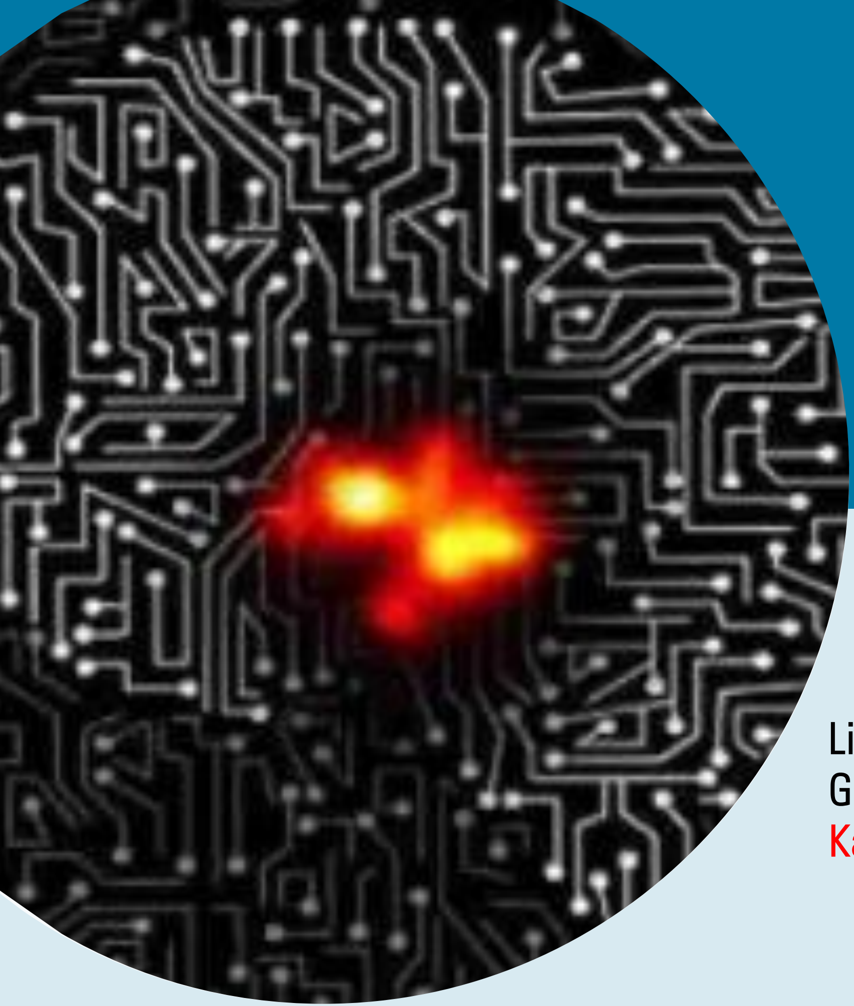


香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

Diffusion Model as Stochastic Quantization in lattice QFT

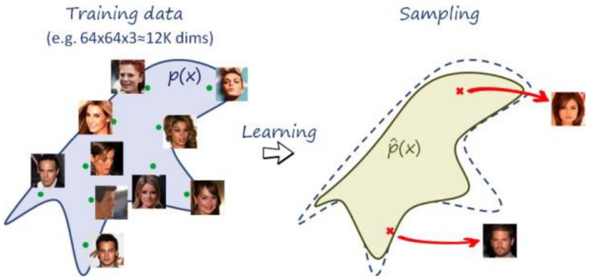
Lingxiao Wang (RIKEN-iTHEMS)
Gert Aarts (Swansea Uni.)
Kai Zhou (CUHK-Shenzhen) ← we are hiring

ML & RG (ECT*, Trento)

Generative Models Problem Set

Want to **model** the observed data's underlying but unknown **distribution**, to further :

- Understand/Inference the data (inherent structure, properties, features...)
- Sample according to the distribution



Suppose observation dataset :

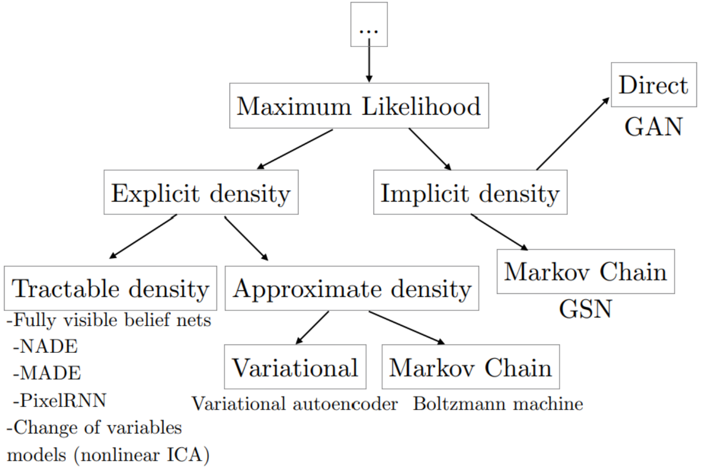
$$\mathbf{X} = \{x^{(1)}, x^{(2)}, ..., x^{(N)}\} \overset{i.i.d}{\sim} p_{data}(x)$$

We use parametric model to approach the data distribution :

$$p_\theta(x) \rightarrow p_{data}(x)$$

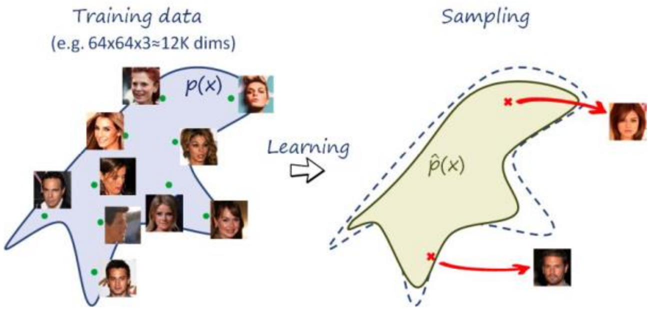
- Maximize Likelihood Estimation :

$$\theta^* = \arg\max_\theta \log p_\theta(\mathbf{X}) = \arg\max_\theta \frac{1}{N} \sum_{i=1}^{N} \log p_\theta(x^{(i)})$$

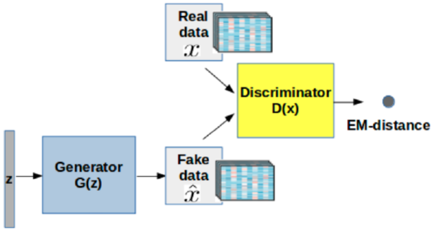


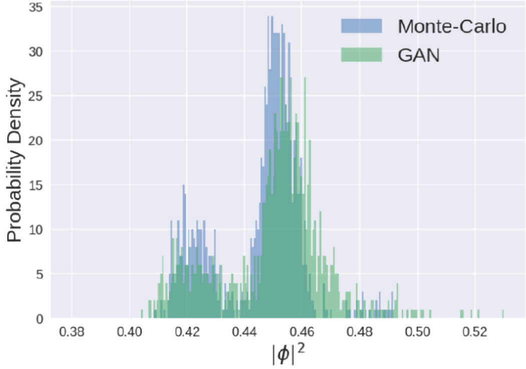
I. Goodfellow, arXiv:1701.00160 (2017)

