Simulating the Hubbard Model with Equivariant Normalizing Flows

Dominic Schuh,[a,b,*] **Janik Kreit,**[a,b,*] **Evan Berkowitz,**[c,d,e,f] **Lena Funcke,**[a,b] **Thomas Luu,**[b,d] **Kim A. Nicoli**[a,b] **and Marcel Rodekamp**[b,d,e,g]

[a] *Transdisciplinary Research Area ''Building Blocks of Matter and Fundamental Interactions'' (TRA Matter), University of Bonn, Bonn, Germany*

[b] *Helmholtz-Institut für Strahlen- und Kernphysik (HISKP), University of Bonn, Bonn, Germany*

[c] *Jülich Supercomputing Center (JSC), Forschungszentrum Jülich, Jülich, Germany*

[d] *Institute for Advanced Simulation 4 (IAS-4), Forschungszentrum Jülich, Jülich, Germany*

[e] *Center for Advanced Simulation and Analytics (CASA), Forschungszentrum Jülich, Jülich, Germany*

[f] *College of Science and Mathematics, University of the Virgin Islands, RR1 Box 10000 Kingshill, St. Croix 00850, US Virgin Islands*

[g] *Institut für Theoretische Physik, Universität Regensburg, 93053 Regensburg, Germany*

E-mail: schuh@hiskp.uni-bonn.de, kreit@hiskp.uni-bonn.de, evan.berkowitz@uvi.edu, lfuncke@uni-bonn.de, t.luu@fz-juelich.de, knicoli@uni-bonn.de, marcel.rodekamp@ur.de

Generative models, particularly normalizing flows, have shown exceptional performance in learning probability distributions across various domains of physics, including statistical mechanics, collider physics, and lattice field theory. In the context of lattice field theory, normalizing flows have been successfully applied to accurately learn the Boltzmann distribution, enabling a range of tasks such as direct estimation of thermodynamic observables and sampling independent and identically distributed (i.i.d.) configurations.

In this work, we present a proof-of-concept demonstration that normalizing flows can be used to learn the Boltzmann distribution for the Hubbard model. This model is widely employed to study the electronic structure of graphene and other carbon nanomaterials. State-of-the-art numerical simulations of the Hubbard model, such as those based on Hybrid Monte Carlo (HMC) methods, often suffer from ergodicity issues, potentially leading to biased estimates of physical observables. Our numerical experiments demonstrate that leveraging i.i.d. sampling from the normalizing flow effectively addresses these issues.

---

*Speaker

1. Introduction

The Hubbard model is a condensed matter model describing the interaction of strongly coupled electronic systems. Studying this model often poses severe challenges for numerical simulations, primarily due to its multi-modal nature and intricate energy landscape. Standard sampling methods, such as Hybrid Monte Carlo (HMC), often suffer from ergodicity problems [1], potentially leading to biased and inaccurate estimates of physical observables. Novel techniques have been developed towards having more efficient samplers, such as the combination of HMC with radial updates [2].

In these proceedings, we deploy a deep generative machine learning approach to overcome the ergodicity issue in the context of the Hubbard model. Deep generative models, also known as generative neural samplers (GNSs), in particular normalizing flows [3–5] and autoregressive neural networks [6, 7], have demonstrated great capabilities for modeling Boltzmann distributions of physical and chemical systems. These generative approaches, often known as Boltzmann generators [8], have lately emerged in various fields, ranging from lattice quantum field theory [9–12] to statistical mechanics [13, 14] to string theory [15, 16] to quantum chemistry [8, 17, 18]. Generative neural samplers hold promise not only for allowing faster and more efficient sampling, but also for directly estimating thermodynamic observables [10] as well as entanglement entropies [19, 20], and Feynman propagators [21]. Furthermore, recent studies demonstrate that GNSs are highly effective in sampling from distributions with challenging topologies, such as bimodal distributions [22, 23] and gauge theories affected by topological freezing [24]. This positions GNSs as a well-suited alternative for sampling from challenging probability densities where HMC faces ergodicity issues.

Our work represents the first attempt in applying GNSs to the Hubbard model. Specifically, we propose to use an equivariant normalizing flow [25] to efficiently learn the underlying probability distribution. This approach allows us to directly incorporate an arbitrary number of symmetries in the neural network design.

The remainder of this paper is organized as follows: First, in Sec. 2, we introduce the Hubbard model and discuss how standard HMC methods struggle with ergodicity. In Sec. 3, we introduce equivariant normalizing flows and elaborate on their application to the Hubbard model. We present numerical experiments in Sec. 4 to prove the advantages of the proposed method compared to standard HMC. Finally, we provide conclusions and an outlook in Sec. 5.

2. Hubbard Model

2.1 Theoretical Background

The Hubbard model [26] describes the interaction of fermions on a spatial lattice of fixed ions, where the atomic grid is considered static. It is capable of describing a vast amount of physical systems [27], most widely applied to model carbon nanostructures and other graphene formations.

The dynamics of the Hubbard model at half-filling, i.e., with vanishing chemical potential, are

described by the Hamiltonian

$$H_\alpha = H_{\text{TB}} + \alpha \frac{U}{2} \sum_x \left(n_{x,\uparrow} - n_{x,\downarrow} \right) - (1 - \alpha) \frac{U}{2} \sum_x \left(n_{x,\uparrow} - n_{x,\downarrow} \right)^2, \tag{1}$$

$$H_{\text{TB}} = -\kappa \sum_{\langle x,y \rangle} \left(a_x^\dagger a_y - b_x^\dagger b_y \right), \tag{2}$$

where Eq. (2) is often referred to as the *tight-binding* Hamiltonian. It describes the binding and hopping of particles between adjacent atomic sites in a solid state, with κ being responsible for the hopping dynamics. The operators $a_x^\dagger (a_x)$ and $b_x^\dagger (b_x)$ create (annihilate) a spin-up and spin-down particle, respectively, with the number operator $n_{x,\uparrow} = a_x^\dagger a_x$ for a spin-up particle. In Eq. (1), the parameter U describes the on-site interaction strength, while $\alpha \in [0, 1]$ continuously parametrizes the choice of basis, i.e., $\alpha = 0$ denotes the spin basis and $\alpha = 1$ the particle/hole basis. Thus, the parameter α defines the physical interpretation of the degrees of freedom, i.e., particles with spin up or down versus particles and holes, respectively. In this work, we will restrict ourselves to the particle/hole (ph) basis, while in future work we will extend our analysis to the spin basis.

