Diffusion models for lattice field theory

Gert Aarts

with Lingxiao Wang, Kai Zhou, **Qianteng Zhu**, Wei Wang and **Diaa Habibi**

L Wang, GA, K Zhou, JHEP 05 (2024) 060 [2309.17082 [hep-lat]]

GA, D Habibi, L Wang, K Zhou, Mach.Learn.Sci.Tech. **6** (2025) 2, 025004 [arXiv:2410.21212 [hep-lat]]

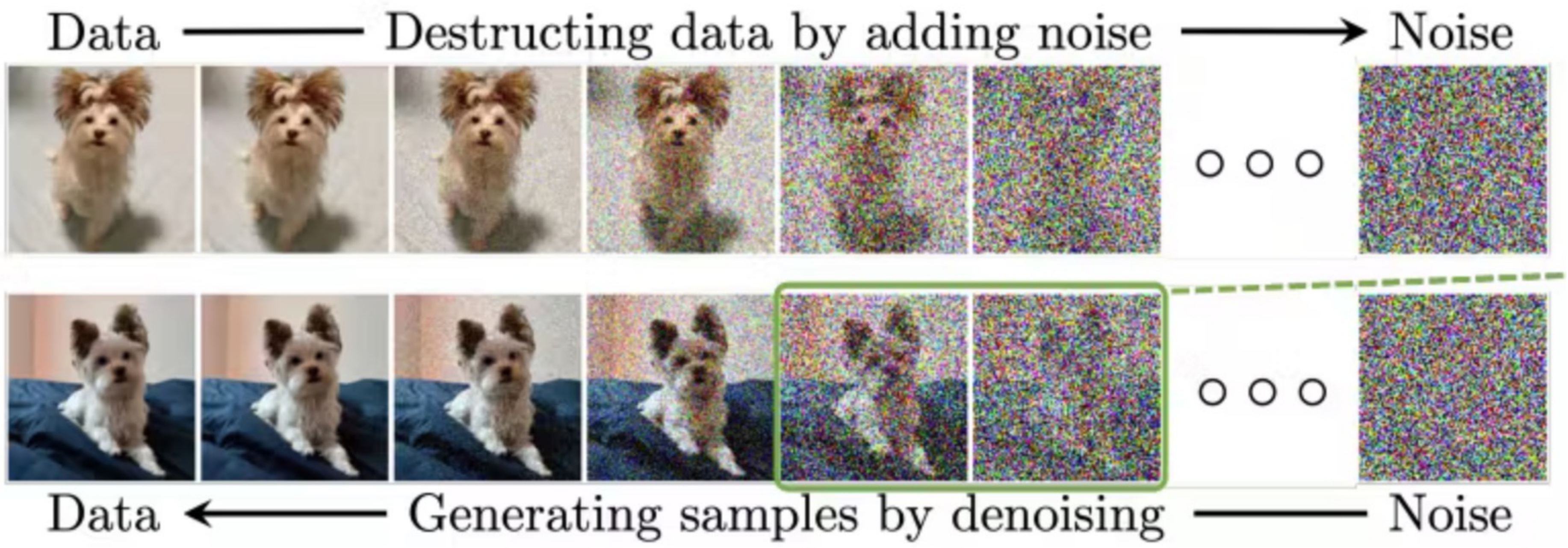
Q Zhu, W Wang, GA, K Zhou, L Wang, 2502.05504 [hep-lat]

Swansea University Prifysgol Abertawe

Wuhan Sept 2025

THE ROYAL SOCIETY

Generative AI using diffusion models



https://encord.com/blog/diffusion-models/

denoising

Generative Modeling by Estimating Gradients of the Data Distribution
Yang Song, Stefano Ermon
1907.05600 [cs.LG]

interpolation

Score-Based Generative Modeling through Stochastic Differential Equations
Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, Ben Poole, 2011.13456 [cs.LG]

Motivation: lattice field theory

- generating configurations is one of the bottlenecks in lattice field theory
- images are two-dimensional configurations from some unknown probability distribution
- machine learning algorithms are usually fast and flexible
- we know the distribution $\sim e^{-S}$: can we incorporate ML algorithms in LFT?

Flow-based generative models for Markov chain Monte Carlo in lattice field theory
MS Albergo, G Kanwar, PE Shanahan, Phys. Rev. D 100 (2019) **3**, 034515 [1904.12072 [hep-lat]]

Applications of machine learning to lattice quantum field theory
D Boyda, et al, Snowmass 2021, 2202.05838 [hep-lat]

Advances in machine-learning-based sampling motivated by lattice quantum chromodynamics
K Cranmer, G Kanwar, S Racanière, DJ Rezende, PE Shanahan, Nature Rev. Phys. **5**, 526 (2023)

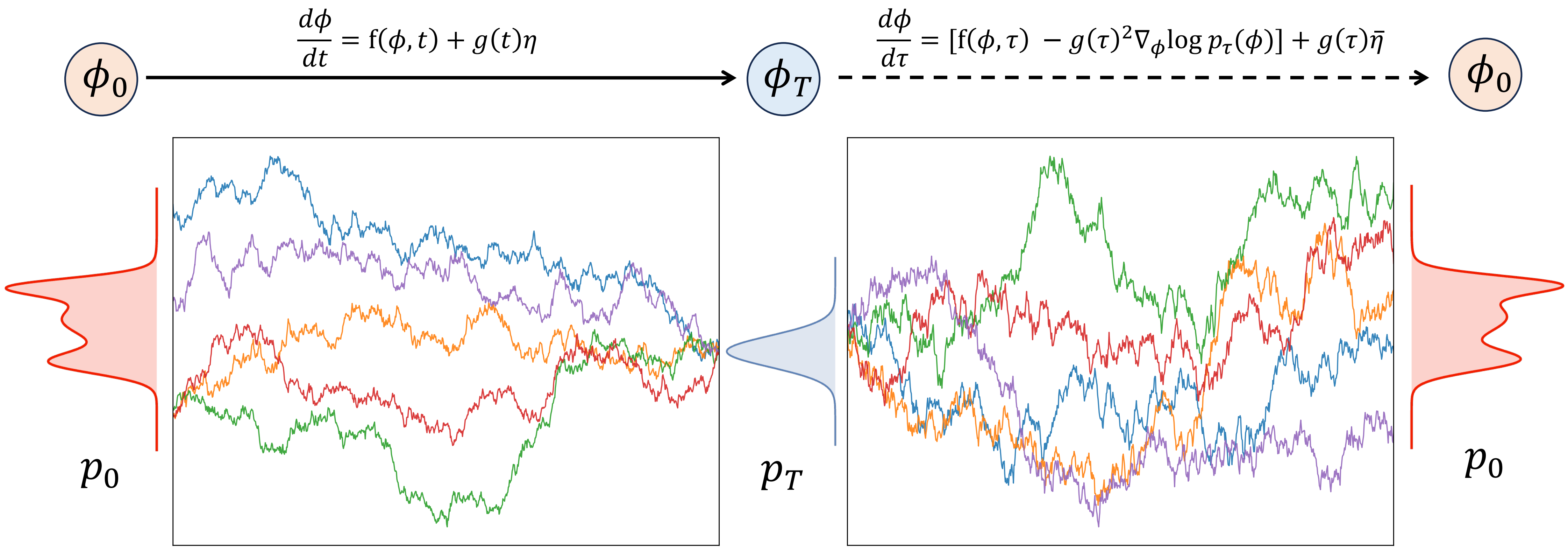
Physics-driven learning for inverse problems in QCD
G Aarts, K Fukushima, T Hatsuda, A Ipp, S Shi, L Wang, K Zhou, Nature Rev. Phys. **7** (2025) 154 [2501.05580 [hep-lat]]

Outline

- diffusion models as stochastic processes
- relation to stochastic quantisation
- theoretical analysis of evolution of cumulants
- recent variations and improvements:
  - Metropolis adjusted Langevin algorithm (MALA) / annealing / physics conditioning
  - applied to U(1) gauge theory in two dimensions

Diffusion models: prior and target distributions

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



forward process

backward process

Diffusion models and stochastic quantisation

- images/configurations are generated during backward process

- stochastic process with time-dependent drift and noise strength

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

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

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

Diffusion models and stochastic quantisation

- then $$\frac{\partial \phi(x,\tau)}{\partial \tau} = -g^2(\tau)\nabla_\phi S(\phi,\tau) + g(\tau)\eta(x,\tau)$$
- very familiar to (lattice) field theorists
- stochastic quantisation (Parisi & Wu 1980)
- path integral quantisation via a stochastic process in fictitious time $$\frac{\partial \phi(x,\tau)}{\partial \tau} = -\nabla_\phi S(\phi) + \sqrt{2}\eta(x,\tau)$$
- stationary solution of associated Fokker-Planck equation $P(\phi) \sim e^{-S(\phi)}$

