Diffusion models and stochastic quantisation in lattice field theory

Gert Aarts

with Lingxiao Wang, Kai Zhou and Diaa Habibi

JHEP 05 (2024) 060 [2309.17082 [hep-lat]]
NeurIPS 2023[2311.03578 [hep-lat]]

Swansea University
Prifysgol Abertawe

Lattice 2024, Liverpool, July 2024

Swansea ML-LFT group

Chanju Park, Diaa Habibi, Shiyang Chen, GA, Biagio Lucini, Matteo Favoni

Presentations

- Monday:

Exploring Generative Networks for Manifolds with Non-Trivial Topology	*Shiyang Chen*	12:15 - 12:35
Diffusion models learn distributions generated by complex Langevin dynamics	*Diaa Eddin Habibi*	15:15 - 15:35

- Wednesday:

Random Matrix Theory for Stochastic Gradient Descent	*Chanju Park*	11:15 - 11:35

- Poster:

Matteo Favoni

Towards the application of random matrix theory to neural networks

- Friday:

An introduction to topological data analysis for lattice field theory	*Jeffrey Giansiracusa*	11:55 - 12:15
Topological Data Analysis of Monopole Currents in U(1) Lattice Gauge Theory	*Xavier Crean*	12:15 - 12:35
Topological Data Analysis, Monopoles and Colour Confinement in SU(3) Yang-Mills	*Biagio Lucini*	12:35 - 12:55

Generative AI and LFT

- in recent years, rich programme to apply methods of AI/ML to lattice field theories

- in particular, employ ML to generate LFT configurations beyond standard (well-tested and well-understood) approaches, such as HMC

- why? reduce auto-correlations, critical slowing down, and because it is really cool!

two schemes: devise ML algorithms to approximate

- target distribution, $\sim e^{-S}$, directly, e.g. normalising flow
- underlying distribution by learning from data, e.g. diffusion models

Diffusion models

- very popular ML method: used in DALL-E, Stable Diffusion, …
- used to generate "fake" images on the internet 🎉
- based on concepts of non-equilibrium physics

Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, Surya Ganguli *Proceedings of the 32nd International Conference on Machine Learning*, PMLR 37:2256-2265, 2015.

a dolphin in an astronaut suit on saturn, artstation — a propaganda poster depicting a cat dressed as french emperor napoleon holding a piece of cheese — a teddy bear on a skateboard in times square

- can we use DMs in LFT?
- physics connection with existing methods?
- competitive with other approaches?

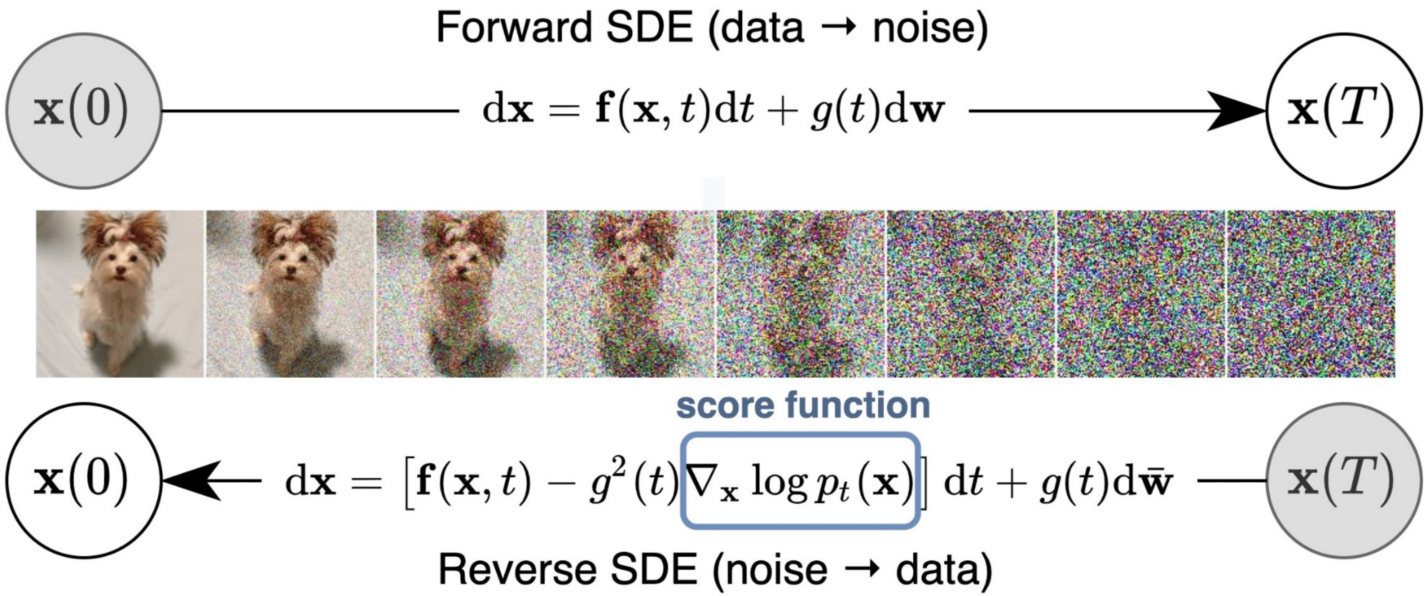
JHEP 05 (2024) 060 [2309.17082 [hep-lat]]