A probabilistic interpretation of the partition function

$$Z_{\text{ph}} = \text{tr} \left(e^{-\beta H_{\alpha=1}} \right) \tag{3}$$

requires the transition to a path integral formulation, which is achieved by discretizing the inverse temperature β into N_t evenly spaced, so-called *time-slices* and performing a Suzuki-Trotter decomposition of second order [28, 29]. Furthermore, to decouple a quartic fermionic term appearing in the path integral, a continuous Hubbard-Stratonovich (HS) transformation [30] is needed,

$$e^{-\frac{1}{2}\tilde{U}n^2} = \frac{1}{\sqrt{2\pi\tilde{U}}} \int_{-\infty}^{\infty} \mathrm{d}\phi \, e^{-\frac{1}{2\tilde{U}}\phi^2 \pm i\phi n}, \tag{4}$$

where $\tilde{U} \equiv U\beta/N_t$ is the rescaled on-site interaction strength, $n = n_\downarrow - n_\uparrow$ is a fermionic number operator, and ϕ is a bosonic auxiliary field. This transformation reduces the power of the Grassmanian fields from four to two at the cost of introducing one bosonic auxiliary field per fermionic degree of freedom. Lastly, by inserting coherent fermionic states and integrating out the remaining fermionic degrees of freedom using standard Gaussian integrals, the partition function is rendered purely dependent on the bosonic fields [31–34],

$$Z_{\text{ph}} = \int_{-\infty}^{\infty} \left[\prod_{x,t} \mathrm{d}\phi_{xt} \right] e^{-S[\phi]}, \tag{5}$$

with the Hubbard action defined as

$$S[\phi] = \frac{1}{2\tilde{U}} \sum_{x,t \in \Lambda} \phi_{xt}^2 - \log \det M[i\phi] - \log \det M[-i\phi], \tag{6}$$

where $\Lambda = N_x \times N_t$ is the space-time volume of the lattice and xt represents a site on the lattice. The fermion matrix M in the exponential discretization[1] reads

$$M^e[\phi]_{x't',xt} = \delta_{x',x}\delta_{t',t} - [e^h]_{x',x} e^{\phi_{xt}} \mathcal{B}_{t'} \delta_{t',t+1}, \tag{7}$$

---

[1] There are other choices for incorporating the hopping term in the fermion matrix, such as the diagonal [31, 34] or the linear [33] discretization. However, all discretizations agree in the continuum limit $N_t \to \infty$ [1].

with the hopping matrix $h_{x',x} = \kappa \delta_{\langle x',x \rangle}$. Here, $\mathcal{B}$ explicitly incorporates anti-periodic boundary conditions in the temporal direction t, i.e., $\mathcal{B}_t = +1$ for $0 < t < N_t$ and $\mathcal{B}_0 = -1$. Taking a closer look at the Hubbard action in Eq. (6), we see that it contains two different components: a Gaussian term and a fermion determinant. The first one stems from the Hubbard-Stratonovich transformation in Eq. (4) and accounts for the on-site interaction previously introduced in Eq. (1), while the latter one encodes the hopping dynamics described by the tight-binding Hamiltonian in Eq. (2).

In these proceedings, we focus on a $1 + 1$D lattice, where the spatial lattice extent is fixed to two, i.e. $N_x = 2$, and we consider different temporal lattice extents, N_t. Unless specified otherwise, we work in the particle/hole basis with the fermion matrix in the exponential discretization.

2.2 Symmetries of the Hubbard Action

The Hubbard action in Eq. (6) obeys a large set of symmetries. Generally, these can be symmetries of the entire action or symmetries only preserving the fermion determinant. For a comprehensive discussion of the symmetries of the Hubbard model, we refer the reader to [1]. In the following, we outline the specific symmetries that will be relevant for these proceedings:

- **$\mathbf{Z_2}$-Symmetry**: The entire action is invariant under a sign-flip transformation

$$\phi \to -\phi,$$

  i.e., in this choice of basis, the action is invariant under the exchange of particles and holes. Note that this is only a symmetry on bipartite lattices at half-filling.

- **Space-Translation Symmetry**: Under the condition that the on-site interaction strength U is uniform across the entire lattice, the action is invariant under the exchange of **all** spatial sites, i.e.,

$$(\phi_{x_1}, \phi_{x_2}) \to (\phi_{x_2}, \phi_{x_1}).$$

- **Periodicity Symmetry**: The only symmetry of relevance that is not preserved by the Gaussian part of the action is the 2π periodicity symmetry of the fermion determinant [1]. As can be seen in Eqs. (6) and (7), the determinant $M^e[i\phi]$ only depends on $e^{i\phi}$ and is therefore invariant under

$$\phi_{xt} \to \phi_{xt} + n \cdot 2\pi, \qquad\qquad n \in \mathcal{Z}.$$

These symmetries will become important in Sec. 3.1 for designing equivariant normalizing flows.

