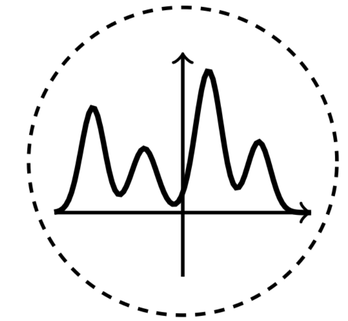
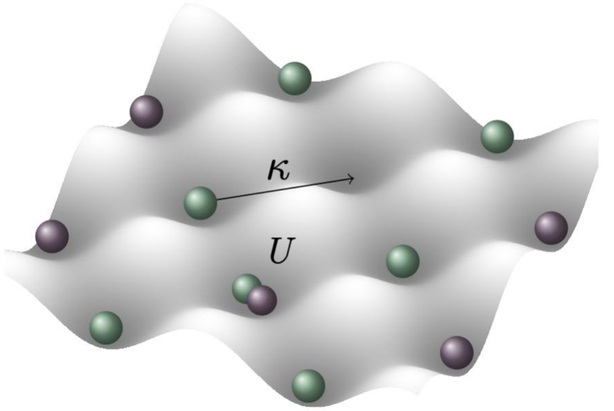




Normalizing Flows for the Hubbard Model

Lena Funcke

ECT* Workshop “Bridging machine learning and statistical field theory through generative and diffusion processes”, 10/09/2026

Slides courtesy of Dominic Schuh

UNIVERSITÄT BONN

TRA MATTER TRANSDISCIPLINARY RESEARCH AREA

HISKP

LAMARR INSTITUTE FOR MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

color meets flavor

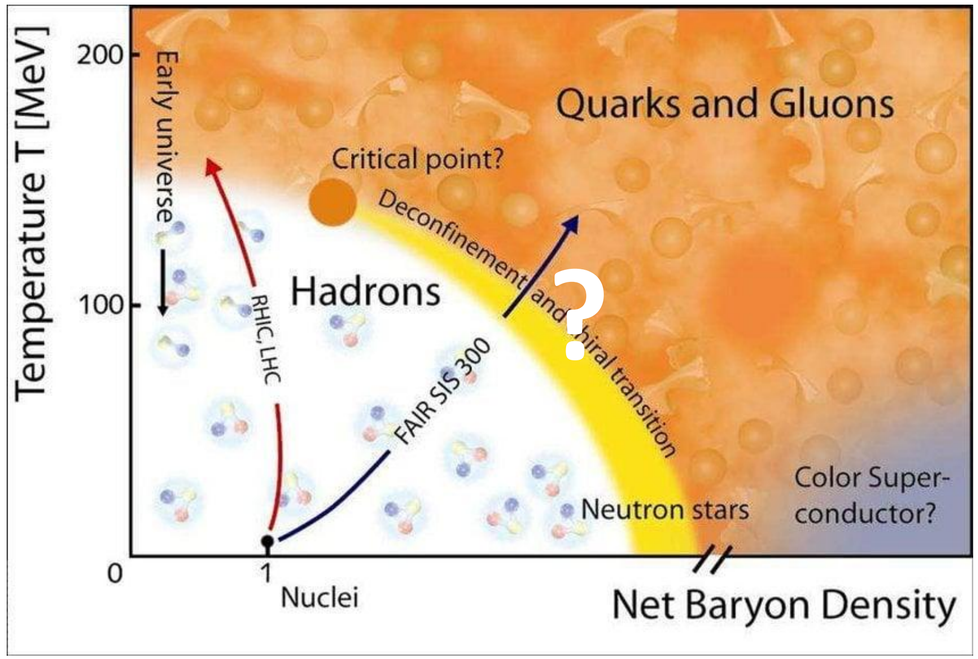
ML4Q MATTER AND LIGHT FOR QUANTUM COMPUTING

⟨NUMERIQS⟩

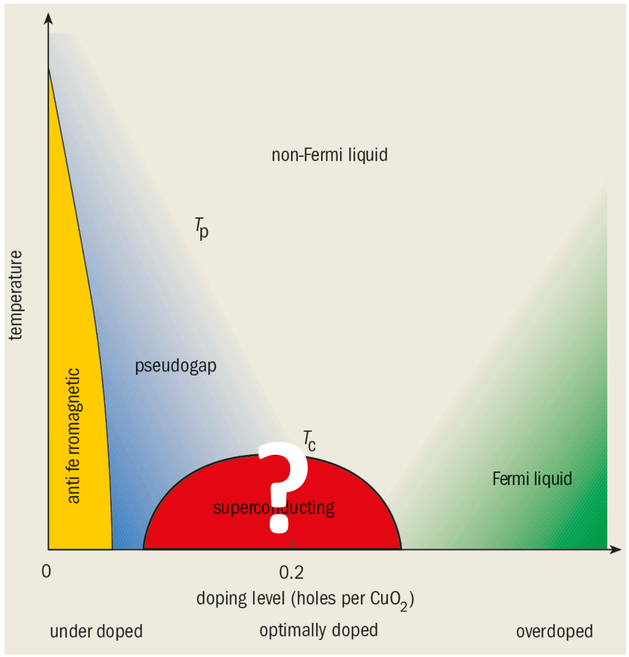
Motivation

Phase diagrams at finite density

QCD



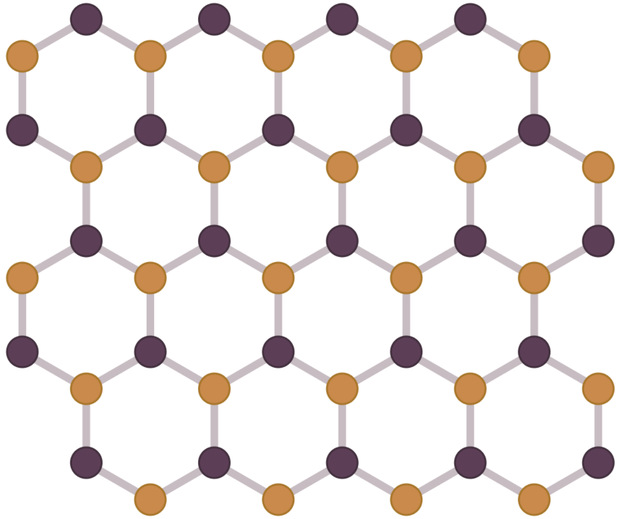
CuO_2 (cuprate) superconductors



→ candidate minimal Hamiltonian: Hubbard model

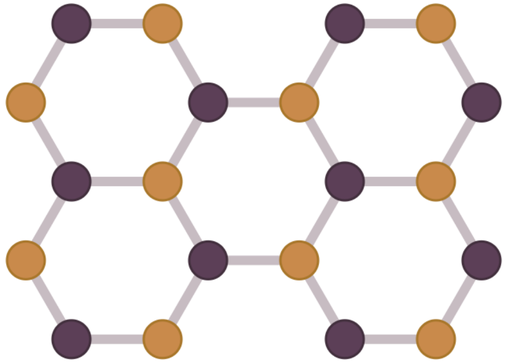
Image courtesy of GSI Helmholtzzentrum für Schwerionenforschung GmbH, Darmstadt/Germany;
National Research Council (2005), *Opportunities in High Magnetic Field Science,* National Academies Press

Carbon-based systems



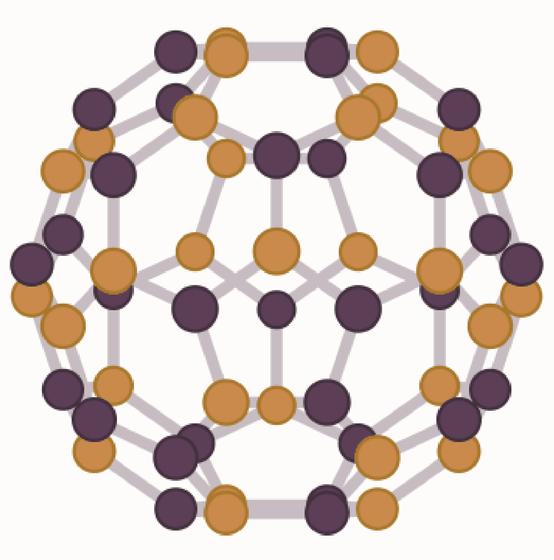
Graphene

Ostmeyer et al., PRB 2021



$C_{20}H_{12}$

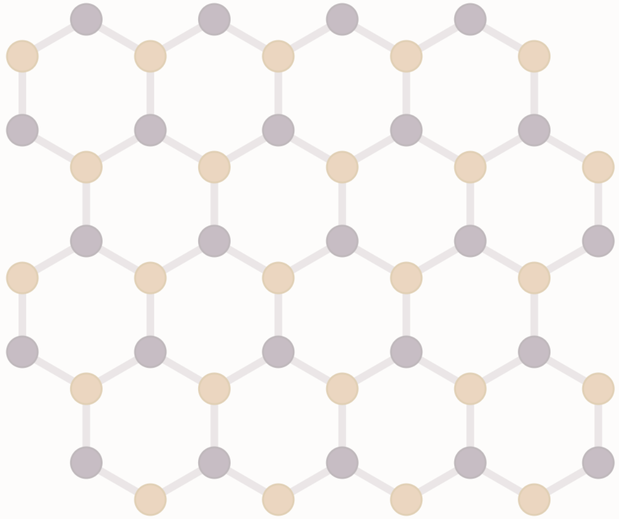
Rodekamp et al., EPJB 2025



C_{60} Fullerene

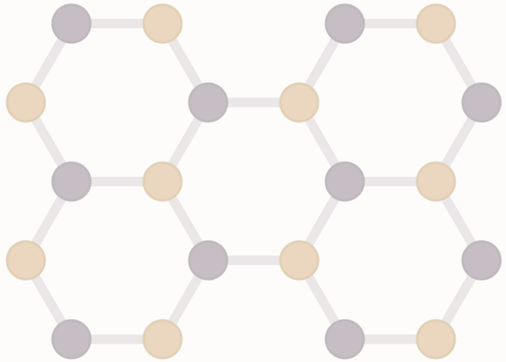
Gäntgen et al., PRB 2024

Carbon-based systems



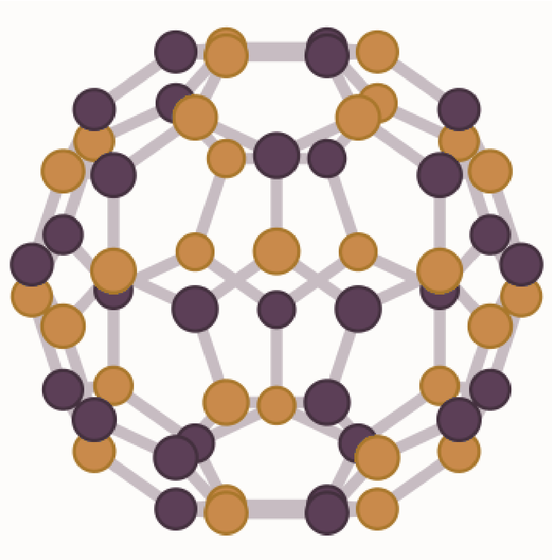
Graphene

Ostmeyer et al., PRB 2021



$C_{20}H_{12}$

Rodekamp et al., EPJB 2025

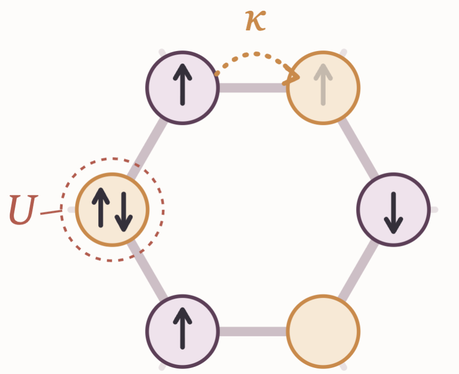


C_{60} Fullerene

Gäntgen et al., PRB 2024

The Hubbard model

$$H = -\kappa \sum_{\langle x,y \rangle} \left(a^{\dagger}_{x,\uparrow} a_{y,\uparrow} + a^{\dagger}_{x,\downarrow} a_{y,\downarrow} \right) - \frac{U}{2} \sum_{x} \left(n_{x,\uparrow} - n_{x,\downarrow} \right)^2 - \mu \sum_{x} q_x$$



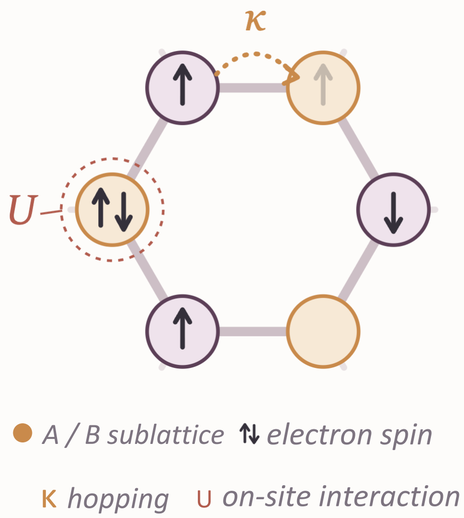
● *A / B sublattice* ⇅ *electron spin*

K *hopping* U *on-site interaction*

The Hubbard model

$$H = -\kappa \sum_{\langle x,y \rangle} \left(a^{\dagger}_{x,\uparrow} a_{y,\uparrow} + a^{\dagger}_{x,\downarrow} a_{y,\downarrow} \right) - \frac{U}{2} \sum_{x} \left(n_{x,\uparrow} - n_{x,\downarrow} \right)^2 - \mu \sum_{x} q_x$$

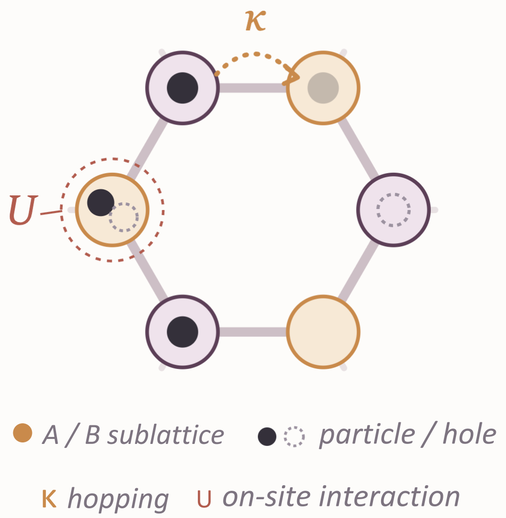
- Mott insulator
- Magnetism
- Superconductivity



The Hubbard model

$$H = -\kappa \sum_{\langle x,y \rangle} \left(a^{\dagger}_{x,\uparrow} a_{y,\uparrow} - a^{\dagger}_{x,\downarrow} a_{y,\downarrow} \right) - \frac{U}{2} \sum_{x} \left(n_{x,\uparrow} - n_{x,\downarrow} \right)^2 - \mu \sum_{x} q_x$$

- Mott insulator
- Magnetism
- Superconductivity



Computational methods

Computational methods

Hamiltonian methods

- Matrix product states
- Projected entangled pair states
- Neural quantum states

Computational methods

Hamiltonian methods

- Matrix product states
- Projected entangled pair states
- Neural quantum states

Path-integral methods

- Auxiliary field quantum Monte Carlo
  - Hybrid Monte Carlo (HMC)

Computational methods

Hamiltonian methods

- Matrix product states
- Projected entangled pair states
- Neural quantum states

Path-integral methods

- Auxiliary field quantum Monte Carlo
  - Hybrid Monte Carlo (HMC)
- Normalizing flows *(this work)*

Computational methods

Hamiltonian methods

- Matrix product states
- Projected entangled pair states
- Neural quantum states

Path-integral methods

- Auxiliary field quantum Monte Carlo
  - Hybrid Monte Carlo (HMC)
- Normalizing flows *(this work)*

There is *no* one-fits-all method: each trades off system size, interaction strength and sign structure.

Path-integral formulation

Path-integral formulation

HUBBARD AUXILIARY-FIELD ACTION

$$S[\phi] = \frac{1}{2\widetilde{U}} \sum_{x,t} \phi_{x,t}^2 - \log det M[\phi] - \log det M[-\phi]$$

Path-integral formulation

HUBBARD AUXILIARY-FIELD ACTION

$$S[\phi] = \frac{1}{2\widetilde{U}} \sum_{x,t} \phi_{x,t}^2 - \log det M[\phi] - \log det M[-\phi]$$

OBSERVABLES & MONTE CARLO ESTIMATOR

$$\langle O \rangle = \lim_{N\to\infty} \sum_{i=0}^{N} O(\phi_i)$$

$$\phi \sim Z^{-1} e^{-S[\phi]}$$

Flow-Based Generative Sampling

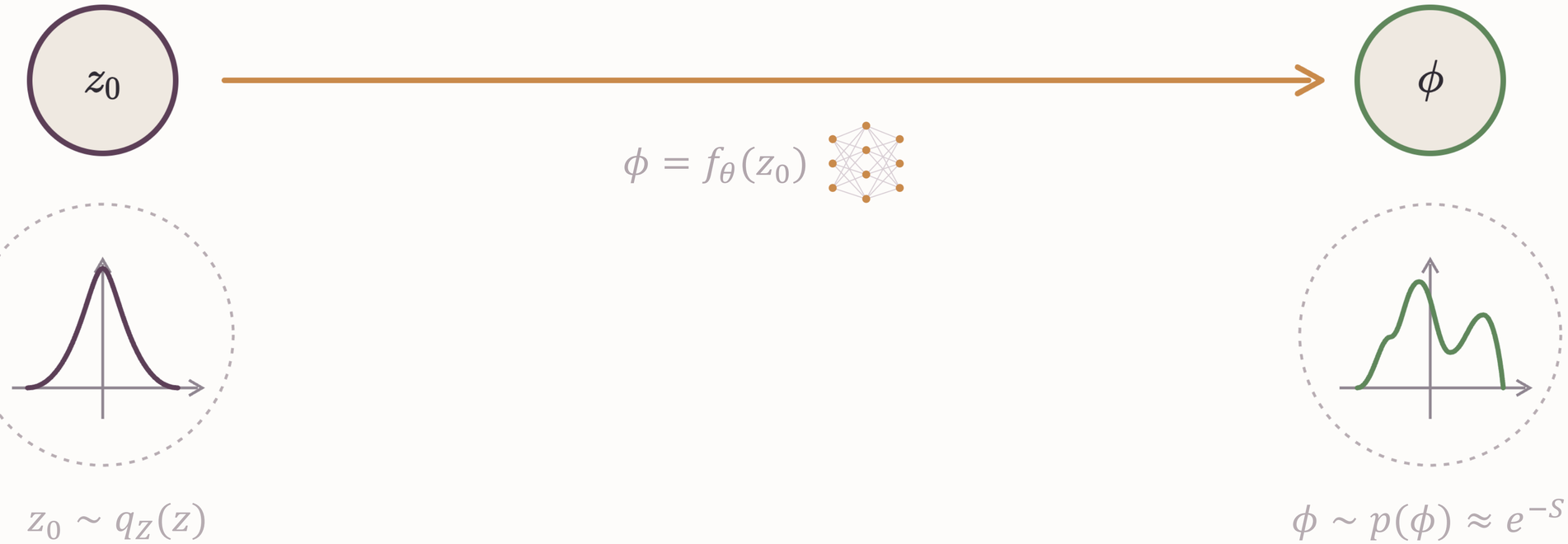
Flow-based generative sampling

Flow-based generative sampling

Simple distribution

Normalizing flow

(Approximate) Path integral distribution



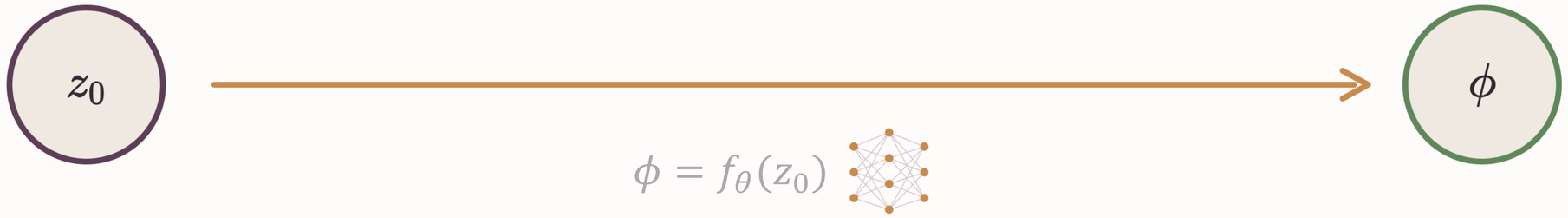
D. Rezende & S. Mohamed, ICML 2016 · M. Albergo et al., PRD 2019