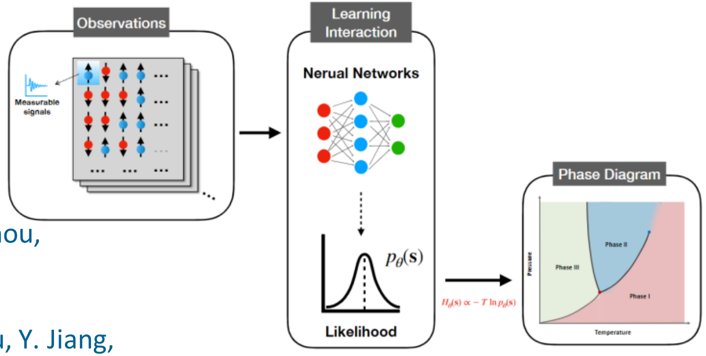
Given an ensemble of data from the target distribution



K. Zhou, G. Endrődi, L.-G. Pang, and H. Stöcker, **PRD 100, 011501 (2019)**

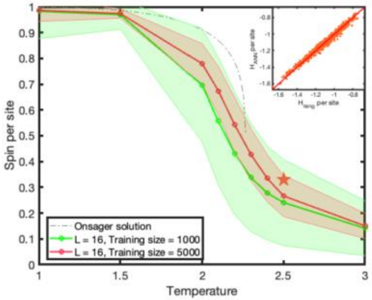


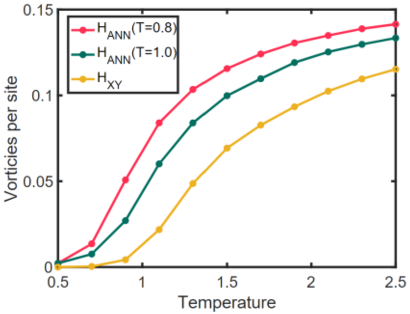




L. Wang, L. He, Y. Jiang, K. Zhou,
arXiv:2007.01037

T. Xu, L. Wang, L. He, K. Zhou, Y. Jiang,
:2405.10493





See **Lingxiao**'s talk on Friday !

$$p(\mathrm{s}) = \frac{e^{-\beta E(\mathrm{s})}}{Z},$$

- Reverse KL divergence

$$D_{\mathrm{KL}}(q_\theta \| p) = \sum_{\mathrm{s}} q_\theta(\mathrm{s}) \ln\left(\frac{q_\theta(\mathrm{s})}{p(\mathrm{s})}\right) = \beta(F_q - F).$$

$$F_q = \frac{1}{\beta}\sum_{\mathrm{s}} q_\theta(\mathrm{s})\left[\beta E(\mathrm{s}) + \ln q_\theta(\mathrm{s})\right]$$

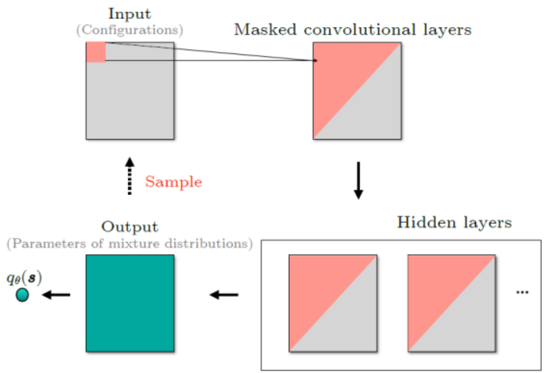
- Autoregressive

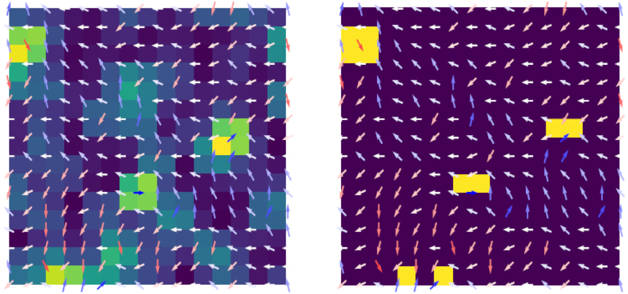
$$q_\theta(\mathrm{s}) = \prod_{i=1}^{N} q_\theta(s_i \mid s_1, \ldots, s_{i-1})$$

D. Wu, Lei Wang and P. Zhang, **PRL122,080602(2019)**

- Continuous Autoregressive for XY model

L. Wang, Y. Jiang, L. He, K. Zhou, **CPL 39, 120502 (2022)**





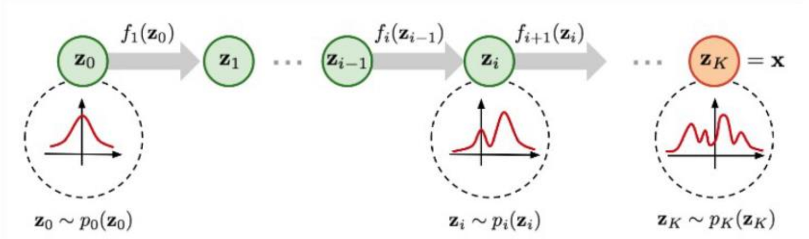
Probability distributions from CANs

Vortices

Flow based generative model given unnormalized distribution

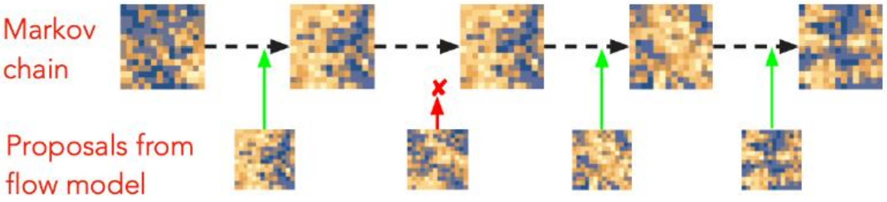
A series **(Flow)** of invertible/bijective transformations for p(z)

<u>compose several invertible transformations to form the flow :</u>



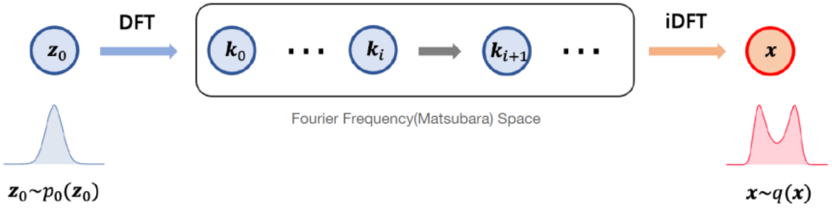
$$p_i(\mathbf{z}_i) = p_{i-1}(f_i^{-1}(\mathbf{z}_i))|\det J_{f_i^{-1}}| = p_{i-1}(\mathbf{z}_{i-1})|\det J_{f_i}|^{-1}$$

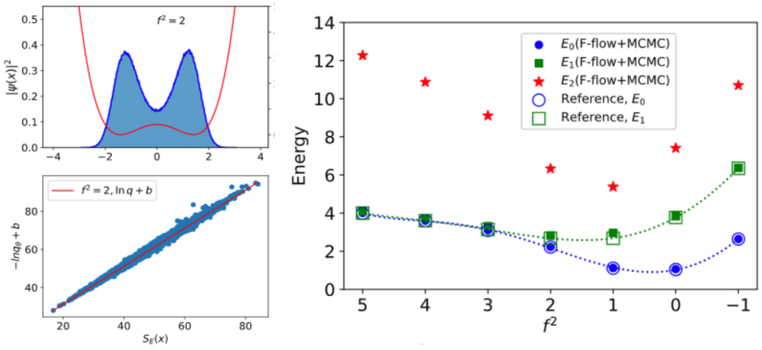
$$\to \log p(\mathbf{x}) = \log p_0(f^{-1}(\mathbf{x})) + \sum_{i=1}^{K} \log|\det J_{f_i^{-1}}| = \log p_0(\mathbf{z}_0) - \sum_{i=1}^{K} \log|\det J_{f_i}|$$