Diffusion models and stochastic quantisation

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

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

similarities and differences:

✓ SQ: fixed drift, determined from known action

constant noise variance (but can be generalised using kernels)

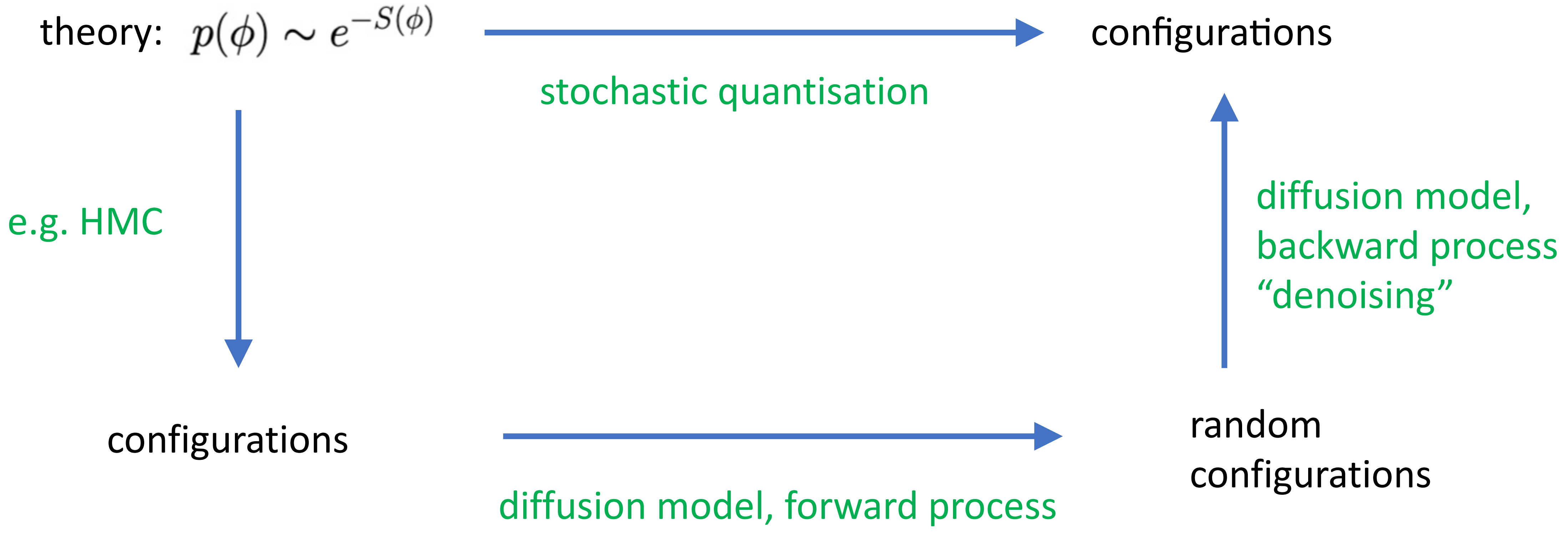
thermalisation followed by long-term evolution in equilibrium

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

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

Diffusion models and stochastic quantisation

- diffusion models as an alternative approach to stochastic quantisation



L Wang, GA, K Zhou, *JHEP* 05 (2024) 060 [2309.17082 [hep-lat]]

see also Y Hirono, A Tanaka, K Fukushima, 2403.11262

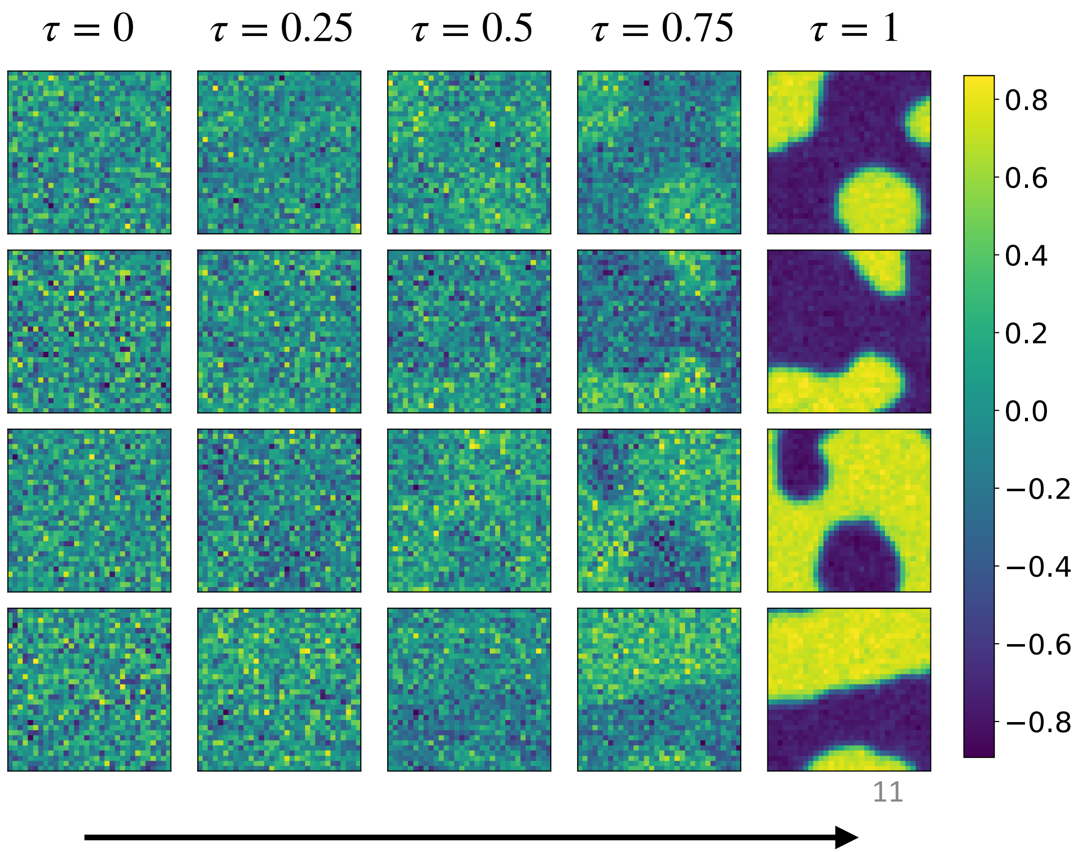
Diffusion model for 2d ϕ^4 lattice scalar theory

- 32^2 lattice, choice of action parameters in symmetric and broken phase
- training data set generated using Hybrid Monte Carlo (HMC)

- first application of diffusion models in lattice field theory

generating configurations:

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



L Wang, GA, K Zhou, *JHEP* 05 (2024) 060 [2309.17082 [hep-lat]]

Diffusion models: generation of correlations

- forward process $\dot{x}(t) = K(x(t), t) + g(t)\eta(t)$ $0 \leq t \leq T$

noise profile $g(t) = \sigma^{t/T}$

- backward process

$$x'(\tau) = -K(x(\tau), T-\tau) + g^2(T-\tau)\partial_x \log P(x, T-\tau) + g(T-\tau)\eta(\tau)$$

score

$\tau = T - t$

two main schemes

- variance-expanding (VE): no drift $K(x,t) = 0$
- variance-preserving (VP) or denoising diffusion probabilistic models (DDPMs):

linear drift $K(x(t), t) = -\frac{1}{2}k(t)x(t)$

assume first moment vanishes

$x_0 \to x_0 - \mathbb{E}_{P_0}[x_0]$

Solve forward process

- forward process $\dot{x}(t) = K(x(t), t) + g(t)\eta(t)$ $K(x(t), t) = -\frac{1}{2}k(t)x(t)$
- initial data from target ensemble $x_0 \sim P_0(x_0)$
- solution $x(t) = x_0 f(t, 0) + \int_0^t ds\, f(t, s)g(s)\eta(s)$ $f(t, s) = e^{-\frac{1}{2}\int_s^t ds'\, k(s')}$
- second moment/cumulant/variance $\kappa_2(t) = \mu_2(t) = \mu_2(0)f^2(t, 0) + \Xi(t)$

$$\Xi(t) = \int_0^t ds \int_0^t ds'\, f(t, s)f(t, s')g(s)g(s')\mathbb{E}_\eta[\eta(s)\eta(s')] = \int_0^t ds\, f^2(t, s)g^2(s)$$

$$f(t,s) = e^{-\frac{1}{2}\int_s^t ds'\, k(s')}$$

Higher-order moments and cumulants

- moments $\mu_n(t) = \mathbb{E}[x^n(t)]$ and cumulants $\kappa_n(t)$: straightforward algebra

$$\kappa_3(t) = \mu_3(t) = \kappa_3(0) f^3(t,0)$$

$$\mu_4(t) = \mu_4(0) f^4(t,0) + 6\mu_2(0) f^2(t,0)\Xi(t) + 3\Xi^2(t)$$

$$\kappa_4(t) = \mu_4(t) - 3\mu_2^2(t) = \left[\mu_4(0) - 3\mu_2^2(0)\right] f^4(t,0) = \kappa_4(0) f^4(t,0)$$

$$\kappa_5(t) = \left[\mu_5(0) - 10\mu_3(0)\mu_2(0)\right] f^5(t,0) = \kappa_5(0) f^5(t,0)$$