Generative AI: Diffusion models

underlying model is based on Brownian motion, i.e. Langevin or SDEs

- start with data set of images
- make the images more blurred by applying noise (forward process)
- learn steps in this process

  … and then revert it
- create new images from noise



https://theaisummer.com/diffusion-models/

Prior and target distributions

- **target distribution** describes the data: not known in real-world applications

$$P(\text{cats}, \text{dogs}) \sim \exp[-S(\text{cats}, \text{dogs})] \quad ?$$

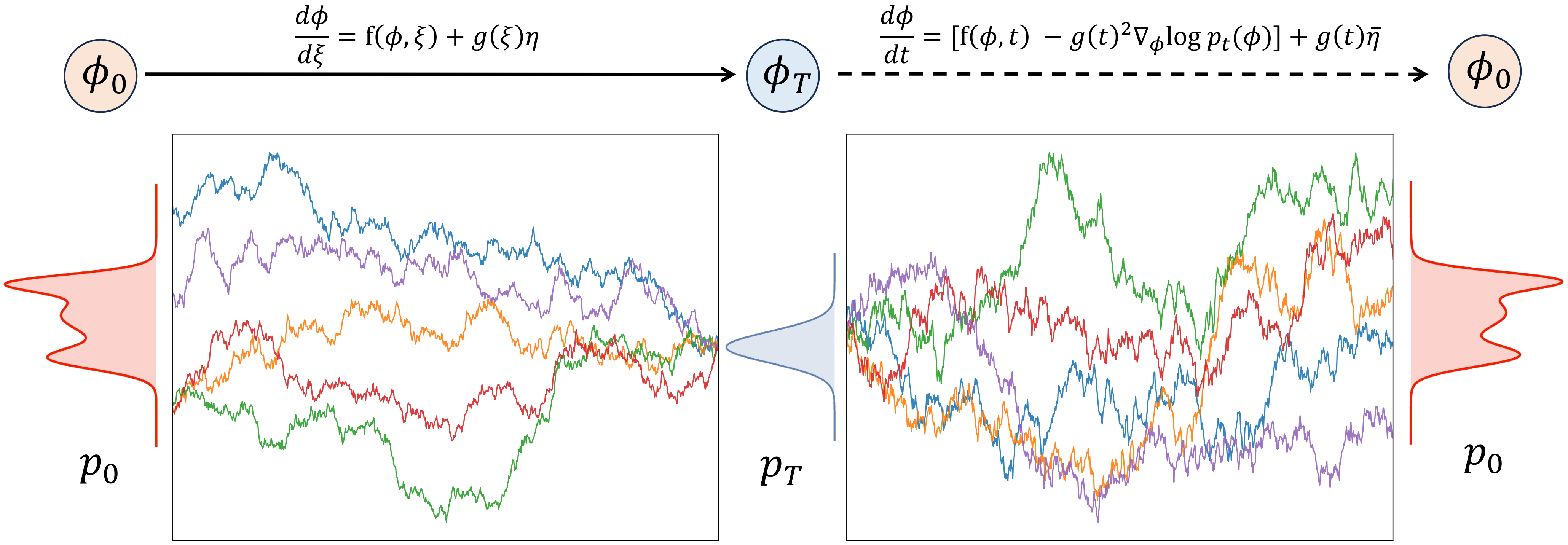
- learn $\text{grad} \log P$ under application of noise with increasing variance: score matching
- in practice achieved using some ML architecture (not discussed here)

- **prior distribution**: simple, e.g. Gaussian
- backward or denoising process: apply $\text{grad} \log P$ to retrieve target distribution

- **after training** (score determination): generate new images using backward process

