

Machine learning to alleviate Hubbard-model sign problems

Jan-Lukas Wynen¹, Evan Berkowitz^{2,1}, Stefan Krieg^{1,3,4}, Thomas Luu^{1,5,4} and Johann Ostmeyer⁴

¹*Institute for Advanced Simulation, Forschungszentrum Jülich, 54245 Jülich, Germany*

²*Maryland Center for Fundamental Physics, University of Maryland, College Park, Maryland 20742, USA*

³*JARA-HPC, Jülich Supercomputing Center, Forschungszentrum Jülich, 54245 Jülich, Germany*

⁴*Helmholtz-Institut für Strahlen- und Kernphysik, Rheinische Friedrich-Wilhelms-Universität Bonn, 53012 Bonn, Germany*

⁵*Institut für Kernphysik, Forschungszentrum Jülich, 54245 Jülich, Germany*



(Received 13 July 2020; revised 12 March 2021; accepted 15 March 2021; published 25 March 2021)

Lattice Monte Carlo calculations of interacting systems on nonbipartite lattices exhibit an oscillatory imaginary phase known as the phase or sign problem, even at zero chemical potential. One method to alleviate the sign problem is to analytically continue the integration region of the state variables into the complex plane via holomorphic flow equations. For asymptotically large flow times, the state variables approach manifolds of constant imaginary phase known as Lefschetz thimbles. However, flowing such variables and calculating the ensuing Jacobian is a computationally demanding procedure. In this paper, we demonstrate that neural networks can be trained to parametrize suitable manifolds for this class of sign problem and drastically reduce the computational cost for different severely afflicted small volume systems. In particular, we apply our method to the Hubbard model on the triangle and tetrahedron, both of which are nonbipartite. At strong interaction strengths and modest temperatures, the tetrahedron suffers from a severe sign problem that cannot be overcome with standard reweighting techniques, while it quickly yields to our method. We benchmark our results with exact calculations and comment on future directions of this work.

DOI: [10.1103/PhysRevB.103.125153](https://doi.org/10.1103/PhysRevB.103.125153)

I. INTRODUCTION

Lattice field theories allow for a first-principles construction of nonperturbative interacting quantum field theories. Beyond being a mathematical footing, they provide a computational strategy for solving such systems numerically, typically via Markov-chain Monte Carlo (MCMC). However, for this numerical approach to succeed, a Euclidean field theory requires a real-valued action, providing a positive-definite integration measure. When the action is complex, additional steps must be taken, because the integrand can oscillate wildly. Hence, intricate cancellations are required for numerical estimates to yield accurate results, thereby rendering otherwise successful lattice methods powerless.

This oscillating complex phase problem (“sign problem”) is prevalent in many areas of computational physics that rely on lattice stochastic methods to tackle nonperturbative phenomena. Lattice quantum chromodynamics (LQCD) calculations at finite baryon chemical potential [1,2], for example, suffer from a sign problem, which precludes any numerical investigation of quark matter in dense astrophysical objects such as neutron stars and supernovae [3–5]. Furthermore, including the strong- θ term [6] in the QCD action induces a sign problem as well. Without the development of methods to alleviate the sign problem, one is forced to assume the affected terms are (perturbatively) small [1,7,8].

In nuclear lattice effective field theory (NLEFT) [9–11], where nucleons are degrees of freedom as opposed to quarks, the sign problem prevents lattice MCMC studies of very

neutron-rich nuclei. These nuclei, near the neutron drip line, play an important role in today’s nuclear reactors and in the industrial and astrophysical synthesis of heavy elements.

A wide variety of condensed-matter systems, including the doped fermionic Hubbard model, exhibit a rich multi-quasiparticle spectrum (see, for example, [12–15]) and comprise tantalizing theoretical systems with commercial relevance. Unfortunately, many of these systems exhibit a sign problem as well. Indeed, the sign problem poses a major stumbling block for MCMC studies in all computational subdisciplines of physics.

Finding a general solution to the phase problem with polynomial scaling in the severity of the problem is NP-hard [16]. However, techniques that take advantage of a particular model’s structure may still be achievable. Therefore various strategies for alleviating the sign problem have been developed.

The most obvious and widespread of these is *reweighting*, which we describe in more detail Sec. II C. Reweighting can be applied when the sign problem is mild, but can fail spectacularly when the problem is severe, much as perturbation theory fails when the interactions become strong.

Another commonly used strategy is analytic continuation. For example, in QCD one may simulate with purely imaginary baryon chemical potential, removing the complex phase completely (see, for example, Refs. [17–19]). Here, the uncertainty lies in the functional form chosen to analytically continue results back to the real axis, which is not known *a priori* and thus relies on model assumptions. Moreover,

analytically continuing Monte Carlo data with uncertainties is no easy task.

Alternatively, Complex Langevin methods do away completely with MCMC and have had various degrees of success (see, for example, Refs. [20–23], and more recently, [24]). Unfortunately, there currently seems to be no consensus on which systems have Langevin methods that are guaranteed to be correct and which do not, though progress has been made in quantifying certain conditions for success [25].

Recently, tensor networks [26–28] have shown promise in tackling many-body systems in both one and two dimensions. Their formalism is agnostic to the presence of a chemical potential. First results are available for fermionic systems [29], such as spinless fermions on the honeycomb lattice [30] and the Hubbard model on a square lattice [31]. It remains to be seen, however, whether such calculations are preferable in terms of precision, scalability, and computational complexity, which also applies to methods that involve direct integration over the group manifold providing polynomial exactness [32].

The method we leverage here is related to integration on Lefschetz thimbles. By analytically continuing the integration variables into the complex plane [33], one can locate higher-dimensional steepest-descent analogues called *Lefschetz thimbles* for each critical point.¹ On a thimble the phase is not oscillatory, but constant, up to the residual phase of the Jacobian (which can reintroduce wild oscillations if the thimble is strongly curved [34]). Each thimble has its own constant phase; lattice Monte Carlo on a given thimble can be performed because the phase is global and can be factored out of the integral. A combination of all thimbles which can be reached by *holomorphic flow* yields the original integral over the real variables, but finding those thimbles (or equivalently, identifying their critical points) is usually difficult.

Early investigations with this method looked at bosonic gauge systems in low dimensions [35–37]. The Thirring model—nonrelativistic fermions with a chemical potential in one dimension—was studied using only one “main” thimble [38–40] before yielding to a flow-based method [41–43] that approaches a thimble in the limit of long flow times. Moreover, Lefschetz thimbles have found use in higher dimensions, including the 1+1-dimensional Thirring model [44] and small examples of the doped 2+1-dimensional Hubbard model [45]. Gauge symmetry complicates the story but the efficacy of Lefschetz thimbles in gauge theories is a field of active research [46–49]. Beyond Lefschetz thimbles one may find sign-optimized manifolds where even the residual Jacobian phase is handled cleanly [50–52].

The most difficult aspect of attempting a Lefschetz decomposition is that the thimbles’ locations and shapes are not generally known *a priori*, and that their determination is a complicated and numerically intensive endeavor, especially as systems get larger in higher dimensions. Even when locating critical points is straightforward, deciding whether their

thimbles must be included to reproduce the integral of interest is not.

Numerical difficulties abound, as well. For example, the calculation of the Jacobian associated with the transformation from the real plane to these complex manifolds and its determinant can be numerically prohibitive as the system size becomes larger. These issues make an exact Lefschetz Thimble decomposition potentially as difficult as the original sign problem.

Instead one can approximate the thimbles by use of holomorphic flow equations. Rather than completely solving the sign problem, this strategy is to merely alleviate it [41,53]. Moreover, numerical techniques to estimate the determinant of the Jacobian in an efficient manner fix scaling problems [54]. Still, the determination of these approximate manifolds is less computationally intensive than determining the exact manifolds, especially since they live in high-dimensional spaces.

We use neural networks to define a complex manifold, or *learnifold*, instead. This can be done by approximating holomorphic flow [55] which is possible (at least for large networks and training data) because of the universal approximation theorem [56,57] and because the flowed manifolds are continuous and smooth. Evaluating a neural network is typically very fast, making them ideal candidates for modeling holomorphic flow in Monte Carlo.

In this paper we study the Hubbard model on small nonbipartite lattices which suffer from a severe phase problem by incorporating a learnifold into HMC. These small systems afford exact numerical solutions with which we can benchmark our method. Even though standard reweighting techniques are completely ineffective for some of these problems, we find that the learnifold-HMC allows us to extract correlation functions well, reproducing exact results. Unfortunately the computational demands of evaluating the determinant of the induced Jacobian presently precludes us from simulating larger systems. Thus, strictly speaking, our method only demonstrates that the sign problem can be alleviated sufficiently for these systems so as to obtain statistically correct observables. However, it is encouraging that our method produces correct results even in cases where the sign problem is too severe for standard reweighting techniques; the severity of the sign problem scales more favorably as we take the zero-temperature limit with our method than with standard reweighting. Indeed, in the future we plan to leverage this technique to study larger fullerene systems, such as buckyballs. We anticipate a straightforward applicability of our method to doped systems as well.

Our paper is organized as follows. In Sec. II, we describe our formalism in detail. In Sec. III, we then describe our algorithm, giving details about how we incorporate machine learning into HMC without compromising the algorithm’s exactness. In Sec. IV, we leverage our method, and show that it reproduces exact results for a number of different observables on some simple systems that may be solved exactly, even when standard reweighting might fail. We point out that there is generally a trade-off between ergodicity and the sign problem, and we demonstrate that the exponential decay of the statistical power is much less steep with our method. Finally, we give some conclusions in Sec. V.

¹In this context, a critical point is defined as a point in field space where the derivative of the action with respect to the (complex) field vanishes, see (22).

II. FORMALISM

In this section, we introduce the Hamiltonian used in our studies and discuss the consequences of having a nonbipartite lattice. We then explain *reweighting*—an exact method for handling complex actions. Finally, we discuss correlation functions and mention the operators we use to construct them.

A. The Hamiltonian

We use the Hubbard model in the particle/hole basis [58–62] to perform simulations. The Hubbard model consists of a tight-binding Hamiltonian,

$$H_0 = - \sum_{x,y} (a_{x,\uparrow}^\dagger h_{xy} a_{y,\uparrow} + a_{x,\downarrow}^\dagger h_{xy} a_{y,\downarrow}), \quad (1)$$

where h_{xy} is nonzero if x and y are nearest neighbors, coupled with an onsite interaction of the form

$$H = H_0 - \frac{U}{2} \sum_x (n_{x,\uparrow} - n_{x,\downarrow})^2, \quad (2)$$

where the number operator $n_{x,s} \equiv a_{x,s}^\dagger a_{x,s}$ counts electrons of spin s at position x . We now change to the hole basis for the spin- $\downarrow$ electrons,

$$b_{x,\downarrow}^\dagger \equiv a_{x,\downarrow}, \quad b_{x,\downarrow} \equiv a_{x,\downarrow}^\dagger, \quad (3)$$

which gives, up to an irrelevant constant,

$$H = - \sum_{x,y} (a_x^\dagger h_{xy} a_y - b_x^\dagger h_{xy} b_y) + \frac{U}{2} \sum_x \rho_x^2, \quad (4)$$

$$\rho_x = n_x^a - n_x^b, \quad (5)$$

where $n_x^a = a_x^\dagger a_x$ counts the number of (spin- $\uparrow$) particles $n_x^b = b_x^\dagger b_x$ counts the number of spin- $\downarrow$ holes at site x ; we use the convention of positively-charged particles.

B. Nonbipartite lattices

Bipartite graphs are those that admit a two-coloring, such that no vertex has a neighbor of the same color. Examples include the standard square lattice (consisting of two underlying square lattices), the honeycomb lattice (consisting of two underlying triangular lattices), or simply two connected sites. A nonbipartite graph, in contrast, cannot be so colored.

Graphs with odd-length cycles are not bipartite. Examples include fullerene refinements of a 2-sphere, such as the buckyball (C_{60}) or the dodecahedron (C_{20}). In this case the presence of 12 pentagonal faces (required to make the geometry closed) destroys the bipartiteness. The simplest, nontrivial nonbipartite graph is a single triangle, which is small enough for exact diagonalization. The tetrahedron, topologically the complete graph on four vertices, is similarly tractable but not bipartite. The Hubbard model has been studied on nonbipartite quasi-one-dimensional chains [63–65] and has been exactly solved on small clusters [66] because solutions on such chains and clusters may be taken as input data for many-body methods. The discovery of unusual heavy-fermion behavior of LiV_2O_4 triggered direct studies of the Hubbard model on the pyrochlore lattice with tetrahedral unit cells [67] and other extended nonbipartite structures (for example, see Refs. [68–70]).

As discussed in Ref. [62], when the Hubbard model is formulated on a bipartite graph, its formulation as a lattice field theory exhibits special features. Discretizing the Euclidean time β into N_t time slices yields a temporal lattice spacing

$$\delta = \frac{\beta}{N_t} \quad (6)$$

and we denote quantities made dimensionless with factors of δ with a tilde, so that $\tilde{U} = U\delta$. The partition function can be cast into the form of a path integral [58–61] yielding

$$\mathcal{Z} = \int \left[\prod_{x,t} d\phi_{xt} \right] e^{-\beta H[\phi]} = \int \left[\prod_{x,t} d\phi_{xt} \right] W[\phi] \quad (7)$$

$$W[\phi] = \det(M[\phi, h] M^\dagger[\phi, -h]) \exp \left(-\frac{1}{2\tilde{U}} \sum_{x,t} \phi_{xt}^2 \right), \quad (8)$$

where we assume $\tilde{U} > 0$ and the fermion matrix is in the *exponential discretization* (see Ref. [62] for a comparison with other discretizations),

$$M[\phi, h]_{x't',xt} = \delta_{x',x} \delta_{t',t} - [e^{\tilde{h}}]_{x',x} e^{i\phi_{xt}} \mathcal{B}_{t'} \delta_{t',t+1}, \quad (9)$$

where $\tilde{h}$ is the dimensionless hopping matrix and $\mathcal{B}_t = +1$ for $0 < t < N_t$ and $\mathcal{B}_0 = -1$ explicitly encodes antiperiodic temporal boundary conditions.

In the case of bipartite lattices, the particle-hole transformation (3) can be modified to include an additional sign

$$b_{x,\downarrow}^\dagger \equiv \mathcal{P}_x a_{x,\downarrow}, \quad b_{x,\downarrow} \equiv \mathcal{P}_x a_{x,\downarrow}^\dagger, \quad (10)$$

where $\mathcal{P}_x$ is the parity of the sublattice, so that we perform a site-dependent sign flip— $\mathcal{P}_x = +1$ if x is on one sublattice, -1 if on the other—for holes. This flips the sign of the hopping term in the hole matrix such that the weight becomes

$$W[\phi] = \det(M[\phi, h] M^\dagger[\phi, h]) \exp \left(-\frac{1}{2\tilde{U}} \sum_{x,t} \phi_{xt}^2 \right) \quad (11)$$

as shown in, for example, Refs. [58–61]. Since $MM^\dagger$ is positive semidefinite, W is real and positive-semi-definite as well.² Such systems are thus easily amenable to standard Monte Carlo simulations.

On nonbipartite lattices, the signed particle-hole transformation (10) does not exist, as the bipartitioning fails, so we cannot apply it and the weight (8) is of indefinite sign. The next section describes how Monte Carlo techniques can nonetheless be applied in this case and discusses the problems such calculations typically face.

We restrict our attention to cases where the hopping matrix h is always just a constant times the graph's adjacency matrix,

$$h_{xy} = \kappa \delta_{(x,y)} \quad (12)$$

though this assumption could be relaxed to model realistic molecules where bond lengths and corresponding hopping strengths vary. In general, the symmetry of the underlying

²In the case of nonzero chemical potential, W is complex-valued even for bipartite lattices. This has been investigated recently for small bipartite systems in the context of holomorphic flow [45,71].

lattice should be considered when determining the Hamiltonian. For example, the regular dodecahedron is vertex- and edge-transitive, so it is natural to assign uniform interaction strengths to every site and uniform hopping strengths along every edge. In contrast, while the truncated icosahedron (C_{60} buckyball) is vertex-transitive and thus every site should have the same interaction strength, it is not edge-transitive: some edges separate two hexagons, while others separate a hexagon from a pentagon and these two different kinds of edges could have different hopping strengths (which physically reflects the fact that the bond lengths differ). In larger fullerenes with I_h symmetry, there are a wider variety of bond lengths, even between hexagonal faces, because some hexagons are closer to or farther from pentagons; in fullerenes with smaller symmetry groups (such as C_{70} , which enjoys a D_{5h} symmetry) we can naturally incorporate its structure by adjusting hopping strengths in h and adjusting U from site to site in a way that respects its symmetry. Since we here are interested in proof-of-principle work we ignore all effects of this kind.

C. Reweighting

The expectation value of an observable $\hat{O}$ is given by

$$\langle \hat{O} \rangle = \frac{1}{\mathcal{Z}} \int \mathcal{D}\phi \hat{O}[\phi] e^{-S[\phi]}, \quad (13)$$

where S is the action, ϕ the fields, and $\mathcal{Z}$ the integral without the operator. For the purpose of this discussion we include the fermionic determinants in the action S by taking the log and casting them up into the exponential.

When the action is real-valued we can construct a Monte Carlo method, sampling configurations of the field ϕ according to their probability given by the Boltzmann weight $\exp(-S)/\mathcal{Z}$ (“importance sampling”). Generated in such a way, we can estimate the expectation value in (13) by computing the mean of the observable measured separately on each configuration.

When the action is complex-valued the Boltzmann weight is complex as well, and does not directly provide a probability distribution. *Reweighting* is an exact, straightforward procedure by which we can overcome this difficulty, given sufficient computational resources. Rather than sampling according to the action, one samples according to the real part of the action S^R and incorporates the phase associated with the imaginary part of the action S^I into each observable, estimating

$$\langle \hat{O} \rangle = \frac{\langle \hat{O} e^{-iS^I} \rangle_R}{\langle e^{-iS^I} \rangle_R} \approx \frac{\sum_j \hat{O}[\phi_j] e^{-iS^I[\phi_j]}}{\sum_j e^{-iS^I[\phi_j]}}, \quad (14)$$

where j runs over the ensemble and the R -subscripted angle brackets indicate an expectation value with respect to the real part of the action only.³

If the phase given by the imaginary part of the action is constant or narrowly distributed, this procedure can successfully estimate observables. However, if the phase is widely

distributed or, in the worst case, evenly covers the unit circle, the expectation value in the denominator nears zero and reweighting becomes computationally intractable. We call the absolute value of the denominator the *statistical power*

$$\Sigma = |\langle e^{i\theta} \rangle_R|, \quad \theta \equiv \arg W. \quad (15)$$

When all configurations have the same imaginary action, $\Sigma = 1$ and each configuration is valuable. When $\theta[\phi]$ varies strongly for different field configurations ϕ , Σ is near zero and the configurations tend to cancel; each additional configuration contributes only marginally to the expectation value. *Quenching* the phase, that is, estimating expectation values by the uncontrolled approximation $\langle \hat{O} \rangle \simeq \langle \hat{O} \rangle_R$, can lead to a dramatic distortion of observables.