2.3 Hybrid Monte Carlo Simulations

One widely used method to generate samples for the Hubbard model is the Hybrid Monte Carlo (HMC) algorithm [35]. While HMC is extremely successful in various applications, it faces substantial challenges when being applied to the Hubbard model. As illustrated in Fig. 1, the density for a $(N_x, N_t) = (2, 1)$ lattice consists of diagonal bands separated by infinite potential barriers, with an underlying periodic blob structure, ''wrapped'' by a Gaussian. The multi-modal nature of this distribution presents significant ergodicity challenges, as the leapfrog integrator is repelled by
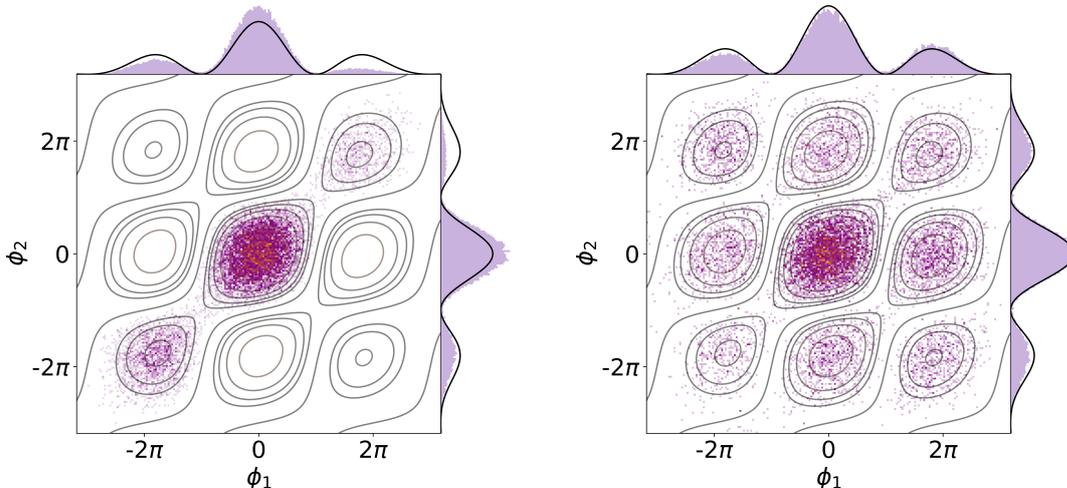
Figure 1: HMC configurations for a $(N_x, N_t) = (2, 1)$ lattice with parameters $\beta = \kappa = 1$ and $U = 18$, generated with a leapfrog integrator with an integration step of $\epsilon = 0.1$ (left) and $\epsilon = 1.0$ (right). The integrated autocorrelation time τ_{int} and the acceptance rate a are $\tau_{int} = 83 \pm 14$ and $a = 99.9\%$ (left) and $\tau_{int} = 346 \pm 100$ and $a = 86.9\%$ (right), respectively. Histograms of the magnetization, obtained by summing along each dimension, respectively, are shown at the top and on the right hand side of the figures. The analytical contours are exact for $N_t = 1$ and in the strong-coupling limit, i.e. $U/\kappa \to \infty$, for $N_t > 1$. The colormap ranging from purple to yellow represents the areas of low and high density, respectively.

high potential barriers. This makes it hard to tunnel between the modes and reach every point in configuration space, as shown in the left panel of Fig. 1. Furthermore, the separation of the different regions in the $N_x N_t$-dimensional configuration space is determined by manifolds where the fermion determinant vanishes. These manifolds are of codimension 1 [1], i.e., they are of dimension $N_x N_t - 1$, and, as a result, are present at all lattice sizes.

This issue may be addressed by a coarser integrator since the proposed configurations in the Markov chain have a larger spread. However, a coarser integrator leads to a substantial decrease in the acceptance rate, along with an increase in the integrated autocorrelation time τ_{int}, as demonstrated in the right panel of Fig. 1.

3. Normalizing Flows

A normalizing flow [4, 5] is a parametric bijective map, $f_\theta : \mathcal{Z}, \to \mathcal{K}$ from a latent space $\mathcal{Z}$ with a simple prior distribution $q_Z(z)$, e.g., a Gaussian, to a target space $\mathcal{K}$ with a target distribution $p(\phi)$. In the case of coupling-based normalizing flow [36, 37], this map consists of a series of invertible and differentiable transformations f^i (see Fig. 2), each parametrized by neural networks with parameters θ_i. This ensures bijectivity of the entire map, yielding

$$f_\theta(z) = \left(f^l \circ f^{l-1} \circ \cdots \circ f^1 \right)(z), \tag{8}$$

where the dependence on the parameters θ_i for each block was omitted for ease of notation. During the training of a normalizing flow, the parameters θ_i of the neural networks parametrizing each block
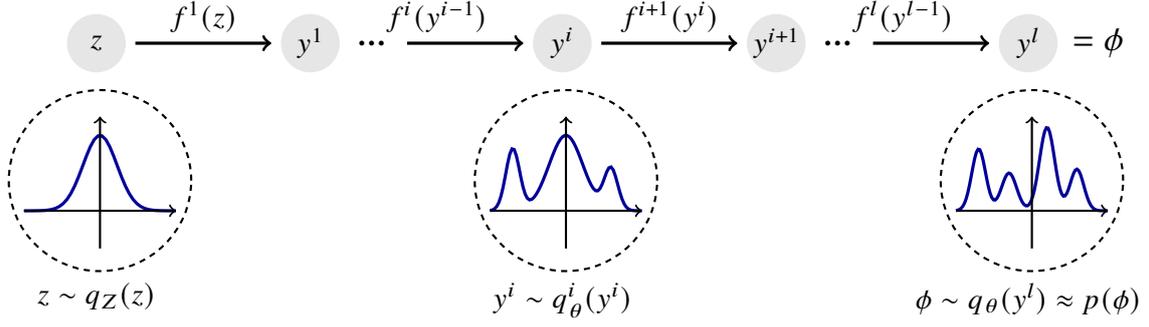
Figure 2: Schematic representation of a coupling-based normalizing flow. Starting with a Gaussian prior distribution $q_Z(z)$ (left), the normalizing flow $f_\theta(z)$ maps the prior samples z to the target samples ϕ through a set of invertible functions f^i, each taking as input the output of the previous layer, y^{i-1}, ultimately approximating the target distribution $p(\phi)$ (right). In the context of lattice field theories, this target distribution corresponds to the path integral distribution $p(\phi) = Z^{-1}e^{-S[\phi]}$, where $S[\phi]$ is the action of the theory.

