Probing diffusion models for lattice field theory with a configurational thermometer

Diaa Eddin Habibi

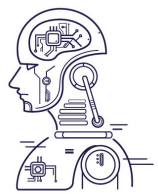
In collaboration with: Gert Aarts, Andreas Ipp, David I. Müller, Thomas R. Ranner, Yang-yang Tan, Lingxiao Wang

ECT*,
7 Sep 2026,
Trento



Swansea University
Prifysgol Abertawe

AIMLAC
UKRI CENTRE FOR DOCTORAL TRAINING IN ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND ADVANCED COMPUTING



References

Wang, Lingxiao, Gert Aarts, and Kai Zhou (2023a). "Diffusion models as stochastic quantization in lattice field theory". In: *JHEP* 05, p. 060. DOI: 10.1007/JHEP05(2024)060. arXiv: 2309.17082 [hep-lat].

— (Nov. 2023b). "Generative Diffusion Models for Lattice Field Theory". In: *37th Conference on Neural Information Processing Systems*. arXiv: 2311.03578 [hep-lat].

Aarts, Gert, Diaa E. Habibi, Lingxiao Wang, and Kai Zhou (2024). "On learning higher-order cumulants in diffusion models". In: *Mach. Learn. Sci. Tech.* 6.2, p. 025004. DOI: 10.1088/2632-2153/adc53a. arXiv: 2410.21212 [hep-lat].

Zhu, Qianteng, Gert Aarts, Wei Wang, Kai Zhou, and Lingxiao Wang (2024). "Diffusion models for lattice gauge field simulations". In: *Machine Learning and the Physical Sciences: Workshop at NeurIPS 2024*. arXiv: 2410.19602 [hep-lat].

— (2025). "Physics-conditioned diffusion models for lattice gauge theory". In: *JHEP* 03, p. 111. DOI: 10.1007/JHEP03(2026)111. arXiv: 2502.05504 [hep-lat].

Aarts, Gert, Diaa E. Habibi, Andreas Ipp, David I. Müller, Thomas R. Ranner, Lingxiao Wang, Wei Wang, and Qianteng Zhu (Jan. 2026). "Generalizable Equivariant Diffusion Models for Non-Abelian Lattice Gauge Theory". In: arXiv: 2601.19552 [hep-lat].

Tan, Yang-yang, Gert Aarts, Diaa E. Habibi, Biagio Lucini, and Lingxiao Wang (July 2026). "Diffusion Models for Sampling Near Criticality in Lattice Field Theories". In: arXiv: 2607.08505 [hep-lat].

Introduction

- Machine learning generative approaches to LFT are emerging as promising auxiliary methods.
- Today's talks already showcase this:

2. Combining annealing and diffusion for efficient transport in LFT
Gurtej Kanwar (University of Edinburgh)
07/09/2026, 10:15

1. Generalizable Equivariant Diffusion Models for LGT
Thomas Ranner (TU Wien)
07/09/2026, 11:30

4. Reconstructing Physical Observables from Sparse Data with Generative Diffusion Models
Dr Herzallah Alharazin (Ruhr Universität Bochum)
07/09/2026, 11:55

5. Generative modeling toward critical behaviors
Yang-yang Tan
07/09/2026, 14:30

7. Stochastic Path Sampler for LFT
Biagio Lucini (Queen Mary University of London (UK))
07/09/2026, 15:45

- How is the correctness of a generative method assessed?
- Can we define a quantity that informs how close a learned distribution is to the target?

Configurational Temperature: Introduction

- Introduced as a geometric approach to thermodynamics in classical statistical mechanics. H.H. Rugh [Phys.Rev. Lett. 78 (1997) 772]
- Written as a ratio of derivative of a Hamiltonian or action

$$\frac{1}{\hat{T}} = \hat{\beta} = \nabla \cdot \left(\frac{\nabla S}{|\nabla S|^2} \right) \Longrightarrow \beta_{\text{conf}} = \langle \hat{\beta} \rangle .$$

  It relies on geometric considerations of the phase space in the microcanonical ensemble.
- Has recently been adapted to lattice field theories. N.S. Dhindsa, A. Joseph, V. Longia [2508.05595]
- Shown to reproduce the input temperature while remaining sensitive to sampling inconsistencies, metastability and algorithmic artefacts.
- Has been applied as a correctness criterion in complex Langevin dynamics. A. Joseph, A. Kumar [2509.08287, 2509.13314]

Configurational Temperature

- Rescale the action $S \to \beta S$ and write the path integral,

$$\mathcal{Z} = \int \mathrm{D}\phi \, e^{-\beta S[\phi]}.$$

  β is the coupling in e.g. pure gauge theories and $\beta \equiv 1$ when rescaling is not possible.

- The gradient and Hessian of the action over configurations

$$g_i = \nabla_i S, \qquad H_{ij} = \nabla_i g_j = \nabla_i \nabla_j S.$$

- The configurational temperature is given by

$$\hat{\beta} = \nabla \cdot \left(\frac{\mathbf{g}}{\mathbf{g}^2} \right) = \frac{\mathrm{Tr}\, H}{\mathbf{g}^2} - 2 \frac{\mathbf{g}^{\mathrm{T}} H \mathbf{g}}{(\mathbf{g}^2)^2}.$$

- $\hat{\beta}$ is intensive, while the second term is suppressed by lattice volume.

Example 1: ϕ^4 scalar field theory

A single scalar field on a d-dimensional lattice, with the action

$$S = \sum_x \left[-2\kappa \sum_{\mu=1}^{d} \phi_x \phi_{x+\hat{\mu}} + (1-2\lambda)\phi_x^2 + \lambda \phi_x^4 \right]$$

- Evaluate its derivatives w.r.t. ϕ_x,

$$g_x = -2\kappa \sum_{\mu} (\phi_{x+\hat{\mu}} + \phi_{x-\hat{\mu}}) + 2(1-2\lambda)\phi_x + 4\lambda\phi_x^3,$$