D. Correlation functions

Two-point correlation functions between two operators $\hat{O}_x$ and $\hat{O}_y^\dagger$ at different times and sites x, y

$$\begin{aligned} C_{xy}(\tau) &= \frac{1}{\mathcal{Z}} \text{tr} \left[e^{-\beta H} \frac{1}{N_t} \sum_t \hat{O}_x(t + \tau) \hat{O}_y^\dagger(t) \right] \\ &= \frac{1}{N_t} \sum_t \langle \hat{O}_x(t + \tau) \hat{O}_y^\dagger(t) \rangle \end{aligned} \quad (16)$$

can be computed as a function of temporal separation by solving for propagators on each configuration and tying them together in the required Wick contractions. They admit spectral decompositions

$$\begin{aligned} C_{xy}(\tau) &= \frac{1}{\mathcal{Z}} \text{tr} \left[e^{-\beta H} \frac{1}{N_t} \sum_t \hat{O}_x(t + \tau) \hat{O}_y^\dagger(t) \right] \\ &= \frac{1}{N_t} \sum_t \frac{1}{\mathcal{Z}} \text{tr} [e^{-\beta H} e^{+H(t+\tau)} \hat{O}_x e^{-H(t+\tau)} e^{+Ht} \hat{O}_y^\dagger e^{-Ht}] \\ &= \frac{1}{\mathcal{Z}} \text{tr} [e^{-(\beta-\tau)H} \hat{O}_x e^{+H(t+\tau)} e^{-Ht} \hat{O}_y^\dagger] \\ &= \frac{1}{\sum_a e^{-E_a\beta}} \sum_{bc} e^{-E_b\beta} e^{-(E_c-E_b)\tau} z_{axb} z_{by^\dagger a} \end{aligned} \quad (17)$$

where we moved to the Heisenberg picture and inserted resolutions of the identity in the energy eigenbasis and defined the overlap factors

$$z_{axb} = \langle a | \hat{O}_x | b \rangle, \quad z_{by^\dagger a} = \langle b | \hat{O}_y^\dagger | a \rangle = z_{ayb}^*, \quad (18)$$

where a and b label energy eigenstates. In the low temperature limit $\beta \rightarrow \infty$, the sum is dominated by the lowest energy state; otherwise thermal artifacts may be seen. In the low temperature and late-(Euclidean-)time limit $\tau \rightarrow \infty$, the correlator’s τ dependence gives the energy gap between the ground and the first excited state (with the observable’s quantum numbers), as all the heavier states decay more quickly.

Hence, we can extract the single-particle/single-hole spectrum by, e.g., setting $\hat{O}_x^\dagger = a_x^\dagger$. $C_{xy}(\tau)$ is quadratic in the volume. However, after diagonalization one obtains only $O(V)$ single-particle eigenfunctions. For the small lattices we study here, it suffices to project x and y to the same irreducible representation of the lattice symmetry in order to diagonalize C .

³Unless otherwise mentioned, uncertainties presented here via reweighting come from a correlated bootstrap procedure, where, on each bootstrap resampling, the numerator and denominator are both measured and divided.

We can also calculate correlation functions between composite operators at additional computational expense. The composite operators we consider are the number operators n_x^a and n_x^b , their sum the total number operator n_x , their difference the charge operator ρ_x , spin raising and lowering operators $S_x^\pm$, and the spin operators S_x^i (where i runs over all three spatial directions); we also consider the two doubly charged local bilinears. In Appendix A, we detail the operators, the correlation functions we measure and how to construct conserved quantities from them.

III. ALGORITHM

We gave an extensive overview of our application of Hybrid Monte Carlo (HMC) to the Hubbard model in Ref. [62]. In all our reweighting-only examples, we run HMC in a standard way, performing the Metropolis accept-reject step according to the real part of the action. In the rest of this section, we detail how we incorporate learnifolds into HMC.

In Sec. III A, we provide a summary of why integrating over a manifold given by holomorphic flow is advantageous for reducing the sign problem. Then in Sec. III B, we show how we use machine learning to quickly compute the learnifold, over which we will integrate. Afterwards we explain how to incorporate the learnifold into HMC and, finally, comment on our update scheme's ergodicity in Sec. III C.

A. Holomorphic flow

Holomorphic flow is a generalization of the steepest descent method to multi-dimensional complex space. Given a holomorphic functional, in our case the action S , of the N -dimensional complex ϕ , the flow equations are

$$\frac{d\phi}{d\tau_f} = \pm \left(\frac{\partial S[\phi]}{\partial \phi} \right)^*, \quad (19)$$

where τ_f is the *flow time*. A minus sign indicates *downward* flow, while a plus sign *upward* flow. Splitting the components $\phi_i = \phi_i^R + i\phi_i^I$ and the action $S[\phi] = S^R[\phi] + iS^I[\phi]$ into their real and imaginary parts, yields

$$\frac{d\phi^R}{d\tau_f} = \pm \frac{\partial S^R}{\partial \phi_i^R} = \pm \frac{\partial S^I}{\partial \phi_i^I}, \quad (20)$$

$$\frac{d\phi^I}{d\tau_f} = \pm \frac{\partial S^R}{\partial \phi_i^I} = \mp \frac{\partial S^I}{\partial \phi_i^R}, \quad (21)$$

which are the Cauchy-Riemann equations. The equations containing S^R essentially implement the gradient flow, whereas the other equations are Hamilton's equations for the imaginary part of the action. That is, S^I remains a constant of motion during the flow. It is easily seen that downward holomorphic flow is the generalization of gradient flow, or steepest descent, for real fields.

The critical points $\hat{\phi}_{\text{cr}}$ of $S[\phi]$ are vectors of complex numbers and satisfy

$$\left. \frac{\partial S[\phi]}{\partial \phi} \right|_{\hat{\phi}_{\text{cr}}} = 0. \quad (22)$$

The $\hat{\phi}_{\text{cr}}$ are saddle points; the associated *Lefschetz thimble* is the manifold in $\mathbb{C}^N$ which flows to the given critical point

under downward flow, while the *dual thimble* flows to the critical point under upward flow. Models with many variables have many critical points and associated thimbles that must be integrated over to produce the same result as the integral on the real manifold.

Flowing the integration region of $\phi \in \mathbb{R}^N$ to the manifold(s) $\hat{\phi}$ eliminates the sign problem since S^I is constant on each manifold. Because there are no poles in the integration kernel, Cauchy's integral formula guarantees that the integral over the thimbles $\hat{\phi} \in \mathbb{C}^N$ will be exactly equal to the original integral over $\phi \in \mathbb{R}^N$,

$$\int \mathcal{D}\phi e^{-S[\phi]} = \sum_{\sigma} e^{-iS^I[\hat{\phi}_{\text{cr},\sigma}]} \int \mathcal{D}\hat{\phi}_{\sigma} e^{-S^R[\hat{\phi}_{\sigma}]}, \quad (23)$$

where σ runs over included thimbles and $\hat{\phi}_{\text{cr},\sigma}$ is the critical point associated with this thimble. The number of thimbles, their critical points, and their relative weights depend on the lattice action and the original integration region (typically $\mathbb{R}^N$). Not all the thimbles in $\mathbb{C}^N$ are included in the sum; only those that are required to preserve the integral's homology class.

Thimbles do not cross each other, but are connected at places where the action diverges so that the integration kernel vanishes. The Gaussian part of the action diverges when $|\hat{\phi}| \rightarrow \infty$; the fermionic part of the action diverges when $\det(M[\hat{\phi}]M[-\hat{\phi}]) = 0$. We call these zero-weight places in the complex space *neverland*, as they never appear in an importance-sampling scheme.

Because the locations of the needed critical points and their associated thimbles are not known *a priori*, their determination requires extensive numerical resources; an analytic determination would amount to a solution of the lattice model. Instead we follow Alexandru, Basar, and Bedaque [42] and use the fact that the thimbles are fixed points of the flow to our advantage: we flow only a modest amount to get an integration manifold that approaches the set of relevant thimbles, not solving the sign problem entirely but alleviating it to the point where standard reweighting techniques are sufficient to address any remaining sign problem.

To ensure our method remains exact, it is important that our ultimate integration manifold is in the same homology class as the original. Because the holomorphic flow preserves the homology class, the manifold resulting from any finite flow time is in the right class. Moreover, because the flow's fixed points are thimbles, a finite-flow manifold approaches those thimbles that contribute to the integral [42]; we need not analytically decide which thimbles to include—the flow discovers this automatically.

One point on the original manifold flows to a thimble's critical point, and only a vanishing neighborhood around that point flows to the rest of the thimble—most of the original manifold flows to *neverland*—the place where thimbles meet and vanish in the integration kernel, typically due to zeros of the fermion determinant. These zeros act as attractors, and most configurations flow to these zeros after a finite amount of flow time [40,42,72]. Because the thimbles must meet at these zeros, attempting an update method like HMC on the thimbles themselves in a naive way will be obstructed by these zeros, causing an ergodicity problem. However, by restricting

ourselves to modest flow times, we can ensure that the manifold will not touch these zeros. We provide some visual evidence of ergodicity in Sec. IV A 1 and reproduce a wide range of exactly known correlation functions. So, deciding how much to flow is a balancing act—one hopes to flow enough that the sign problem is alleviated but not so much that an ergodicity problem emerges.

Any transformation of integration variables, including the one provided by the flow, comes with an associated Jacobian, and this Jacobian must be included in the Monte Carlo weight (or incorporated by reweighting) for a correct method. There are flow equations for the Jacobian as well [41],

$$\frac{dJ}{d\tau_f}(t) = [H[\phi(t)]J(t)]^*, \quad (24)$$

where H is the Hessian,

$$H_{ij}[\phi] \equiv \left. \frac{\partial^2 S}{\partial \phi_i \partial \phi_j} \right|_{\phi} = (\text{tr}[M^{-1}(\partial_i M)M^{-1}(\partial_j M)] - \text{tr}[M^{-1}\partial_i \partial_j M])|_{+\kappa, +\phi} + (\kappa, \phi \rightarrow -\kappa, -\phi) + \frac{\delta_{ij}}{\bar{U}}. \quad (25)$$

Flowing the Jacobian is the most time-consuming aspect of this calculation and thus the reason for considering neural networks, as we discuss in the following section.

To perform the numerical integration of the flow equations (19), and when necessary of the Jacobian (24), we use a fourth order adaptive Runge-Kutta method. Because the flow preserves the imaginary part of the action, we monitor the latter during the numerical integration and adjust the integration stepsize to keep deviations within a prescribed tolerance.

B. Networks

1. Architecture

Here, we only briefly describe the networks that we used. Interested readers can find a wealth of literature on neural networks online, see for instance [73] for an excellent introduction. We use feed-forward neural networks of dense layers to tackle the sign problem. Like Refs. [52,55] these networks produce the imaginary part of a field configuration from the real part and keep the latter fixed. The transformation into complex space is thus

$$\tilde{\phi} = \phi + i \text{NN}(\phi), \quad (26)$$

where NN is the neural network. In addition to the ability of the network to avoid dealing with complex numbers directly, Ref. [55] points out two advantages this formulation enjoys over $\tilde{\phi} = \text{NN}(\phi)$: it might ameliorate the ergodicity problem that flowing can induce (for an enlightening discussion and illustration see their Fig. 2); and it tends to yield a more stable Jacobian (Fig. 3 of Ref. [55]).

We encode ϕ as a space-time vector and use neural networks with a single hidden layer, consisting of twice the number of neurons than the input and output layers. The hidden layer has a Softplus activation function

$$\text{Softplus}(x) = \ln(1 + \exp(x)), \quad (27)$$

and the output layer has none. We found this to be sufficient for all cases we tested. Wider or deeper networks as well as convolutional layers or different activation functions did not yield significantly better results. It remains to be seen whether this setup scales to larger spatial lattices.

Neural networks of the chosen architecture can be evaluated very efficiently, which is, unfortunately, only half of the story. We also need to compute the Jacobian determinant for the change of variables, which will appear in the transformed integral (35). Given (26), it is

$$\det J[\phi] = \det \left(\mathbb{1} + i \frac{\partial \text{NN}_i(\phi)}{\partial \phi_j} \right). \quad (28)$$

This operation is $\mathcal{O}((N_s N_t)^3)$ (cubic in the space-time volume), because both the determinant and matrix multiplications in the derivative need to be performed with a general algorithm for dense matrices. No speedup seems possible for vanilla dense networks as the weight matrices are unconstrained. We still find our network-based transformation to significantly outperform flow-based transformations in terms of run time, however. However, this approach does not scale to larger lattices. It is possible to improve on the scaling by estimating the Jacobian as in Ref. [55] which uses a different network that produces only one component of $\tilde{\phi}_l$ at a time. Alternatively, one can use coupling layers which were designed to have simple Jacobians. This requires complex valued networks in our case, however, as described in Appendix B.

We implemented the neural networks and training procedures using PYTORCH [74] and used ISLE [75] to implement Monte Carlo for the Hubbard model. Section III C describes the HMC scheme we used to accommodate neural networks.

2. Training procedure

We train our models using a supervised approach. To that end, we generate random real configurations ϕ and flow them upward according to Eq. (19) for a fixed flow time to $\hat{\phi}$. It is nontrivial how to generate useful data and we provide explicit details on our approach in Sec. III B 3. Once the data is generated, we use $\text{Re } \hat{\phi}$ as inputs and $\text{Im } \hat{\phi}$ as target outputs to train the networks. We do so by using the ADAM algorithm [76] to minimise a smooth L1 loss

$$\text{loss}(x, y) = \frac{1}{n} \sum_i \begin{cases} (x_i - y_i)^2/2, & \text{for } |x_i - y_i| < 1 \\ |x_i - y_i| - 1/2, & \text{otherwise} \end{cases} \quad (29)$$

as defined in PYTORCH [74].

We know our action has exact symmetries (see Ref. [62] for a listing), comprising transforms T that change field configurations but leave the action invariant, $S[T\phi] = S[\phi]$. A uniformly-flowed manifold exhibits many of those symmetries, the most obvious being the temporal and spatial translation symmetries. Naturally it is desirable that our neural networks preserve as many of these symmetries as possible.

Reference [55] accomplishes this by training a network that takes a whole space-time vector and produces only the flowed configuration at the space-time origin and using translational invariance to construct the other vector elements. This technique suffers a number of constraints, most notably, it requires

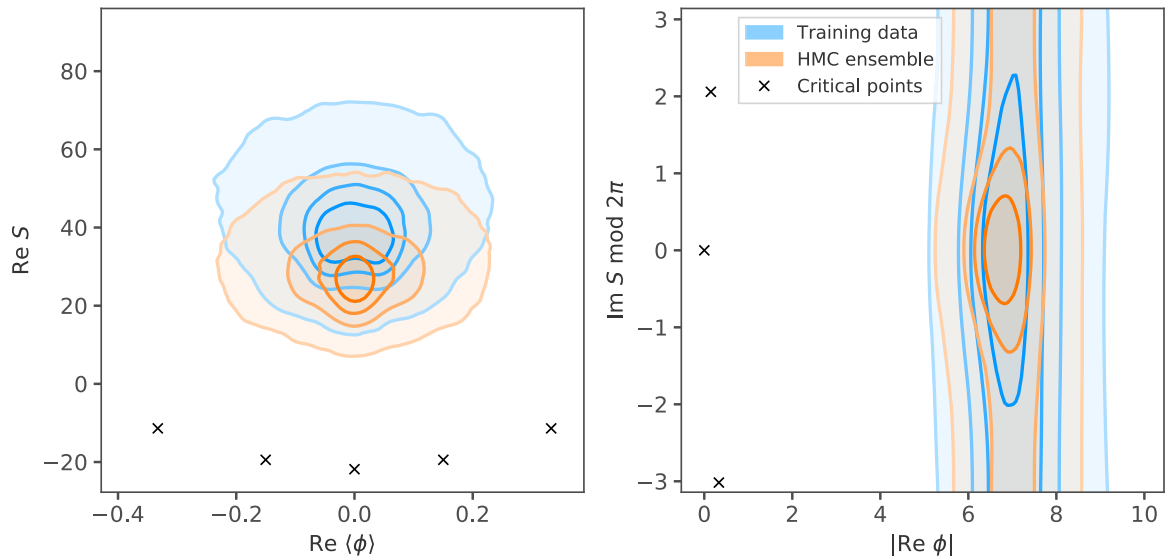


FIG. 1. Distributions of configurations used for training data (blue-shaded region) and an ensemble produced via HMC with a neural network (orange-shaded region) trained on this data. The system is a triangle with $N_t = 32$, $\kappa\beta = 6$, and $U/\kappa = 3$. The distributions were estimated from 10^5 training and 10^6 HMC configurations. The black crosses mark critical points of a space-time constant ϕ .

full translational invariance of the lattice. This is not the case for many interesting nonbipartite lattices, however, as they are not necessarily vertex-transitive. In such a case, a network with a single output is not enough because spatial symmetries cannot take every element to every other.⁴

As described above, our neural networks produce a full space-time vector as output and we encode symmetries by augmenting the training data. One symmetry that all lattices have is temporal translational invariance. We thus train our networks not only on the configurations generated according to Sec. III B 3 but also on all possible temporal shifts of the input and target output configurations. This effectively increases the size of the data set by a factor of N_t . The lattices considered in this work, triangle and tetrahedron, are vertex-transitive and we tried augmenting the training data by all possible spatial permutations but did not find an improvement. Adding temporal translations proves very useful for increasing model quality, however.

Every model was trained on minibatches of size 16 drawn from 1000 random configurations plus temporal translations. The only exception is the tetrahedron with $N_t = 32$ whose model was trained with 4000 configurations. We found this larger number necessary to train the network well. For larger N_t , 1000 configurations were enough for the tetrahedron which might be due to the higher number of added configurations from augmentation by temporal translations.

3. Training: data

Several different methods have been used to generate training data. Reference [52] uses HMC with a partially trained model to generate data for the next iteration of training. This is expensive as new configurations have to be generated ev-

ery time model hyperparameters are changed. Because it is self-reinforcing, it is possible that such an approach can find and remain in a local minimum or overfit a part of the phase space.

Reference [55] uses HMC without a neural network to generate training data before fitting a model. In order to achieve large phase space coverage, several MC ensembles with different temperatures were generated. This approach needs to generate data only once per parameter set and therefore allows for faster tuning of hyperparameters. It is still susceptible to autocorrelations, though, meaning that a large number of configurations might be needed for sufficient phase space coverage.

One of the goals of this work is to simplify the generation of training data in order to develop a more scalable approach.