$f^i_{\theta_i}$ are optimized with the goal of enabling the normalizing flow to approximate a target distribution $p(\phi)$. Once the flow has been trained, sampling becomes efficient, as it only requires drawing samples from the base distribution $q_Z(z)$ and transforming them according to the parametrized map $f_\theta(z)$, thus allowing for embarrassingly parallelizable generation of independent and identically distributed (i.i.d) samples from the approximated density $q_\theta(z) \approx p(\phi)$.

When considering lattice field theories, the goal is to generate samples from the path integral distribution of the theory. Therefore, the target density to sample from is given by $p(\phi) = Z^{-1}e^{-S[\phi]}$, where Z is the partition function. To approximate this distribution, the parameters θ of the normalizing flow are optimized by gradient descent while minimizing the so-called reverse Kullback-Leibler divergence [38]

$$\mathrm{KL}(q_\theta||p) = \int \mathcal{D}[\phi]q_\theta(\phi) \ln\left(\frac{q_\theta(\phi)}{p(\phi)}\right), \tag{9}$$

which provides a measure of how much the variational distribution q_θ generated by the normalizing flow differs from the target distribution p. The likelihood of q_θ is known analytically,

$$q_\theta(\phi) = q_z(f_\theta^{-1}(\phi))\left|\frac{\mathrm{d}f_\theta}{\mathrm{d}z}\right|^{-1} = q_z(z)\left|\frac{\mathrm{d}f_\theta}{\mathrm{d}z}\right|^{-1}, \tag{10}$$

which allows to rewrite Eq. (9) as

$$\mathrm{KL}(q_\theta||p) = \mathbb{E}_{z\sim q_Z}\left[S[f_\theta(z)] - \ln\left|\frac{\mathrm{d}f_\theta}{\mathrm{d}z}\right|(z) + \ln q_Z(z) + \ln Z\right]. \tag{11}$$

This expression contains the logarithm of the partition function, which is unknown. However, being a constant shift, this quantity vanishes upon taking the gradient with respect to the model parameters.

Each component of f_θ has to be designed such that each transformation is invertible and has a tractable Jacobian to ensure the efficient computation of its determinant. Many architectures have

been proposed that fulfill these requirements, we refer to Ref. [4] for an overview. In this work, we use the so-called *real-valued non-volume preserving* (RealNVP) [37] transformation. This particular type of affine transformation requires two neural networks s_θ and t_θ, responsible for scaling and shifting the input y^l, respectively. Furthermore, to preserve the invertibility of the transformation, the input sites y^l are divided into two partitionings, y^l_{on} and y^l_{off}, and the transformation acting on the *on* sites is restricted to depend on the complementary *off* sites,

$$
\begin{aligned}
y^{l+1}_{\mathrm{off}} &= y^l_{\mathrm{off}}, \\
y^{l+1}_{\mathrm{on}} &= y^l_{\mathrm{on}} \cdot \exp\left[s_\theta\left(y^l_{\mathrm{off}} \right) \right] + t_\theta\left(y^l_{\mathrm{off}} \right).
\end{aligned}
\tag{12}
$$

This type of transformation ensures a tractable Jacobian determinant

$$
\det \frac{\partial y^{l+1}}{\partial y^l} = \begin{vmatrix} \mathbf{I}_d & 0 \\ * & \mathrm{diag}\left(\exp\left[s_\theta\left(y^l_{\mathrm{off}} \right) \right] \right) \end{vmatrix},
\tag{13}
$$

where d refers to the dimensionality of the *off* partitioning and y^l is the output of the l-th layer.

Every block $f^i(y^{i-1})$, as shown in Fig. 2, is an instance of a RealNVP transformation, i.e., a RealNVP block, with an alternating splitting between *on* and *off* partitionings, to ensure that the entire input is transformed within two blocks.

3.1 Equivariant Normalizing Flows for the Hubbard Model

Equivariant normalizing flows [25, 39] have been proposed to integrate prior knowledge, e.g., known symmetries, of the target distribution into the deep generative model, in order to make the bijective map equivariant by design. Reducing redundancies in the distribution in this way leads to more efficient training as well as more accurate results.

Let T be some transformation that represents a symmetry of the target distribution. A neural network $f_\theta(z)$ is *equivariant* with respect to T if it fulfills [25]

$$
f_\theta(z) = T^{-1} f_\theta(Tz).
\tag{14}
$$

This way, the input z gets transformed into the canonical cell of the theory, leaving the neural network with the task of learning only the canonicalized distribution.

In the case of the Hubbard model, we leverage the symmetries introduced in Sec. 2.2 and define the following set of transformations:

$$
\mathbf{Z}_2\text{-Symmetry} \quad : \quad (z_1, z_2) \mapsto \begin{cases} (z_1, z_2) & \text{if } z_1 + z_2 \geq 0 \\ -(z_1, z_2) & \text{else} \end{cases}
\tag{15}
$$

$$
\text{Space-Translation Symmetry} \quad : \quad (z_1, z_2) \mapsto \begin{cases} (z_1, z_2) & \text{if } z_1 - z_2 \leq 0 \\ (z_2, z_1) & \text{else} \end{cases}
\tag{16}
$$

$$
\text{Periodicity Symmetry} \quad : \quad z_{xt} \mapsto z_{xt} - 2\pi \cdot k, \quad k = \mathrm{round}\left(\frac{z_{xt}}{2\pi} \right),
\tag{17}
$$