Flow-based generative sampling

Simple distribution

Normalizing flow

(Approximate) Path integral distribution



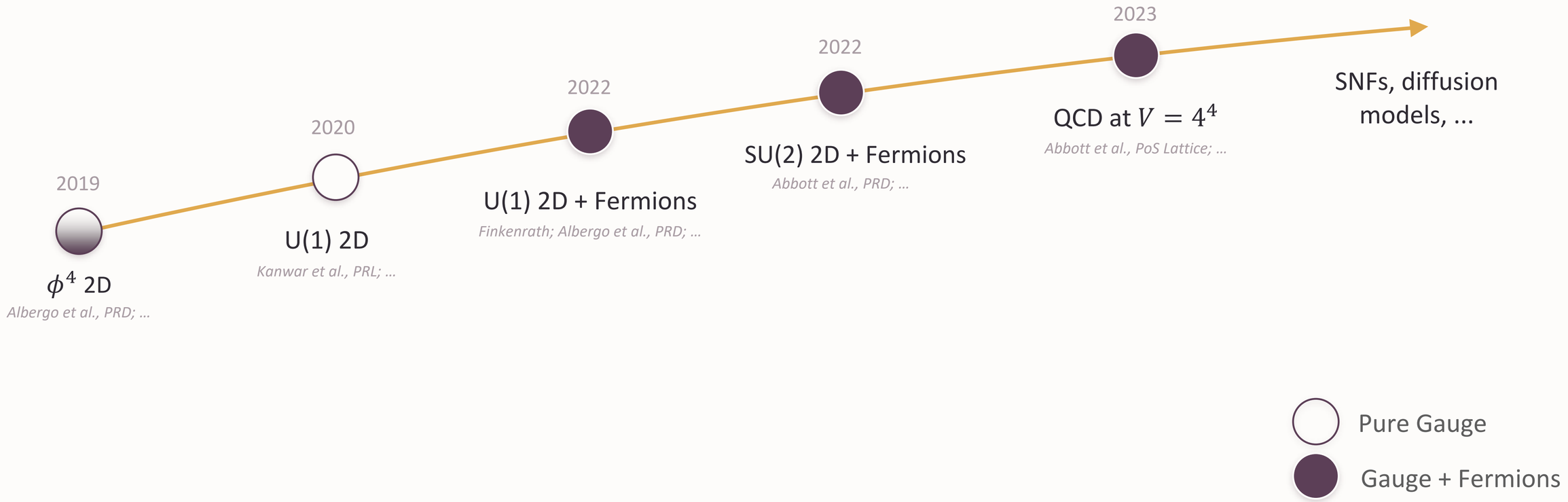
> Invertibility allows for tractable and exact probability estimation, enabling reweighting and and an *asymptotically unbiased* estimator.

- Sampling is i.i.d. and highly parallelizable
- Data free training

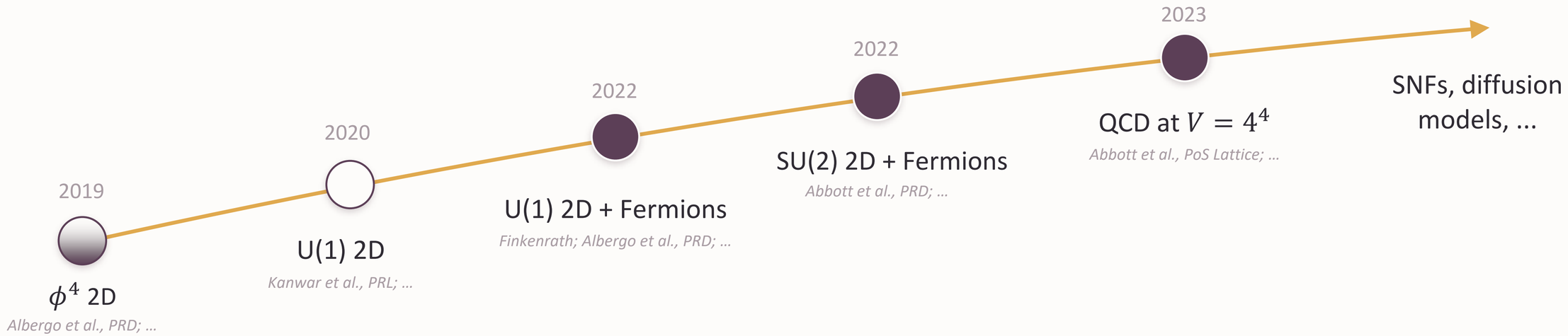
D. Rezende & S. Mohamed, ICML 2016 · M. Albergo et al., PRD 2019 · K. Nicoli et al., PRE 2019

Applications to lattice field theories

Applications to lattice field theories

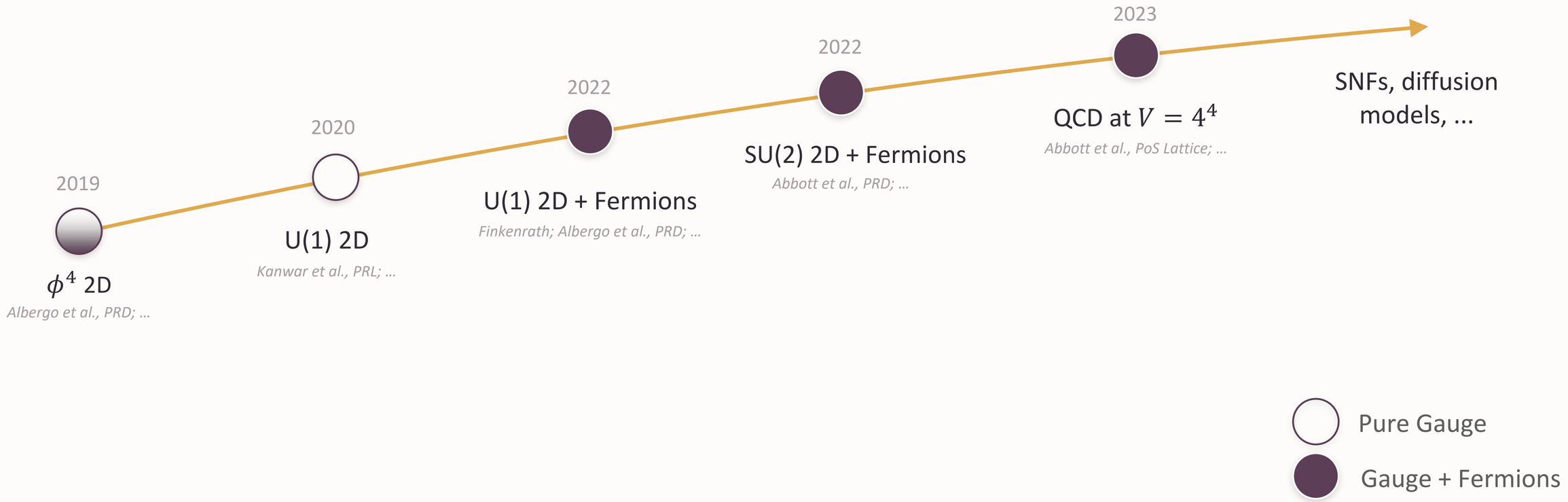


Applications to lattice field theories



Flow-based sampling for full QCD with fermions! Small volumes so far.

Applications to lattice field theories



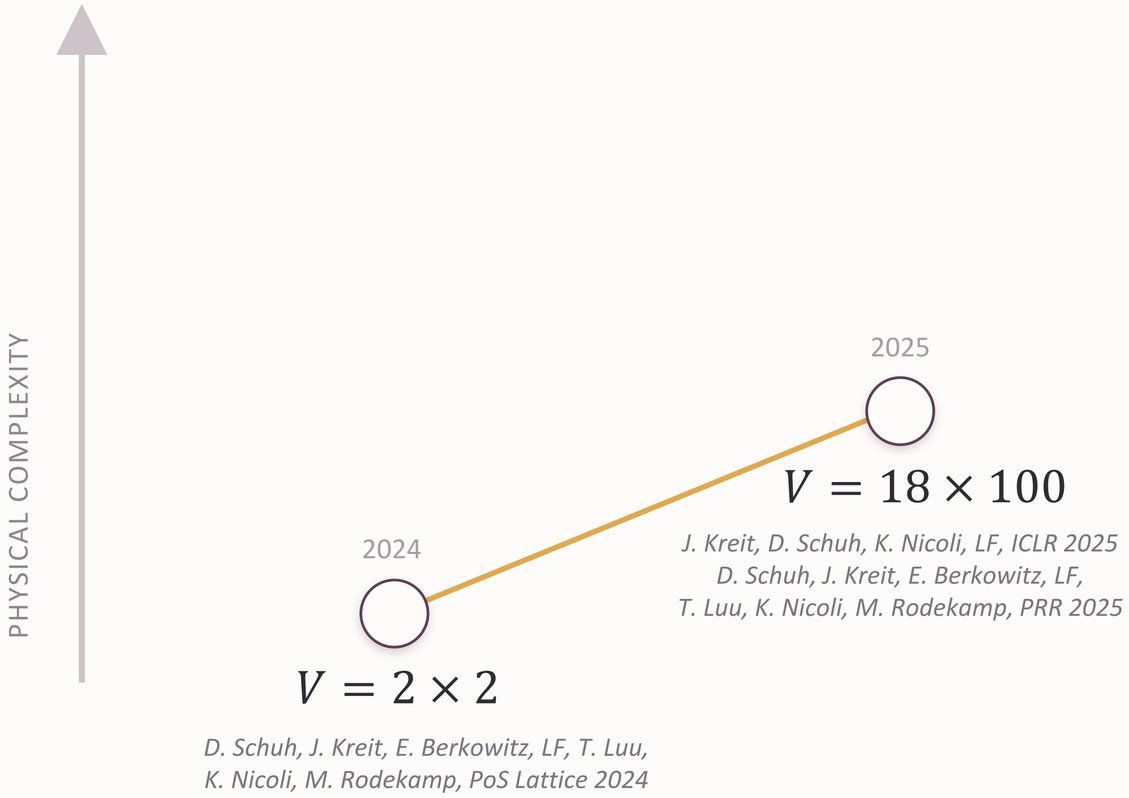
Flow-based sampling for full QCD with fermions! Small volumes so far.

Many condensed-matter systems *are* small: C_{60} has exactly 60 sites.

Normalizing Flows for the Auxiliary-Field Hubbard Model

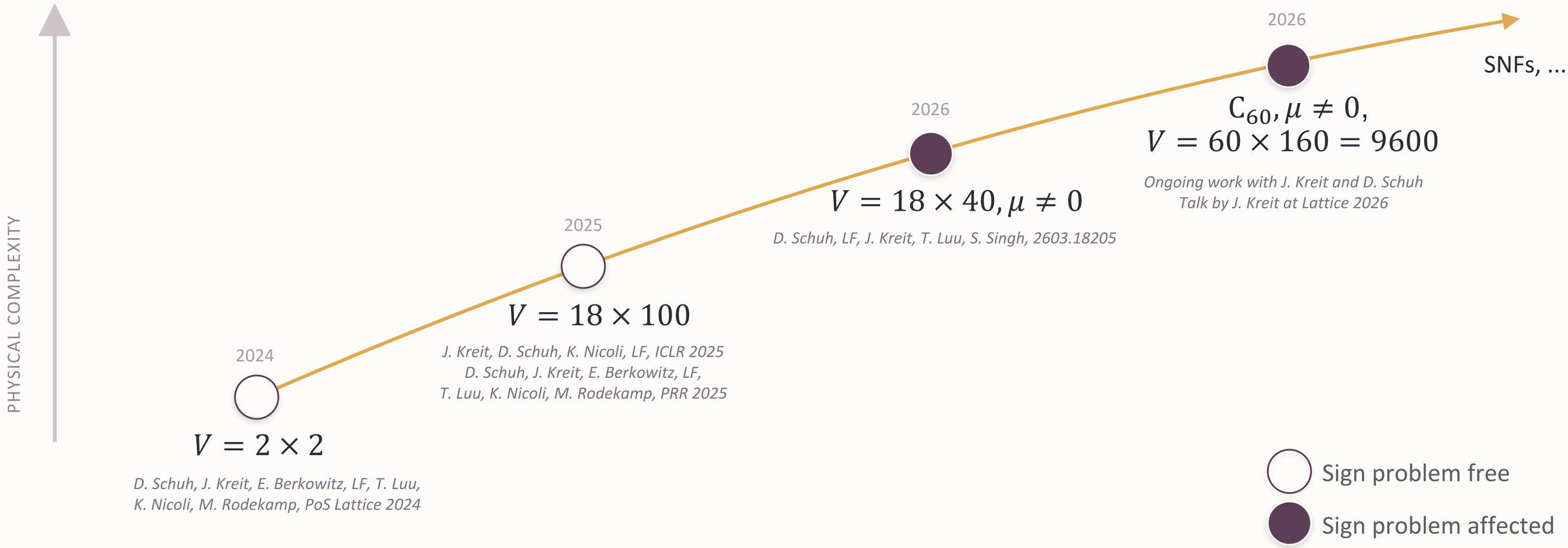
Results

Results

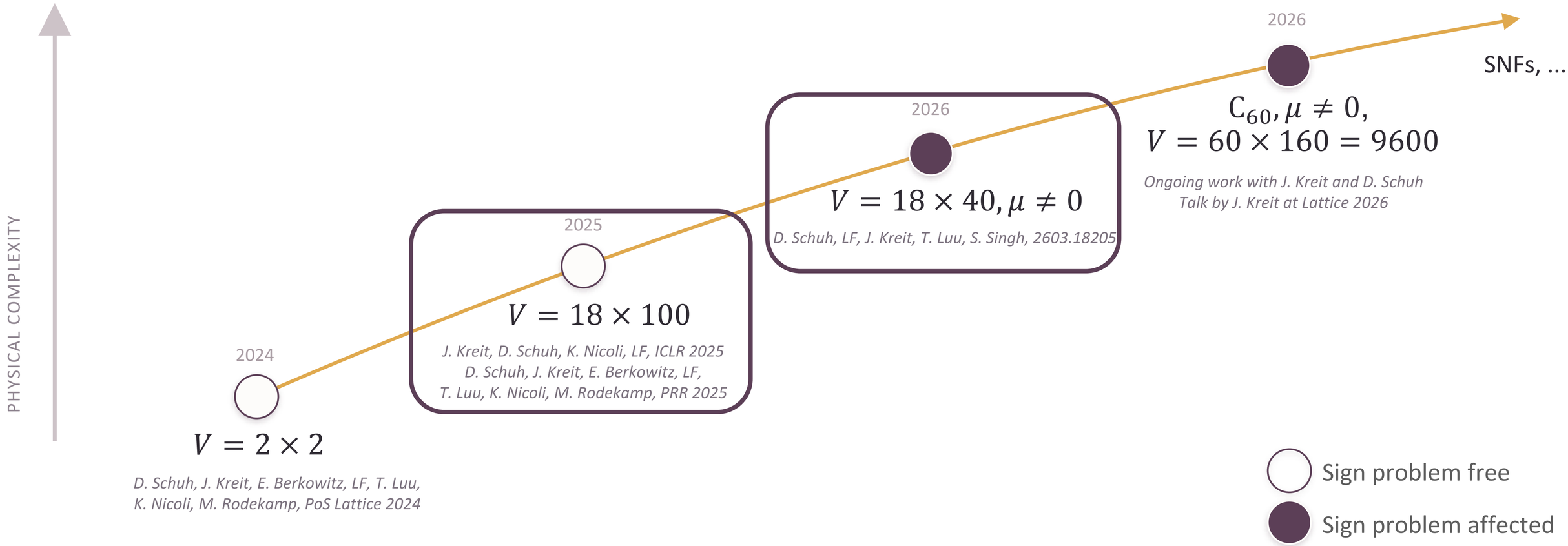


Sign problem free

Results


PHYSICAL COMPLEXITY
2024
$V = 2 \times 2$
D. Schuh, J. Kreit, E. Berkowitz, LF, T. Luu, K. Nicoli, M. Rodekamp, PoS Lattice 2024
2025
$V = 18 \times 100$
J. Kreit, D. Schuh, K. Nicoli, LF, ICLR 2025
D. Schuh, J. Kreit, E. Berkowitz, LF, T. Luu, K. Nicoli, M. Rodekamp, PRR 2025
2026
$V = 18 \times 40, \mu \neq 0$
D. Schuh, LF, J. Kreit, T. Luu, S. Singh, 2603.18205
2026
$C_{60}, \mu \neq 0,$
$V = 60 \times 160 = 9600$
Ongoing work with J. Kreit and D. Schuh
Talk by J. Kreit at Lattice 2026
SNFs, ...
Sign problem free
Sign problem affected

