu dx^\mu)\right\rangle$$

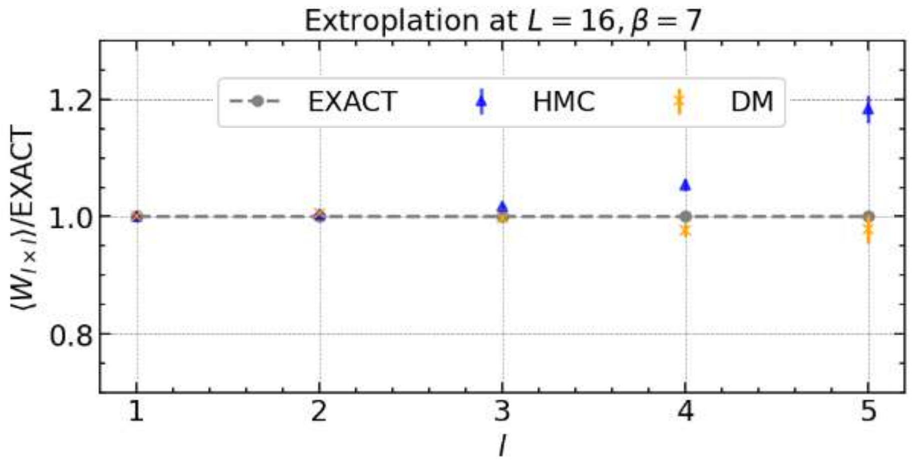
$$= C\exp(-t\boxed{V(r)})$$

Extrapolation to larger coupling

1024 configurations sampled at β=7, L=16

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



Results: Mitigating CSD

Autocorrelation time

Integrated autocorrelation time $$\tau_{X,\mathrm{int}} = \frac{1}{2} + \sum_{t=1}^{N} \Gamma_X(t)$$

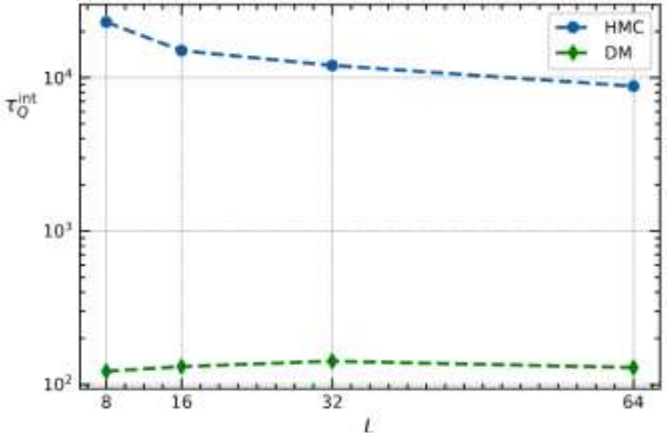
Normalized Correlation function $$\Gamma_X(t) \equiv \frac{C_X(t)}{C_X(0)}$$

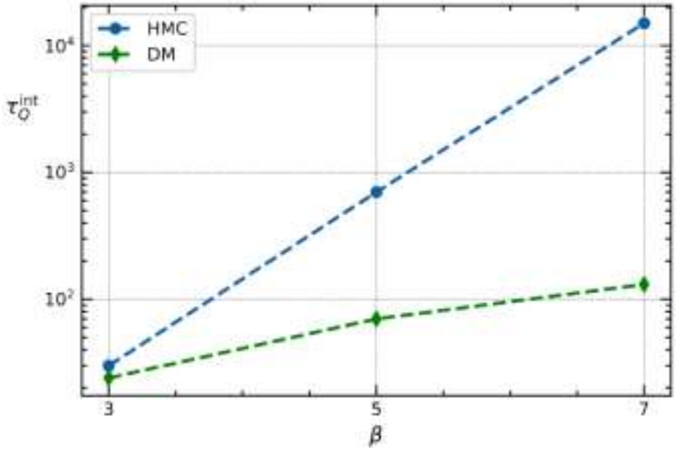
Autocorrelation function $$C_X(t) = C_X(X_i,\, X_{i+t})$$

$$C_X(X_i,\, X_{i+t}) = \langle X_i\, X_{i+t}\rangle - \langle X_i\rangle \langle X_{i+t}\rangle$$