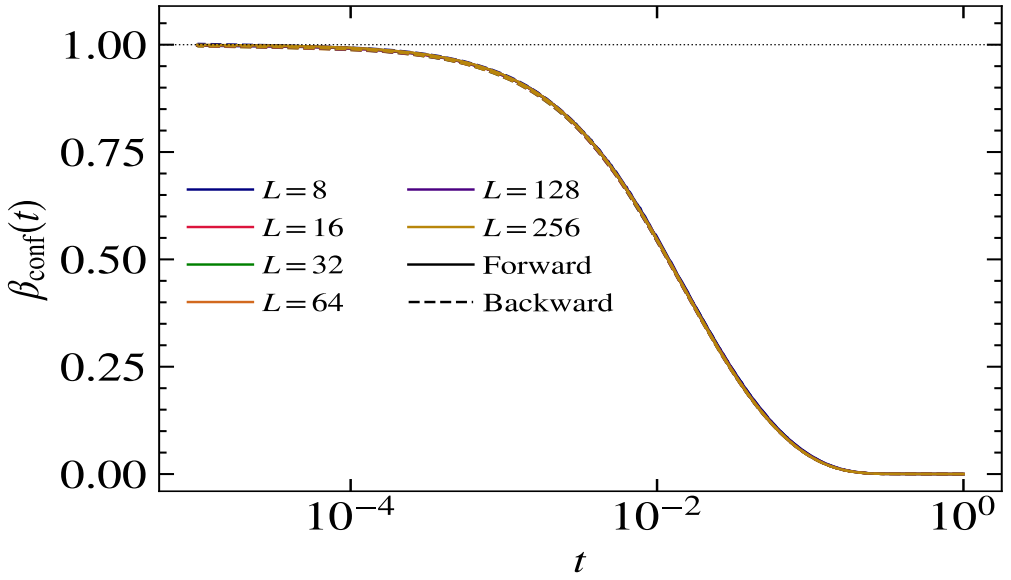
$$H_{xy} = -2\kappa \sum_{\mu} (\delta_{y,x+\hat{\mu}} + \delta_{y,x-\hat{\mu}}) + \left[2(1-2\lambda) + 12\lambda\phi_x^2 \right] \delta_{xy}.$$

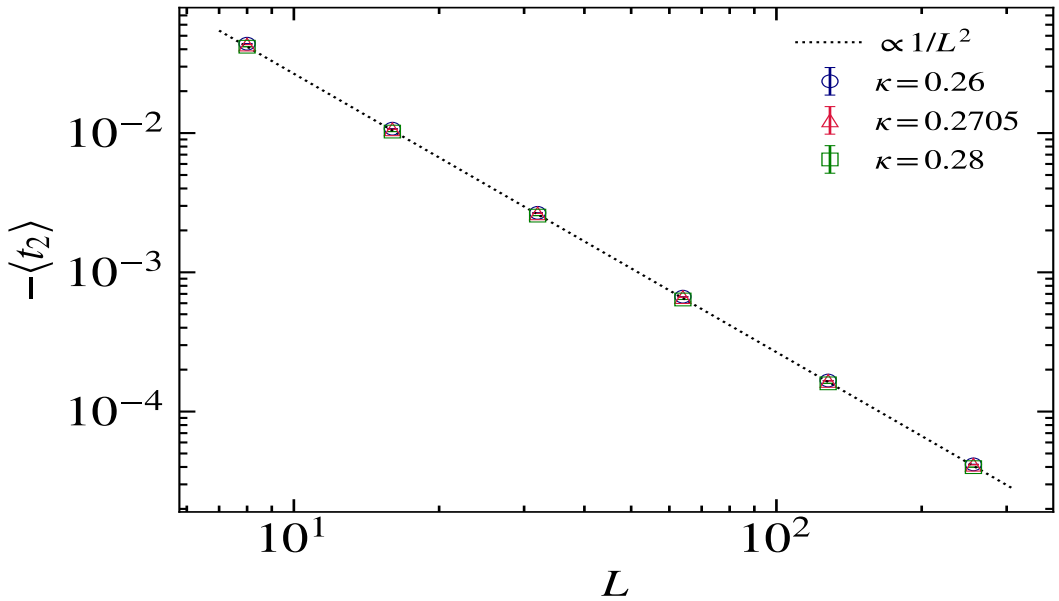
- Evaluate the terms of configurational temperature

$$\mathbf{g}^2 = \sum_x g_x^2, \quad \mathrm{Tr}\, H = \sum_x H_{xx} = \sum_x 2(1-2\lambda) + 12\lambda\phi_x^2,$$

$$\mathbf{g}^{\mathrm{T}} H \mathbf{g} = \sum_{xy} g_x H_{xy} g_y = \sum_x \left[-4\kappa \sum_{\mu} g_x g_{x+\hat{\mu}} + g_x^2 H_{xx} \right].$$

- How does $\beta_{\text{conf}}(t)$ evolve during the diffusion process?
- Data in the broken phase @ $(\kappa, \lambda) = (0.28, 0.022)$ [2607.08505]
- $\beta_{\text{conf}}(t)$ flows from 0 to $\beta = 1$, with the majority of the change occurring at $t < 0.1$.
- The off-diagonal term $t_2 = -2\frac{\mathbf{g}^{\mathrm{T}} H \mathbf{g}}{(\mathbf{g}^2)^2}$ is related inversely to the lattice volume, $-\langle t_2 \rangle \propto L^{-2}$.





Example 2: $U(1)$ Gauge field theory

Consider the compact $U(1)$ in d-dimensions, with the action

$$S[\theta] = \sum_{x,\mu<\nu} [1 - \cos(\theta_{x,\mu\nu})]$$

written with the plaquette $\theta_{x,\mu\nu} = \theta_{x,\mu} + \theta_{x+\hat{\mu},\nu} - \theta_{x+\hat{\nu},\mu} - \theta_{x,\nu}$.

- Evaluate its derivatives w.r.t. $\theta_{x,\mu}$, and terms

$$g_{x,\mu} = \sum_{y,\rho\neq\nu} \frac{\partial\theta_{y,\rho\nu}}{\partial\theta_{x,\mu}} \sin\theta_{y,\rho\nu} = \sum_{\nu\neq\mu} \left(\sin\theta_{x,\mu\nu} - \sin\theta_{x-\hat{\nu},\mu\nu}\right),$$

$$H_{xy,\mu\nu} = \sum_{z,\rho\neq\sigma} \frac{\partial\theta_{z,\rho\sigma}}{\partial\theta_{x,\mu}} \frac{\partial\theta_{z,\rho\sigma}}{\partial\theta_{y,\nu}} \cos\theta_{z,\rho\sigma},$$

$$\mathbf{g}^2 = \sum_{x,\mu} g_{x,\mu}^2, \quad \mathrm{Tr}\, H = \sum_{x,\mu} H_{xx,\mu\mu} = 4 \sum_{x,\mu<\nu} \cos\theta_{x,\mu\nu},$$

$$\mathbf{g}^{\mathrm{T}} H \mathbf{g} = \sum_{x,\mu<\nu} \cos\theta_{x,\mu\nu} \left(\sum_{y,\rho} \frac{\partial\theta_{x,\mu\nu}}{\partial\theta_{y,\rho}} g_{y,\rho} \right)^2.$$

Example 3: $U(N_c)$ Gauge field theory

Consider $U(N_c)$ in d-dimensions, with the action

$$S[U] = \sum_{x,\mu} \left(1 - \frac{1}{N_c} \mathrm{Re}\,\mathrm{Tr}\left[U_{x,\mu} X^\dagger_{x,\mu}\right]\right),$$

$$X^\dagger_{x,\mu} = \sum_{\nu \neq \mu} \left(U_{x+\mu,\nu} U^\dagger_{x+\nu,\mu} U^\dagger_{x,\nu} + U^\dagger_{x+\mu-\nu,\nu} U^\dagger_{x-\nu,\mu} U_{x-\nu,\nu}\right).$$