where z is a sample taken from the Gaussian distribution and $z_x = \sum_{t=1}^{N_t} z_{xt}$. An intuitive understanding of how these symmetries act on the base distribution is illustrated in Fig. 3: The periodicity
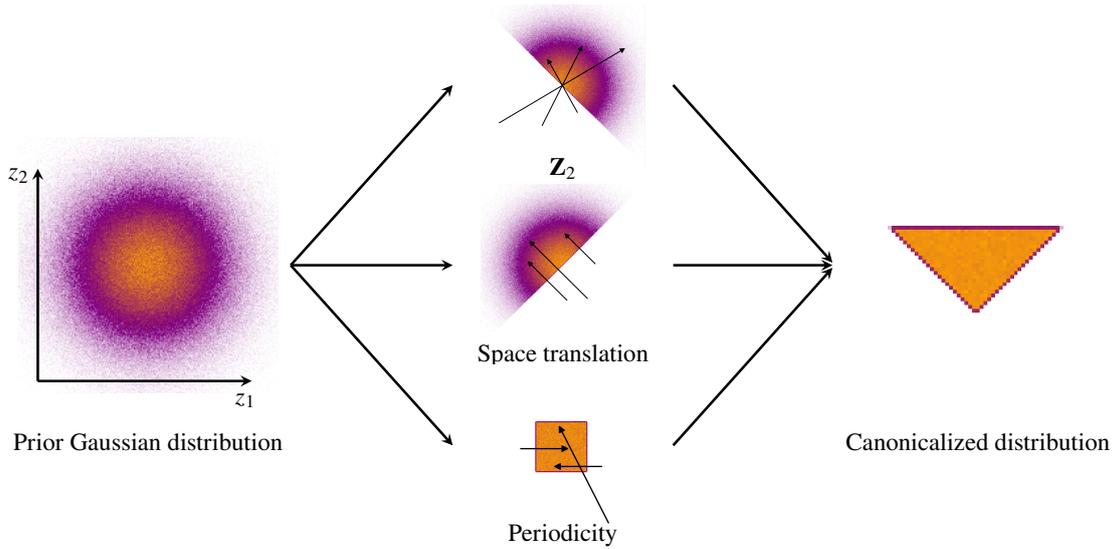
Figure 3: Transformation of the prior distribution into the canonical cell of the theory. Starting from a Gaussian prior distribution (left), we apply three types of symmetry transformations (middle): $\mathbf{Z}_2$ symmetry (top), space-translation symmetry (center), and periodicity symmetry (bottom). This results in a canonical cell of triangular shape (right). The small black arrows indicate the mapping of three exemplary samples under these transformation, indicating their position before and after the transformation has been applied.

symmetry maps every sample into a square of length 2π, while the $\mathbf{Z}_2-$ and space-translation symmetries *fold* this square, resulting in a triangle. Inverting these symmetries can therefore be understood as *unfolding* this triangle and mapping every sample back to its origin, i.e., going backward by applying the inverse transformation T^{-1}.

4. Results

In this section, we apply the aforementioned techniques to the Hubbard model by training a normalizing flow f_θ to approximate the target distribution $p(\phi) = Z^{-1}e^{-S[\phi]}$, where Z and $S[\phi]$ are given by Eq. (5) and Eq. (6), respectively. As there is no guarantee that $q_\theta \equiv p$, one should enforce asymptotic unbiasedness [14]. To this end, one could either use the so-called NeuralMCMC or metropolization technique by applying an accept-reject step to the i.i.d. samples from the normalizing flow, or perform neural importance sampling (NIS) [10]. The numerical experiments shown in these proceedings are based on NeuralMCMC and have been obtained using a preliminary version of the NeuLat software framework [40].

The top row of Fig. 4 shows our results for a single time-slice, i.e., a $(N_x, N_t) = (2, 1)$ system. We see that, in this case, even a *non*-equivariant normalizing flow is able to capture the underlying probability density, as displayed in the left column. However, when making the flow equivariant, i.e., when including the equivariant layer to incorporate the symmetries into the model by design, we observe a notable improvement in both training efficiency as well as the quality of the learned distribution, as shown in the middle column. While the non-equivariant normalizing flow takes 25
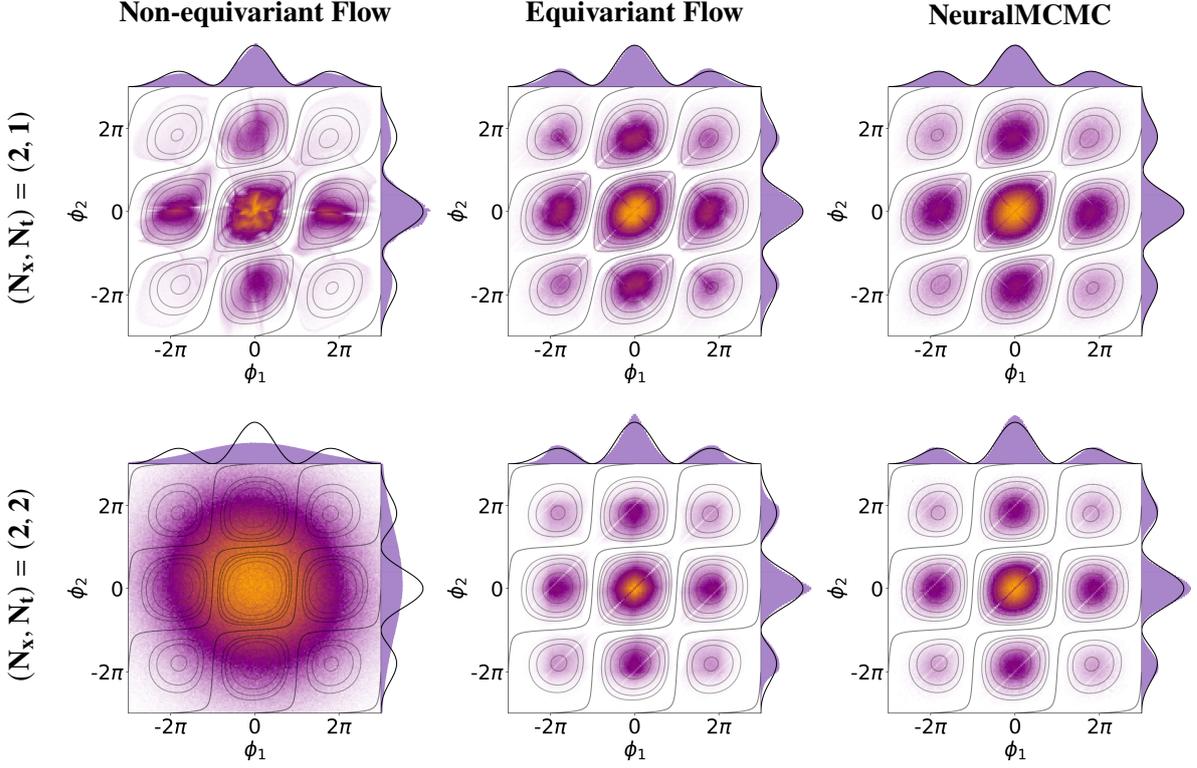
Figure 4: Field configurations obtained with non-equivariant (left), equivariant (middle), and equivariant and metropolized (right) normalizing flows. The marginalized magnetization, shown at the top and on the right of each plot, is exact for $N_t = 1$ and in the strong-coupling limit for $N_t > 1$. The top and bottom rows show results for $(N_x, N_t) = (2, 1)$ and $(N_x, N_t) = (2, 2)$ lattices, respectively. For $N_t \geq 1$, $\phi_{\{1,2\}}$ is understood to be the sum in the temporal direction, i.e., $\phi_{\{1,2\}} = \sum_{i=1}^{N_t} \phi_{\{1,2\}t}$. While the equivariant approach is able to learn the structure for both lattice sizes to high precision, the non-equivariant flow is unable to learn an approximate to the target distribution for $N_t > 1$. Note that the slightly visible lines in the equivariant distributions originate from a penalty term necessary to keep the normalizing flow bijective while applying the symmetry transformation T. The right-most plots on both rows have been obtained by performing metropolization, i.e., filtering i.i.d. samples through a metropolis accept-reject step. This ensures that the sampling from the approximate model q_θ is asymptotically unbiased.