→ $$\kappa_{n>2}(t) = \kappa_n(0) f^n(t,0)$$

variance-expanding scheme: no drift

$$f(t,0) = 1$$

higher cumulants conserved!

$$\Xi(t) = \int_0^t ds\, f^2(t,s) g^2(s)$$

Proof to all orders

- generating functionals: average over both noise and target distributions

moments $Z[J] = \mathbb{E}[e^{J(t)x(t)}]$ cumulants $W[J] = \log Z[J]$

- noise average $$Z_\eta[J] = \mathbb{E}_\eta[e^{J(t)x(t)}] = \frac{\int D\eta\, e^{-\frac{1}{2}\int_0^t ds\, \eta^2(s) + J(t)\left[x_0 f(t,0) + \int_0^t ds\, f(t,s) g(s) \eta(s)\right]}}{\int D\eta\, e^{-\frac{1}{2}\int_0^t ds\, \eta^2(s)}}$$

- full average $$Z[J] = \mathbb{E}[e^{J(t)x(t)}] = e^{\frac{1}{2}J^2(t)\Xi(t)} \int dx_0\, P_0(x_0) e^{J(t)x_0 f(t,0)}$$

- cumulant generator $$W[J] = \log Z[J] = \frac{1}{2}J^2(t)\Xi(t) + \log \int dx_0\, P_0(x_0) e^{J(t)x_0 f(t,0)}$$

$$f(t,s) = e^{-\frac{1}{2}\int_s^t ds'\, k(s')}$$

$$\Xi(t) = \int_0^T ds\, f^2(t,s) g^2(s)$$

Proof to all orders: cumulants

- cumulant generator $$W[J] = \log Z[J] = \frac{1}{2} J^2(t)\Xi(t) + \log \int dx_0\, P_0(x_0) e^{J(t) x_0 f(t,0)}$$
- 2nd cumulant $$\kappa_2(t) = \frac{d^2 W[J]}{dJ(t)^2}\Big|_{J=0} = \Xi(t) + \mathbb{E}_{P_0}[x_0^2] f^2(t,0)$$ ✓
- higher-order cumulants $$\kappa_{n>2}(t) = \frac{d^n W[J]}{dJ(t)^n}\Big|_{J=0} = \frac{d^n}{dJ(t)^n} \log \mathbb{E}_{P_0}[e^{J(t) x_0 f(t,0)}]\Big|_{J=0} = \kappa_n(0) f^n(t,0)$$ ✓

Two-dimensional scalar fields

extension to scalar fields trivial: each lattice point is treated separately

- forward $\partial_t \phi(x,t) = K[\phi(x,t),t] + g(t)\eta(x,t)$
- backward $\partial_\tau \phi(x,\tau) = -K[\phi(x,\tau),T-\tau] + g^2(T-\tau)\nabla_\phi \log P(\phi,T-\tau) + g(T-\tau)\eta(x,\tau)$
- two-point function $G(x,y;t) \equiv \mathbb{E}[\phi(x,t)\phi(y,t)] = \mathbb{E}_{P_0}[\phi_0(x)\phi_0(y)]f^2(t,0) + \Xi(t)\delta(x-y)$
- moments $\mu_n(x,t) = \mathbb{E}[\phi^n(x,t)]$ independent of x

$$\Xi(t) = \int_0^t ds\, f^2(t,s) g^2(s)$$

Generating functionals

full path integral
with sources

- moment generating

$$Z[J] = \mathbb{E}[e^{J(x,t)\phi(x,t)}] = e^{\frac{1}{2}J^2(x,t)\Xi(t)} \int D\phi_0\, P_0[\phi_0] e^{J(x,t)\phi_0(x) f(t,0)}$$

- cumulant generating

$$W[J] = \log Z[J] = \frac{1}{2} J^2(x,t)\Xi(t) + \log \int D\phi_0\, P_0[\phi_0] e^{J(x,t)\phi_0(x) f(t,0)}$$

- higher-order cumulants

$$\kappa_{n>2}(t) = \frac{\delta^n W[J]}{\delta J(x,t)^n}\Big|_{J=0} = \frac{\delta^n}{\delta J(x,t)^n} \log \mathbb{E}_{P_0}[e^{J(x,t)\phi_0(x) f(t,0)}]\Big|_{J=0}$$

variance preserving

$$f(t,0) \to 0$$

variance expanding

$$f(t,0)) = 1$$

$$\Xi(t) = \int_0^t ds\, f^2(t,s) g^2(s)$$

Generating functionals: summary

- euclidean path integral/target distribution is always there in the background

$$W[J] = \log Z[J] = \frac{1}{2} J^2(x,t)\Xi(t) + \log \int D\phi_0\, P_0[\phi_0] e^{J(x,t)\phi_0(x) f(t,0)}$$

- correlations are being destroyed/overwhelmed and retrieved

- if score is determined exactly, full theoretical control

GA, D Habibi, L Wang, K Zhou, Mach.Learn.Sci.Tech. **6** (2025) 2, 025004 [arXiv:2410.21212 [hep-lat]]

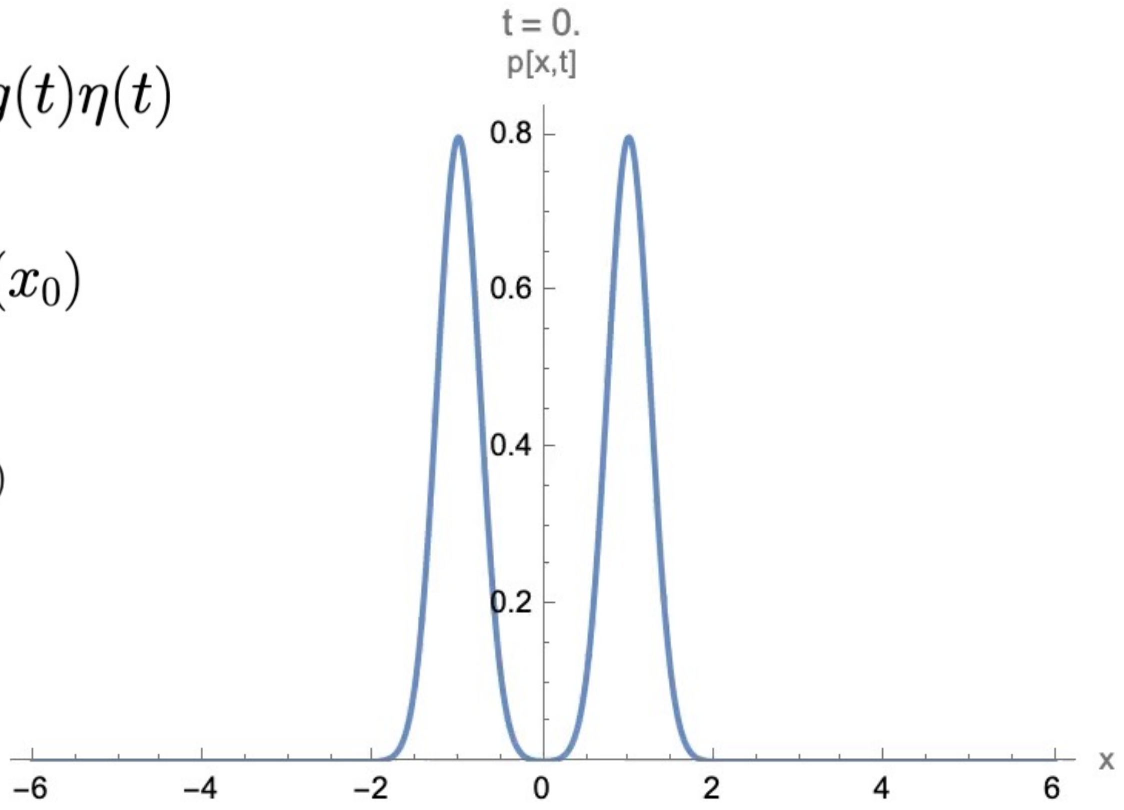
FPE: $\partial_t P_t(x) = \frac{1}{2} g^2(t) \partial_x^2 P_t(x)$