One thimble of the Hubbard model on nonbipartite lattices is easy to find—the image of $\text{Re } \phi \equiv 0$ under the flow. It connects to a critical point $\hat{\phi}_{\text{cr}} \equiv ic$ for some U , β , and N_t dependent value $c < 0$. In concordance with literature, we call this the “main” thimble. There is an ongoing discussion whether a single thimble Ref. [33] or multiple thimbles are required to solve the integral [40–43]. To be on the safe side, we, therefore, identified additional thimbles, all with a space-time constant $\hat{\phi}_{\text{cr}} = z \in \mathbb{C}$, but close to the main thimble. Fig. 1 shows the critical points for a sample system.

We produced training data by first generating random field configurations close to the critical points that we found and then flowing them for a short, fixed flow time. In order to reduce the required flow time, we sampled the initial configurations on the plane tangent to the main thimble in the critical point ($\phi = \phi^R + ic$); this is referred to as the “tangent plane” from now on. Models trained on these configurations for a triangle lattice can sometimes reduce the sign problem but are prone to producing learnifolds with a wrong homology class. This problem becomes worse the larger N_t is and we did not obtain any successful models for the tetrahedron. The reason is the following.

⁴It may still be possible for the network to produce a much smaller vector—for example, the size of a unit cell, or just a single time slice.

Empirically, we found that the dependence of the typical set of the probability distribution W on the vector norm $|\text{Re } \phi|$ is almost the same as that of a normal distribution. In high dimensions, the typical set of a Gaussian, that is the region of phase space that contains most of the probability, is a thin hyperspherical shell around the origin. Importance sampling, by definition, generates fields that lie in the typical set. We can thus use HMC to visualize the set, an example is shown in orange in Fig. 1. The critical points of all thimbles that we used are shown as crosses in the figure. It can be seen that those points are close to the origin and far away from the typical set. Therefore it should not be surprising that neural networks trained on fields near the critical points do not generalize well to the typical set and yield poor performance of HMC.

The approach used in this work is motivated by the dominance of the Gaussian part of the action. We draw $\text{Re } \phi$ randomly from a Gaussian and set $\text{Im } \phi$ to be on the tangent plane of the main critical point. These ϕ are then flowed upwards. This method has two numerical parameters that need to be tuned. One is the width of the Gaussian σ , the other is the flow time. Plots like Fig. 1 provide good estimates for the quality of training data. The better the overlap between training data and HMC in these projections the better the neural network performs. Fortunately, the distributions explored by HMC with and without neural networks are largely identical in these projections for well trained models. It is thus sufficient to sample configurations on the real plane in order to estimate the HMC distribution—even though the figure shows an ensemble generated *with* a network. The overlap shown in the figure, while not perfect, is enough and the figure shows data from a successful run with reduced sign problem shown in more detail below.

Samples drawn from a normal distribution of fixed width $\sigma = \sqrt{\bar{U}}$ have poor overlap in figures like Fig. 1. To remedy this, we draw the width of every sample randomly from a uniform distribution $\mathcal{U}$

$$\sigma \sim \mathcal{U}(\sqrt{\bar{U}}/x, \sqrt{\bar{U}}) \quad (30)$$

and then draw ϕ from a normal distribution $\mathcal{N}$ with that width: $\text{Re } \phi \sim \mathcal{N}(0, \sigma)$. The factor x is chosen such that the overlap of HMC ensemble and training data is maximized.

Similarly, the other parameter, the flow time τ_f , needs to be chosen such that plots like Fig. 1 show good overlap. Additionally, one needs to strike a balance between reducing the sign problem and avoiding ergodicity problems as described in Sec. III A. It is easy to show that for space-time-constant ϕ the gradient of the action is invariant under simultaneous $N_t \rightarrow \alpha N_t$ and $\phi \rightarrow \phi/\alpha$. The starting position for flow, which is on the main tangent plane, has, therefore, the same N_t dependence. We use this scaling for the flow time as well and choose it to be

$$\tau_f^{\max} = 0.1 \times 16/N_t \quad (31)$$

in all cases. As with σ , this scaling was found purely based on numerical experiments.

We found significant improvements in both the time required to make training data and the performance of the final neural networks by not requiring a fixed flow time. When using holomorphic flow directly within MC, it is necessary

to fix the flow time in order for all configurations to be on the same manifold. In our case, the neural network ensures this regardless of how the training data was created. Many configurations generated from a Gaussian quickly flow into neverland and flowing becomes numerically unstable. It is thus impossible to reach the targeted flow time. Instead of discarding these configurations, we monitor the integrator for stability, abort early in such a case, and add the last configuration that could be reached in a stable way to the set of training data. However, it does not make sense to use configurations that could only be flowed for very short times as those provide little information to the neural network. We thus require a minimum flow time of

$$\tau_f^{\min} = 0.04 \times 16/N_t \quad (32)$$

for all training configurations.

C. HMC

Rather than implementing HMC on the curved manifold of Lefschetz thimbles defined through holomorphic flow [77, 78] or learnifolds, we pull back to the real plane before performing molecular dynamics, as will now be explained.

Given a transformation into complex space as implemented by holomorphic flow or neural networks, we need to incorporate that transformation into Hybrid Monte Carlo. Let

$$f : \mathbb{R}^N \rightarrow \mathbb{C}^N, \phi \mapsto \tilde{\phi} \quad \text{and} \quad J_{ij}[\phi] \equiv \frac{\partial f_i(\phi)}{\partial \phi_j}, \quad (33)$$

with $\mathcal{M} \equiv f(\mathbb{R}^N)$ the image of f . We want to calculate observables by integrating over the manifold $\mathcal{M}$ because with suitably chosen f the sign problem is alleviated on that manifold. Since $\mathcal{M}$ is in the same homology class as $\mathbb{R}^N$ and the integrand is regular in ϕ , by Cauchy's theorem the expectation value of an observable $\hat{O}$ is

$$\langle \hat{O} \rangle = \frac{1}{\mathcal{Z}_{\mathbb{R}^N}} \int_{\mathbb{R}^N} \mathcal{D}\phi \hat{O}[\phi] e^{-S[\phi]} = \frac{1}{\mathcal{Z}_{\mathcal{M}}} \int_{\mathcal{M}} \mathcal{D}\tilde{\phi} \hat{O}[\tilde{\phi}] e^{-S[\tilde{\phi}]}. \quad (34)$$

We do not know $\mathcal{M}$, however, and can thus not evaluate the integral on the right hand side directly. Instead, we parametrize it using f . That is, we perform a transformation of integration variables:

$$\begin{aligned} & \frac{1}{\mathcal{Z}_{\mathcal{M}}} \int_{\mathcal{M}} \mathcal{D}\tilde{\phi} \hat{O}[\tilde{\phi}] e^{-S[\tilde{\phi}]} \\ &= \frac{1}{\mathcal{Z}_{\mathbb{R}^N}} \int_{\mathbb{R}^N} \mathcal{D}\phi \hat{O}[\tilde{\phi}(\phi)] e^{-S[\tilde{\phi}(\phi)]} \det J[\phi]. \end{aligned} \quad (35)$$

Now define the effective action as

$$S_{\text{eff}}[\tilde{\phi}(\phi)] \equiv S[\tilde{\phi}(\phi)] - \ln \det J[\phi]. \quad (36)$$

We can estimate the integral stochastically by generating an ensemble $\{\phi \sim \exp(-\text{Re} S_{\text{eff}}[\tilde{\phi}(\phi)])\}$, where we use the real part of S_{eff} as per the reweighting procedure described in Sec. II C. This effectively produces an ensemble $\{\tilde{\phi} = f(\phi) \sim \exp(-\text{Re} S[\tilde{\phi}])\}$ on which the observables $\hat{O}$ can be measured.

We use HMC to generate the ensembles. To this end, we augment the integral by multiplying with a one in the form of

an integral over the artificial conjugate momentum p such that

$$\langle \hat{O} \rangle = \frac{1}{Z} \int \mathcal{D}\phi \mathcal{D}p \hat{O}[\tilde{\phi}(\phi)] e^{-i\text{Im}S_{\text{eff}}[\tilde{\phi}(\phi)]} e^{-H[p, \tilde{\phi}(\phi)]},$$

$$H[p, \tilde{\phi}(\phi)] \equiv \frac{p^2}{2} + \text{Re}S_{\text{eff}}[\tilde{\phi}(\phi)]. \quad (37)$$

All integrals are to be understood as integrating over $\mathbb{R}^N$ from now on. We can encapsulate all dependencies on f in a modified HMC algorithm such that it produces configurations on $\mathcal{M}$ suitable for measurements. The following summarizes the algorithm:

complexified HMC in: $\tilde{\phi}$ out: $\tilde{\phi}'$

$\phi \leftarrow f^{-1}(\tilde{\phi})$	# transform to $\mathbb{R}^N$
$p \leftarrow \mathcal{N}_{0,1}$	# draw random momentum
$\psi', p' \leftarrow \text{molecular_dynamics}(\phi, p)$	# generate candidate on $\mathbb{R}^N$
$\tilde{\psi}' \leftarrow f(\psi')$	# transform to $\mathcal{M}$
$\tilde{\phi}' \leftarrow \text{accept_reject}(H[p', \tilde{\psi}'], H[p, \tilde{\phi}])$	# pick new field

A new configuration is obtained from an old one by first transforming the old complex field $\tilde{\phi}$ to the real plane⁵ where it is then updated using molecular dynamics (MD), or any other suitable updating scheme, such as the large jumps explained in Ref. [62]. The new candidate field ψ is then transformed to $\mathcal{M}$ where it is accepted or rejected using a Metropolis-Hastings step.

It remains to prove that this algorithm produces a Markov chain. For this, we adopt the picture that we are producing real fields $\{\phi \sim \exp(-\text{Re}S_{\text{eff}}[\tilde{\phi}(\phi)])\}$. Such a proof is equivalent to showing that $\{\tilde{\phi}\}$ is a Markov chain sampled from $\exp(-\text{Re}S[\tilde{\phi}])$. We do so by first proving detailed balance

$$\mathbb{P}(\phi)\mathbb{P}(\phi'|\phi) = \mathbb{P}(\phi')\mathbb{P}(\phi|\phi') \quad (38)$$

with (ignoring normalization factors)

$$\mathbb{P}(\phi) = e^{-\text{Re}S_{\text{eff}}[\tilde{\phi}(\phi)]}, \quad (39)$$

$$\begin{aligned} \mathbb{P}(\phi'|\phi) &= \int \mathcal{D}p' \mathcal{D}p e^{-p'^2/2} \mathbb{P}_{\text{MD}}(p', \phi'|p, \phi) \\ &\times \mathbb{P}_{\text{atr}}(p', \tilde{\phi}(\phi')|p, \tilde{\phi}(\phi)). \end{aligned} \quad (40)$$

$\mathbb{P}_{\text{MD}}$ and $e^{-p^2/2}$ are the same molecular dynamics and Gaussian probabilities as in standard HMC. The prior $\mathbb{P}(\phi)$ and accept/reject

$$\begin{aligned} &\mathbb{P}_{\text{atr}}(p', \tilde{\phi}(\phi')|p, \tilde{\phi}(\phi)) \\ &= \min \left(1, \frac{\exp(-\frac{p'^2}{2} - \text{Re}S_{\text{eff}}[\tilde{\phi}(\phi')])}{\exp(-\frac{p^2}{2} - \text{Re}S_{\text{eff}}[\tilde{\phi}(\phi)])} \right) \end{aligned} \quad (41)$$

probabilities use the effective action which encapsulates $\tilde{\phi}$. Thus the proof of detailed balance (38) proceeds as usual for HMC. The only ingredient missing to fully prove correctness of our algorithm is a proof of ergodicity. Such a proof is generally not available even for standard HMC. We thus rely on

⁵The inverse transformation f^{-1} can be unstable and/or expensive to evaluate. However, there is no need to perform this calculation if we just keep track of ϕ as well as $\tilde{\phi}$.

a posteriori analyses to verify ergodicity. Certainly with a long flow time we expect many configurations to flow to neverland, creating large zero-probability regions that separate important islands of configurations. However, if we only flow a little, few if any configurations flow to neverland and the manifold of integration is not partitioned. Our ability to reproduce exact results in Sec. IV suggests that our method successfully explores fields whose images are near different thimbles. A well trained network inherits these properties from the flowed manifold of its training data.

IV. RESULTS

To demonstrate the efficacy of the neural network methods we will explore small lattices, leaving larger lattices for future work. In particular, we here consider the triangle and tetrahedron, both maximally connected and therefore, where we expect the worst sign problem. Finally, we demonstrate that our method yields improved scaling of the statistical power in time, alleviating the sign problem. Focusing on the small examples we study, see the greatest improvement for the most difficult: the tetrahedron. We leave the question of how much our method improves scaling in space for future study.

A. The triangle

Three spatial sites is the smallest nontrivial nonbipartite graph we might study—the two-site problem is bipartite and the one-site problem has no hopping at all. We studied those problems extensively using the lattice methods applied here in Ref. [62].

The hopping matrix is given by

$$h = \kappa \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad (42)$$

and the Hamiltonian has D_3 dihedral symmetry. When diagonalized, the hopping matrix reveals one spatially uniform (trivial) irrep with eigenvalue 2κ and a dimension-two irrep with eigenvalue $-\kappa$.

The unitary site permutation operator P

$$\begin{aligned} P a_x P^\dagger &= a_{x+1}, & P b_x P^\dagger &= b_{x+1}, \\ P a_x^\dagger P^\dagger &= a_{x+1}^\dagger, & P b_x^\dagger P^\dagger &= b_{x+1}^\dagger \end{aligned} \quad (43)$$

rotates the sites into one another (the indices on the ladder operators are understood mod 3). P trivially commutes with the potential and nontrivially with the hopping (1).

The three irreducible single-particle destruction operators are

$$\hat{O}_k = \frac{1}{\sqrt{3}} \sum_{j=0}^2 e^{2\pi i j k/3} a_j, \quad (44)$$

which are labeled by $k = \pm 1, 0$, which corresponds to their transformation properties,

$$P \hat{O}_k P^\dagger = e^{2\pi i k/3} \hat{O}_k, \quad (45)$$

a spherical-tensor-like relation, and the hopping Hamiltonian can be decomposed

$$H_0 = \kappa[(\hat{O}_{-1}^\dagger \hat{O}_{-1} + \hat{O}_{+1}^\dagger \hat{O}_{+1} - 2\hat{O}_0^\dagger \hat{O}_0) - (a \rightarrow b)], \quad (46)$$

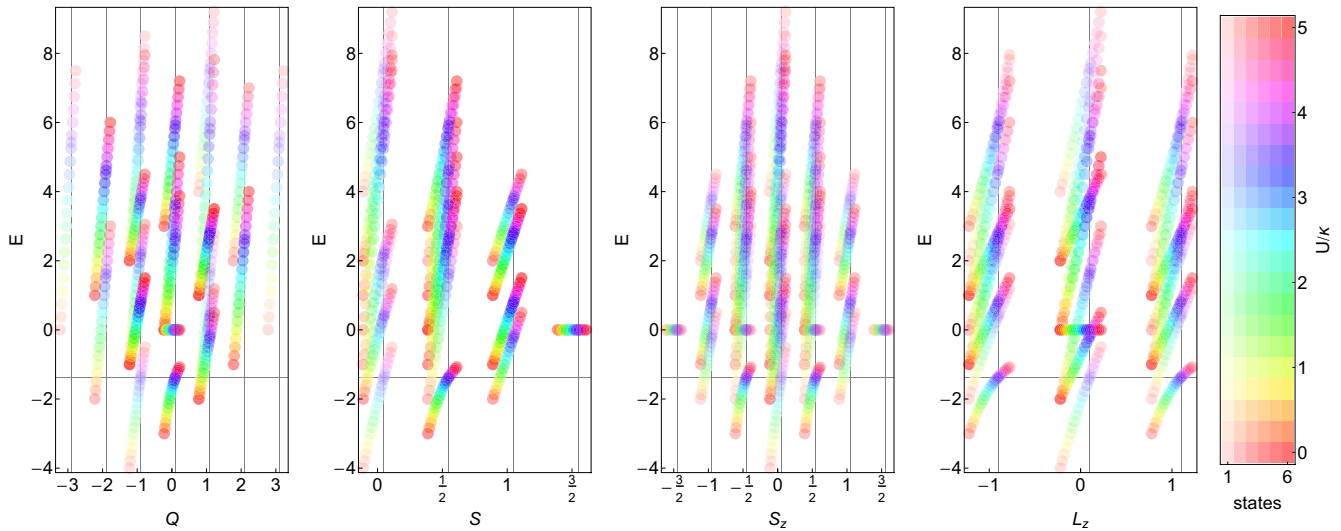


FIG. 2. The entire spectrum of the Hubbard model on a triangle as a function of U/κ , changing from 0 to 5 in 20 equal steps, with colors corresponding to different U/κ , corresponding to the vertical direction in the color bar at right. One translucent point is plotted per state; more opaque circles correspond to more highly degenerate states, corresponding to the horizontal direction in the color bar. Energy eigenfunctions carry Q , S , S_z , and L_z quantum numbers; each panel projects the five-dimensional space onto one quantum number and the energy. Each U/κ is slightly offset, so that each point in the spectrum for $U/\kappa = 0$ is to the left of that point's quantum numbers, while $U/\kappa = 5$ is to the right. The ground-state manifold qualitatively changes at $U^\Delta/\kappa = 3.61775 \dots$, shown in a deep blue. We draw a thin horizontal line at the corresponding degenerate ground-state energy $E_0^\Delta = -1.36845 \dots$ to make the quantum numbers of the states apparent, and thin vertical lines for each value of each quantum number at $U = U^\Delta$.

corresponding to the irreps described above. We label the spatial permutation quantum number by $L_z = k$, since it corresponds to an angular momentum around the center of the triangle.

The spectrum for this system, which can be obtained from direct diagonalization of the Hamiltonian, consists of $4^3 = 64$ states in the entire Fock space. When $U = 0$, a single state with $Q = -1$, $S = 0$, $S_z = 0$, $L_z = 0$ has the lowest energy, and a partner with $Q = +1$, $S = 0$, $S_z = 0$, $L_z = 0$ has the highest energy. Explicitly, the spectrum is not symmetric in Q , though of course the spectrum is symmetric in S_z and L_z . When U becomes large, the lowest energy is shared by a degenerate quadruplet of states with $Q = 0$, $S = 1/2$, $S_z = \pm 1/2$, and $L_z = \pm 1$. At approximately $U^\Delta/\kappa = 3.61775$ the low- U/κ ground state and high- U/κ four-plet are degenerate. When U/κ is very large the spectrum nearly exhibits symmetry in Q .

Spectra, like those shown in Fig. 2 and operator overlap factors (18), can be used to directly calculate the single-particle correlator via the spectral decomposition (17). Using the single-particle operators (44), we can also use Monte Carlo to compute the single-particle correlators numerically.