hours of training to achieve an acceptance rate of 75%, the equivariant counterpart converges to an acceptance rate of 85% already after 16 minutes of training. This significant speedup in training is illustrated in Fig. 5. Furthermore, when comparing the integrated autocorrelation times in Tab. 1, we see a difference of a factor of two between the equivariant and non-equivariant approaches. Remarkably, both approaches show a substantial improvement in the integrated autocorrelation time of roughly two orders of magnitude compared to the standard HMC approach (see the caption of Fig. 1).

The bottom row of Fig. 4 shows our results for increasing the temporal extent of the lattice by one, i.e., $N_t = 2$. In this case, the non-equivariant normalizing flow is unable to learn the target probability distribution, even after more than 24 hours of training, as shown in the left column. The equivariant normalizing flow, in contrast, is able to approximately learn the target distribution, as shown in the middle column. Here, it achieves an acceptance rate of 69.7% and an integrated
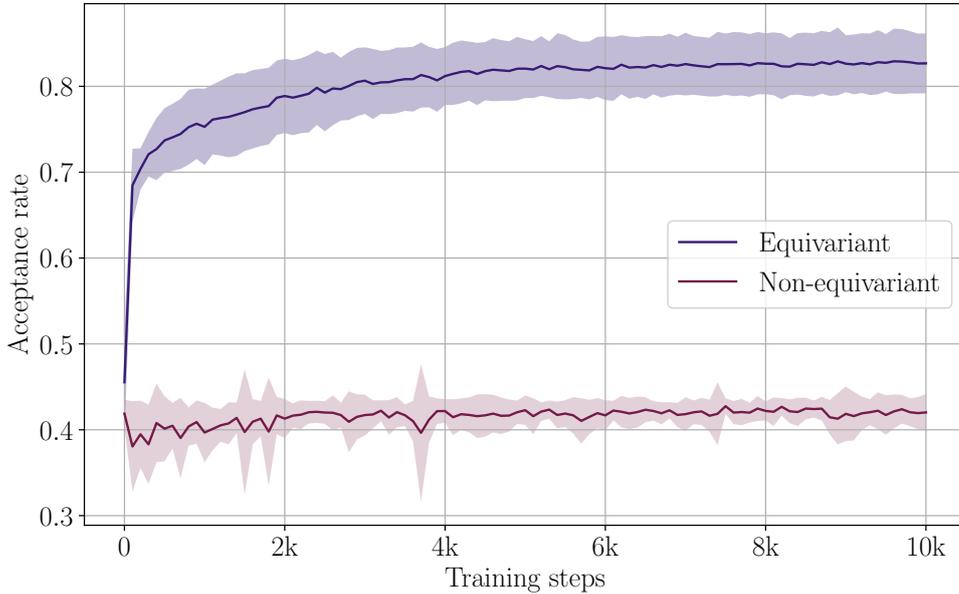
Figure 5: Acceptance rate vs. training steps for a non-equivariant (red) and an equivariant (purple) normalizing flow for a $(N_x, N_t) = (2, 1)$ lattice. The equivariant approach yields acceptance rates above 80% after four thousand training steps, while the non-equivariant only reaches 40%. In order to to reach an acceptance rate of 70%, the non-equivariant flow would need more than seven-hundred thousand training steps.

autocorrelation time that is two orders of magnitude smaller than in the standard HMC approach (see Tab. 1 and Fig. 1), thus indicating a much better sampler.

