

Diffusion Sampling with Dynamical Fermions

Physics-factorized scores and coupling-conditioned generation

A two-dimensional Gross–Neveu study

How much of the target measure is learned?

Physical observables, momentum-mode bias and finite corrections.

Chuanyang Li

Department of Physics, Swansea University

In collaboration with

Gert Aarts, Simon Hands, Lingxiao Wang, Kai Zhou and Yang-yang Tan

6th China Lattice QCD Workshop · Shenzhen · 11 October 2026

AI for lattice: a brief overview



Normalizing flows

2023–2025

Reshape simple random noise into field configurations through a reversible map.

Diffusion models

2024–2026

Start from noise and gradually denoise it into a configuration. This is our route.

Learned Dirac solvers

2023–2026

Give the solver a learned head start, reducing the iterations needed for fermion calculations.

Physics extraction

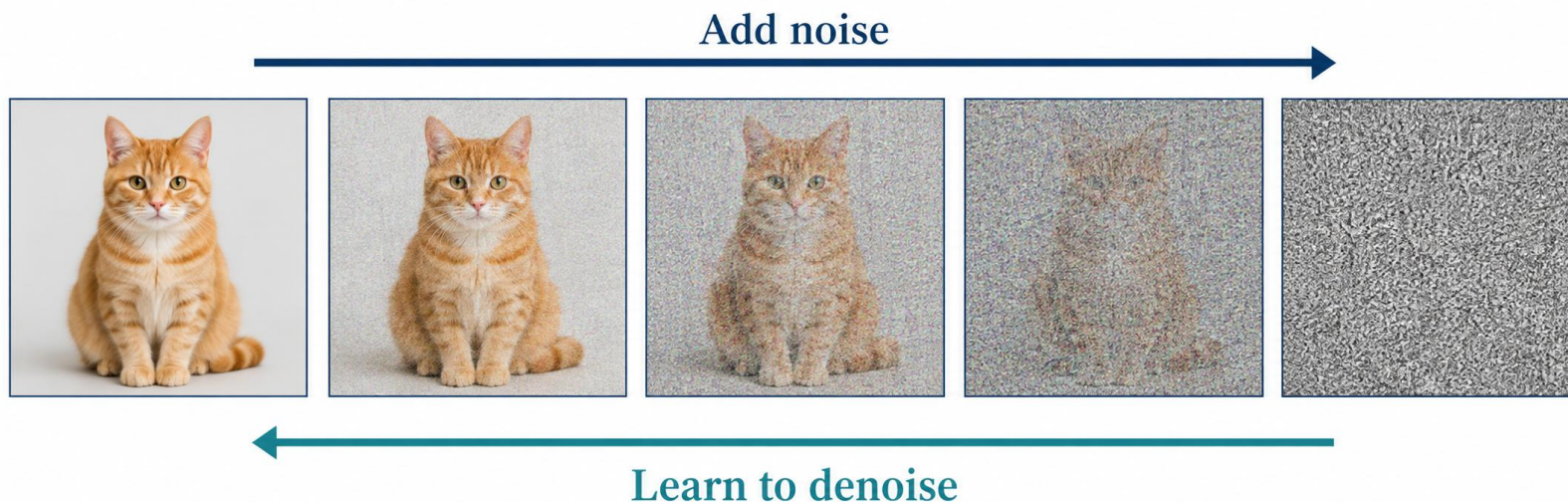
Example: 2024

Learn to reconstruct spectral information from Euclidean correlation functions.

This talk: diffusion sampling with dynamical fermions



Diffusion Models: Learning to Denoise



$$x_{\tau} = x_0 + v(\tau)\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I)$$

Training: learn from known noise

Add known noise to examples. Train a network to predict it at many noise levels.

Generation: start from fresh noise

Apply many small denoising updates. Different random starts produce different samples.

Transformers: Learning from Context

The **animal** crossed the road because **it** was tired.



The diagram illustrates the concept of self-attention in a Transformer model. It shows the sentence "The animal crossed the road because it was tired." with the words "animal" and "it" highlighted in teal. A thick teal curved arrow points from the token "it" back to the token "animal", representing a strong attention weight. Two thinner grey curved arrows point from "it" to "crossed" and "road", representing lower attention weights. This visualizes how the model uses context to interpret the pronoun "it".

