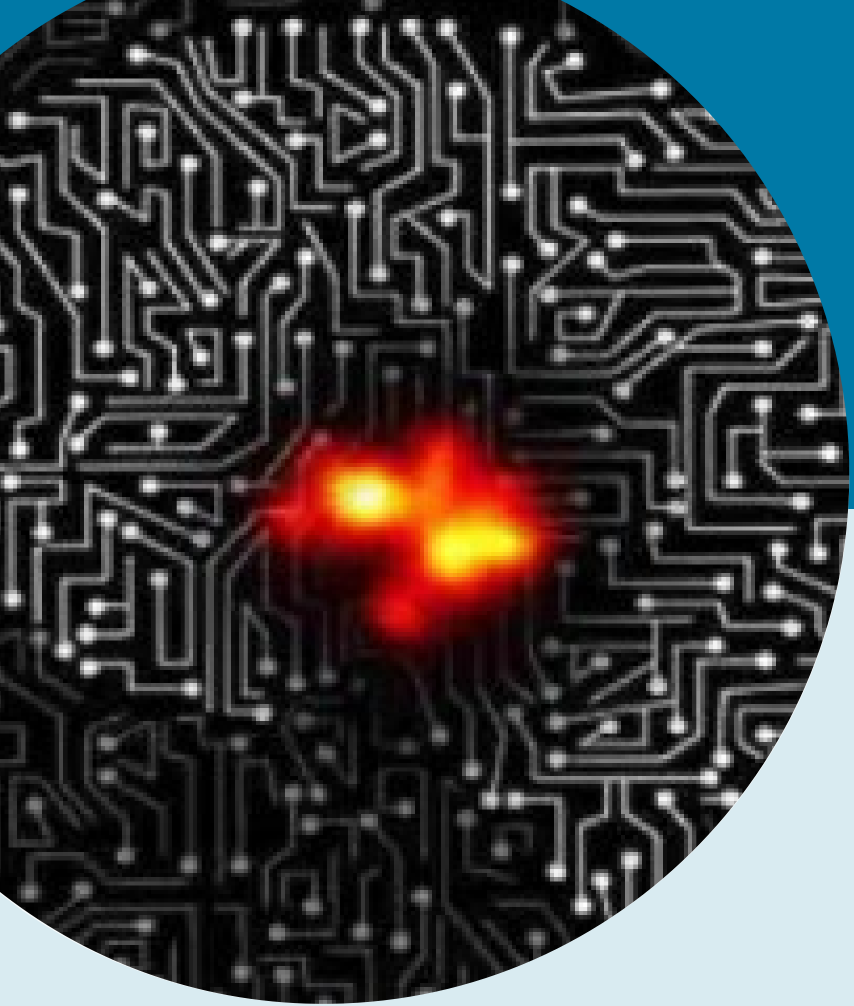


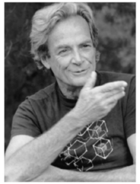
香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

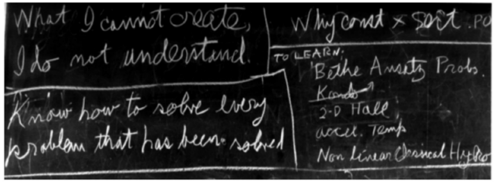
Generative Models meets Physics

Kai Zhou (CUHK - Shenzhen)

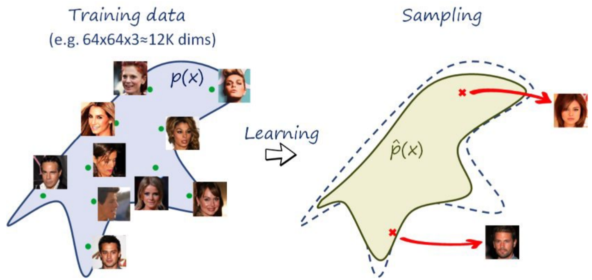
极端核物质前沿研讨会，宜昌，中国

Generative Models





"What I can not create, I do not understand"



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 :

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

Often use NN to parametrize transformation $\log p_\theta(\mathbf{x}) = \log p_\theta(f_\theta(\mathbf{z}_0))$

- Maximize Likelihood Estimation : **(given training samples)**

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

Reverse KL Divergence : Sample many $\mathbf{z}_0 \sim p_0(\mathbf{z}_0)$
(given unnormalized target distribution, e.g., Action)

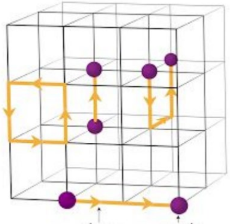
$$\theta^* = \arg\min_\theta \frac{1}{N}\sum_{i=1}^{N}[\log p_\theta(f_\theta(\mathbf{z}_0)) - \log \tilde{p}_{target}(f_\theta(\mathbf{z}_0))]$$

Why GenAI matters for – **high-dim distribution problems**

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

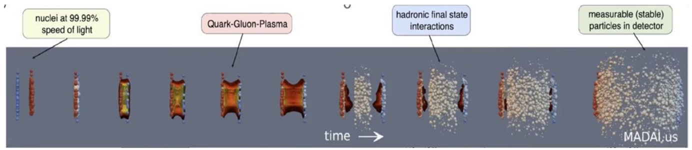
Lattice QFT

critical slowing down
topological freezing



HIC Simulation

Multi-stage EbE slow
Fluctuations/correlations



Inverse Inference

Recover physics from noisy data
Calibration and uncertainty

lattice sampler
$$p(\phi) \propto e^{-S[\phi]}$$

→

HIC surrogate
$$p(\text{event} \mid c)$$

→

posterior engine
$$p(\theta \mid y)$$

Bottlenecks differ, but all revolve around high-dimensional probability measures.

GenAI - Transport between probability measures

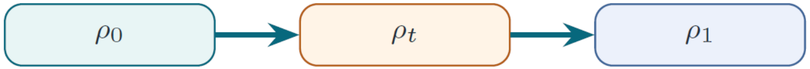
1. Generation = sampling

$$x_0 \sim \rho_0 \xrightarrow{\text{learned dynamics}} x_1 \sim \rho_1.$$

- data-driven: $\rho_1 = p_{\mathrm{data}}(x \mid y)$;
- action-driven: $\rho_1 = Z^{-1} e^{-S(x)}$;
- inference-driven: $\rho_1 = p(\theta \mid y)$.

2. A probability path is chosen

$\{\rho_t\}_{t\in[0,1]}$, ρ_0 simple, ρ_1 physical target.



path in distribution space

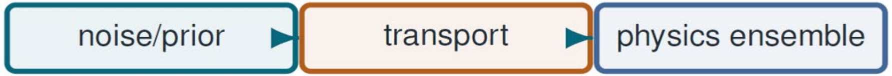
3. Different GenAI models realize the path differently

flow / FM:
$$dx_t = v_t(x_t)dt,$$
$$\partial_t \rho_t = -\nabla\cdot(\rho_t v_t),$$

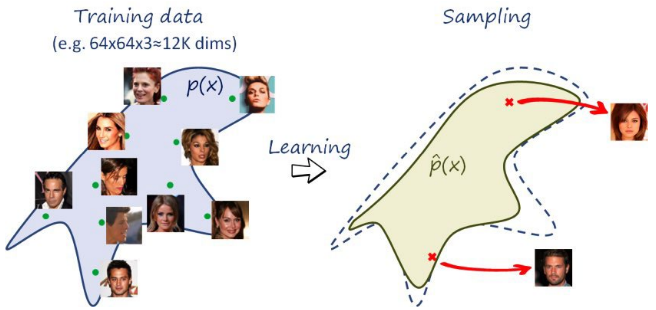
diffusion:
$$dx_t = b_t(x_t)dt + g_t dW_t,$$
$$\partial_t \rho_t = -\nabla\cdot(b_t \rho_t) + \tfrac{1}{2}\nabla^2(g_t^2 \rho_t).$$

Physical reading

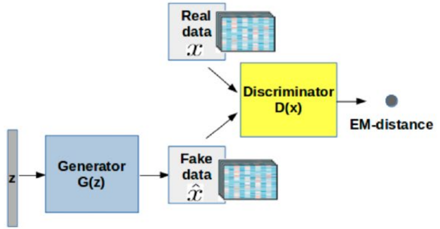
- **state** = fields, particles, detector responses, or inferred parameters;
- **probability path** = evolution of uncertainty from prior/noise to physics ensemble;
- **dynamics** = transport that should respect action, symmetry, conservation laws, and response;
- **output** = an ensemble/posterior with calibrated observables, not a single plausible picture.

