Score-based diffusion models for lattice gauge theory and prospects for TMD/FF reconstruction

Bao-Dong Sun

Based on: Herzallah Alharazin, Julia Yu. Panteleeva, BDS: arXiv:2602.09045

QCD Evolution 2026, 11-15 May 2026

Outline：

1. **HMC for gauge field configurations**
2. **Score-based diffusion models for U(1)/SU(2)/SU(N)**
3. **More applications: reconstruction of effective mass, gravitational form factors and prospects for TMD/FF**

RUHR
UNIVERSITÄT
BOCHUM

HMC for gauge field configurations

Lattice QCD calculate expectation values:

gauge field configurations

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}U\, \mathcal{O}[U]\, e^{-S[U]} \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}[U^{(i)}],$$

Hybrid-Monte-Carlo (HMC) Algorithms is usually used to generate the gauge fields

- It introduce fictitious momenta π conjugate to gauge fields.
- It evolves (U, π) via Hamiltonian dynamics: $H = \frac{1}{2}\pi^2 + S[U]$.
- It applies Accept/reject with Metropolis step.

Issues for HMC

1. It suffers from **Critical Slowing Down**: Autocorrelation time τ grows as $\tau \sim \xi^z \sim a^{-z}$, where ξ =correlation length, a= lattice spacing, $z \approx 2$(dynamical critical exponent).

Issues for HMC

1. It suffers from **Critical Slowing Down**: Autocorrelation time τ grows as $\tau \sim \xi^z \sim a^{-z}$, where ξ =correlation length, a= lattice spacing, $z \approx 2$(dynamical critical exponent).

2. Gauge configurations are classified by topological charge $Q \in \mathbb{Z}$:

   - In the continuum, different topological sectors are separated by infinite action barriers.
   - On the lattice, barriers are finite but grow as $a \to 0$.
   - HMC cannot tunnel between sectors efficiently.
   - That is, ensemble gets *frozen* in a single Q sector
   - **This is a serious systematic error**: physics depends on summing over all topological sectors!

Issues for HMC

1. It suffers from **Critical Slowing Down**: Autocorrelation time τ grows as $\tau \sim \xi^z \sim a^{-z}$, where ξ =correlation length, a= lattice spacing, $z \approx 2$(dynamical critical exponent).

2. Gauge configurations are classified by topological charge $Q \in \mathbb{Z}$:

- In the continuum, different topological sectors are separated by infinite action barriers.
- On the lattice, barriers are finite but grow as $a \to 0$.
- HMC cannot tunnel between sectors efficiently.
- That is, ensemble gets *frozen* in a single Q sector
- **This is a serious systematic error**: physics depends on summing over all topological sectors!

3. At finite baryon density, the Boltzmann weight becomes complex:

- $p(U) \propto e^{-S[U]}$ is no longer a probability distribution on the lattice.
- This is the so-called *sign problem* physics depends on summing over all topological sectors!

Lattice and Machine Learning

- A variety of approaches already exist in the literature, including open boundary conditions, parallel tempering, metadynamics, and normalizing flows.
- But none of them solve all three problems (sign problem, topological freezing, and critical slowing down) entirely.

Q. Zhu, G. Aarts, W. Wang, K. Zhou, and L. Wang
2410.19602, 2502.05504 → **Diffusion models**

Lattice and Machine Learning

- A variety of approaches already exist in the literature, including open boundary conditions, parallel tempering, metadynamics, and normalizing flows.
- But none of them solve all three problems (sign problem, topological freezing, and critical slowing down) entirely.

`Q. Zhu, G. Aarts, W. Wang, K. Zhou, and L. Wang`
`2410.19602, 2502.05504` → **Diffusion models**

- A diffusion model learns to reverse a gradual noising process.
- The training operates in two phases:
  - **Forward process**: Gradually add noise to data until it becomes pure noise.
  - **Reverse process**: Learn to undo the noise, step by step.
- The sampling process is performed as:

  Start from random noise → iteratively denoise → generate realistic sample.

Lattice and Machine Learning

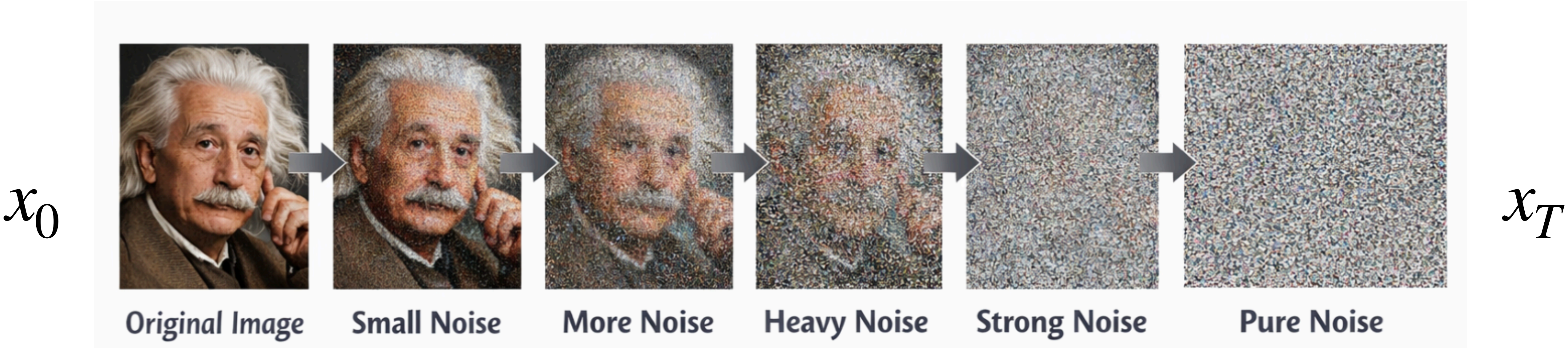
- A variety of approaches already exist in the literature, including open boundary conditions, parallel tempering, metadynamics, and normalizing flows.
- But none of them solve all three problems (sign problem, topological freezing, and critical slowing down) entirely.

`Q. Zhu, G. Aarts, W. Wang, K. Zhou, and L. Wang`
`2410.19602, 2502.05504` → **Diffusion models**

- A diffusion model learns to reverse a gradual noising process.
- The training operates in two phases:
  - **Forward process**: Gradually add noise to data until it becomes pure noise.
  - **Reverse process**: Learn to undo the noise, step by step.
- The sampling process is performed as:
  Start from random noise → iteratively denoise → generate realistic sample.
- The diffusion models were originally developed for Image generation.

Diffusion Models

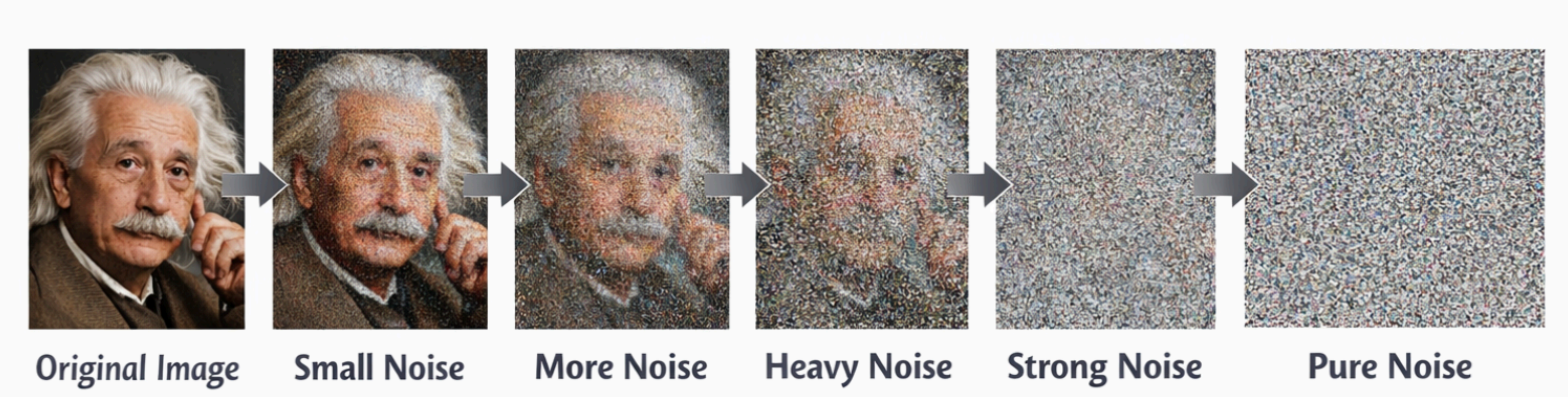
For example, one can train the Diffusion Model using pictures of Einstein:



Diffusion Models

For example, one can train the Diffusion Model using pictures of Einstein:

x_0

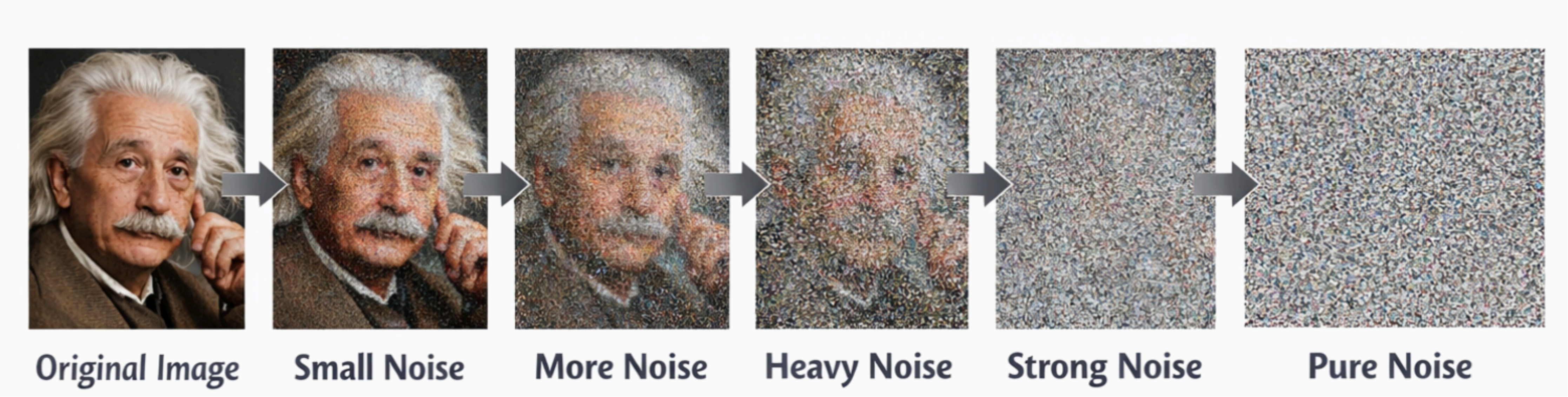


x_T

Diffusion Models

For example, one can train the Diffusion Model using pictures of Einstein:

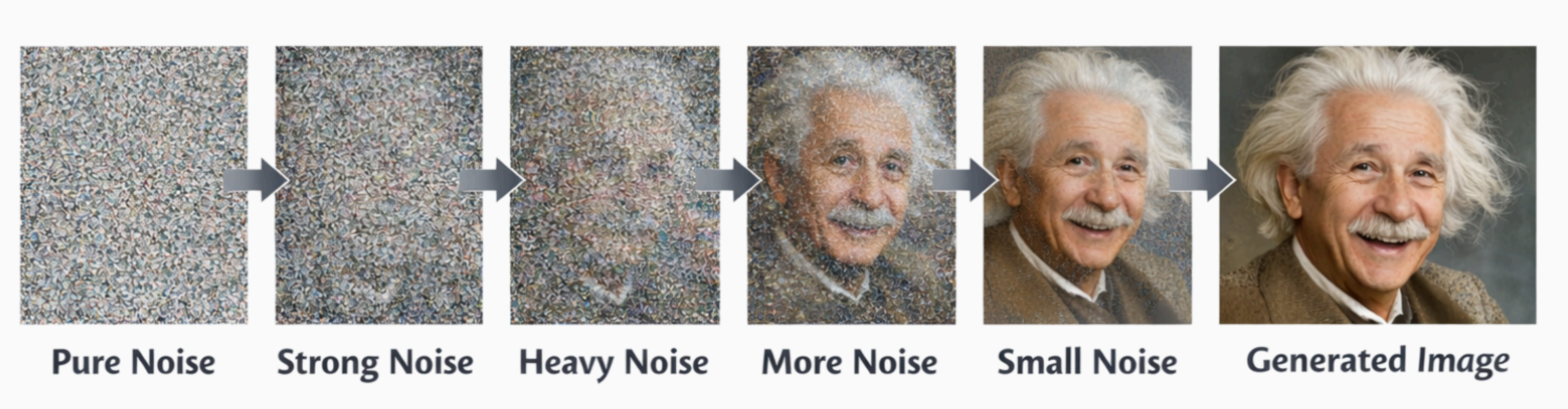
x_0



x_T

After training, we can generate completely new pictures of Einstein:

x_T



$\hat{x}_0$

Diffusion Model for Pure Gauge Theory

- Stochastic quantization, an alternative method for quantizing field theories where field evolve according to a Langevin equation:

$$\frac{\partial\phi(x,\tau)}{\partial\tau} = -\underbrace{\frac{\delta S[\phi]}{\delta\phi(x,\tau)}}_{\text{drift term}} + \underbrace{\sqrt{2}\,\overset{\boxed{\text{Gaussian white noise term}}}{\eta(x,\eta)}}_{\text{stochastic term}},$$

- This dynamics drives the system toward a stationary distribution proportional to $e^{-S[\phi]}$, exactly reproducing the Euclidean path integral measure.

Diffusion Model for Pure Gauge Theory

- Stochastic quantization, an alternative method for quantizing field theories where field evolve according to a Langevin equation:

Gaussian white noise term

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

drift term **stochastic term**

- This dynamics drives the system toward a stationary distribution proportional to $e^{-S[\phi]}$, exactly reproducing the Euclidean path integral measure.

- Diffusion models operate similarly to generates new samples from:

$$\frac{\partial\phi}{\partial t} = -\underbrace{g^2(t)\,\nabla_\phi \log p_t(\phi)}_{\text{drift term}} + g(t)\eta(t)$$

$g(t)$ controls the amount of noise added/removed at each time step;
$p_t(\phi)$ is the distribution at time t, i.e. $p_t(\phi) \propto e^{-S[\phi]}$

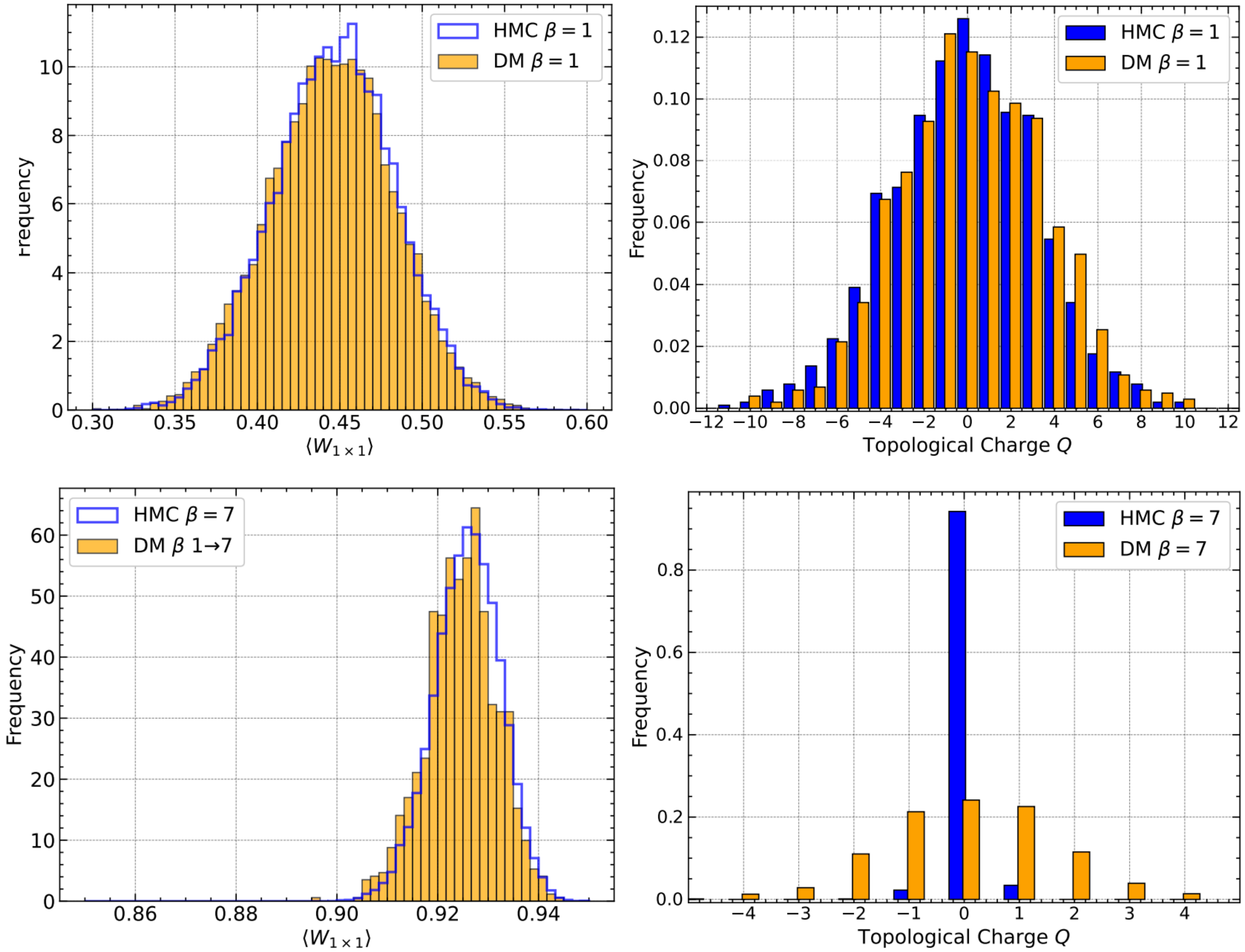
intractable to compute directly,
approximate it by deep neural network $s_{\hat{\theta}}$

$$\frac{\partial\phi(x,t)}{\partial t} = s_{\hat{\theta}}(\phi(x,t),t;\beta_0) + g(t)\bar{\eta}(x,t),$$

- Pure gauge theories $$\frac{\partial\phi(x,t)}{\partial t} = \frac{\beta}{\beta_0} s_{\hat{\theta}}(\phi(x,t),t;\beta_0) + g(t)\bar{\eta}(x,t).$$ ←action is separable: $S(\phi;\beta) = \beta\tilde{S}(\phi)$

Diffusion Model for $U(1)$ in two dimensional spacetime

`Q. Zhu, G. Aarts, W. Wang, K. Zhou, and L. Wang 2410.19602, 2502.05504`



- Local observables correctly reproduced → validates the physics-conditioned approach.