Which earlier information helps interpret “it”?

Tokens exchange context

Tokens are pieces of the input. Self-attention combines their information with input-dependent weights.

A Diffusion Transformer (DiT)

Uses Transformer blocks as the network that predicts the denoising field at each noise level.

Schematic illustration



Diffusion generation and MALA correction

Reverse diffusion: from noise (time 1) to a field (time 0)

$$d\sigma = -g^2(\tau)s_\tau(\sigma; \beta) d\tau + g(\tau) d\bar{W}, \quad \tau : 1 \rightarrow 0$$

First term: learned drift

The score is the log-density gradient. It guides denoising; the noise schedule sets its strength.

Second term: random fluctuations

Brownian noise adds random fluctuations. Its amplitude decreases during denoising.

MALA: Metropolis-adjusted Langevin algorithm

$$\sigma' = \sigma - \varepsilon \nabla_\sigma S_{\text{eff}}[\sigma; \beta] + \sqrt{2\varepsilon} \xi, \quad \xi \sim \mathcal{N}(0, I)$$

Propose a small downhill step in the action plus Gaussian noise, then accept or reject it using the full action and proposal probabilities. The step size controls both drift and noise.

Why sample across couplings?

What we have

HMC ensembles at five couplings. Each includes the dynamical fermion determinant.

What we want

New configurations at other couplings, using the same trained model and fresh noise.

Train at a few couplings. Generate at the requested coupling.

$$q_{\theta}(\sigma; \beta) \approx \pi(\sigma; \beta)$$

The generated distribution should match the physical target at each coupling.

A controlled dynamical-fermion target

$$S_{\text{eff}}[\sigma; \beta] = \frac{N_f \beta}{2} \sum_x \sigma(x)^2 - N_f \log |\det M[\sigma]|$$

Analytical bosonic structure

The quadratic term gives a known linear restoring force. Its coupling dependence can be handled explicitly in the noisy score.

Nonlocal fermionic response

The determinant couples field modes through the Dirac operator. Its mass term uses a 2×2 plaquette average of the auxiliary field.

$$N_f = 12, \quad L = 36, \quad V = 1296, \quad \beta \in [0.10, 1.60]$$

Zero bare mass · global Z_2 symmetry · periodic space / antiperiodic fermion time

Physics-factorized noisy score

Combine the bosonic action with Gaussian diffusion noise

The quadratic action favors small fields; noise blurs them. Combining these two Gaussian factors gives known coefficients for the network.

$$\kappa = \frac{1}{1 + N_f \beta v_\tau^2}, \quad \beta_{\text{eff}} = N_f \beta + \frac{1}{v_\tau^2}$$

Kappa: Gaussian shrinkage

Rescales the noisy field before it enters the network. It tends to one when the added noise vanishes.

Effective precision

Combines bosonic stiffness with inverse noise variance. The noise scale is the standard deviation introduced earlier.

Conditional DiT: training and generation

01 Sparse HMC anchors

Train at five couplings: 0.10, 0.30, 0.60, 1.00 and 1.50.
Each anchor contributes 2,000 configurations.

02 Conditional residual learning

A 21.6M-parameter DiT learns with denoising score matching; odd-score symmetry is imposed explicitly.

