

Avoiding ergodicity problems in lattice discretizations of the Hubbard model

Jan-Lukas Wynen¹,² Evan Berkowitz,^{1,2} Christopher Körber,^{1,2,3,4} Timo A. Lähde,^{1,2} and Thomas Luu^{1,2,3}

¹*Institute for Advanced Simulation, Forschungszentrum Jülich, 52445 Jülich, Germany*

²*Institut für Kernphysik, Forschungszentrum Jülich, 52445 Jülich, Germany*

³*Helmholtz-Institut für Strahlen- und Kernphysik, Rheinische Friedrich-Williams-Universität Bonn, 53012 Bonn, Germany*

⁴*Department of Physics, University of California, Berkeley, California 94720, USA*



(Received 23 January 2019; revised manuscript received 26 July 2019; published 22 August 2019)

The Hubbard model arises naturally when electron-electron interactions are added to the tight-binding descriptions of many condensed matter systems. For instance, the two-dimensional Hubbard model on the honeycomb lattice is central to the *ab initio* description of the electronic structure of carbon nanomaterials, such as graphene. Such low-dimensional Hubbard models are advantageously studied with Markov chain Monte Carlo methods, such as hybrid Monte Carlo (HMC). HMC is the standard algorithm of the lattice gauge theory community, as it is well suited to theories of dynamical fermions. As HMC performs continuous, global updates of the lattice degrees of freedom, it provides superior scaling with system size relative to local updating methods. A potential drawback of HMC is its susceptibility to ergodicity problems due to so-called exceptional configurations, for which the fermion operator cannot be inverted. Recently, ergodicity problems were found in some formulations of HMC simulations of the Hubbard model. Here, we address this issue directly and clarify under what conditions ergodicity is maintained or violated in HMC simulations of the Hubbard model. We study different lattice formulations of the fermion operator and provide explicit, representative calculations for small systems, often comparing to exact results. We show that a fermion operator can be found which is both computationally convenient and free of ergodicity problems.

DOI: [10.1103/PhysRevB.100.075141](https://doi.org/10.1103/PhysRevB.100.075141)

I. INTRODUCTION

Modern lattice QCD simulations are performed using the hybrid Monte Carlo (HMC) algorithm [1,2]. While the basic structure of the HMC algorithm has remained unchanged since it was introduced, much effort has been directed toward the development of efficient iterative solvers, with accelerated convergence [3]. Such highly optimized solvers have been instrumental in tearing down the so-called computational “Berlin wall,” which for a long time prevented simulations with dynamical fermions at the physical pion mass, for systems of a realistic size [4–8]. An interesting aspect of lattice QCD is that a large freedom of choice exists for the formulation of the lattice action, provided that the proper continuum limit is recovered as the lattice spacing $a \rightarrow 0$. This freedom has been exploited to formulate lattice actions that conserve more symmetries at finite a , or sacrifice exact symmetries in order to gain computational performance or scaling with system size V . This is particularly true for chiral symmetry in lattice QCD, which is intimately connected to the fermion “doubling problem” through the Nielsen-Ninomiya theorem [9,10]. While lattice actions have been formulated which circumvent the doubling problem, and maintain exact or near-exact chiral symmetry at nonzero a (such as overlap [11–14] or domain wall [15–17] fermions), they typically come at a high computational cost compared with lattice actions where exact chiral symmetry needs to be recovered by extrapolation. For this reason, the development of lattice QCD has tended to emphasize computational simplicity and efficiency, together

with Symanzik improvement [18,19] of the lattice action, which accelerates the approach to the continuum limit by the systematic removal of lattice artifacts.

Remarkably, HMC has rarely been applied to problems in condensed matter physics. In part, this can be traced back to the higher dimensionality of QCD—lattice QCD researchers have few options besides relying on the expected $\sim V^{5/4}$ scaling of HMC [20]. In contrast, Monte Carlo (MC) simulations of Hamiltonian theories with electron-electron interactions, such as the Hubbard model [21–23], are usually performed using one of the many possible formulations of auxiliary field quantum Monte Carlo (AFQMC). In the Blankenbecler-Sugar-Scalapino (BSS) algorithm [24–27], the four-fermion interactions are split by means of a Hubbard-Stratonovich (HS) or “auxiliary” field ϕ , which is sometimes taken to be discrete [25], in analogy with lattice models of spin systems. Unlike the case of lattice QCD, early attempts at combining BSS with HMC updates [28,29] were not pursued further, largely because the configuration space of auxiliary fields becomes increasingly fragmented into regions with positive and negative fermion determinant at low temperatures. The HMC algorithm requires a continuous (rather than discrete) auxiliary field, and furthermore $M[\phi]$ should be invertible at every point during the HMC Hamiltonian update, or *trajectory*, for a new configuration proposal. At the boundaries between regions of different sign $\det M[\phi]$ vanishes, which causes HMC to become trapped in the region of the starting configuration. As a result, algorithms which combine BSS

with HMC updates are in general not ergodic.¹ Note that similar ergodicity problems appear for overlap fermions in Lattice QCD, which has prompted the development of a class of HMC algorithms which can reflect from or refract through boundaries where $\det M[\phi]$ changes sign [30–32]. However, such algorithms are rather costly computationally, which suggests that a more amenable approach is to find a lattice action free from ergodicity problems.

The Hubbard and Hubbard/Coulomb models are key components of the *ab initio* description of electron-electron interactions in low-dimensional materials [33,34] (e.g., graphene), so the problem of combining HMC with the Hubbard Hamiltonian has recently been revisited.² The dimensionality of such systems is between that of QCD and problems of great interest in applied condensed matter physics and materials science, such as quantum dots and nanoribbons [35,36]. For any 2d lattice, one therefore expects that the (expected) superior scaling with V of HMC over BSS would be advantageous in the study of critical phenomena, such as high- T_c superconductivity for the square lattice or the antiferromagnetic Mott insulating (AFMI) phase for the honeycomb lattice (and possibly other types of spin-liquid phases) which appears at moderate values of the on-site electron-electron coupling U [37–39]. Recently, new attempts have been made by Beyl *et al.* in Ref. [40] to circumvent the ergodicity problem of the BSS + HMC method by making the auxiliary field ϕ complex-valued. While this doubles the number of lattice degrees of freedom, it would represent an acceptable trade-off, if the computational scaling would be dramatically improved. However, it was found that HMC simulations of the complexified theory failed to deliver the expected $\sim V^{5/4}$ scaling. For the Su-Schrieffer-Heeger (SSH) electron-phonon Hamiltonian [41], which lacks the ergodicity problem of the Hubbard case, this scaling was realized. The recent studies of Ref. [42] also found favorable scaling with HMC when using Hasenbusch preconditioning.