- Using the group derivative $D^a f(U) = \frac{\mathrm{d}}{\mathrm{d}\epsilon} f(e^{i\epsilon T^a} U)\big|_{\epsilon=0}$,

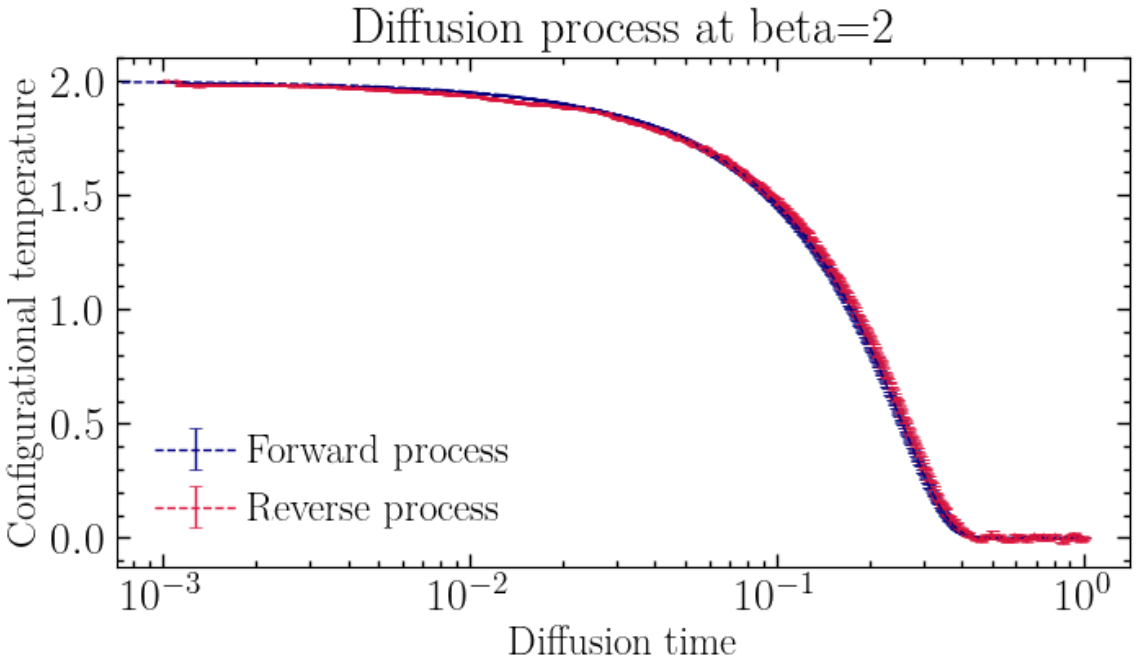
$$g^a_{x,\mu} = D^a_{x,\mu} S[U] = \frac{1}{N_c} \mathrm{Im}\,\mathrm{Tr}\left[T^a U_{x,\mu} X^\dagger_{x,\mu}\right],$$

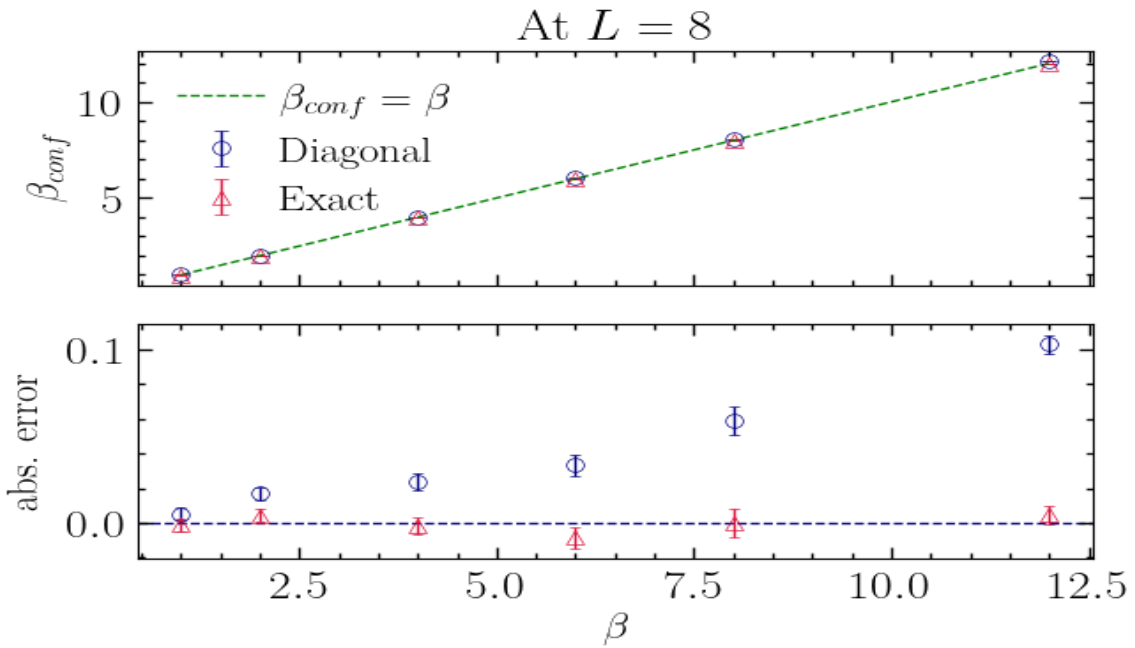
$$\mathrm{Tr}\, H = \sum_{x,\mu,a} D^a_{x,\mu} D^a_{x,\mu} S[U] = \frac{1}{N_c} \sum_{x,\mu,a} \mathrm{Re}\,\mathrm{Tr}\left[T^a T^a U_{x,\mu} X^\dagger_{x,\mu}\right]$$

$$= \frac{C_F}{N_c} \sum_{x,\mu} \mathrm{Re}\,\mathrm{Tr}\left[U_{x,\mu} X^\dagger_{x,\mu}\right], \qquad \sum_a T^a T^a = C_F \mathbb{I}.$$

- Generated for 2d $U(2)$.

  [2602.19552]
- $\beta_{\text{conf}}(t)$ behaves similarly for gauge theories, interpolating between strong coupling and target.
- The diagonal term $\frac{\text{Tr}\, H}{\mathbf{g}^2}$ degrades as it approaches the continuum limit.
- The exact estimator $\hat{\beta}$ remains insensitive.