As mentioned in Sec. 2.2, the $\mathbf{Z}_2$- and space-translation symmetries leave the entire action invariant, while the periodicity symmetry is only a symmetry of the fermion determinant, but not of the Gaussian component of the action. In this sense, the periodicity symmetry can be seen as an *approximate* symmetry. The effect of this ''approximation'' can be observed in the magnetization of the equivariant distribution for $(N_x, N_t) = (2, 1)$, as shown in the second column of the top row of Fig. 4, where the two outer peaks are tilted inwards and, therefore, do not follow the same Gaussian envelope as the analytically obtained contours. However, the third column of the top row of Fig. 4 demonstrates that this artifact can be removed by using the NeuralMCMC technique.[2]

5. Conclusion and Outlook

Well-established sampling methods like HMC often suffer from severe ergodicity problems when being applied to the Hubbard model. In this work, we propose to use equivariant normalizing flows to overcome these issues. We present a proof-of-concept demonstration that a normalizing flow can be trained to approximate the target Boltzmann distribution of the Hubbard model and allows

---

[2]Note that the outer peaks for the equivariant distribution of $(N_x, N_t) = (2, 2)$ exhibit a similar tilting, see the second column of the bottom row of Fig. 4. However, in this case, the analytical solution (solid black line) is not exact for $N_t > 1$, thus preventing a similar comparison with the ground truth as for the $(N_x, N_t) = (2, 1)$ case.

Table 1: Integrated autocorrelation times τ_{int} of the equivariant and non-equivariant normalizing flow models for different system sizes of the Hubbard model. Since the non-equivariant approach fails to produce a result for $N_t > 1$, a value for τ_{int} cannot be determined in this case.

(N_x, N_t)	$\tau_{\text{int, non-equivariant}}$	$\tau_{\text{int, equivariant}}$
(2,1)	1.52 ± 0.04	0.71 ± 0.02
(2,2)	-	1.17 ± 0.03

to efficiently sample configurations for a 2×1 lattice. Furthermore, we show that incorporating symmetries of the action into an equivariant normalizing flow architecture lowers the training cost, allows to achieve higher acceptance rates, and reduces the integrated autocorrelation times, compared to the non-equivariant approach.

Nevertheless, making the flow equivariant generally carries the risk of breaking bijectivity, an essential property of a normalizing flow. In order for the normalizing flow to retain this property, one has to restrict the samples to remain *inside* the canonical cell through a penalty term in the loss. While this works well for small systems, the canonical cell is expected to grow with the temporal extent N_t, thus reducing the advantage from canonicalizing the prior distribution for a larger number of time-slices. In future work, we will focus on finding a rigorous way to preserve bijectivity of the normalizing flow while incorporating symmetries, in order to use this framework to scale to larger systems. Moreover, we will investigate the performance of normalizing flows for the Hubbard model in different physical regimes, i.e., for various values of the coupling parameters of the theory.

Acknowledgements

References

[1] J.-L. Wynen, E. Berkowitz, C. Körber, T.A. Lähde and T. Luu, *Avoiding ergodicity problems in lattice discretizations of the Hubbard model*, *Phys. Rev. B* **100** (2019) 075141 [1812.09268].

[2] F. Temmen, E. Berkowitz, A. Kennedy, T. Luu, J. Ostmeyer and X. Yu, *Overcoming Ergodicity Problems of the Hybrid Monte Carlo Method using Radial Updates*, *PoS* **LATTICE2024** (2025) 068 [2410.19148].

[3] D. Rezende and S. Mohamed, *Variational Inference with Normalizing Flows*, in *Proceedings of the 32nd International Conference on Machine Learning*, vol. 37, pp. 1530–1538, PMLR, 2015 [1505.05770].

[4] I. Kobyzev, S.J. Prince and M.A. Brubaker, *Normalizing Flows: An Introduction and Review of Current Methods*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43** (2021) 3964 [1908.09257].

[5] G. Papamakarios, E. Nalisnick, D.J. Rezende, S. Mohamed and B. Lakshminarayanan, *Normalizing Flows for Probabilistic Modeling and Inference*, *J. Mach. Learn. Res.* **22** (2021) [1912.02762].

[6] A. van den Oord, N. Kalchbrenner and K. Kavukcuoglu, *Pixel Recurrent Neural Networks*, in *Proceedings of The 33rd International Conference on Machine Learning*, vol. 48 of *Proceedings of Machine Learning Research*, pp. 1747–1756, PMLR, 20–22 Jun, 2016 [1601.06759].

[7] A. van den Oord, N. Kalchbrenner, L. Espeholt, K. Kavukcuoglu, O. Vinyals and A. Graves, *Conditional Image Generation with PixelCNN Decoders*, in *Advances in Neural Information Processing Systems*, vol. 29, pp. 4797–4805, 2016 [1606.05328].

[8] F. Noé, S. Olsson, J. Köhler and H. Wu, *Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning*, *Science* **365** (2019) eaaw1147 [1812.01729].

[9] M.S. Albergo, G. Kanwar and P.E. Shanahan, *Flow-based generative models for Markov chain Monte Carlo in lattice field theory*, *Phys. Rev. D* **100** (2019) 034515 [1904.12072].

[10] K.A. Nicoli, C.J. Anders, L. Funcke, T. Hartung, K. Jansen, P. Kessel et al., *Estimation of Thermodynamic Observables in Lattice Field Theories with Deep Generative Models*, *Phys. Rev. Lett.* **126** (2021) 032001 [2007.07115].

[11] M. Caselle, E. Cellini, A. Nada and M. Panero, *Stochastic normalizing flows as non-equilibrium transformations*, *JHEP* **07** (2022) 015 [2201.08862].

[12] K. Cranmer, G. Kanwar, S. Racanière, D.J. Rezende and P.E. Shanahan, *Advances in machine-learning-based sampling motivated by lattice quantum chromodynamics*, *Nature Reviews Physics* **5** (2023) 526 [2309.01156].

[13] D. Wu, L. Wang and P. Zhang, *Solving Statistical Mechanics Using Variational Autoregressive Networks*, *Phys. Rev. Lett.* **122** (2019) 080602 [1809.10606].