Single Markov chain

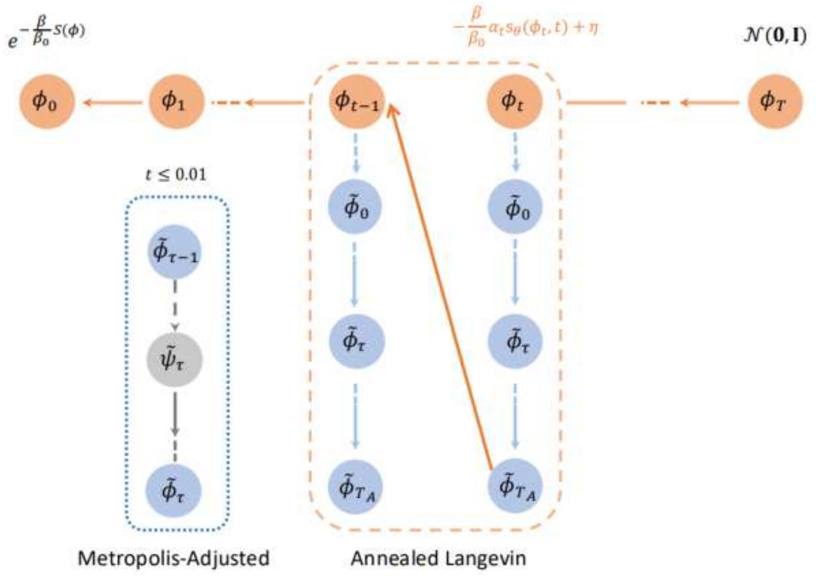
1024 configurations sampled at β=7, L=16





Summary and Outlooks

- DM as Stochastic quantization
  - Estimate gradients of data distribution
  - Better observables
- Physics-conditioned sampler
  - Annealing to different couplings and sizes
  - Detailed balance for exactness
- Future works
  - Non-Abelian gauge group
  - Fermions with non-local action
  - Renormalization group



Back-Ups

Numerical Tables

Q. Zhu et al., arXiv: 2502.05504

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

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

Numerical Tables

Q. Zhu et al., arXiv: 2502.05504

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)	0.0045(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

Sampler

Q. Zhu et al., arXiv: 2502.05504

Algorithm 2 Metropolis-Adjusted Annealed Langevin Sampler

Require: $\{t_i\}_{i=1}^{N_T}, \epsilon, N_A$

Initialize ϕ_0
for $i \leftarrow 1$ to N_T **do**
    $\alpha_i \leftarrow \epsilon \cdot g \cdot t_i^2 / t_{N_T}^2$ ▷ α_i is the step size.
    $\tilde{\phi}_0 \leftarrow \phi_{i-1}$
    for $\tau \leftarrow 1$ to N_A **do**
        Draw $\eta \sim \mathcal{N}(0, I)$
        Propose $\tilde{\psi}_\tau = \tilde{\phi}_{\tau-1} + \frac{\beta}{\beta_0}\alpha_i s_\theta(\tilde{\phi}_{\tau-1}, t_i) + \sqrt{2\alpha_i}\eta$
        if $t_i < 0.01$ **then**
            $P_i \leftarrow \min\left\{1, \frac{p(\tilde{\psi}_{\tau+1})q(\tilde{\phi}_\tau|\tilde{\psi}_{\tau+1})}{p(\tilde{\phi}_\tau)q(\tilde{\psi}_{\tau+1}|\tilde{\phi}_\tau)}\right\}$
            Draw $k \sim \text{Uniform}(0, 1)$
            if $k_i < P_i$ **then**
                $\tilde{\phi}_\tau \leftarrow \tilde{\psi}_\tau$
            else
                $\tilde{\phi}_\tau \leftarrow \tilde{\phi}_{\tau-1}$
            end if
        else
            $\tilde{\phi}_\tau \leftarrow \tilde{\psi}_\tau$
        end if
    end for
    $\phi_i \leftarrow \tilde{\phi}_{N_A}$
end for
return ϕ_{N_T}