From the perspective of the lattice QCD community, MC simulations of the Hubbard model have several attractive features. Some of these features are the apparent simplicity of the Hubbard model in comparison to the QCD Lagrangian, the possibility to identify the spatial lattice discretization a with the physical lattice spacing, and the potential of applying the versatile and sophisticated toolbox of numerical methods developed for Lattice QCD to a problem with many promising applications. The seminal work of Ref. [43] introduced the Brower-Rebbi-Schaich (BRS) algorithm for the Hubbard model, which is inspired by lattice QCD methods. The BRS algorithm has recently been

applied not only to graphene [44,45], but also to carbon nanotubes [46,47]. The BSS and BRS algorithms are closely related. The main differences are the treatment of the hopping matrix h in the fermion operator M and that BRS uses a purely imaginary auxiliary field. Specifically, BSS uses a “compact” operator $M[\phi] \sim \mathbb{1} - \exp(h) \exp(\phi)$ (as in AFQMC) where both the hopping term and the auxiliary field appear as arguments of exponentials, while BRS uses the “half-compact” operator $M[\phi] \sim \mathbb{1} - h - \exp(i\phi)$ (as in lattice QCD, where the phase is replaced by the parallel transporter, or gauge link). Reference [45] found that “noncompact” formulations with $M[\phi] \sim \mathbb{1} - h - i\phi$ are numerically unstable due to round-off error. These studies in terms of BRS found no indication of an ergodicity problem.

Still, the ergodicity of BRS remains controversial. It has been noted [40,48,49] that the ergodicity problem related to the use of HMC with BSS cannot be eliminated simply by switching to an imaginary auxiliary field, as in BRS. Irrespective of whether the auxiliary field is purely real or purely imaginary, the BSS fermion determinant $\det M[\phi]$ factorizes, becoming proportional to real-valued function which is not positive definite. Hence, the configuration space of BSS should remain fragmented into regions separated by boundaries of *exceptional configurations*, which have vanishing determinant, that HMC cannot cross. However, Ref. [40] found that using a complex auxiliary field solved the ergodicity problem of BSS. The argument for factorization of Ref. [48] only applies to the compact version (BSS) of $M[\phi]$, which may explain why no ergodicity problem was found in previous BRS simulations. Here, we show in detail that BRS avoids the ergodicity problem associated with factorization of $\det M[\phi]$, and we also reproduce the ergodicity problems reported in earlier work using BSS + HMC (with both real-valued and imaginary-valued auxiliary fields). We note that the compact and half-compact versions of $M[\phi]$ are equivalent, up to terms of higher order in the Euclidean time step δ . We also discuss the relative merits of each formulation including the extrapolation of observables to the temporal continuum limit $\delta \rightarrow 0$. It can be argued that BRS represents a case where symmetries of the continuum theory are sacrificed on the lattice, in order to improve computational scaling and gain the applicability of the computational toolbox of lattice QCD. Unlike the complexified BSS formulation, we find no indication of adverse computational scaling with BRS. A thorough analysis of the computational scaling will be given in an upcoming publication.

This paper is organized as follows. We describe HMC and the Hubbard model in Sec. II, including different choices of basis and discretization, the associated symmetries, and the properties of the fermion matrix M and its eigenvalues. We study the ergodicity problem in Sec. III, and show its connection to the eigenvalues of M . We also give explicit examples for small numbers of lattice sites, which demonstrates how and when ergodicity issues appear. We explore possible ways to circumvent such issues in Sec. IV. By taking into account the symmetries of the system, we propose novel ways to effect large jumps between configurations, thereby crossing regions of low or zero probability. We recapitulate and conclude in Sec. V.

¹We stress that the BSS algorithm with discrete auxiliary fields is ergodic, and we do not consider such algorithms further in this paper.

²We restrict our presentation for simplicity to the Hubbard model, but our conclusions hold for the Hubbard/Coulomb model because, as we will see, the ergodicity problems we address here depend only on the fermion operator, and not on the form of the Gaussian part of the action of the auxiliary field, where fermion-fermion interactions ultimately appear.

II. FORMALISM

We start by giving a cursory description of the HMC algorithm and describing the different discretizations of the Hubbard model in the literature. Throughout, we assume the system is at half-filling (with zero chemical potential) and when we discretize the number of time slices N_t is even.

A. Hybrid Monte Carlo

The hybrid Monte Carlo (HMC) algorithm is a Markov chain Monte Carlo (MCMC) method which can be used to estimate multidimensional integrals

$$\left(\prod_{d=1}^{N_d} \int d\phi_d \right) W[\phi] O[\phi] \simeq \frac{1}{N_c} \sum_{i=1}^{N_c} O[\phi^{(i)}], \quad \phi^{(i)} \sim W, \quad (1)$$

using importance sampling according to $W[\phi]$. Each HMC step generates a new *configuration*, or integration point $\phi^{(i)}$ in the N_d -dimensional space, based on the previous configuration $\phi^{(i-1)}$ in the Markov chain. The larger the integration probability density $W[\phi^{(i)}]$ for a configuration $\phi^{(i)}$, the higher the probability that $\phi^{(i)}$ will be generated during the MC evolution. Once an *ensemble* consisting of N_c configurations $\{\phi^{(1)}, \dots, \phi^{(N_c)}\}$ has been generated, operator expectation values can be estimated stochastically, by performing the sum over i in (1).

HMC is a global algorithm: all N_d field components of a configuration ϕ are updated simultaneously. Each field component ϕ_d is assigned a canonically conjugate momentum component π_d , and the resulting (ϕ, π) system is evolved in a fictitious time by numerical integration of the Hamiltonian equations of motion. This is done using the hybrid Molecular dynamics (HMD) algorithm, which combines the stochastic Langevin and deterministic molecular dynamics (MD) methods. Specifically, each Langevin update (where the conjugate momenta are refreshed from a random Gaussian distribution) is interspersed with a number of MD integration steps, where the field ϕ follows a *trajectory* through the field space. The key advantage of HMC is the treatment of the HMD update as the proposal machine for the Metropolis algorithm. In principle, energy is conserved during an MD trajectory, but as numerical integration schemes have finite truncation errors energy conservation is violated. This violation is incorporated into the acceptance criterion of the Metropolis test—if the energy were exactly conserved, every proposed configuration would be accepted. Unlike HMD and similar algorithms, HMC does not require extrapolation of the step size of the MD integration rule to the continuum. The computational scaling of HMC as a function of system size V is expected to be $\sim V^{5/4}$ [20,42], superior to the cubic (or nearly cubic) scaling of local updates in theories of dynamical fermions. For this reason, HMC is the method of choice for computing ensembles of configurations in high-dimensional theories, such as lattice QCD.

Viewed as a Markov process, HMC converges to the desired equilibrium probability distribution $W[\phi]$ if

(1) the detailed balance condition $W[\phi]\Omega(\phi \rightarrow \phi') = W[\phi']\Omega(\phi' \rightarrow \phi)$ is satisfied, where $W[\phi]$ is the normalized Boltzmann distribution $\exp(-S[\phi])/\mathcal{Z}$, and $S[\phi]$ the

Euclidean action of the theory. Also, $\Omega(\phi \rightarrow \phi')$ is the transition probability from configuration ϕ to ϕ' .

(2) The Markov chain is ergodic meaning that the equilibrium distribution $W[\phi]$ is unique and independent of the starting configuration of the chain. In other words, given a configuration ϕ for which $W[\phi] \neq 0$, every other configuration ϕ' for which $W[\phi'] \neq 0$ should be reachable from ϕ in a finite number of steps (or amount of MC time).

For detailed balance to be satisfied, the MD integration should be performed with an integration rule which is reversible and symplectic (such as the leapfrog and Omelyan integrators). Such integrators also ensure that the acceptance rate of HMC only depends weakly on V , as there is *a priori* no guarantee that HMC can perform large global updates with significant decorrelation between successive configurations.