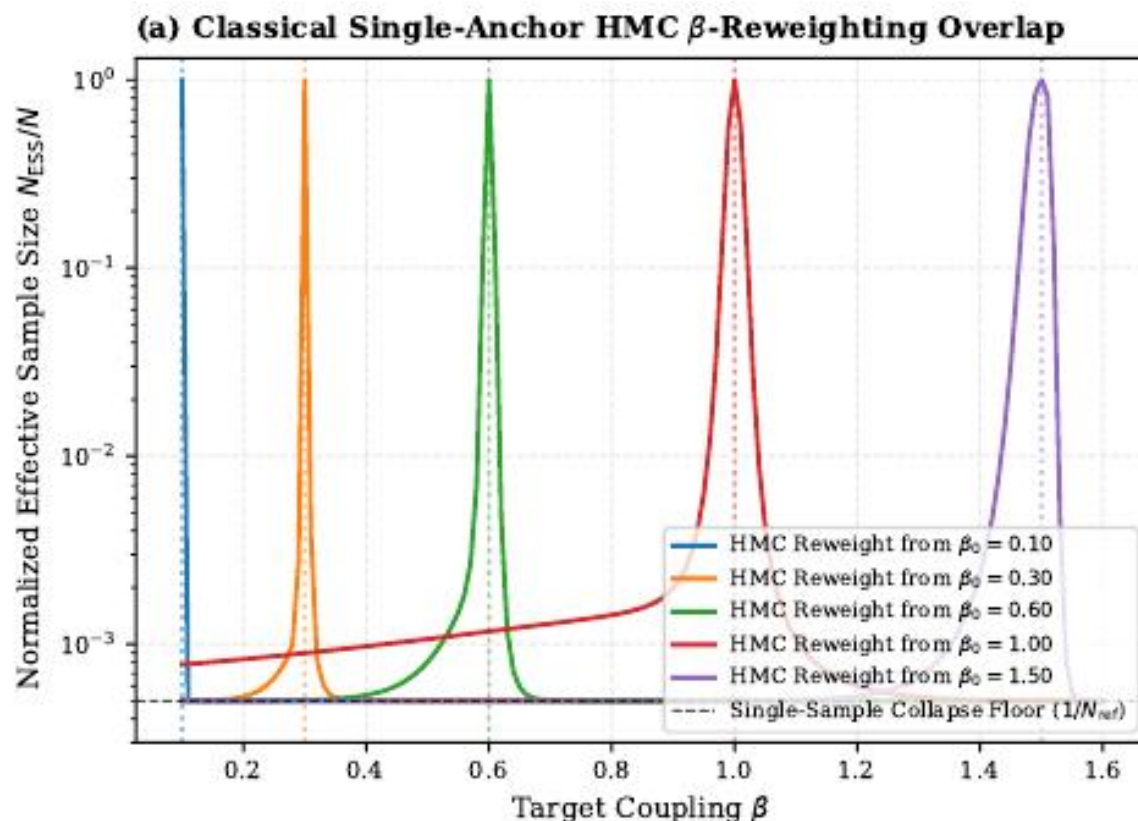
03 Reverse-SDE generation

Use 200 Euler–Maruyama steps to obtain raw samples at the requested coupling.

04 Optional correction

Apply MALA or separate zero / nonzero-mode Metropolis updates, then measure the remaining bias.

Why reweighting is not enough



Five peaks, five HMC anchors

At its own coupling: all 2,000 configurations contribute. Away from it: a few weights dominate.

Narrow peaks, little coverage

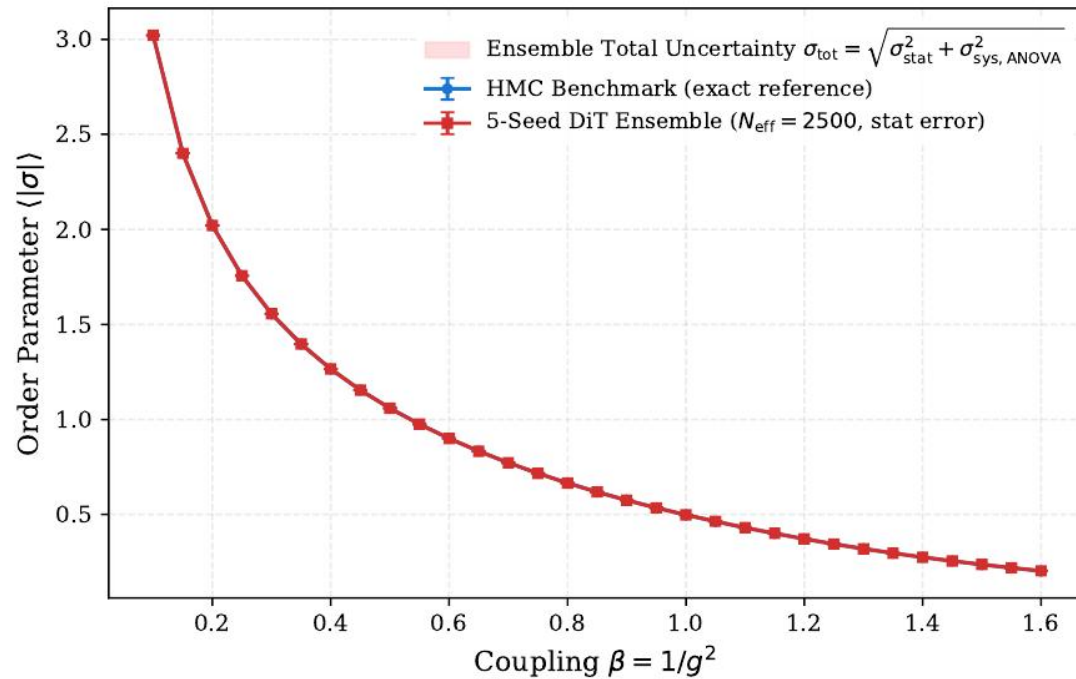
Narrow peaks leave poor coverage between anchors. At the floor, only about one configuration contributes.