Results


PHYSICAL COMPLEXITY
2024
$V = 2 \times 2$
D. Schuh, J. Kreit, E. Berkowitz, LF, T. Luu, K. Nicoli, M. Rodekamp, PoS Lattice 2024
2025
$V = 18 \times 100$
J. Kreit, D. Schuh, K. Nicoli, LF, ICLR 2025
D. Schuh, J. Kreit, E. Berkowitz, LF, T. Luu, K. Nicoli, M. Rodekamp, PRR 2025
2026
$V = 18 \times 40, \mu \neq 0$
D. Schuh, LF, J. Kreit, T. Luu, S. Singh, 2603.18205
2026
$C_{60}, \mu \neq 0,$
$V = 60 \times 160 = 9600$
Ongoing work with J. Kreit and D. Schuh
Talk by J. Kreit at Lattice 2026
SNFs, ...
Sign problem free
Sign problem affected

RESULTS · SECTION ONE

Simulating Correlated Electrons with Symmetry-Enforced Normalizing Flows

BASED ON

J. Kreit, D. Schuh, K. Nicoli, LF — ICLR 2025

D. Schuh, J. Kreit, E. Berkowitz, LF, T. Luu, K. Nicoli, M.Rodekamp — PRR 2025

The two-site Hubbard model

The two-site Hubbard model

$$S = \frac{1}{2\widetilde{U}} \sum_{x,t} \phi_{x,t}^2 - \log det M[\phi] - \log det M[-\phi]$$

The two-site Hubbard model

$$S = \frac{1}{2\widetilde{U}} \sum_{x,t} \phi_{x,t}^2 - \log det M[\phi] - \log det M[-\phi]$$

$$U/\kappa \gg 1$$

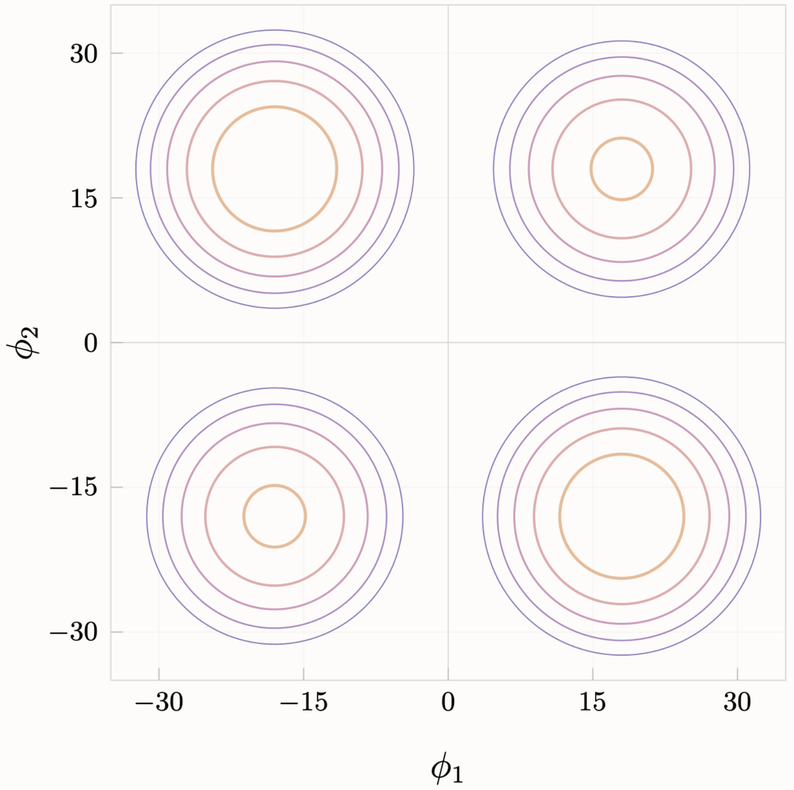
$$p[\phi_1, \phi_2] \sim \left[\cosh\left(\frac{\phi_1 + \phi_2}{2}\right) + \cosh\left(\frac{\phi_1 - \phi_2}{2}\right) \cosh(\tilde{\kappa}) \right]$$

The two-site Hubbard model

$$S = \frac{1}{2\widetilde{U}} \sum_{x,t} \phi_{x,t}^2 - \log det M[\phi] - \log det M[-\phi]$$

$$U/\kappa \gg 1$$

$$p[\phi_1, \phi_2] \sim \left[\cosh\left(\frac{\phi_1 + \phi_2}{2}\right) + \cosh\left(\frac{\phi_1 - \phi_2}{2}\right) \cosh(\tilde{\kappa}) \right]$$

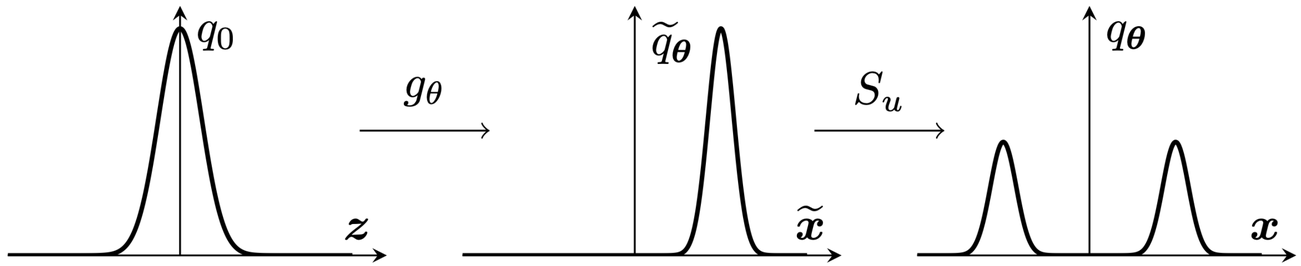


Incorporating symmetries

Incorporating symmetries

THIS WORK: Symmetry Enforcing Stochastic Modulation (SESaMo)

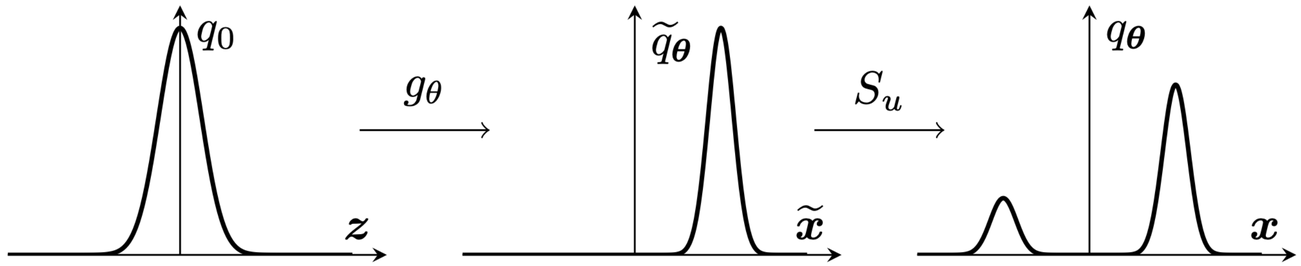
Learn a single mode, then apply a stochastic symmetry operation S_u that replicates it across all symmetric copies with correct weights.



Incorporating symmetries

THIS WORK: Symmetry Enforcing Stochastic Modulation (SESaMo) for broken symmetries

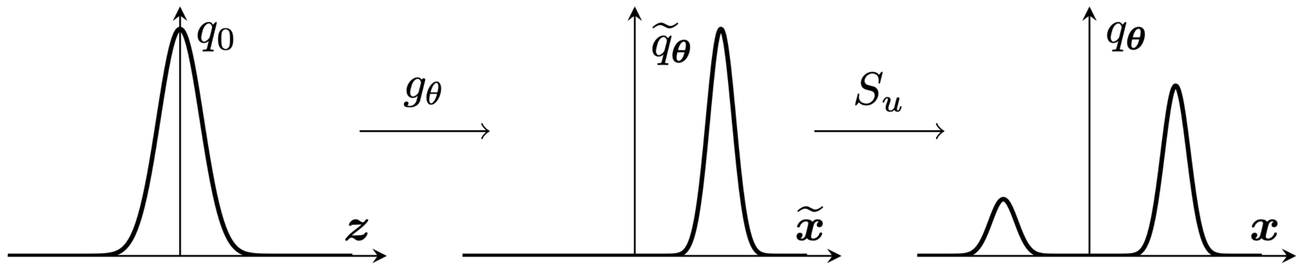
Learn a single mode, then apply a stochastic symmetry operation S_u that replicates it across all symmetric copies with correct weights.



Incorporating symmetries

THIS WORK: Symmetry Enforcing Stochastic Modulation (SESaMo) for broken symmetries

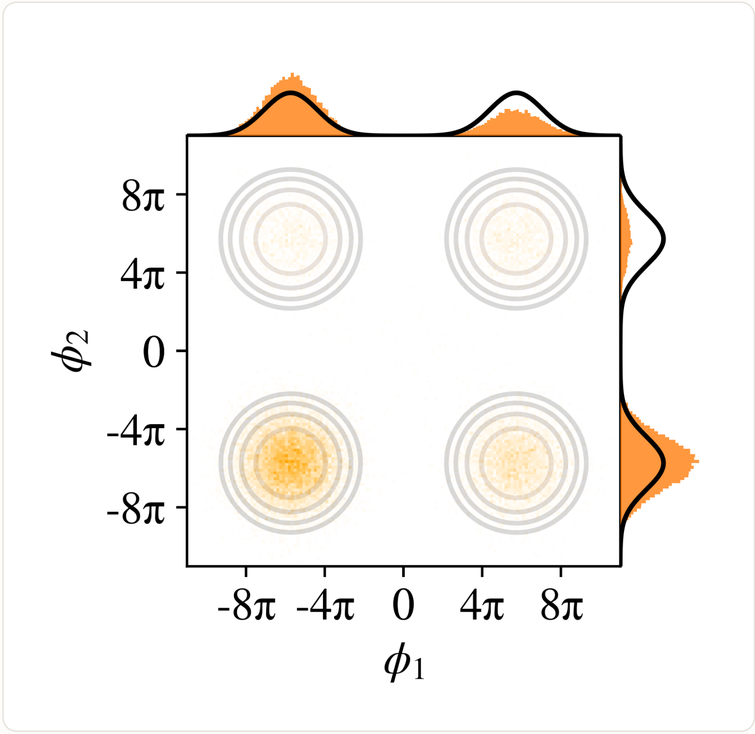
Learn a single mode, then apply a stochastic symmetry operation S_u that replicates it across all symmetric copies with correct weights.



For Hubbard model application: g_θ is a coupling-based NF with RealNVP architecture, trained with reverse KL.

Two-site distribution $U = 18, \beta = 1$

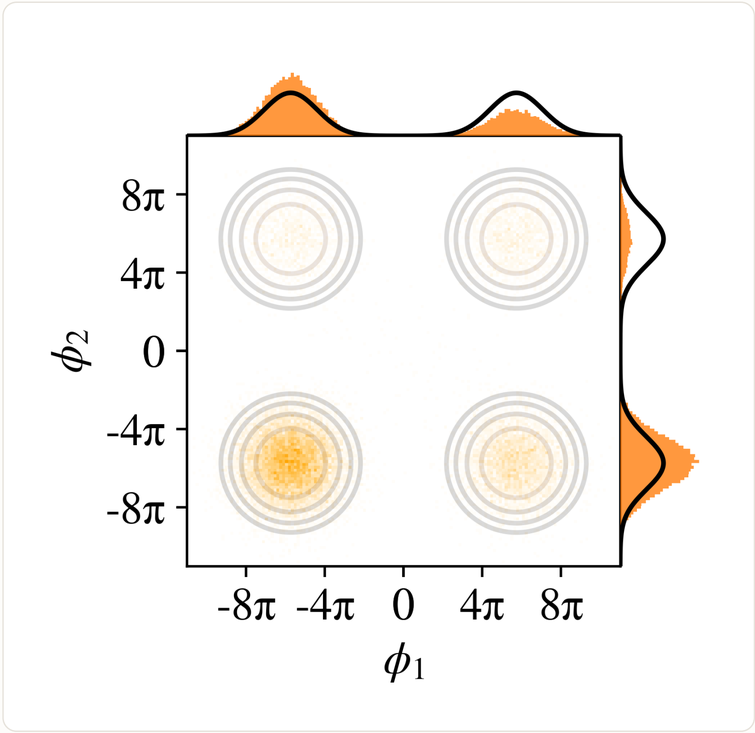
HMC



Trapped, misses modes

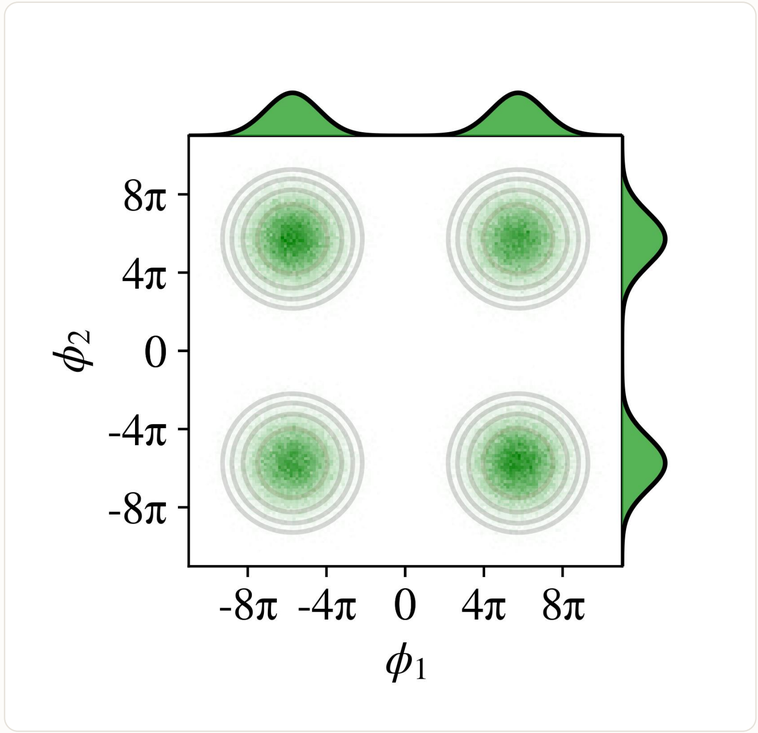
Two-site distribution $U = 18, \beta = 1$

HMC



Trapped, misses modes

Flow with SESaMo



All four modes, correct weights

One-body correlation function

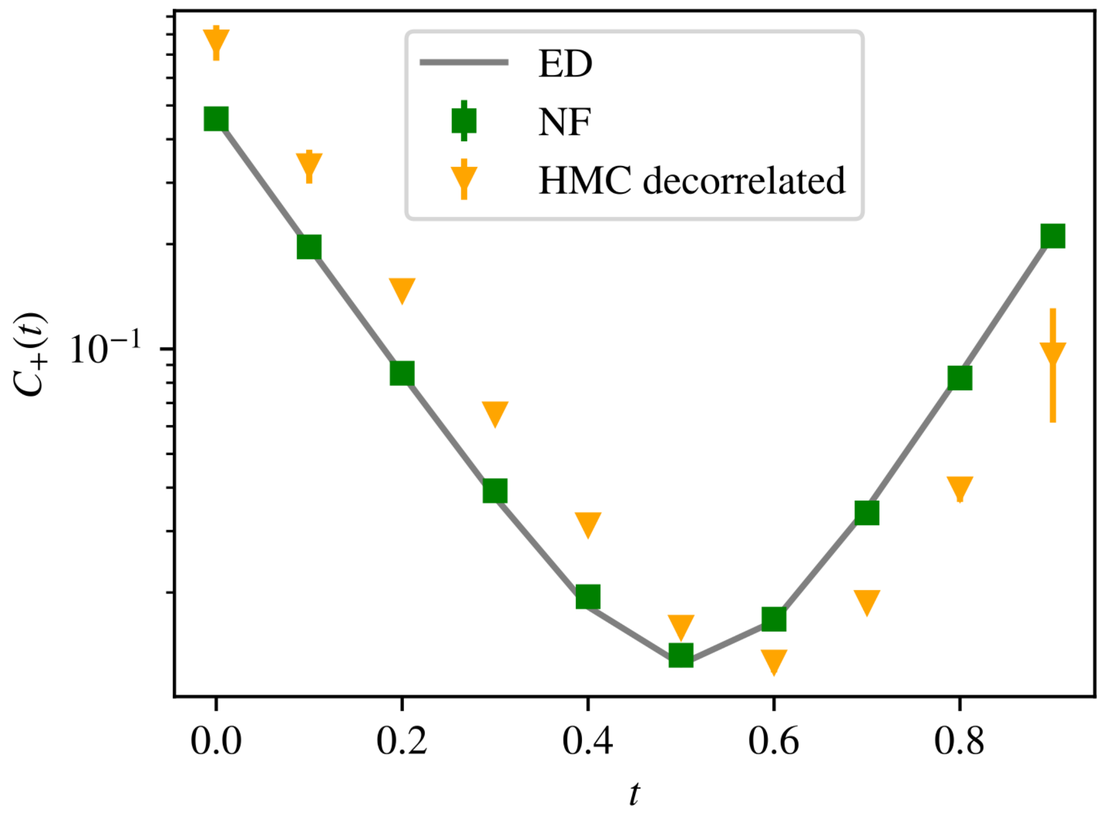
One-body correlation function

$$C_{xy}(t) = \langle a_x(t) a_y^\dagger(0) \rangle$$

One-body correlation function

$$C_{xy}(t) = \langle a_x(t) a_y^\dagger(0) \rangle = \langle M^{-1}_{(x,t);(y,0)}[\phi] \rangle$$

One-body correlation function



NORMALIZING FLOW

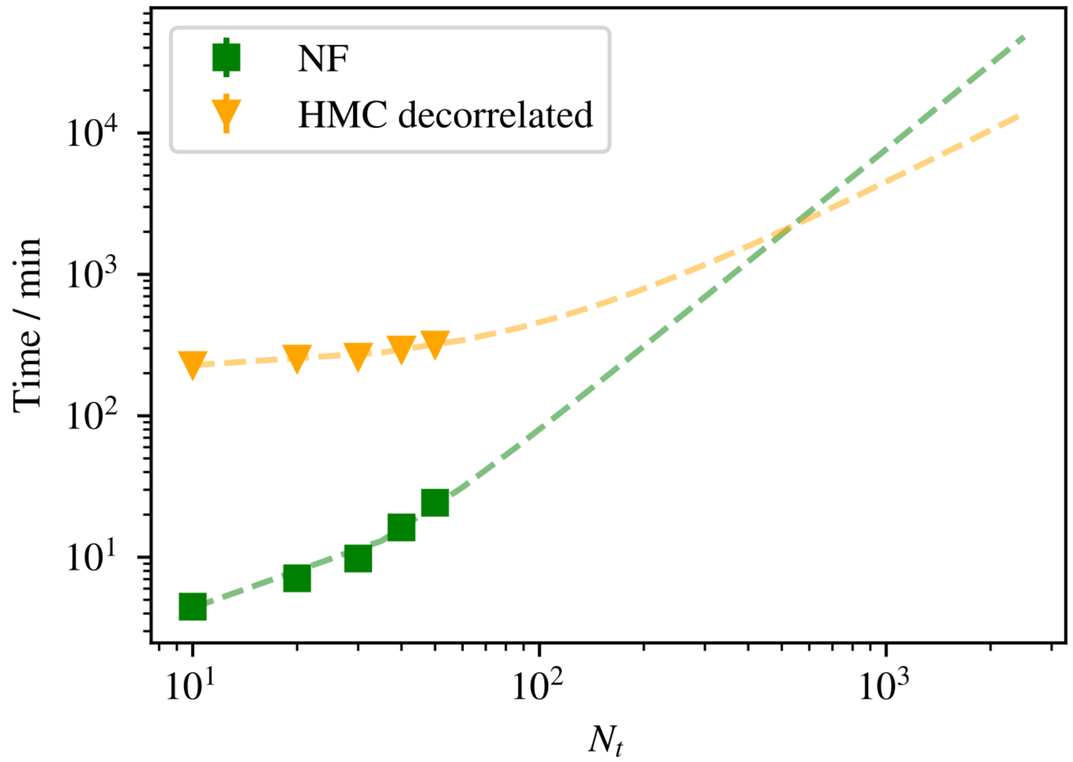
Correctly reproduces the exact (ED) ground truth.