The second criterion is much harder to enforce, especially for multidimensional probability densities. While indicators for ergodicity issues can be monitored during the generation of configurations (for example one can monitor the force, watch for large changes in the acceptance rate, or the freezing of observables), a formal proof that HMC is ergodic for a particular system is usually not available. In some cases, a physical understanding of ergodicity problems is possible, such as the difficulty of tunneling between different topological sectors in lattice QCD. Note that the ergodicity problems referred to here should not be confused with the lack of ergodicity in algorithms that effect updates in terms of pure MD trajectories, with no periodic refreshment of the conjugate momenta or Metropolis accept/reject step.

The violation of energy conservation during the MD trajectory should remain small if an HMC update should be accepted with high likelihood. The classical MD evolution is driven by the functional derivative $F[\phi] = -\delta S[\phi]/\delta\phi$, the *HMC force term*. HMC is susceptible to barriers or discontinuities in the landscape of $W[\phi]$. Such barriers can occur when $W[\phi_0] = 0$ and *exceptional configurations* ϕ_0 can separate the integration domain into disconnected regions. As exceptional configurations correspond to singularities in the force term, attempts to cross the barrier generate a large energy violation, and the HMC update is rejected with a high probability. The inability of the standard HMC algorithm to cross barriers between disconnected regions of $W[\phi]$ leads to an ergodicity problem. In other words, the HMC Markov chain becomes locked in a region with boundaries ϕ_0 , where $W[\phi_0] = 0$. It should be noted that HMC can cross such boundaries infrequently if the numerical integration of the Hamiltonian equations of motion is sufficiently coarse. In general too coarse MD updates cannot maintain a high acceptance rate as V is increased.

As we have already noted, an ergodicity problem appears whenever the fermion determinant $\det M[\phi]$ becomes proportional to a real-valued function $f[\phi]$ which is not positive definite [48]. While such a problem can appear also for complex-valued ϕ and $\det M[\phi]$, it is usually more severe in theories where ϕ and $\det M[\phi]$ are real-valued, such as the overlap formulation of chiral fermions in lattice QCD. Another example is the AFQMC treatment of Ref. [28], which combined the HMC algorithm with the BSS formulation of the Hubbard model. There, the fermion matrix $M[\phi]$ is not

guaranteed to satisfy $\det M[\phi] > 0$, which becomes apparent at low T and at strong on-site coupling U , where the energy landscape fragments into multiple regions of the positive and negative $\det M[\phi]$. For overlap fermions, specialized HMC algorithms have been developed which can tunnel through or reflect from infinite-force barriers in a reversible manner, maintaining ergodicity and a high acceptance rate [30–32]. Here, we take a different approach and instead seek an optimal representation of $M[\phi]$, which minimizes or eliminates ergodicity problems altogether.

B. Choice of basis

The nearest-neighbor tight-binding Hamiltonian H_0 ,

$$H_0 = -\kappa \sum_{\langle x,y \rangle} (a_{x,\uparrow}^\dagger a_{y,\uparrow} + a_{x,\downarrow}^\dagger a_{y,\downarrow}) \quad (2)$$

contains a kinetic term only, which describes free electrons of spin $\uparrow$ and spin $\downarrow$ hopping between different lattice sites with hopping parameter κ . The bracket $\langle x, y \rangle$ denotes pairs of nearest neighbors. The Hubbard model adds on-site interactions

$$H = H_0 - \frac{U}{2} \sum_x (n_{x,\uparrow} - n_{x,\downarrow})^2 \quad (3)$$

where number operator $n_{x,s} \equiv a_{x,s}^\dagger a_{x,s}$ counts electrons of spin s at position x .

We can change basis via a particle-hole transformation on the spin- $\downarrow$ electrons

$$b_{x,\downarrow}^\dagger \equiv a_{x,\downarrow}, \quad b_{x,\downarrow} \equiv a_{x,\downarrow}^\dagger. \quad (4)$$

Up to an overall irrelevant constant, the Hamiltonian under this transformation is

$$H = -\kappa \sum_{\langle x,y \rangle} (a_{x,\uparrow}^\dagger a_{y,\uparrow} - b_{x,\downarrow}^\dagger b_{y,\downarrow}) + \frac{U}{2} \sum_x (n_x - \tilde{n}_x)^2, \quad (5)$$

where the number operator $\tilde{n}_x \equiv b_x^\dagger b_x$ counts spin- $\downarrow$ holes at position x . The degrees of freedom here are electrons of spin $\uparrow$ and holes with spin $\downarrow$. This lets us drop the spin indices. Both bases describe the same system and give the same relative spectrum.

In the case of bipartite lattices (like the honeycomb lattice of graphene and carbon nanotubes), it is possible to modify

transformation (4) to switch signs on one sublattice,

$$b_{x,\downarrow}^\dagger \equiv \mathcal{P}_x a_{x,\downarrow}, \quad b_{x,\downarrow} \equiv \mathcal{P}_x a_{x,\downarrow}^\dagger, \quad (6)$$

where $\mathcal{P}_x$ is $+1$ if x is on one sublattice and -1 if it is on the other. This keeps H_0 invariant under the particle-hole transformation, but still flips the sign of the on-site interaction compared to (3),

$$H = -\kappa \sum_{\langle x,y \rangle} (a_{x,\uparrow}^\dagger a_{y,\uparrow} + b_{x,\downarrow}^\dagger b_{y,\downarrow}) + \frac{U}{2} \sum_x (n_x - \tilde{n}_x)^2; \quad (7)$$

we recognize the first term as simply the tight binding Hamiltonian H_0 given in (2) with b in lieu of $a_\downarrow$ and spin labels dropped. We henceforth specialize to bipartite lattices. We say that this Hamiltonian is written in the *particle/hole basis*, while (3) is in the *spin basis*. The only difference between (3) and (7) is the sign in front of the on-site interaction term.

It is possible to write down a Hamiltonian that includes both types of interactions, parameterized via $\alpha \in [0, 1]$ as

$$H = H_0 + \alpha \frac{\mathcal{U}}{2} \sum_x (n_{x,\uparrow} - n_{x,\downarrow})^2 - (1 - \alpha) \frac{U}{2} \sum_x (n_{x,\uparrow} - n_{x,\downarrow})^2. \quad (8)$$

Ignoring the superficial difference in labeling of spin- $\downarrow$ electrons or spin- $\downarrow$ holes, when $\alpha = 0$ one recovers the Hamiltonian of the spin basis (3) while $\alpha = 1$ yields the Hamiltonian in the particle/hole basis (7) if $\mathcal{U} = U$. For arbitrary $\alpha \in (0, 1)$, Hubbard-Stratonovich transformations will introduce auxiliary fields with both real and imaginary components, as thoroughly investigated in Ref. [40], which found ergodicity problems for the extreme values 0 and 1.

As our investigations revolve around issues related to ergodicity, in what follows we concentrate only on the extreme values $\alpha = 0$, the spin basis, and $\alpha = 1$, the particle/hole basis. Sometimes we use α to label the different bases for brevity.