Fourier Flow Model

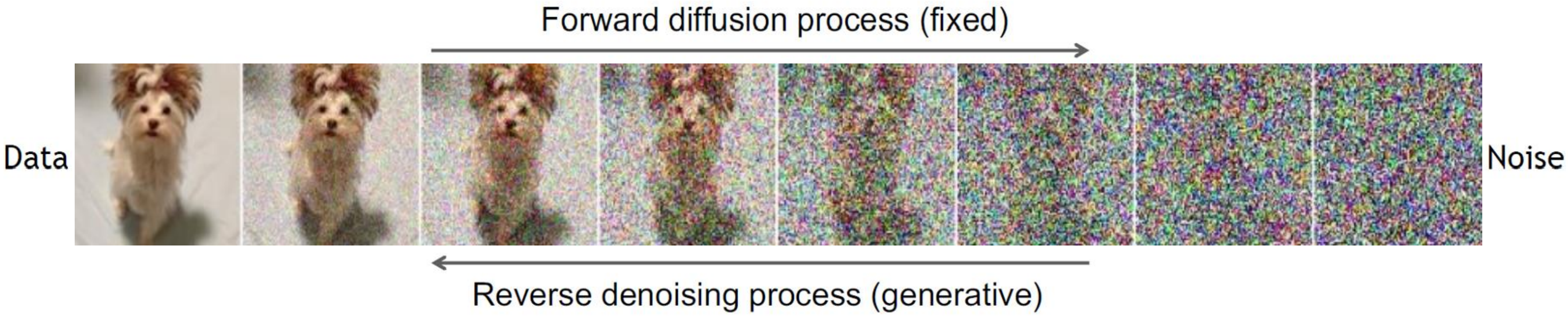
S.Chen, O. Savchuk, S. Zheng, B. Chen, H. Stoecker, L. Wang, K. Zhou, **PRD107, 056001(2023)**



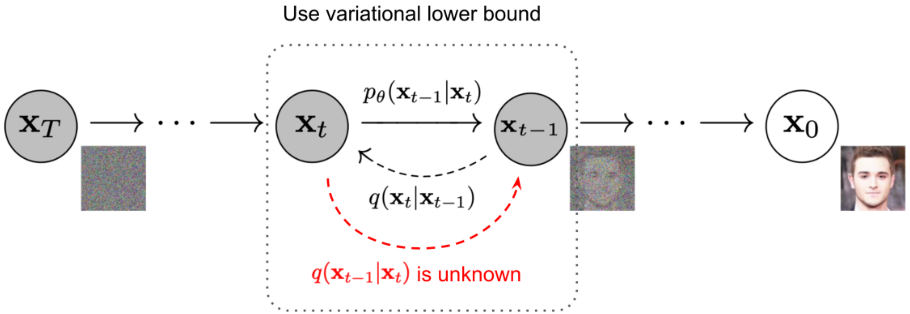


"A heavy quark move inside quark-gluon plasma"

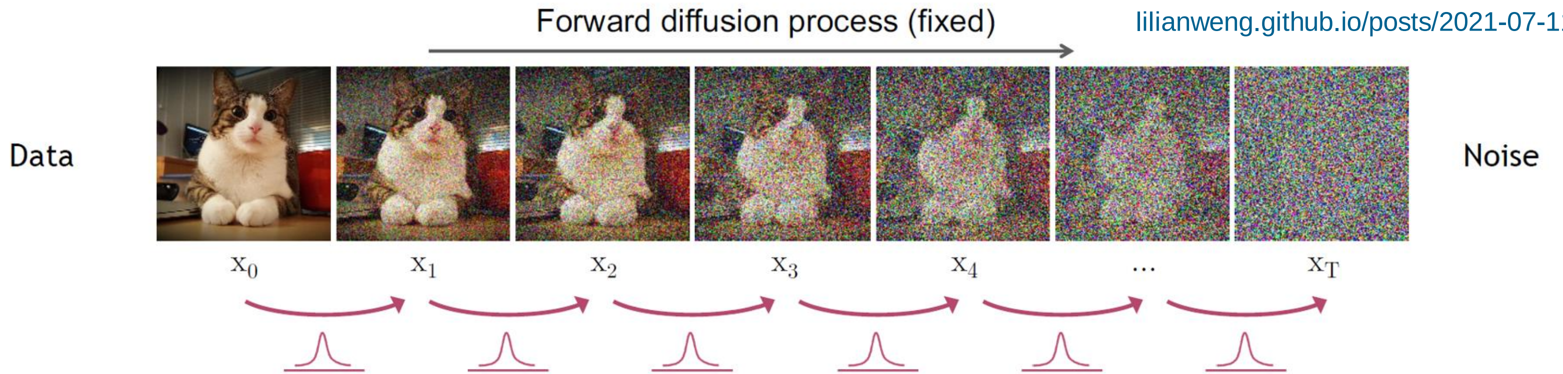
- **Forward diffusion process (<u>fixed</u>)**: gradually introduce noise into data



- **Reverse diffusion process (<u>learned</u>)**: gradually denoise to generate data



lilianweng.github.io/posts/2021-07-11-diffusion-models/



$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \quad \Rightarrow \quad q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})$$ (joint)

Let $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$ $\Rightarrow$ $q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$

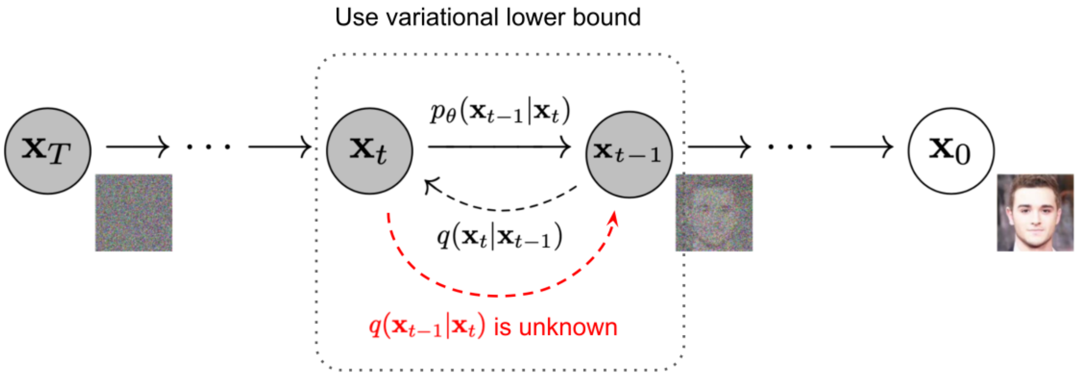
$\Updownarrow$

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{(1-\bar{\alpha}_t)}\,\epsilon$$ where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

$\beta_1 < \beta_2 < \cdots < \beta_T$ therefore $\bar{\alpha}_1 > \cdots > \bar{\alpha}_T$ $\quad$ $\bar{\alpha}_T \to 0$ and $q(\mathbf{x}_T|\mathbf{x}_0) \approx \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I}))$