Example: forward evolution

- initial distribution $P_0(x_0)$: Gaussian mixture (two Gaussian peaks)
- add noise in variance-expanding scheme $\dot{x}(t) = g(t)\eta(t)$
- analytical description $P_t(x) = \int dx_0 \, P_t(x|x_0) P_0(x_0)$

$$P_t(x|x_0) = \mathcal{N}(x; x_0, \sigma^2(t)) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} e^{-(x-x_0)^2/(2\sigma^2(t))}$$

- peak structure erased



$0 \leq t \leq T.$

noise profile $g(t) = \sigma^{t/T}$

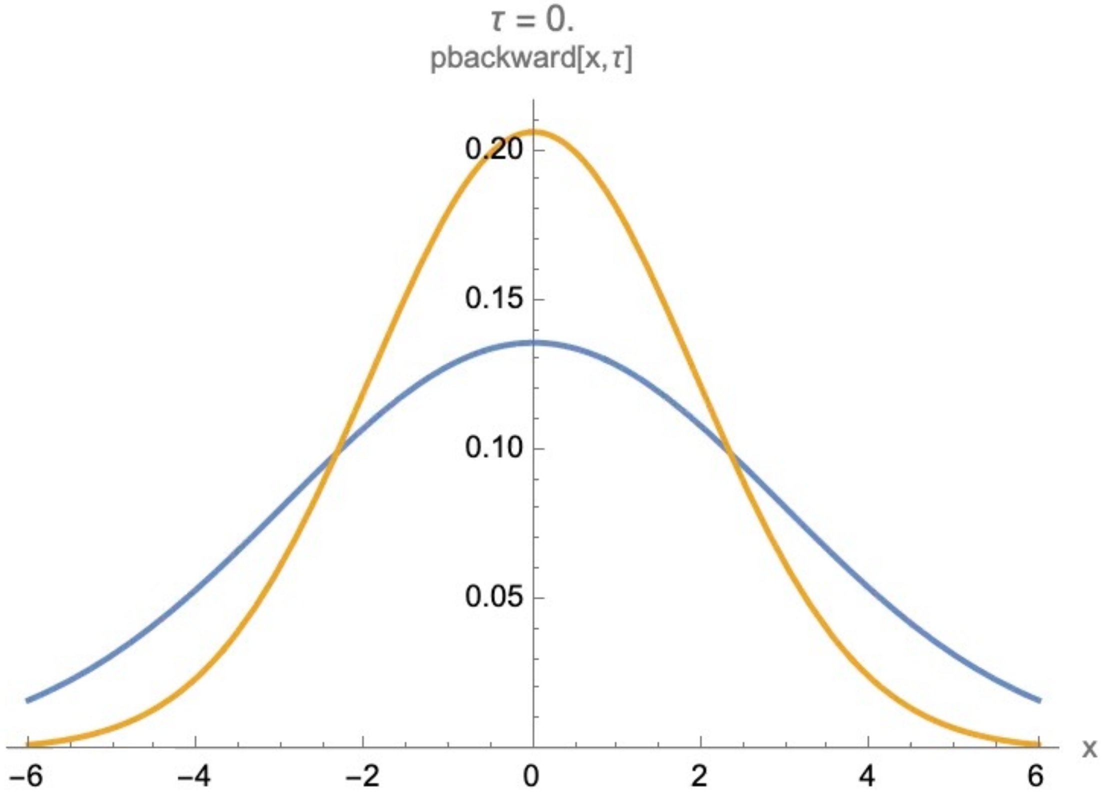
Example: backward evolution

- target distribution: two Gaussian peaks
- forward process $\dot{x}(t) = g(t)\eta(t)$
- corresponding backward process

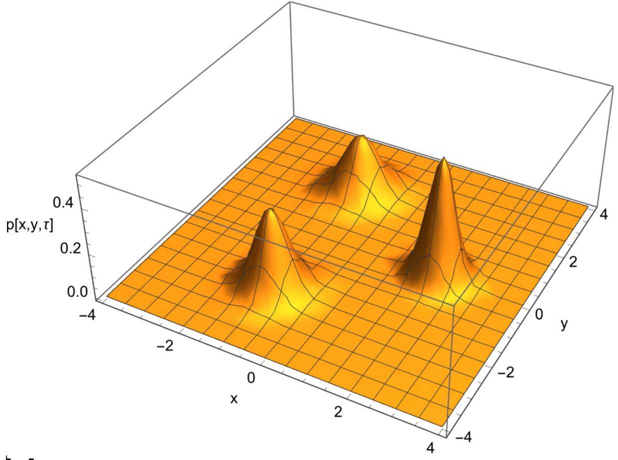
$$x'(\tau) = g^2(T-\tau)\partial_x \log(P(x, T-\tau) + g(T-\tau)\eta(\tau)$$

with $\tau = T - t$

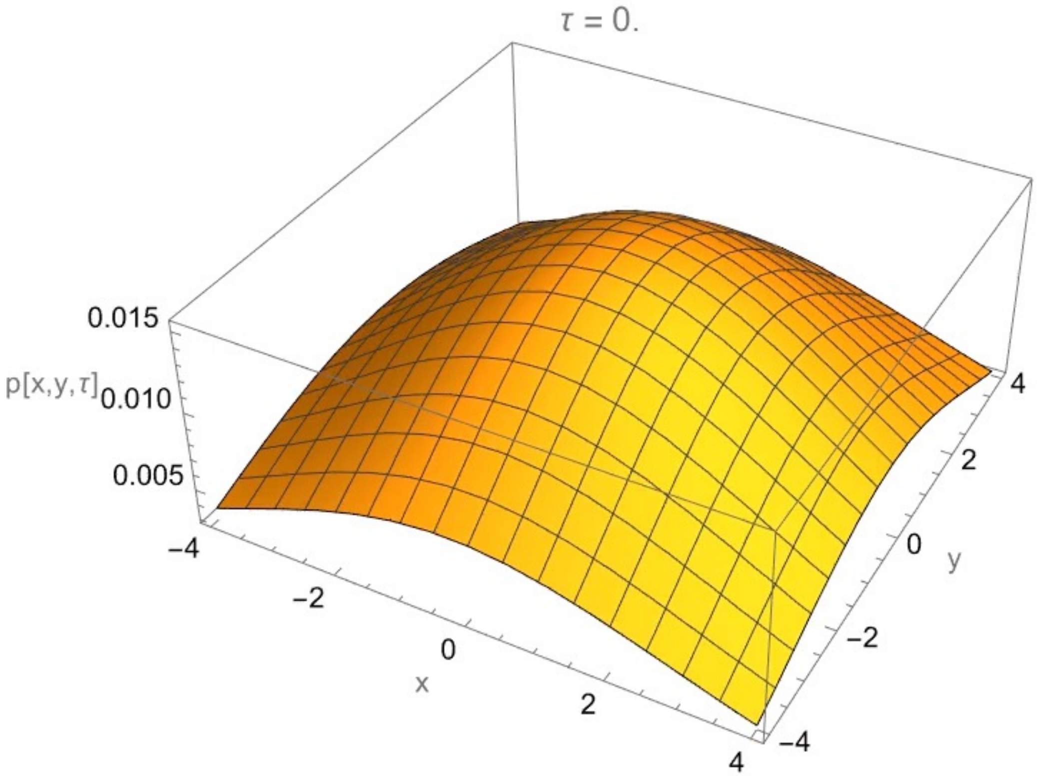
solve FPE for backward process using two initial distributions