C. Discretization

Discretizing the Hubbard model path integral, and the introduction of auxiliary fields ϕ by Hubbard-Stratonovich transformation, has been discussed before (see Refs. [43–46], for example). After discretization, the partition function in the spin basis, up to an overall normalization, can be written

$$\mathcal{Z}_{\uparrow\downarrow} = \int \left[\prod_{x,t} d\phi_{xt} \right] W[\phi] = \int \left[\prod_{x,t} d\phi_{xt} \right] \det M[\phi] \det M[-\phi] \exp \left(-\frac{1}{2\tilde{U}} \sum_{x,t} \phi_{xt}^2 \right) \quad (9)$$

$$= \int \left[\prod_{x,t} d\phi_{xt} \right] \exp \left(-\frac{1}{2\tilde{U}} \sum_{x,t} \phi_{xt}^2 + \ln \det M[\phi] + \ln \det M[-\phi] \right), \quad (10)$$

where $W[\phi]$ is the probability weight of a configuration ϕ and $M[\phi]$ is the fermion matrix. M is also a function of the hopping matrix $h_{x',x} = \tilde{\kappa} \delta_{\langle x',x \rangle}$ which corresponds to the nearest-neighbor connections in the tight-binding Hamiltonian H_0 with hopping strength $\tilde{\kappa} = \kappa \delta$, with $\delta = \beta/N_t$ the discretization of the inverse temperature into N_t evenly spaced slices. The interaction strength is $\tilde{U} = U \delta$.

The partition function in the particle/hole basis uses imaginary fields in the fermion matrix which is a consequence of the different sign in front of the on-site interaction term in (7) compared to (3). The partition function is otherwise identical to the

one in the spin basis

$$\mathcal{Z}_{\text{ph}} = \int \left[\prod_{x,t} d\phi_{xt} \right] \exp \left(-\frac{1}{2\bar{U}} \sum_{x,t} \phi_{xt}^2 + \ln \det M[i\phi] + \ln \det M[-i\phi] \right). \quad (11)$$

A negative sign in front of the interaction requires completely real auxiliary fields, whereas a positive sign requires completely imaginary fields.

Differences in discretizations manifest themselves in the structure of the fermion matrix M . In Refs. [38,40,48], for example, matrix elements of the fermion operator have the form

$$M^e[\phi]_{x't',xt} = \delta_{x',x} \delta_{t',t} - [e^h]_{x',x} e^{\phi_{xt}} \mathcal{B}_{t'} \delta_{t',t+1} \quad (\text{exponential discretization}), \quad (12)$$

where space and time directions are combined and $x't'$ denote the row and xt the column index. $\mathcal{B}_{t'} = +1$ for $0 < t' < N_t$ and $\mathcal{B}_0 = -1$ explicitly encodes the antiperiodic boundary condition in time. In Ref. [45], on the other hand, the matrix elements are

$$M^\ell[\phi]_{x't',xt} = \delta_{x',x} \delta_{t',t} - (h_{x',x} + e^{\phi_{xt}} \delta_{x',x}) \mathcal{B}_{t'} \delta_{t',t+1} \quad (\text{linear discretization}), \quad (13)$$

Finally, in Refs. [43,46,47], the hopping term is moved to the time diagonal,³

$$M^d[\phi]_{x't',xt} = (\delta_{x',x} - h_{x',x}) \delta_{t',t} - e^{\phi_{xt}} \delta_{x',x} \mathcal{B}_{t'} \delta_{t',t+1} \quad (\text{diagonal discretization}). \quad (14)$$

This last discretization is more akin to what is done in lattice gauge theories, where the gauge links (parallel transporters) reside between discretization slices. The exponential, linear, and diagonal discretizations in Eqs. (12), (13), and (14) formally agree up to $\mathcal{O}(\delta^2)$, and thus have the same continuum $\delta \rightarrow 0$ ($N_t \rightarrow \infty$) limit. Observables calculated with these different discretizations should only be compared after a continuum limit extrapolation. In this work, we will focus on the exponential and diagonal discretizations, only occasionally commenting on the linear discretization.

It will prove useful to consider the matrix $S = M - \mathbb{1}$ for the various discretizations with matrix elements

$$S^e[\phi]_{x't',xt} = -[e^h]_{x',x} e^{\phi_{xt}} \mathcal{B}_{t'} \delta_{t',t+1} \quad (\text{exponential}), \quad (15)$$

$$S^d[\phi]_{x't',xt} = -h_{x',x} \delta_{t',t} - e^{\phi_{xt}} \delta_{x',x} \mathcal{B}_{t'} \delta_{t',t+1} \quad (\text{diagonal}), \quad (16)$$

which in the exponential case is entirely off-diagonal. Each eigenvalue of M differs from an eigenvalue of S by 1.

D. Symmetries, fermion determinants, and fermion matrix eigenvalues

Understanding the symmetries and limits of the physical problem and the discretizations will prove valuable for later discussion and inspiration for how to alleviate some ergodicity problems. The impatient reader may prefer to skip this detailed discussion, though we do rely on observations here throughout the rest of the paper.

It is useful to consider the probability weight of a field configuration ϕ

$$W[\phi] = \det M[\phi] \det M[-\phi] \exp \left(-\frac{1}{2\bar{U}} \sum_{x,t} \phi_{xt}^2 \right) \quad (17)$$

that appears in the spin-basis partition function (10) and its analog in (11) where the arguments of the fermion matrices get an i .

1. Charge conjugation

The first, most obvious symmetry of the probability weight is the change of the sign of ϕ . When one sends $\phi \rightarrow -\phi$ the quadratic piece is invariant and the determinants change roles, so that

$$W[\phi] = W[-\phi]. \quad (18)$$

The two determinants arise from the different spins or species, depending on the basis. Thus, sending $\phi \rightarrow -\phi$ exchanges the spins, or exchanges particles and holes. This symmetry is broken when away from half filling, or, put another way, with nonzero chemical potential. Away from half filling one must also negate the chemical potential to achieve equality of W . This is the analog to charge conjugation symmetry C . Interestingly, for nonbipartite lattices the sign of the tight binding Hamiltonian differs between the two determinants, and $\phi \rightarrow -\phi$ fails to be a symmetry, even at half filling.

2. Characteristic polynomials

To go beyond this observation it will prove useful to have a firm understanding of the fermion matrix, its eigenvalues, and its determinant in the four different cases described in the previous section. It is simpler to consider the eigenvalues of $S = M - \mathbb{1}$, given in (15) and (16). We restrict our attention to even N_t for simplicity.

We will demonstrate equalities of the characteristic polynomial of S ,

$$P[\phi](s) = \det(S[\phi] - s\mathbb{1}). \quad (19)$$

When s is a root of $P[\phi]$ it is an eigenvalue of $S[\phi]$, and $\lambda = s + 1$ is an eigenvalue of $M[\phi]$. Note that since $M = S + \mathbb{1}$, we know $\det M[\phi] = P[\phi](-1)$. In the fully general case P also depends on the chemical potential and on the sign of the adjacency matrix. We suppress these dependencies for clarity

³In addition to moving the hopping term to the time diagonal, in Refs. [46,47], a mixed forward and backward differencing scheme was applied to the underlying sublattices.

and focus on the bipartite, half-filling case and only comment when ignoring these assumptions invalidates a conclusion.