Reverse Diffusion process

lilianweng.github.io/posts/2021-07-11-diffusion-models/



$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \quad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

$$\mathbb{E}_{q(\mathbf{x}_0)}\left[-\log p_\theta(\mathbf{x}_0)\right] \leq \mathbb{E}_{q(\mathbf{x}_0)q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\left[-\log \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\right] = L = \mathbb{E}_q\Bigg[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)||p(\mathbf{x}_T))}_{L_T} + \sum_{t>1}\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{- \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}\Bigg]$$

$$\begin{aligned}
L_t &= \mathbb{E}_{\mathbf{x}_0,\epsilon}\Big[\frac{1}{2\|\boldsymbol{\Sigma}_\theta(\mathbf{x}_t,t)\|_2^2}\|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) - \boldsymbol{\mu}_\theta(\mathbf{x}_t,t)\|^2\Big] \\
&= \mathbb{E}_{\mathbf{x}_0,\epsilon}\Big[\frac{1}{2\|\boldsymbol{\Sigma}_\theta\|_2^2}\Big\|\frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\Big) - \frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\Big)\Big\|^2\Big] \\
&= \mathbb{E}_{\mathbf{x}_0,\epsilon}\Big[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\|^2\Big] \\
&= \mathbb{E}_{\mathbf{x}_0,\epsilon}\Big[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, t)\|^2\Big]
\end{aligned}$$

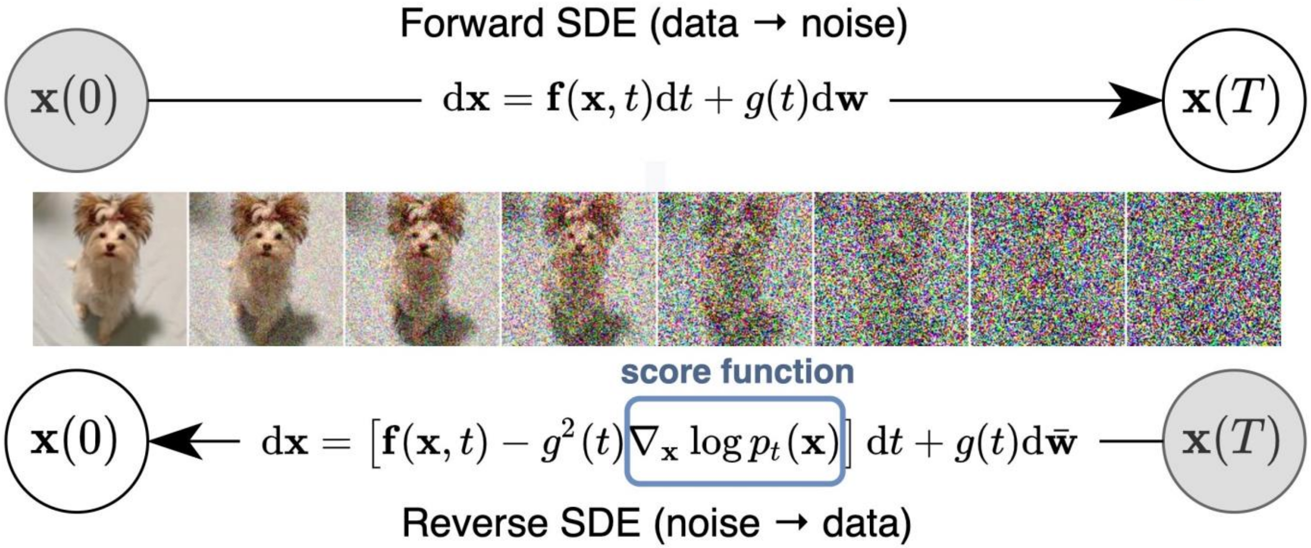
Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \mathrm{Uniform}(\{1,\ldots,T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$
5: Take gradient descent step on
$\nabla_\theta \left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\right\|^2$
6: **until** converged

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0},\mathbf{I})$
2: **for** $t = T,\ldots,1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\right) + \sigma_t\mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

$$d\mathbf{x}_t = -\frac{1}{2}\beta(t)\mathbf{x}_t\,dt + \sqrt{\beta(t)}\,d\boldsymbol{\omega}_t$$



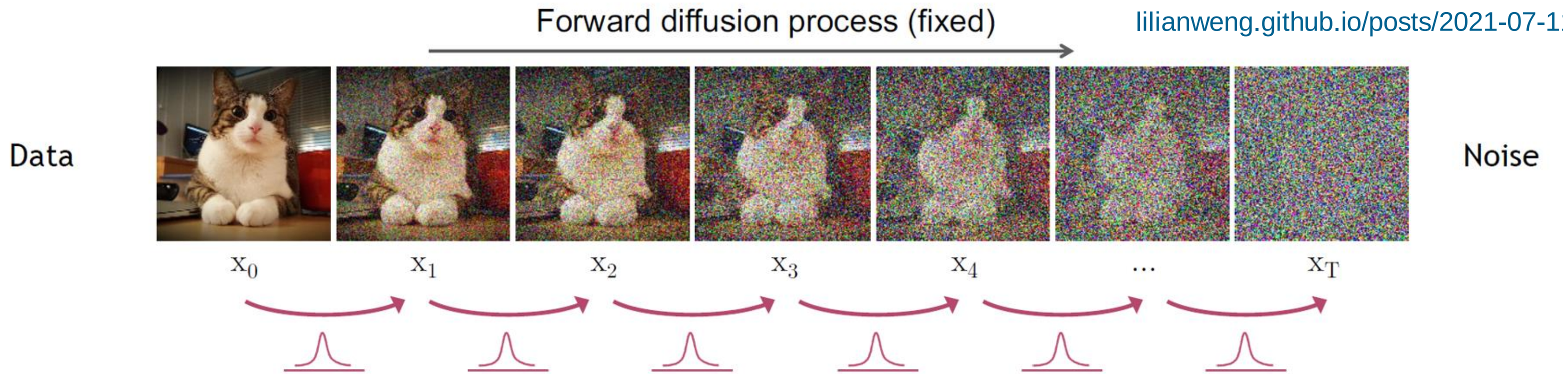
Song et al., Score-Based Generative Modeling through Stochastic Differential Equations, ICLR 2021

- Fisher Divergence

Via score matching

$$\mathbb{E}_{p(\mathbf{x})}[||\nabla_{\mathbf{x}} \log p(\mathbf{x}) - \mathbf{s}_\theta(\mathbf{x})||_2^2]$$

lilianweng.github.io/posts/2021-07-11-diffusion-models/



$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x_t}; \sqrt{1-\beta_t}\mathbf{x_{t-1}}, \beta_t\mathbf{I}) \Rightarrow q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}) \quad \text{(joint)}$$