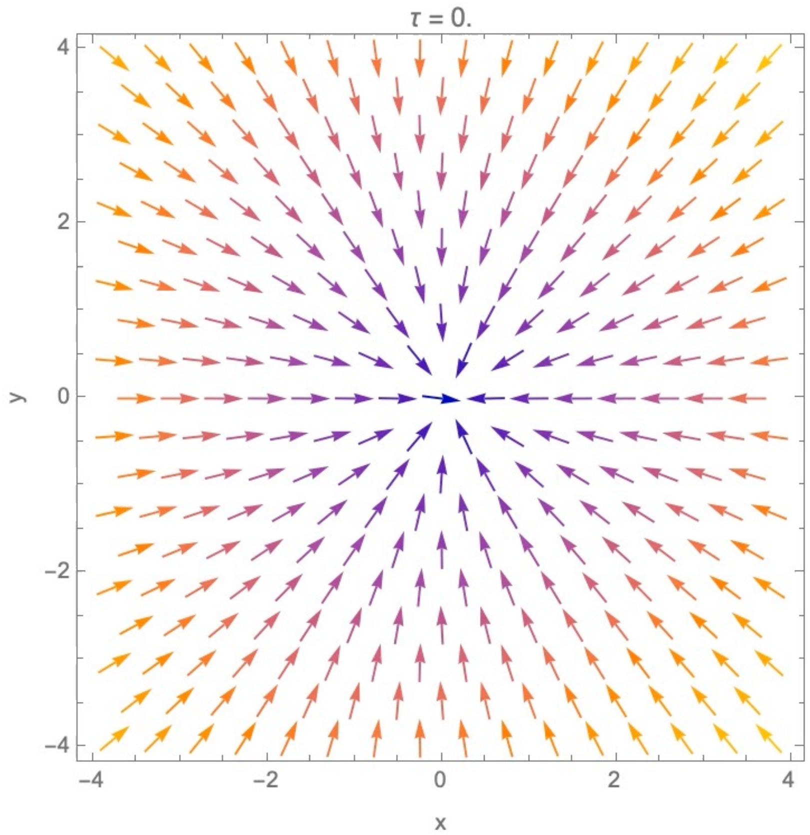
2D example: three Gaussian peaks



backward process, starting
from wide normal distribution



score $\nabla \log P_t(x, y)$
during backward process



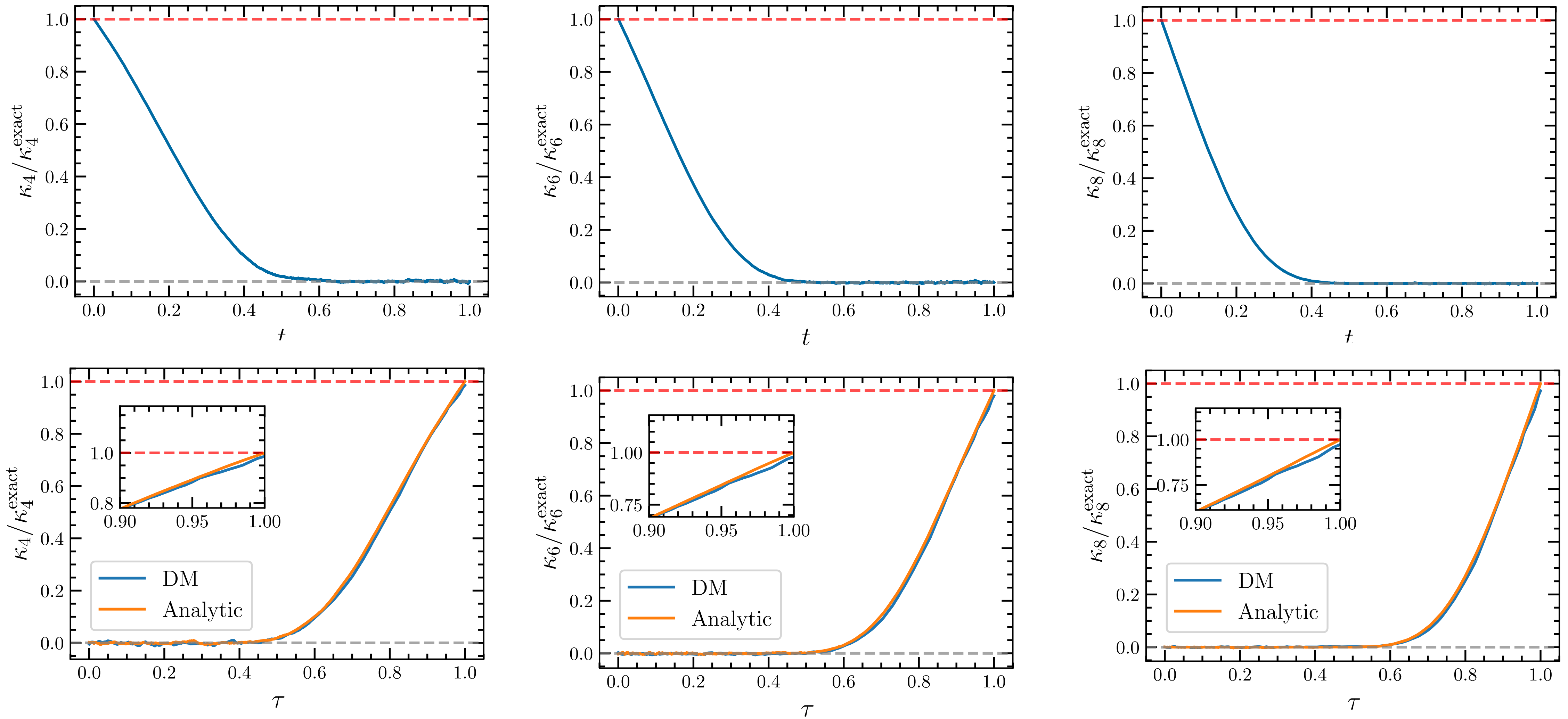
$$\kappa_{n>2}(t) = \kappa_n(0) f^n(t,0)$$

$$f(t,0) \to 0$$

4th, 6th, 8th cumulant with drift (DDPM)

forward

backward



analytic = analytic score

Incorporate (new/old) ideas in diffusion models

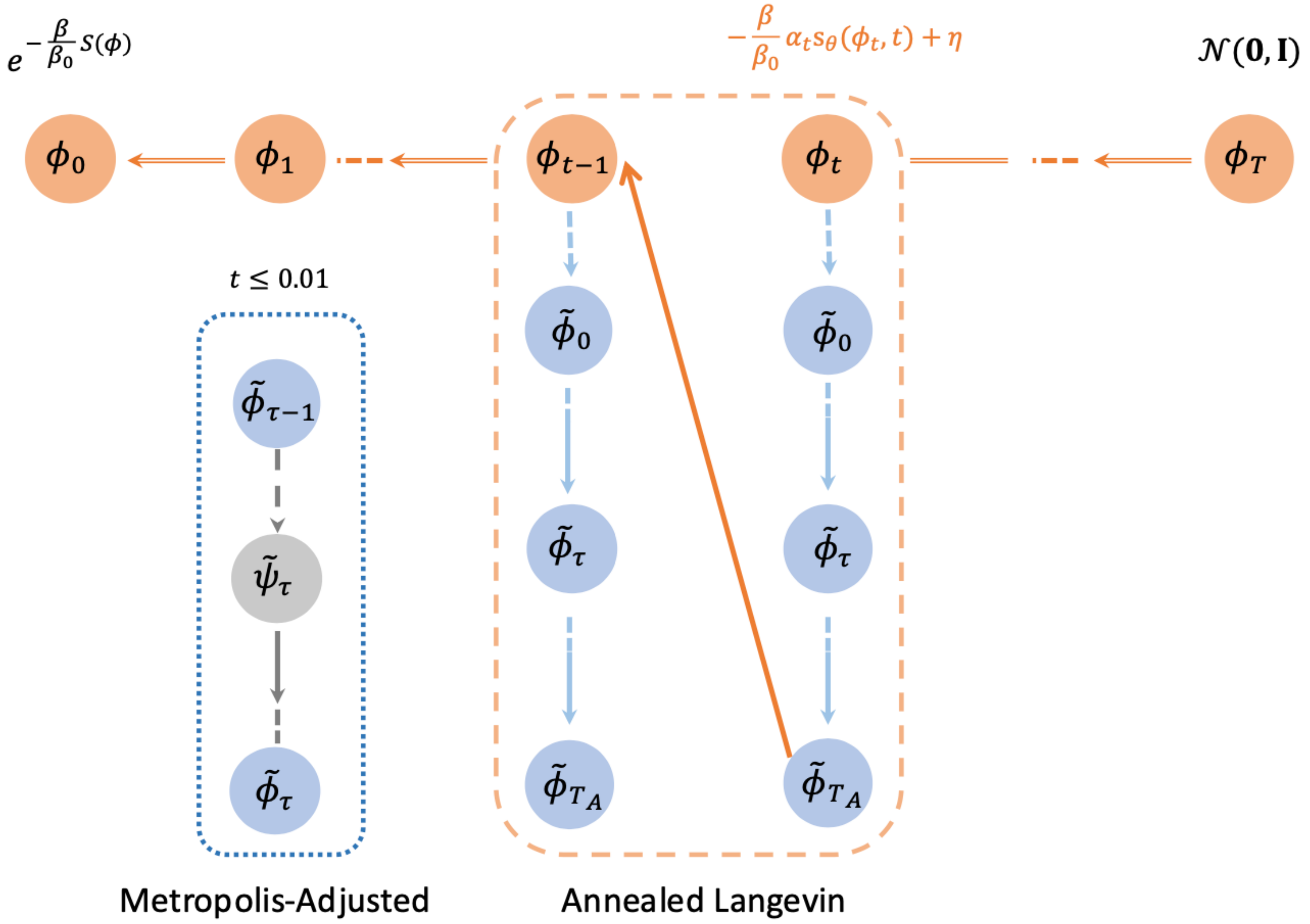
- exactness → include an accept/reject step
- thermalisation: score is time dependent, system never thermalises → annealing
- train at one set of parameters, apply trained model at different set → conditioning
- apply to 2D U(1) gauge theory