First let us consider the exponential case with even N_t . We will use the identity (A11) shown in Appendix A, letting

$$D_{x',x} = -s\delta_{x',x}, \quad [T_t]_{x',x} = -[e^h]_{x',y} F_t[\phi]_{y,x} \quad (20)$$

where

$$F_t[\phi]_{x',x} = \begin{cases} e^{-\phi_{x(t-1)}} \delta_{x',x} & \alpha = 0 \\ e^{-i\phi_{x(t-1)}} \delta_{x',x} & \alpha = 1 \end{cases} \quad (21)$$

is a diagonal matrix of auxiliary fields on a given time slice and the t index is understood modulo N_t .⁴ Note that F has the property

$$F_t[\phi]^{-1} = F_t[-\phi] \quad (22)$$

generally and

$$F_t[\phi]^* = \begin{cases} F_t[\phi] & \alpha = 0 \\ F_t[-\phi] & \alpha = 1 \end{cases}. \quad (23)$$

Then the characteristic polynomial is given by

$$\begin{aligned} P^e[\phi](s) &= \det(-s\mathbb{1}_{N_x})^{N_t} \det\left(\mathbb{1}_{N_x} + \frac{\mathbb{1}_{N_x}}{s} e^h F_{N_t-1} \frac{\mathbb{1}_{N_x}}{s} e^h F_{N_t-2} \cdots \frac{\mathbb{1}_{N_x}}{s} e^h F_0\right) \\ &= \det(s^{N_t} \mathbb{1}_{N_x}) \det(\mathbb{1}_{N_x} + s^{-N_t} e^h F_{N_t-1} e^h F_{N_t-2} \cdots e^h F_0) \\ &= \det(s^{N_t} \mathbb{1}_{N_x} + e^h F_{N_t-1} e^h F_{N_t-2} \cdots e^h F_0), \end{aligned} \quad (24)$$

which is a polynomial in the variable s^{N_t} . So, if s is a root of $P^e[\phi]$, any other s with the same N_t th power is also a root. This establishes that, in the fully interacting exponential case, if s is an eigenvalue, so is $s \exp(2\pi i/N_t)$, which can be bootstrapped all the way around the circle. That is, the eigenvalues come equally spaced around rings in the exponential case, independent of α .

In the diagonal case we use (A12), an equivalent determinant identity also shown in Appendix A but now with

$$D_{x',x} = -s\delta_{x',x} - h_{x',x} \quad [T_t]_{x,x} = -F_t(\phi)_{x',x}. \quad (25)$$

One finds, dropping N_t powers of minus signs,

$$\begin{aligned} P^d[\phi](s) &= \det(F_{N_t-1} \cdots F_0) \det(\mathbb{1} + F_0^{-1}(s\mathbb{1}_{N_x} + h) \cdots F_{N_t-2}^{-1}(s\mathbb{1}_{N_x} + h) F_{N_t-1}^{-1}(s\mathbb{1}_{N_x} + h)) \\ &= \left(\prod_t \det(F_t) \right) \det(\mathbb{1} + F_0^{-1}(s\mathbb{1}_{N_x} + h) \cdots F_{N_t-2}^{-1}(s\mathbb{1}_{N_x} + h) F_{N_t-1}^{-1}(s\mathbb{1}_{N_x} + h)). \end{aligned} \quad (26)$$

This is not a polynomial in s^{N_t} and we do not expect to find perfect rings—were $h = 0$ we would.

In the linearized case (13), the diagonal of S is again zero, s can be gathered in the characteristic polynomial as for $P^e[\phi]$, and we expect perfect rings. We have verified this expectation numerically.

3. Field shifts and periodicity

From (21) it is clear that in the $\alpha = 1$ case, F remains invariant if any field component changes by 2π , and the determinant is therefore invariant under such a shift. In fact, we can make a more generic transformation,

$$\phi_{xt} \rightarrow \phi_{xt} + \theta_t \quad (27)$$

shifting all the fields on time slice t by a constant θ_t . Under that transformation, F changes by an overall phase

$$F_t[\phi + \theta] = F_t[\phi] e^{-i\theta_{t-1}}. \quad (28)$$

In the exponential case, the characteristic polynomial of the transformed configuration,

$$\begin{aligned} P^e[\phi_{xt} + \theta_t](s) &= \det(s^{N_t} \mathbb{1} + e^h F_{N_t-1} e^{-i\theta_{N_t-2}} \cdots e^h F_1 e^{-i\theta_0} e^h F_0 e^{-i\theta_{N_t-1}}) \\ &= \det(s^{N_t} \mathbb{1} + e^{-i \sum_t \theta_t} e^h F_{N_t-1} \cdots e^h F_0) \end{aligned} \quad (29)$$

which is equal to $P^e[\phi](s)$, so long as

$$\sum_t \theta_t \equiv 0 \pmod{2\pi}. \quad (30)$$

The same condition holds for the diagonal discretization, analogously. Of course, these shift transformations are only symmetries of the fermion determinant; the Gaussian part of the action changes, in general. These shifts leave each Polyakov loop, the time-ordered product of links around the temporal direction,

$$\mathcal{P}_x = \prod_t e^{i\phi_{xt}} = e^{i\Phi_x} \quad \Phi_x = \sum_t \phi_{xt} \quad (31)$$

invariant.

In the $\alpha = 0$ case, we can make analogous field shifts to each time slice. However, rather than finding a periodic requirement on the sum of the shifts, the requirement to get the same eigenvalues is that the sum of the shifts vanishes, $\sum_t \theta_t = 0$, as the links in the Polyakov loop lose their factor of i .

We also see that if we send an even number of F s to minus themselves, the determinant is invariant. When $\alpha = 1$ we can flip a single F by shifting all of the fields on a single time slice by $\pm\pi$, making an independent choice for each,

$$\phi_x \rightarrow \phi_x + \pi j_x, \quad (32)$$

where each j_x is an odd integer. In fact, we can think of this transformation as a composition of individual 2π jumps and

⁴Note that here we have explicitly included the α -dependent factor of i in F so that we can always just think of ϕ as real.