HMC, DECORRELATED

Broken ergodicity leads to the wrong result.

Scaling

Scaling



NORMALIZING FLOW

Scaling proportional to N_t^2

HMC, DECORRELATED

Scaling proportional to $N_t^{5/4}$

Shortcomings

Shortcomings

> The $(Z_2)^{N_x}$ symmetry is only present at large U. At intermediate couplings, the performance degrades.

Shortcomings

The $(Z_2)^{N_x}$ symmetry is only present at large U. At intermediate couplings, the performance degrades.

Number of breaking parameters grows exponentially with the spatial size of the system.

Shortcomings

The $(Z_2)^{N_x}$ symmetry is only present at large U. At intermediate couplings, the performance degrades.

Number of breaking parameters grows exponentially with the spatial size of the system.

We need a more scalable way of overcoming ergodicity issues at intermediate couplings.

RESULTS · SECTION TWO

Tackling the Sign Problem in the Doped Hubbard Model with Normalizing Flows

BASED ON

D. Schuh, LF, J. Kreit, T. Luu, S. Singh, arXiv:2603.18205

The sign problem

The sign problem

Monte Carlo treats each configuration's weight as a probability, so it has to be real and positive.

> At finite density the weights can turn negative and complex.

The sign problem

Monte Carlo treats each configuration's weight as a probability, so it has to be real and positive.

At finite density the weights can turn negative and complex.

→

Reweighting: $\langle O \rangle = \frac{\langle O e^{-\mathrm{Im}[S]} \rangle_{\mathrm{Re}[S]}}{\langle e^{-\mathrm{Im}[S]} \rangle_{\mathrm{Re}[S]}}$

The sign problem

Monte Carlo treats each configuration's weight as a probability, so it has to be real and positive.

At finite density the weights can turn negative and complex.

→

Reweighting: $\langle O \rangle = \frac{\langle O e^{-\mathrm{Im}[S]} \rangle_{\mathrm{Re}[S]}}{\langle e^{-\mathrm{Im}[S]} \rangle_{\mathrm{Re}[S]}}$ Σ

The sign problem

Monte Carlo treats each configuration's weight as a probability, so it has to be real and positive.

At finite density the weights can turn negative and complex.

→

Reweighting: $\langle O \rangle = \frac{\langle O e^{-\mathrm{Im}[S]} \rangle_{\mathrm{Re}[S]}}{\langle e^{-\mathrm{Im}[S]} \rangle_{\mathrm{Re}[S]}} \Sigma$

SEVERITY

$$N_{\mathrm{eff}} = |\Sigma|^2 N \qquad \Sigma \propto e^{-\beta V}$$

The sign problem in the Hubbard model

The sign problem in the Hubbard model

PROBLEM

Finite chemical potential μ breaks the charge symmetry that kept the probability weight positive and real.

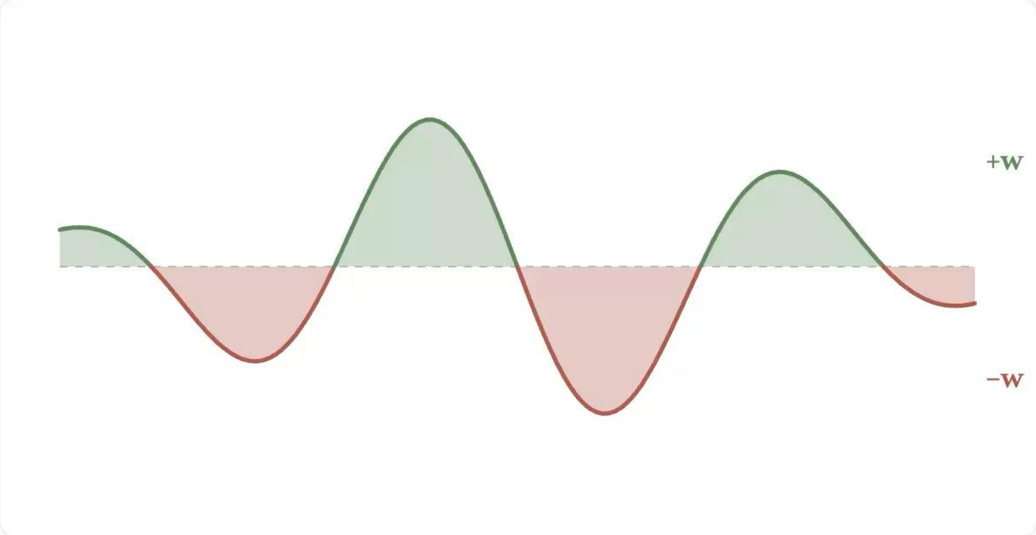
The sign problem in the Hubbard model

PROBLEM

Finite chemical potential μ breaks the charge symmetry that kept the probability weight positive and real.

SOLUTION

We sample the sign-quenched theory $|w|$, then restore the sign by reweighting.



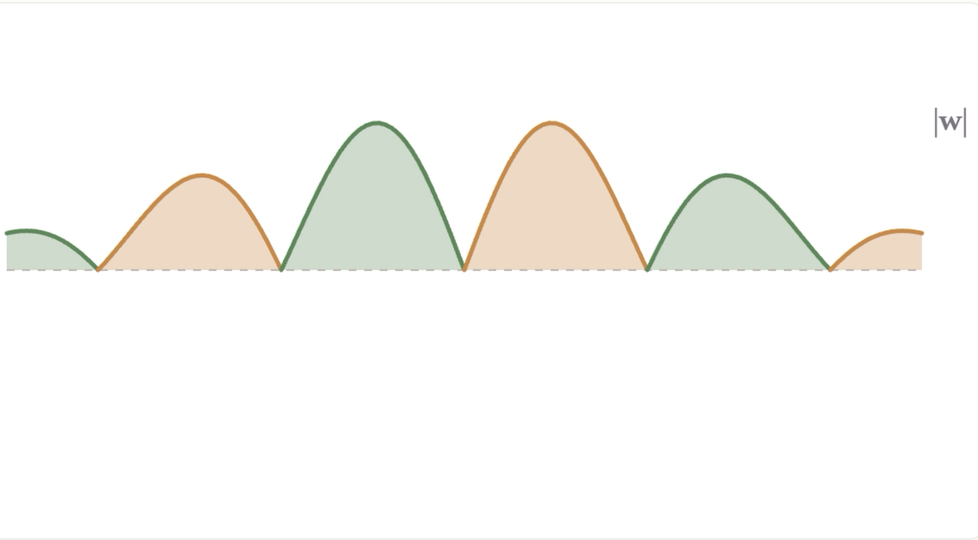
The sign problem in the Hubbard model

PROBLEM

Finite chemical potential μ breaks the charge symmetry that kept the probability weight positive and real.

SOLUTION

We sample the sign-quenched theory $|w|$, then restore the sign by reweighting.



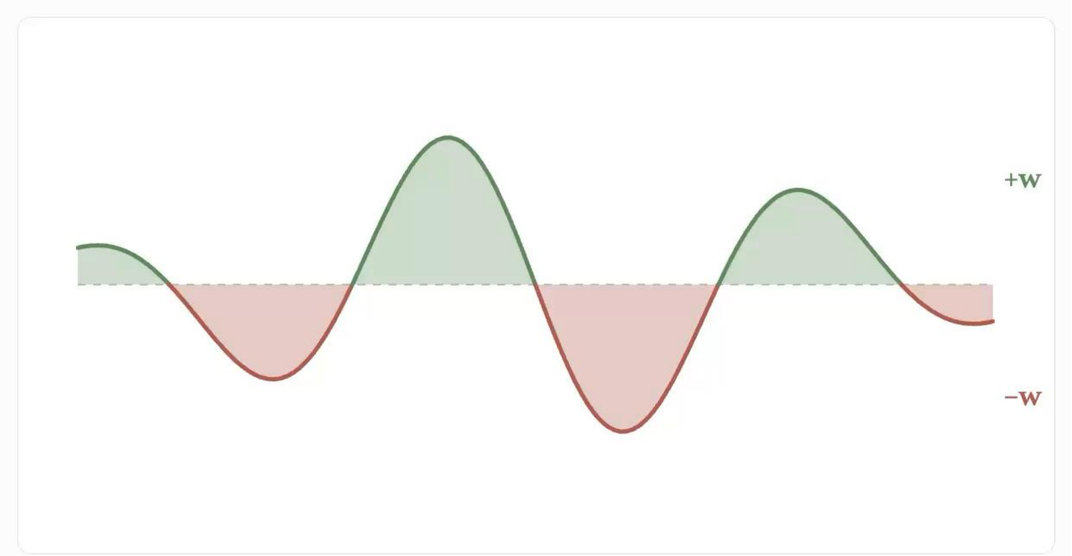
The sign problem in the Hubbard model

PROBLEM

Finite chemical potential μ breaks the charge symmetry that kept the probability weight positive and real.

SOLUTION

We sample the sign-quenched theory $|w|$, then restore the sign by reweighting.



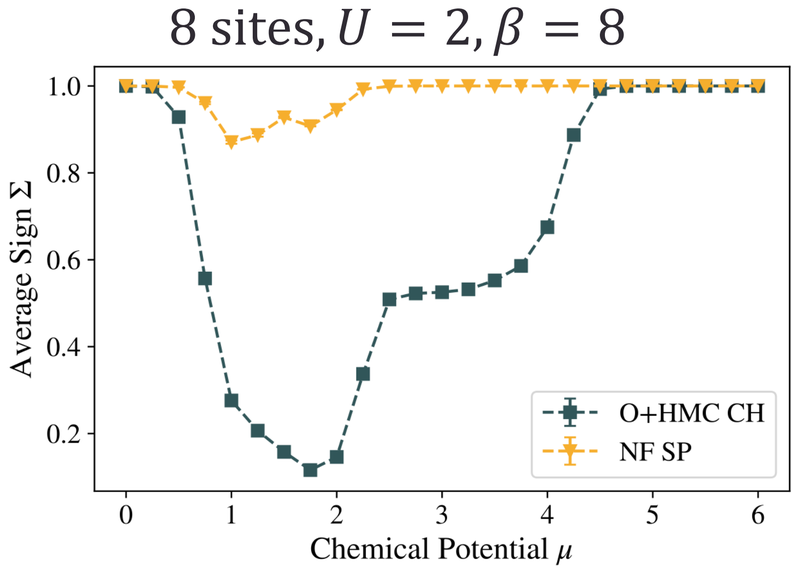
The sign problem in the Hubbard model

PROBLEM

Finite chemical potential μ breaks the charge symmetry that kept the probability weight positive and real.

SOLUTION

We sample the sign-quenched theory $|w|$, then restore the sign by reweighting.



Schuh et al. 2026 · HMC from Gäntgen et al. PRB 2024

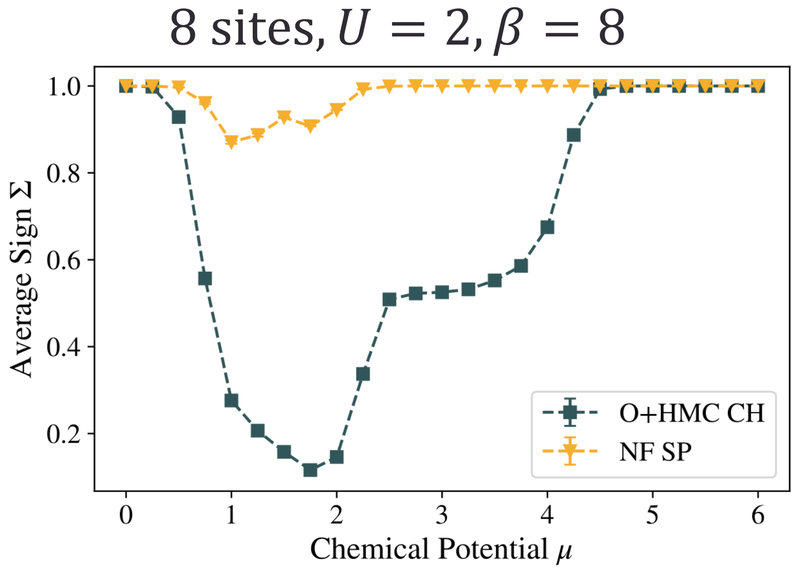
The sign problem in the Hubbard model

PROBLEM

Finite chemical potential μ breaks the charge symmetry that kept the probability weight positive and real.

SOLUTION

We sample the sign-quenched theory $|w|$, then restore the sign by reweighting.



BASIS TRADEOFF

HMC focuses on charge basis, due to ergodicity issues in spin basis. But sign problem is milder in spin basis!

Schuh et al. 2026 · HMC from Gäntgen et al. PRB 2024

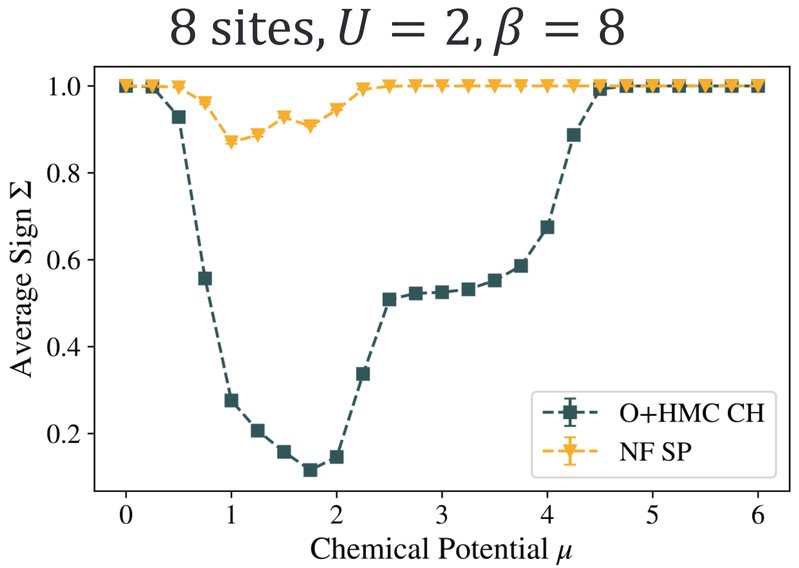
The sign problem in the Hubbard model

PROBLEM

Finite chemical potential μ breaks the charge symmetry that kept the probability weight positive and real.

SOLUTION

We sample the sign-quenched theory $|w|$, then restore the sign by reweighting.



BASIS TRADEOFF

HMC focuses on charge basis, due to ergodicity issues in spin basis. But sign problem is milder in spin basis!

GOAL

Using NFs, find a way to *ergodically* simulate in spin basis at intermediate couplings and finite density.

Schuh et al. 2026 · HMC from Gäntgen et al. PRB 2024

Annealing to the target distribution

Annealing to the target distribution

Goal: get training signal from every mode, without deterministic symmetry transformations.

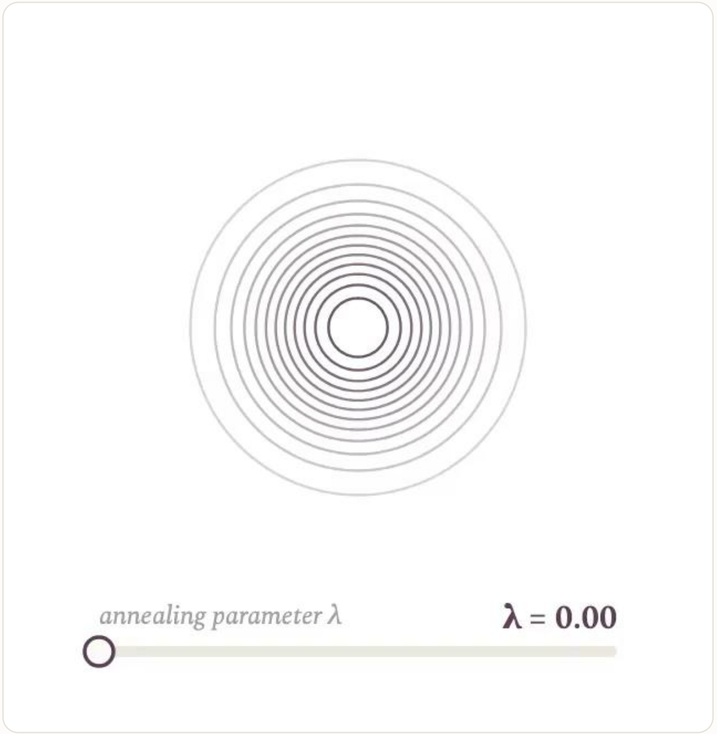
$$S_q[\phi] = \sum_{x,t\in\Lambda} \frac{\phi_{x,t}^2}{2\widetilde{U}} - \log|det(M[\phi]M[-\phi])|$$

Annealing to the target distribution

Goal: get training signal from every mode, without deterministic symmetry transformations.

$$S_{q,\lambda}[\phi] = \sum_{x,t\in\Lambda} \frac{\phi_{x,t}^2}{2\widetilde{U}} - \lambda \cdot \log|det(M[\phi]M[-\phi])|$$

Smoothly introduce multimodality by slowly ramping determinant contributions from $\lambda = 0$ to $\lambda = 1$ at the beginning of training.