Q Zhu, W Wang, GA, K Zhou, L Wang, 2502.05504 [hep-lat]

Incorporate (new/old) ideas in DM dynamics

backward process
(after model has been trained)

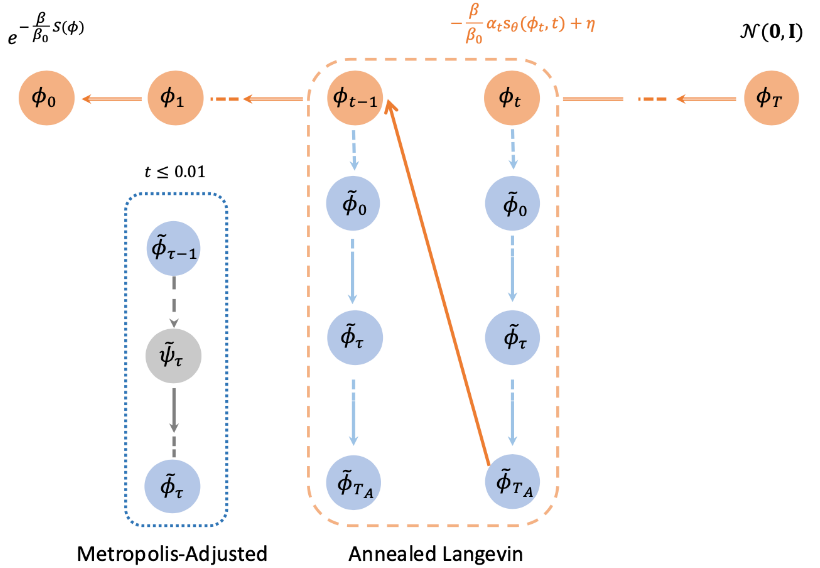
- Metropolis-adjusted Langevin algorithm (MALA)
- annealing stage: thermalisation
- reweighting from β_0 to β



Metropolis-adjusted Langevin algorithm (MALA)

- include an accept/reject step: well-known in Langevin dynamics *

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



- include ratio of target distributions $p(\phi) \sim e^{-S(\phi)}$
- and ratios of transition amplitudes

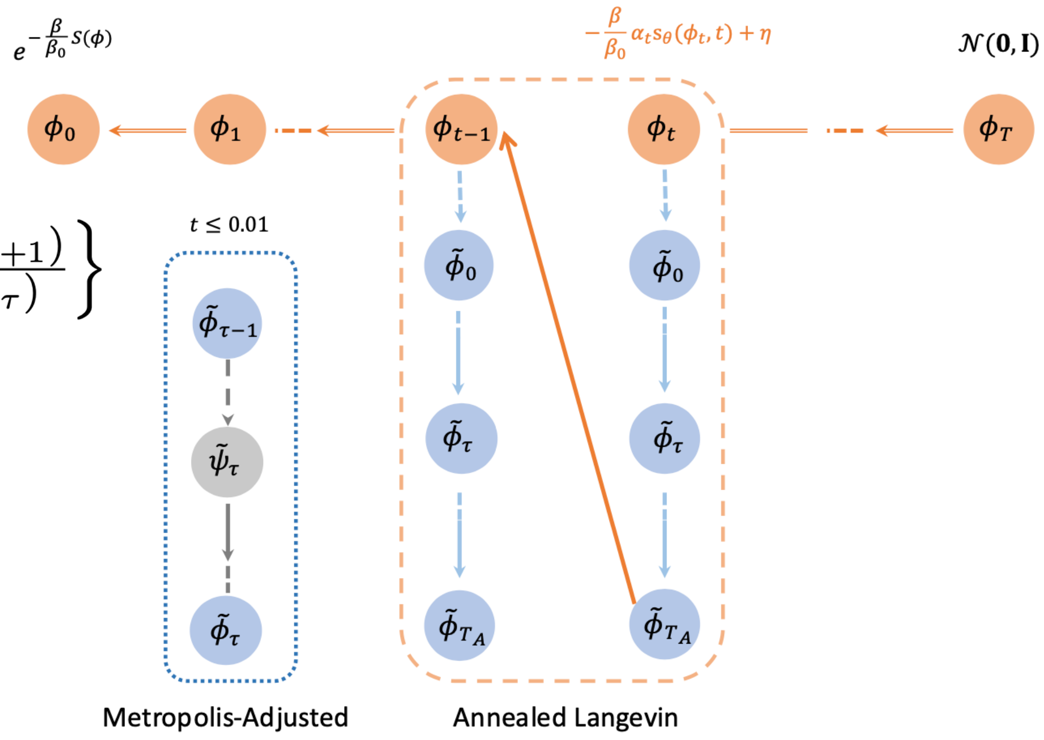
$$q(\phi_\tau|\psi_{\tau+1}) = \frac{1}{(4\pi\alpha_i)^{n/2}} \exp\left(-\frac{1}{4\alpha_i}\left\|\phi_\tau - (\psi_{\tau+1} + \alpha_i f(\psi_{\tau+1}, \tau+1))\right\|_2^2\right)$$

* G.O. Roberts and J.S. Rosenthal, Optimal scaling of discrete approximations to Langevin diffusions, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60 (1998) 255

Metropolis-adjusted Langevin algorithm (MALA)