[14] K.A. Nicoli, S. Nakajima, N. Strodthoff, W. Samek, K.-R. Müller and P. Kessel, *Asymptotically unbiased estimation of physical observables with neural samplers*, *Phys. Rev. E* **101** (2020) 023304 [1910.13496].

[15] M. Caselle, E. Cellini and A. Nada, *Sampling Nambu-Goto theory using Normalizing Flows*, *PoS* **LATTICE2023** (2024) 015 [2309.14983].

[16] M. Caselle, E. Cellini and A. Nada, *Sampling the lattice Nambu-Goto string using Continuous Normalizing Flows*, *JHEP* **02** (2024) 048 [2307.01107].

[17] N. Gebauer, M. Gastegger and K. Schütt, *Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules*, in *Advances in Neural Information Processing Systems*, vol. 32, 2019 [1906.00957].

[18] N.W.A. Gebauer, M. Gastegger, S.S.P. Hessmann, K.-R. Müller and K.T. Schütt, *Inverse design of 3d molecular structures with conditional generative neural networks*, *Nature Commun.* **13** (2022) 973.

[19] P. Białas, P. Korcyl, T. Stebel and D. Zapolski, *Rényi entanglement entropy of a spin chain with generative neural networks*, *Phys. Rev. E* **110** (2024) 044116 [2406.06193].

[20] A. Bulgarelli, E. Cellini, K. Jansen, S. Kühn, A. Nada, S. Nakajima et al., *Flow-based Sampling for Entanglement Entropy and the Machine Learning of Defects*, *ArXiv e-prints* (2024) [2410.14466].

[21] S. Chen, O. Savchuk, S. Zheng, B. Chen, H. Stoecker, L. Wang et al., *Fourier-flow model generating Feynman paths*, *Phys. Rev. D* **107** (2023) 056001 [2211.03470].

[22] K.A. Nicoli, C.J. Anders, L. Funcke, T. Hartung, K. Jansen, P. Kessel et al., *Machine Learning of Thermodynamic Observables in the Presence of Mode Collapse*, *PoS* **LATTICE2021** (2022) 338 [2111.11303].

[23] K.A. Nicoli, C.J. Anders, T. Hartung, K. Jansen, P. Kessel and S. Nakajima, *Detecting and mitigating mode-collapse for flow-based sampling of lattice field theories*, *Phys. Rev. D* **108** (2023) 114501 [2302.14082].

[24] G. Kanwar, M.S. Albergo, D. Boyda, K. Cranmer, D.C. Hackett, S. Racanière et al., *Equivariant Flow-Based Sampling for Lattice Gauge Theory*, *Phys. Rev. Lett.* **125** (2020) 121601 [2003.06413].

[25] J. Köhler, L. Klein and F. Noé, *Equivariant flows: exact likelihood generative learning for symmetric densities*, in *International conference on machine learning*, pp. 5361–5370, PMLR, 2020 [2006.02425].

[26] J. Hubbard, *Electron correlations in narrow energy bands*, *Proc. Roy. Soc. A* **276** (1963) 238.

[27] D.P. Arovas, E. Berg, S.A. Kivelson and S. Raghu, *The Hubbard Model*, *Annu. Rev. Condens. Matter Phys.* **13** (2022) 239 [2103.12097].

[28] H.F. Trotter, *On the product of semi-groups of operators*, *Proc. Amer. Math. Soc.* **10** (1959) 545.

[29] M. Suzuki, *Relationship between d-Dimensional Quantal Spin Systems and (d+1)-Dimensional Ising Systems: Equivalence, Critical Exponents and Systematic Approximants of the Partition Function and Spin Correlations*, *Prog. Theor. Phys.* **56** (1976) 1454.

[30] J. Hubbard, *Calculation of Partition Functions*, *Phys. Rev. Lett.* **3** (1959) 77.

[31] R.C. Brower, D. Schaich and C. Rebbi, *Hybrid Monte Carlo simulation on the graphene hexagonal lattice*, *PoS* **Lattice 2011** (2012) 056 [1204.5424].

[32] M.V. Ulybyshev, P.V. Buividovich, M.I. Katsnelson and M.I. Polikarpov, *Monte Carlo Study of the Semimetal-Insulator Phase Transition in Monolayer Graphene with a Realistic Interelectron Interaction Potential*, *Phys. Rev. Lett.* **111** (2013) 056801 [1304.3660].

[33] D. Smith and L. von Smekal, *Monte Carlo simulation of the tight-binding model of graphene with partially screened Coulomb interactions*, *Phys. Rev. B* **89** (2014) 195429 [1403.3620].

[34] T. Luu and T.A. Lähde, *Quantum Monte Carlo calculations for carbon nanotubes*, *Phys. Rev. B* **93** (2016) 155106 [1511.04918].

[35] S. Duane, A.D. Kennedy, B.J. Pendleton and D. Roweth, *Hybrid Monte Carlo*, *Phys. Lett. B* **195** (1987) 216.

[36] L. Dinh, D. Krueger and Y. Bengio, *NICE: non-linear independent components estimation*, in *3rd International Conference on Learning Representations (Workshop Track Proceedings)*, 2015 [1410.8516].

[37] L. Dinh, J. Sohl-Dickstein and S. Bengio, *Density estimation using Real NVP*, in *International Conference on Learning Representations*, 2017 [1605.08803].

[38] S. Kullback and R.A. Leibler, *On Information and Sufficiency*, *Annals Math. Stat.* **22** (1951) 79.

[39] D. Boyda, G. Kanwar, S. Racanière, D.J. Rezende, M.S. Albergo, K. Cranmer et al., *Sampling using* SU(N) *gauge equivariant flows*, *Phys. Rev. D* **103** (2021) 074504 [2008.05456].

[40] K.A. Nicoli, C.J. Anders, L. Funcke, K. Jansen, S. Nakajima and P. Kessel, *NeuLat: a toolbox for neural samplers in lattice field theories*, *PoS* **LATTICE2023** (2024) 286.