Single Markov-Chain

Q. Zhu et al., arXiv: 2502.05504

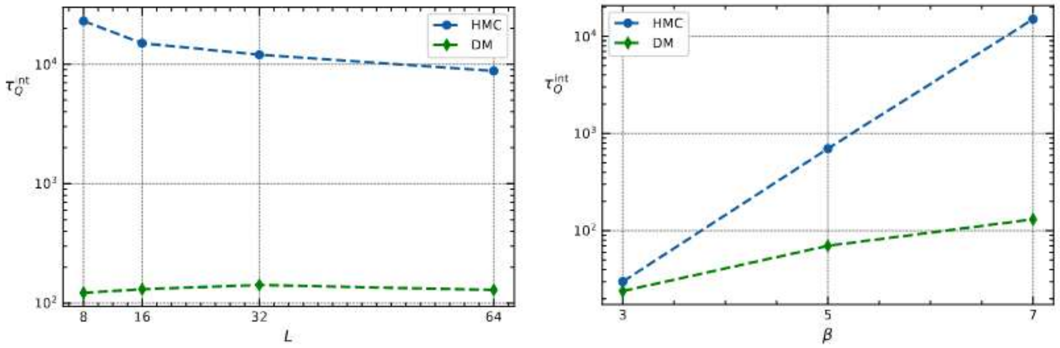


Figure 8. Comparison of integrated autocorrelation time in a single Markov chain for HMC and DM at $\beta = 7$ in different littice sizes L (left), and at $L = 16$ in different couplings β (right)

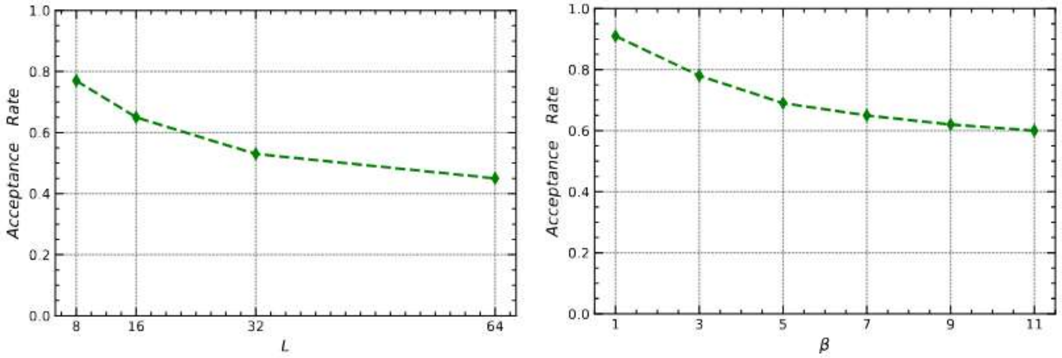


Figure 9. Dependence of acceptance Rate in a single Markov chain of DM at $\beta = 7$ on different littice sizes L (left), and at $L = 16$ on different couplings β (right)

Why SDE, Not ODE Sampler?

Stochastic Differential Equation

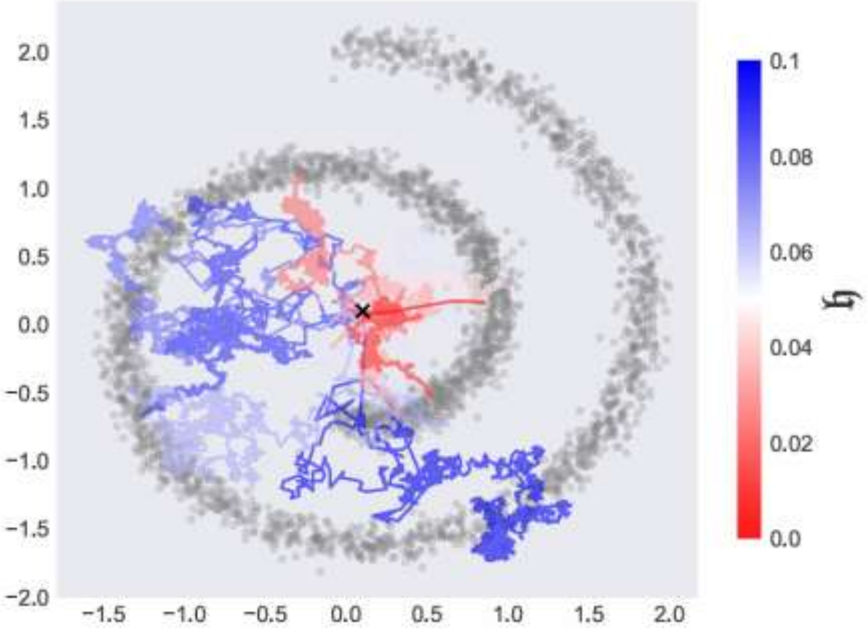
$$\mathrm{d}\mathbf{x}_t = [\boldsymbol{f}(\mathbf{x}_t, t) - g(t)^2 \boldsymbol{s}_\theta(\mathbf{x}_t, t)]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}_t$$

$$\mathrm{d}\mathbf{x}_t = \left[\boldsymbol{f}(\mathbf{x}_t, t) - \frac{1+\mathfrak{h}}{2} g(t)^2 \boldsymbol{s}_\theta(\mathbf{x}_t, t)\right]\mathrm{d}t + \sqrt{\mathfrak{h}}\, g(t)\mathrm{d}\bar{\mathbf{w}}_t$$