Application: Testing a sampler

We use β_{conf} as a general test of a sampling scheme's convergence of the SDE

$$\mathrm{d}x = g^2(t) s_\theta(x,t)\mathrm{d}t + g(t)\mathrm{d}W_t.$$

- Euler-Maruyama, $\sim N_t$

$$x_{i+1} = x_i + g^2(t) s_\theta(x,t)\mathrm{d}t + g(t_i)\sqrt{\mathrm{d}t}\; z_i, \quad z_i \sim \mathcal{N}(0,1).$$

- Modified stochastic Heun, $\sim 2N_t$ Karras et al. [2206.00364]

$$\begin{aligned}
\hat{x}_i &= x_i + \sqrt{g^2(\hat{t}_i) - g^2(t_i)}\sqrt{\mathrm{d}t}\; z_i, \quad z_i \sim \mathcal{N}(0,1),\\
\hat{x}_{i+1} &= \hat{x}_i + f(\hat{x}_i, \hat{t}_i)\,\mathrm{d}t, \quad f(x,t) = g^2(\hat{t}_i)\, s_\theta(\hat{x}_i, \hat{t}_i),\\
x_{i+1} &= \hat{x}_i + \frac{1}{2}\Big[f(\hat{x}_{i+1}, t_{i+1}) + f(\hat{x}_i, \hat{t}_i)\Big]\,\mathrm{d}t.
\end{aligned}$$

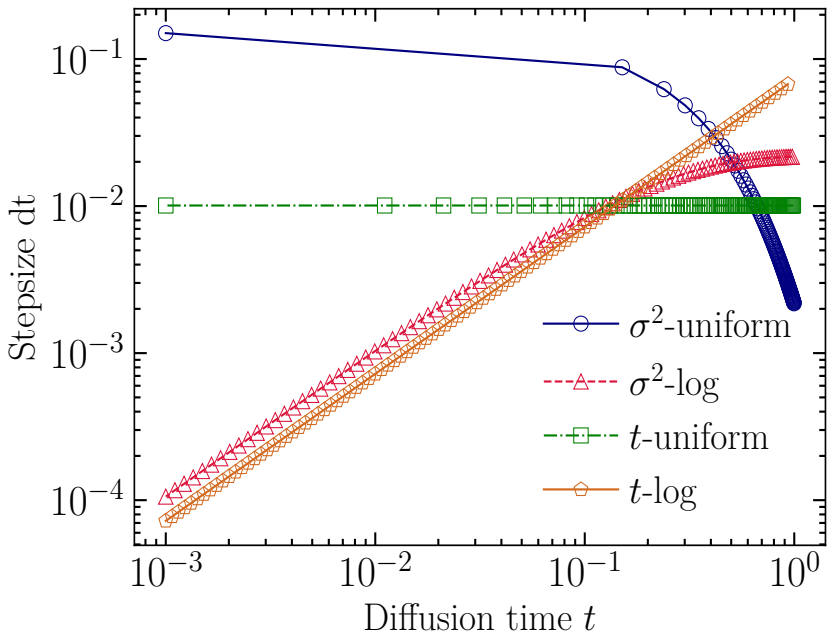
- Metropolis-adjusted annealed Langevin (MAALA), $\sim N_\ell N_t$

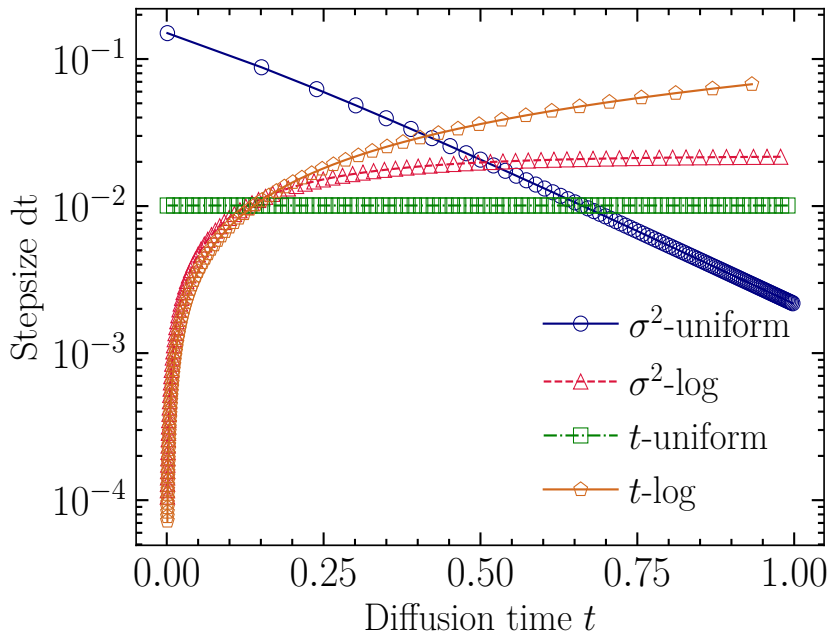
[2601.19552, 2502.05504]

$$x_{i+1} = x_i + s_\theta(x_i, t_i)\mathrm{d}t + \sqrt{2\mathrm{d}t}\; z_i, \quad z_i \sim \mathcal{N}(0,1).$$

Scheduling the generative process

- The reverse process requires scheduling the time steps when integrating the SDE.
- This discretization controls the step size of the integrator at each noise scale.





Schedule comparison - ϕ^4

Compare ϕ^4 in two- and three-dimensions for a linear and logarithmic time schedule.

Figure: Diffusion process @ $d = 2$, in the broken phase $(\kappa, \lambda) = (0.28, 0.022)$.