Prior and target distributions

- in pictures: p_0 is target (non-trivial), p_T is the prior (easy)



DMs and stochastic quantisation

- dynamics of backward process is stochastic process with time-dependent drift and noise variance

$$\frac{\partial\phi(x,\tau)}{\partial\tau} = g^2(\tau)\nabla_\phi \log P(\phi;\tau) + g(\tau)\eta(x,\tau)$$

- if $P(\phi;\tau) = \frac{e^{-S(\phi,\tau)}}{Z}$ such that $\nabla_\phi \log P(\phi,\tau) = -\nabla_\phi S(\phi,\tau)$

- then

$$\frac{\partial\phi(x,\tau)}{\partial\tau} = -g^2(\tau)\nabla_\phi S(\phi,\tau) + g(\tau)\eta(x,\tau)$$

- stochastic quantisation (Parisi & Wu 1980)
- path integral quantisation via a stochastic process in fictitious time

$$\frac{\partial\phi(x,\tau)}{\partial\tau} = -\nabla_\phi S(\phi) + \eta(x,\tau)$$

DMs and stochastic quantisation

$$\frac{\partial \phi(x,\tau)}{\partial \tau} = g^2(\tau)\nabla_\phi \log P(\phi;\tau) + g(\tau)\eta(x,\tau)$$

$$\frac{\partial \phi(x,\tau)}{\partial \tau} = -\nabla_\phi S(\phi) + \eta(x,\tau)$$

similarities and differences:

✓ SQ: fixed drift, determined from known action

constant noise variance (but can be generalised using kernels)

thermalisation followed by long-term evolution in equilibrium

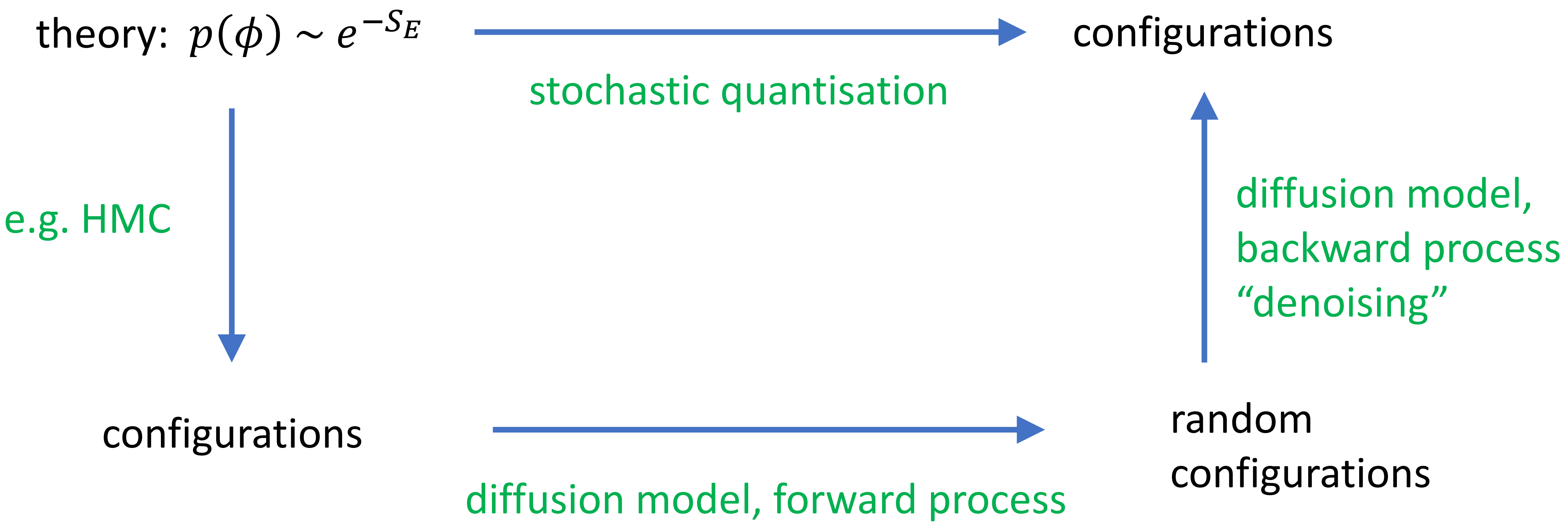
✓ DM: drift and noise variance time-dependent, learn from data

evolution between $0 \leq \tau \leq T = 1$ many short runs, very rapid thermalisation