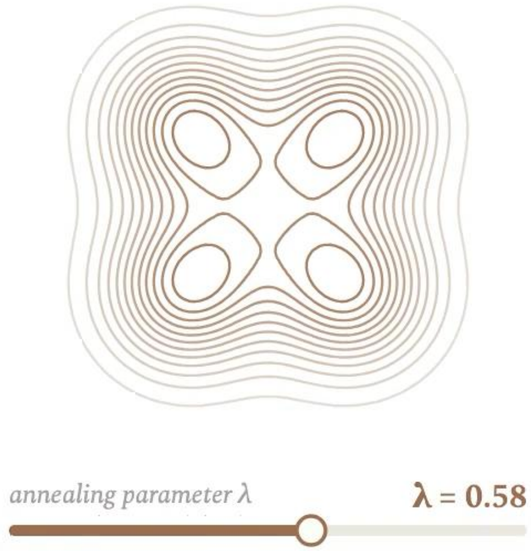


Annealing to the target distribution

Goal: get training signal from every mode, without deterministic symmetry transformations.

$$S_{q,\lambda}[\phi] = \sum_{x,t\in\Lambda} \frac{\phi_{x,t}^2}{2\widetilde{U}} - \lambda \cdot \log|det(M[\phi]M[-\phi])|$$

Smoothly introduce multimodality by slowly ramping determinant contributions from $\lambda = 0$ to $\lambda = 1$ at the beginning of training.

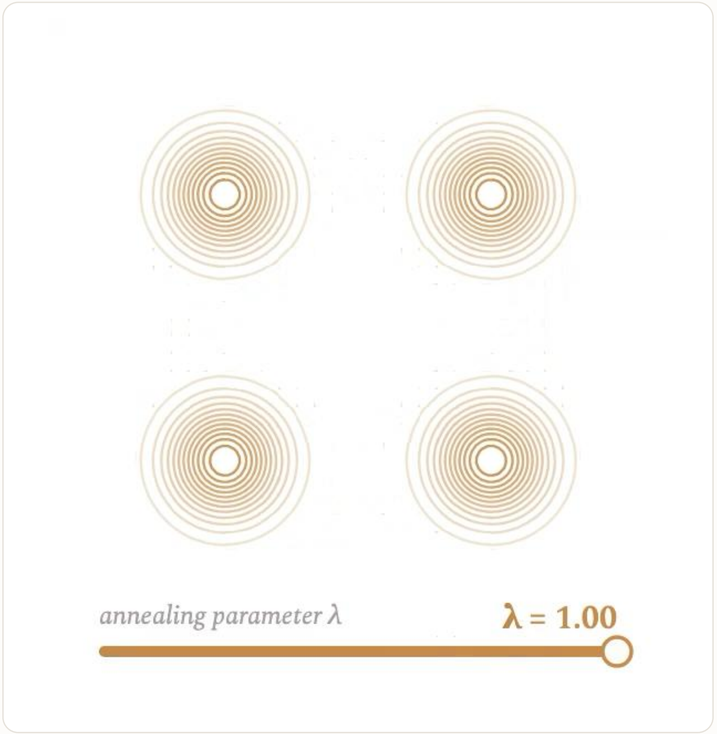


Annealing to the target distribution

Goal: get training signal from every mode, without deterministic symmetry transformations.

$$S_{q,\lambda}[\phi] = \sum_{x,t\in\Lambda} \frac{\phi_{x,t}^2}{2\widetilde{U}} - \lambda \cdot \log|det(M[\phi]M[-\phi])|$$

Smoothly introduce multimodality by slowly ramping determinant contributions from $\lambda = 0$ to $\lambda = 1$ at the beginning of training.

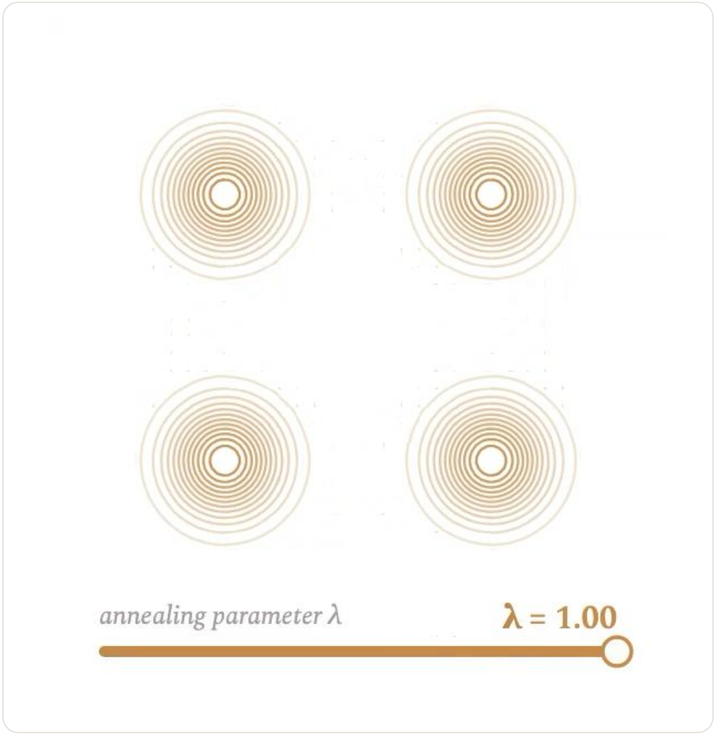


Annealing to the target distribution

Goal: get training signal from every mode, without deterministic symmetry transformations.

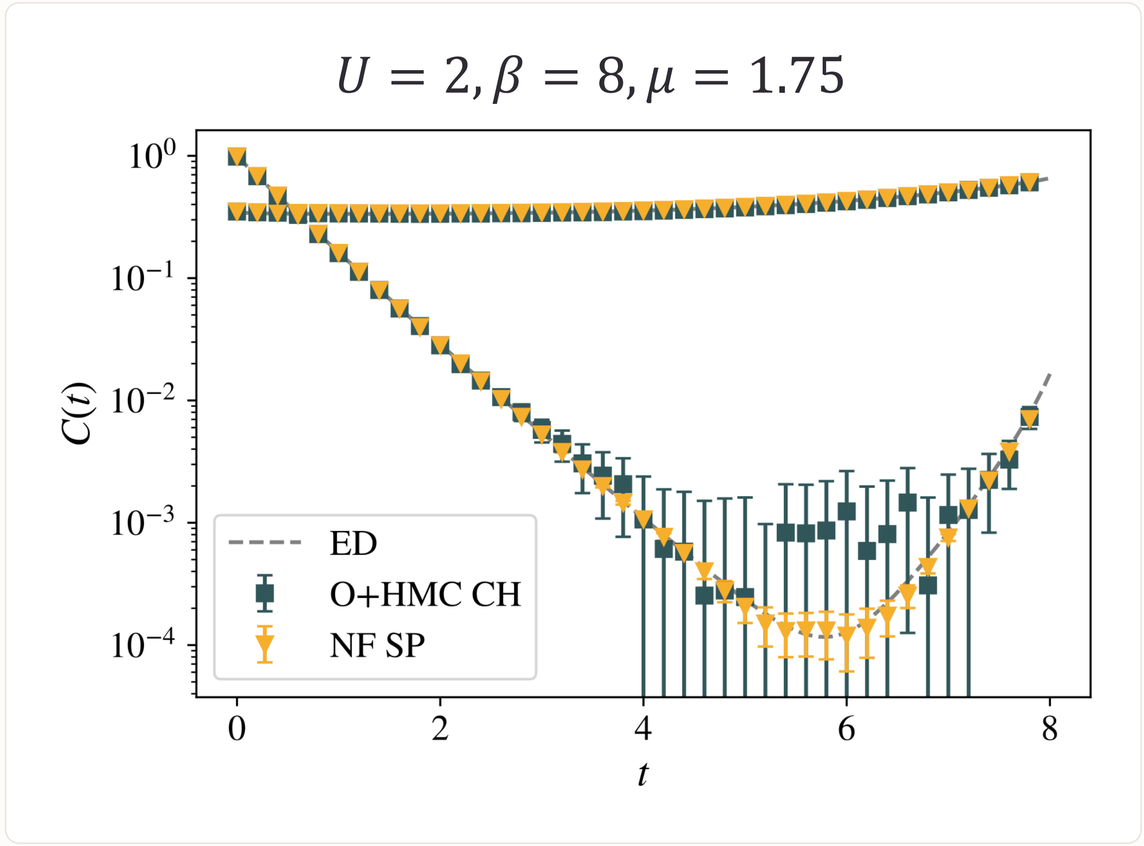
$$S_{q,\lambda}[\phi] = \sum_{x,t\in\Lambda} \frac{\phi_{x,t}^2}{2\widetilde{U}} - \lambda \cdot \log|det(M[\phi]M[-\phi])|$$

Smoothly introduce multimodality by slowly ramping determinant contributions from $\lambda = 0$ to $\lambda = 1$ at the beginning of training.



> Signal from *every mode*, independent of volume and parameter regime.

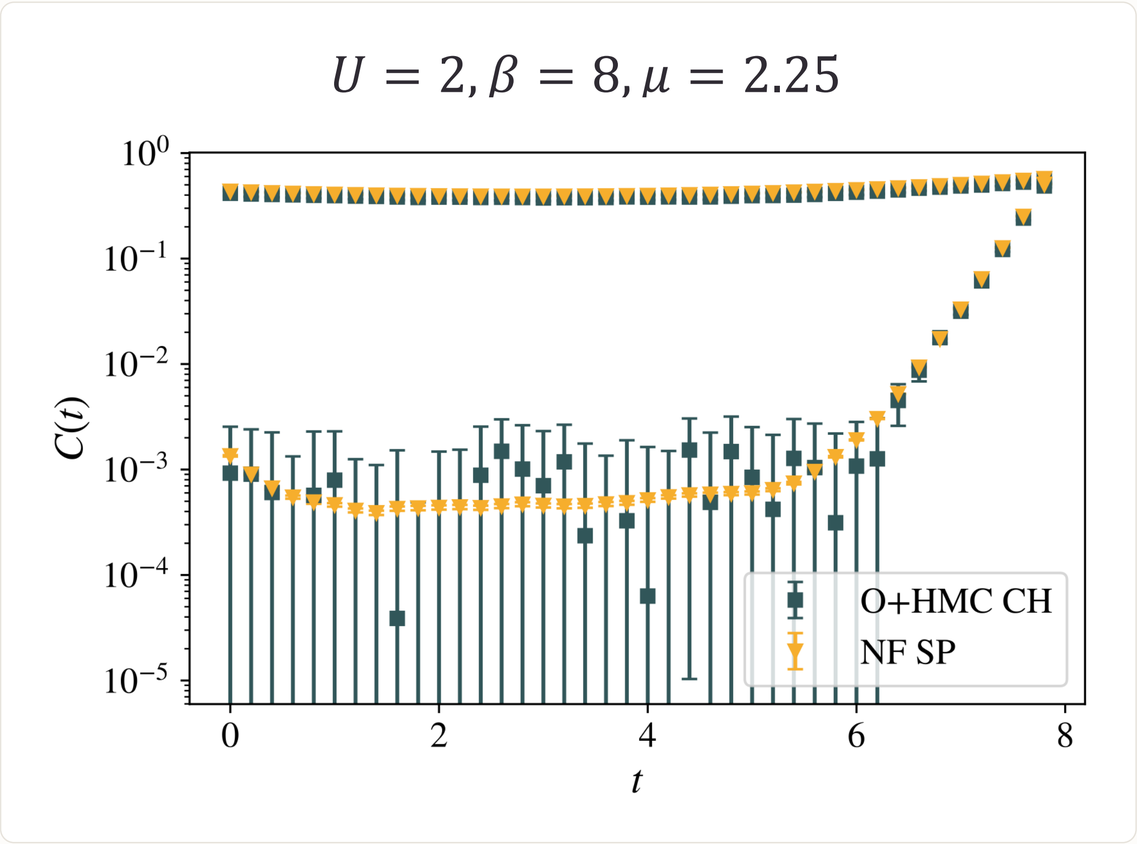
8-sites one-body correlator



$C(t)$ vs. Euclidean time — orange ▽: NF (spin basis); dark ■: HMC with optimized shifts (charge basis)

Schuh et al. 2026 · HMC from Gäntgen et al. PRB 2024

18-sites one-body correlator



$C(t)$ vs. Euclidean time — orange ▽: NF (spin basis); dark ■: HMC with optimized shifts (charge basis)

Schuh et al. 2026 · HMC from Gäntgen et al. PRB 2024

Open questions

Open questions

> Annealing is empirically robust, but we have no proof that ergodicity is always restored.

Open questions

Annealing is empirically robust, but we have no proof that ergodicity is always restored.

Expressivity and scaling are inherited from the flow model: same bottleneck as in the ML-for-LQCD community.

Open questions

Annealing is empirically robust, but we have no proof that ergodicity is always restored.

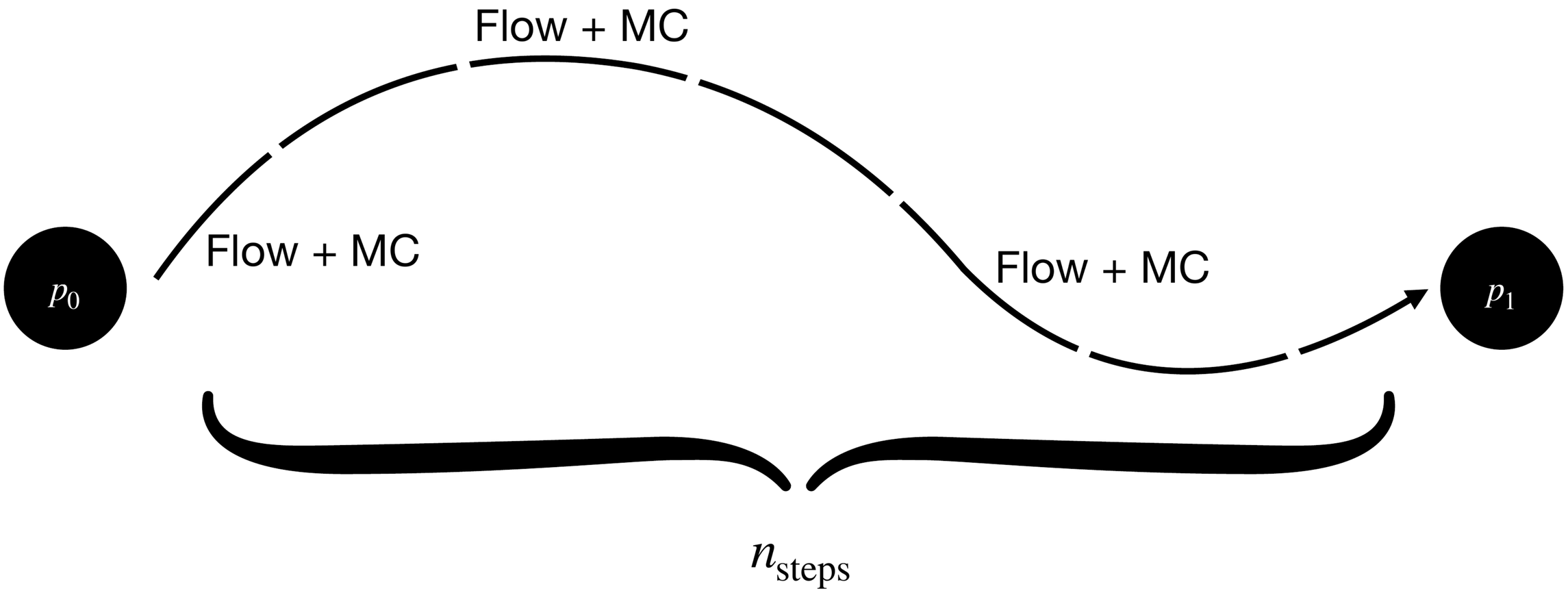
Expressivity and scaling are inherited from the flow model: same bottleneck as in the ML-for-LQCD community.

Similar questions but different target: transfer of LQCD ideas to Hubbard model, use small Hubbard systems as benchmark.

Outlook:
SNFs for the Auxiliary-Field Hubbard Model

Stochastic Normalizing Flows

Ongoing work with: A. Bulgarelli, J. Kreit, J. Ostmeyer, D. Schuh, L. Verzichelli



Slides courtesy of Janik Kreit

Applied to the Hubbard Model

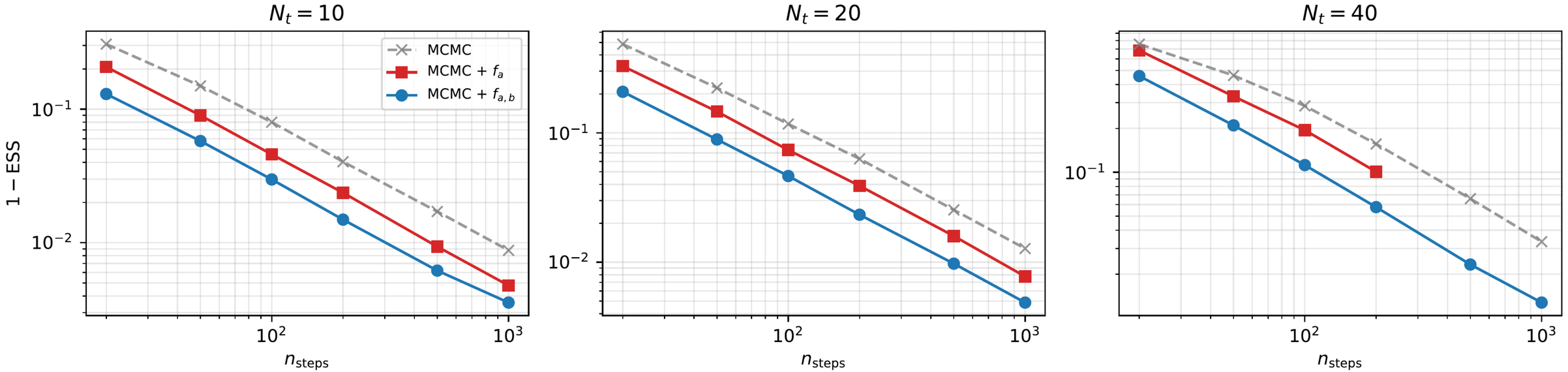
Use **very simple** flows:

$$f_a(\phi) = e^{a \cdot \square} \phi$$

$$f_{a,b}(\phi) = e^{a \cdot \square + b \cdot h} \phi$$

h: Hopping matrix

Training $a, b \in \mathbb{R}$

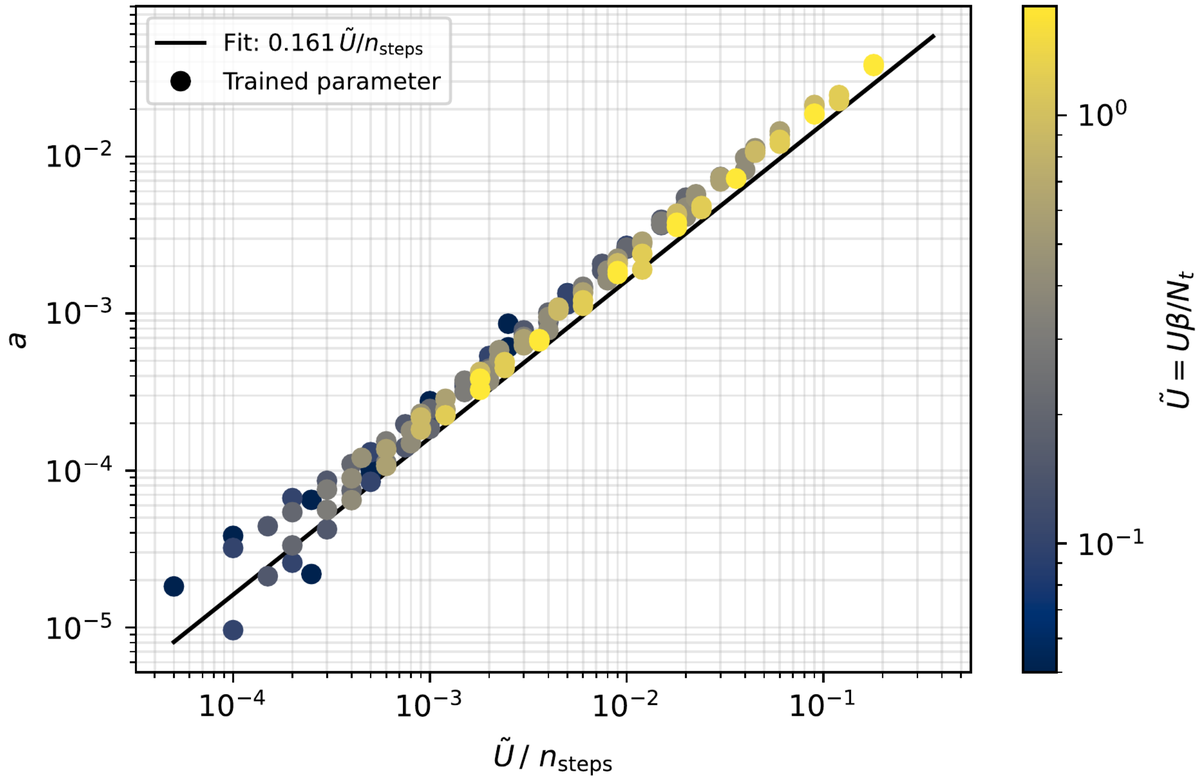


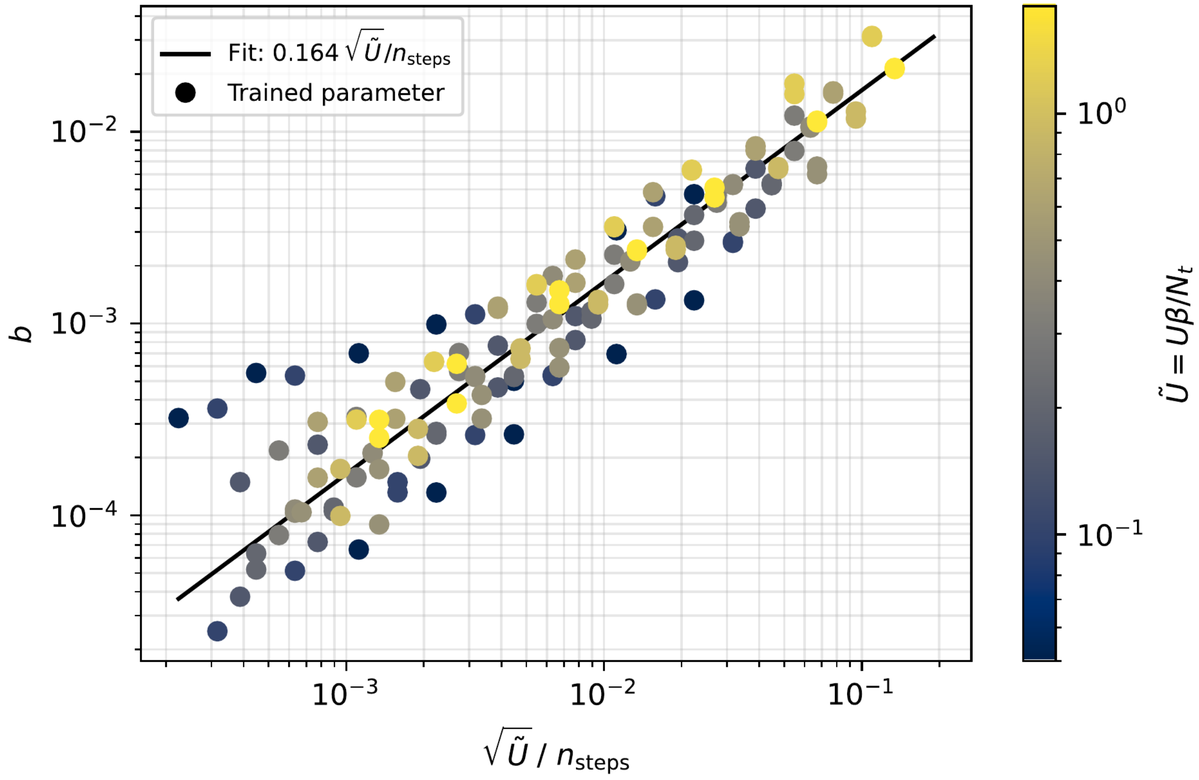
$U = 2, N_x = 4, \beta = N_t/4$

Slides courtesy of Janik Kreit

Determine a, b without Training

at different physical parameters U, β, N_t





Slides courtesy of Janik Kreit

Determine a, b without Training

at different physical parameters U, β, N_t



We train once, then never again

Slides courtesy of Janik Kreit

Conclusion

Conclusion & outlook

For the first time, applied flow-based sampling to condensed-matter system.

Developed a new architecture to incorporate broken symmetries and an annealing scheme.

Simulated ergodically in the spin basis, across a range of couplings and dopings.

Conclusion & outlook

For the first time, applied flow-based sampling to condensed-matter system.

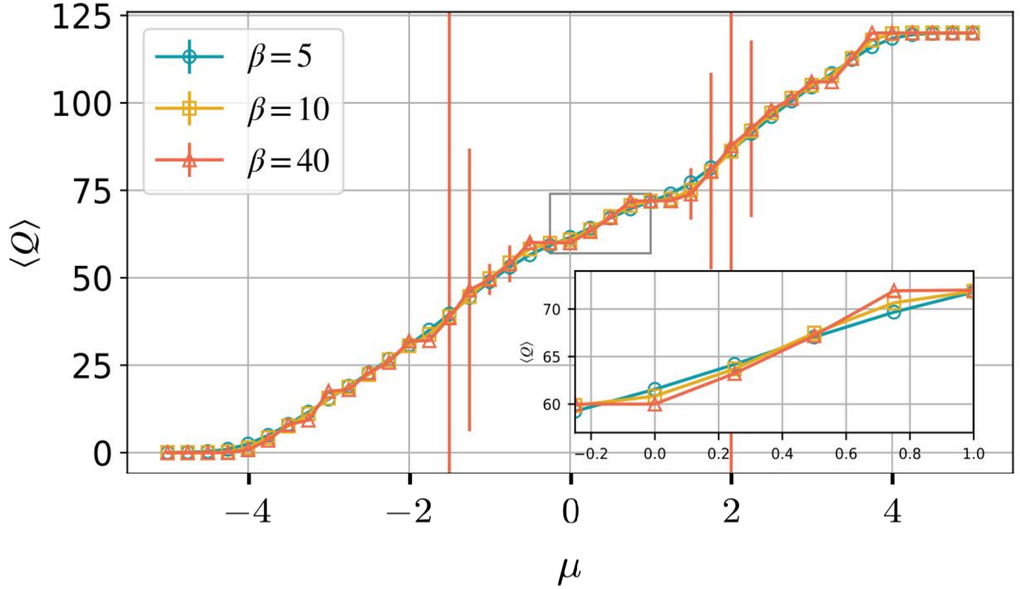
Developed a new architecture to incorporate broken symmetries and an annealing scheme.

Simulated ergodically in the spin basis, across a range of couplings and dopings.

Extension towards larger systems and lower temperatures, using NE-MCMC, SNFs, …

Ongoing work with A. Bulgarelli, J. Kreit, J. Ostmeyer, D. Schuh, L. Verzichelli

Conclusion & outlook



Talk by J. Kreit at Lattice 2026

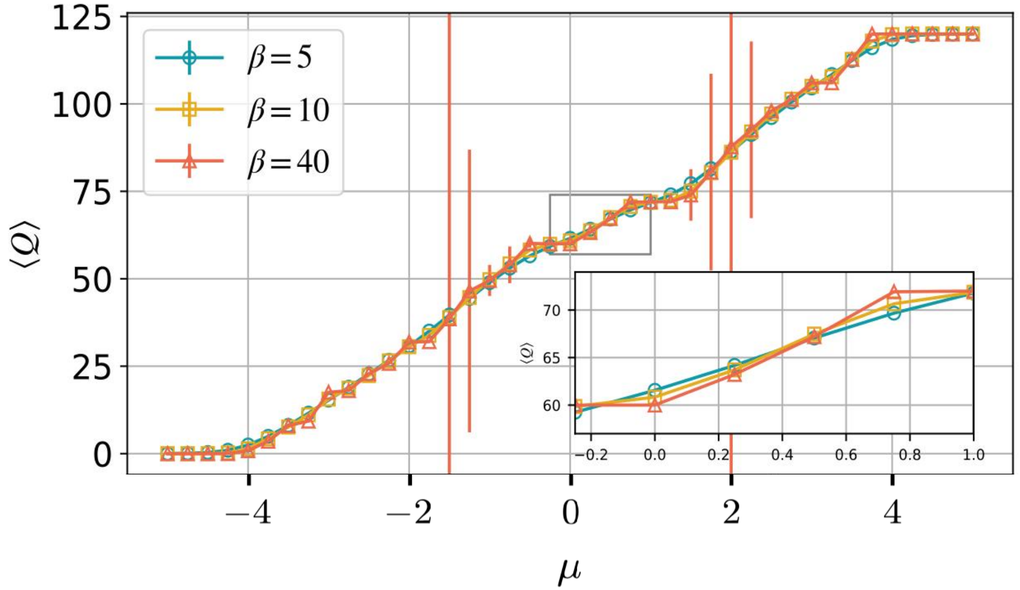
Extension towards larger systems and lower temperatures, using NE-MCMC, SNFs, …

Ongoing work with A. Bulgarelli, J. Kreit, J. Ostmeyer, D. Schuh, L. Verzichelli

Simulations of the discrete auxiliary-field Hubbard model, up to $V = 60 \times 160 = 9600$.

Ongoing work with J. Kreit and D. Schuh

Conclusion & outlook



Talk by J. Kreit at Lattice 2026

Extension towards larger systems and lower temperatures, using NE-MCMC, SNFs, …

Ongoing work with A. Bulgarelli, J. Kreit, J. Ostmeyer, D. Schuh, L. Verzichelli

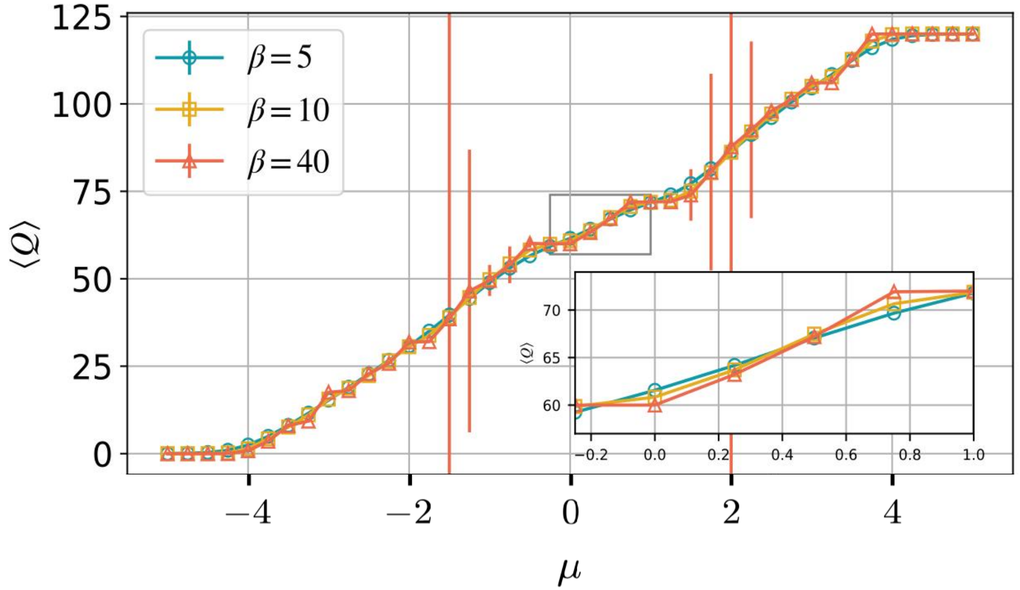
Simulations of the discrete auxiliary-field Hubbard model, up to $V = 60 \times 160 = 9600$.

Ongoing work with J. Kreit and D. Schuh

Density-of-states method → talk by Simran!

Ongoing work with S. Singh, F. Freche, S. Kreit, D. Schuh, C. Schmidt-Sonntag

Conclusion & outlook



Talk by J. Kreit at Lattice 2026

Extension towards larger systems and lower temperatures, using NE-MCMC, SNFs, …

Ongoing work with A. Bulgarelli, J. Kreit, J. Ostmeyer, D. Schuh, L. Verzichelli

Simulations of the discrete auxiliary-field Hubbard model, up to $V = 60 \times 160 = 9600$.

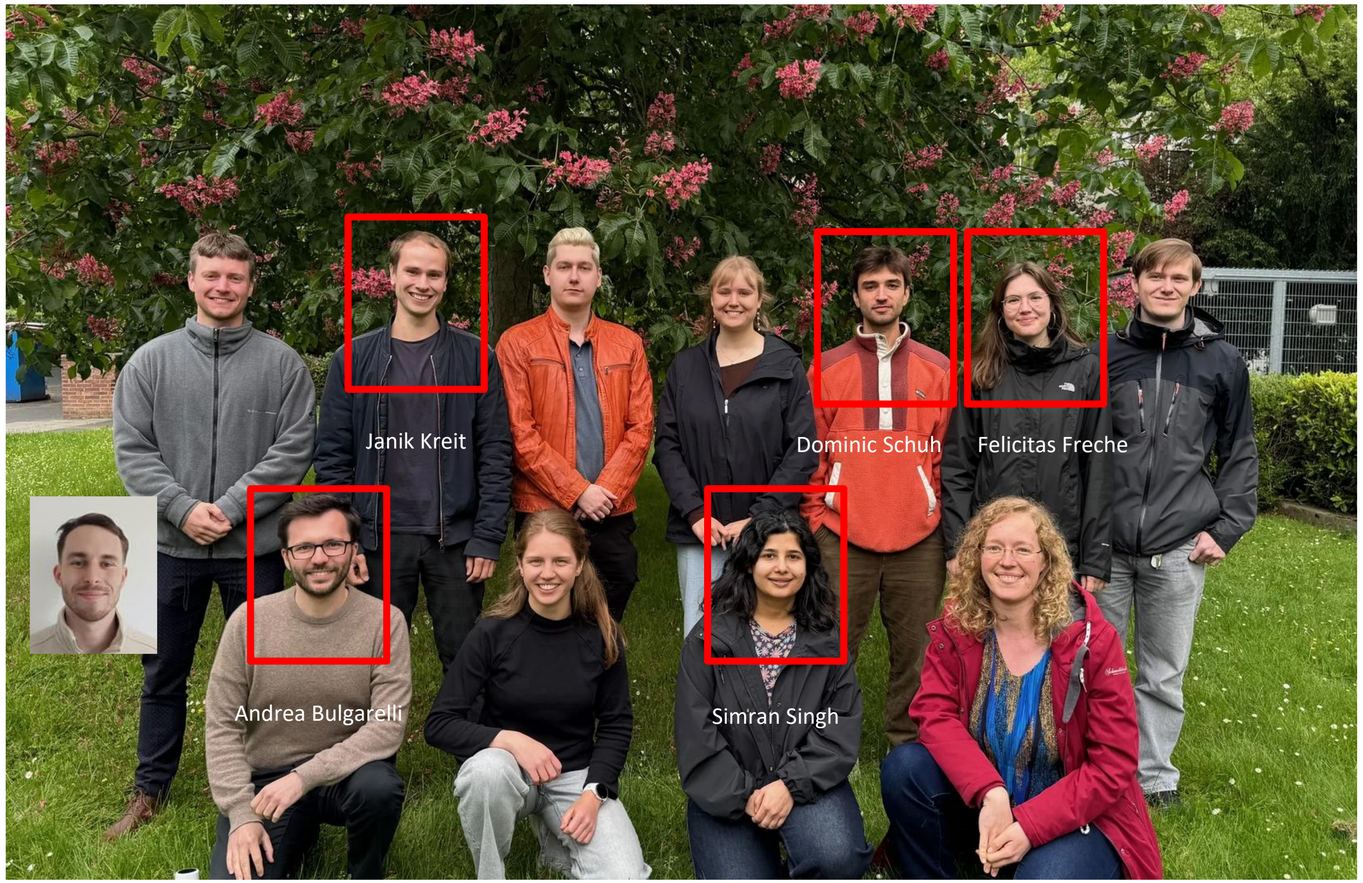
Ongoing work with J. Kreit and D. Schuh

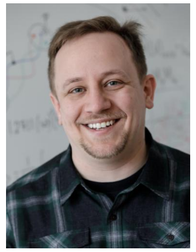
Density-of-states method → talk by Simran!

Ongoing work with S. Singh, F. Freche, S. Kreit, D. Schuh, C. Schmidt-Sonntag

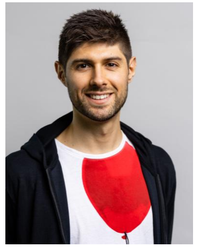
Cuprate superconductors and Hubbard ladders.

Thanks to our group and collaborators!