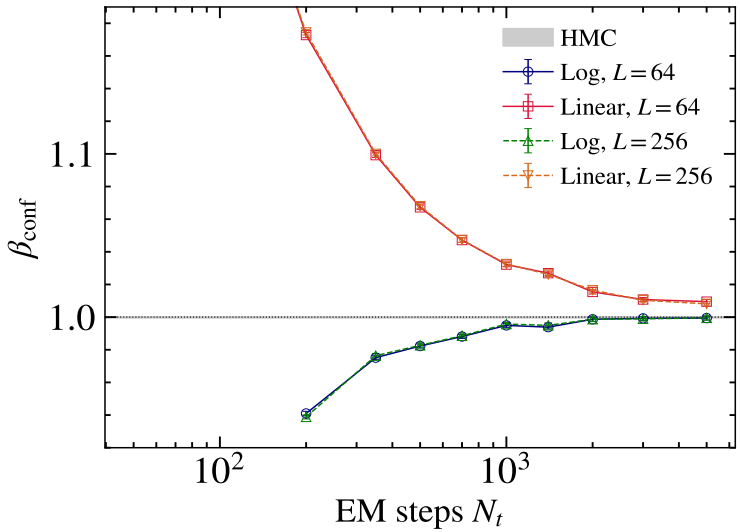
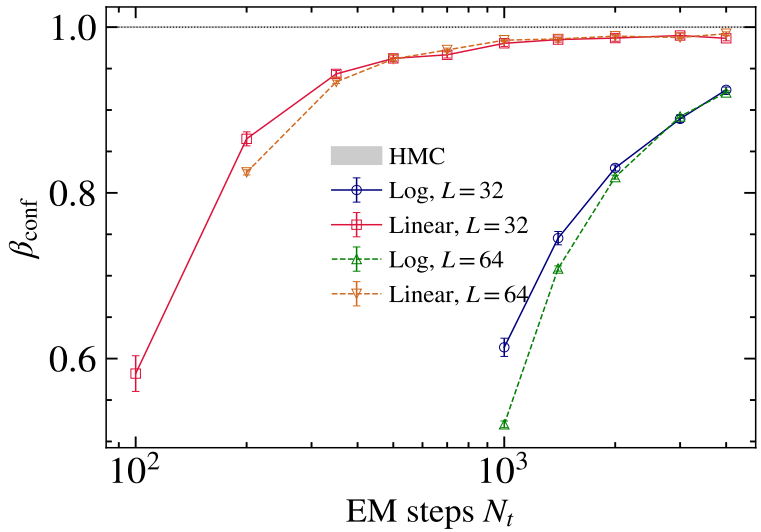
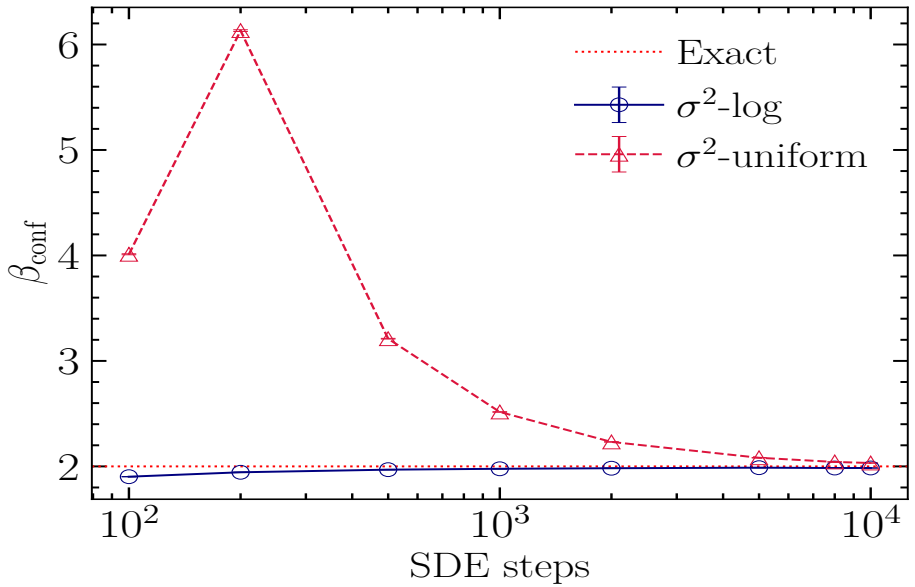


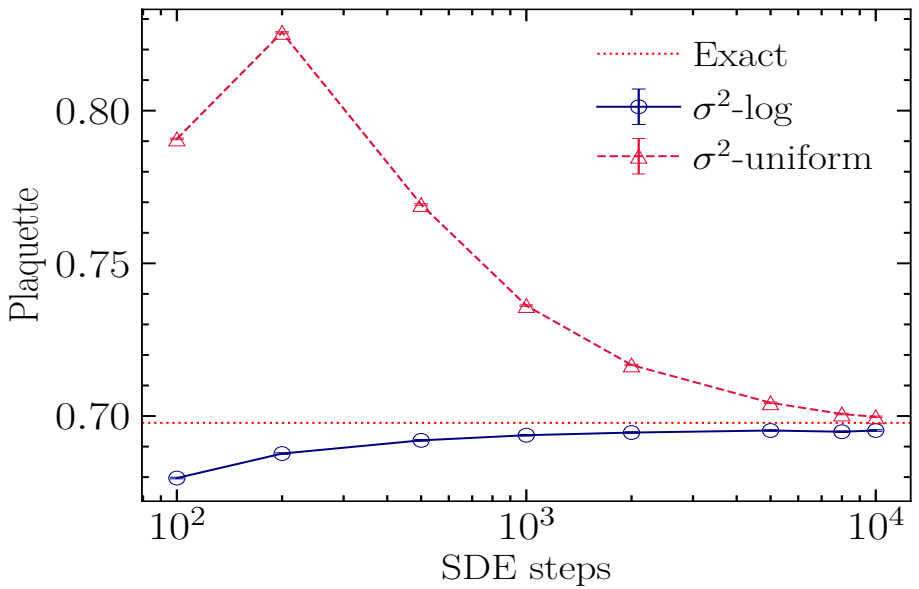
Figure: Diffusion process @ $d = 3$, $(\kappa, \lambda) = (0.1923, 0.9)$.



Schedule comparison: Euler-Maruyama

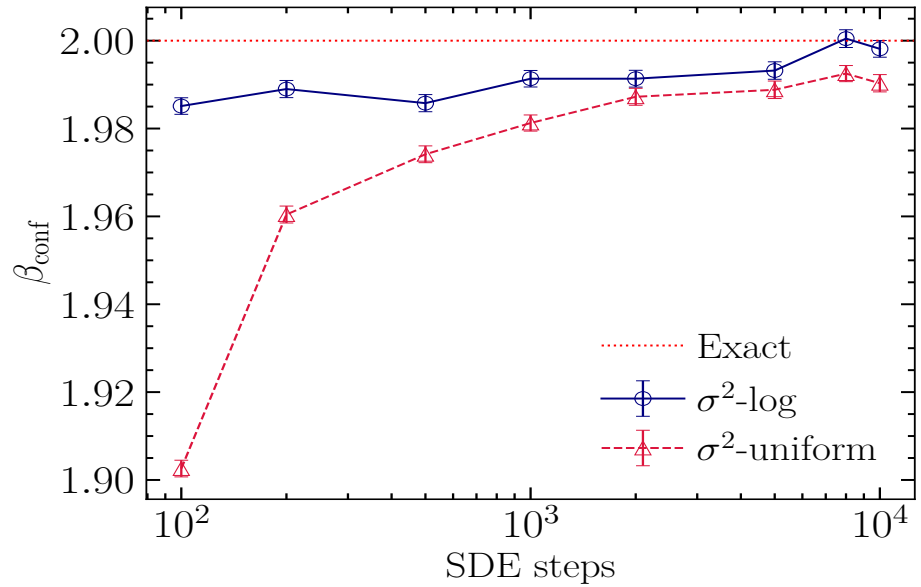
- Model is trained on 2d $U(1)$ @ $\beta = 2, L = 16$.
- Time schedule based on variance $\sigma^2(t)$.
- Uniform schedule shows clear inefficiency while logarithmic schedule achieves convergence swiftly.