1. Ergodicity and the sign problem

In the language of Refs. [62,79], we use the exponential $\alpha = 1$ discretization. As detailed in Ref. [62], there is a formal ergodicity problem on bipartite lattices. When the lattice is not bipartite the codimension-1 manifolds of exceptional configurations are reduced in dimension and there is no formal ergodicity problem. First, we give a small toy problem confirming this claim and then reproduce the exact results obtained through direct diagonalization. Then, we examine the statistical power for HMC alone as a function of U/κ and

$\kappa\beta$ and find that, for a given temperature, the sign problem is worst when $U/\kappa = U^\Delta/\kappa \approx 3.61775$, the value where the vacuum changes character, as shown in Fig. 2.

To visually appreciate that the codimension-1 manifolds that prevent HMC alone from being formally ergodic in the bipartite case are reduced in dimension, consider a problem with $N_t = 1$ and let ϕ_x live on spatial site x . Then, the product of the fermion determinants is

$$\det M_p M_h = \frac{4}{9} \left[3 \cos \left(\frac{1}{2} \Phi \right) + \sum_{j=1}^3 \left(2 \cosh \left(\kappa + \frac{i}{2} (\Phi - 2\phi_j) \right) + \cosh \left(2\kappa - \frac{i}{2} (\Phi - 2\phi_j) \right) \right) \right]^2 \quad (47)$$

where $\Phi = \sum_x \phi_x$. Figure 3 shows the absolute value (left panel) and complex argument (right panel) of this determinant at $\Phi = 0$ and two orthogonal combinations of the field variables. The only zeros are where the lines of different phase meet—in the two-dimensional projection of the phase in Fig. 3, points around which the phase winds. As Φ changes, those points move but they never become extended. So, even with an exponential discretization, the codimension-1 zeros are reduced when the lattice is not bipartite; the fermion determinant allows free exploration the complex plane, rather than constraining it by the reality condition that arises in the bipartite case, as explained in Ref. [62].

The Gaussian part of the weight encourages the fields to stay in the central mode. When the Gaussian becomes wider, more than one mode in Fig. 3 might become important. There is no formal ergodicity problem, even in the exponential case,

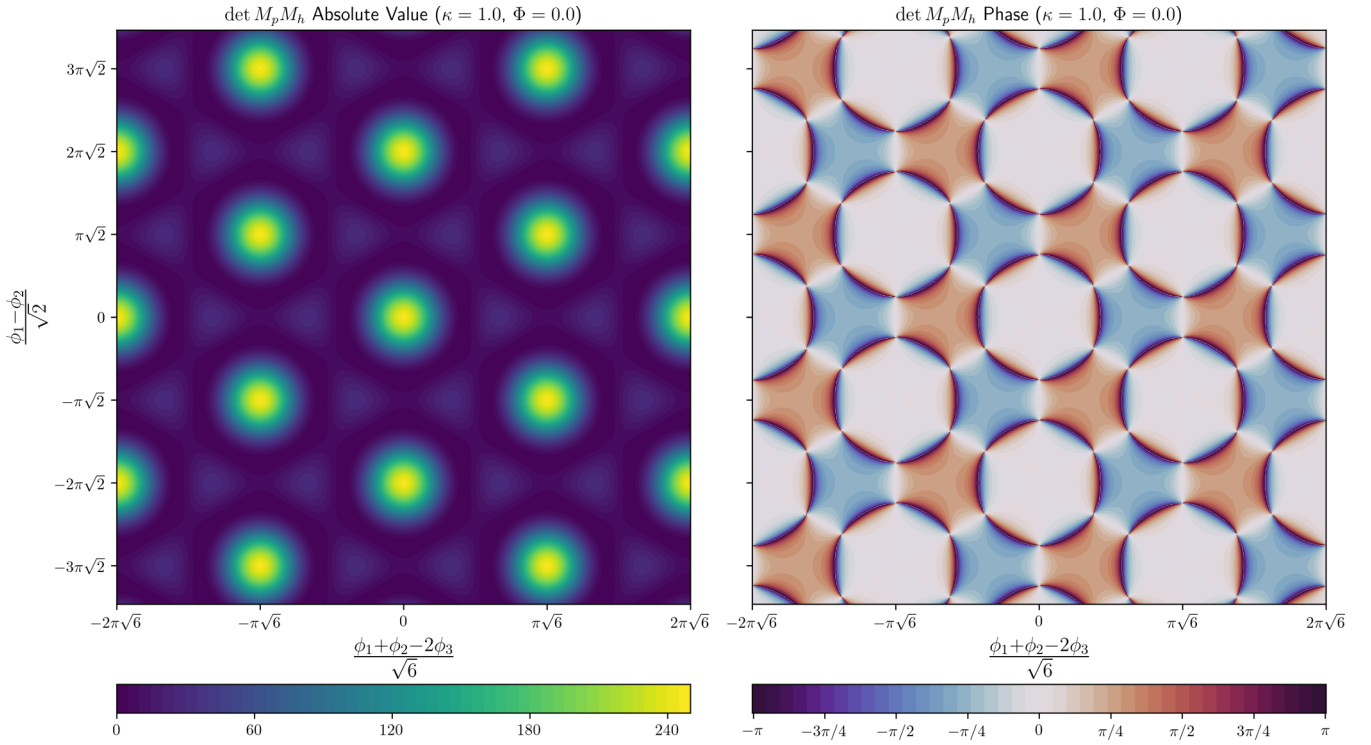


FIG. 3. The absolute value and phase of (47) for $\kappa = 1$ and $\Phi = 0$, shown as a function of two directions in field space orthogonal to the Φ direction. In the left panel, yellow indicates large absolute values and purple small ones. In the right panel the color scheme is periodic; the exact zeros occur at the six points on each dark circle where the phase wraps around the point.

and HMC can take us from mode to mode. However, at the mode boundaries, the phase changes rapidly, causing a sign problem. In this small example, if the Gaussian is wide enough, HMC can sample trajectories that near this rapid change we would expect to encounter a sign problem, but if the width were very narrow we would not. In examples with more lattice sites and time slices, the huge growth of phase space of configurations further from $\phi = 0$ can counterbalance a narrow Gaussian (the width is controlled by $U\beta/N_t$), so it is *a priori* unclear whether increasing N_t will help or hurt; since the action is extensive, one expects a sign problem exponentially bad with β . However, increasing N_t also increases the number of variables and the odds that some are near the mode boundaries.

In Fig. 4, we show the phase histogram (left panels) and the statistical power (right panels) of ensembles generated with HMC with real valued fields for two values of U/κ as functions of $\beta\kappa$. We generated 100 000 trajectories with one molecular dynamics time unit and the number of steps in the leapfrog integrator to give better than 75% acceptance. The histograms are actually results for multiple N_t s superimposed, to show the very mild sensitivity to the discretization scale. We also generated additional ensembles at U/κ from one to nine in integer steps and U^Δ/κ , and found that the sign problem modestly improved, for fixed $\kappa\beta$, for couplings further from U^Δ/κ . By analogy, we expect worse sign problems for critical points, where the system must tunnel between qualitatively different ground states.

2. Results

Since we can exactly calculate the spectrum and overlap factors for this small problem, we can generate the exact, continuum-limit correlation functions according to the spectral decomposition (17). This provides us a means to directly check the accuracy of our NN method.

We ran HMC in three different ways—on the real plane, on the tangent plane of the main thimble, and on the learnifold. The left panel of Fig. 5 shows running averages of the statistical power as a function of Monte Carlo time for $U/\kappa = 3$ and $\kappa\beta = 8$. After many configurations, Σ on real and tangent planes converges to the same small but nonvanishing value, while the neural network produces a markedly greater statistical power. Remember, an improved statistical power indicates an exponential reduction of the sign problem. So, while the sign problem is not *solved*, per se, we provide evidence here that it is significantly alleviated on the lattices we are considering. This chosen set of parameters shows the worst statistical power on real and tangent plane out of the sets we tested. See the right panel of Fig. 5 for a summary for different parameters. The large variations in Σ for neural network based calculations stem from different qualities of the trained models. It should be possible to tune the networks better and thus increase statistical power. These networks perform well enough, however. Generally, we found that the sign problem is not a complete impediment on the triangle and a plain calculation on the real plane with reweighting can suffice.

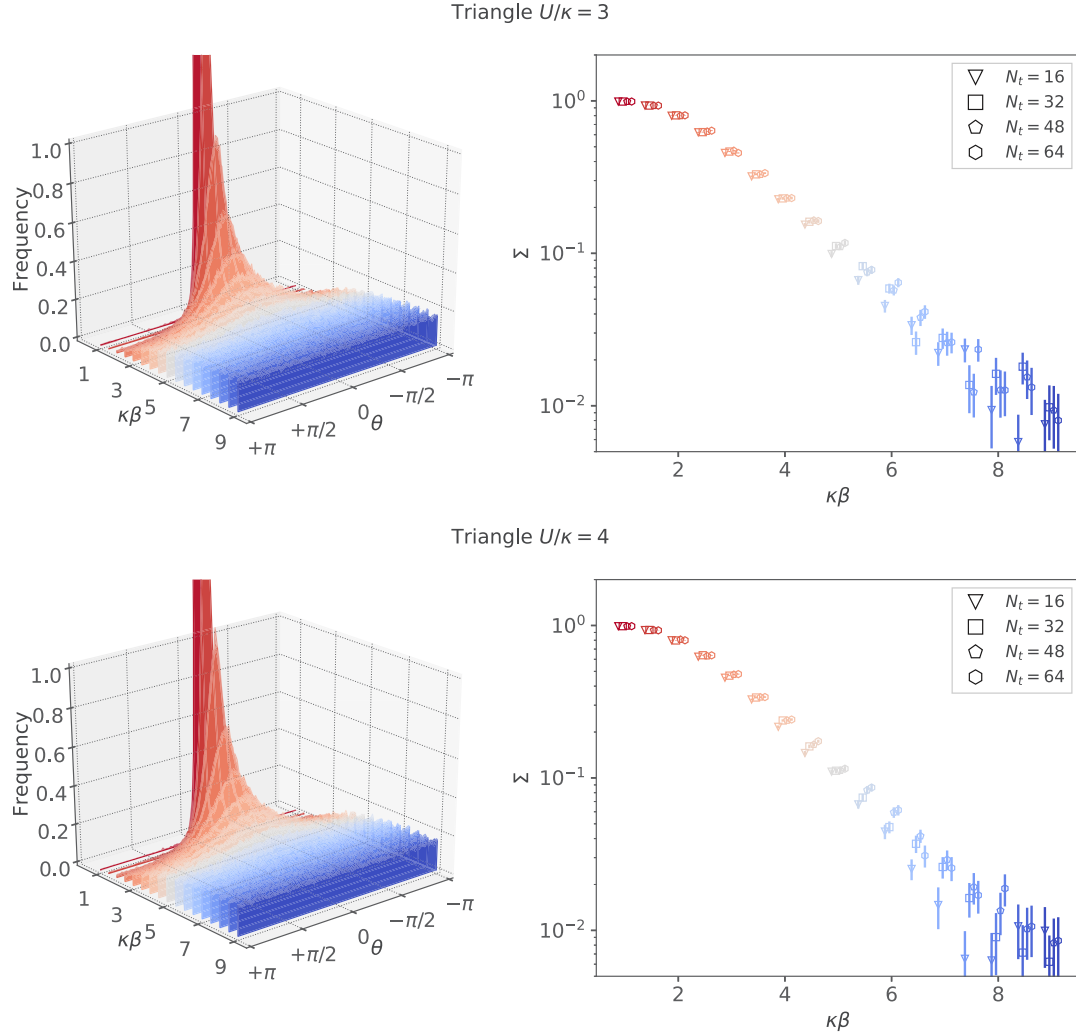


FIG. 4. The histograms (left) of the phase for two different values of U/κ as a function of β and the corresponding statistical power (right) with bootstrap errors. In the right panel, different N_t are shown from left to right: 16, 32, 48, and 64. The histograms (left) and statistical powers (right) are colored from red, a hot temperature, to blue, a cold temperature.

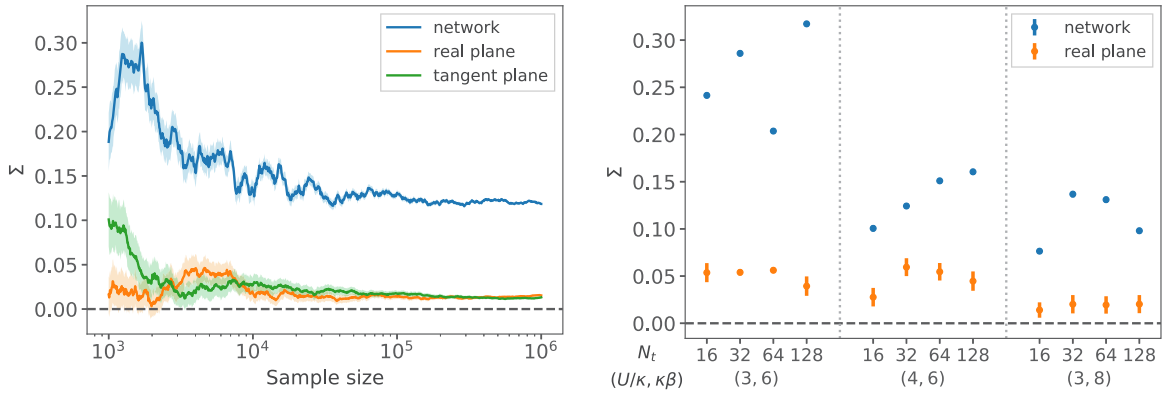


FIG. 5. Left panel shows statistical power as a function of sample size on a triangle lattice with $N_t = 64$, $U/\kappa = 3$, $\kappa\beta = 8$. Right panel shows statistical power for different parameters on a triangle lattice. Each point was estimated using 10^5 configurations. The dashed line is at $\Sigma = 0$.

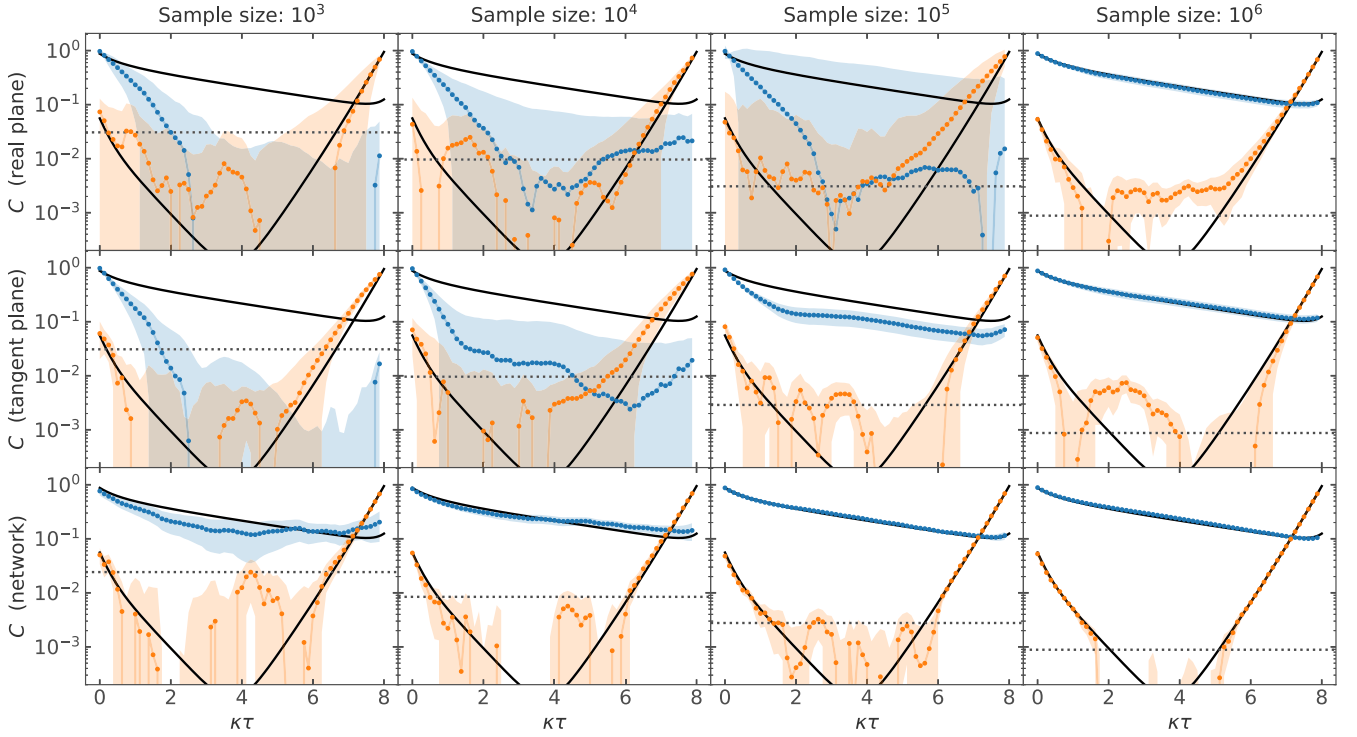


FIG. 6. Single-particle correlators $\langle aa^\dagger(\tau) \rangle$ on a triangle lattice with $N_t = 64$, $U/\kappa = 3$, and $\kappa\beta = 8$. Each column shows correlators obtained from ensembles of the given number of configurations, while each row shows a different implementation of HMC. Colored points and areas show averages and error bands from HMC and black lines show exact results. The two unique correlators on a triangle are shown in blue and orange respectively—the former, low-energy correlator is the average of the doublet of two-point correlators of $O_{\pm 1}$ while the latter, high-energy correlator is a two-point correlator of the singlet O_0 . Sample sizes list the total number of Monte Carlo configurations but correlators were measured only on every 10^{th} configuration. The dotted lines are placed at $\max(C)/\sqrt{\text{Sample size}}$ and indicate the scale at which even a sign-problem-free method would show sizable statistical fluctuations.

In Fig. 6, we show how the different methods' measurements of diagonalized⁶ single-particle correlation functions converge as a function of the number of configurations. With many configurations all three methods reproduce the exact correlators (shown in black). However, the network reproduces the exact results (up to the expected $1/\sqrt{N}$ relative scale from statistical noise) with fewer configurations.

Focusing on learnifold-enhanced HMC calculations, Fig. 7 shows single-particle correlators for several different parameters and discretization scales measured on ensembles of 10^5 configurations. The figure also shows deviations from the exact results as

$$E = C_{\text{MC}}/C_{\text{exact}} - 1. \quad (48)$$

In addition, we compute and diagonalize a variety of correlation functions between bilinear operators, as mentioned in Sec. IID and detailed in Appendix A 1. In Fig. 8, we show two—the diagonalized charge-charge and spin-spin correlators. Additional correlators can be found in Fig. 16.

⁶All correlators shown in this work are projected to irreducible representations of the lattice, as in (44). We found this method suitable to produce diagonal all-to-all correlation matrices on triangle and tetrahedron lattices. There are two degenerate correlators on the triangle and three on the tetrahedron. Thus all correlator plots show only two distinct lines.