Evan Berkowitz


Thomas Luu

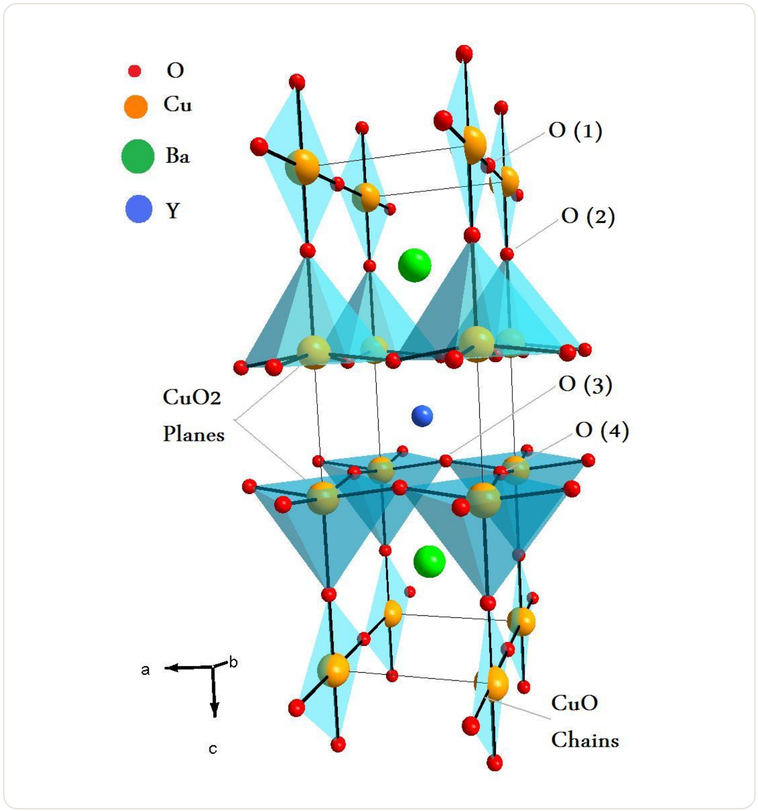

Kim Nicoli


Marcel Rodekamp

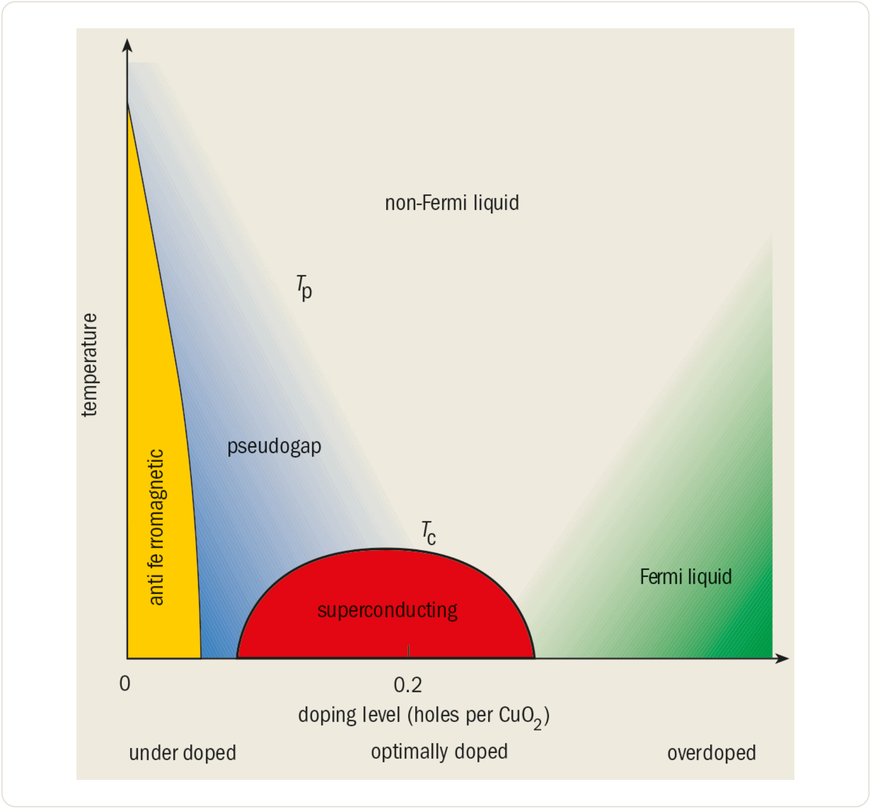
Thanks to you for listening!
Questions?

Backup

Cuprate high-T_C superconductivity



Layered CuO_2 planes (YBCO)



Temperature–doping phase diagram

Auxiliary-field formulation

Linearize the interaction on each time slice with Hubbard–Stratonovich

$$\exp\left[\frac{\delta U}{2}\sum_{x\in\Lambda} q_x^2\right] \;\propto\; \int \mathcal{D}[\phi_t]\; \exp\left[-\frac{1}{2\delta U}\sum_{x}\phi_{x,t}^2 \;+\; \sum_{x}\phi_{x,t}\,q_x\right]$$

Fermions enter bilinearly → Grassmann integral done exactly

$$Z = \int \mathcal{D}[\phi]\, e^{-S[\phi]}\,, \qquad S[\phi] = \frac{1}{2\tilde{U}}\sum_{x,t}\phi_{x,t}^2 \;-\; \log\det M[\phi] \;-\; \log\det M[-\phi]$$

Exponential discretization of the fermion matrix

$$M^{e}_{x't',\,xt}[\phi] \;=\; \delta_{x,x'}\,\delta_{t,t'} \;-\; \left[e^{\delta K}\right]_{x,x'} e^{\phi_{xt}}\; \mathscr{B}_{t'}\,\delta_{t',t+1}$$

NOTATION

$\phi_{x,t} \in \mathbb{R}^{N_x\times N_t}$	real bosonic auxiliary field	K	hopping matrix (incl. chemical potential)
q_x	charge operator on site x	$\mathscr{B}_{t'}$	anti-periodic temporal boundary conditions
$\delta = \beta/N_t$	time-slice spacing, $\tilde{U} = \delta U$	$M[\phi]$, $M[-\phi]$	fermion matrices for spin ↑ / spin ↓

Single-particle correlator

Propagator = inverse fermion matrix, averaged over configurations

$$C_{x,y}(t) \;\equiv\; \big\langle c_x(t)\, c_y^\dagger(0) \big\rangle \;=\; \frac{1}{N_{\mathrm{cfg}}} \sum_{\{\phi\}} M^{-1}_{(t,x);(0,y)} \big[\phi \,\big|\, \tilde{\kappa}, \tilde{\mu}\big] \tag{6}$$

Projected onto the eigenvectors of h, Fourier transform over unit cells

$$C_\pm(\mathbf{k}, t) \;=\; \frac{1}{N_{\mathrm{cell}}} \sum_{\mathbf{x}} e^{i\mathbf{k}\cdot\mathbf{x}}\, C_\pm(\mathbf{x}, t) \tag{7}$$

Decay rate = single-particle energy measured from the chemical potential

$$C_\pm(\mathbf{k}, t) \;\sim\; e^{-(E_\pm(\mathbf{k}) - \mu)t}$$

NOTATION

$c_x,\ c_y^\dagger$	annihilation / creation operators	h	hopping matrix; its eigenvectors define the projection
$M^{-1}[\phi \,\vert\, \tilde{\kappa}, \tilde{\mu}]$	propagator on configuration ϕ	N_x	number of sites $\Rightarrow$ N_x eigen-correlators
N_{cfg}	number of Monte Carlo configurations	$N_{\mathrm{cell}} = N_x/2$	unit cells (bipartite lattice)
$\tilde{\kappa},\ \tilde{\mu}$	hopping and chemical potential	$\pm$	symmetric / antisymmetric sublattice combination

Real-valued non-volume preserving transformation

Split the state $x = (x_a, x_b)$. An affine coupling layer transforms one half conditioned on the other.

TRANSFORMATION

$$\mathbf{y}_a = \mathbf{x}_a, \mathbf{y}_b = \mathbf{x}_b \odot \exp(s_\theta(\mathbf{x}_a)) + t_\theta(\mathbf{x}_a)$$

INVERSE

$$\mathbf{x}_a = \mathbf{y}_a, \mathbf{x}_b = (\mathbf{y}_b - t_\theta(\mathbf{y}_a)) \odot \exp(-s_\theta(\mathbf{y}_a))$$

JACOBIAN

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{pmatrix} 1_d & 0 \\ \dfrac{\partial \mathbf{y}_b}{\partial \mathbf{x}_a} & \operatorname{diag}(\exp s_\theta(\mathbf{x}_a)) \end{pmatrix}$$

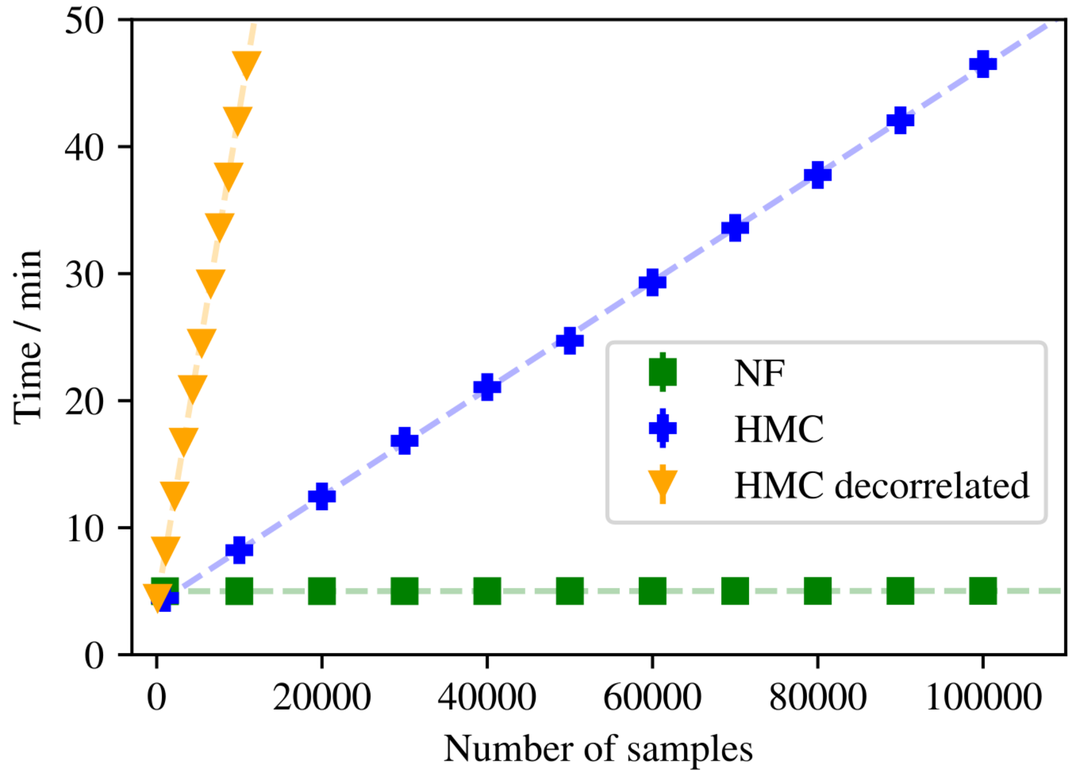
Training

$$\mathrm{KL}(q_\theta||p) = \int \mathcal{D}[\phi] q_\theta(\phi) \log\left(\frac{q_\theta(\phi)}{p(\phi)}\right)$$

↓ substitute $p(\phi) = Z^{-1} e^{-S[\phi]}$

$$\mathrm{KL}(q_\theta||p) = \mathbb{E}_{\phi\sim q_\theta(\phi)}(\log q_\theta(\phi) + S[\phi] + \log Z)$$

Sampling cost



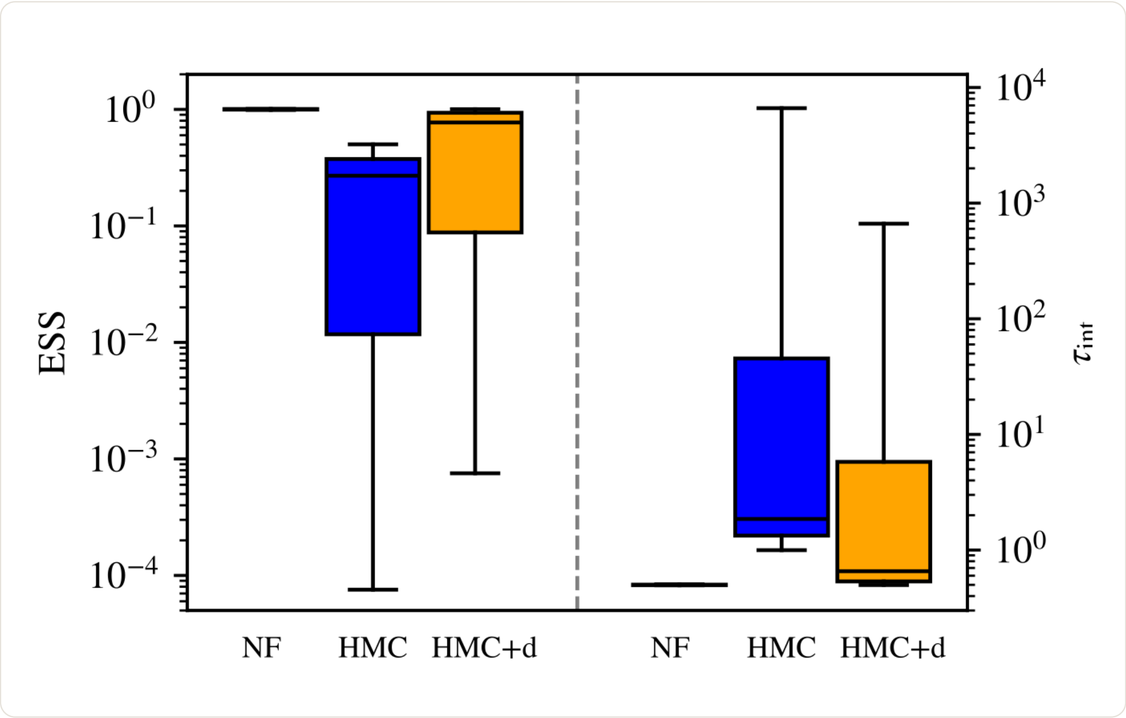
NORMALIZING FLOW

~5 min of training, then essentially free sampling

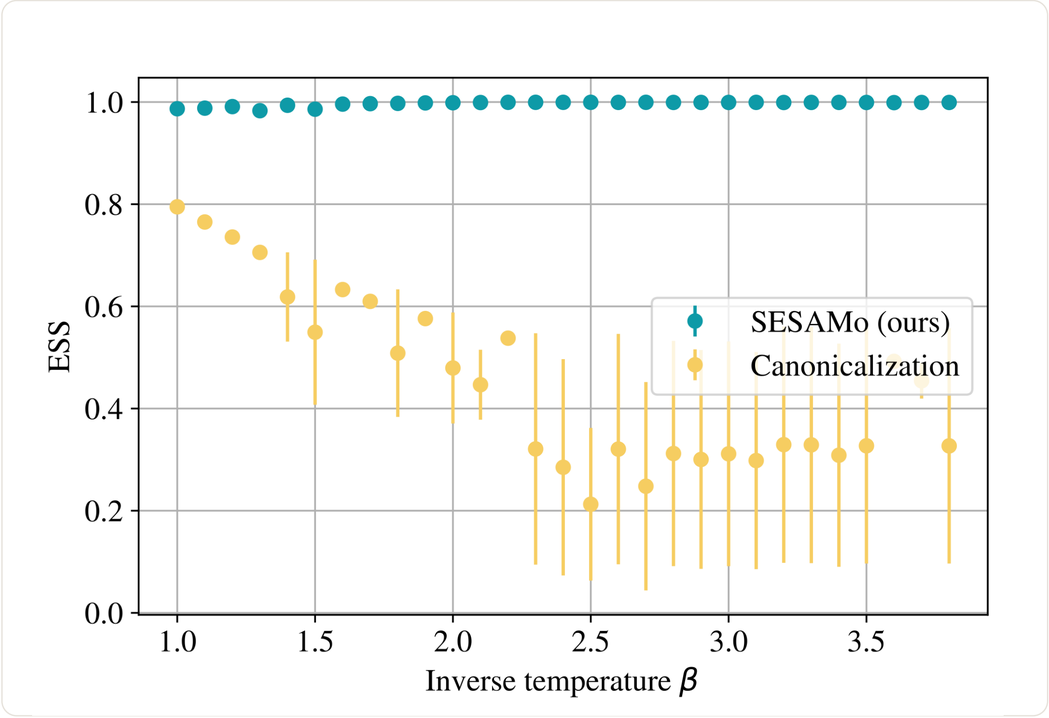
HMC, DECORRELATED

Cost grows linearly with samples and steeper once decorrelated

Sampling efficiency



NF vs. HMC

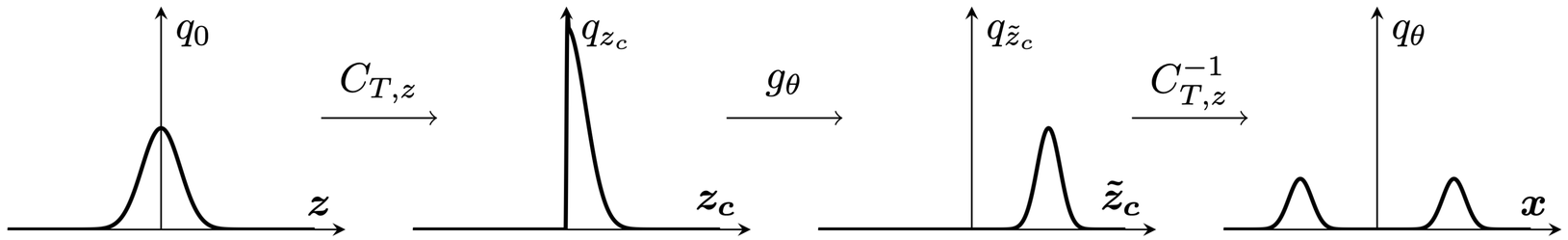


SESAMo vs. Canonicalization

Incorporating symmetries

APPROACH A **Canonicalization**

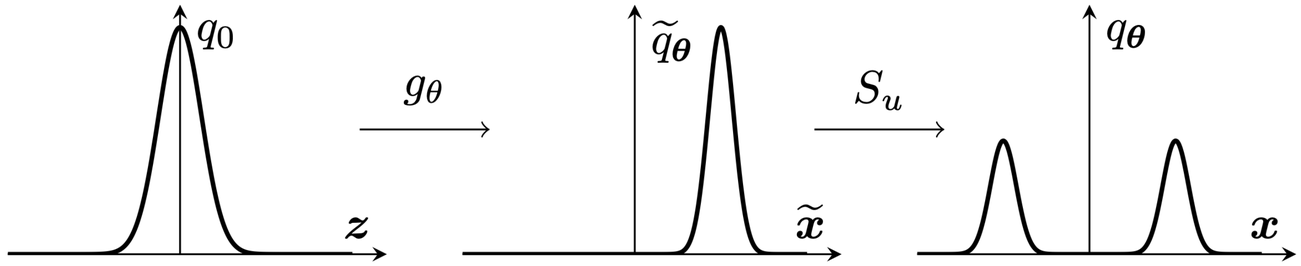
Fold every configuration into one canonical mode, learn there, then unfold back.