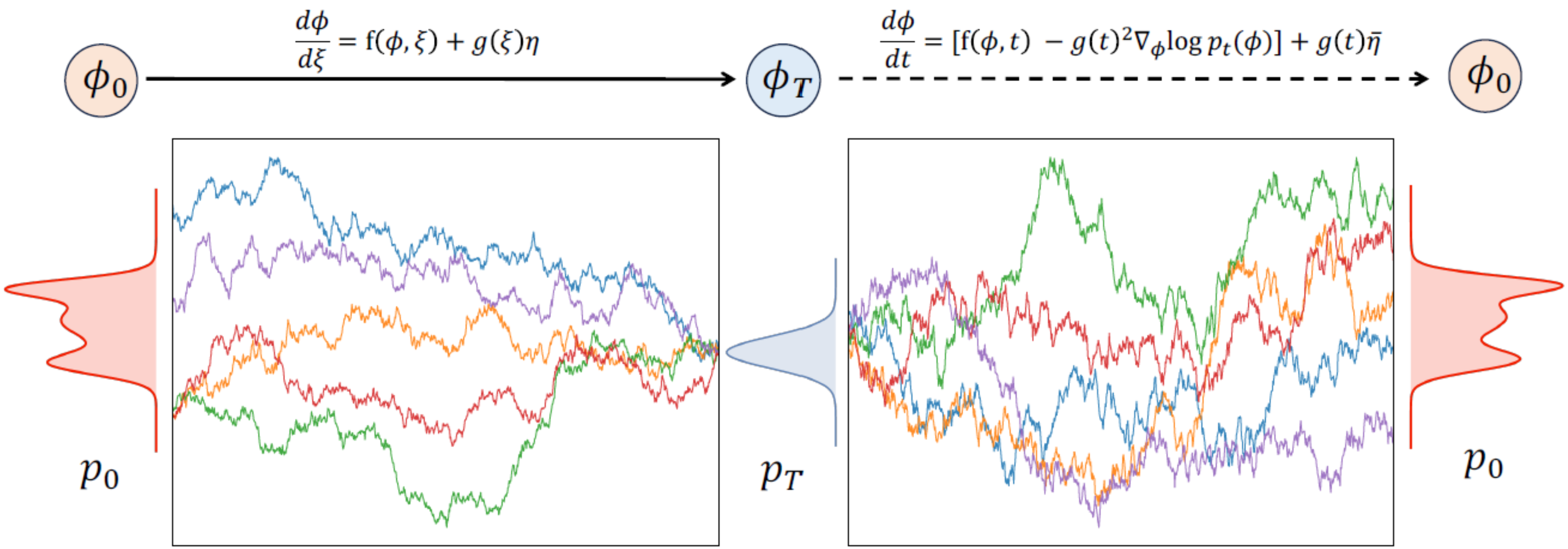
Let $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$ $\Rightarrow$ $q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$

$\Updownarrow$

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{(1-\bar{\alpha}_t)}\,\epsilon \quad \text{where } \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$\beta_1 < \beta_2 < \cdots < \beta_T$ therefore $\bar{\alpha}_1 > \cdots > \bar{\alpha}_T$ $\quad$ $\bar{\alpha}_T \to 0$ and $q(\mathbf{x}_T|\mathbf{x}_0) \approx \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I}))$

Apply DM on lattice QFT configurations



L. Wang, G. Arts, K. Zhou, JHEP 05 (2024) 060
L. Wang, G. Arts, K. Zhou, arXiv:2311.03578 (NeurIPS 2023 workshop “ML&Physical Sciences”)

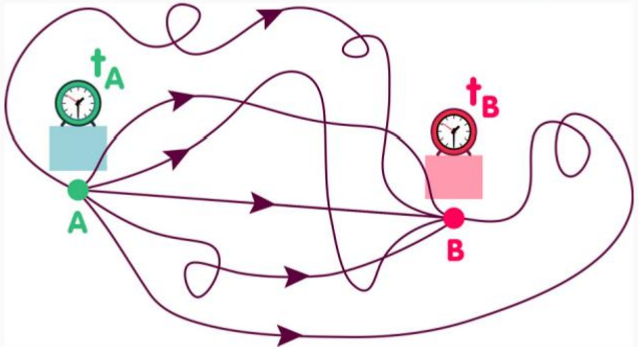
Stochastic Quantization

- <u>Stochastic vibration</u>

$$i\hbar\frac{\partial}{\partial t}\Psi = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2}$$

$$\frac{\partial\phi(x,t)}{\partial t} = D\frac{\partial^2\phi(x,t)}{\partial x^2}$$

In Feynman's formulation of quantum mechanics in Euclidean space:



$$\mathcal{Z} = \int \mathcal{D}x\, e^{-S_E[x]/\hbar}$$

$$\langle 0|\hat{x}^N|0\rangle = \frac{1}{\mathcal{Z}}\int \mathcal{D}x\, x^N e^{-S_E[x]/\hbar}$$

One can construct stochastic process to reproduce the quantum path Integral with its equilibrium:

$$e^{-S_E[x]/\hbar} \rightarrow \frac{\partial x}{\partial \tau} = -\frac{\delta S_E[x]}{\delta x} + \eta \begin{cases} \langle\, \eta(t,\tau)\, \rangle_\eta = 0 \\ \langle\, \eta(t,\tau)\eta(t',\tau')\, \rangle_\eta = 2\hbar\delta(t-t')\delta(\tau-\tau') \end{cases}$$

Parisi G. and Wu Y. S., **Sci. China, A 24**, ASITP-80-004 (1980).

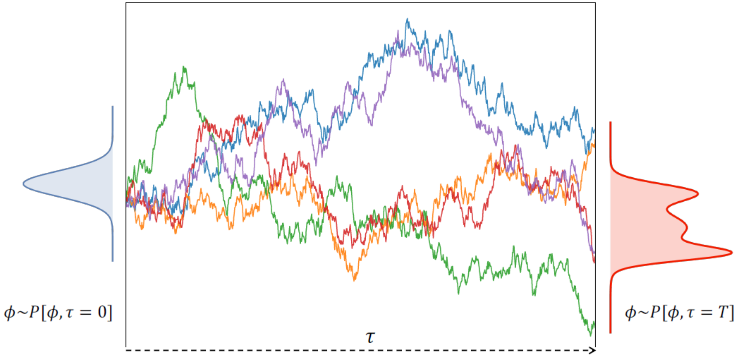
Stochastic Quantization

- **Stochastic quantization** $Z = \int D\phi\, e^{-S_E}$ $p(\phi) = \frac{e^{-S_E(\phi)}}{Z}$

$$\frac{\partial \phi(x,\tau)}{\partial \tau} = -\frac{\delta S_E[\phi]}{\delta \phi(x,\tau)} + \eta(x,\tau) \qquad \langle \eta(x,\tau) \rangle = 0, \qquad \langle \eta(x,\tau)\eta(x',\tau') \rangle = 2\alpha\delta(x-x')\delta(\tau-\tau')$$

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

long time equilibrium limit $P_{\mathrm{eq}}[\phi] \propto e^{-\frac{1}{\alpha} S_E[\phi]}$

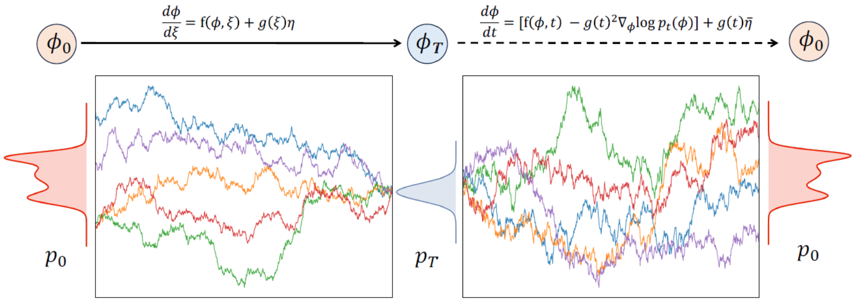


- **Observables** $\langle \mathcal{O}[\phi] \rangle_\tau = \int D\phi\, \mathcal{O}[\phi] P[\phi,\tau]$

$$\langle \mathcal{O}[\phi] \rangle_{\tau\to\infty} = \frac{\int D\phi\, \mathcal{O}(\phi) e^{-\frac{1}{\hbar}S_E(\phi)}}{\int D\phi\, e^{-\frac{1}{\hbar}S_E(\phi)}} = \langle \mathcal{O}[\phi] \rangle_{\mathrm{quantum}}$$

Parisi G. and Wu Y. S., **Sci. China, A 24**, ASITP-80-004 (1980).