Generate new configurations instead of reweighting old ones.

31 target couplings, including 26 unseen points. We now compare their observables with HMC.

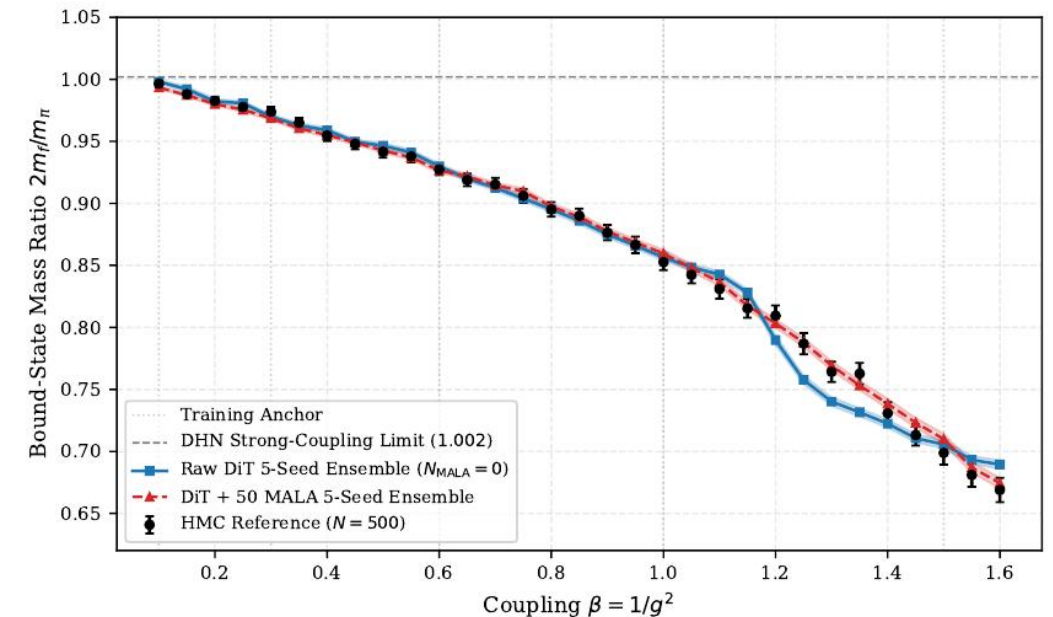
Physical observables across couplings

Order parameter: DiT + MALA



The raw mass ratio agrees to -0.12% on average, with no target coupling beyond 3σ . The corrected order parameter follows HMC across the full scan.