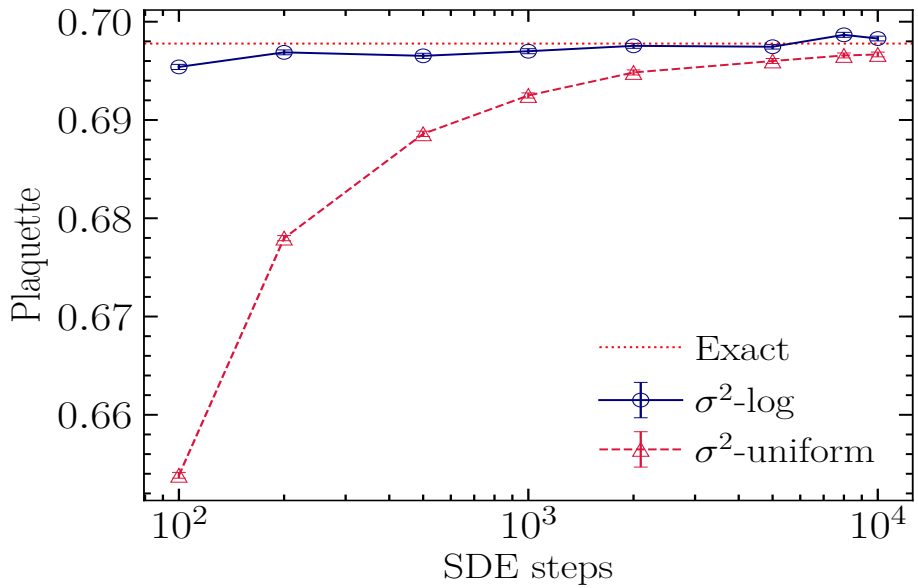




Schedule comparison: Stochastic Heun

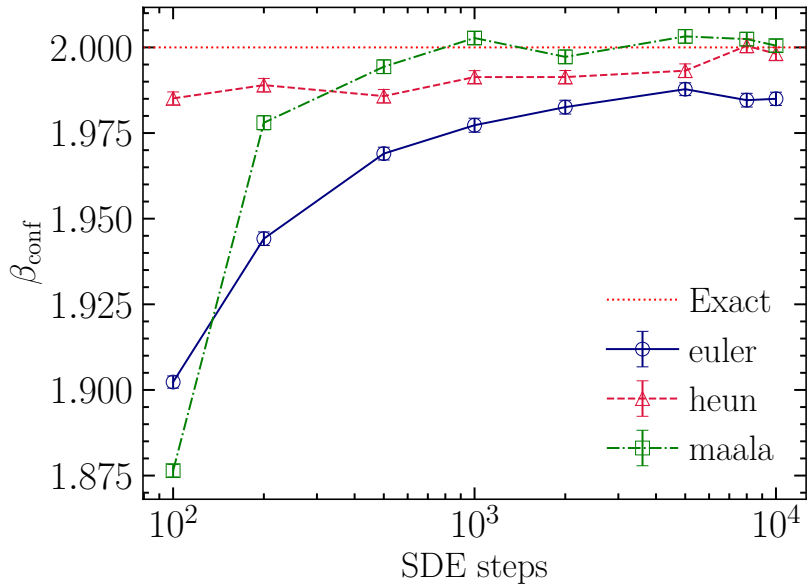
- Model is trained on 2d $U(1)$ @ $\beta = 2, L = 16$.
- Time schedule based on variance $\sigma^2(t)$.
- Uniform schedule shows clear inefficiency while logarithmic schedule achieves convergence swiftly.

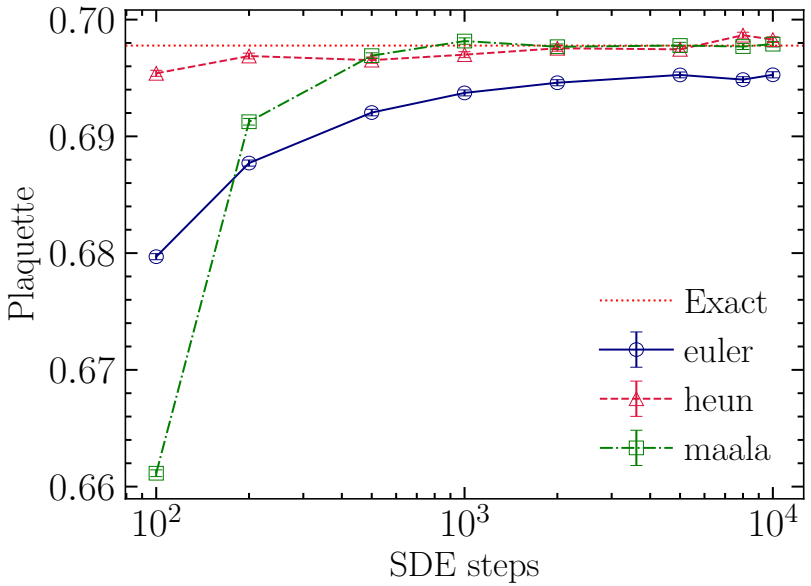




Scheme comparison - $U(1)$

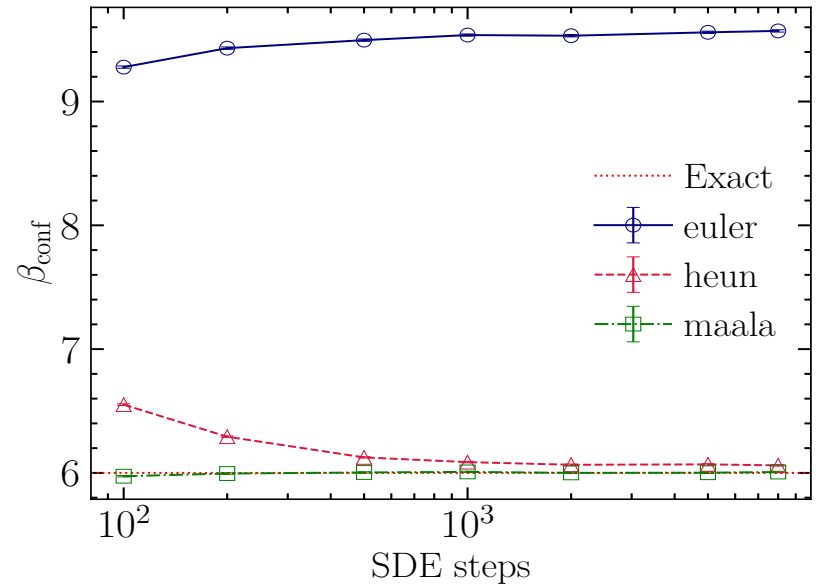
- Data for 2d $U(1)$ is generated @ $\beta = 2, L = 16$.
- Use logarithmic variance schedule.
- MAALA was run with $N_\ell = 100$, $\delta t = 10^{-4}$.
- MAALA and Heun reach similar performance; Heun is less expensive.