Ordinary Differential Equation

$$\mathrm{d}\mathbf{x} = \underbrace{\left\{\mathbf{f}(\mathbf{x}, t) - \frac{1}{2}\nabla \cdot [\mathbf{G}(\mathbf{x}, t)\mathbf{G}(\mathbf{x}, t)^{\mathsf{T}}] - \frac{1}{2}\mathbf{G}(\mathbf{x}, t)\mathbf{G}(\mathbf{x}, t)^{\mathsf{T}}\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}, t)\right\}}_{=:\tilde{\mathbf{f}}_{\boldsymbol{\theta}}(\mathbf{x}, t)} \mathrm{d}t$$

$$\log p_0(\mathbf{x}(0)) = \log p_T(\mathbf{x}(T)) + \int_0^T \nabla \cdot \tilde{\mathbf{f}}_{\boldsymbol{\theta}}(\mathbf{x}(t), t)\mathrm{d}t$$



2403.11262

ODE Likelihood

Q. Zhu et al., arXiv: 2502.05504

Log-likelihood

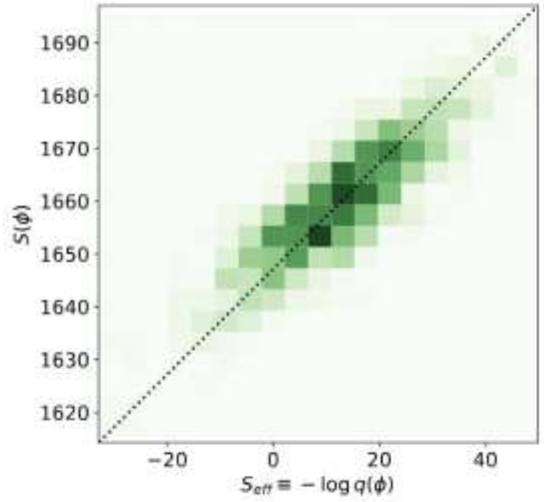
Ode likelihood estimator：

$$\log p_0(\mathbf{x}(0)) = \log p_T(\mathbf{x}(T)) + \int_0^T \nabla \cdot \tilde{\mathbf{f}}_{\boldsymbol{\theta}}(\mathbf{x}(t), t)\mathrm{d}t$$