- At $\beta = 1$ both methods sample broad distribution($Q \approx -10$ to $+10$), no freezing.
- At $\beta = 7$ HMC freezes near $Q = 0$, while DM explores $Q = -4$ to 4.

- This is because higher β leads to larger energy barriers between topological sectors → Local updates cannot overcome these barriers.

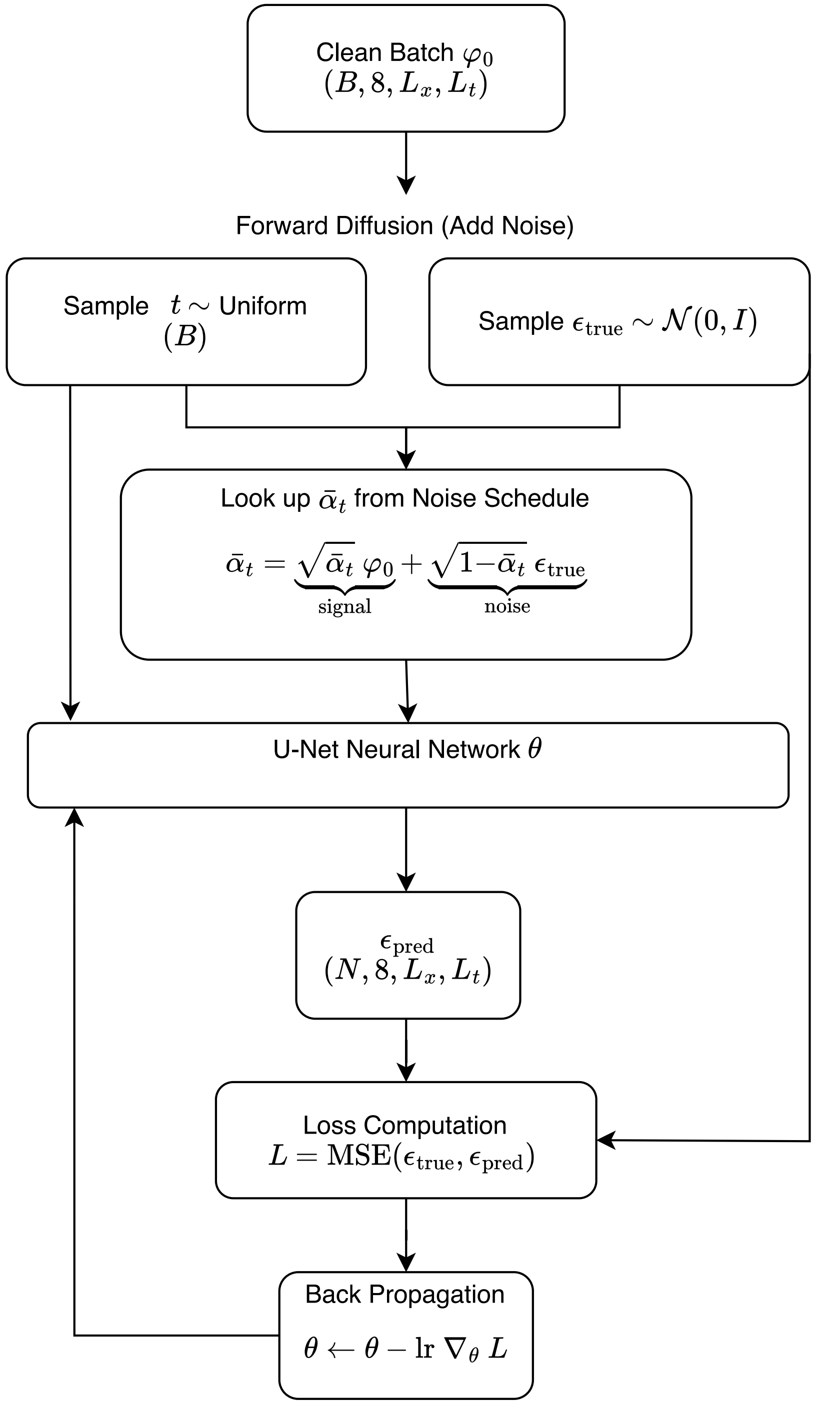
- **DM samples globally from noise**: → independent access to all sectors.

- General strategy: Train at easy β_0 (no freezing), generate at difficult β (would freeze with HMC).

Diffusion Models for $SU(2)$ in two dimensional spacetime

H. Alharazin, J. Yu. Panteleeva, B.-D. Sun: arXiv:2602.09045

Training procedure:



Clean Batch φ_0
$(B, 8, L_x, L_t)$

Forward Diffusion (Add Noise)

Sample $t \sim$ Uniform (B)

Sample $\epsilon_{\text{true}} \sim \mathcal{N}(0, I)$

Look up $\bar{\alpha}_t$ from Noise Schedule

$$\bar{\alpha}_t = \underbrace{\sqrt{\bar{\alpha}_t}\,\varphi_0}_{\text{signal}} + \underbrace{\sqrt{1-\bar{\alpha}_t}\,\epsilon_{\text{true}}}_{\text{noise}}$$

U-Net Neural Network θ

ϵ_{pred}
$(N, 8, L_x, L_t)$

Loss Computation
$L = \text{MSE}(\epsilon_{\text{true}}, \epsilon_{\text{pred}})$

Back Propagation
$\theta \leftarrow \theta - \text{lr}\,\nabla_\theta\, L$

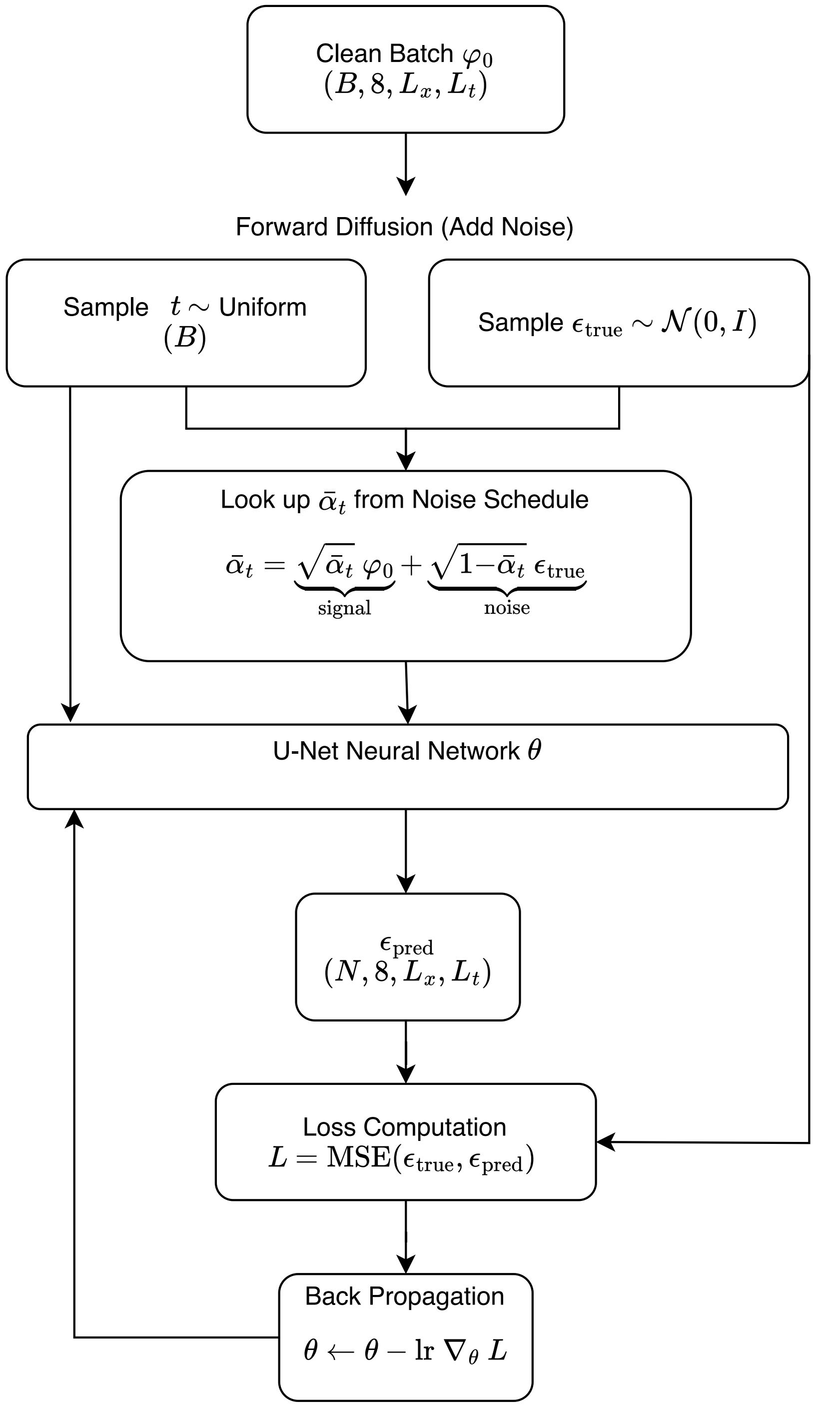
From U(1) to SU(2):

1. Configuration space: $S^1 \to S^3$.
2. Adopt quaternion parameterization,
   $$U = a_0 \mathbf{1} + i\left(a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3\right) \to (a_0, a_1, a_2, a_3)$$
3. Non-commutativity of SU(2) is not built-in, but learned from data
4. Never evaluates the action during sampling: easy to generate for complex action