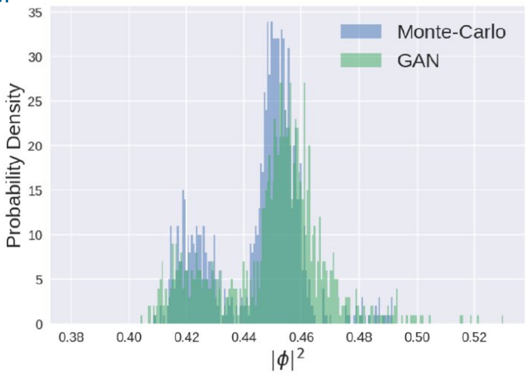


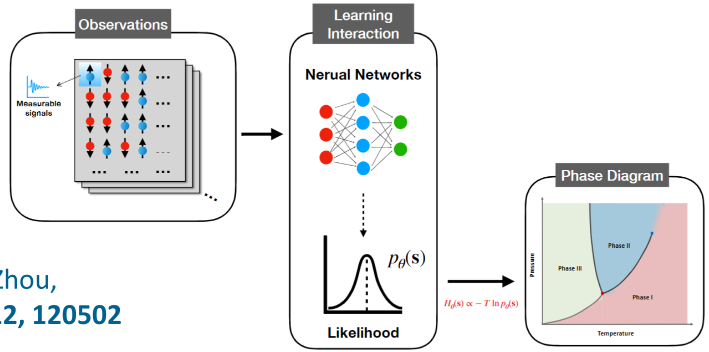
Data-driven lattice GenAI – early proof of principle from ensembles

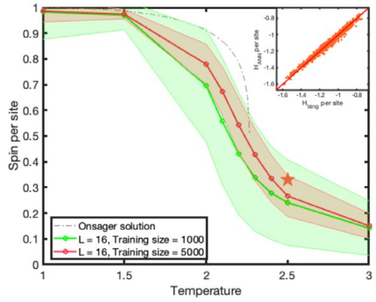


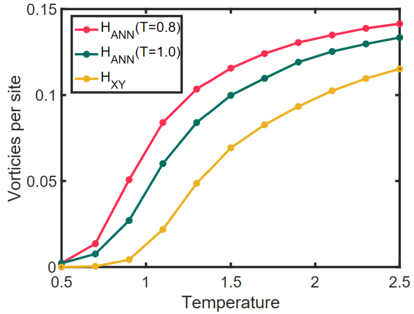
K. Zhou, G. Endrődi, L.-G. Pang, and H. Stöcker
PRD 100, 011501 (2019)











L. Wang, L. He, Y. Jiang, K. Zhou,
Chin.Phys.Lett. 39 (2022) 12, 120502

T. Xu, L. Wang, L. He, K. Zhou, Y. Jiang,
Chin.Phys.C 48 (2024) 10, 103101, arXiv:2007.01037

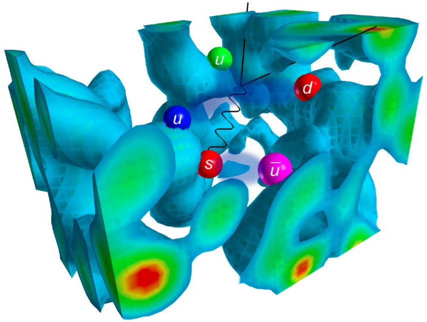
Diffusion Model

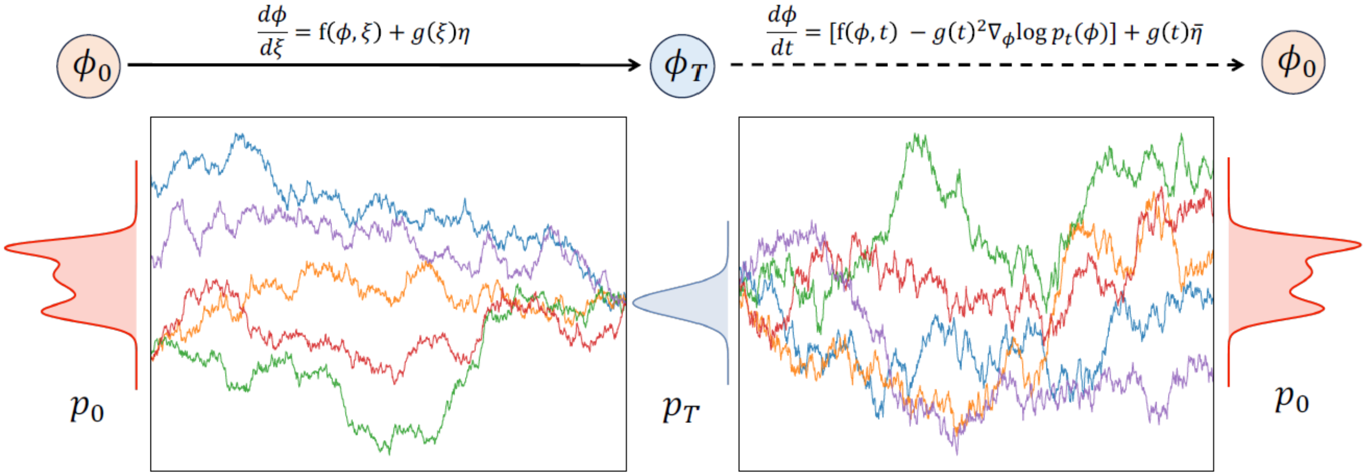
"A heavy quark move inside quark-gluon plasma"

Diffusion Model on lattice QFT configurations

$$p(\phi) = e^{-S(\phi)}/Z$$

$$\langle O \rangle \approx \frac{1}{N}\sum_i O_i$$



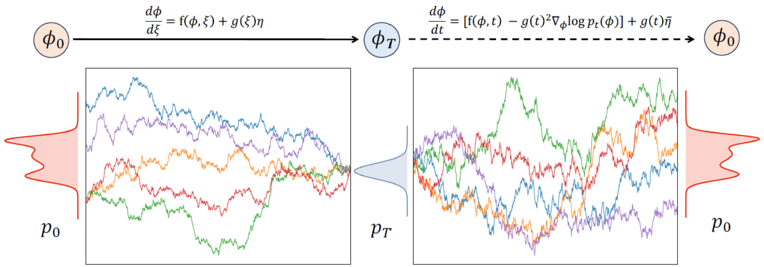


L. Wang, G. Arts, K. Zhou, JHEP 05 (2024) 060
L. Wang, G. Arts, K. Zhou, arXiv:2311.03578 (NeurIPS 2023 workshop "ML&Physical Sciences")
G. A, D. E. H, L. W, K. Z, arXiv:2410:21212 (NeurIPS 2024 workshop "ML&Physical Sciences) → **"Best Physics for AI Paper" Award**
Q. Zhu, G. Aarts, W. Wang, K. Zhou, L. Wang, arXiv:2410.19602 (NeurIPS 2024 workshop "ML&Physical Sciences)

Diffusion Model as Stochastic Quantization

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

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



- Backward diffusion SDE

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

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

- **DM generation SDE and Stochastic Quantization**

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

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

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

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

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

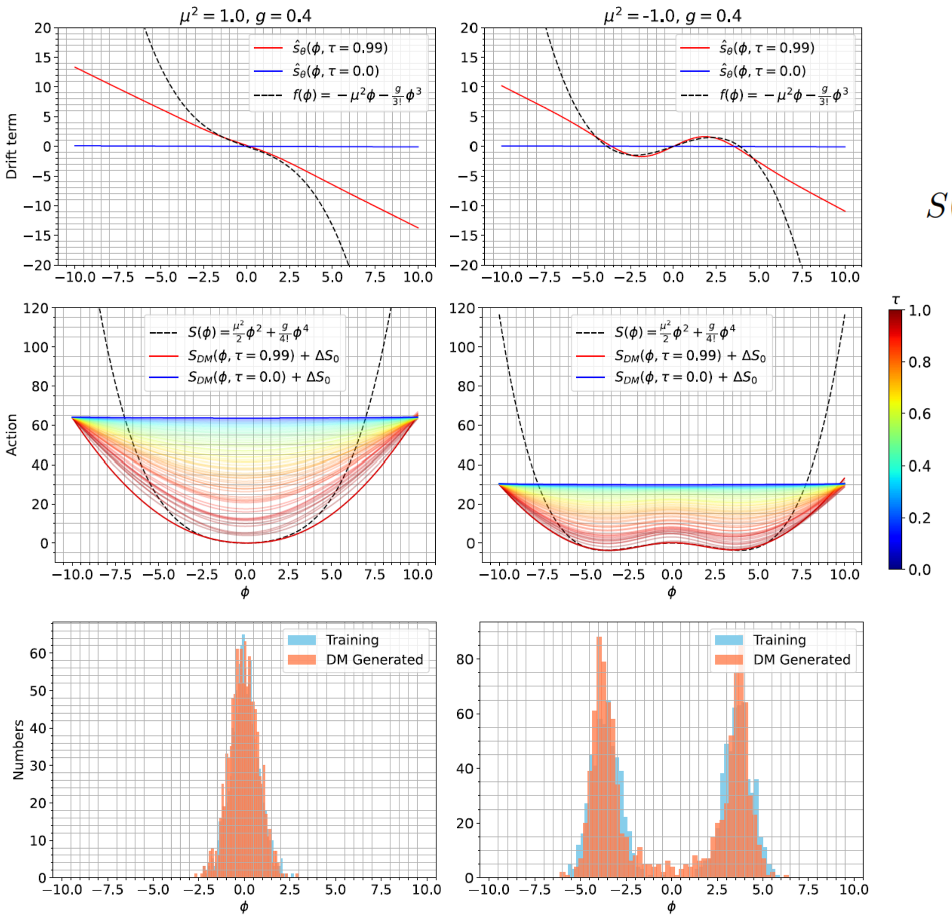
$$\frac{\partial P[\phi,\tau]}{\partial \tau} = \alpha \int d^n x \left\{ \frac{\delta}{\delta\phi}\left(\frac{\delta}{\delta\phi} + \frac{\delta S_E}{\delta\phi}\right)\right\} P[\phi,\tau]$$