We can use the constant correlators to extract $\langle Q^2 \rangle$ and $\langle S^2 \rangle$, as explained in Appendix A 2. In Fig. 9, we show results for different couplings, temperatures, and discretizations. The learnifold approach yields reduced errors and results consistent with the exact results of Q^2 for $U/\kappa = 3$. The $(U/\kappa, \kappa\beta) = (4, 6)$ case seems to have a systematic deviation. It is unsurprising that these parameters yield the worst result as they are closest to U^Δ/κ . The errors of S^2 are also reduced by the machine learning approach but a significant systematic deviation from the exact result remains in all cases.

B. The tetrahedron

Many arrangements of four sites are not bipartite. Of these the “most nonbipartite”, and therefore the one where we expect the worst sign problem, is the tetrahedron, where the connectivity matrix is proportional to the adjacency matrix of the complete graph on four sites,

$$h = \kappa \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad (49)$$

so that each subset of 3 sites forms a frustrated triangle; we label the sites 0–3. The Hamiltonian commutes with the permutation operator P that acts on any triangular face (43) and leaves the other site alone. We conventionally pick

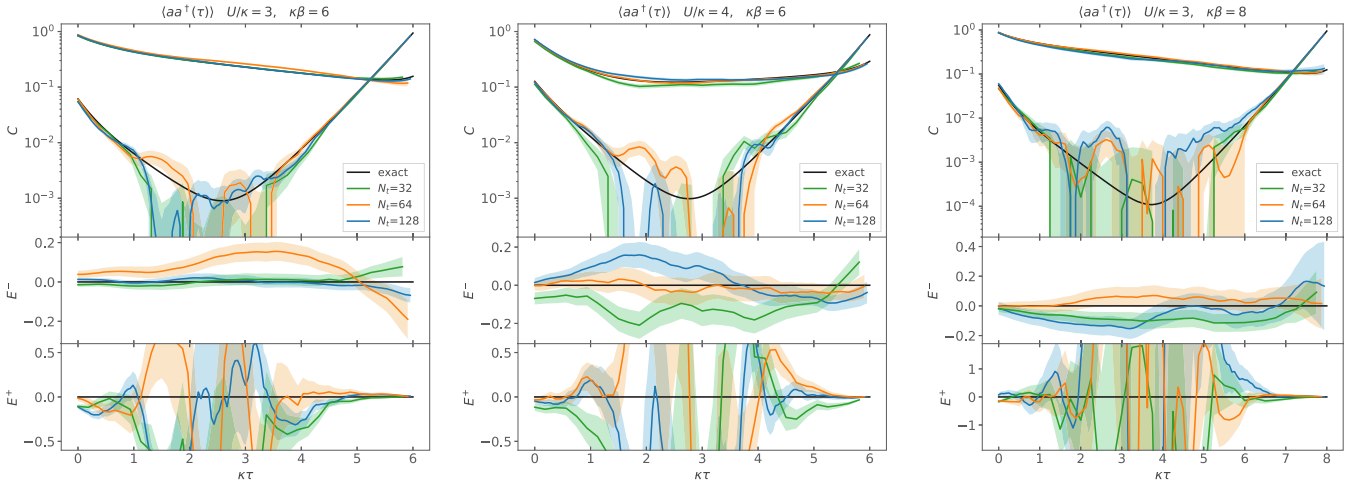


FIG. 7. Single-particle correlators $\langle aa^\dagger(\tau) \rangle$ on a triangle lattice for a sample size of 10^5 . The low energy correlators are averages of two degenerate results. The top panels show the correlators for different N_t from HMC with a neural network and the results from exact diagonalization of the Hamiltonian. The lower panels show the relative error (48). Errors for the low and high energy correlators are labeled E^- and E^+ , respectively.

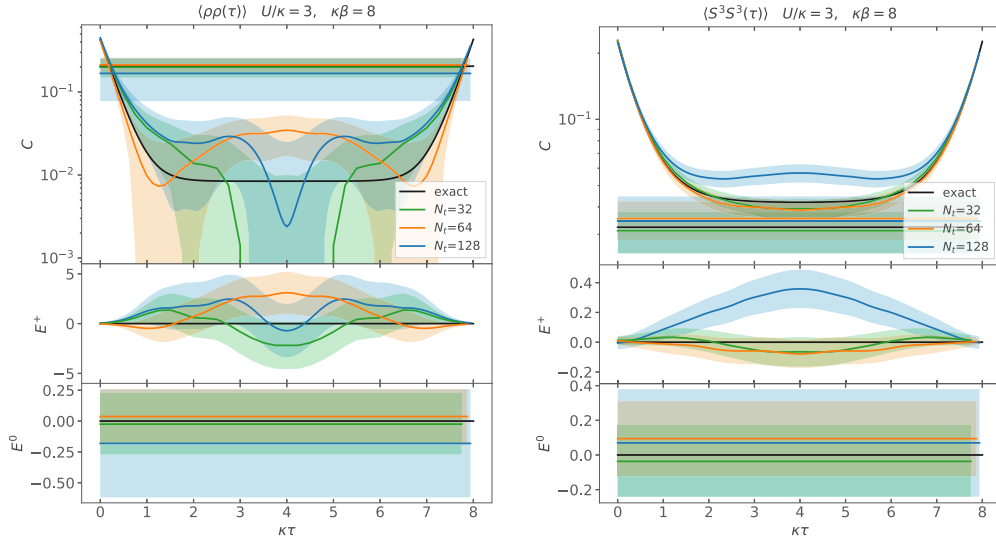


FIG. 8. Correlation functions between two charge operators ρ or two S^3 spin operators separated by euclidean time $\kappa\tau$ on a triangle for different discretization scales given by N_t and the exact result (in black). The bottom panels show the relative error (48) for the constant correlator (0) and the average of two heavy correlators (+) which have $k = 0$ and ± 1 , respectively.

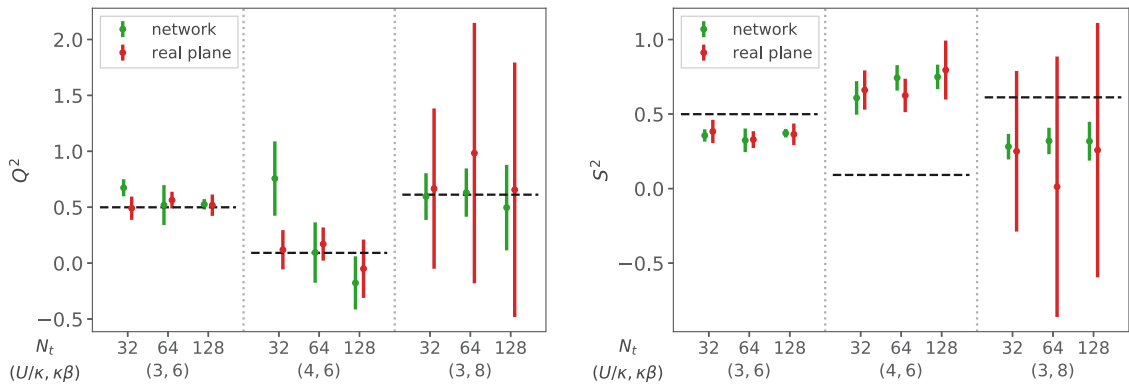


FIG. 9. The order parameters Q^2 (left) and S^2 (right) calculated with the triangle system with various couplings U and inverse temperatures β . Shown is a comparison between results calculated with the neural network and on the standard real plane. The dashed lines show results from exact diagonalization.

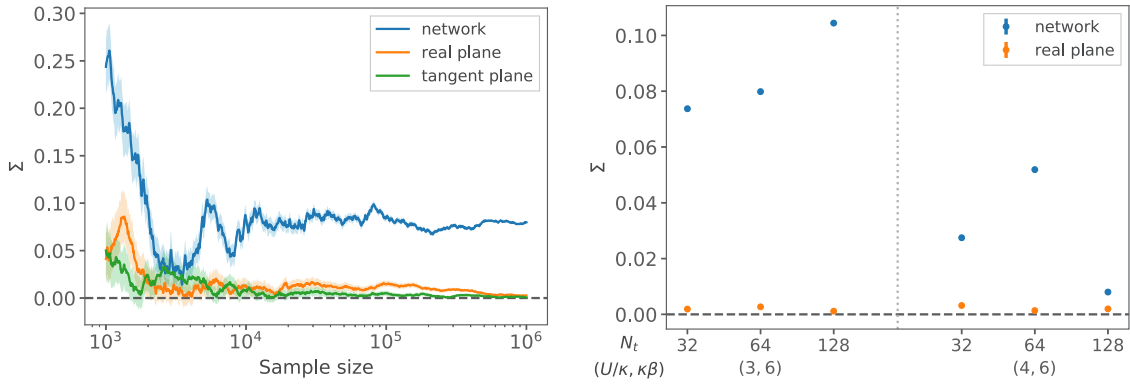


FIG. 10. Left panel shows the statistical power as a function of sample size on a tetrahedron lattice with $N_t = 64$, $U/\kappa = 3$, and $\kappa\beta = 6$. Right panel shows statistical power for different parameters on a triangle lattice. Each point was estimated using 10^6 configurations. The dashed line is at $\Sigma = 0$.

the symmetry axis through the fourth site to be the axis of symmetry around which we have a rotational quantum number.

The four irreducible single-particle destruction operators are

$$\begin{aligned}\hat{O}_0^0 &= \frac{1}{2} \sum_{j=0}^3 a_j, & \hat{O}_1^0 &= \frac{1}{2\sqrt{3}} \left(3a_3 - \sum_{j=0}^2 a_j \right), \\ \hat{O}_1^{\pm 1} &= \frac{1}{\sqrt{3}} \sum_{j=0}^2 e^{\pm \frac{2\pi i}{3} j} a_j,\end{aligned}\quad (50)$$

where the lower index indicates an L^2 -like quantum number ℓ and the upper index an L_z -like quantum number m . The free Hamiltonian may be written in terms of these operators,

$$H_0 = \kappa \left[\left(-3\hat{O}_0^{0\dagger} \hat{O}_0^0 + \sum_{m=-1}^1 \hat{O}_1^{m\dagger} \hat{O}_1^m \right) - (a \rightarrow b) \right] \quad (51)$$

and the translationally invariant interaction term transforms as an $\ell = 0$, $m = 0$ “spherical tensor.” Therefore these are good quantum numbers and correlation functions put into this basis are diagonal.

1. Results

We expect the sign problem to be worse than on a triangle because the tetrahedron is substantially more frustrated since it has four triangular faces. This is an opportunity to test our method in a system where calculations on the real plane are, as far as we can tell, just not possible. Indeed, HMC on both the real and tangent planes has essentially 0 statistical power as shown in the left panel of Fig. 10, an extremely difficult sign problem. In contrast, the neural network method converges to a finite statistical power. Even though this value is small, it is sufficient as shown below. The right panel of Fig. 10 shows this improvement in statistical power holds for all the ensembles we consider.

Figure 11 shows the drastic improvement obtained when simulating with the neural network. By studying the single-particle correlators, it is apparent that while the network method converges to the exact answer, the real and tangent

plane methods are completely ineffective—their uncertainties remain large and their match to the exact results poor.

In Fig. 12, we show network-method results for different discretizations, and their relative errors E (48). Those correlators were computed on an ensemble of 10^6 configurations, measuring on only every tenth configuration to reduce auto-correlations. The light correlator is an average of the $\ell = 1$ triplet of correlation functions. The error of the heavy correlator grows for intermediate euclidean time, but this is expected given the concrete sample size, see Fig. 11.

In addition, we computed bilinear correlation functions as described in Sec. IID and Sec. A 1. Figure 13 shows the continuum-limit convergence of the charge-charge and S^3 - S^3 correlators projected to the singlet and triplet [as in (50)] towards the exact result. Additional examples can be found in Fig. 17.

In Fig. 14, we show $\langle Q^2 \rangle$ and $\langle S^2 \rangle$, see Appendix A 2 for their derivation. Calculations on the real plane show significant systematic deviations from the exact result. Calculations on learnifolds, however, have improved Q^2 for all U/κ and S^2 for $U/\kappa = 3$ to the point where they agree with the exact result. Curiously, S^2 for $(U/\kappa, \kappa\beta) = (4, 6)$ is worse on the learnifold. Note, however, that while the real plane results are consistent with the exact value of S^2 , they are also consistent with zero.

C. Scaling

It is widely expected that the statistical power decays with space-time volume. This follows from the observation that the power is the ratio of two partition functions—that of the full theory divided by that of the phase-quenched theory—and therefore is exponential in a difference of free energies, which is extensive in the space-time volume [80].

However, because the particular sign problem treated in this work is generated by nonbipartiteness rather than a chemical potential, the extensiveness in space may depend on the space’s structure. For example, in a triangular lattice, every face is frustrated and the ‘amount of nonbipartiteness’ scales with the lattice volume, so a natural guess is that the severity of the sign problem scales with space, like a local chemical potential. In contrast, on a topologically-spherical fullerene, Euler’s theorem guarantees that no matter how large

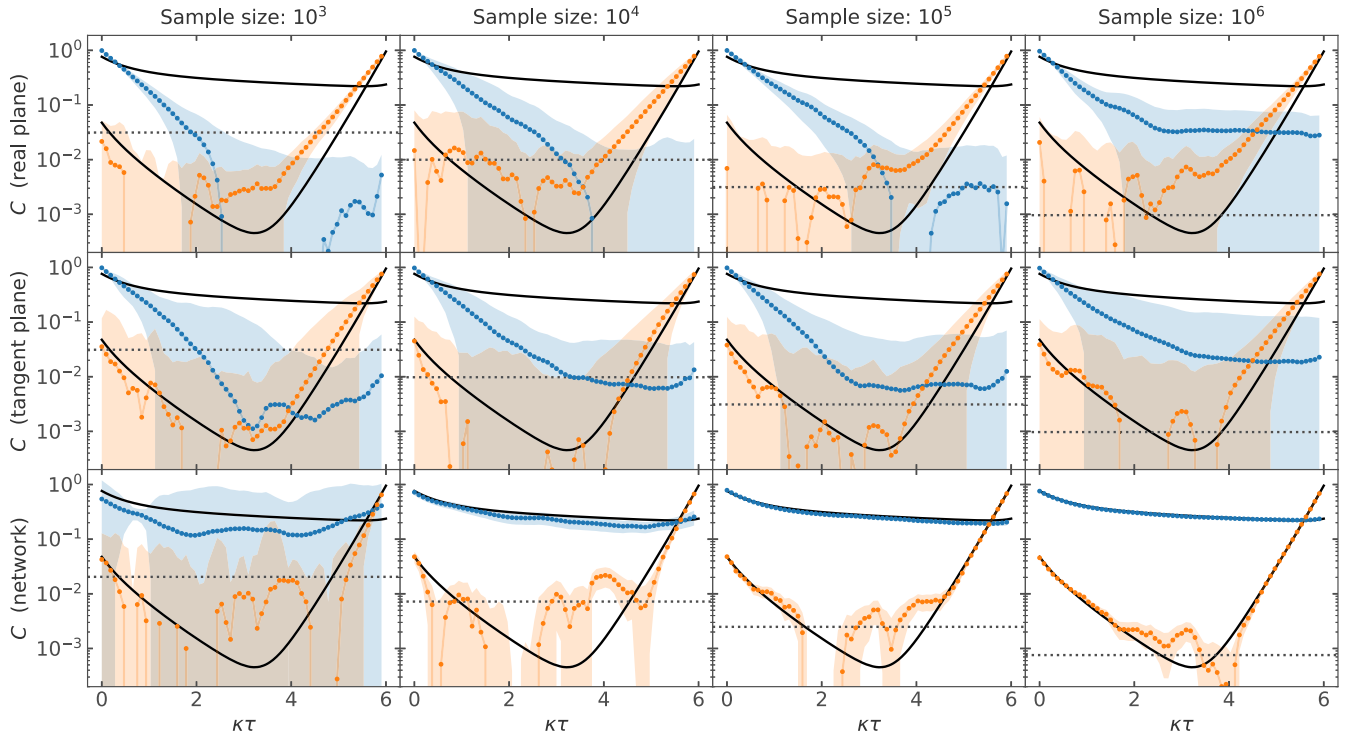


FIG. 11. Single-particle correlators $\langle aa^\dagger(\tau) \rangle$ on a tetrahedron lattice with $N_t = 64$, $U/\kappa = 3$, $\kappa\beta = 6$. Each column shows correlators obtained from ensembles of the given number of configurations, while each row shows a different implementation of HMC. Colored points and areas show averages and error bands from HMC and black lines show exact results. The two unique correlators on a tetrahedron are shown in blue and orange respectively—the former, low-energy correlator is the average of the three degenerate two-point correlators of the triplet of O_1 while the latter, high-energy correlator is a two-point correlator of the O_0 singlet. Correlators are computed only on every tenth configuration. The dotted lines are placed at $\max(C)/\sqrt{\text{Sample size}}$ and indicate the scale at which even a sign-problem-free method would show sizable statistical fluctuations.

the fullerene becomes there will always be 12 pentagons—the frustrated faces that are the only defects in the otherwise-bipartite fullerenes; their number does not scale with the fullerene’s size.

However the spatial part scales with the number of sites, the scaling with Euclidean time continues to hold. For example, in Fig. 4, we showed the statistical power Σ of HMC for the triangle as a function of $\kappa\beta$, which exhibits

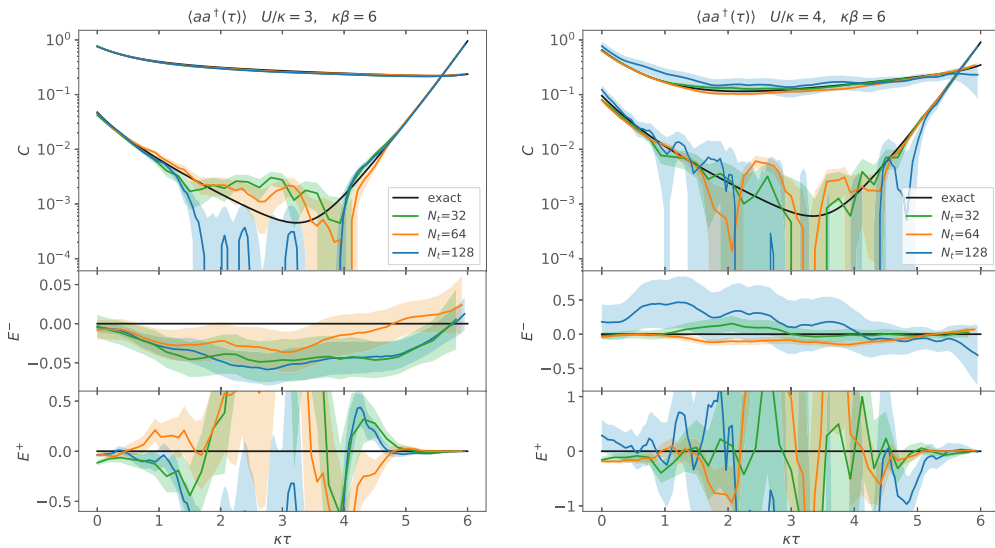


FIG. 12. Single-particle correlators $\langle aa^\dagger(\tau) \rangle$ on a tetrahedron lattice. The low energy correlators are averages of three degenerate results. The top panels show the correlators for different N_t from HMC with neural network and the results from exact diagonalization of the Hamiltonian. The lower panels show the relative error (48). Errors for the low and high energy correlators are labeled E^- and E^+ , respectively.