Mass ratio: raw and corrected



$$R = 2m_f/m_\pi$$

Agreement across physical observables

DiT + MALA ensemble: 5 seeds, 31 couplings, 26 unseen points

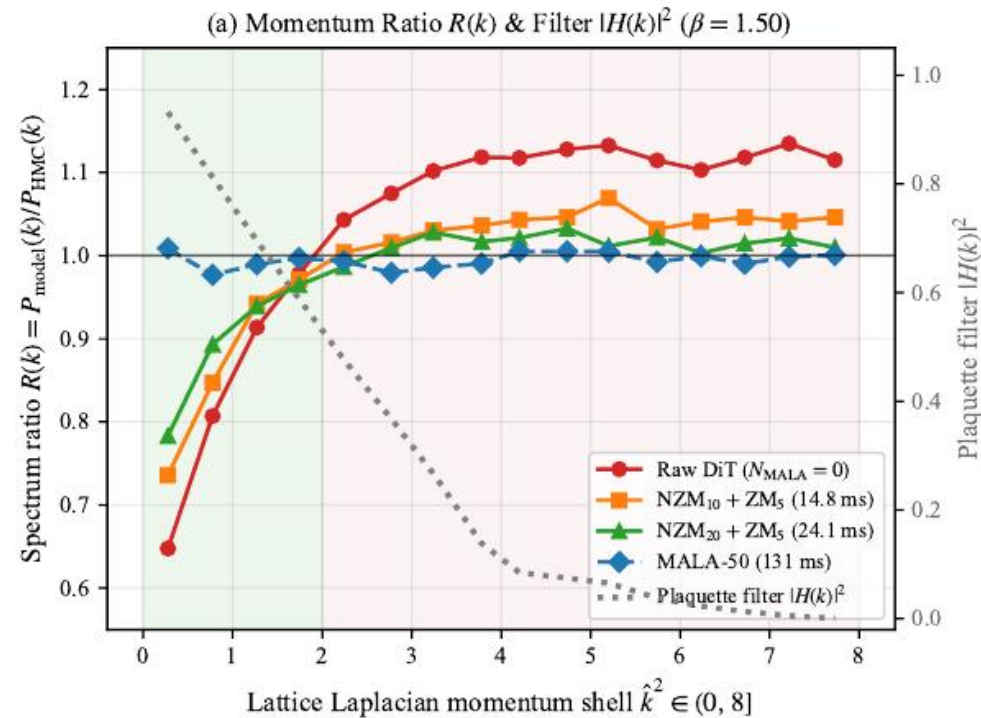
Observable		Mean pull	Largest pull	Points > 3σ
Order parameter	$\langle \bar{\sigma} \rangle$	0.83σ	2.46σ	0 / 31
Chiral condensate	$\langle \bar{\psi} \psi \rangle$	0.84σ	2.32σ	0 / 31
Susceptibility	$\chi/V = \langle \bar{\sigma}^2 \rangle - \langle \bar{\sigma} \rangle^2$	1.03σ	2.26σ	0 / 31
Effective action density	$\langle S_{\text{eff}} \rangle / V$	0.56σ	1.54σ	0 / 31

Agreement covers both field and fermionic observables.

All four observables stay within 3σ of HMC across the scan. Pulls include statistical and seed-to-seed uncertainty; results use 50 MALA updates.

How MALA restores the fluctuation spectrum

Fluctuation power at coupling 1.50



$$R(k) = P_{\text{generated}}(k) / P_{\text{HMC}}(k)$$

$P(k)$: fluctuation power at momentum k .

$R = 1$: HMC agreement. $R > 1$: stronger. $R < 1$: weaker.

What the plot tests

Low momentum probes long distances; high momentum probes short distances. MALA brings the fluctuation power close to HMC.

Why this matters for fermions

$$|H(k)|^2 = \cos^2\left(\frac{k_0}{2}\right) \cos^2\left(\frac{k_1}{2}\right)$$

The arguments are lattice momenta. Plaquette averaging filters ultraviolet power out of the Dirac mass.

Current limitations and improvements

Current limitations

Residual mode-dependent bias

Raw samples do not reproduce every momentum mode. MALA improves agreement but adds physical updates.

One volume, sparse training anchors

Current results use a 36×36 lattice and five training couplings. Independent draws can still share a biased distribution.

Possible improvements

A more accurate fermionic score

Improve the learned residual, especially the short-distance fluctuations revealed by the spectrum.

Broader tests of physical accuracy

Test transfer across volumes and couplings. Add independent physics identities and correction-convergence checks.

Summary and Outlook

Highlights from this work

Learning dynamical fermion effects

Exact bosonic score + learned fermionic residual in the 2D Gross–Neveu model.

Five anchors, 31 target couplings

One conditional model generates at 26 unseen couplings, beyond narrow reweighting overlap.

Physical observables reproduced

After MALA correction, field and fermionic observables agree with HMC across the coupling scan.

Value for dynamical-fermion sampling

Fresh noise, no inherited chain memory

Raw generation avoids HMC burn-in and inherited chain autocorrelations by using independent noise seeds.

Toward shorter equilibration

Warm starts could shorten equilibration with dynamical fermions.

Reuse learned fermionic physics

Parallel draws and coupling reuse may lower cost. Bias still needs validation.

Thank you for your attention !

Questions and discussion