- A flow of **effective action** will be learned in DMs

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

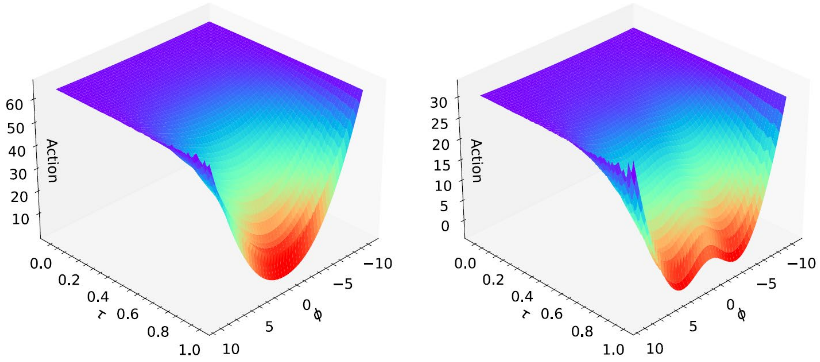
L. Wang, G. Arts, K. Zhou, JHEP 05(2024) 060

Effective Action Flow



- Flow of the effective action

$$S(\phi) = \frac{\mu^2}{2}\phi^2 + \frac{g}{4!}\phi^4, \qquad f(\phi) = -\frac{\partial S(\phi)}{\partial \phi} = -\mu^2\phi - \frac{g}{3!}\phi^3$$



$$S_{\mathrm{DM}}(\phi, \tau) = \int^{\phi} \hat{s}_\theta(\tilde{\phi}, \tau) d\tilde{\phi}$$

DM Relation to (inverse) RG

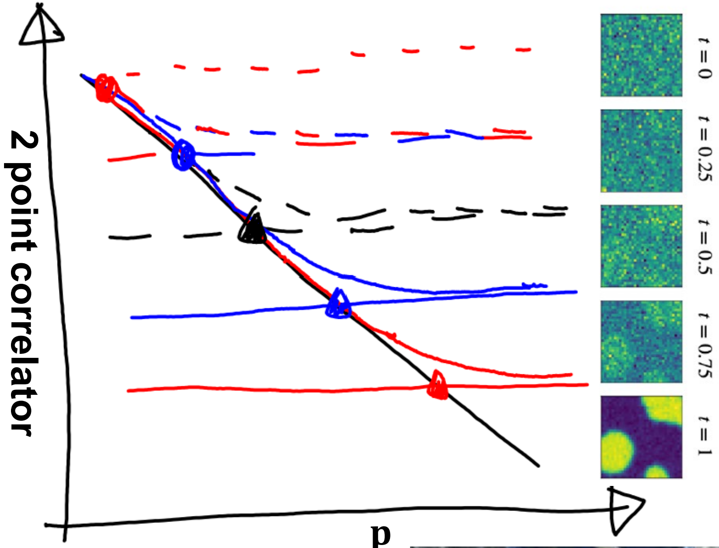
- Forward diffusion kernel: **gaussian smoothing**

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

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

- In Fourier space:

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



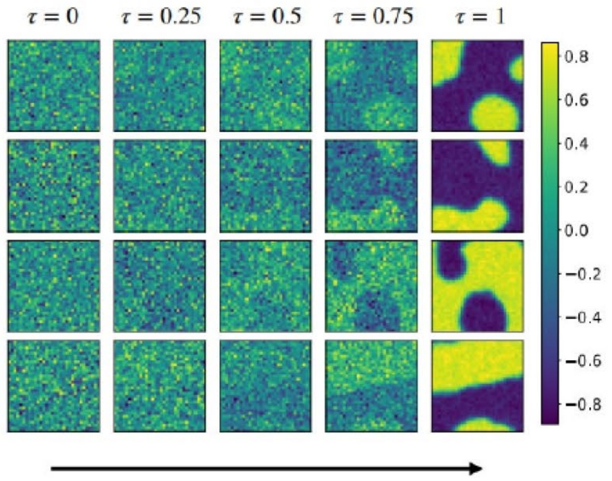
- ! the above evolution will perturb (smear) higher momentum modes first,

With decreasing cut scale because of the gradually increasing noise level !

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

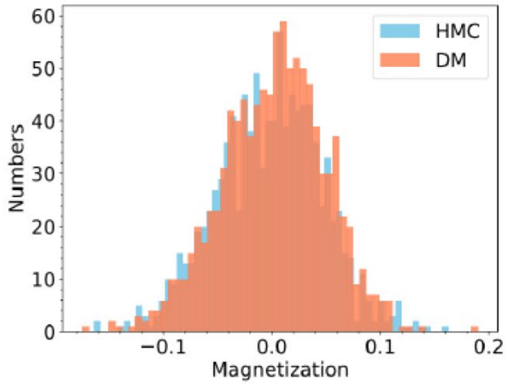
DM on 2d scalar ϕ^4 model - first nontrivial benchmark

Broken phase :



numerous “bulk” patterns emerge

symmetric phase :



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

- In the broken phase, the generated fields develop the expected bulk structures; in the symmetric phase, integrated observables agree with HMC in this benchmark setup.

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

How correlations are destroyed and rebuilt in DM ?

- forward process: $\dot{x}(t) = K(x(t), t) + g(t)\eta(t)$ $\quad 0 \leq t \leq T$
- linear (or zero) drift: $K(x(t), t) = -\frac{1}{2}k(t)x(t)$ $\quad$ noise profile $g(t) = \sigma^{t/T}$
- initial data from target ensemble $x_0 \sim P_0(x_0)$
- solution: $x(t) = x_0 f(t, 0) + \int_0^t ds\, f(t, s)g(s)\eta(s)$
- with $f(t, s) = e^{-\frac{1}{2}\int_s^t ds'\, k(s')}$
- moments $\mu_n(t) = \mathbb{E}[x^n(t)]$ and cumulants or connected n-point functions
- second moment/cumulant:

$$\kappa_2(t) = \mu_2(t) = \mu_2(0)f^2(t, 0) + \Xi(t)$$

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

- In a Gaussian theory, all connected cumulants with $n > 2$ vanish.
- In an interacting theory, nonzero $\kappa_4, \kappa_6, \ldots$ are direct fingerprints of non-Gaussian dynamics.
- So the real question is not only whether the model reproduces κ_2, but whether it reconstructs the connected hierarchy.

Why this matters for diffusion

- forward noising quickly hides non-Gaussian structure behind added variance
- the learned score must therefore carry that information implicitly
- this is a much more physical diagnostic than image-like quality metrics

Higher order cumulants

$$W[J] = \log Z[J], \qquad \kappa_n = \frac{\delta^n W[J]}{\delta J^n}\Big|_{J=0}$$

- proof to all orders: generating functionals $Z[J] = \mathbb{E}[e^{J(t)x(t)}]$ $\qquad W[J] = \log Z[J]$

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

- noise average: $$Z_\eta[J] = e^{J(t)x_0 f(t,0) + \frac{1}{2}J^2(t)\Xi(t)}$$

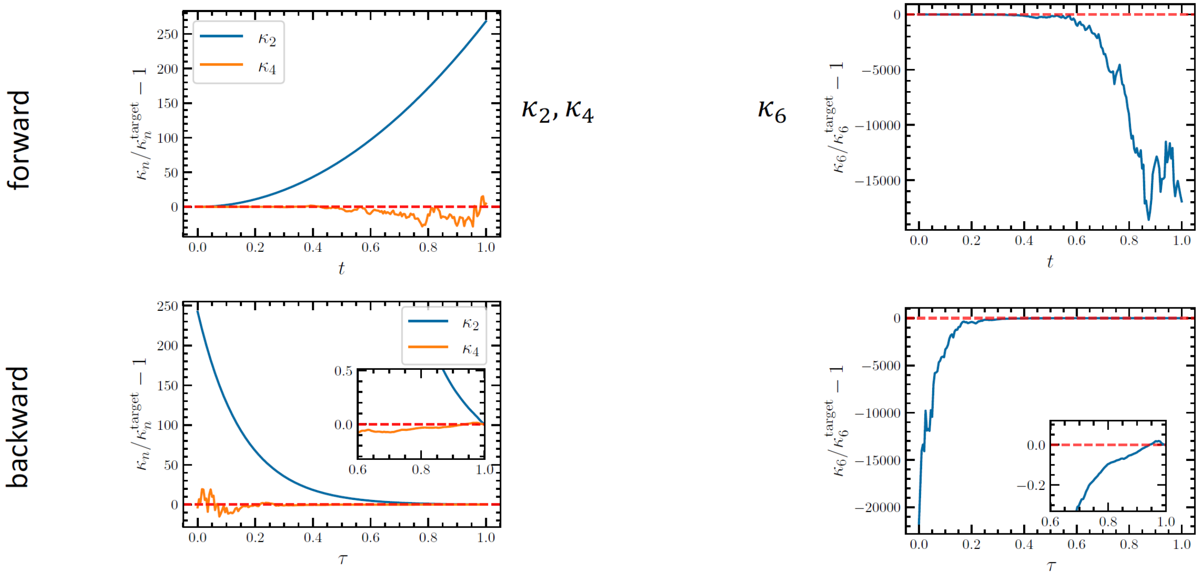
- total average: $$Z[J] = \mathbb{E}[e^{J(t)x(t)}] = e^{\frac{1}{2}J^2(t)\Xi(t)} \int dx_0\, P_0(x_0) e^{J(t)x_0 f(t,0)}$$

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

- 2nd cumulant: $$\kappa_2(t) = \frac{d^2 W[J]}{dJ(t)^2}\Big|_{J=0} = \Xi(t) + \mathbb{E}_{P_0}[x_0^2] f^2(t,0) \quad ✓$$

- higher-order cumulants: $$\kappa_{n>2}(t) = \frac{d^n W[J]}{dJ(t)^n}\Big|_{J=0} = \frac{d^n}{dJ(t)^n} \log \mathbb{E}_{P_0}[e^{J(t)x_0 f(t,0)}]\Big|_{J=0} = \kappa_n(0) f^n(t,0) \quad ✓$$

2d scalar field theory in variance expanding scheme



	κ_2	κ_4	κ_6	κ_8
HMC (normalised)	0.39597(4)	−0.29453(6)	0.90108(28)	−5.8689(25)
Diffusion model	0.39598(4)	−0.29454(7)	0.90113(32)	−5.8694(28)

$\phi^4: 32^2, \kappa = 0.4, \lambda = 0.022, 10^5$ configurations

Aarts, Habibi, Wang, Zhou, MLST 6 (2025) 025004.