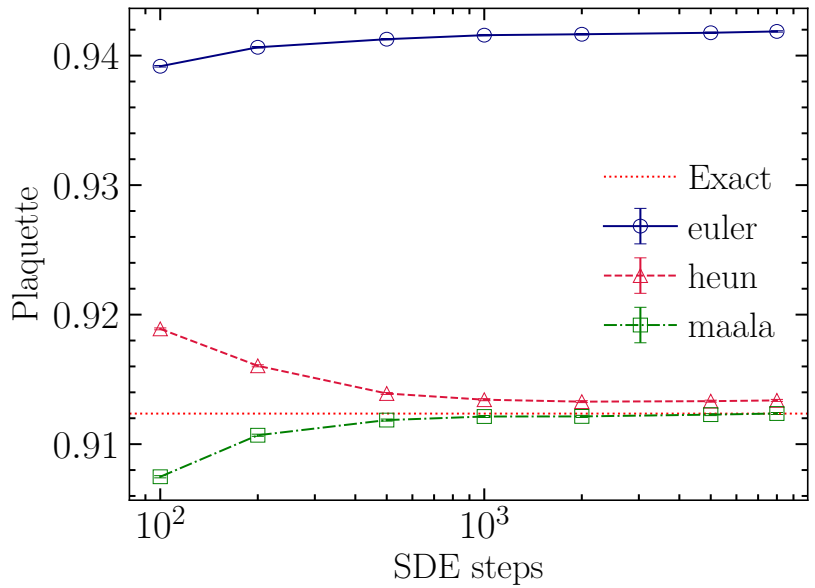




Scheme comparison - $U(1)$

- Data for 2d $U(1)$ is generated @ $\beta = 6, L = 16$ using score-rescaling.
- Use logarithmic variance schedule.
- MAALA was run with $N_\ell = 100$, $\delta t = 10^{-4}$.
- Euler fails completely; Heun performs drastically better in comparison.



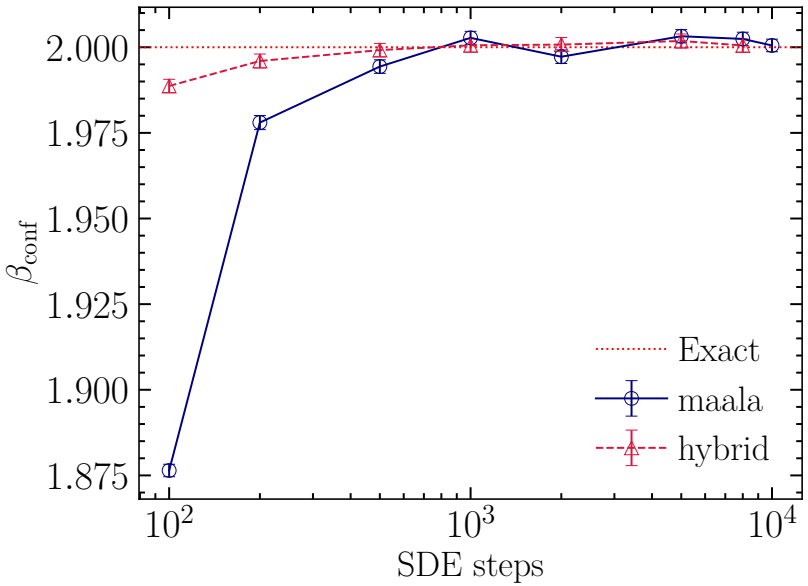


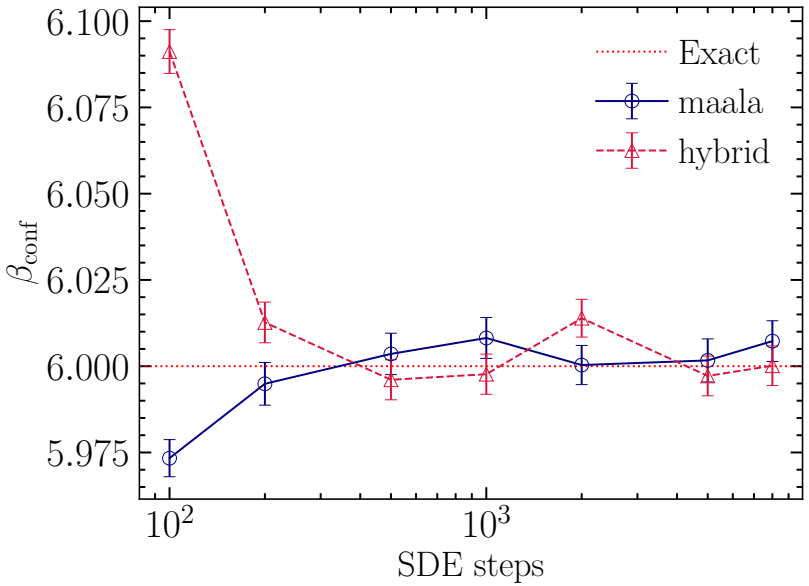
Hybrid sampler

- The performance of the samplers motivates a hybrid approach that reaches similar sampling quality while keeping expense reasonable.
- Heun sampler manages to approach the target distribution efficiently thanks to its predictor-corrector scheme.
- Proceed with the generative process using the Heun method while using the MAALA methodology for the last 5% of the SDE steps.