Diffusion Models for $SU(2)$ in two dimensional spacetime

`H. Alharazin, J. Yu. Panteleeva, B.-D. Sun: arXiv:2602.09045`

Training procedure:



Clean Batch φ_0
$(B, 8, L_x, L_t)$

Forward Diffusion (Add Noise)

Sample $t \sim$ Uniform (B)

Sample $\epsilon_{\text{true}} \sim \mathcal{N}(0, I)$

Look up $\bar{\alpha}_t$ from Noise Schedule

$$\bar{\alpha}_t = \underbrace{\sqrt{\bar{\alpha}_t}\,\varphi_0}_{\text{signal}} + \underbrace{\sqrt{1-\bar{\alpha}_t}\,\epsilon_{\text{true}}}_{\text{noise}}$$

U-Net Neural Network θ

ϵ_{pred}
$(N, 8, L_x, L_t)$

Loss Computation
$L = \text{MSE}(\epsilon_{\text{true}}, \epsilon_{\text{pred}})$

Back Propagation
$\theta \leftarrow \theta - \text{lr}\, \nabla_\theta\, L$

From U(1) to SU(2):

1. Configuration space: $S^1 \rightarrow S^3$.
2. Adopt quaternion parameterization,
$$U = a_0 \mathbf{1} + i \left(a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 \right) \rightarrow (a_0, a_1, a_2, a_3)$$
3. Non-commutativity of SU(2) is not built-in, but learned from data
4. Never evaluates the action during sampling: easy to generate for complex action

Comparison with Aarts et al. arXiv:2601.19552:
（A parallel work for same SU（2） physics with DM）

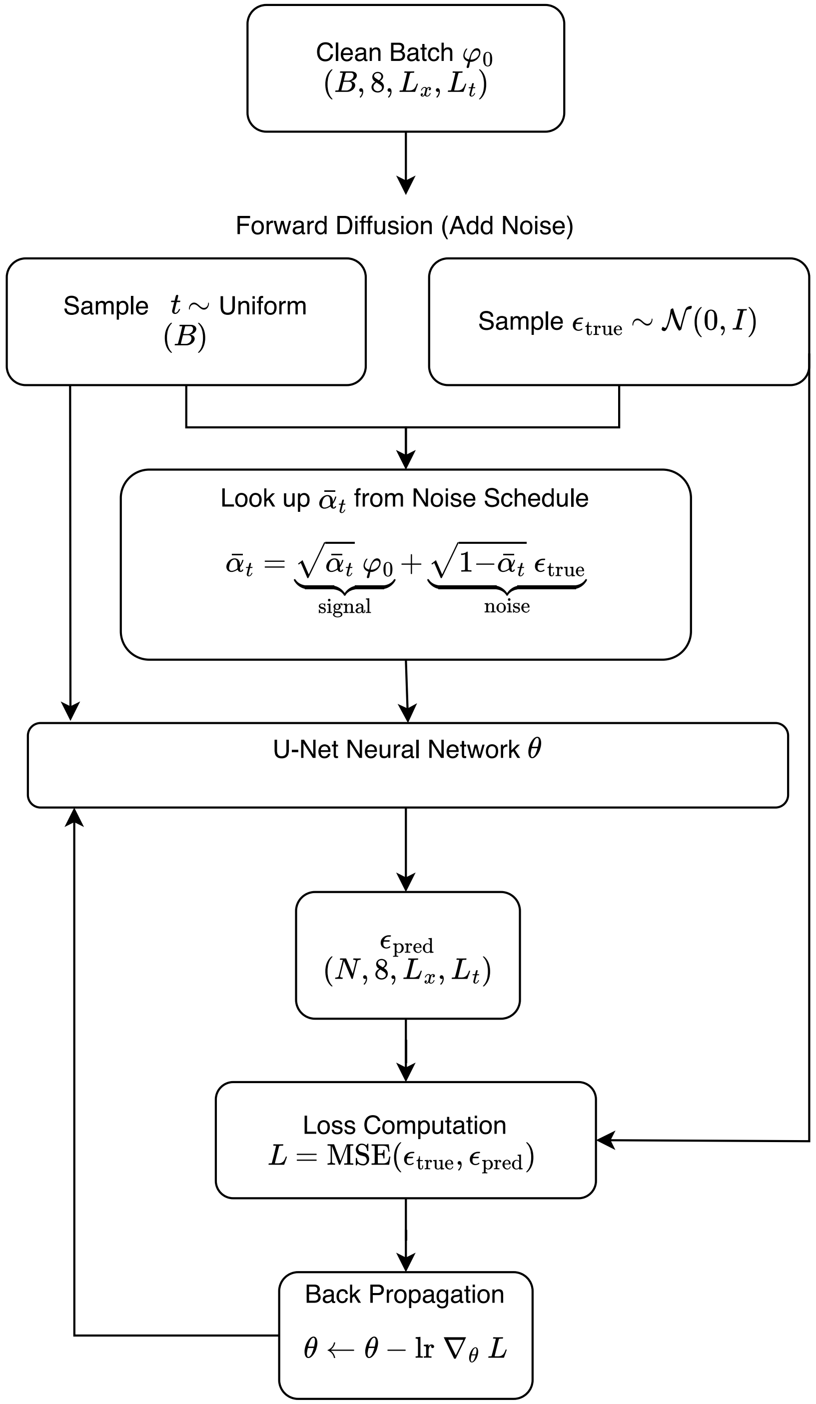
Three fundamental differences, they：

1. Enforcing exact local gauge equivariance
2. Target Wilson action needed to be calculated in sampling
3. Training setups：10 networks, ~ 15 000 parameters each → much more training data needed ours

Diffusion Models for $SU(2)$ in two dimensional spacetime

H. Alharazin, J. Yu. Panteleeva, B.-D. Sun: arXiv:2602.09045

Training procedure:



Clean Batch φ_0
$(B, 8, L_x, L_t)$

Forward Diffusion (Add Noise)

Sample $t \sim$ Uniform
(B)

Sample $\epsilon_{\text{true}} \sim \mathcal{N}(0, I)$

Look up $\bar{\alpha}_t$ from Noise Schedule

$$\bar{\alpha}_t = \underbrace{\sqrt{\bar{\alpha}_t}\,\varphi_0}_{\text{signal}} + \underbrace{\sqrt{1-\bar{\alpha}_t}\,\epsilon_{\text{true}}}_{\text{noise}}$$

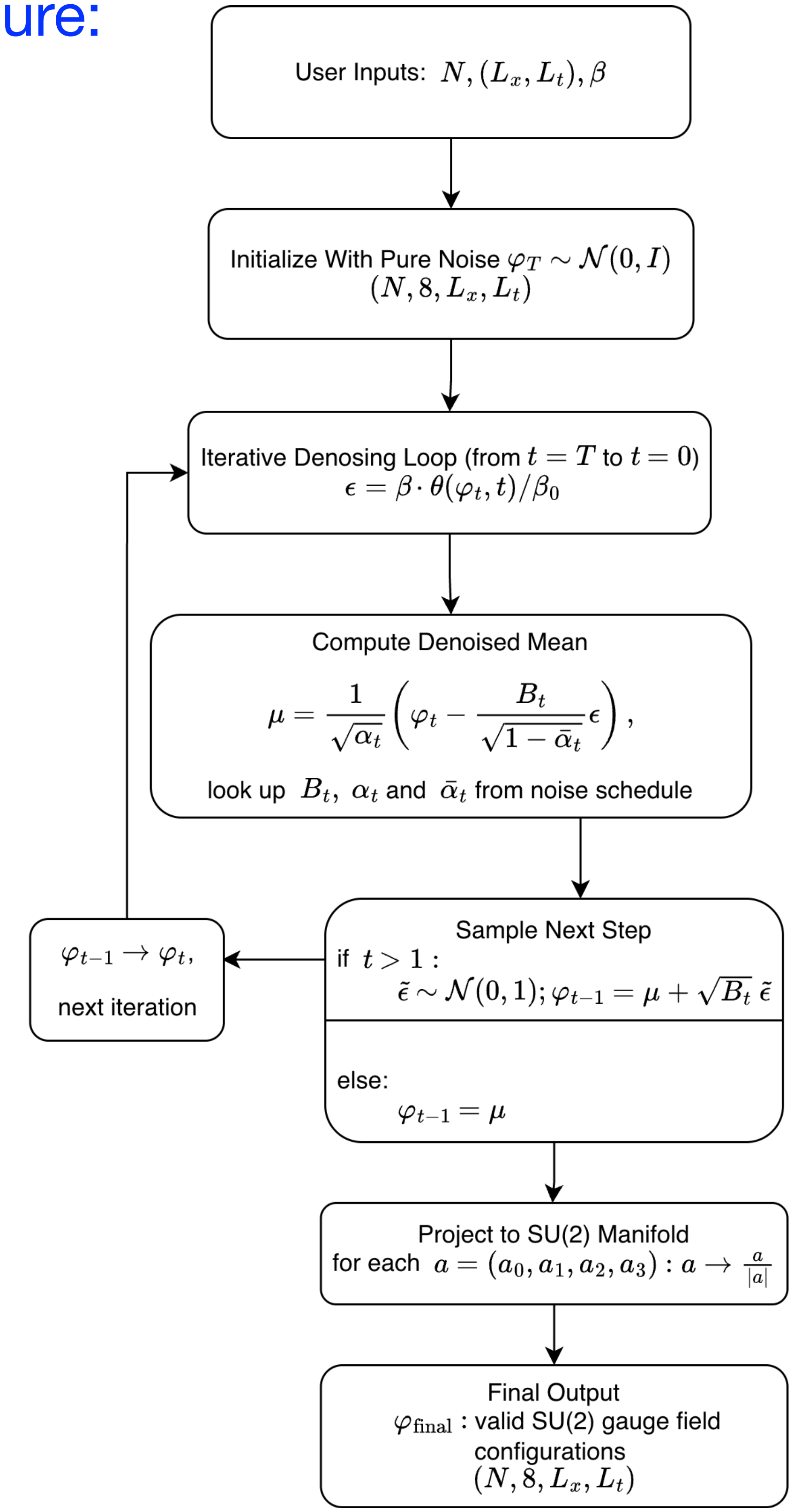
U-Net Neural Network θ

ϵ_{pred}
$(N, 8, L_x, L_t)$

Loss Computation
$L = \text{MSE}(\epsilon_{\text{true}}, \epsilon_{\text{pred}})$