Can it learn the real effective distribution behind Complex Langevin?

- With a sign problem, the original Boltzmann weight is complex and importance sampling fails.
- Complex Langevin can sample a real effective distribution $P(x, y)$ in the complexified space.
- Diffusion or energy-based models can then learn *that* effective distribution from CL trajectories.

$$\langle O(z) \rangle = \int dx\, dy\, P(x, y)\, O(x + iy)$$

This does not solve the sign problem, but it gives a new handle on what CL actually samples.

- Quartic model with a complex mass $S = \frac{1}{2}\sigma_0 x^2 + \frac{1}{4}\lambda x^4, \qquad \sigma_0 = A + iB.$
- Exact results for partition function $Z = \int dx\, e^{-S(x)} = \sqrt{\frac{4\xi}{\sigma_0}} e^{\xi} K_{-\frac{1}{4}}(\xi) \qquad \xi = \sigma_0^2/(8\lambda)$
- Sampling from CL process yields empirical **histogram 2d-distribution**, with no analytical expression

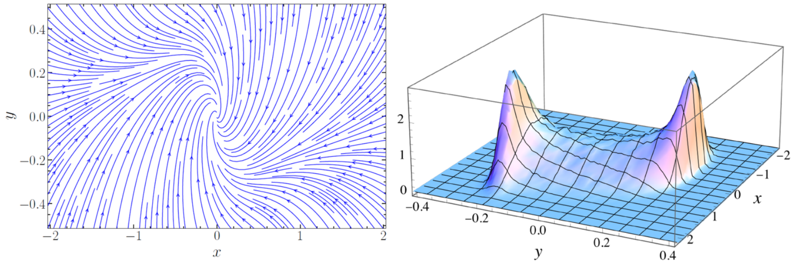


Figure 1: Complex-valued quartic model with parameters $\sigma_0 = 1 + i$ and $\lambda = 1$: CL drift in the complex plane (left) and histogram $P(x, y)$ obtained by sampling the CL process (right)

- EBM learned the Energy directly, the score (from gradient of energy) is conservative, with direct distribution

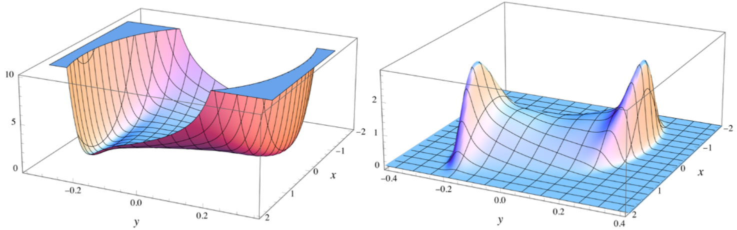


Figure 4: Quartic model: Energy $E_\theta(\mathbf{x})$ learned in the energy-based diffusion model (left) and the corresponding distribution $p_\theta(\mathbf{x}) \sim \exp[-E_\theta(\mathbf{x})]$ (right).

first time that a parametrization of distribution from a CL process is obtained in non-trivial case wo histrogram way

G. Aarts, D. H, L. W and K. Zhou, **JHEP12(2025)160**

Given Hamiltonian/Action – Variational Free Energy

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

- Reverse KL divergence

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

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

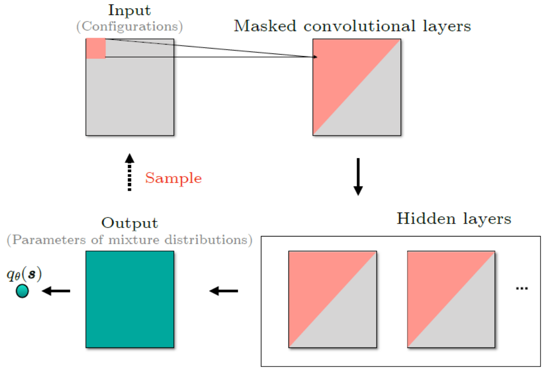
- Autoregressive

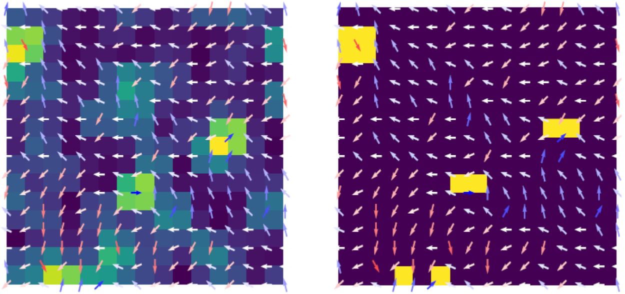
$$q_\theta(\mathbf{s}) = \prod_{i=1}^{N} q_\theta(s_i \mid s_1, \ldots, s_{i-1})$$

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)**





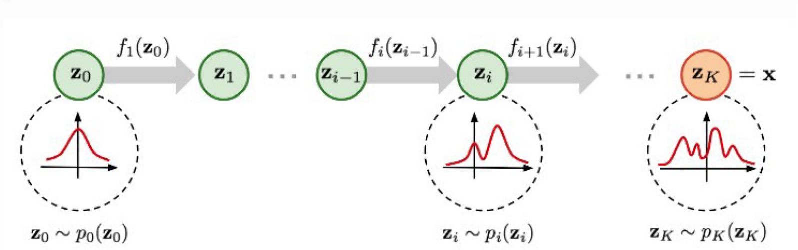
Probability distributions from CANs

Vortices

Flow based generative model given unnormalized distribution

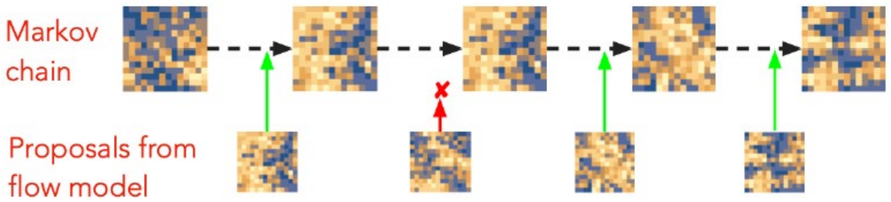
A series **(Flow)** of invertible/bijective transformations for p(z)

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



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

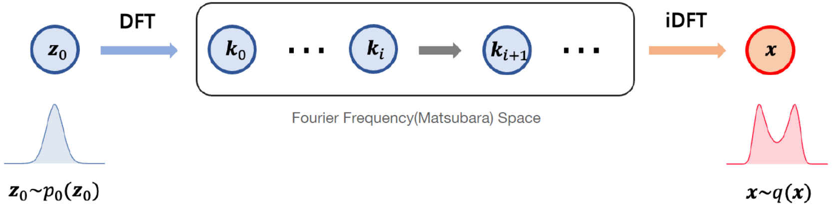
$$\rightarrow \log p(\mathbf{x}) = \log p_0(f^{-1}(\mathbf{x})) + \sum^K \log|\det J_{f_i^{-1}}| = \log p_0(\mathbf{z}_0) - \sum^K \log|\det J_{f_i}|$$

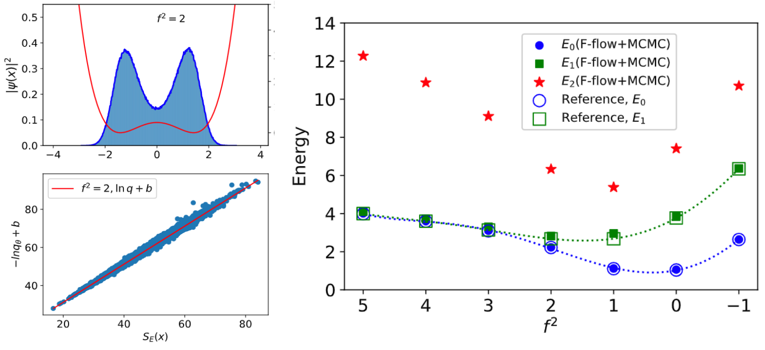


Albergo +, 1904.12072; Boyda +, 2008.05456; Favoni +, 2012.12901; Abbott +, 2208.03832; Abbott +, 2211.07541; Abbott +, 2305.02402; Bulgarelli+ 2412.00200 (SU(3)); Abbott +, arXiv:2502.00263

K.C, G. K., S. R., D. R., P. S., **Nature Reviews Physics** 5, 526-535 (2023)

Fourier Flow Model





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

Stochastic Path Sampler for HMC-data-free GenAI

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

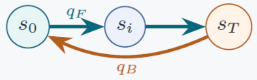
Target and path

Known physics target, unknown normalizer:

$$\widehat{p}(\phi) = e^{-S(\phi)}, \qquad p = \widehat{p}/Z.$$

SPS learns a stochastic trajectory

$$\tau : s_0 \sim q_0 \longrightarrow s_T \equiv \phi.$$



Trajectory balance

Instead of fitting a local score, match whole path measures:

$$D_{\mathrm{KL}} = \mathbb{E}_{q_F}\left[\log \frac{q_F(\tau)}{q_B(\tau)}\right].$$

The desired endpoint condition is

$$\frac{q_0 \prod_i q_F}{\prod_i q_B} \propto \widehat{p}(s_T).$$

global transport **no MCMC labels**

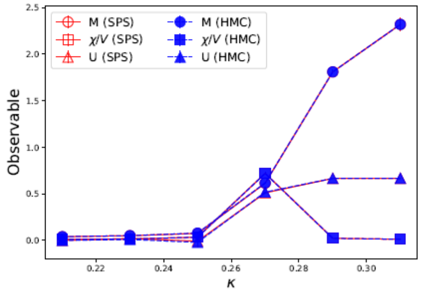
Physics leverage

- Dataless: only pointwise evaluations of $S(\phi)$ are required.
- Global: long-horizon credit assignment helps multimodal or symmetry-related sectors.
- Symmetry injection: random physical symmetry moves can be inserted along trajectories.
- Validation: compare $|M|$, χ, Binder cumulant and $G(r)$ with HMC.