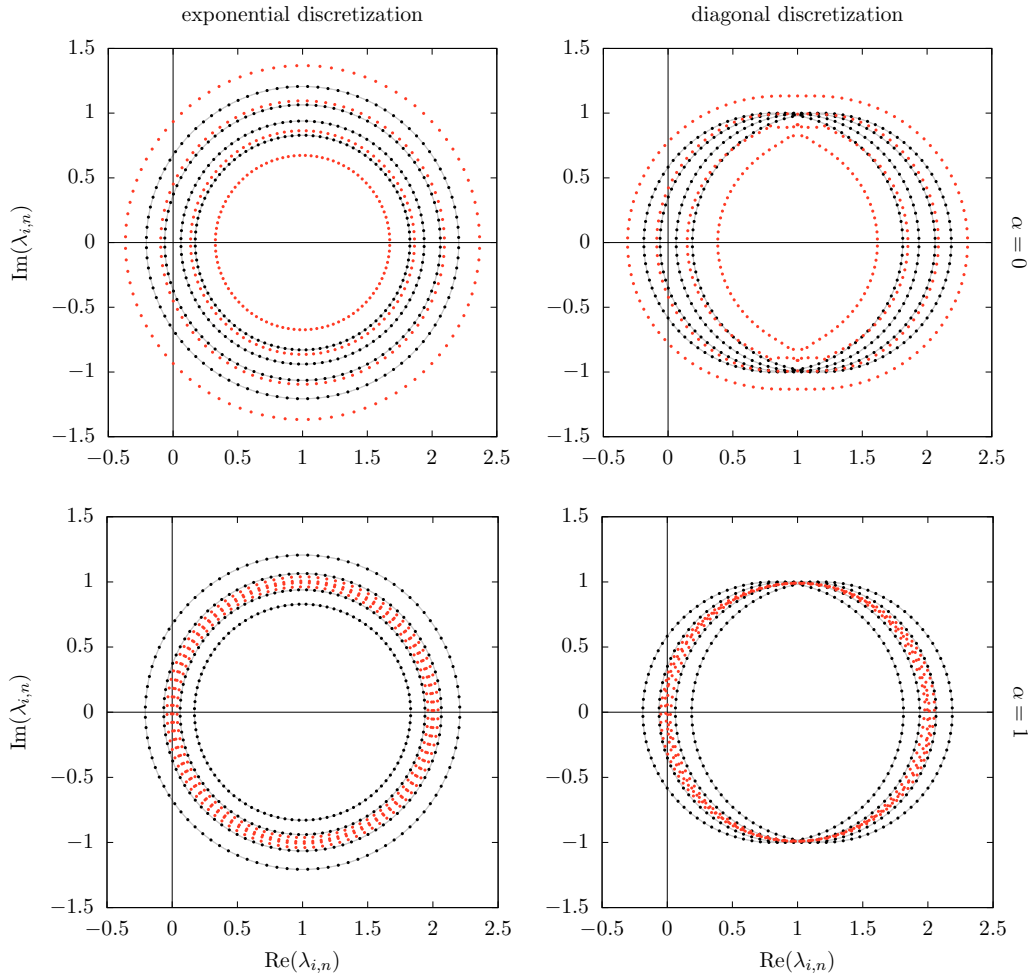


FIG. 1. In the four panels, we show as lighter red points the spectrum of a fermion matrix with a particular α and discretization, as indicated, for a four-site honeycomb lattice with $N_t = 96$. The darker, black points are the spectrum of the noninteracting case ($\phi = 0$), which depends only on which discretization is used. We show light gray rings to guide the eye for the eigenvalue rings as described by (33). There are four rings because there are four spatial sites; each ring has $N_t = 96$ points. In all four cases, we use the same configuration ϕ which was randomly sampled from a Gaussian with width $\sqrt{U\beta/N_t}$, where $U\beta = 60$.

the coordinated field shift (27), setting $\theta_t = \pi$. One choice that leaves the sum of all the field variables invariant on a bipartite lattice is $j_x = \pm \mathcal{P}_x$, where $\mathcal{P}_x$ is $+1$ on one sublattice and -1 on the other, as in the particle-hole transformation (6). As long as an even number of time slices get so transformed, the eigenvalues are invariant, though the Gaussian part of the action may change.

4. The noninteracting case

In the noninteracting $U = 0$ case, we can solve the fermion matrix exactly. The Gaussian controlling the auxiliary field ϕ in Eqs. (10) and (11) becomes infinitely narrow and we need only consider $\phi = 0$, so all F s are the identity matrix. In this case, we can find the exact spectrum of the fermion matrix. Independent of whether α is 0 or 1, the eigenvalues are given by

$$s_{i,n} = \begin{cases} e^{\delta\epsilon_i + i\omega_n} & \text{exponential} \\ -\delta\epsilon_i + e^{i\omega_n} & \text{diagonal} \end{cases}, \quad \lambda_{i,n} = s_{i,n} + 1, \quad (33)$$

where the Matsubara frequencies $\omega_n = \frac{2\pi}{N_t}(n + \frac{1}{2})$ for n in the integers from 0 to $N_t - 1$ and ϵ_i are the noninteracting eigenvalues of H_0/κ .

In the exponential case, the eigenvalues of M come in rings concentric around 1, as discussed following (24), with radii $r_i = \exp \delta\epsilon_i$. When the hopping is bipartite, the eigenvalues ϵ come in additive-inverse pairs and the corresponding radii are multiplicative inverses. That is, if one ring has a radius r , another has radius $1/r$ —if s is an eigenvalue of S , so is $1/s^*$. We will show, below, that this remains true in the exponential $\alpha = 1$ case and that generally, in the interacting exponential case, if s is an eigenvalue for one species, $1/s$ is an eigenvalue for the other, a statement of exact chiral symmetry.

In the diagonal case, the eigenvalues of M come in rings, all of radius 1, centered on $1 - \delta\epsilon_i$. In the continuum limit $\delta \rightarrow 0$ both discretizations give the same N_x -degenerate ring of eigenvalues, as expected. In both cases, the eigenvalues are evenly distributed around their respective rings, at angles given by the Matsubara frequencies; with interactions this perfect spacing is true for the exponential case, as already

shown. Figure 1 shows in black the noninteracting eigenvalues for the four-site honeycomb lattice, which has ϵ in $\{\pm 1, \pm 3\}$, with $N_t = 96$.

In both discretizations, the eigenvalues in (33) come in complex conjugate pairs. However, this simple picture of perfect rings is broken when $\phi \neq 0$, corresponding to $U \neq 0$. Moreover, once interactions are turned on the spectra differ depending on α . Figure 1 shows in red an example fermion matrix spectrum for the same example lattice, using the same field configuration for each discretization.

5. Reality of the probability weight

Starting from (24) and using the fact that $\det(A^*) = \det(A)^*$, one finds

$$\begin{aligned} P[+\phi](s) &= \det((s^*)^{N_t} \mathbb{1} + e^h F_{N_t-1}^* \cdots e^h F_0^*)^* \\ &= \begin{cases} P[+\phi](s^*)^* & \alpha = 0 \\ P[-\phi](s^*)^* & \alpha = 1 \end{cases}, \end{aligned} \quad (34)$$

where we used the properties of F under complex conjugation in (23) and the fact that h is real. In the $\alpha = 0$ case, the eigenvalues of a single fermion matrix come in complex-conjugate pairs or as real singletons, each determinant is independently real, and so the probability weight is real. In the $\alpha = 1$ case, the probability weight is guaranteed to be real and positive because, as (34) shows, the particle eigenvalues are the complex conjugates of the hole eigenvalues. An analogous argument starting from (26) shows that (34) holds for the

diagonal discretization as well. We will return to the question of positivity in the $\alpha = 0$ in Sec. II D 11.

In the case of nonzero chemical potential, (34) it is not enough to show the reality of the probability as the exchange of determinants also requires flipping the chemical potential and the right-hand side is not the characteristic polynomial for the other species. Similarly, in the case of nonbipartite lattices, the sign of h flips for the other species but is unchanged by this manipulation and we again find an identity between characteristic polynomials, but not the ones we need to demonstrate reality. Thus we expect a sign problem at nonzero chemical potential or on nonbipartite lattices; the difficulty of those sign problems is a question for future work.

6. Particle-hole symmetry

We can take advantage of the bipartite structure of the adjacency matrix. Let Σ be a diagonal matrix that is $+1$ on one sublattice and -1 on the other. Then

$$\Sigma^2 = \mathbb{1}_{N_x}, \quad \Sigma h \Sigma = -h, \quad \Sigma F \Sigma = F. \quad (35)$$

By repeatedly inserting Σ^2 , we see that

$$P[+h, \phi](s) = P[-h, \phi](s) \quad (36)$$

independent of discretization or choice of α . In the general case without applying (6) one species' characteristic polynomial naturally occurs with a $-h$; in the bipartite case, the sign of h can be flipped without repercussion for the eigenvalues of the fermion matrix.