no correlations between runs

Stochastic quantisation and diffusion models

- diffusion models as an alternative approach to stochastic quantisation

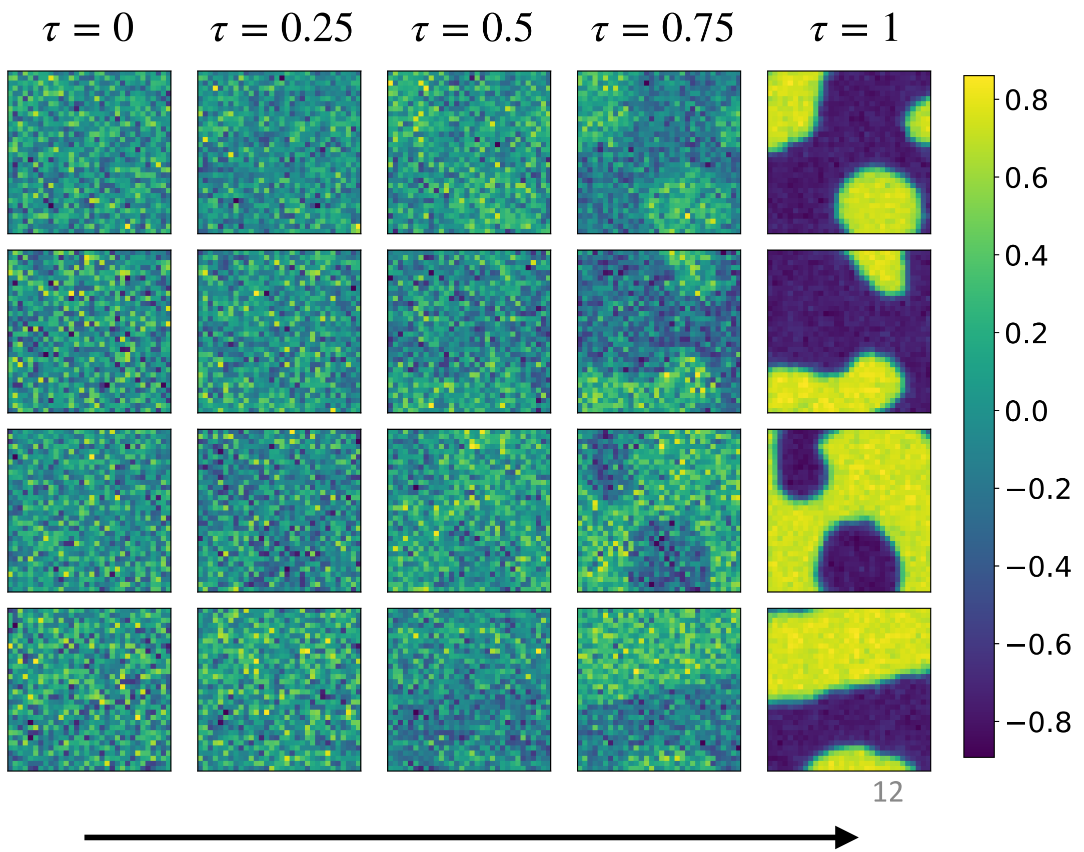


Diffusion model for 2d ϕ^4 scalar theory

- 32^2 lattice, choice of action parameters in symmetric and broken phase
- training data set generated using Hybrid Monte Carlo (HMC)
- variance expanding DM trained using U-Net architecture

generating configurations:

- broken phase
- "denoising" (backward process)
- large-scale clusters emerge, as expected