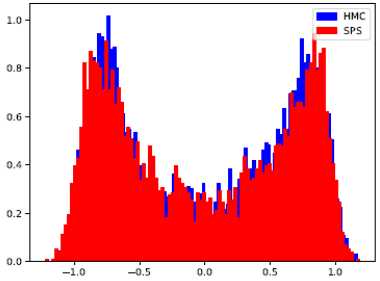
Forward stochastic path

$$s_{i+1} = s_i + \sigma_i^2 K_{\theta,F}(s_i, t_i)\Delta t + \sigma_i \xi_i \sqrt{\Delta t}.$$

A backward path uses another learned drift $K_{\theta,B}$ on the same trajectory space.



(b) $L = 32$



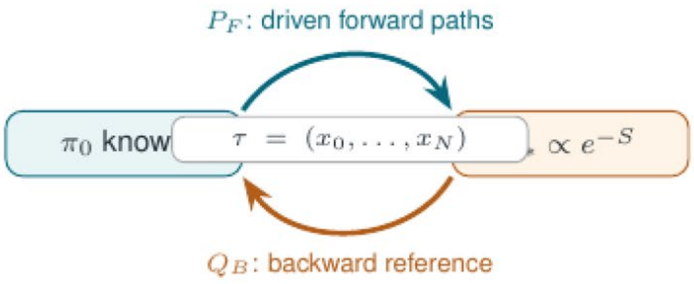
(b) $L = 32$

S. Chen, M. Qian, B. Lucini, G. Aarts, K. Zhou, **in preparation**

Forward/backward path measures

$$P_F(\tau) = \pi_0(x_0) \prod_{k=0}^{N-1} q_{F,k}(x_{k+1} \mid x_k)$$

$$\widetilde{Q}_B(\tau) = e^{-S(x_N)} \prod_{k=0}^{N-1} q_{B,k}(x_k \mid x_{k+1}), \qquad Q_B = \widetilde{Q}_B / Z.$$



Work and entropy production

$$W(\tau) = \log \frac{P_F(\tau)}{\widetilde{Q}_B(\tau)}, \qquad \Sigma(\tau) = \log \frac{P_F(\tau)}{Q_B(\tau)} = W(\tau) + \log Z.$$

W **is the nonequilibrium work-like functional:** a dimensionless accumulated log-ratio along a driven path, analogous to βW_{phys} in stochastic thermodynamics.

Change-of-measure identities

$$\left\langle e^{-\Sigma} \right\rangle_{P_F} = 1, \qquad \left\langle e^{-W} \right\rangle_{P_F} = Z,$$

$$\langle \Sigma \rangle_{P_F} = D_{\mathrm{KL}}(P_F \| Q_B) \geq 0.$$

SPS / trajectory balance: suppress fluctuations of $W(\tau)$ so that forward and backward path measures become more reversible.

The nonequilibrium view gives diagnostics: work, entropy production, path-space KL, and reversibility.

M. Qian, S. Chen, B. Lucini, G. Aarts, K. Zhou, **in preparation**

Forward noising : $\boxed{\frac{d\phi}{dt} = g(t)\,\eta(t)}$, $\phi(t=0) \sim \frac{1}{Z} e^{-S[\phi]} = P_*(\phi)$

$$\Rightarrow P_t(\phi) = \int P_*(\phi_0)\, N(\phi; \phi_0, g(t))\, d\phi_0$$

F-P equation : $\frac{\partial P_t(\phi)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial \phi^2} \left(g^2(t) P_t(\phi) \right)$

$$\Rightarrow \partial_t \log P_t(\phi) = \frac{\partial_t P_t(\phi)}{P_t} = \frac{1}{2} g^2(t) \frac{\partial^2_{\phi\phi} P_t(\phi)}{P_t}$$

$$= \frac{1}{2} g^2(t) \left(\partial^2_{\phi\phi} \log P_t(\phi) + (\partial_\phi \log P_t(\phi))^2 \right)$$

$$= \frac{1}{2} g^2(t) \left(\partial_\phi S_\theta(\phi, t) + S_\theta(\phi, t)^2 \right)$$

The Chinese University of Hong Kong, Shenzhen

Take derivative w.r.t. ϕ $\Longrightarrow$

$$\frac{\partial s_\theta(\phi,t)}{\partial t} = \frac{1}{2} g^2(t) \left(\frac{\partial^2 s_\theta(\phi,t)}{\partial \phi^2} + 2 s_\theta(\phi,t) \frac{\partial s_\theta(\phi,t)}{\partial \phi} \right)$$

Backward denoising (Generation) process :

$$\frac{d\phi}{dt} = -\frac{1}{2} g^2(t) s_\theta(\phi,t) , \quad \phi(t=T) \sim \mathcal{N}(0, g(t))$$

$$= v_\theta(\phi,t)$$

GenAI – flow PDEs

General SDE ladder

假设各向同性扩散：a_t = α(t)I

1 SDE + Fokker-Planck

$$dX_t = b_t(X_t)\,dt + \sigma_t(X_t)\,dW_t, \quad a_t = \sigma_t\sigma_t^\top$$
$$\partial_t\rho_t = -\nabla\cdot(b_t\rho_t) + \tfrac{1}{2}\partial_i\partial_j(a_{ij,t}\rho_t)$$

↓

2 log transform: ρ_t = exp(-S_t)

$$\partial_t S_t = \nabla\cdot b_t - b_t\cdot\nabla S_t + \frac{\alpha(t)}{2}\left(\Delta S_t - \|\nabla S_t\|^2\right)$$

标量 action 比直接向量场少掉旋转 / gauge-like 冗余。

↓

3 score 与 probability-flow velocity

$$s_t = \nabla\log\rho_t = -\nabla S_t, \qquad v_t = b_t - \frac{\alpha(t)}{2}s_t$$
$$\partial_t s_t = \nabla\left[-\nabla\cdot b_t - b_t\cdot s_t + \frac{\alpha(t)}{2}\left(\nabla\cdot s_t + \|s_t\|^2\right)\right]$$

采样 ODE/SDE 只需 score；物理求解更适合从 S_t 或势函数出发。

VE DM case

b_t=0, α(t)=g(t)^2

forward heat semigroup

$$dX_t = g(t)dW_t, \quad \tau = \int_0^t g(r)^2dr = \sigma(t)^2$$
$$\partial_\tau\rho_\tau = \tfrac{1}{2}\Delta\rho_\tau$$

effective action PDE = HJB / KPZ

$$\boxed{\partial_\tau S_\tau = \tfrac{1}{2}\left(\Delta S_\tau - \|\nabla S_\tau\|^2\right)}$$

Cole-Hopf: ρ=e^{-S} 让热方程变成非线性 action flow。

score / Burgers / sampling velocity

$$s_\tau = -\nabla S_\tau, \quad v_\tau = -\tfrac{1}{2}s_\tau = \tfrac{1}{2}\nabla S_\tau$$
$$u_\tau := \nabla S_\tau \quad \Rightarrow \quad \partial_\tau u_\tau + (u_\tau\cdot\nabla)u_\tau = \tfrac{1}{2}\Delta u_\tau$$