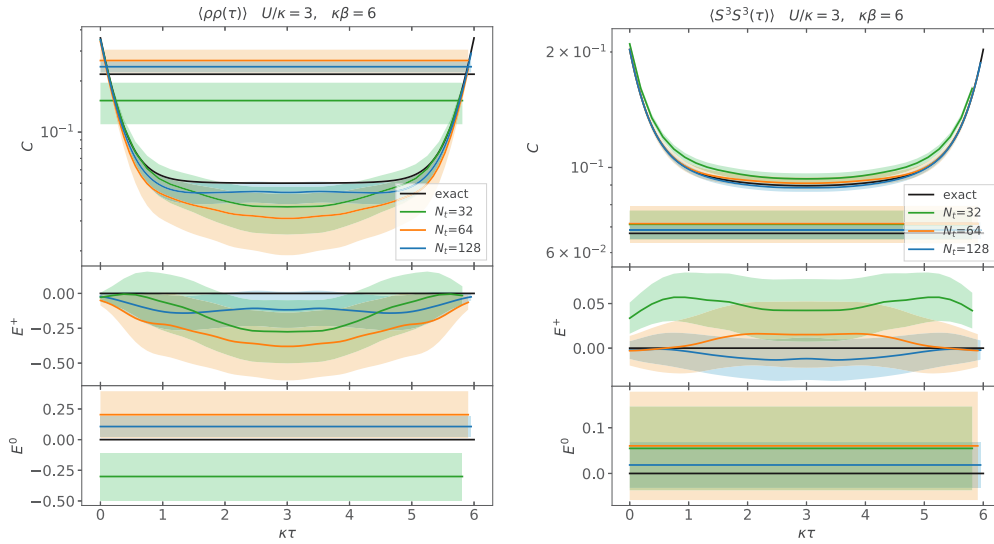


FIG. 13. Correlation functions between two charge operators ρ or two S^3 spin operators on a tetrahedron. The bottom panels show the relative error (48) for the constant correlator E^0 and the average of three heavy correlators E^+ , respectively.

a clear exponential decay once $\kappa\beta \gtrsim 3$. In Fig. 15, we show the results of a scaling study on the triangle, a diamond made of two equilateral triangles with a common edge (or, equivalently, a tetrahedron with one edge deleted), and tetrahedron comparing straight HMC and our neural-network-enhanced HMC, trained as described above, all with a discretization of molecular dynamics time targeting a 75% acceptance rate at least. For the HMC-only ensembles we produced ensembles of 10^5 configurations, except for the tetrahedron with temperatures above 3.75, where the statistical power could not be reliably resolved without 10^6 configurations, and measured the imaginary part of the action on every fourth configuration, computing the statistical power Σ (15) with uncertainties determined by bootstrap. For each machine-learning-enhanced ensemble we trained on 1000 flowed configurations as described above, and produced 60 000 configurations, measuring the imaginary part of the action on every other.

It is apparent that with HMC the tetrahedron has a much worse sign problem than the triangle which, as we have

already seen, has a relatively mild sign problem. This jibes with our intuition that additional frustrated faces should present a worse problem. Surprisingly, the triangle and diamond have almost the same scaling, counter to our intuition that the sign problem scales with the number of frustrated faces.

Also clear is that the statistical power of HMC+NN scales much more favorably. For example, the tetrahedron with neural-network-enhanced HMC scales slightly better than the triangle without. It is also clear that the statistical power can gain from a well-trained network or suffer because of a poorly trained network. For example, the tetrahedron $\kappa\beta = 5$ point is slightly lower than the rest of the linear trend, while the $\kappa\beta = 6.5$ point is above the trend. We emphasize that the uncertainties in Fig. 15 are only statistical error bars—they do not account for systematics that depend on the training of the neural network, which explains why the trend is not as smooth. Nevertheless, the clear takeaway is that the method provides a dramatically easier sign problem, improving the exponential decay of the statistical power markedly.

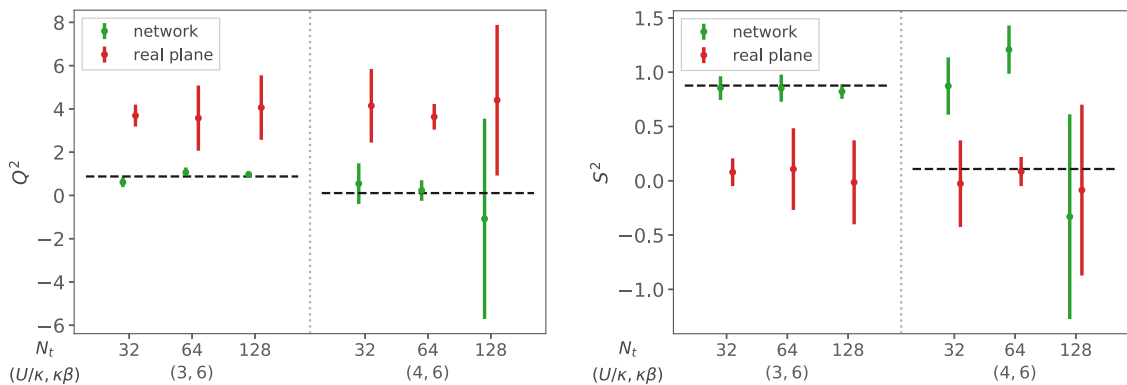


FIG. 14. The order parameters Q^2 (left) and S^2 (right) calculated with the tetrahedron system with different couplings U and inverse temperature β . Shown is a comparison between results calculated with the neural network and on the standard real plane. The dashed lines show results from exact diagonalization.

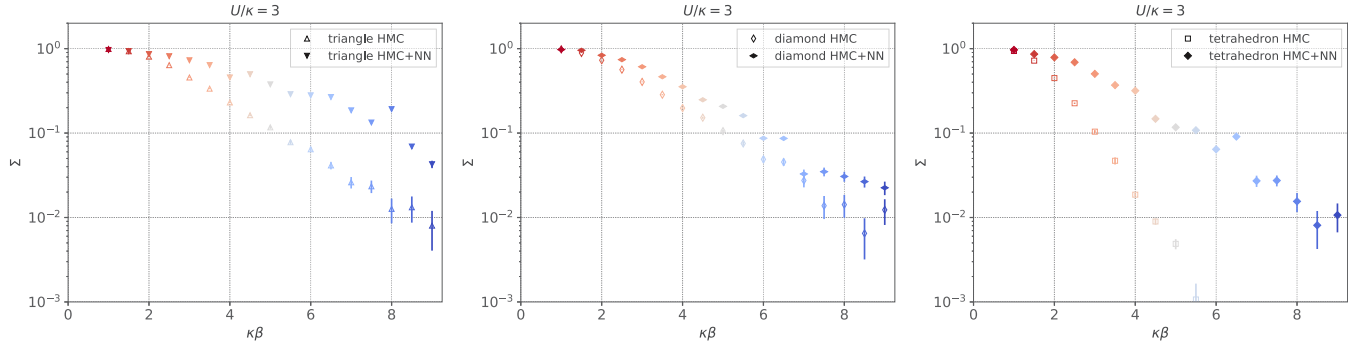


FIG. 15. The statistical power Σ for the triangle, diamond, and tetrahedron, with and without a neural-network-enhanced HMC, with $U/\kappa = 3$ and $N_t = 64$, as a function of inverse temperature $\kappa\beta$. The histograms (left) and statistical powers (right) are colored from red, a hot temperature, to blue, a cold temperature.

V. CONCLUSIONS

In this work, we adapted the learnifold method proposed by Alexandru, Bedaque, Lamm, and Lawrence [55] to alleviate the sign problem in small, frustrated Hubbard model examples. As shown in Figs. 5 and 10, our method definitively improves the statistical power, thus providing an exponential improvement in the sign problem. With the NN, it reproduces results obtained from exact diagonalization on a variety of correlation functions, cf. Figs. 7, 8, 12, 13, 16, 17. The agreement with exact results provides *a posteriori* evidence that we have an accurate, ergodic method.

By approximating the holomorphic flow with a neural network, we have developed an HMC method well-suited for tackling sign problems in the Hubbard model. As opposed to most previous work involving Lefschetz thimbles, our method does not rely on obtaining the precise location of the main thimble nor on the locations of less critical thimbles.

In fact, our neural networks perform better when not using prior knowledge of thimble locations. The reason is twofold. First, during HMC evolution, the Gaussian part of our action drives configurations to regions that are not dominated by a single thimble, but rather multiple thimbles. If we instead only trained on the main thimble, HMC would quickly force configurations into regions outside the trained network and we immediately encounter a numerical runaway problem—the imaginary part of the fields become arbitrarily large in an attempt to minimize the action. Second, our goal is to alleviate the sign problem to a point where standard reweighting techniques can be applied to obtain sufficiently accurate observables. Thus our maximum flow time was relatively small, yet sufficient to obtain accurate results. The modest flow times also prevented us from coming too close to points where the integration kernel vanishes, which in turn kept any ergodicity issues at bay. Though integrating directly on the

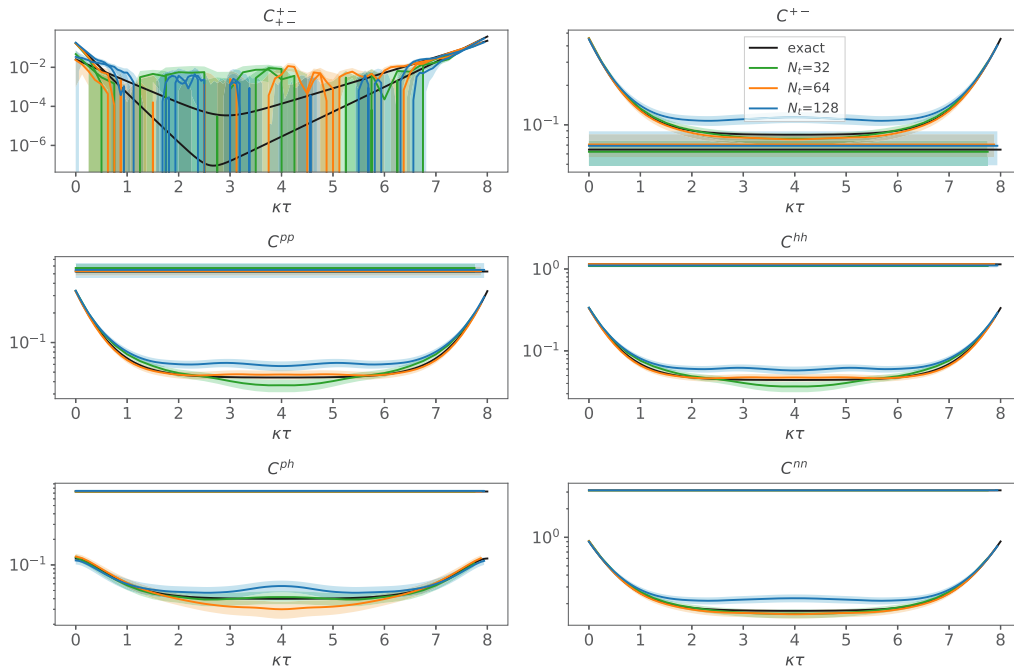


FIG. 16. Numerical results for some correlation functions on a triangle lattice with $U/\kappa = 3$, $\kappa\beta = 8$, and 10^5 configurations.

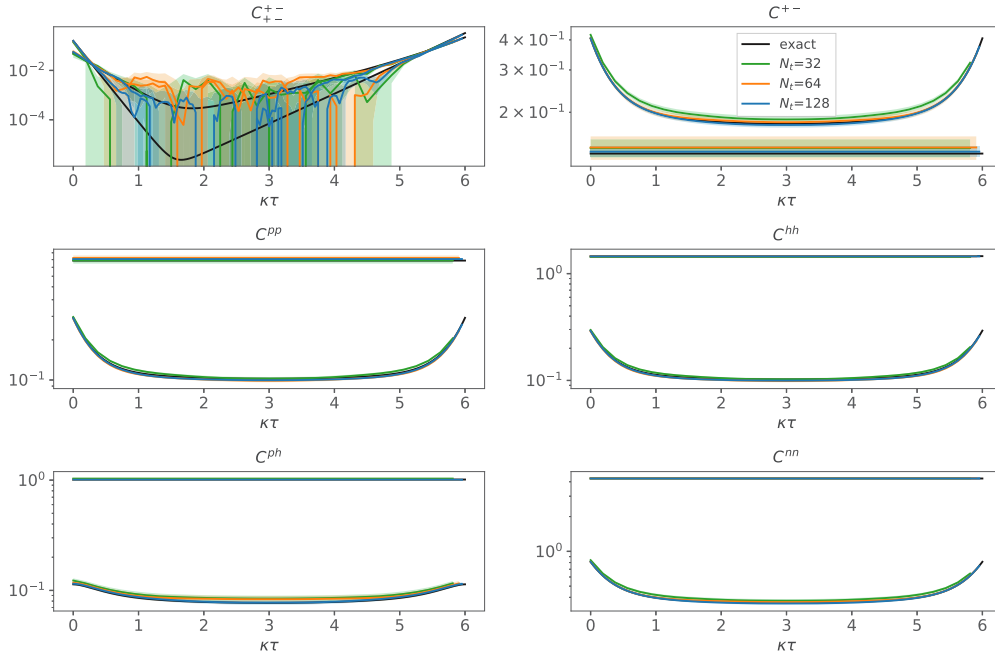


FIG. 17. Numerical results for some correlation functions on a tetrahedron lattice with $U/\kappa = 3$, $\kappa\beta = 8$, and 10^5 configurations.

thimbles eliminates the sign problem altogether, we believe that determining the exact location of these manifolds and the subsequent manifold integration is a daunting task (perhaps as difficult as the original sign problem) and becomes only more difficult if one considers gauge theories in higher dimensions.

A drawback of our method is still the need to calculate the determinant of a Jacobian induced by the complexified fields. Though our network was able to calculate the Jacobian much more quickly than a direct numerical flowing of the Jacobian as needed by holomorphic flow, the calculation still scales with the cubic power of the space-time volume. Thus calculations for systems with a larger spatial extent or additional time slices with our network method will ultimately run into this scaling barrier. Indeed, this was the driving reason for concentrating only on small system sizes in this study, and, therefore, our claim of a mitigated sign problem can only be made for such systems. However, the improved scaling, especially of the tetrahedron, towards the zero-temperature limit shown in Fig. 15 is heartening. We are actively investigating methods to improve the computational scaling, the goal being to test our method on larger systems. Already, we see some promise in networks that work directly with complex variables, see Appendix B.

The sign problem we addressed was due to the nonbipartite spatial lattices, thus providing a basis for extensions to quasi-zero-dimensional systems like C_{20} and the buckyball C_{60} .⁷ Our framework is nonetheless easily adapted to other sign problems. In future work, we anticipate simulating systems away from half filling and extended models with sufficiently strong nonlocal couplings [81], both of which suffer a sign problem not only extensive in euclidean time but also in space.

⁷We do not anticipate the calculation of the determinant of the Jacobian to be too onerous for these systems.

ACKNOWLEDGMENTS

This work was supported in part by the NSFC and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the funds provided to the Sino-German Collaborative Research Center TRR110 “Symmetries and the Emergence of Structure in QCD” (NSFC Grant No. 12070131001, DFG Project-ID 196253076 - TRR 110). E.B. is supported by the U.S. Department of Energy under Contract No. DE-FG02-93ER-40762. The authors gratefully acknowledge the computing time granted through JARA-HPC on the supercomputer JURECA [82] at Forschungszentrum Jülich.

APPENDIX A: CORRELATION FUNCTIONS OF BILINEAR OPERATORS

Spin-spin correlation functions are correlation functions between local spin operators S_x^i , where x is a lattice site and i runs over the indices of the Pauli matrices,

$$S_x^i = \frac{1}{2} \sum_{ss'} c_{xs} \sigma_{ss'}^i c_{xs'}^\dagger \quad (\text{A1})$$

and where c is a doublet of operators,

$$c_{xs} = \begin{pmatrix} a_x \\ (-\sigma_\kappa)^x b_x^\dagger \end{pmatrix} \quad (\text{A2})$$

where σ_κ is $+1$ on bipartite lattices and must be -1 on non-bipartite lattices, following the convention of Isle (bipartite graphs can also have $\sigma_\kappa = -1$). In Ref. [83], the authors also define, just after (1), the electric charge operator

$$\rho_x = c_{x,\uparrow}^\dagger c_{x,\uparrow} + c_{x,\downarrow}^\dagger c_{x,\downarrow} - 1, \quad (\text{A3})$$

which can be rewritten as $\rho_x = 1 - 2S_x^0$, where the 0^{th} Pauli matrix is the 2×2 identity matrix. The S operators are Hermitian. Rewriting those operators into the ISLE

basis,

$$\begin{aligned} S_x^0 &= \frac{1}{2}[a_x a_x^\dagger - b_x b_x^\dagger + 1], \\ S_x^1 &= \frac{1}{2}(-\sigma_\kappa)^x [b_x^\dagger a_x^\dagger + a_x b_x], \\ \rho_x &= n_x^a - n_x^b = 1 - 2S_x^0 = b_x b_x^\dagger - a_x a_x^\dagger, \\ S_x^2 &= \frac{i}{2}(-\sigma_\kappa)^x [b_x^\dagger a_x^\dagger - a_x b_x], \\ n_x &= n_x^a + n_x^b = 1 - 2S_x^3 = a_x^\dagger a_x + b_x^\dagger b_x, \\ S_x^3 &= \frac{1}{2}[a_x a_x^\dagger + b_x b_x^\dagger - 1], \end{aligned} \quad (\text{A4})$$

where the σ_κ squares away when two b operators are multiplied and no sum is implied on the right-hand sides. The other two spin bilinears have absolute charge 2,

$$(S_+^+)_x = a_x^\dagger b_x, \quad (S_-^-)_x = b_x^\dagger a_x. \quad (\text{A5})$$

There are, of course, other doubly charged operators but none that live on a single site, by Pauli exclusion.