Back Propagation
$\theta \leftarrow \theta - \text{lr}\,\nabla_\theta L$

Sampling procedure:



User Inputs: $N, (L_x, L_t), \beta$

Initialize With Pure Noise $\varphi_T \sim \mathcal{N}(0, I)$
$(N, 8, L_x, L_t)$

Iterative Denosing Loop (from $t = T$ to $t = 0$)
$\epsilon = \beta \cdot \theta(\varphi_t, t)/\beta_0$

Compute Denoised Mean

$$\mu = \frac{1}{\sqrt{\alpha_t}}\left(\varphi_t - \frac{B_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right),$$

look up B_t, α_t and $\bar{\alpha}_t$ from noise schedule

Sample Next Step
if $t > 1$:
$\tilde{\epsilon} \sim \mathcal{N}(0, 1); \varphi_{t-1} = \mu + \sqrt{B_t}\,\tilde{\epsilon}$

else:
$\varphi_{t-1} = \mu$

$\varphi_{t-1} \to \varphi_t$,
next iteration

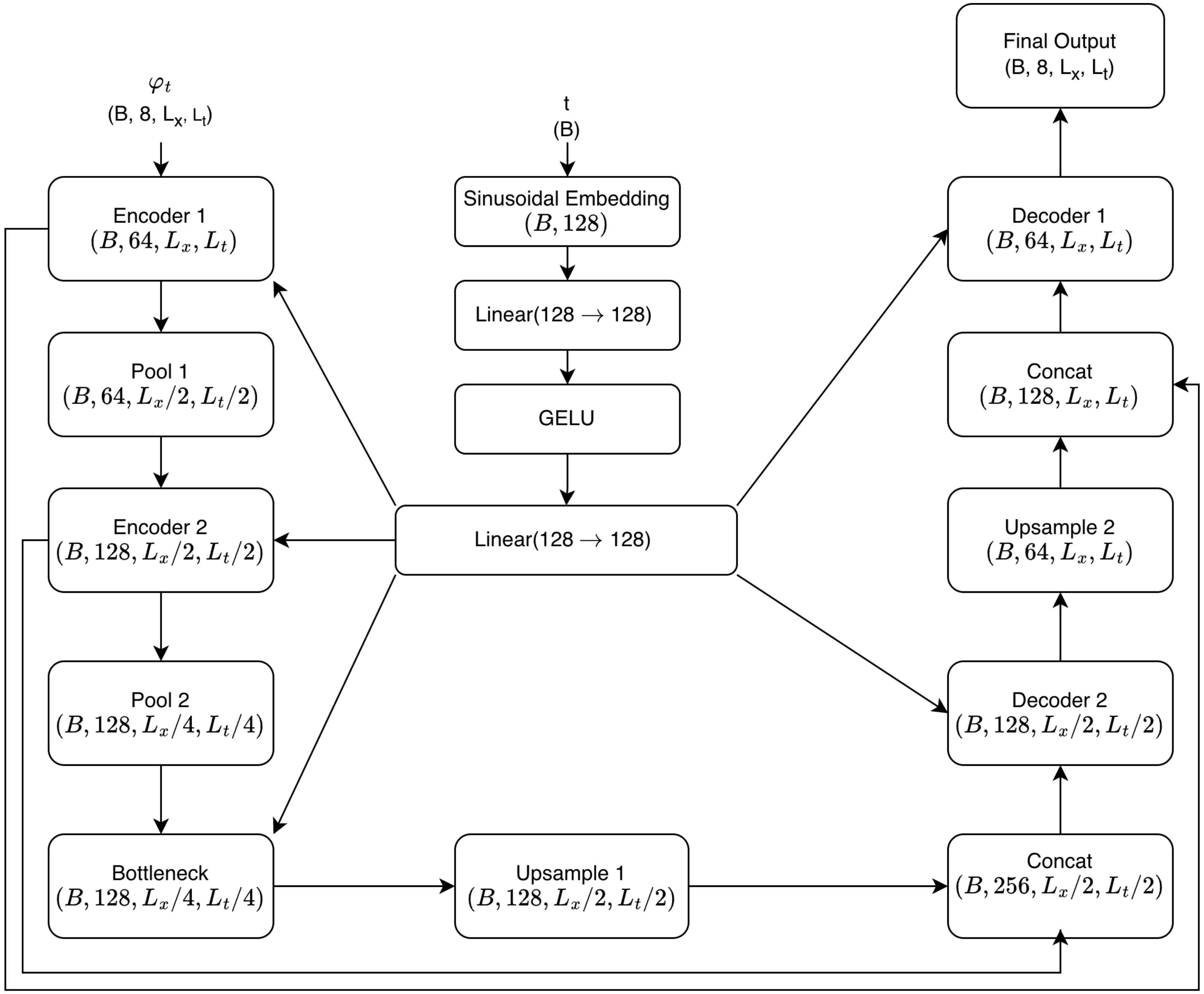
Project to SU(2) Manifold
for each $a = (a_0, a_1, a_2, a_3) : a \to \frac{a}{|a|}$

Final Output
φ_{final} : valid SU(2) gauge field configurations
$(N, 8, L_x, L_t)$

Diffusion Models for SU(2) Lattice Gauge Theory in Two Dimensions

The network $\epsilon_\theta(\phi_t^i, t^i)$ contains various building blocks. The most important ones are:

1. Encoder Blocks.
2. Decoder Blocks.
3. Pooling Blocks.
4. Upsampling Blocks.
5. Sinusodial Embedding.
6. Concat Blocks.



φ_t
(B, 8, L_x, L_t)

Encoder 1
$(B, 64, L_x, L_t)$

Pool 1
$(B, 64, L_x/2, L_t/2)$

Encoder 2
$(B, 128, L_x/2, L_t/2)$

Pool 2
$(B, 128, L_x/4, L_t/4)$

Bottleneck
$(B, 128, L_x/4, L_t/4)$

t
(B)

Sinusoidal Embedding
$(B, 128)$

Linear(128 → 128)

GELU

Linear(128 → 128)

Upsample 1
$(B, 128, L_x/2, L_t/2)$

Concat
$(B, 256, L_x/2, L_t/2)$

Decoder 2
$(B, 128, L_x/2, L_t/2)$

Upsample 2
$(B, 64, L_x, L_t)$

Concat
$(B, 128, L_x, L_t)$

Decoder 1
$(B, 64, L_x, L_t)$

Final Output
(B, 8, L_x, L_t)

Plaquette in 2D SU（2）：（gauge-invariant, elementary closed loop around a single lattice cell）

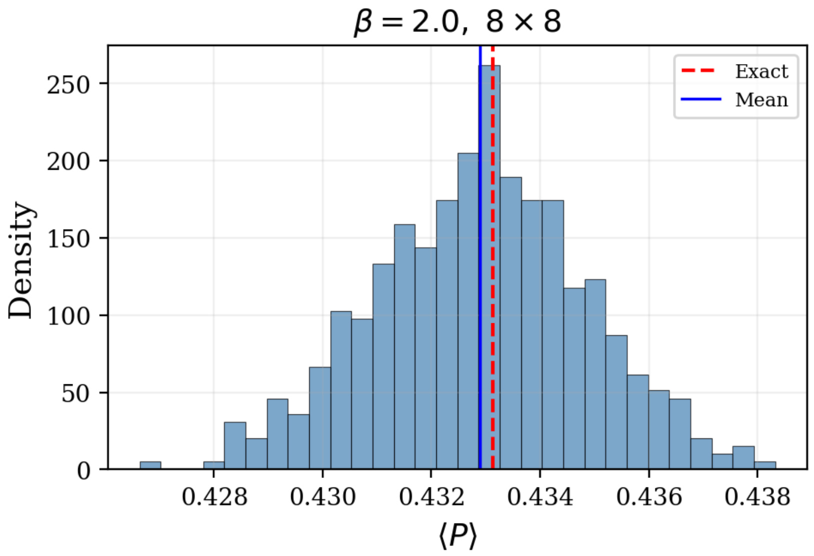
$$P_{01}(x) = U_0(x)U_1(x+\hat{0})U_0^{\dagger}(x+\hat{1})U_1^{\dagger}(x)$$