Diffusion Model for field configurations



- Forward diffusion SDE

$$\frac{d\phi}{d\xi} = f(\phi,\xi) + g(\xi)\eta(\xi) \qquad \langle \eta(\xi)\eta(\xi')\rangle = 2\alpha\delta(\xi-\xi')$$

- Backward diffusion SDE

$$\frac{d\phi}{dt} = \left[f(\phi,t) - g^2(t)\nabla_\phi \log p_t(\phi)\right] + g(t)\bar{\eta}(t) \qquad t \equiv T - \xi$$

- Score match training

$$\mathcal{L}_\theta = \sum_{i=1}^{N} \sigma_i^2 \mathbb{E}_{p_0(\phi_0)} \mathbb{E}_{p_i(\phi_i|\phi_0)} \left[\| s_\theta(\phi_i,\xi) - \nabla_{\phi_i} \log p_i(\phi_i|\phi_0)\|_2^2\right] \qquad p_\xi(\phi_\xi|\phi_0) = \mathcal{N}\left(\phi_\xi;\phi_0,\frac{1}{2\log\sigma}(\sigma^{2\xi}-1)\mathbf{I}\right)$$

- Sample generation SDE

$$\frac{d\phi}{dt} = \left[f(\phi,t) - g^2(t)s_{\hat{\theta}}(\phi,t)\right] + g(t)\bar{\eta}(t).$$

L. Wang, G. Arts, K. Zhou, JHEP 05(2024) 060

Denoising within DM as Stochastic Quantization

- Backward diffusion SDE in **variance expanding** scheme (i.e., vanishing drift in Forward)

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

- Redefine time $\tau \equiv T - t$ and denoting $g_\tau = g(T-\tau),\ q_\tau(\phi) = p_{T-\tau}(\phi)$

$$\frac{d\phi}{d\tau} = g_\tau^2 \nabla_\phi \log q_\tau(\phi) + g_\tau \bar{\eta}(\tau)$$

$$\phi(\tau_{n+1}) = \phi(\tau_n) + g_{\tau_n}^2 \nabla_\phi \log q_{\tau_n}[\phi(\tau_n)]\Delta\tau + g_{\tau_n}\sqrt{\Delta\tau}\,\bar{\eta}(\tau_n)$$

- The corresponding Fokker-Planck equation and equilibrium

$$\frac{\partial p_\tau(\phi)}{\partial \tau} = \int d^n x \left\{ g_\tau^2 \frac{\delta}{\delta\phi}\left(\bar{\alpha}\frac{\delta}{\delta\phi} + \nabla_\phi S_{\rm DM}\right)\right\} p_\tau(\phi),$$

$$\nabla_\phi S_{\rm DM} \equiv -\nabla_\phi \log q_\tau(\phi)$$

$$p_{\rm eq}(\phi) \propto e^{-S_{\rm DM}/\bar{\alpha}}$$

- A flow of effective action will be learned in DMs

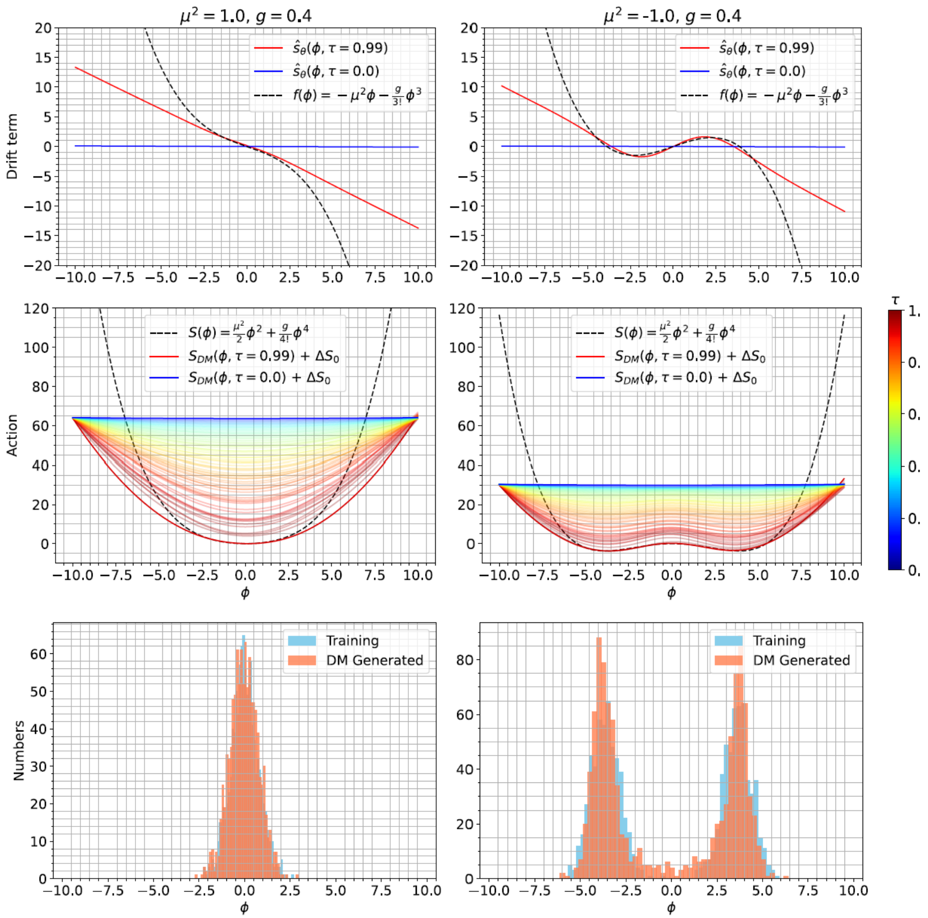
$$p_{\tau=T}(\phi) \to P[\phi, T]$$

$$O(\bar{\alpha}) \sim O(\hbar)$$

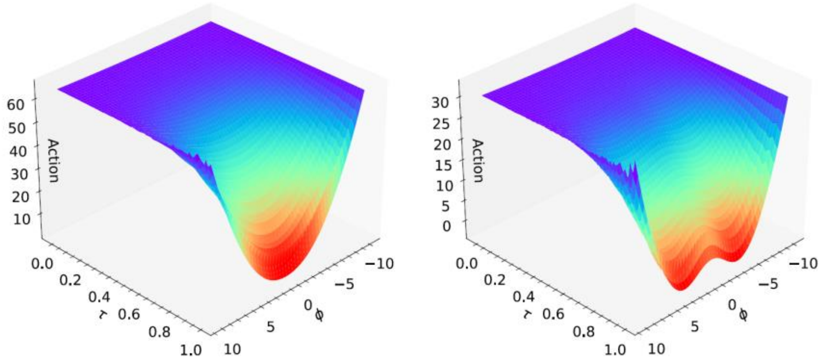
sampling from a DM is equivalent to optimizing a stochastic trajectory to approach the "equilibrium state"