Burgers 形式说明 score 场不是任意向量场，而是梯度流。

Outcome: path → PDE type → feasible solver

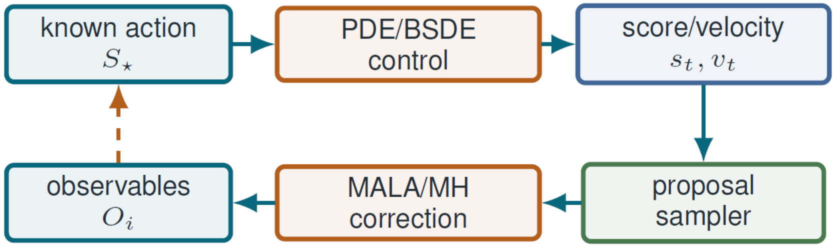
GenAI – flow PDEs

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

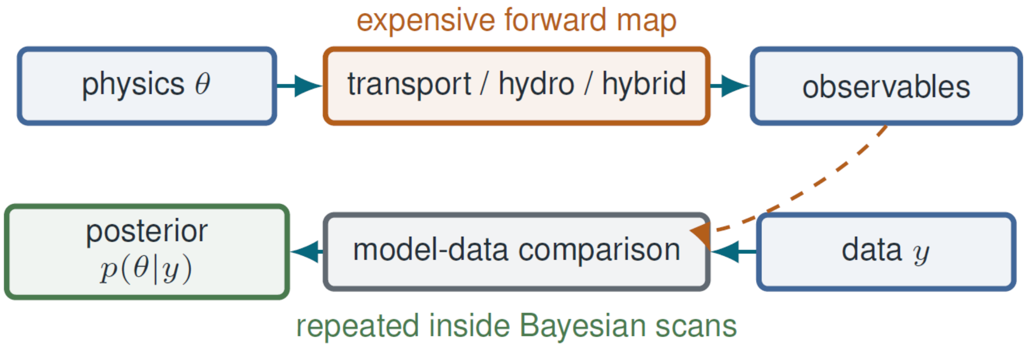
模型	path 动力学	scalar action S_t	score PDE	velocity PDE
VE	$dX = g(t)dW$	$\partial_\tau S = \frac{1}{2}(\Delta S - \|\nabla S\|^2)$	$\partial_\tau s = \frac{1}{2}\nabla(\nabla \cdot s + \|s\|^2)$	$v = -\frac{1}{2}s,\ \partial_\tau v + 2(v \cdot \nabla)v = \frac{1}{2}\Delta v$
DDPM/VP	$dX = -\frac{1}{2}\beta X dt + \sqrt{\beta} dW$	$\partial_t S = -\kappa d + \kappa x \cdot \nabla S + \kappa \Delta S - \kappa\|\nabla S\|^2$	$\partial_t s = \kappa[s + (x \cdot \nabla)s + \nabla(\nabla \cdot s + \|s\|^2)]$	$v = -\kappa(x + s)$
Flow matching	$X_t = (1 - t)X_0 + tX_1$	$\partial_t S = \nabla \cdot v - v \cdot \nabla S$	$\partial_t s = -\nabla(\nabla \cdot v + v \cdot s)$	$\partial_t v + (v \cdot \nabla)v = -\rho^{-1}(\rho\Sigma)$
Schrödinger bridge	$dX = \varepsilon \nabla \eta\, dt + \sqrt{\varepsilon} dW$	$\partial_t S = \Delta\psi - \nabla S \cdot \nabla\psi$	$\partial_t s = -\nabla(\Delta\psi + \nabla\psi \cdot s)$	$\partial_t v + (v \cdot \nabla)v = -\frac{\varepsilon^2}{2}\nabla(\Delta\sqrt{\rho}/\sqrt{\rho})$

From functional PDE to tractable control

- solve layerwise and center unknown normalizers;
- fit scalar potentials or local operator coefficients, not arbitrary high-dimensional vector fields;
- sample collocation points from the current typical set;
- use Hutchinson/JVP estimators for traces and Burgers terms;
- close with MALA/MH and physical observables.

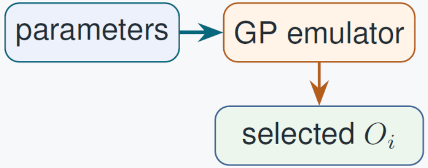


GenAI for HIC physics bottleneck – inference loop



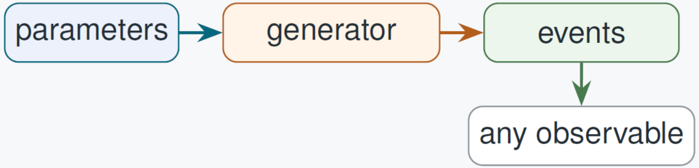
- Infer EoS, initial conditions, transport coefficients, and phase structure from observables.
- Forward map: transport, hydrodynamics, hybrid evolution, detector response.
- Bottleneck: repeated simulator calls inside likelihood or posterior loops.

Observable emulator



- Efficient for a small observable vector.
- Harder to scale when many differential observables, correlations, and cumulants are needed.

Event-level generator



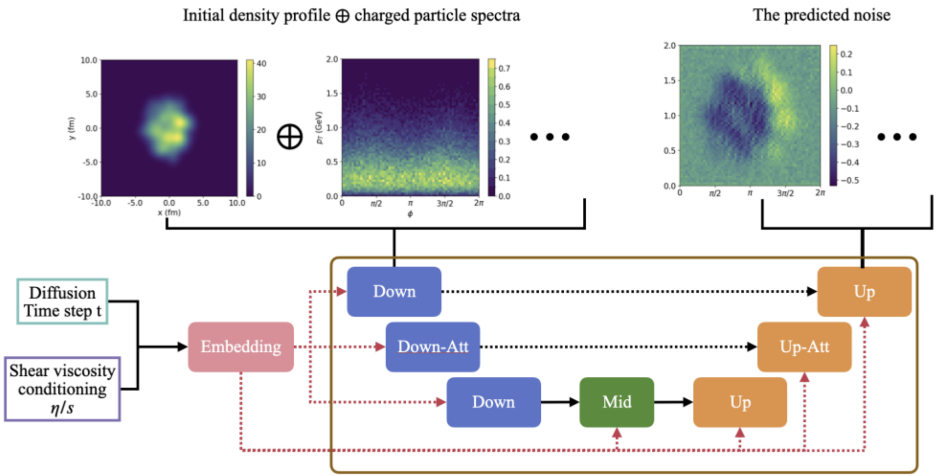
- Samples complete events; observables are computed afterwards.
- Better aligned with event-by-event Bayesian inference and multi-observable systematics.

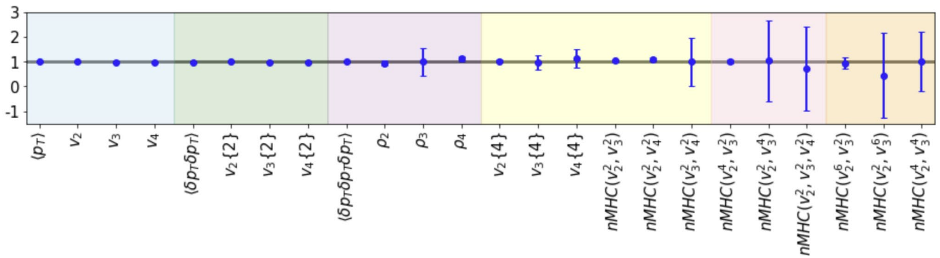
Generative diffusion model to heavy-ion collisions

An end-to-end generative diffusion model for heavy-ion collisions

Phys.Rev.C 112 (2025) 5, L051903

Jing-An Sun,[1,2] Li Yan,[1,3] Charles Gale,[2] and Sangyong Jeon[2]





tor. We carried out (2+1)D minimum bias simulations of Pb-Pb collisions at 5.02 TeV, choosing the shear viscosity η/s to be one of three distinct values: 0.0, 0.1, and 0.2. For each value of η/s, we generate 12,000 pairs of initial entropy density profiles and final particle spectra, corresponding to 12,000 simulated events, as the training dataset. 70% of the total events are used for training and the rest are used for validation.

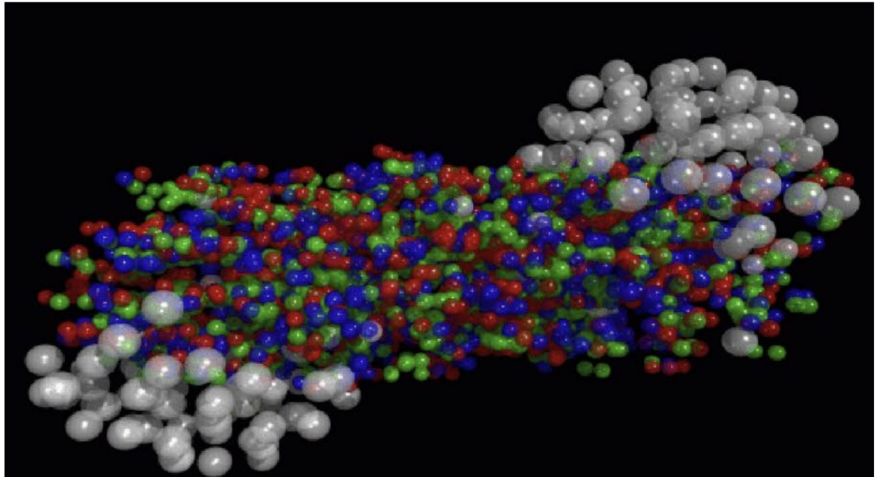
Considering that the spectra $\boldsymbol{S}_0$ depend on the initial entropy density profiles $\boldsymbol{I}$ and the shear viscosity η/s, we train a conditional reverse diffusion process $p(\boldsymbol{S}_0|\boldsymbol{I}, \eta/s)$ without modifying the forward process.

one single central collision event in just 10^{-1} seconds on a GeForce GTX 4090 GPU.

ble precision, as the traditional numerical simulation of hydrodynamics for one central event typically takes approximately 120 minutes (10^4 seconds) on a single CPU.

Point Cloud Diffusion Model for HICs – UrQMD cascade model

- Event-by-event collision output
- Microscopic non-equilibrium description
- hadrons on classical trajectories
  - stochastic binary scatterings
  - color string formation
  - resonance excitation and decays

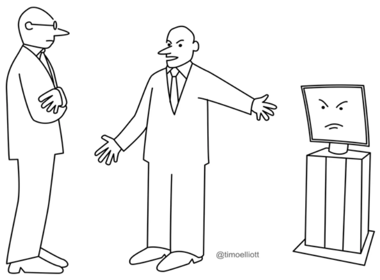

https://itp.uni-frankfurt.de/~bleicher/urqmd-cern.gif

- interactions based on scattering cross sections
- default setup effective EoS: Hadron Resonance Gas
- Non-trivial interactions can be added through QMD approach

Can we emulate UrQMD with DL?