- include an accept/reject step

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



- only done towards end of backward process

- learned score should be fairly close to “exact” score

$$\nabla \log p(\phi) \qquad\qquad p(\phi) \sim e^{-S(\phi)}$$

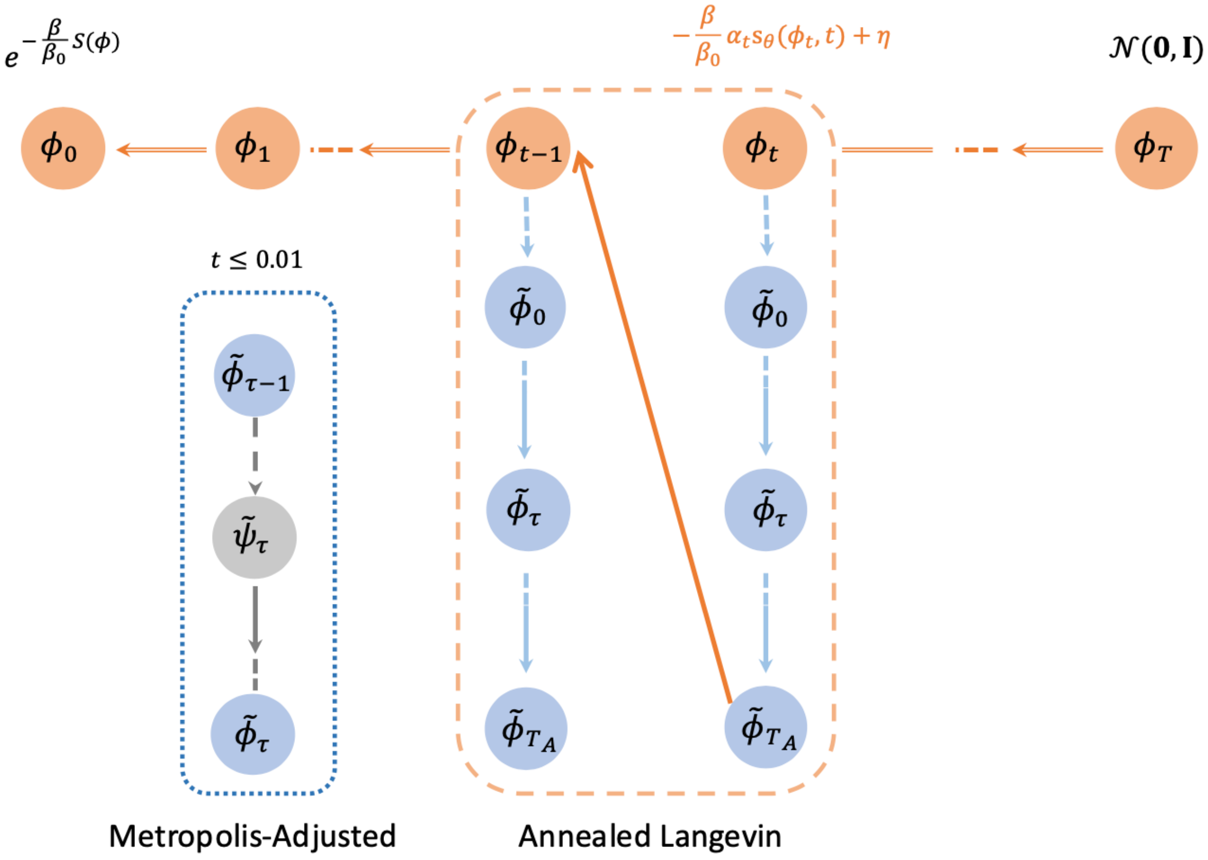
- Markov chain starting from each configuration towards end of backward process

Annealing

- score (drift or force in Langevin equation) is time dependent
- system never thermalises
- allow for additional steps at fixed score

→ annealing

- strictly speaking not needed, but seems useful

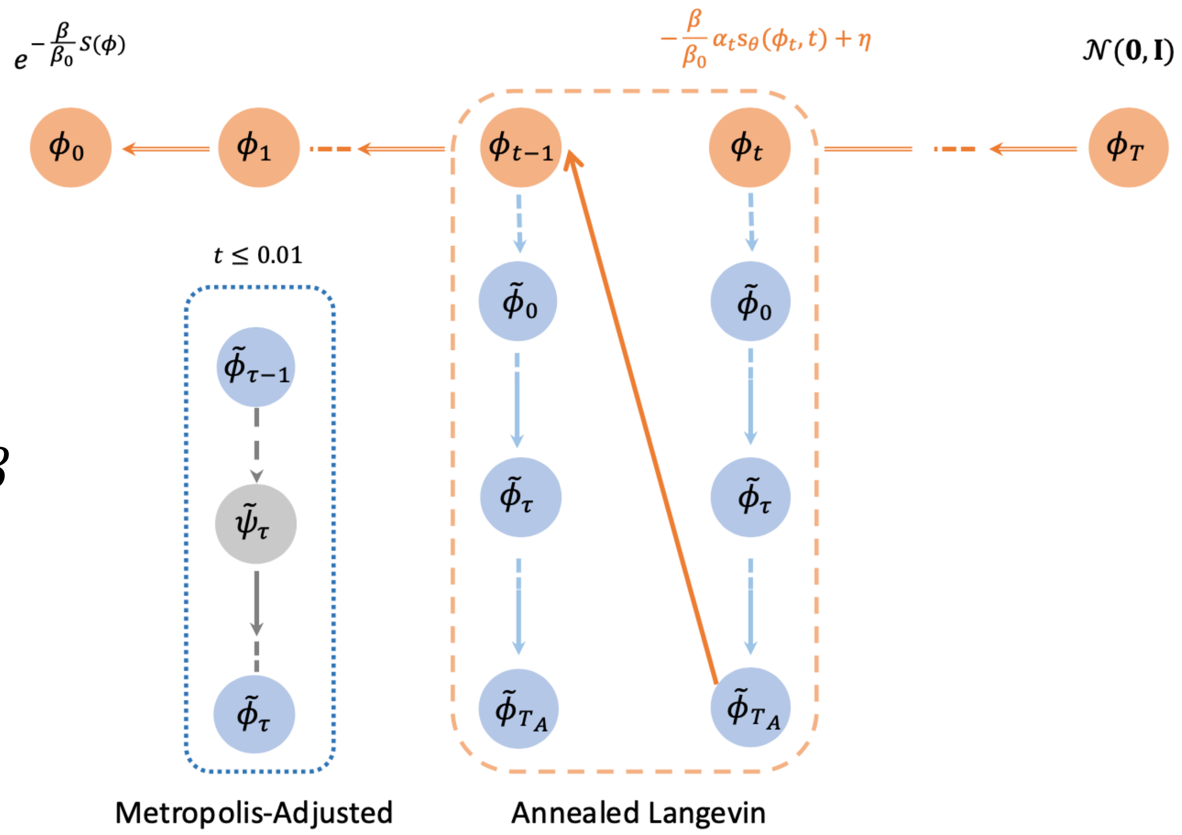


Physics conditioning (gauge theory)

- train using data generated at β_0
- employ at different β values
- applied to U(1) gauge theory: action scales with β

motivated by stochastic quantisation:

- drift is proportional to β

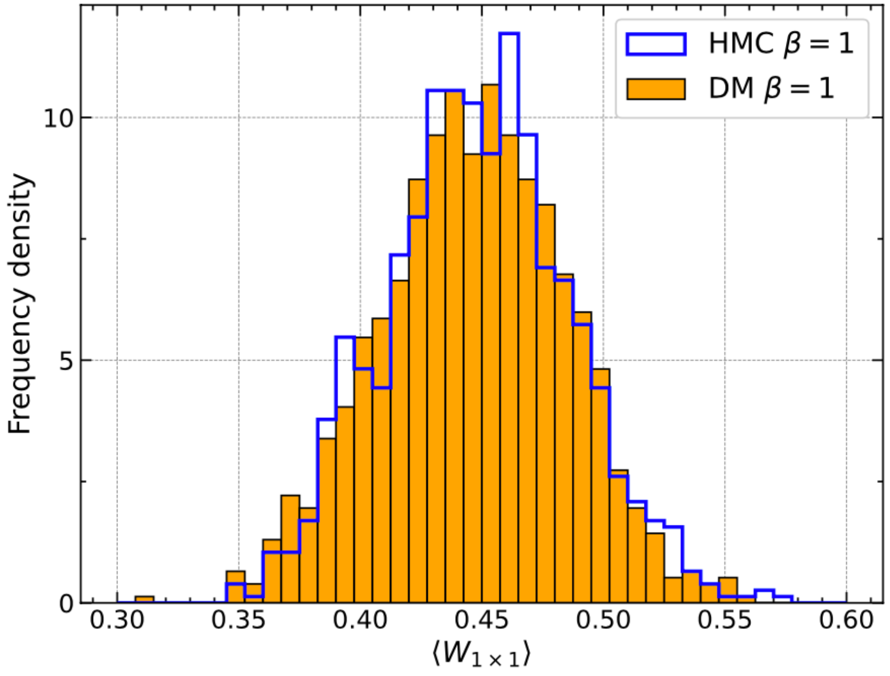


Two-dimensional U(1) gauge theory

- training: 30k configurations at $\beta = 1$ on 16^2 obtained using HMC
- generating: 1024 configs at $\beta = 1, 3, 5, 7, 9, 11$ on $8^2, 16^2, 32^2$

2D U(1) gauge theory: vary the volume

- training: 30k configurations at $\beta = 1$ on 16^2 obtained using HMC
- generating: 1024 configs at $\beta = 1, 3, 5, 7, 9, 11$ on $8^2, 16^2, 32^2$