L. Wang, G. Arts, K. Zhou, JHEP 05 (2024) 060

Effective Action on toy model



- Flow of an effective action



DM on scalar phi4 model

- Consider a real scalar field with action:

$$S = \int \mathrm{d}^d x \mathrm{d}t \mathcal{L} = \int \mathrm{d}^d x \mathrm{d}t (\frac{1}{2}(\partial^2 \phi_0^2 - m^2 \phi_0^2) - \frac{\lambda_0}{4!}\phi_0^4),$$

- In Euclidean space, the discretized action with dimension less form:

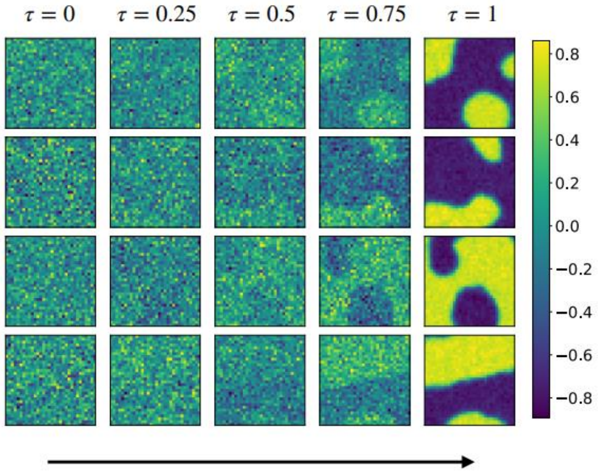
$$S_E = \sum_x [-2\kappa \sum_{\mu=1}^{d} \phi(x)\phi(x+\hat{\mu}) + (1-2\lambda)\phi(x)^2 + \lambda\phi(x)^4].$$

$$a^{\frac{d-2}{2}}\phi_0 = (2\kappa)^{1/2}\phi, (am_0)^2 = \frac{1-2\lambda}{\kappa} - 2d, a^{-d+4}\lambda_0 = \frac{6\lambda}{\kappa^2},$$

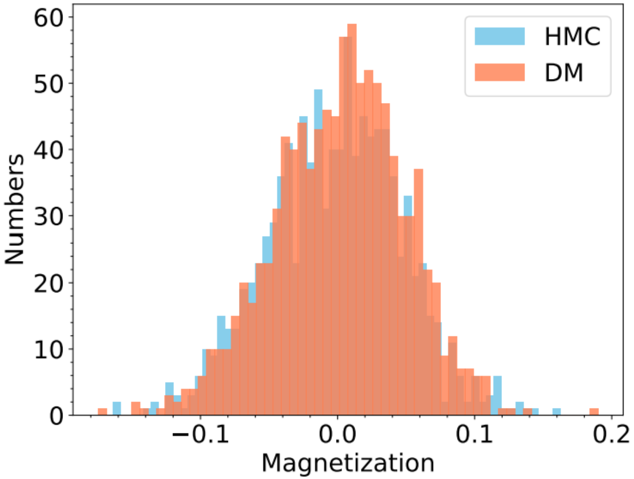
- Broken phase and symmetric phase

$$\kappa_c(\lambda) = \frac{1-2\lambda}{2d}.$$

- 2d 32x32 lattice size, HMC generated 5120 configurations for training
- Broken phase :   symmetric phase :



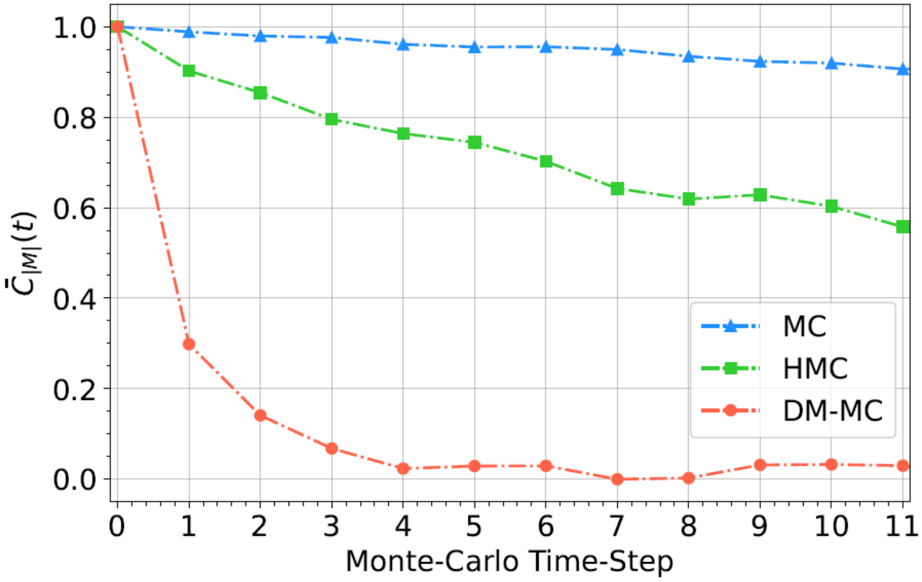
numerous “bulk” patterns emerge

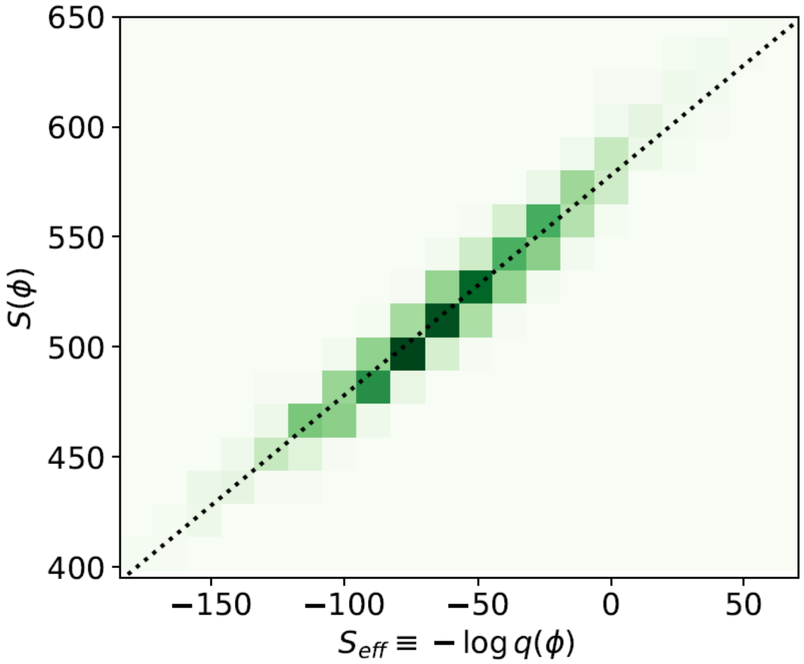


data-set	$\langle M \rangle$	χ_2	U_L
Training (HMC)	0.0012± 0.0007	2.5160 ± 0.0457	0.1042 ± 0.0367
Testing (HMC)	0.0018 ± 0.0015	2.4463 ± 0.1099	-0.0198 ± 0.1035
Generated (DM)	0.0017± 0.0015	2.4227 ± 0.1035	0.0484 ± 0.0959