The three spin operators obey the commutation relation

$$[S_x^i, S_y^j] = i\delta_{xy}\epsilon^{ijk}S_x^k, \quad (\text{A6})$$

which may be checked explicitly by writing out the operators and using the anticommutation properties of a and b . By a similar exercise, one may show

$$[S_x^0, S_y^j] = 0. \quad (\text{A7})$$

Single-particle and single-hole operators $\mathcal{O}$ with a definite third component of spin s_3 obey the operator eigenvalue equation

$$[S_x^3, \mathcal{O}_y] = s_3 \mathcal{O}_y \delta_{xy}. \quad (\text{A8})$$

This equation is satisfied when $(\mathcal{O}, s_3) = (a, +\frac{1}{2}), (a^\dagger, -\frac{1}{2}), (b, +\frac{1}{2}),$ and $(b^\dagger, -\frac{1}{2})$. Single-particle and single-hole operators $\mathcal{O}$ with a definite electric charge q obeys the operator eigenvalue equation

$$[\rho_x, \mathcal{O}_y] = q \mathcal{O}_y \delta_{xy}. \quad (\text{A9})$$

This equation is satisfied when $(\mathcal{O}, q) = (a, -1), (a^\dagger, +1), (b, +1),$ and $(b^\dagger, -1)$. Note that the signs differ from the S^3 case.

One may also construct spin raising and lowering operators in the standard way,

$$S_x^+ = S_x^1 + iS_x^2 = (-\sigma_\kappa)^x a_x b_x, \quad S_x^- = S_x^1 - iS_x^2 = (-\sigma_\kappa)^x b_x^\dagger a_x^\dagger, \quad (\text{A10})$$

which obey the eigenvalue relations

$$[S_x^3, S_y^\pm] = \pm S_y^\pm \delta_{xy}, \quad (\text{A11})$$

which can be shown using the single-particle and single-hole eigenvalue equations and the Leibniz rule.

The construction of the number operators proceeds in a similar fashion,

$$\delta_{xx} - n_x^a = S_x^0 + S_x^3 = a_x a_x^\dagger = \delta_{xx} - a_x^\dagger a_x, \quad (\text{A12})$$

$$n_x^b = S_x^0 - S_x^3 = -b_x b_x^\dagger + \delta_{xx} = -\delta_{xx} + b_x^\dagger b_x + \delta_{xx} = b_x^\dagger b_x. \quad (\text{A13})$$

We can of course drop the constant term in the first definition. The number operators obey the equations

$$[n_x^a, a_y] = -a_y \delta_{xy}, \quad [n_x^a, a^\dagger] = +a_y^\dagger \delta_{xy}, \quad (\text{A14})$$

and similarly for holes. It is easy to see using the eigenoperator equations (A8) and (A9) and the Leibniz rule that these operators commute with the local electric charge and spin, so that they have vacuum quantum numbers, while the doubly charged operators satisfy

$$[\rho_x, (S_+^+)_y] = +2(S_+^+)_y \delta_{xy}, \quad [\rho_x, (S_-^-)_y] = -2(S_-^-)_y \delta_{xy} \quad (\text{A15})$$

and are spin-0 because they commute with the spin operators. The one-point functions may be computed by Wick contraction

$$\begin{aligned} \frac{1}{N_t} \sum_t \langle n_x^a \rangle &= \left\langle 1 - \frac{1}{N_t} \sum_t P_{xtxt} \right\rangle, \\ \frac{1}{N_t} \sum_t \langle n_x^b \rangle &= \left\langle 1 - \frac{1}{N_t} \sum_t H_{xtxt} \right\rangle, \end{aligned} \quad (\text{A16})$$

where we denoted the Wick contraction of $a_{x,t_f} a_{y,t_i}^\dagger = (M^p)_{xt_f y t_i}^{-1} \equiv P_{xt_f y t_i}$, defining the particle propagator P , and similarly for holes $b_{x,t_f} b_{y,t_i}^\dagger = (M^h)_{xt_f y t_i}^{-1} \equiv H_{xt_f y t_i}$ the hole propagator. These may be combined according to (A4) to get one-point expectation values for n_x , ρ_x , and S_x^3 . Bilinears not having vacuum quantum numbers have vanishing one-point expectation values.

1. Correlation functions

Now we can write two-point correlation functions

$$C_{xy}^{uv}(\tau) = \frac{1}{N_t} \sum_t \langle S_{x,t+\tau}^u S_{y,t}^v \rangle \quad (\text{A17})$$

and we do not need to track time separately, until we start analyzing how to actually analyze these correlation functions via their spectral decompositions, though we always put the y position at the initial time $i = t$ and the x position at the final time $f = t + \tau$, so one can read x and y as superindices. When calculating numerically we sum over all initial time slices t to ensure the only time dependence is on the time difference τ . In these correlator expressions, x and y are unsummed.

The simplest correlation function is between S^+ and S^- ,

$$C_{xy}^{+-} = \langle S_x^+ S_y^- \rangle = (-\sigma_\kappa)^{x+y} \langle a_x b_x b_y^\dagger a_y^\dagger \rangle = (-\sigma_\kappa)^{x+y} \langle P_{xy} H_{xy} \rangle, \quad (\text{A18})$$

$$\begin{aligned} C_{xy}^{-+} &= \langle S_x^- S_y^+ \rangle = (-\sigma_\kappa)^{x+y} \langle b_x^\dagger a_x^\dagger a_y b_y \rangle \\ &= (-\sigma_\kappa)^{x+y} \langle (\delta_{yx} - b_y b_x^\dagger)(\delta_{yx} - a_y a_x^\dagger) \rangle \\ &= (-\sigma_\kappa)^{x+y} \langle (\delta_{yx} - H_{yx})(\delta_{yx} - P_{yx}) \rangle, \end{aligned} \quad (\text{A19})$$

where we have taken advantage of the anticommutator rules and that the Wick contractions yield the particle and hole propagators P and H (suppressing the time dependence for clarity). At half filling on a bipartite lattice, the cost to create or destroy a spin from the vacuum should be equal and the correlators should match, in the limit of large statistics. At

equal time $\tau = 0$, these correlation functions provide access to the spin-flip information [84].

Correlations between the number operator n^p and itself or n^h are also simple to write,

$$\begin{aligned} C_{xy}^{ph} &= \langle n_x^p n_y^h \rangle = \langle (\delta_{xx} - a_x a_x^\dagger)(\delta_{yy} - b_y b_y^\dagger) \rangle \\ &= \langle (\delta_{xx} - P_{xx})(\delta_{yy} - H_{yy}) \rangle, \end{aligned} \quad (\text{A20})$$

$$\begin{aligned} C_{xy}^{pp} &= \langle n_x^p n_y^p \rangle = \langle (\delta_{xx} - a_x a_x^\dagger)(\delta_{yy} - a_y a_y^\dagger) \rangle \\ &= \langle \delta_{xx} \delta_{yy} - a_x a_x^\dagger \delta_{yy} - \delta_{xx} a_y a_y^\dagger + a_x a_x^\dagger a_y a_y^\dagger \rangle \\ &= \langle \delta_{xx} \delta_{yy} - a_x a_x^\dagger \delta_{yy} - \delta_{xx} a_y a_y^\dagger + a_x (\delta_{xy} - a_y a_x^\dagger) a_y^\dagger \rangle \\ &= \langle \delta_{xx} \delta_{yy} - P_{xx} \delta_{yy} - \delta_{xx} P_{yy} + P_{xy} \delta_{xy} + P_{xx} P_{yy} - P_{xy} P_{yx} \rangle, \end{aligned} \quad (\text{A21})$$

and we can interchange the p/h species superscripts by exchanging the P and H propagators. While these correlators are between operators as simple as S^+ and S^- , computationally these Wick contractions are tougher to compute because they are “quark-line disconnected.”

We can also build correlators between the spin operators S^i . For example, C^{11} is given by

$$\begin{aligned} C_{xy}^{11} &= \langle S_x^1 S_y^1 \rangle \\ &= \frac{1}{4} (-\sigma_\kappa)^{x+y} \langle [b_x^\dagger a_x^\dagger + a_x b_x] [b_y^\dagger a_y^\dagger + a_y b_y] \rangle \\ &= \frac{1}{4} (-\sigma_\kappa)^{x+y} \langle a_x a_y^\dagger b_x b_y^\dagger + (\delta_{yx} - a_y a_x^\dagger)(\delta_{yx} - b_y b_x^\dagger) \rangle \\ &= \frac{1}{4} (-\sigma_\kappa)^{x+y} \langle P_{xy} H_{xy} + (\delta_{yx} - P_{yx})(\delta_{yx} - H_{yx}) \rangle, \end{aligned} \quad (\text{A22})$$

where we have used the fact that we will only get a nonzero result if we have the same number of a s as $a^\dagger$ s (and likewise for b) to drop the four-dagger and no-dagger terms. Computing C_{xy}^{22} requires

$$\langle S_x^2 S_y^2 \rangle = \frac{1}{4} (-\sigma_\kappa)^{x+y} \langle [b_x^\dagger a_x^\dagger - a_x b_x] [a_y b_y - b_y^\dagger a_y^\dagger] \rangle \quad (\text{A23})$$

though when written out in their complete glory, only the non-vanishing operator content in $\langle S_x^1 S_y^1 \rangle$ remains, so $C_{xy}^{22} = C_{xy}^{11}$ configuration by configuration (the vanishing operators have the opposite sign). In fact, using the definition of the spin raising and lowering operators (A10) one concludes

$$C_{xy}^{11} + C_{xy}^{22} = \frac{1}{2} (C_{xy}^{+-} + C_{xy}^{-+}), \quad C_{xy}^{12} - C_{xy}^{21} = \frac{i}{2} (C_{xy}^{+-} - C_{xy}^{-+}), \quad (\text{A24})$$

and we have explicitly checked the first identity by computing the Wick contractions (A18), (A19), and (A22), and the fact that $C_{xy}^{22} = C_{xy}^{11}$. It is easy to show that once the all-dagger or no-dagger operators are dropped,

$$\langle S_x^1 S_y^2 \rangle = -\langle S_x^2 S_y^1 \rangle \quad \text{so that} \quad C_{xy}^{12} = \frac{i}{4} (C_{xy}^{+-} - C_{xy}^{-+}) \quad (\text{A25})$$

which may be checked explicitly, and is true configuration by configuration.

The other two spins S^0 and S^3 do not enjoy such simplifications, because unlike the raising and lowering operators the number operators (A13) have vacuum quantum numbers, so there are no zero- or four-dagger terms which may be dropped from the Wick contractions. We are stuck computing four correlators:

$$\begin{aligned} C_{xy}^{00} &= \frac{1}{4} (C_{xy}^{pp} + C_{xy}^{hh} - C_{xy}^{ph} - C_{xy}^{hp} + \langle 1 - n_x^a - n_y^a + n_x^b + n_y^b \rangle), \\ C_{xy}^{03} &= \frac{1}{4} (C_{xy}^{pp} - C_{xy}^{hh} + C_{xy}^{ph} - C_{xy}^{hp} + \langle 1 - n_x^a - n_y^a + n_x^b - n_y^b \rangle), \\ C_{xy}^{30} &= \frac{1}{4} (C_{xy}^{pp} - C_{xy}^{hh} - C_{xy}^{ph} + C_{xy}^{hp} + \langle 1 - n_x^a - n_y^a - n_x^b + n_y^b \rangle), \\ C_{xy}^{33} &= \frac{1}{4} (C_{xy}^{pp} + C_{xy}^{hh} + C_{xy}^{ph} + C_{xy}^{hp} + \langle 1 - n_x^a - n_y^a - n_x^b - n_y^b \rangle), \end{aligned}$$

and we define

$$C_{xy}^{\rho\rho} = C_{xy}^{pp} + C_{xy}^{hh} - C_{xy}^{ph} - C_{xy}^{hp}, \quad (\text{A26})$$

$$C_{xy}^{\rho n} = C_{xy}^{pp} - C_{xy}^{hh} + C_{xy}^{ph} - C_{xy}^{hp}, \quad (\text{A27})$$

$$C_{xy}^{n\rho} = C_{xy}^{pp} - C_{xy}^{hh} - C_{xy}^{ph} + C_{xy}^{hp}, \quad (\text{A28})$$

$$C_{xy}^{nn} = C_{xy}^{pp} + C_{xy}^{hh} + C_{xy}^{ph} + C_{xy}^{hp}, \quad (\text{A29})$$

so that a ρ superscript indicates the charge operator (A4) and an n superscript the total number operator. The Wick contractions may be explicitly computed or built by rewriting the definition of the number operators (A13) as

$$S_x^0 = \frac{1}{2} (1 + n_x^b - n_x^a), \quad S_x^3 = \frac{1}{2} (1 - n_x^b - n_x^a), \quad (\text{A30})$$

and using the particle-hole (A20) and particle-particle (A21) correlators and the one-point functions (A16). Note that $S^{1,2}$ cannot be correlated with $S^{0,3}$ because each term would not have the right constituent operator content to contract completely, so those correlators automatically vanish.

The doubly charged operators have simple Wick contractions. Note that $(S_+^\pm)^\dagger = (S_-^\mp)$ so that

$$(C_{+-}^{+-})_{xy} = \langle a_x^\dagger b_x b_y^\dagger a_y \rangle = \langle a_x^\dagger a_y b_x b_y^\dagger \rangle = \langle (\delta_{yx} - a_y a_x^\dagger) b_x b_y^\dagger \rangle = \langle (\delta_{yx} - P_{yx}) H_{xy} \rangle, \quad (\text{A31})$$

$$C_{-+}^{-+} = [C_{+-}^{+-} \text{ with } P \leftrightarrow H]. \quad (\text{A32})$$

Based on the exact results, a Lepage-style argument [85] suggests these doubly-charged correlators should suffer from a signal-to-noise problem.

2. Conserved quantities

When the Hamiltonian takes a Hubbard-Coulomb-like form,

$$H = - \sum_{xy} (a_x^\dagger h_{xy} a_y + \sigma_\kappa b_x^\dagger h_{xy} b_y) + \frac{1}{2} \sum_{xy} \rho_x V_{xy} \rho_y, \quad (\text{A33})$$

some of the bilinears may correspond to conserved quantities. For example, we can calculate the commutator with a local charge density operator,

$$[H, \rho_z] = \left[\sum_{xy} a_x^\dagger h_{xy} a_y + \sigma_\kappa b_x^\dagger h_{xy} b_y, \rho_z \right] \quad (\text{A34})$$

$$= \sum_x -a_x^\dagger h_{xz} a_z + \sum_y a_z^\dagger h_{zy} a_y - \sigma_\kappa (a \leftrightarrow b), \quad (\text{A35})$$

where we immediately dropped the interaction term since the charge operator commutes with itself. If we sum z over all space the two terms cancel, so that the total charge

$$Q = \sum_z \rho_z \quad (\text{A36})$$

is conserved. One similarly finds the total spins in each direction conserved,

$$[H, S^i] = 0, \quad S^i = \sum_z S_z^i \quad (\text{A37})$$

for $i \in \{1, 2, 3\}$ and in fact the total spin also commutes with the Hamiltonian

$$[H, S^2] = 0, \quad S^2 = \sum_i (S^i)^2. \quad (\text{A38})$$

When the operators are conserved, their two-point correlation functions are constant,

$$\begin{aligned} C^{QQ}(\tau) &= \frac{1}{Z} \text{tr}[Q(\tau)Q(0)e^{-\beta H}] \\ &= \frac{1}{Z} \text{tr}[e^{+H\tau} Q(0)e^{-H\tau} Q(0)e^{-\beta H}] \\ &= \frac{1}{Z} \text{tr}[Q(0)e^{-H\tau} Q(0)e^{-(\beta-\tau)H}] \\ &= \frac{1}{Z} \text{tr}[Q(0)Q(0)e^{-\beta H}], \end{aligned} \quad (\text{A39})$$

where we wrote the Heisenberg-picture $Q(\tau)$ in terms of the zero-time operator and the Hamiltonian and repeatedly used the fact that Q commutes with the Hamiltonian. We can turn this relation on its head and get an estimate for the equal-time correlator $\langle Q^2 \rangle$ by averaging over the temporal separation,

$$\langle Q^2 \rangle = \frac{1}{N_t} \sum_\tau C^{QQ}(\tau) = \frac{1}{N_t} \sum_\tau C_{++}^{\rho\rho}(\tau), \quad (\text{A40})$$

where a $+$ subscript indicates that the spatial index is summed over—in this case, implementing (A36). This same observation holds for the total spin operators S^i (and therefore also for

$S^\pm$), with the Hamiltonian shown above. We can measure the mean-squared magnetization $\langle S^2 \rangle$ by

$$\langle S^2 \rangle = \sum_{i=1}^3 \frac{1}{N_t} \sum_\tau C_{++}^{ii}(\tau). \quad (\text{A41})$$

On small test examples one observes numerically that the correlators are flat with the exponential discretization and seem linear with time in the diagonal discretization; averaging properly still yields good values—for an explicit example see the last Appendix of Ref. [62].

3. Numerical results

This section shows results for additional correlators that are not covered in the main text. Correlators in Figs. 16 and 17 are computed on the same ensembles (“network”) as those in Secs. IV A 2 and IV B 1, respectively.

APPENDIX B: COUPLING LAYERS

As described above, the networks used here have a significant disadvantage, their Jacobians are expensive to compute. Using coupling layers with a suitable coupling rule as introduced in Ref. [86] instead of dense layers, can effectively reduce that cost to zero in most applications.

For $x, y \in \mathbb{R}^\Lambda$, let $\{A, B\}$ be a partition of the integer interval $\llbracket 1, \Lambda \rrbracket$. Λ is the lattice space-time volume $N_x N_t$ and we restrict ourselves to even Λ and $|A| = |B| = \Lambda/2$. A coupling layer is defined as

$$f(x) = \begin{cases} y_A = x_A \\ y_B = g(x_A, x_B) \end{cases}. \quad (\text{B1})$$

Here, we focus on affine coupling layers which have a coupling rule of⁸

$$g(x_A, x_B) = x_B \odot s(x_A) + t(x_A), \quad (\text{B2})$$

where $\odot$ denotes element wise multiplication and s and t are arbitrary functions which can be parameterized through neural networks. The Jacobian determinant of a network made out of layers, meaning $\text{NN}(x) = f^n(f^{n-1}(\dots f^1(x)))$, factorizes into separate determinants for each layer:

$$\begin{aligned} \det J &= \det \frac{\partial \text{NN}}{\partial x} \\ &= \det \left(\frac{\partial f^n(x)}{\partial x} \right) \det \left(\frac{\partial f^{n-1}(x)}{\partial x} \right) \dots \det \left(\frac{\partial f^1(x)}{\partial x} \right). \end{aligned} \quad (\text{B3})$$

In the case of affine coupling layers, the rows and columns of the layer Jacobians can be permuted under the determinants such that for each layer $A = \llbracket 1, \Lambda/2 \rrbracket$, $B = \llbracket \Lambda/2 + 1, \Lambda \rrbracket$ thus making the matrices triangular and the determinant fast

⁸The multiplicative part is typically written as $e^{s(x_A)}$ in the literature in order to simplify inversion. Since we do not need to invert the network here, we do not need the exponential.

to compute:

$$\det\left(\frac{\partial f}{\partial x}\right) = \det\left(\begin{pmatrix} \mathbb{1}_{\Lambda/2} & 0 \\ \frac{\partial y_B}{\partial x_A} & s(x_A) \end{pmatrix}\right) = \prod_{i=1}^{\Lambda/2} s(x_A)_i, \quad (\text{B4})$$

where $s(x_A)$ is to be understood as a diagonal matrix in the second expression.