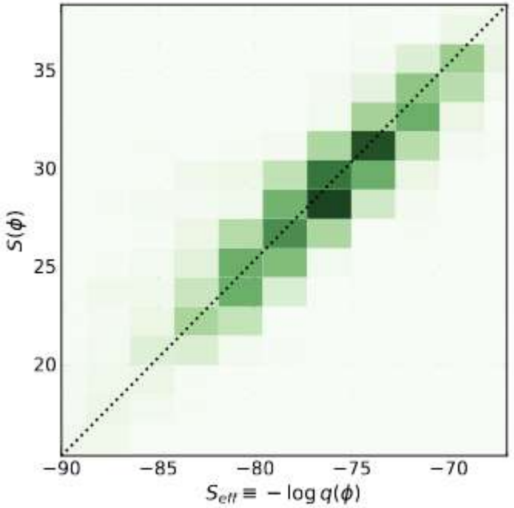
target distribution：

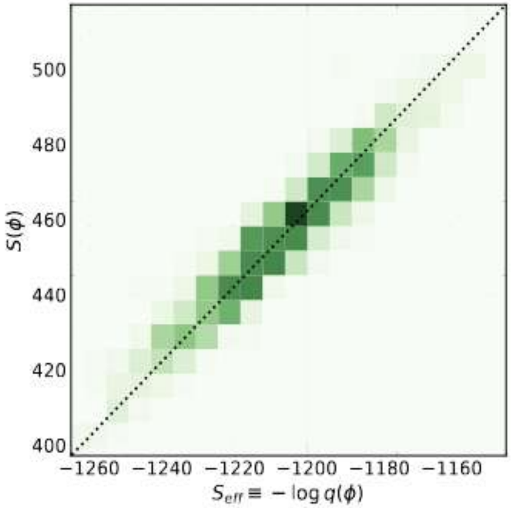
$$p(\phi) = \frac{1}{Z} e^{-S[\phi]}$$

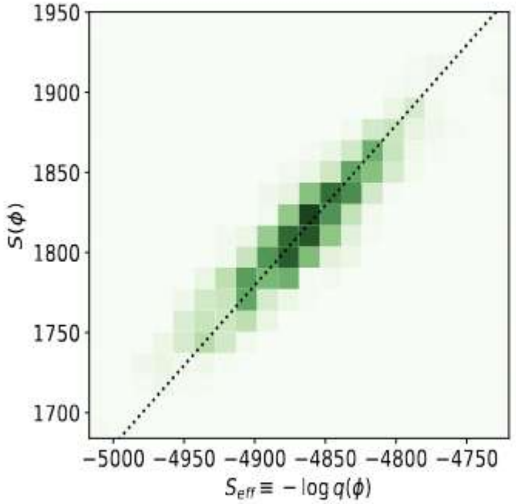
1024 configurations trained and sampled by DM at β=7, L=16



1024 configurations sampled at β=1, L=8(left),32(middle),64(right)



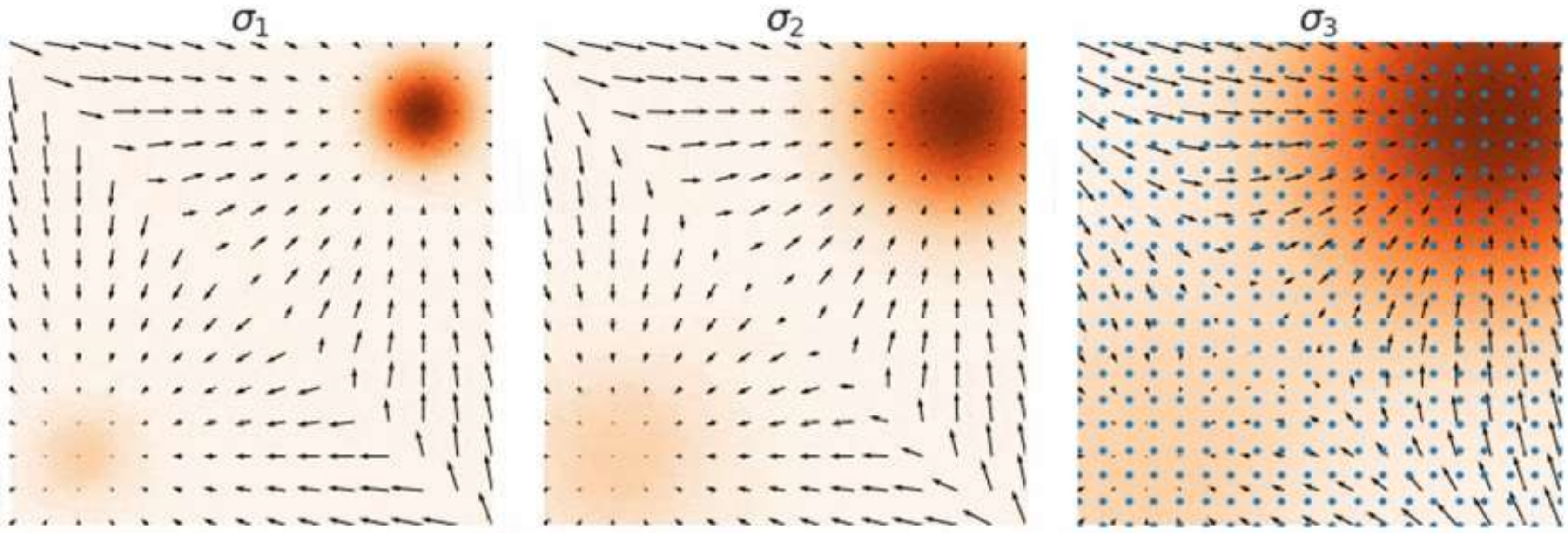




Diffusion Model: Drift Term

Drift term Score(ϕ,t) is Force
Same shape with sample

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



2D Drift map(2 variables) for different noise scales($\sigma_1 < \sigma_2 < \sigma_3$)