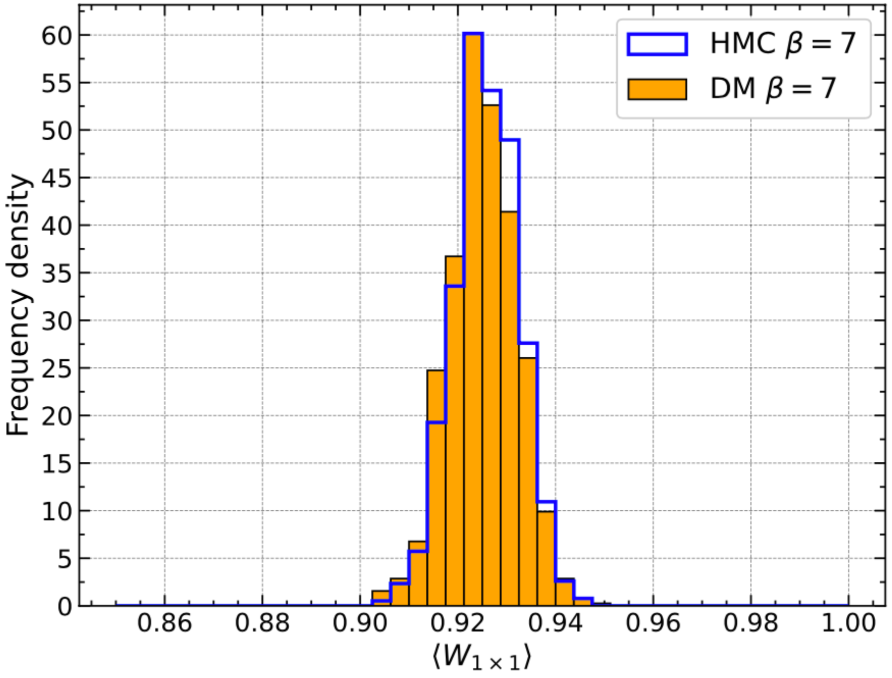
$\beta = 1, L = 16$, HMC vs DM

Lattice Size (L)	1×1 Wilson Loop			
	HMC	DM	Langevin	Exact
8	0.447(72)	0.445(74)	0.443(80)	0.446
16	0.447(37)	0.446(37)	0.444(36)	0.446
32	0.446(18)	0.445(19)	0.445(18)	0.446
64	0.446(9)	0.446(11)	0.445(9)	0.446

increase the volume, after training on $L = 16$

2D U(1) gauge theory: vary the coupling

- training: 30k configurations at $\beta = 1$ on 16^2 obtained using HMC
- generating: 1024 configs at $\beta = 1, 3, 5, 7, 9, 11$ on $8^2, 16^2, 32^2$



$\beta = 7, L = 16$, HMC vs DM

coupling (β)	1 × 1 Wilson Loop			
	HMC	DM	Langevin	Exact
3	0.811(17)	0.811(17)	0.809(17)	0.810
5	0.894(9)	0.894(9)	0.891(10)	0.894
7	0.926(7)	0.926(7)	0.924(6)	0.926
9	0.944(3)	0.942(4)	0.940(6)	0.942
11	0.954(3)	0.953(4)	0.950(5)	0.953

increase the coupling, after training at $\beta = 1$

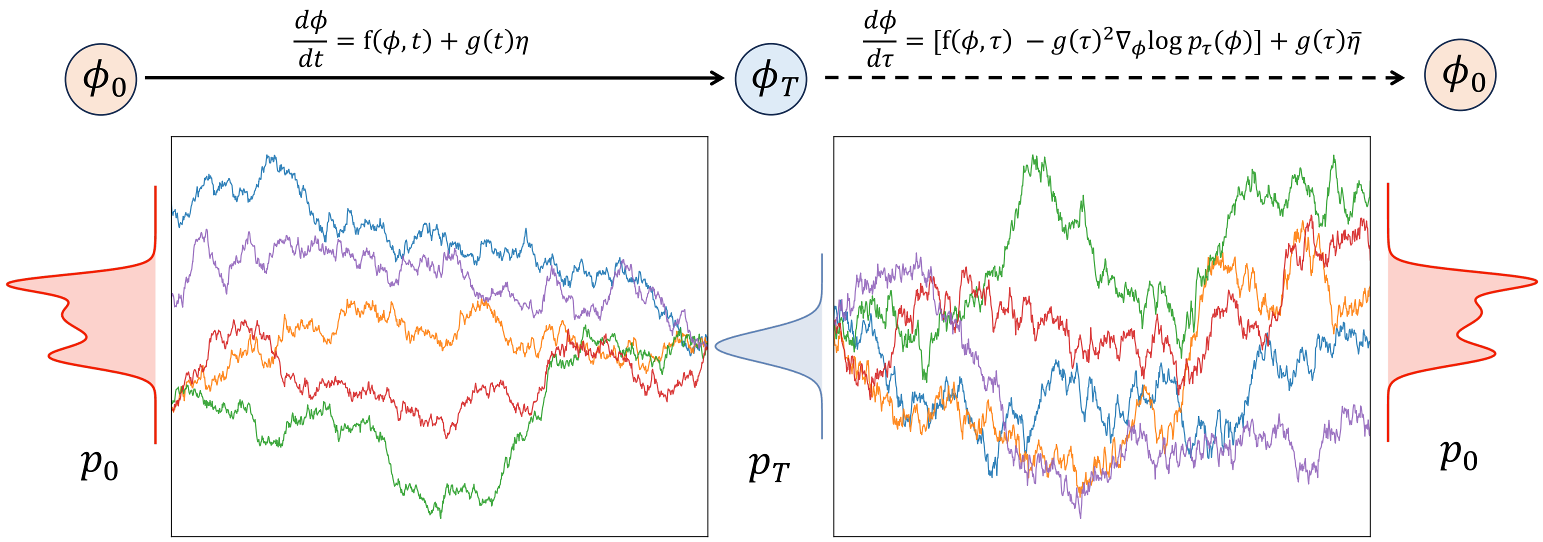
Summary: diffusion models

- offer a new approach for ensemble generation to explore in LFT
- learn from data: requires high-quality ensembles
- closely related to stochastic quantisation
- need better understanding of precision and exactness
- indicated three promising directions to be explored further

BACKUP SLIDES

Score matching: learn the drift for backward process

- one degree of freedom, variance-expanding scheme: $\dot{x}(t) = g(t)\eta(t) \qquad \eta \sim \mathcal{N}(0,1)$
- time-dependent distribution $P(x,t) = P_t(x)$ describes forward and backward process
- so-called **score** $\nabla \log P_t(x)$ is not known, needs to be "learnt" during forward process



Score matching: learn the drift for backward process

- one degree of freedom, variance-expanding scheme: $\dot{x}(t) = g(t)\eta(t) \qquad \eta \sim \mathcal{N}(0,1)$
- time-dependent distribution $P(x,t) = P_t(x)$ describes forward and backward process
- so-called **score** $\nabla \log P_t(x)$ is not known, needs to be "learnt"
- loss function $\mathcal{L}(\theta) = \frac{1}{2}\int_0^T dt\, \mathbb{E}_{P_t(x)}\left[\sigma^2(t)\, \|s_\theta(x,t) - \nabla \log P_t(x)\|^2\right] \qquad \sigma^2(t) = \int_0^t ds\, g^2(s)$
- $s_\theta(x,t)$ approximates score, vector field learnt by some neural network
- introduce conditional distribution $P_t(x) = \int dx_0\, P_t(x|x_0) P_0(x_0)$ initial data $P_0(x_0)$

$$P_t(x) = \int dx_0\, P_t(x|x_0) P_0(x_0)$$

Score matching: learn the drift

- loss function $\mathcal{L}(\theta) = \frac{1}{2}\int_0^T dt\, \mathbb{E}_{P_t(x)}\left[\sigma^2(t)\, \| s_\theta(x,t) - \nabla \log P_t(x) \|^2\right]$
- diffusion process $\dot{x}(t) = g(t)\eta(t)$ easily solved $x(t) = x_0 + \sigma(t)\eta(t)$ $\sigma^2(t) = \int_0^t ds\, g^2(s)$
- conditional distribution $P_t(x|x_0) = \mathcal{N}(x; x_0, \sigma^2(t)) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} e^{-(x-x_0)^2/(2\sigma^2(t))}$
- and hence $\nabla \log P_t(x_t|x_0) = -(x_t - x_0)/\sigma^2(t)$
- loss function

$$\mathcal{L}(\theta) = \frac{1}{2}\int_0^T dt\, \mathbb{E}_{P_t(x_t)}\left[\left\| \sigma(t) s_\theta(x_t,t) + \frac{x_t - x_0}{\sigma(t)} \right\|^2\right]$$

$$= \frac{1}{2}\int_0^T dt\, \mathbb{E}_{P_t(x_t)}\left[\| \sigma(t) s_\theta(x_t,t) + \eta(t) \|^2\right]$$

tractable, computable