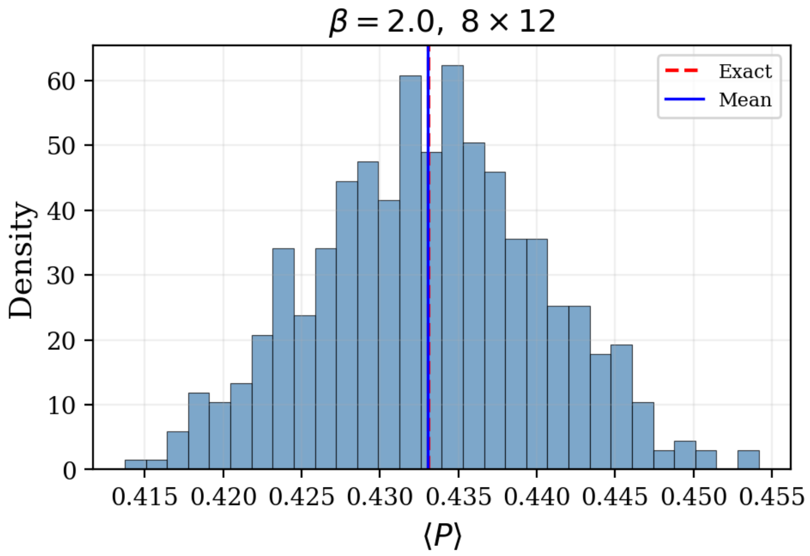
Average plaquette：（fundamental observable）

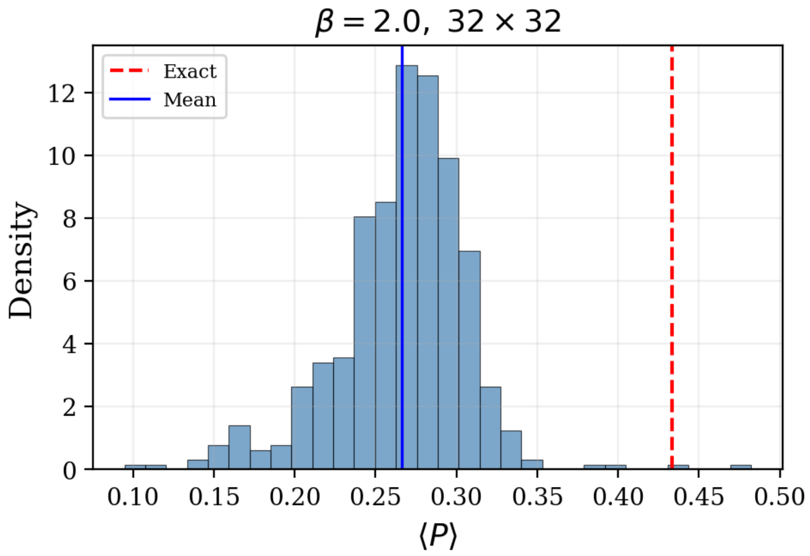
$$\langle P \rangle = \left\langle \frac{1}{2}\,\mathrm{Re}\,\mathrm{Tr}\left[P_{01}\right] \right\rangle = < a_0^{(\text{plaq})} > \xrightarrow{\text{SU（2）}} \langle P \rangle_{\text{exact}} = \frac{I_2(\beta)}{I_1(\beta)}\Big|_{\beta=2} = 0.4331$$

$$\langle P \rangle_{\text{diff}}^{(\beta=2)} = 0.4329 \pm 0.0001(\text{stat}) \qquad \Delta \equiv \langle P \rangle_{\text{diff}} - \langle P \rangle_{\text{exact}} = -0.0002$$

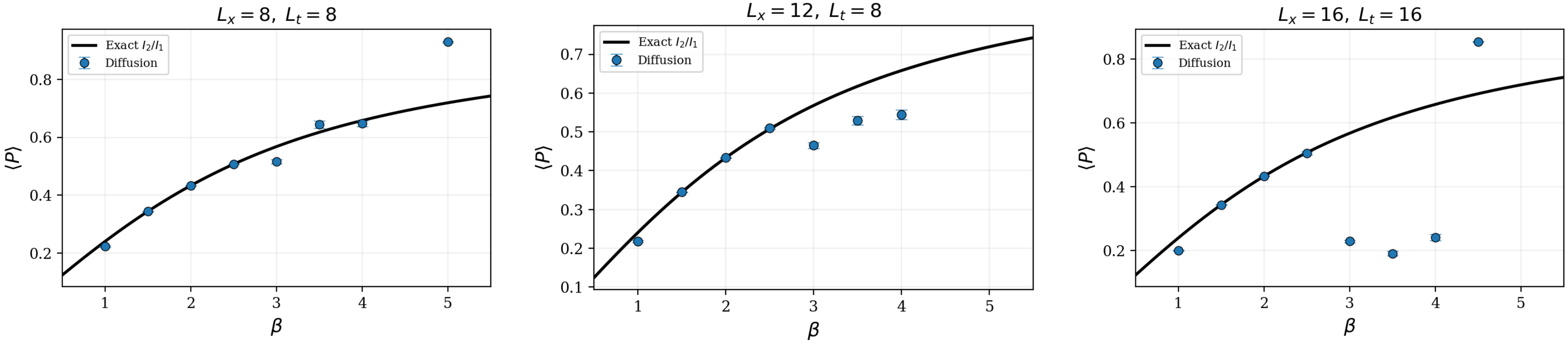
Plaquette distributions：



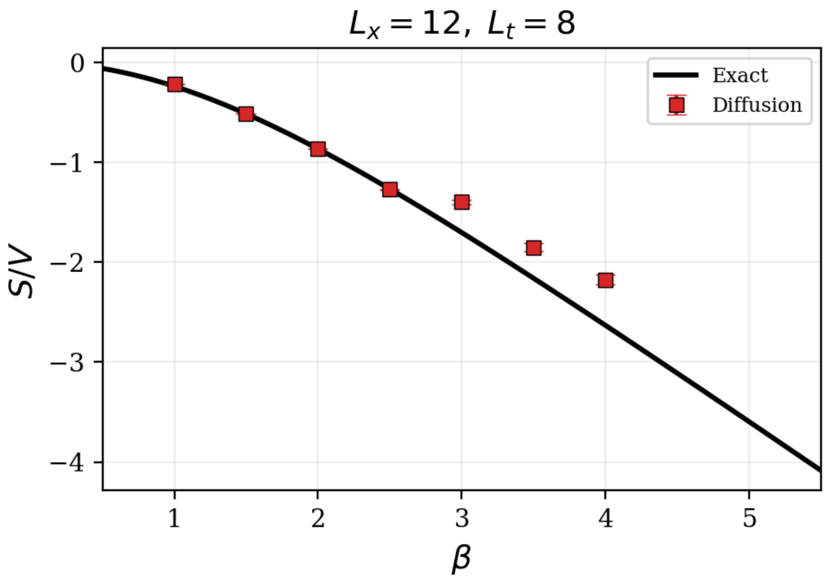




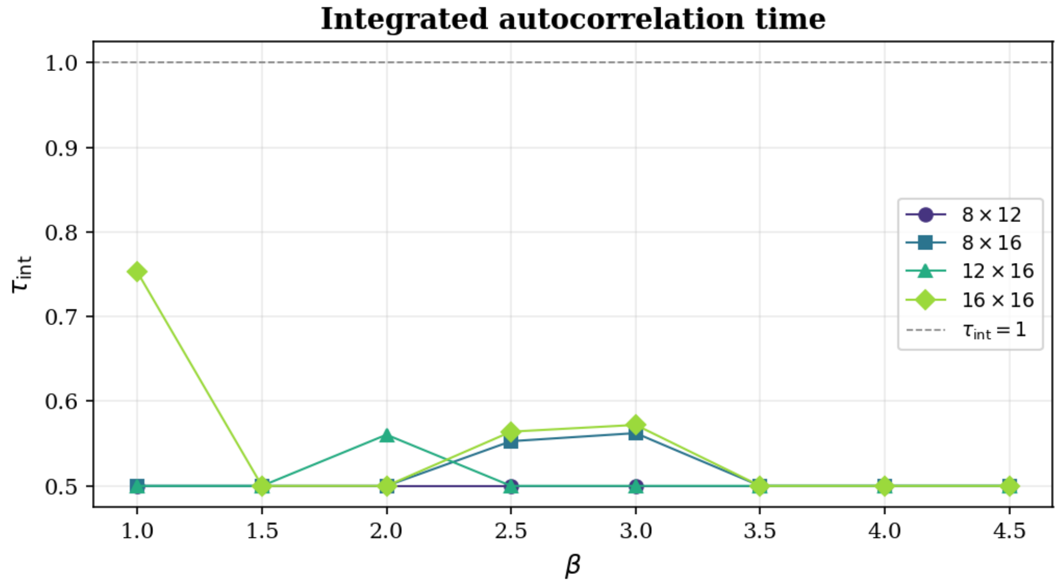
Physics-Conditioned Generation at Different Couplings



Action density



Integrated autocorrelation time



No critical slowing down within tested range!

Diffusion model for SU(N) gauge theories

`Javad Komijani, Marina K. Marinkovic, Lara Turgut: arXiv:2605.06134`

1. Score-matching framework directly on **general SU(N)**, extendable to other Lie groups
2. **Implicit score matching** — avoids the explicit divergence in standard score matching, computationally efficient
3. Tested on **SU(3)** with the Wilson gauge action in **both 2D and 4D**
4. For large β: introduce a **predictor–corrector** scheme with a **Hamiltonian-MD-based corrector** for accurate reverse-time integration
5. Trade-off: corrector visibly improves sampling quality, at higher computational cost

Fermions are not included yet!
→ Not enough for real calculation in lattice.