This procedure can not be applied with the approach taken in this work as the transformation Eq. (26) has Jacobian [see Eq. (28)]

$$\det J[\phi] = \det\left(\mathbb{1} + i \frac{\partial \text{NN}(\phi)}{\partial \phi}\right). \quad (\text{B5})$$

This determinant does not factorize and we can not bring the derivative of each layer into triangular shape. A and B can also not be chosen such that the product of layers $\partial \text{NN}(\phi)/\partial \phi$ itself is triangular because that would mean that not all components of ϕ could influence all components of $\tilde{\phi}$ which would limit the expressive power of the network.

The root cause for the problem is the different treatment of real and imaginary parts in Eq. (26) because it leads to the “ $\mathbb{1}+$ ” in the determinant. A potential solution would be changing the transformation to $\tilde{\phi} = \text{NN}(\phi)$. However, this

requires networks which deal with complex numbers and forgoes the potential benefits of Eq. (26) described in the main text. It is not trivial to formulate complex valued neural networks. One difficulty comes from the loss functions which are typically nonanalytic. This problem can be avoided by using Wirtinger calculus to construct an optimization procedure [87,88]. Another difficulty is finding complex activation functions [89]. Here, the problem is complicated further because the chain rule of Wirtinger derivatives contains a sum:

$$z = f(y), y = g(x) \Rightarrow \frac{\partial z}{\partial x} = \frac{\partial f}{\partial y} \frac{\partial g}{\partial x} + \frac{\partial f}{\partial y^*} \frac{\partial g^*}{\partial x}. \quad (\text{B6})$$

This would ultimately produce a sum in the determinant of the Jacobian, preventing factorization. The efficient Jacobian of Eq. (B4) can thus only be realized with holomorphic neural networks, which implies holomorphic activation functions. For this first stage of our work we relied on professionally-optimized, third-party machine learning libraries but support for complex-valued networks is currently incomplete. We plan to develop an implementation with support for complex-valued coupling layers and pursue the approach described here.

-
- [1] K. Splittorff and J. J. M. Verbaarschot, The QCD sign problem for small chemical potential, *Phys. Rev. D* **75**, 116003 (2007).
 - [2] J. Danzer, C. Gatttringer, L. Liptak, and M. Marinkovic, A Study of the sign problem for lattice QCD with chemical potential, *Phys. Lett. B* **68**, 240 (2009).
 - [3] N. Garron and K. Langfeld, Anatomy of the sign-problem in heavy-dense QCD, *Eur. Phys. J. C* **76**, 569 (2016).
 - [4] S. D. H. Hsu and D. Reeb, On the sign problem in dense QCD, *Int. J. Mod. Phys. A* **25**, 53 (2010).
 - [5] V. A. Goy, V. Bornyakov, D. Boyda, A. Molochkov, A. Nakamura, A. Nikolaev, and V. Zakharov, Sign problem in finite density lattice QCD, *Prog. Theor. Exp. Phys.* **2017**, 031D01 (2017).
 - [6] A. Shindler, T. Luu, and J. de Vries, Nucleon electric dipole moment with the gradient flow: The θ -term contribution, *Phys. Rev. D* **92**, 094518 (2015).
 - [7] R. V. Gavai and S. Gupta, QCD at finite chemical potential with six time slices, *Phys. Rev. D* **78**, 114503 (2008).
 - [8] J. Dragos, T. Luu, A. Shindler, J. de Vries, and A. Youisf, Confirming the existence of the strong CP problem in lattice QCD with the gradient flow, *Phys. Rev. C* **103**, 015202 (2021).
 - [9] T. Lähde and U.-G. Meißner, *Nuclear Lattice Effective Field Theory*, Lecture Notes in Physics 957 (Springer International Publishing, Cham, 2019).
 - [10] T. A. Lähde, T. Luu, D. Lee, U.-G. Meißner, E. Epelbaum, H. Krebs, and G. Rupak, Nuclear lattice simulations using symmetry-sign extrapolation, *Eur. Phys. J. A* **51**, 92 (2015).
 - [11] S. Elhatisari, E. Epelbaum, H. Krebs, T. A. Lähde, D. Lee, N. Li, B.-N. Lu, U.-G. Meißner, and G. Rupak, *Ab initio* Calculations of the Isotopic Dependence of Nuclear Clustering, *Phys. Rev. Lett.* **119**, 222505 (2017).
 - [12] R. Matsunaga, K. Matsuda, and Y. Kanemitsu, Observation of Charged Excitons in Hole-Doped Carbon Nanotubes Using Photoluminescence and Absorption Spectroscopy, *Phys. Rev. Lett.* **106**, 037404 (2011).
 - [13] T. Deilmann and M. Rohlfing, Huge trionic effects in graphene nanoribbons, *Nano Lett.* **17**, 6833 (2017).
 - [14] A. Arora, T. Deilmann, T. Reichenauer, J. Kern, S. Michaelis de Vasconcellos, M. Rohlfing, and R. Bratschitsch, Excited-State Trions in Monolayer WS₂, *Phys. Rev. Lett.* **123**, 167401 (2019).
 - [15] M. Gaulke, A. Janissek, N. A. Peyyety, I. Alamgir, A. Riaz, S. Dehm, H. Li, U. Lemmer, B. S. Flavel, M. M. Kappes, F. Hennrich, L. Wei, Y. Chen, F. Pyatkov, and R. Krupke, Low-temperature electroluminescence excitation mapping of excitons and trions in short-channel monochiral carbon nanotube devices, *ACS Nano* **14**, 2709 (2020).
 - [16] M. Troyer and U.-J. Wiese, Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations, *Phys. Rev. Lett.* **94**, 170201 (2005).
 - [17] M. D’Elia and M.-P. Lombardo, Finite density QCD via imaginary chemical potential, *Phys. Rev. D* **67**, 014505 (2003).
 - [18] R. Bellwied, S. Borsanyi, Z. Fodor, J. Günther, S. D. Katz, C. Ratti, and K. K. Szabo, The QCD phase diagram from analytic continuation, *Phys. Lett. B* **751**, 559 (2015).
 - [19] V. Vovchenko, A. Pasztor, Z. Fodor, S. D. Katz, and H. Stoecker, Repulsive baryonic interactions and lattice QCD observables at imaginary chemical potential, *Phys. Lett. B* **775**, 71 (2017).
 - [20] G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Langevin simulations of lattice field theories, *Phys. Rev. D* **32**, 2736 (1985).
 - [21] G. G. Batrouni, Variations on the Langevin equation for lattice QCD with fermions, *Phys. Rev. D* **33**, 1815 (1986).
 - [22] M. Fukugita, Y. Oyanagi, and A. Ukawa, Langevin simulation of the full QCD hadron mass spectrum on a lattice, *Phys. Rev. D* **33**, 824 (1987).

- [23] P. H. Damgaard and H. Hüffel, Stochastic quantization, *Phys. Rep.* **152**, 227 (1987).
- [24] J. B. Kogut and D. K. Sinclair, Applying complex Langevin simulations to lattice QCD at finite density, *Phys. Rev. D* **100**, 054512 (2019).
- [25] D. Sexty, Progress in complex Langevin simulations of full QCD at non-zero density, *Nucl. Phys. A* **931**, 856 (2014).
- [26] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [27] F. Verstraete and J. I. Cirac, Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions, [arXiv:cond-mat/0407066v1](https://arxiv.org/abs/cond-mat/0407066v1).
- [28] R. Orus, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, *Ann. Phys.* **349**, 117 (2014).
- [29] P. Corboz and G. Vidal, Fermionic multiscale entanglement renormalization ansatz, *Phys. Rev. B* **80**, 165129 (2009).
- [30] P. Corboz, P. Czarnik, G. Kapteijns, and L. Tagliacozzo, Finite Correlation Length Scaling with Infinite Projected Entangled-Pair States, *Phys. Rev. X* **8**, 031031 (2018).
- [31] P. Corboz, Improved energy extrapolation with infinite projected entangled-pair states applied to the two-dimensional Hubbard model, *Phys. Rev. B* **93**, 045116 (2016).
- [32] A. Ammon, T. Hartung, K. Jansen, H. Leövey, and J. Volmer, Overcoming the sign problem in one-dimensional QCD by new integration rules with polynomial exactness, *Phys. Rev. D* **94**, 114508 (2016).
- [33] M. Cristoforetti, F. Di Renzo, and L. Scorzato (AuroraScience Collaboration), New approach to the sign problem in quantum field theories: High density QCD on a Lefschetz thimble, *Phys. Rev. D* **86**, 074506 (2012).
- [34] S. Lawrence, Beyond thimbles: Sign-optimized manifolds for finite density, *Proc. Sci.* **334**, 149 (2018).
- [35] M. Cristoforetti, F. Di Renzo, G. Erucci, A. Mukherjee, C. Schmidt, L. Scorzato, and C. Torrero, An efficient method to compute the residual phase on a Lefschetz thimble, *Phys. Rev. D* **89**, 114505 (2014).
- [36] M. Cristoforetti, F. Di Renzo, A. Mukherjee, and L. Scorzato, Monte Carlo simulations on the Lefschetz thimble: Taming the sign problem, *Phys. Rev. D* **88**, 051501(R) (2013).
- [37] A. Mukherjee, M. Cristoforetti, and L. Scorzato, Metropolis Monte Carlo integration on the Lefschetz thimble: Application to a one-plaquette model, *Phys. Rev. D* **88**, 051502(R) (2013).
- [38] H. Fujii, S. Kamata, and Y. Kikukawa, Lefschetz thimble structure in one-dimensional lattice Thirring model at finite density, *J. High Energy Phys.* **11** (2015) 078 [Erratum: **02** (2016) 036].
- [39] H. Fujii, S. Kamata, and Y. Kikukawa, Monte Carlo study of Lefschetz thimble structure in one-dimensional Thirring model at finite density, *J. High Energy Phys.* **12** (2015) 125 [Erratum: **09** (2016) 172].
- [40] T. Kanazawa and Y. Tanizaki, Structure of Lefschetz thimbles in simple fermionic systems, *J. High Energy Phys.* **03** (2015) 044.
- [41] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, Sign problem and Monte Carlo calculations beyond Lefschetz thimbles, *J. High Energy Phys.* **05** (2016) 053.
- [42] A. Alexandru, G. Basar, and P. Bedaque, Monte Carlo algorithm for simulating fermions on Lefschetz thimbles, *Phys. Rev. D* **93**, 014504 (2016).
- [43] Y. Tanizaki, Y. Hidaka, and T. Hayata, Lefschetz-thimble analysis of the sign problem in one-site fermion model, *New J. Phys.* **18**, 033002 (2016).
- [44] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, Monte Carlo calculations of the finite density Thirring model, *Phys. Rev. D* **95**, 014502 (2017).
- [45] M. Ulybyshev, C. Winterowd, and S. Zafeiropoulos, Taming the sign problem of the finite density Hubbard model via Lefschetz thimbles, [arXiv:1906.02726](https://arxiv.org/abs/1906.02726).
- [46] C. Schmidt and F. Ziesché, Simulating low dimensional QCD with Lefschetz thimbles, *Proc. Sci.* **256**, 076 (2017).
- [47] K. Zambello and F. Di Renzo, Towards Lefschetz thimbles regularization of heavy-dense QCD, *Proc. Sci.* **334**, 148 (2018).
- [48] A. Alexandru, G. Basar, P. F. Bedaque, H. Lamm, and S. Lawrence, Finite density $1 + 1$ QED near Lefschetz thimbles, *Phys. Rev. D* **98**, 034506 (2018).
- [49] J. M. Pawłowski, M. Scherzer, C. Schmidt, F. P. G. Ziegler, and F. Ziesché, Simulating gauge theories on Lefschetz thimbles, *Proc. Sci.* **363**, 223 (2019).
- [50] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence, Finite-density Monte Carlo calculations on sign-optimized manifolds, *Phys. Rev. D* **97**, 094510 (2018).
- [51] A. Alexandru, P. F. Bedaque, H. Lamm, S. Lawrence, and N. C. Warrington, Fermions at Finite Density in $2 + 1$ Dimensions with Sign-Optimized Manifolds, *Phys. Rev. Lett.* **121**, 191602 (2018).
- [52] Y. Mori, K. Kashiwa, and A. Ohnishi, Application of a neural network to the sign problem via the path optimization method, *Prog. Theor. Exp. Phys.* **2018**, 023B04 (2018).
- [53] J. Nishimura and S. Shimasaki, Combining the complex Langevin method and the generalized Lefschetz-thimble method, *J. High Energy Phys.* **06** (2017) 023.
- [54] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, Fast estimator of Jacobians in the Monte Carlo integration on Lefschetz thimbles, *Phys. Rev. D* **93**, 094514 (2016).
- [55] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence, Deep learning beyond lefschetz thimbles, *Phys. Rev. D* **96**, 094505 (2017).
- [56] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signals Syst.* **2**, 303 (1989).
- [57] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks* **2**, 359 (1989).
- [58] R. C. Brower, D. Schaich, and C. Rebbi, Hybrid Monte Carlo simulation on the graphene hexagonal lattice, *Proc. Sci.* **139**, 056 (2011).
- [59] 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 Realistic Inter-Electron Interaction Potential, *Phys. Rev. Lett.* **111**, 056801 (2013).
- [60] 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**, 195429 (2014).
- [61] T. Luu and T. A. Lähde, Quantum Monte Carlo calculations for carbon nanotubes, *Phys. Rev. B* **93**, 155106 (2016).
- [62] 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**, 075141 (2019).

- [63] S. Daul and D. J. Scalapino, Frustrated hubbard ladders and superconductivity in κ -BEDT-TTF organic compounds, *Phys. Rev. B* **62**, 8658 (2000).
- [64] W. Z. Wang, Ground states in a quasi-one-dimensional triangular hubbard model, *Phys. Rev. B* **72**, 125116 (2005).
- [65] O. Rojas, S. M. de Souza, and N. S. Ananikian, Geometrical frustration of an extended hubbard diamond chain in the quasi-atomic limit, *Phys. Rev. E* **85**, 061123 (2012).
- [66] R. Schumann, Rigorous solution of a hubbard model extended by nearest-neighbour coulomb and exchange interaction on a triangle and tetrahedron, *Ann. Phys.* **17**, 221 (2008).
- [67] K. Hattori and H. Tsunetsugu, Effective hamiltonian of a three-orbital hubbard model on the pyrochlore lattice: Application to LiV_2O_4 , *Phys. Rev. B* **79**, 035115 (2009).
- [68] T. Yoshioka, A. Koga, and N. Kawakami, Mott transition in the hubbard model on checkerboard lattice, *J. Phys. Soc. Jpn.* **77**, 104702 (2008).
- [69] T. Yoshioka, A. Koga, and N. Kawakami, Frustration effects in an anisotropic checkerboard lattice Hubbard model, *Phys. Rev. B* **78**, 165113 (2008).
- [70] Y. Furukawa, T. Ohashi, Y. Koyama, and N. Kawakami, Mott transition in the Hubbard model on the anisotropic kagome lattice, *Phys. Rev. B* **82**, 161101(R) (2010).
- [71] M. Ulybyshev, C. Winterowd, and S. Zafeiropoulos, Lefschetz thimbles decomposition for the Hubbard model on the hexagonal lattice, *Phys. Rev. D* **101**, 014508 (2020).
- [72] Y. Mori, K. Kashiwa, and A. Ohnishi, Lefschetz thimbles in fermionic effective models with repulsive vector-field, *Phys. Lett. B* **781**, 688 (2018).
- [73] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).
- [74] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems* 32, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates Inc., Red Hook, NY, 2019), pp. 8024–8035.
- [75] J.-L. Wynen and E. Berkowitz, jl-wynen/isle: Release 0.3, doi:[10.5281/zenodo.3834729](https://doi.org/10.5281/zenodo.3834729).
- [76] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [77] M. Fukuma, N. Matsumoto, and N. Umeda, Implementation of the HMC algorithm on the tempered Lefschetz thimble method, [arXiv:1912.13303](https://arxiv.org/abs/1912.13303).
- [78] M. Fukuma, N. Matsumoto, and N. Umeda, Applying the tempered Lefschetz thimble method to the Hubbard model away from half-filling, *Phys. Rev. D* **100**, 114510 (2019).
- [79] S. Beyl, F. Goth, and F. F. Assaad, Revisiting the hybrid quantum Monte Carlo method for Hubbard and electron-phonon models, *Phys. Rev. B* **97**, 085144 (2018).
- [80] K. Splittorff and J. J. M. Verbaarschot, Phase of the Fermion Determinant at Nonzero Chemical Potential, *Phys. Rev. Lett.* **98**, 031601 (2007).
- [81] P. Buividovich, D. Smith, M. Ulybyshev, and L. von Smekal, Competing order in the fermionic Hubbard model on the hexagonal graphene lattice, *Proc. Sci.* **256**, 244 (2016).
- [82] Jülich Supercomputing Centre, JURECA: Modular supercomputer at Jülich supercomputing centre, *J. Large-Scale Res. Facil.* **4**, A132 (2018).
- [83] P. Buividovich, D. Smith, M. Ulybyshev, and L. von Smekal, Hybrid Monte Carlo study of competing order in the extended fermionic Hubbard model on the hexagonal lattice, *Phys. Rev. B* **98**, 235129 (2018).
- [84] L. Colmenarez, P. A. McClarty, M. Haque, and D. J. Luitz, Statistics of correlations functions in the random Heisenberg chain, *SciPost Phys.* **7**, 64 (2019).
- [85] G. P. Lepage, *Theoretical Advanced Study Institute in Elementary Particle Physics*, Report No. CLNS-89-971 (1989).
- [86] L. Dinh, D. Krueger, and Y. Bengio, NICE: non-linear independent components estimation, [arXiv:1410.8516](https://arxiv.org/abs/1410.8516).
- [87] D. H. Brandwood, A complex gradient operator and its application in adaptive array theory, *IEEE Proc. F* **130**, 11 (1983).
- [88] K. Kreutz-Delgado, The complex gradient operator and the CR-calculus, [arXiv:0906.4835](https://arxiv.org/abs/0906.4835).
- [89] S. Scardapane, S. V. Vaerenbergh, A. Hussain, and A. Uncini, Complex-valued neural networks with non-parametric activation functions, [arXiv:1802.08026](https://arxiv.org/abs/1802.08026).