Results: Autocorrelation time and final captured eff action

validation R2 ~0.96





Relation to (inverse) RG

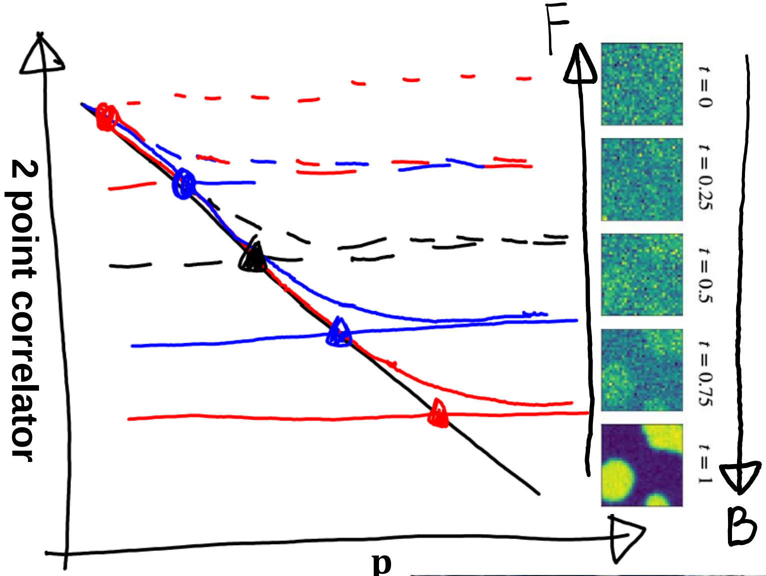
- Forward diffusion kernel: gaussian smoothing

$$p_\xi(\phi_\xi|\phi_0) = \mathcal{N}\left(\phi_\xi; \phi_0, \frac{1}{2\log\sigma}(\sigma^{2\xi}-1)\mathbf{I}\right)$$

$$\phi_\tau(\mathbf{x}) \;=\; \phi_0(\mathbf{x}) + \sqrt{\frac{\sigma^{2\tau}-1}{2\log\sigma}}\epsilon(\mathbf{x}) \;\;\text{with}\;\; \epsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})$$

- In Fourier space:

$$\phi_\tau(p) = \phi_0(p) + \sqrt{\frac{\sigma^{2\tau}-1}{2\log\sigma}}\epsilon(p)$$

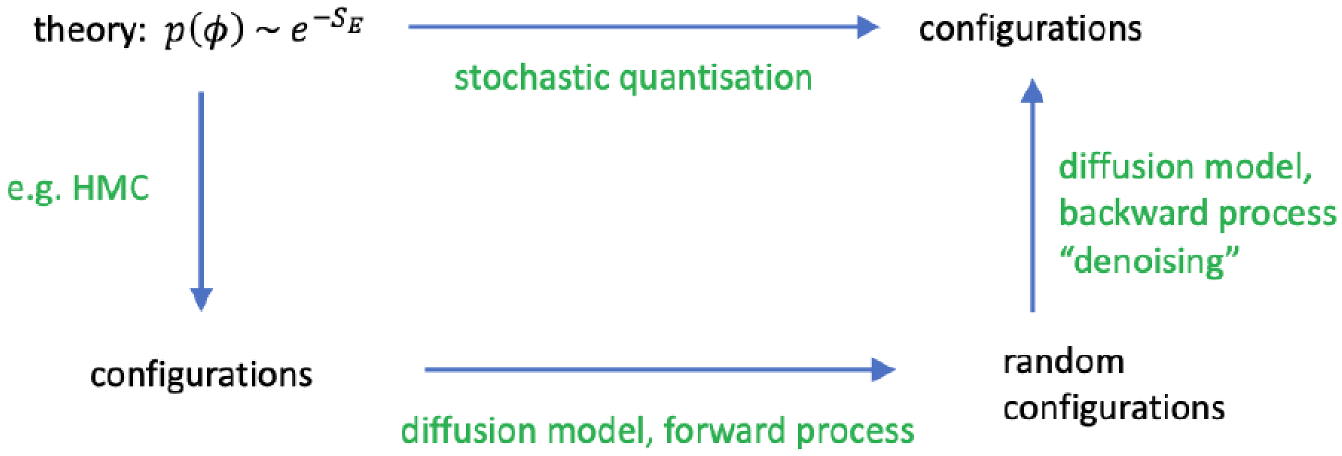


- ! the above evolution will perturb (smear) higher momentum modes faster because of the gradually increasing noise level

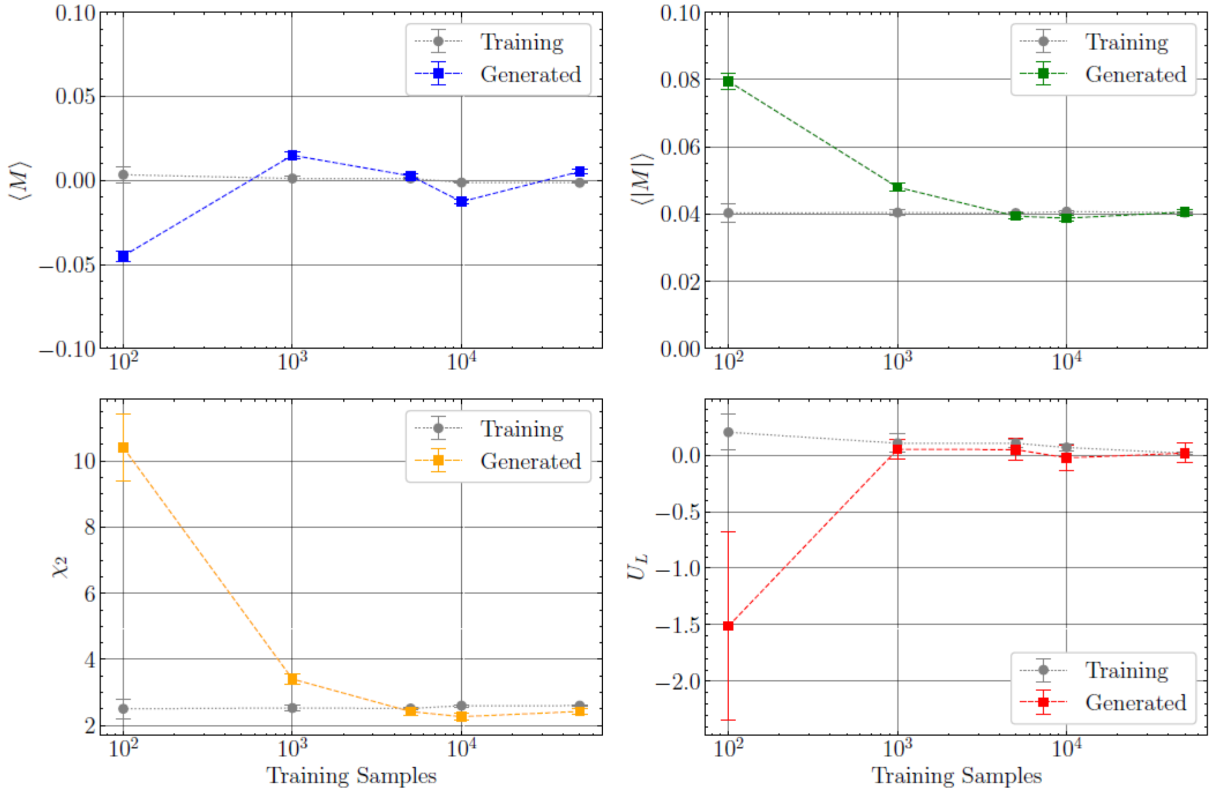
In FRG, the high frequency (short-distance) degrees of freedom is progressively integrated out !

See Semon's and Mathis's talk!

Summary


theory: $p(\phi) \sim e^{-S_E}$
configurations
stochastic quantisation
e.g. HMC
diffusion model,
backward process
“denoising”
configurations
random
configurations
diffusion model, forward process

Training efficiency



Acceptance rate