Diffusion model for effective mass reconstruction

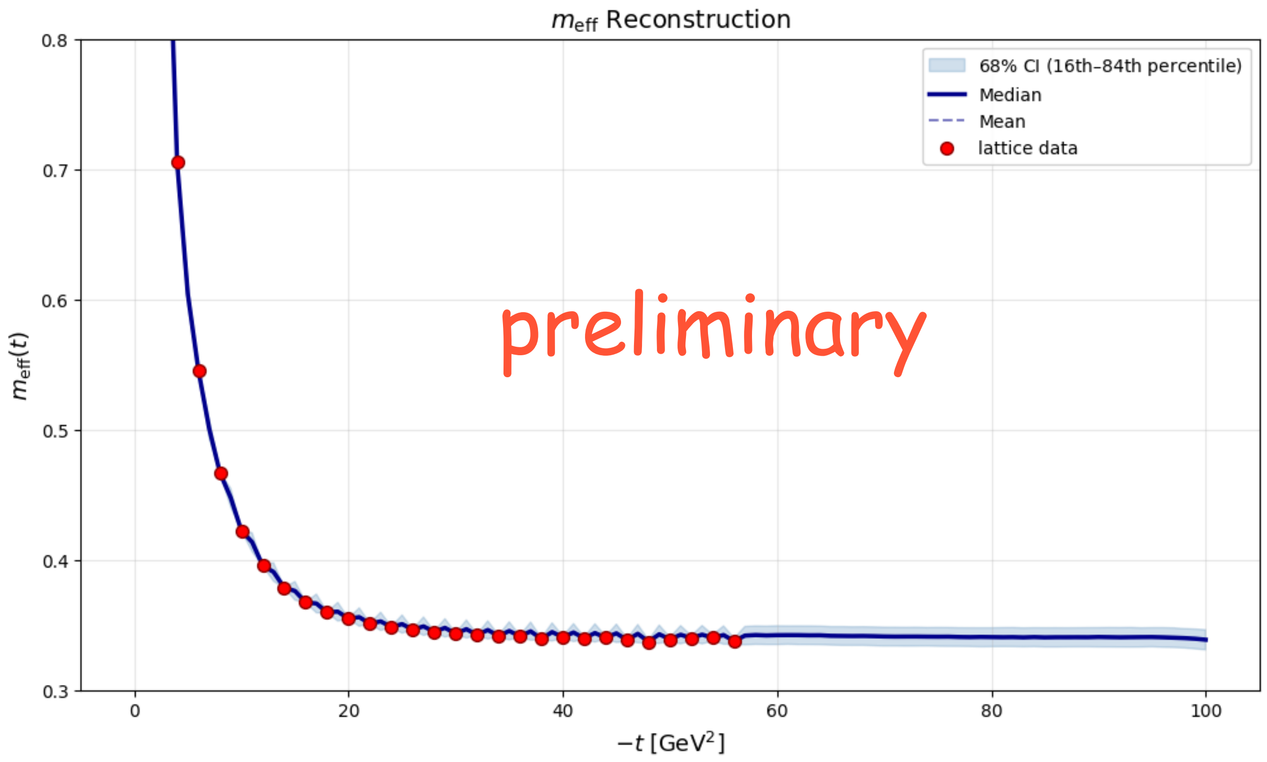
`H. Alharazin, J. M. Bulava, D. Laudicina, B-.D Sun, in preparation.`

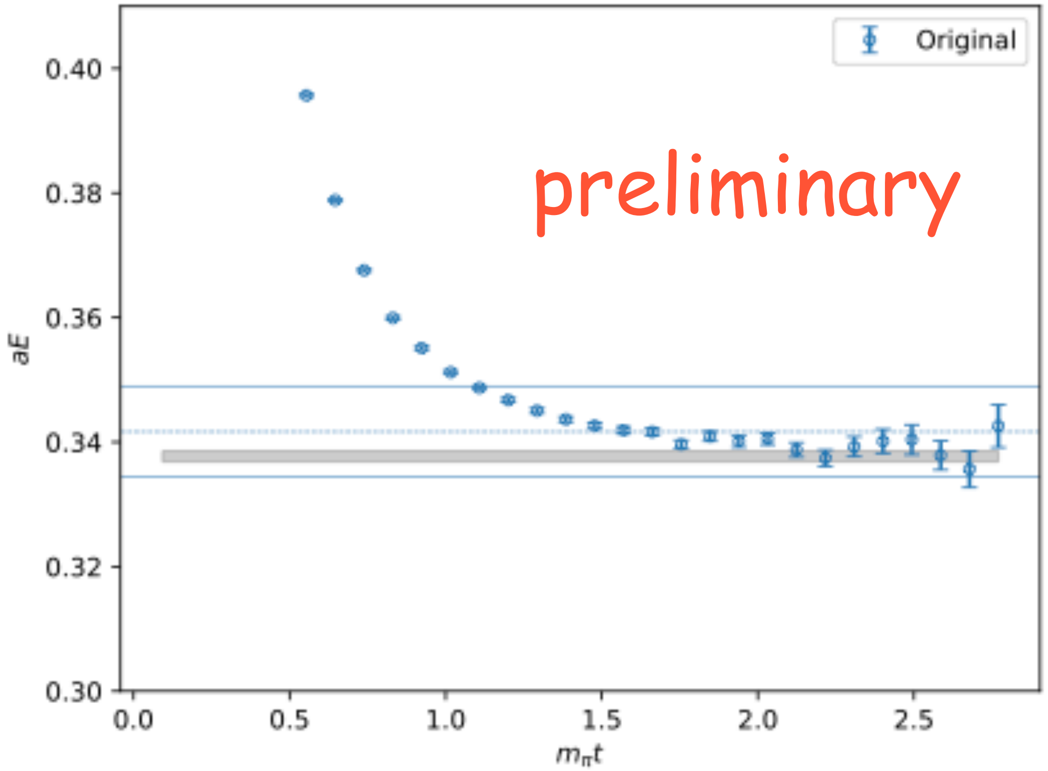
Euclidean correlator takes the spectral decomposition:

$$C(t) = \sum_{n=0}^{\infty} A_n e^{-E_n t}$$

Ground-state energy E_0 (effective mass):

$$E_0 = \left[-\frac{d}{dt} \ln C(t) \right]_{t-}$$



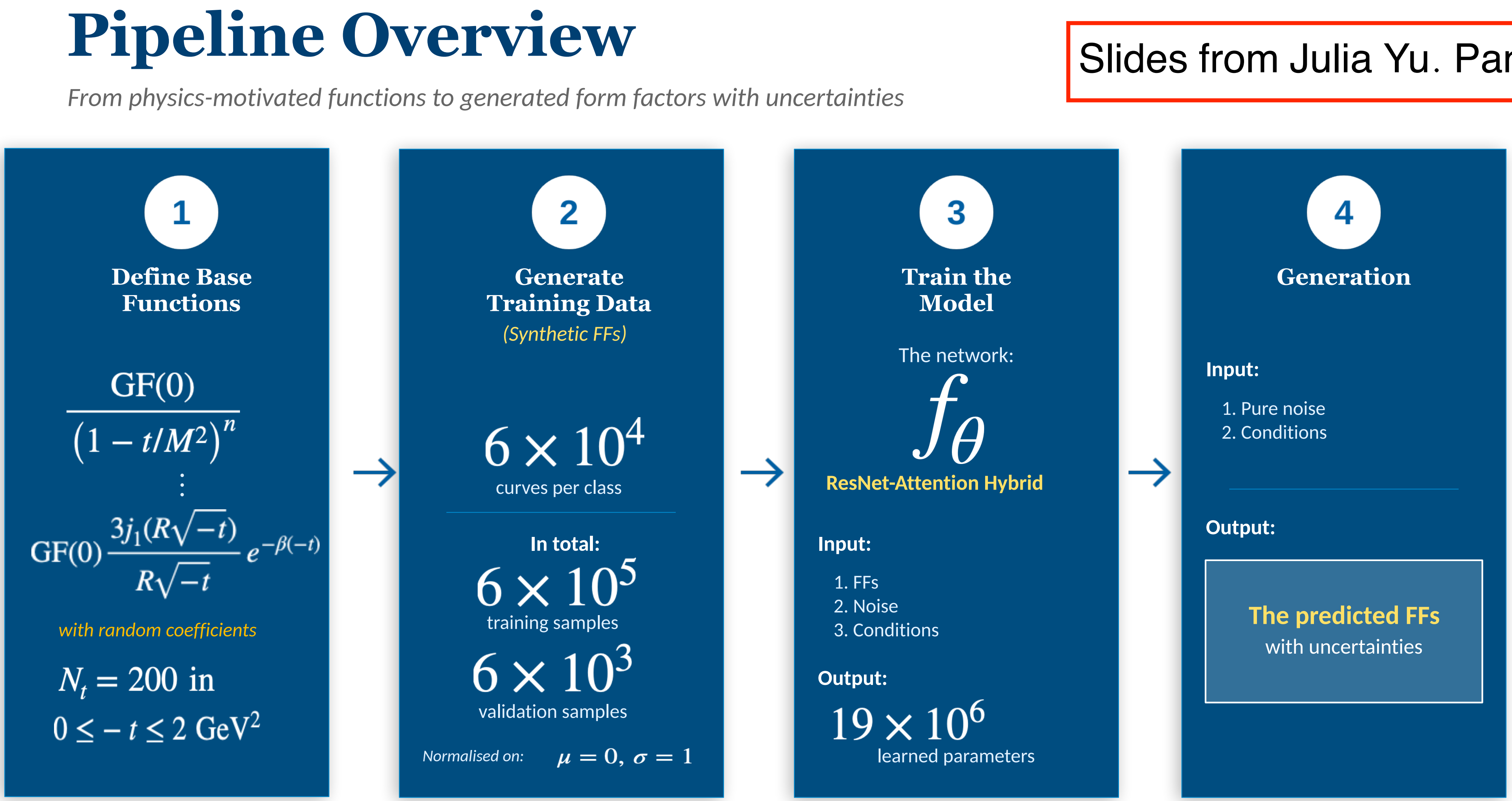


Error is still too large!