When N_t is even we see that the eigenvalues come in additive inverse pairs,

$$\begin{aligned} P^d[-h, \phi](s) &= \left(\prod_t \det(F_t) \right) \det(\mathbb{1}_{N_x} + F_0^{-1}(s\mathbb{1}_{N_x} - h) \cdots F_{N_t-1}^{-1}(s\mathbb{1}_{N_x} - h)) \\ &= \left(\prod_t \det(F_t) \right) \det(\mathbb{1}_{N_x} + (-1)^{N_t} F_0^{-1}(-s\mathbb{1}_{N_x} + h) \cdots F_{N_t-1}^{-1}(-s\mathbb{1}_{N_x} + h)) \\ &= P^d[+h, \phi](-s) \end{aligned} \quad (37)$$

so that if s is a root of $P^d[\phi]$, $-s$ is too. In the exponential case, this is already guaranteed by the fact that the eigenvalues come evenly spaced around rings.

7. Temporal shifts

By using Sylvester's determinant identity

$$\det(\mathbb{1} + AB) = \det(\mathbb{1} + BA), \quad (38)$$

we can cyclically permute matrices around the determinant.

In particular, from both (24) and (26), we see immediately, by shifting both the hopping and the first F , that

$$P[\phi_t](s) = P[\phi_{t+1}](s), \quad (39)$$

where ϕ_{t+1} is the field configuration ϕ_t but shifted by one time slice (modulo N_t). Since ϕ is bosonic, we need not worry about antiperiodic boundary conditions; they are built into the fermion matrix directly through $\mathcal{B}_t$. This shift can

be repeatedly applied and ultimately guarantees the time-translation invariance of the probability weight, independent of discretization scheme and α .

8. Time reversal

We also immediately see that ϕ has a time-reversal symmetry T . Since the determinant is invariant under transposition, starting from (24) one sees

$$\begin{aligned} P[\phi_{+t}](s) &= \det(s^{N_t} \mathbb{1}^\top + F_0^\top e^{h^\top} \cdots F_{N_t-1}^\top e^{h^\top}) \\ &= \det(s^{N_t} \mathbb{1} + F_0 e^h \cdots F_{N_t-1} e^h) \\ &= \det(s^{N_t} \mathbb{1} + e^h F_0 \cdots e^h F_{N_t-1}) \\ &= P[\phi_{-t}](s), \end{aligned} \quad (40)$$

where we used the fact that h is symmetric, F is diagonal, and Sylvester's identity (38). This shows that the field configuration ϕ_{-t} , which is time-reversed with respect to ϕ_{+t} , so

that ϕ_{-t} on the first time slice is ϕ_{+t} on the last and so on, yields the same eigenvalues. An analogous proof holds for the diagonal case, starting from (26). Thus time reversal T holds independent of discretization scheme and α .

9. Spatial symmetries

The spatial lattice may have some rotational, translational, or parity or reflection symmetries. An operation $\mathcal{P}$ of these kinds permutes spatial sites,

$$x \rightarrow \mathcal{P}x \quad (41)$$

and is a symmetry if its application commutes with the Hamiltonian (8).⁵ Since every on-site interaction is of the same strength U it is automatically invariant under any relabeling of sites; the symmetries of the lattice are those permutations that commute with the tight-binding Hamiltonian or h , $[\mathcal{P}, h] = 0$. Two configurations related by one of these symmetries have the same weight in the path integral. They manifestly have the

same Gaussian factor. Note that the action of a permutation $\mathcal{P}$ on $F[\phi]$ is

$$\mathcal{P}F[\phi_x]\mathcal{P}^{-1} = F[\phi_{\mathcal{P}x}], \quad (42)$$

where $\phi_{\mathcal{P}x}$ is configuration where the fields have changed sites according to the permutation. Since $\mathcal{P}$ commutes with h (and also, obviously, e^h), we may insert $\mathbb{1}_{N_x} = \mathcal{P}^{-1}\mathcal{P}$ everywhere, and find

$$P[\phi_x](s) = P[\phi_{\mathcal{P}x}](s) \quad (43)$$

independent of discretization scheme and α .

10. Exact chiral symmetry

We can calculate the characteristic polynomial in the exponential case using the other identity, (A12), and massage it such that we can identify (24),

$$\begin{aligned} P^e[+h, +\phi_{+t}](s) &= \det(e^h F_{N_t-1} \cdots e^h F_0) \det(\mathbb{1}_{N_x} + F_0^{-1} e^{-h} s \mathbb{1}_{N_x} \cdots F_{N_t-1}^{-1} e^{-h} s \mathbb{1}_{N_x}) \\ &= \left(\prod_t \det(F_t) \right) \det(s^{N_t} \mathbb{1}_{N_x}) \det(s^{-N_t} \mathbb{1}_{N_x} + F_0^{-1} e^{-h} \cdots F_{N_t-1}^{-1} e^{-h}) \\ &= \left(\prod_t \det(F_t) \right) \det(s^{N_t} \mathbb{1}_{N_x}) P^e[-h, -\phi_{-t}](1/s) \\ &= \left(\prod_t \det(F_t) \right) \det(s^{N_t} \mathbb{1}_{N_x}) P^e[-h, -\phi_{+t}](1/s), \end{aligned} \quad (44)$$

where we used the fact that the determinant of e^h is unity to simplify the leading factor, and we applied the time reversal identity for the characteristic polynomial (40). This identity shows that the eigenvalues s for one species are the reciprocal of the eigenvalues for the other, even without the assumption of a bipartite lattice. Using the “other” determinant identity for the diagonal case yields no useful identity, because s does not appear alone, but rather in combination with the tight binding Hamiltonian, and cannot be gathered together and factored out. In the linearized case of (13), we also find no useful identity, as suggested in Ref. [49].

Since, in the exponential case, the off-diagonal blocks in S are the time slice-to-time slice transfer matrices, (44) shows that the transfer matrix for one species has the inverse eigenvalues of the transfer matrix for the other. This is the exact chiral symmetry discussed in Ref. [49], and only appears

in the exponential case. Note that this identity does not rely on particle-hole symmetry.

Using (36) in the bipartite case and (34) in the $\alpha = 1$ case, we arrive at

$$P^e[\phi](s) = \left(\prod_t \det(F_t) \right) \det(s^{N_t} \mathbb{1}_{N_x}) P^e[\phi](1/s^*)^* \quad (\alpha = 1). \quad (45)$$

As long as $s \neq 0$, if s is a root of $P^e[\phi]$, so is $1/s^*$. This demonstrates that in the exponential $\alpha = 1$ case each ring of eigenvalues has a partner ring with a reciprocal radius and the same angular alignment, as we observed in the noninteracting exponential case. We say that $P^e[\phi]$ is proportional to its own *conjugate reciprocal polynomial* $P^e[\phi]^\dagger$. The conjugate reciprocal polynomial $p^\dagger$ of a polynomial p is given by

$$p^\dagger(z) = z^n \overline{p(\bar{z}^{-1})} \quad (46)$$

here we use the overbar to indicate complex conjugation to avoid confusion with the asterisk that sometimes indicates the reciprocal polynomial.

Essential to demonstrating this conjugate reciprocity in the interacting case was (34) and therefore the properties of F (23). Although it is true for the noninteracting case and it is visually plausible for the interacting exponential $\alpha = 0$ example in Fig. 1, the radii are not, in fact, multiplicative inverses.

⁵Strictly speaking, an operation commuting with the Hamiltonian is not enough to guarantee an invariance of the discretized action for a particular field configuration. On the honeycomb lattice, for example, some parity symmetries of the lattice exchange sublattices. In the mixed differencing scheme of Refs. [46,47], those sublattices have different differencing operators and the weight is not guaranteed to be invariant under those parity operations. However, those operations followed by a time reversal and charge conjugation keep the overall action invariant.