M. O.K, K. Zhou, J. S, H. S, **PRC 112 (2025), L051902;** **PRC 112 (2025), 054907**

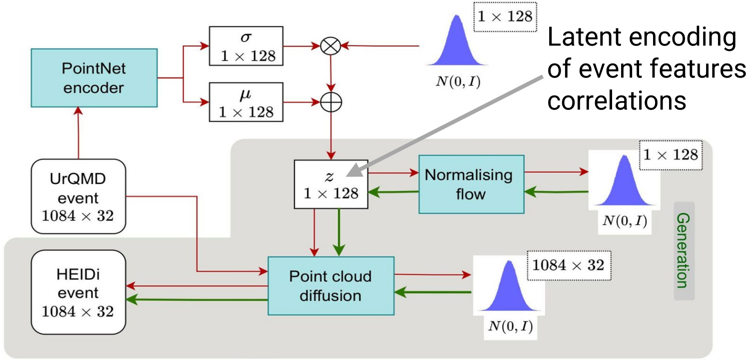
Point Cloud Diffusion Model for HICs – AI clone of simulation

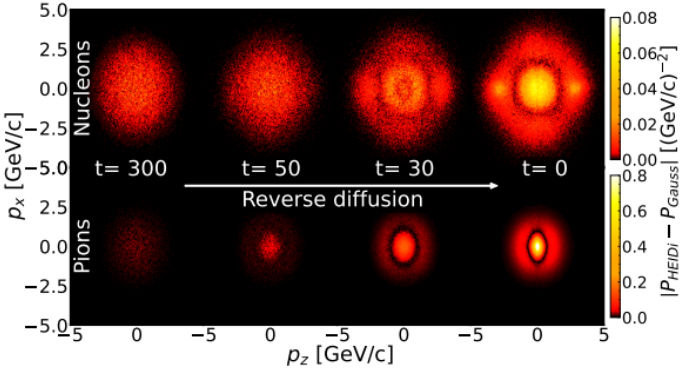


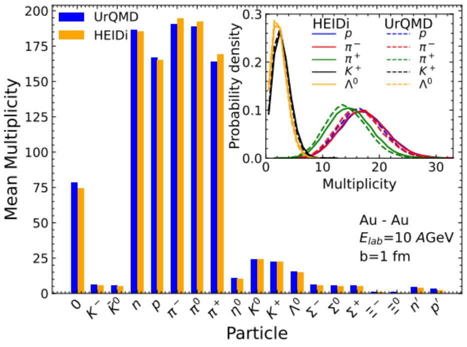
His decisions aren't any better than yours — but they're WAY faster...

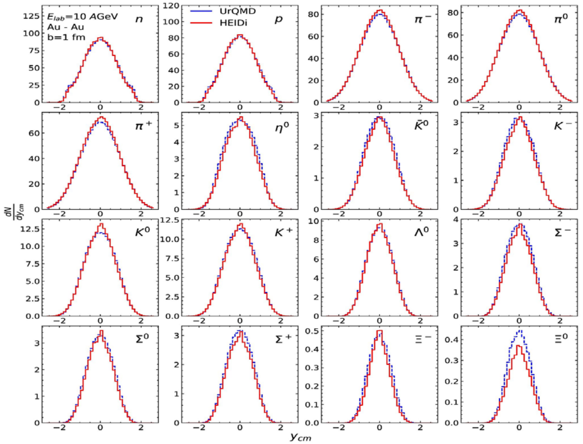
- 18k UrQMD simulation events for central Au-Au@10 AGeV collisions
- HEIDi:
  Heavy-ion Events through Intelligent Diffusion

PointNet encoder + Normalizing flow decoder + Pointcloud diffusion →

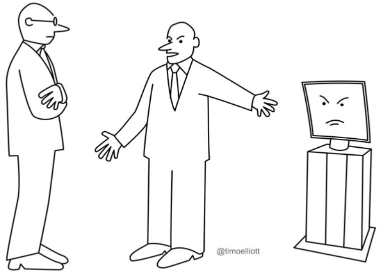








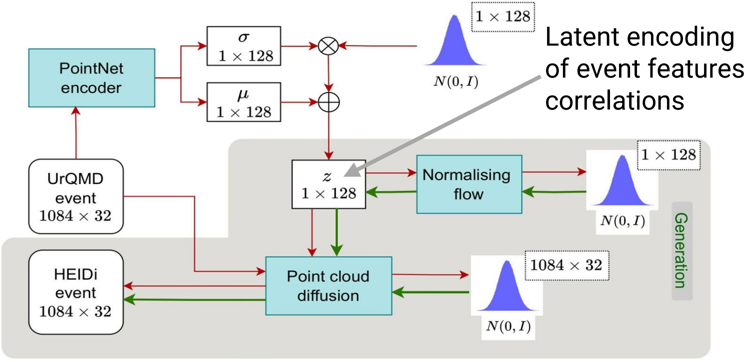
Point Cloud Diffusion Model for HICs – AI clone of simulation

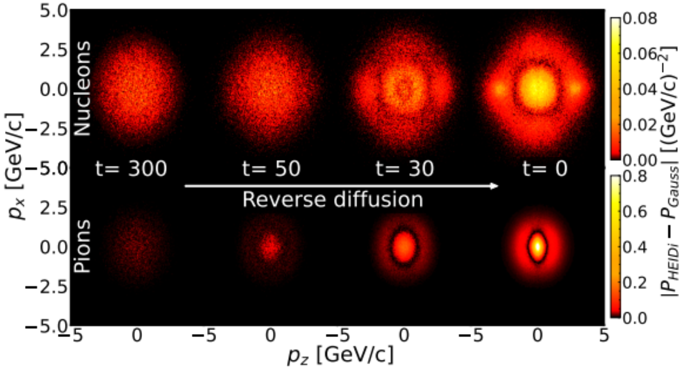


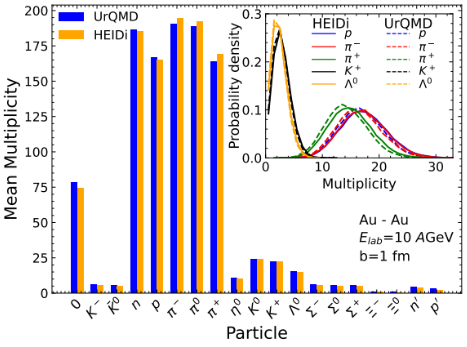
His decisions aren't any better than yours — but they're WAY faster...

- 18k UrQMD simulation events for central Au-Au@10 AGeV collisions
- HEIDi:
  Heavy-ion Events through Intelligent Diffusion

PointNet encoder + Normalizing flow decoder + Pointcloud diffusion →





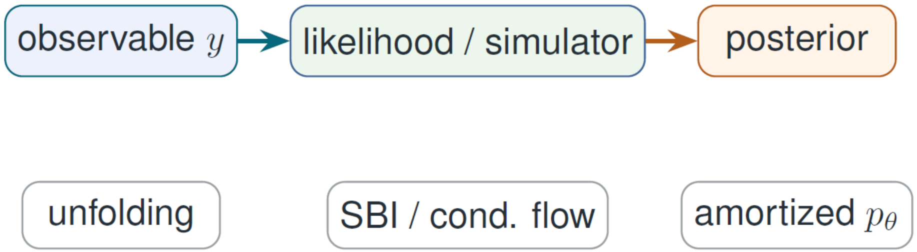


- **Running time of** UrQMD simulation
  cascade : ~ 3 sec/event;
  with potential : ~ 3 min/event;
  hybrid : ~ 1 hour/event
- HEIDI on A100: ~ 30 ms/event
- Speedup **2 ~ 5 orders** of magnitude

M. O.K, K. Zhou, J. S, H. S, **PRC 112 (2025), L051902; PRC 112 (2025), 054907**

Inverse Inference – GenAI as a posterior engine

$$p(\theta \mid y) \propto p(y \mid \theta)\, p(\theta)$$



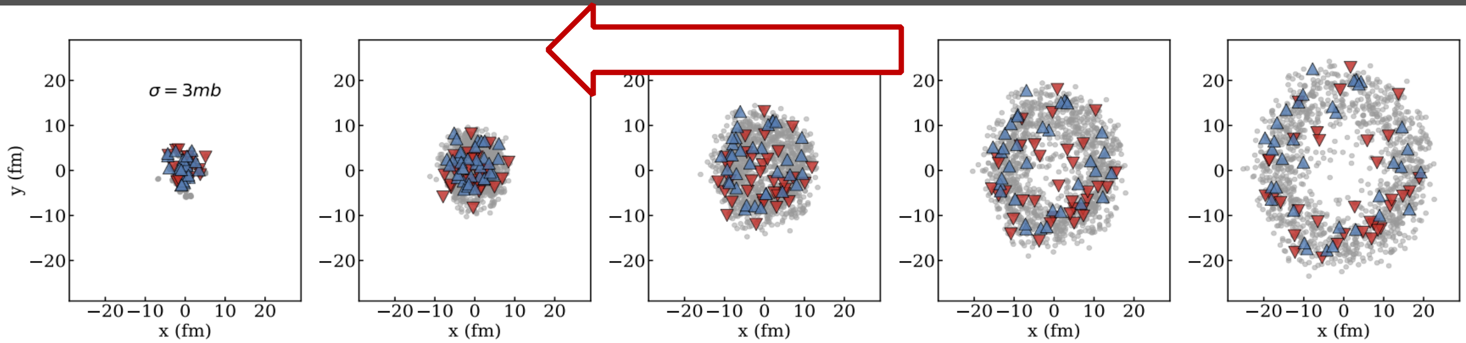
- The physics signal may be hidden, smeared, or only indirectly observed.
- GenAI enters as learned unfolding, conditional density estimation, or simulation-based inference.
- The desired output is a **distribution over explanations**, not only a point estimate.