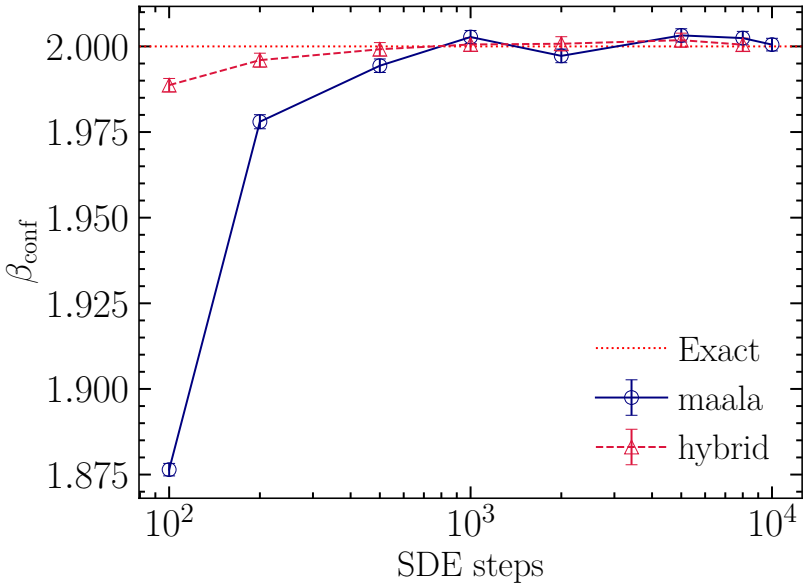
Compare new Sampler

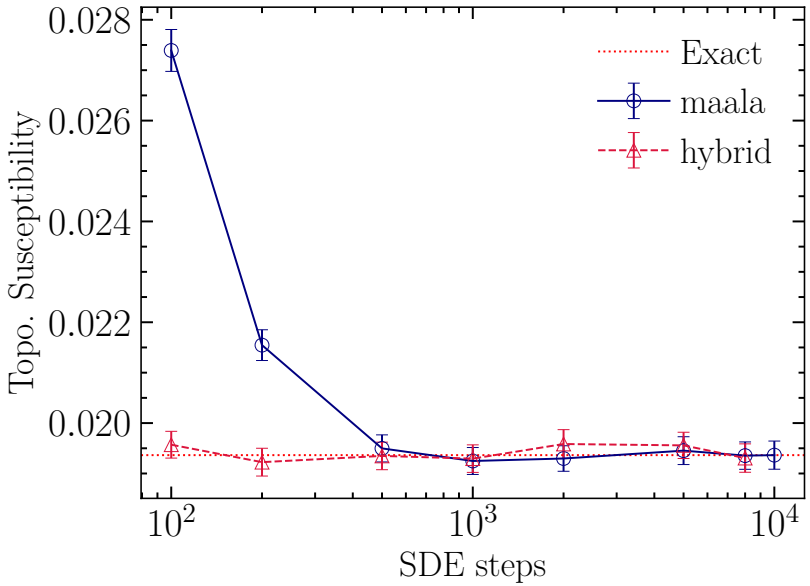
Hybrid sampler performs similarly to MAALA with approximately 5% of the cost.

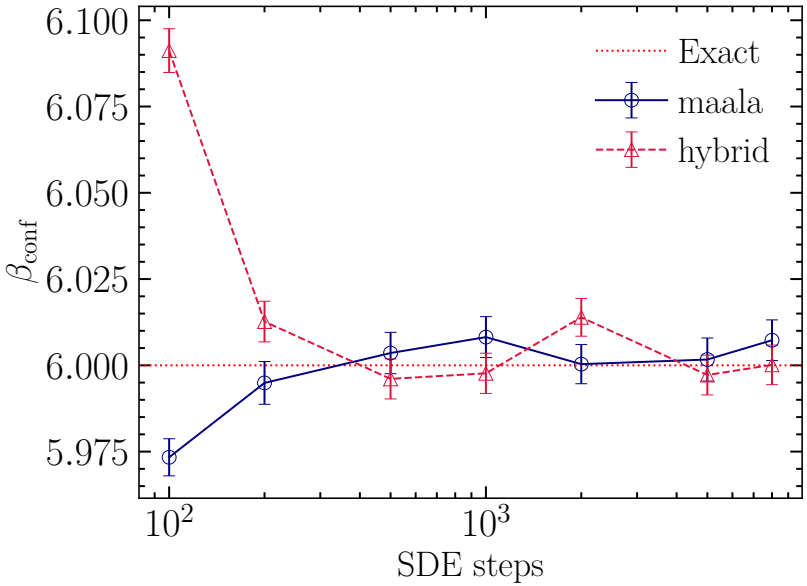


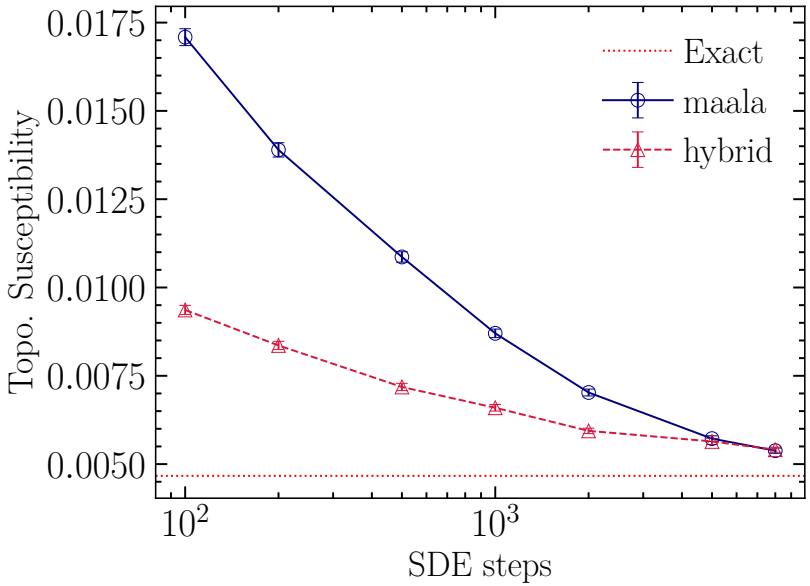


Sensitivity of β_{conf}: Topological susceptibility









Sensitivity of β_{conf}: Schwinger-Dyson

- Measure the violation of the Schwinger-Dyson identity associated with a probe field F under p_t. A. Joseph, 2606.27360

$$\Delta_F(p_t) = \langle \nabla \cdot F - F \cdot \nabla S \rangle_{p_t} = - \left\langle F \cdot \nabla \log \frac{p_t}{P_{\text{eq}}} \right\rangle_{p_t}$$

- The score mismatch $\delta s = \nabla \log \frac{p_t}{P_{\text{eq}}}$ characterizes local deformations of the sampled p_t geometry relative to P_{eq}.
- Probe field F selects a direction in configuration space. $\Delta_F(p_t)$ is the projection of δs onto that direction.
- Different probes differ in sensitivity to the same score mismatch. For $F = \frac{\nabla S}{|\nabla S|^2}$, $\beta_{\text{conf}} = \langle \nabla \cdot F \rangle_{p_t}$ and

$$|\beta_{\text{conf}} - \beta_{\text{target}}|^2 \leq \langle |F|^2 \rangle_{p_t} \langle |\delta s|^2 \rangle_{p_t}.$$

Summary & Outlook

- Configurational temperature provides a general diagnostic tool for assessing thermodynamic consistency in generative approaches to lattice field theory.
- Can be evaluated for any generative method that defines a flow between a prior and a target distribution.
- Consistency check when using techniques like score-rescaling that allow for out-of-training sampling.
- Possible future application to scaling to 4d SU(3) theories.

Summary & Outlook

- Configurational temperature provides a general diagnostic tool for assessing thermodynamic consistency in generative approaches to lattice field theory.
- Can be evaluated for any generative method that defines a flow between a prior and a target distribution.
- Consistency check when using techniques like score-rescaling that allow for out-of-training sampling.
- Possible future application to scaling to 4d SU(3) theories.

Thank you for listening!