Diffusion model for 2d ϕ^4 scalar theory

generating configurations in symmetric phase

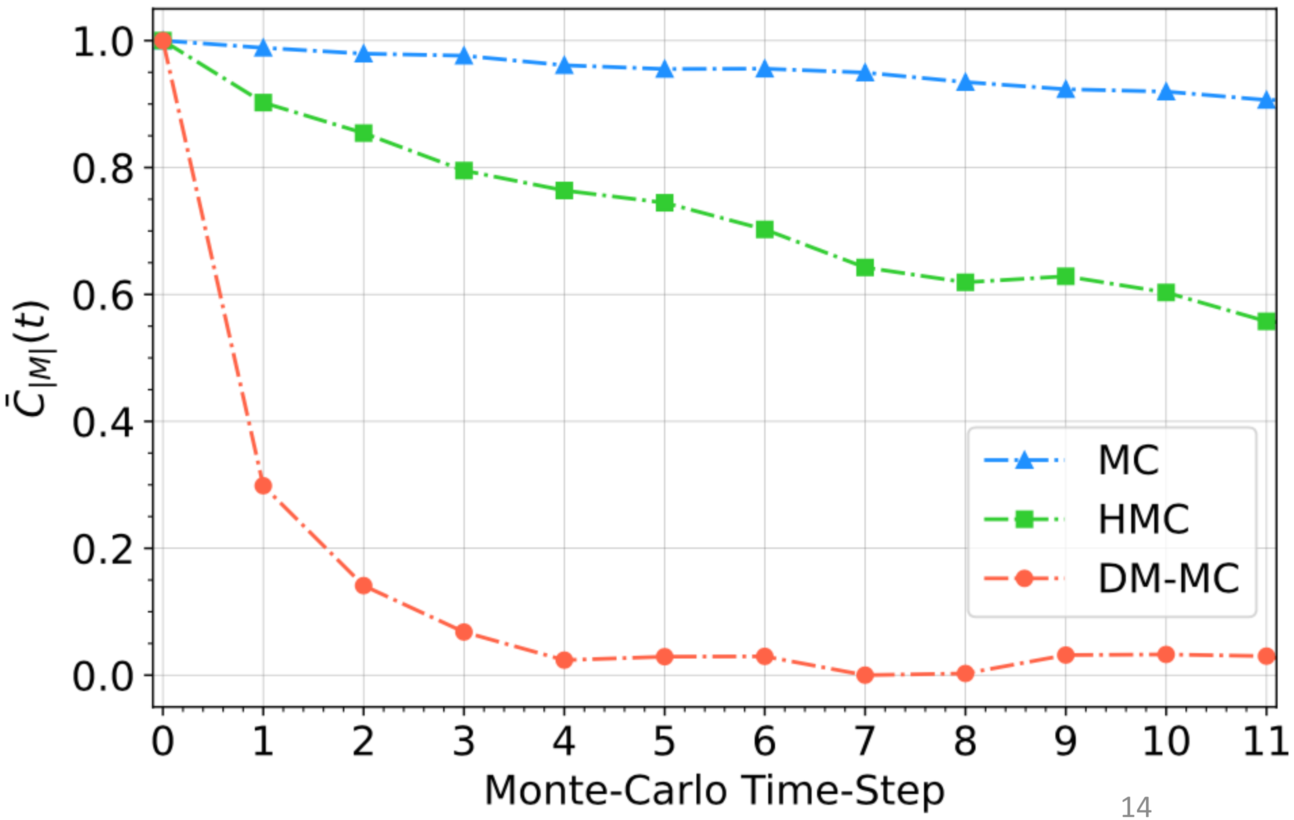
- compute magnetisation $\langle M \rangle$, susceptibility χ_2, Binder cumulant U_L
- compare with test HMC data set (with same statistics)

data-set	$\langle M \rangle$	χ_2	U_L
Training (HMC)	0.0012 ± 0.0007	2.5160 ± 0.0457	0.1042 ± 0.0367
Testing (HMC)	0.0018 ± 0.0015	2.4463 ± 0.1099	-0.0198 ± 0.1035
Generated (DM)	0.0017 ± 0.0015	2.4227 ± 0.1035	0.0484 ± 0.0959

- good agreement is observed

Diffusion model for 2d ϕ^4 scalar theory

- auto-correlation time (first rough comparison)
- normalised auto-correlation function



overall:

- proof of principle
- expected results obtained
- need to do detailed comparison of precision, speed and scalability