More applications: Diffusion Model for gravitational form factors

H. Alharazin, J. Yu. Panteleeva: arXiv:2602.19267, accepted by PRD

Slides from Julia Yu. Panteleeva



Pipeline Overview

From physics-motivated functions to generated form factors with uncertainties

1

Define Base Functions

$$\frac{\mathrm{GF}(0)}{\left(1 - t/M^2\right)^n}$$

$$\vdots$$

$$\mathrm{GF}(0)\,\frac{3 j_1(R\sqrt{-t})}{R\sqrt{-t}}\,e^{-\beta(-t)}$$

with random coefficients

$N_t = 200$ in

$0 \le -t \le 2\ \mathrm{GeV}^2$

→

2

Generate Training Data

(Synthetic FFs)

6×10^4

curves per class

In total:

6×10^5

training samples

6×10^3

validation samples

Normalised on: $\mu = 0,\ \sigma = 1$

→

3

Train the Model

The network:

f_θ

ResNet-Attention Hybrid

Input:

1. FFs
2. Noise
3. Conditions

Output:

19×10^6

learned parameters

→

4

Generation

Input:

1. Pure noise
2. Conditions

Output:

The predicted FFs

with uncertainties

ensuring dense coverage of the accessible shape space without overpopulating any single region

Differ with DM used in SU(2) work:

DM with velocity(v)-prediction: $v_\tau = \sqrt{\bar{\alpha}_\tau}\varepsilon - \sqrt{1-\bar{\alpha}_\tau}x_0$

(avoiding the signal-to-noise imbalance)

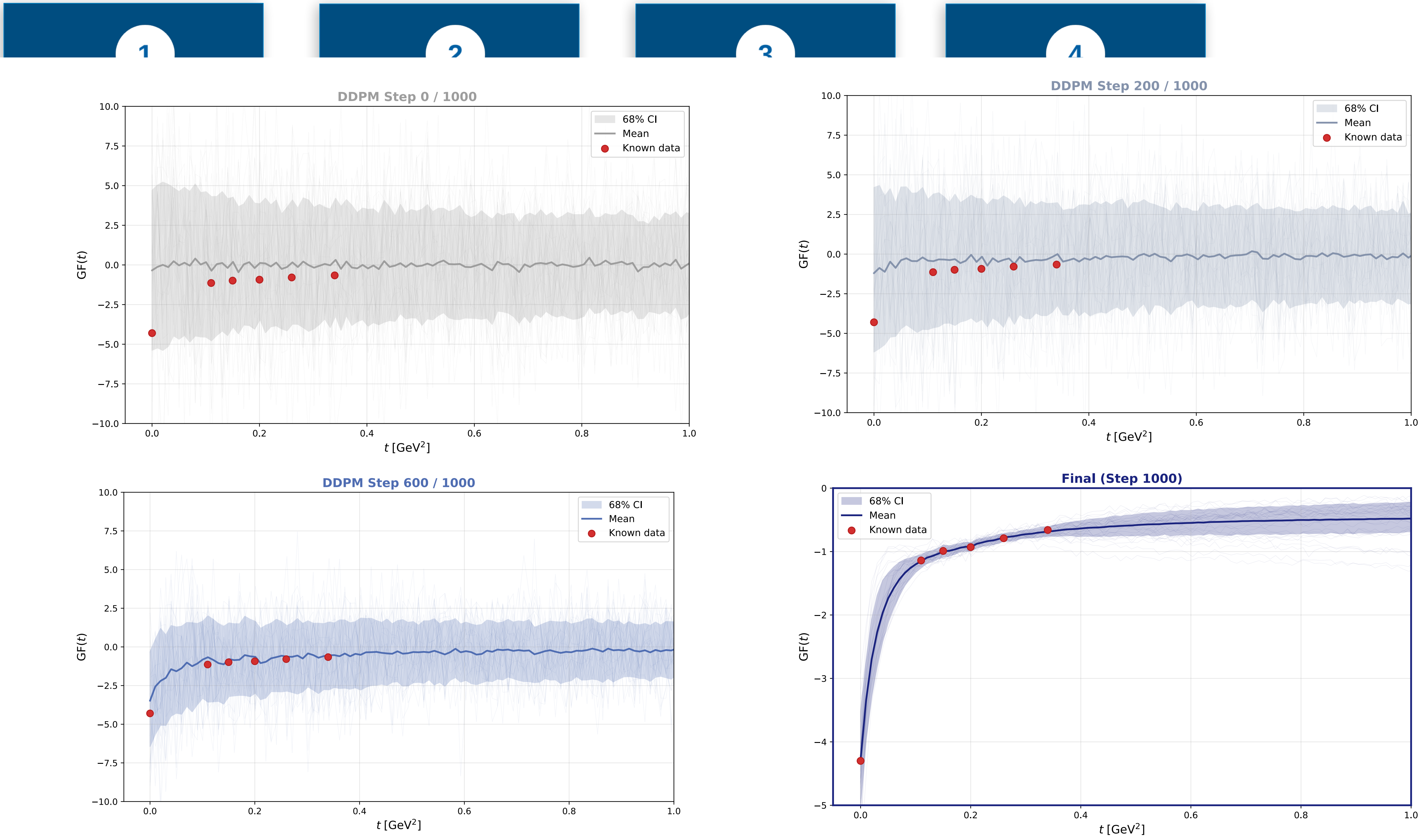
More applications: Diffusion Model for gravitational form factors

H. Alharazin, J. Yu. Panteleeva: arXiv:2602.19267, accepted by PRD

Pipeline Overview

From physics-motivated functions to generated form factors with uncertainties

Slides from Julia Yu. Panteleeva



Prospects for TMD/FF reconstruction using Diffusion Model

Some of recent ML-based TMD extractions:

1. A. Bacchetta, **V. Bertone**, C. Bissolotti, M. Cerutti, M. Radici, S. Rodini, L. Rossi arXiv:**2502.04166** — first NN extraction of unpolarized quark TMDs from Drell-Yan data, NN parametrization of nonperturbative part *(see talk by Valerio Bertone)*
2. I. P. Fernando, D. Keller, arXiv:**2510.17243** — DNN extraction in $k_\perp$ space (E288, E605)
3. Z.-B. Kang, J. R. Sellers, X.-N. Zhang, T. Zhou, arXiv:**2604.14133** — MLP emulator + Bayesian MCMC, N^{3}LO + N^{4}LO
4. M. Zaccheddu, L. Gamberg, W. Melnitchouk, D. Pitonyak, A. Prokudin, J.-W. Qiu, N. Sato arXiv:**2605.06606** *(May 7, 2026)* — non-parametric pixel basis + normalizing-flow MH; first identification of *null TMDs* via SVD *(see talk by Marco Zaccheddu)*
5. *(more ….)*

Prospects for TMD/FF reconstruction using Diffusion Model

H. Alharazin, J. Yu. Panteleeva, Q. Maaz, B-.D Sun, in preparation.

Why score-based diffusion is a complementary tool

1. Trains from samples only. no tractable likelihood, no invertibility constraint *(NF requires both)*
2. Fully amortized: one trained model handles many conditioning sets
3. Less constraint, more flexibility: possible advantage for high-dimensional and multimodal reconstruction of TMDs.

Plans for TMD / CSK / FF reconstruction

1. **CS kernel first.** Apply the GFF-style DDPM pipeline to the new lattice CSK results
    - Tan et al., arXiv:**2511.22547** — continuum + physical pion mass
    - target the nonperturbative window $b_\perp \sim 0.1 - 1$ fm
2. **TMD PDFs in $b_\perp$ space**, conditioning on (lattice CSK) + (DY/SIDIS data); compare against null-TMD directions identified in arXiv:2605.06606
3. **TMD FFs** — same kernel structure, e^+e^-+SIDIS data
4. Long term: simultaneous $\{f_1(x, k_\perp), K(b_\perp), D_1(z, k_\perp)\}$ reconstruction in one diffusion model

Summary & Outlook

1. Our score-based diffusion model for SU(2) gauge theory in 2D, *(arXiv:2602.09045)*
    - Local + topological observables reproduced; not *frozen in topological sector*
    - $\tau \leq 1$across β∈[1,4.5] → no critical slowing down within tested range
2. SU(N) case is also worked out very recently (Komijani et al, arXiv:2605.06134), but Fermion configurations are still missing to be fully useful!
3. Reconstruction of effective mass using DM is on-going
4. Same generative-AI machinery transfers to inverse problems in hadron structure, demonstrated on gravitational form factors (arXiv:2602.19267)
5. Natural next step: TMDs, CS kernel, and TMD FFs — combining lattice + experimental data through a single conditional diffusion model
6. Test diffusion model as a complementary thread (especially) to Zaccheddu et al., arXiv:2605.06606, (**pixel-based non-parametric**, normalizing flows + MH, first identification of null TMDs.)

Thanks for your attention!