Physics interpretation

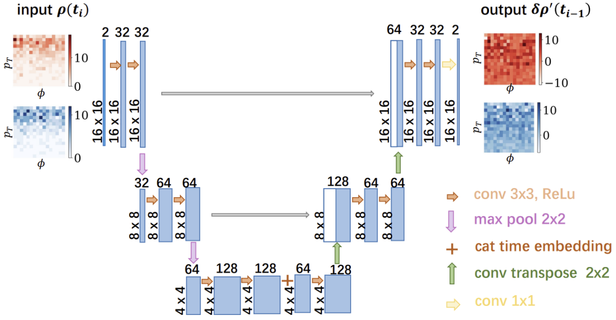
The model should not merely predict a hidden variable; it should quantify which physical explanations remain plausible.

Inverse problems turn GenAI from a generator into an uncertainty-aware inference tool.

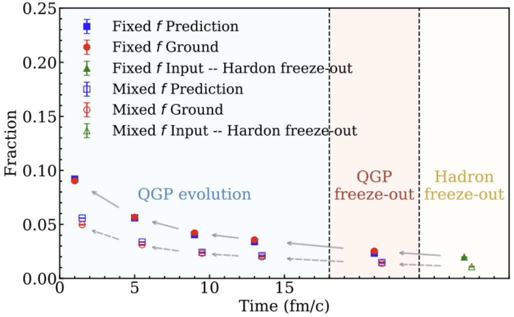
Unfold CME in HICs with time-embedded UNet



$$f = \frac{N^{\pm}_{\uparrow(\downarrow)} - N^{\pm}_{\downarrow(\uparrow)}}{N^{\pm}_{\uparrow(\downarrow)} + N^{\pm}_{\downarrow(\uparrow)}}$$



$$\delta\rho(t_{i-1}) = \rho(t_{i-1}) - \rho(t_i)$$
$$\rho'(t_{i-1}) = \delta\rho'(t_{i-1}) + \rho(t_i)$$



- Reconstruct CME-related charge separation from late-stage momentum-space information.
- The signal is blurred by partonic interactions, hadronization, and hadronic rescattering.
- CME unfolding is therefore a **dynamical inverse problem**: approximate the time-reversed evolution step by step.

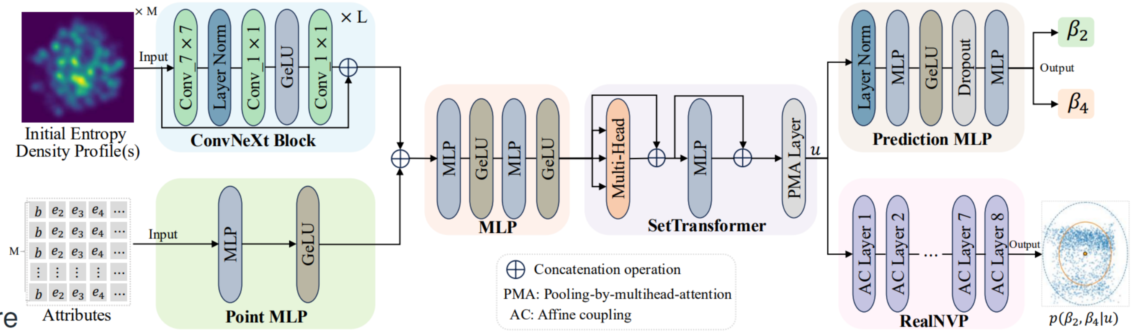
Interpretation

The model is not predicting one summary statistic. It is approximating a time-reversed transport map for the CME signal itself.

S.Guo, L. Wang, K. Zhou, G. Ma, **Chin. Phys. Lett. 42, 110101** (2025)

Inference on nuclear structure in HICs

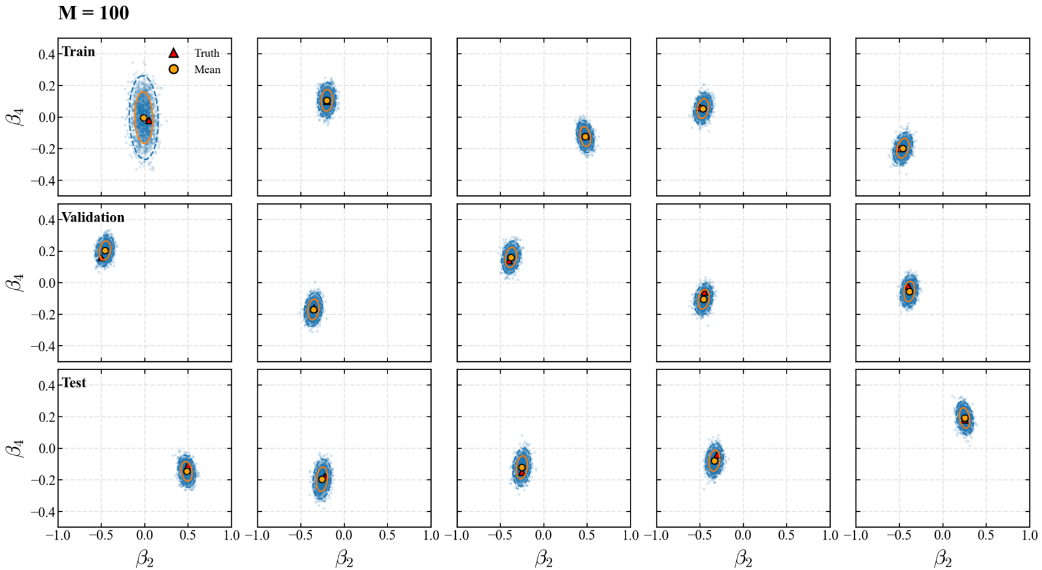
- initial entropy-density or nucleon-configuration features are encoded from event bags;
- conditional flow/SBI returns a calibrated posterior over deformation parameters;
- the GenAI component represents non-Gaussian uncertainty, not only regression.



- Single events are too noisy: participant fluctuations obscure deformation information.
- Grouping many events with the same deformation parameters strongly improves identifiability.
- Conditional normalizing flows provide

$$p_\theta(\beta_2, \beta_4 \mid \text{event bag})$$

  instead of a single estimate.
- GenAI acts as a many-event likelihood-to-posterior map.



J. Tao, Y. Liu, Y. Sha, X. Fan, Y. Tu, K. Zhou, H. Zheng, B. Zhang, **arXiv: 2603.24088**

Summary

Lattice QFT

sampler and PDE testbed

$$\pi[\phi] \propto e^{-S[\phi]}$$

transport must preserve action, symmetries, cumulants, topology, and path-space diagnostics

HIC simulation

conditional surrogate

$$p(\text{event} \mid c)$$

fast generators must preserve event structure relevant for Bayesian calibration and observables

Inverse problems

posterior engine

$$p(\theta \mid y)$$

models should infer hidden variables, uncertainties, and earlier dynamical stages

probability paths + action structure + physical validation

Unifying statement

GenAI is useful here when it becomes a controlled transport tool: a sampler, a surrogate, or a posterior map whose errors are judged by physics.

Thanks！

AI FOR SCIENCE

☛ *Initial Target JIF 10*
☛ *Long Term Target JIF 20*

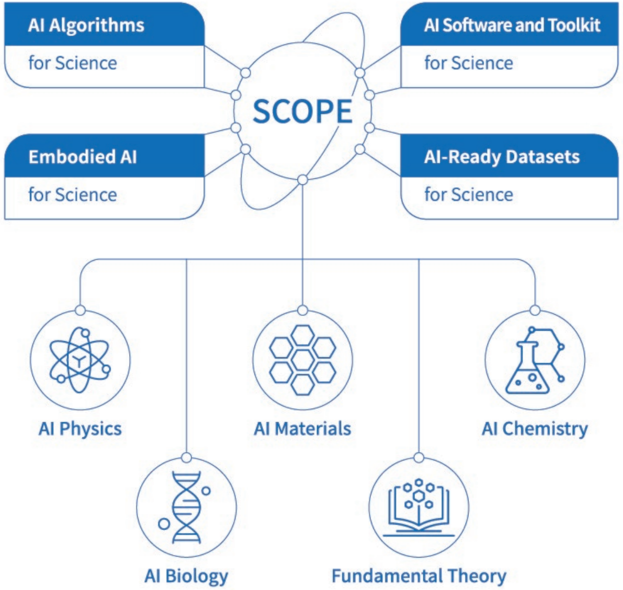
IOP Publishing

中国科学院东莞材料科学与技术研究所
Dongguan Institute of Materials Science and Technology, CAS

Editors-in-Chief

Weihua Wang
Dongguan Institute of Materials Science and Technology, CAS, China

Gian-Marco Rignanese
Ecole Polytechnique de Louvain , Belgium



Unique Features

- Inclusive Scope
- Free and Open Access
- International Editorial Board
- 24hrs Online
- Ease of Formatting
- Data Availability

Submit

mc04.manuscriptcentral.com/ai4s-slab

E-mail

editorialoffice@aiforsci.net

Homepage

iopscience.iop.org/journal/3050-287X
aiforsci.net

Wechat Homepage