Evolution of drift/score in toy model

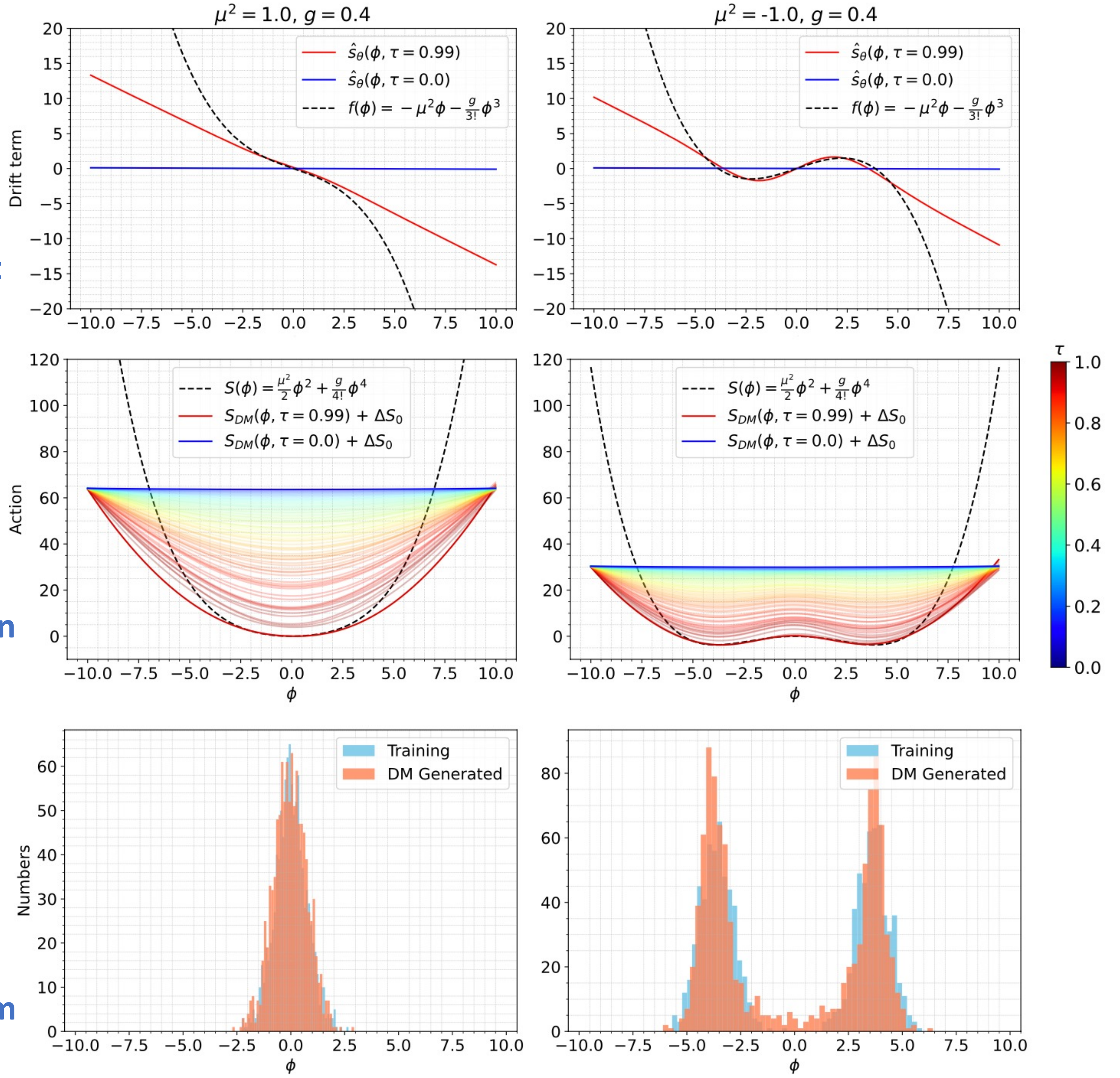
- one degree of freedom
- single/double well
- from constant action to target action as

$$0 \leq \tau \leq T = 1$$

drift

action

histogram



Stochastic quantisation: complex actions

- approach not limited to real-valued distributions/actions
- extend Langevin process to complex manifold: complex Langevin dynamics (Parisi 1981)

$$z \sim \rho(z) \in \mathbb{C} \quad \Rightarrow \quad x, y \sim P(x, y) \in \mathbb{R}$$

- convergence not guaranteed, no general solution of Fokker-Planck equation
- a posteriori justification (GA, Seiler, Stamatescu 2009, Nagata, Nishimura, Shimasaki 2016)
- recent applications in QCD (Sexty et al, 2023, 2024)
- introductory lectures (GA, 1512.05145 [hep-lat])

Complex Langevin and DMs

- distribution sampled in CL process determined by (real-valued) Fokker-Planck equation
- but FPE is not solvable generically (unlike for real Langevin)
- hence distribution, and its properties, remain elusive
- learn distribution from CL data using DMs?

see next talk by Diaa Habibi!

Summary and outlook

- diffusion models offer a new approach for ensemble generation to explore in LFT
- learn from data: requires high-quality ensembles
- use well-trained DMs to enhance statistics, beat critical slowing down, …
- can be incorporated in Markov chain, using accept/reject step

- apply to theories with fermions: DMs learn presence of fermions implicitly?
- apply to complex actions/Langevin: DMs learn elusive real-valued distributions
- …

Stochastic quantisation and diffusion models

- diffusion models as an alternative approach to stochastic quantisation