J. Köhler et al., ICML 2020; D. Boyda et al., PRL 2021

APPROACH B — THIS WORK **Stochastic modulation (SESaMo)**

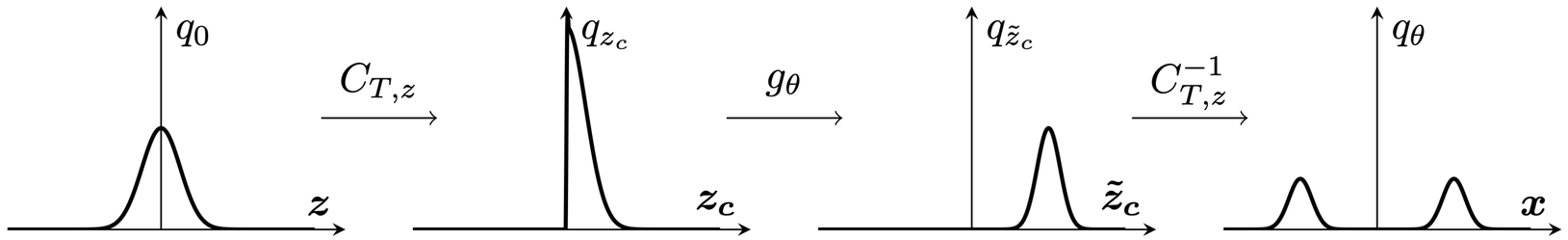
Learn a single mode, then apply a stochastic symmetry operation S_u that replicates it across all symmetric copies with correct weights.



Incorporating symmetries

APPROACH A **Canonicalization**

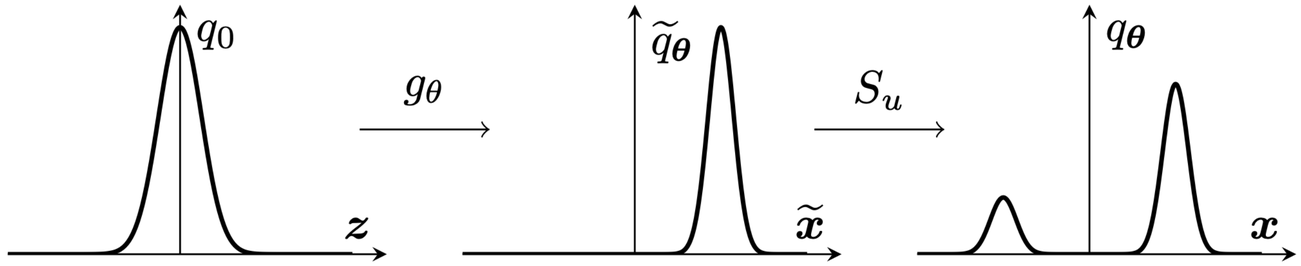
Fold every configuration into one canonical mode, learn there, then unfold back.



J. Köhler et al., ICML 2020; D. Boyda et al., PRL 2021

APPROACH B — THIS WORK **Stochastic modulation (SESaMo) for broken symmetries**

Learn a single mode, then apply a stochastic symmetry operation S_u that replicates it across all symmetric copies with correct weights.



Stochastic modulation

Flow into the canonical cell, then a random transformation

$$\widetilde{g}_{\boldsymbol{\theta},u}(z) \;=\; S_u\big(g_{\boldsymbol{\theta}}(z)\big) \tag{12}$$

Resulting output density

$$q_{\boldsymbol{\theta},b}(\boldsymbol{x}) \;=\; \sum_u p_{S,b}(u) \;\cdot\; q_0\left(\widetilde{g}^{-1}_{\boldsymbol{\theta},u}(\boldsymbol{x})\right) \;\cdot\; \left|\det\left(\frac{\partial \widetilde{g}^{-1}_{\boldsymbol{\theta},u}}{\partial \boldsymbol{x}}\right)\right| \tag{11}$$

Factorization of the Jacobian

$$\det\left(\frac{\partial \widetilde{g}_{\boldsymbol{\theta},u}}{\partial z}\right) \;=\; \det\left(\frac{\partial S_u}{\partial g_{\boldsymbol{\theta}}}\right) \det\left(\frac{\partial g_{\boldsymbol{\theta}}}{\partial z}\right) \tag{13}$$

NOTATION

$z \sim q_0$	prior samples	S_u	stochastic modulation, conditioned on u
$g_{\boldsymbol{\theta}}$	flow into the canonical cell Ω	$p_{S,b}(u)$	modulation probability; learnable b
$\widetilde{x} \sim \widetilde{q}_{\boldsymbol{\theta}},\ \widetilde{x} \in \Omega$	flow output, aligned with one mode	$\widetilde{g}_{\boldsymbol{\theta},u}$	full map, base density $\to$ target

General form and $\mathbb{Z}_2$ example

General modulation over M transformations

$$S_{\mathbf{T},u} : \boldsymbol{x} \mapsto \begin{cases} \boldsymbol{x}, & u = 0 \\ T_1\boldsymbol{x}, & u = 1 \\ \vdots & \\ T_M\boldsymbol{x}, & u = M \end{cases} \qquad u \sim p_{S,b} \qquad T_i \neq T_j \;\; \forall\, i \neq j\,, \qquad T_i\boldsymbol{x} \notin \Omega \tag{14}$$

Log-probability: u is sampled, not marginalized

$$\ln q_{\boldsymbol{\theta}}(\widetilde{g}_{\boldsymbol{\theta},u}(z)) \;=\; \ln p_{S,b}(u) \;+\; \ln q_0(z) \;-\; \ln \left| \det \frac{\partial S_u}{\partial g_{\boldsymbol{\theta}}} \right| \;-\; \ln \left| \det \frac{\partial g_{\boldsymbol{\theta}}}{\partial z} \right| \tag{15}$$

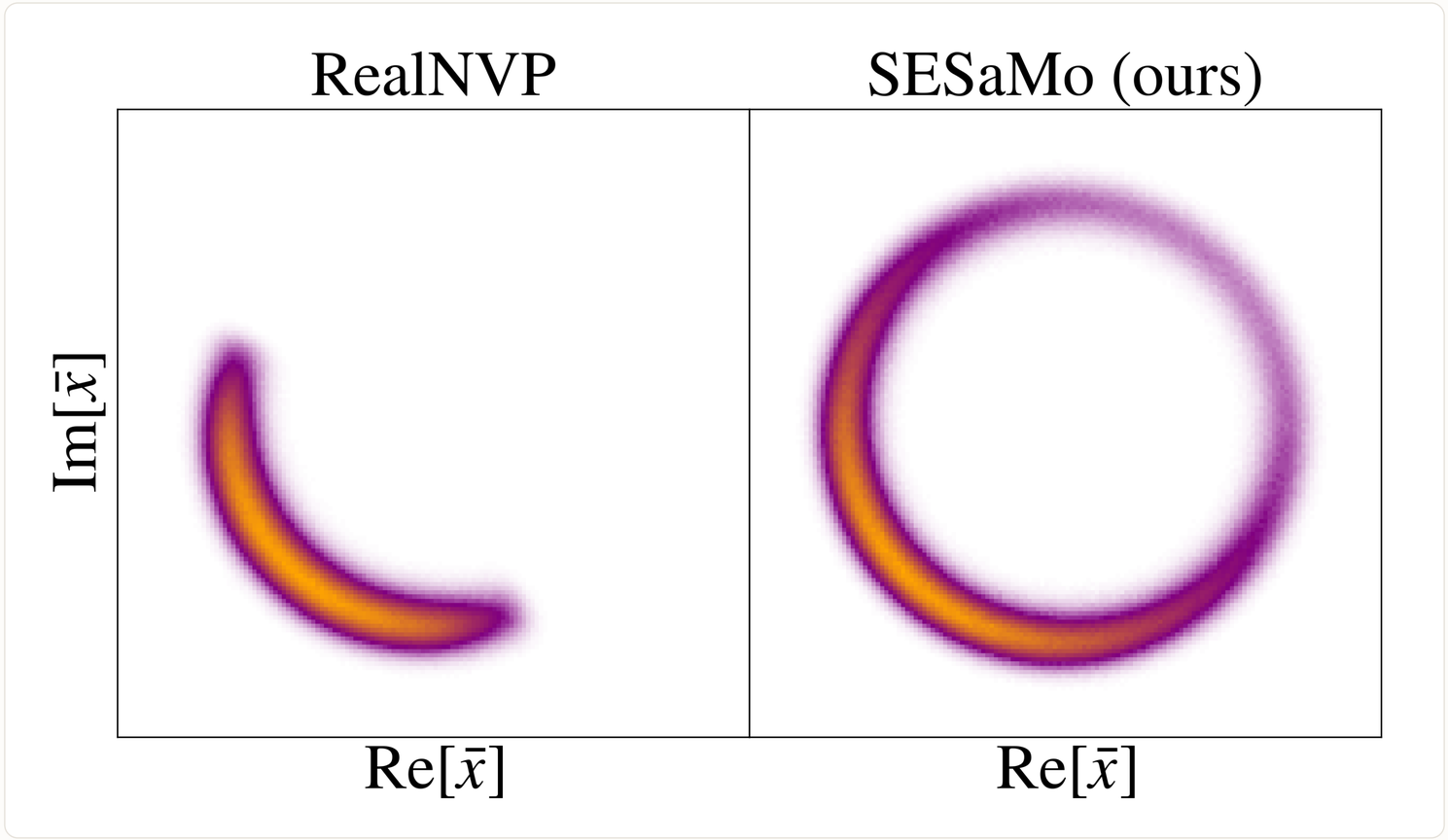
Example: $\mathbb{Z}_2$ symmetry

$$S_u : \boldsymbol{x} \to \begin{cases} \boldsymbol{x} & u = 0 \\ -\boldsymbol{x} & u = 1 \end{cases} \qquad u \sim p_{S,b} = \mathcal{B}\left(e^b\right)\,, \qquad b = \ln 0.5 \tag{16}$$

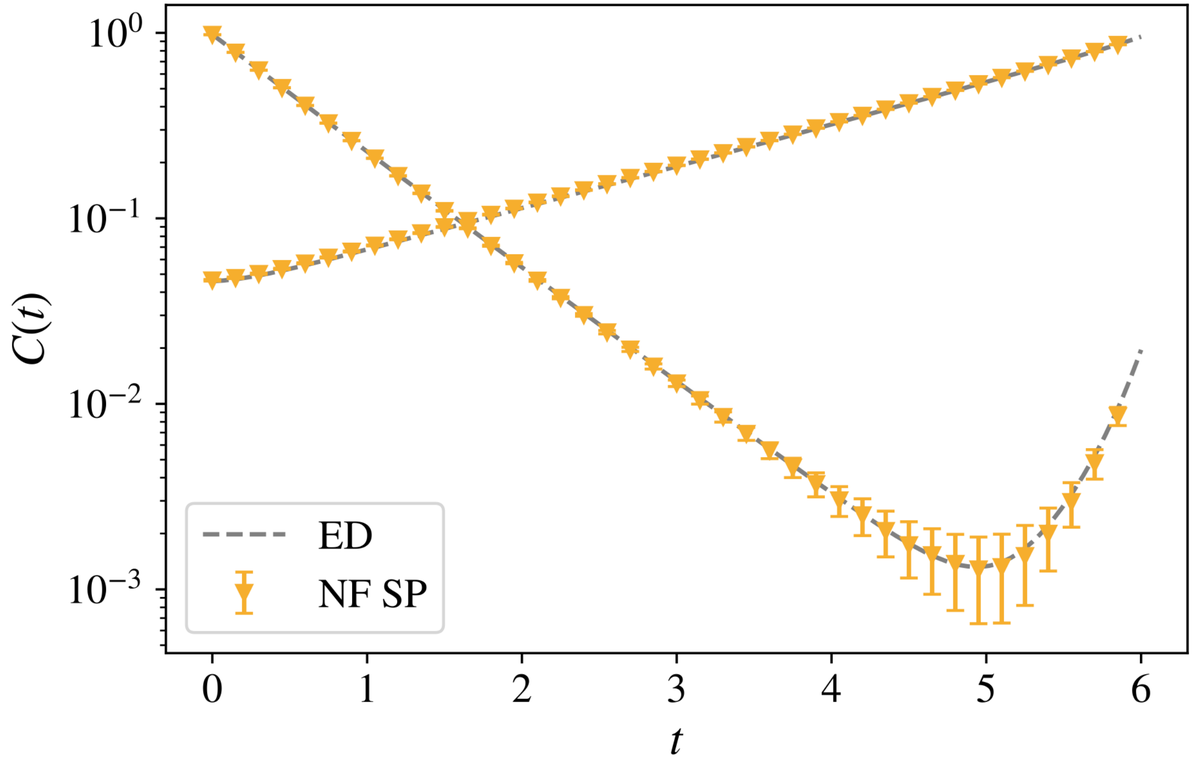
NOTATION

$\{T_i\},\ i = 1 \ldots M$	map samples to distinct modes of $p(\boldsymbol{x})$	$\mathcal{B}(e^b)$	Bernoulli distribution
$T_i \neq T_j,\ T_i\boldsymbol{x} \notin \Omega$	no overlap $\Rightarrow$ bijectivity preserved	$b = \ln 0.5$	equal weight on both $\mathbb{Z}_2$ modes

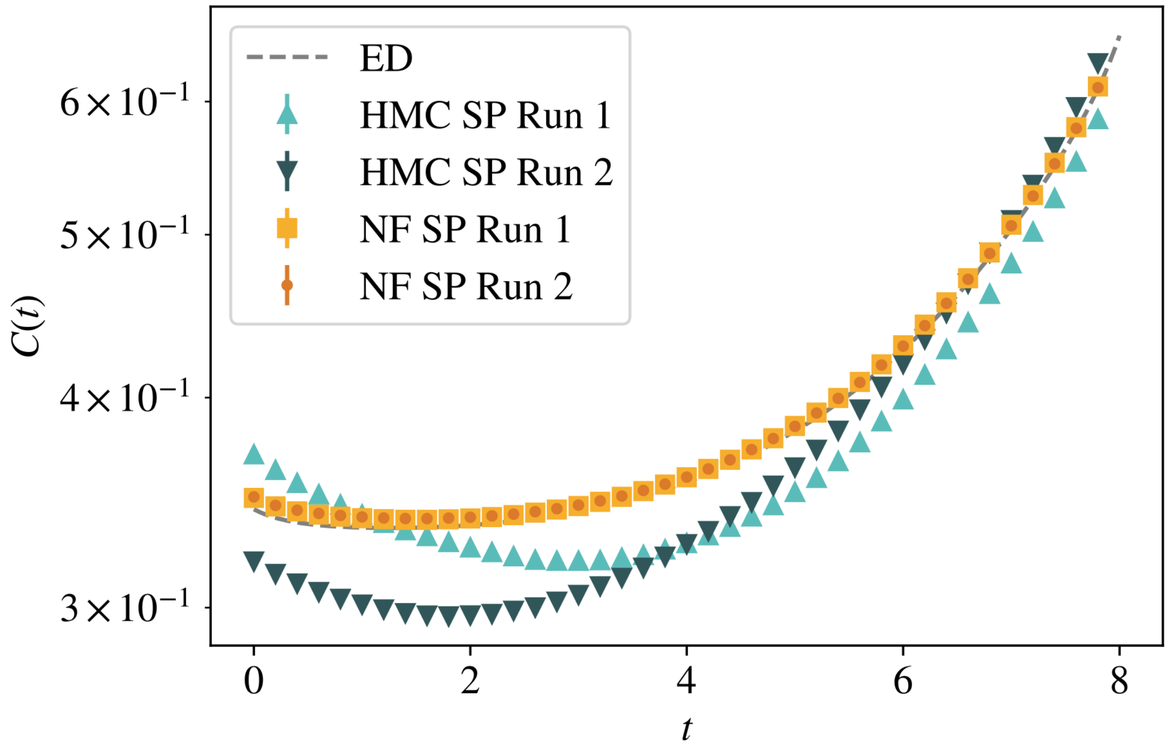
SESaMo for continuous theories



Annealing 8-sites $U = 3$



Annealing HMC comparison 8-sites $U = 2$



Symmetry-informed autoregressive network

Chain of L = $N_x N_t$ layers: one site sampled per layer

$$g_{\boldsymbol{\theta}} \;=\; g_{L,\boldsymbol{\theta}} \circ \cdots \circ g_{2,\boldsymbol{\theta}} \circ g_{1,\boldsymbol{\theta}}\,, \qquad\qquad L = N_x N_t \tag{8}$$

Each layer: symmetry transform, sample, transform back

$$g_{l,\boldsymbol{\theta}} \;=\; T_l^{-1}\, f_{\boldsymbol{\theta}}\, T_l \qquad\qquad g_{l,\boldsymbol{\theta}} : z_{l-1} \mapsto z_l \tag{9}$$

Update rule: only the site at the origin is drawn

$$\begin{aligned} z_l^0 \;&\sim\; 2\,\mathrm{Bern}\Big(n_{\boldsymbol{\theta}}\Big(z_l^{\neq 0}\Big)\Big) - 1\,, \\ z_l^{\neq 0} \;&=\; z_{l-1}^{\neq 0}\,. \end{aligned} \tag{10}$$

NOTATION

$g_{\boldsymbol{\theta}} : z \to \phi$	generative model, $\phi \in \mathbb{D}^{\lvert\Lambda\rvert}$	$z_l^0 \in \mathbb{D}$	the one site drawn in layer l
$\mathbb{D} = \{-1, +1\}$	output values	$z_l^{\neq 0} \in \mathbb{D}_0^{\lvert\Lambda\rvert - 1}$	all other sites, left unchanged
$\mathbb{D}_0 = \{-1, 0, +1\}$	0 = site not yet sampled	$n_{\boldsymbol{\theta}} : \mathbb{D}_0^{\lvert\Lambda\rvert - 1} \to (0, 1)$	feed-forward neural network
$z_{xt} = 0 \;\; \forall x, t$	prior configuration	T_l	symmetry transf. mapping site $l \to$ origin

Exact likelihood

Conditional probability after each layer

$$q_{\boldsymbol{\theta}}\left(z_l^0 \,\middle|\, z_l^{\neq 0}\right) = \begin{cases} 1 - n_{\boldsymbol{\theta}}\left(z_l^{\neq 0}\right) & \text{if} \quad z_l^0 = -1 \\ n_{\boldsymbol{\theta}}\left(z_l^{\neq 0}\right) & \text{if} \quad z_l^0 = +1 \end{cases} \tag{11}$$

Log-density of the final configuration

$$\log q_{\boldsymbol{\theta}}(\phi) \;=\; \sum_l \log q_{\boldsymbol{\theta}}\left(z_l^0 \,\middle|\, z_l^{\neq 0}\right), \qquad \phi \equiv z_L \tag{12}$$

WHY IT WORKS

- Discrete configurations, no Jacobian, log q_θ is exact and cheap.
- The 0 component marks “not yet sampled” → strict autoregressive ordering.
- T_l maps every site to the origin → one parameter set θ shared by all L layers.