Also essential was the particle-hole symmetry that allowed us to flip the sign of h . Without that symmetry, we find a relation between two characteristic polynomials, but not a relation that can be used to show this conjugate reciprocity.

Recall that setting $s = -1$ in a characteristic polynomial gives $\det(M[\phi])$. It is now helpful to return the α dependence to the argument of M rather than implicit in F as in (21). Starting from (44), using the particle-hole identity (36), and plugging in $s = -1$ leads to the identity

$$\begin{aligned} \det M^e[+\phi] &= e^{+\Phi} \det M^e[-\phi], \\ f[\phi] &= e^{-\Phi/2} \det M^e[+\phi] = e^{+\Phi/2} \det M^e[-\phi] = f[-\phi], \end{aligned} \quad (47)$$

where we define $\Phi = \sum_{x,t} \phi_{x,t}$, $f[\phi] = e^{-\Phi/2} \det M^e[\phi]$ and immediately see that f is even. When $\alpha = 0$, $f[\phi]$ must be real, because the determinant is real, as shown in (34). When $\alpha = 1$, we can use (34) and find

$$\begin{aligned} f[i\phi] &= e^{-i\Phi/2} \det M^e[+i\phi] = e^{+i\Phi/2} \det M^e[-i\phi] \\ &= e^{+i\Phi/2} \det M^e[+i\phi]^* = f[i\phi]^* \end{aligned} \quad (48)$$

so that f is also real when $\alpha = 1$. Put another way, starting from (45), one finds

$$\det(M^e[i\phi]) = e^{i\Phi} \det(M^e[i\phi])^*. \quad (49)$$

Writing, in radial coordinates, $\det(M[i\phi]) = f[i\phi]e^{i\theta(\phi)}$, one finds

$$\theta(\phi) = \frac{\Phi}{2} + Q\pi, \quad (50)$$

where Q is an integer. So

$$e^{-i\Phi/2} \det M[i\phi] \in \mathbb{R} \quad (51)$$

but of either sign, as discussed in (28) of Ref. [48].⁶ Since f is continuous in ϕ , changing Q requires passing through $f = 0$. This is the origin of formal ergodicity problems, as we discuss later.

When $\alpha = 0$ it is not guaranteed that f must take both signs. Indeed, as we later show for the one-site problem (65) and a simple two-site problem (78), it may be that the sign of $f[\phi]$ is fixed in this case. That is, at least for some examples $f[\phi]$ does not change sign and there are no formal ergodicity problems, but we stress that this is not necessarily generic. We will see, one way or the other, that there are also in-practice problems in the exponential $\alpha = 0$ case.

In the diagonal discretization, there is no analog of (44), no natural factorization of the determinant emerges, and there are no sectors that are separated by a vanishing determinant, even when $\alpha = 1$. Thus the diagonal discretization does not formally suffer from the ergodicity problems the exponential discretization suffers. In Sec. III C, we provide a simple two-site example where this claim may be directly verified. If one insists on writing $\det M^d[\phi] = e^{\Phi/2} f[\phi]$, in the $\alpha = 0$

case of course f will still be real, but only because the determinant itself is real. We find no restriction forcing $f[i\phi]$ to be real in the diagonal case. We later provide an example of f wandering off-axis in the complex plane in Fig. 2.

As shown, the partitioning of the field space into sectors is a result of the conjugate reciprocity of the bipartite, exponential, $\alpha = 1$ case. It may be that in other cases there are other as-yet formally undemonstrated partitionings. In Sec. III A, we show examples of the determinant flipping sign by crossing zero for both $\alpha = 0$ cases.

Let us continue to focus on the $\alpha = 1$ case and on the properties of $f[i\phi]$. Consider now what happens when we increase one of the auxiliary field variables by 2π ,

$$\phi_{x,t} \rightarrow \phi_{x,t} + 2\pi\delta_{x,x_0}\delta_{t,t_0}, \quad (52)$$

where x_0 and t_0 are the space and time coordinates of the field we are changing. Since F is 2π -periodic in each field variable individually, the determinants must be equal. Then, we find

$$\begin{aligned} f[i\phi_{x,t}]e^{i\Phi/2} &= \det M^e[i\phi] = \det M^e[i\phi + 2\pi i\delta_{x,x_0}\delta_{t,t_0}] \\ &= f[i\phi_{x,t} + 2\pi i\delta_{x,x_0}\delta_{t,t_0}]e^{i(\Phi/2+\pi)} \\ f[i\phi_{x,t}] &= -f[i\phi_{x,t} + 2\pi i\delta_{x,x_0}\delta_{t,t_0}] \end{aligned} \quad (53)$$

so that shifting any field variable by 2π flips the sign of f . Since this flip is independent of all other field variables, this shows the manifolds of zeros are codimension 1.

11. Summary

We collect in Table I constraints on the eigenvalues in the different discretizations and bases. Independent of discretization or basis we have, on a half-filled bipartite lattice, charge conjugation symmetry, particle-hole symmetry, temporal translation symmetry, time reversal symmetry, and whatever spatial symmetries the lattice exhibits. In the exponential case, we have exact chiral symmetry and when $\alpha = 1$ conjugate reciprocity.

Earlier we showed the weight W was real. For a straightforward Monte Carlo method, W should have an interpretation as a probability measure, and should therefore be positive. When $\alpha = 1$, each particle eigenvalue is the complex conjugate of a hole eigenvalue. This guarantees positivity of the weight W . When $\alpha = 0$ the complex conjugate guarantee is not enough. In the exponential case, we can use (47) to demonstrate positivity, though each determinant individually is not guaranteed to be positive—but the chiral symmetry is enough to guarantee that the two species determinants have the same sign.

Consider the $\alpha = 0$ diagonal case. If s is real it is its own complex conjugate. If the corresponding eigenvalue of M , $\lambda = s + 1$, is real and negative it need not have another negative partner in the eigenvalues of either spin species. Thus, positivity is not, in general, guaranteed in this case. We later give a simple example in (78) and find that positivity can be guaranteed with small enough $\tilde{\kappa}^2$. We observe that positivity can be lost when N_t is odd and $\tilde{\kappa}$ is large and have not found an N_t -even example; we conjecture that this is generic and positivity can always be guaranteed by increasing N_t and approaching the continuum limit. Since we know of no large-scale computational effort using this combination of basis, discretization, and odd N_t we leave a precise determination

⁶Actually, Ref. [48] differs by a factor of two in the exponent. For the one-site problem, we later give the explicit form in (66), confirming our result. Note that one does get $\prod_i \det F_i = \exp(-\Phi)$ in forms like (26) but no demonstration of a partitioning of the configuration space follows.

TABLE I. Demonstrated relationships between the eigenvalues s in the half-filled bipartite case. The eigenvalues of the fermion matrix are these eigenvalues plus one. When no subscript is attached the relationship is between eigenvalues for the same species. When $\alpha = 0$ we attach $\uparrow$ and $\downarrow$ subscripts to indicate eigenvalues for the different spin species; for $\alpha = 1$, we attach p and h for particles and holes. The last two relationships for the exponential $\alpha = 1$ case imply conjugate reciprocity and lead to a partitioning of the configuration space into sectors separated by exceptional configurations.

	Exponential	Diagonal
$\alpha = 0$	$s \leftrightarrow se^{2\pi i/N_t}$	$s \leftrightarrow -s$
	$s \leftrightarrow s^*$	$s \leftrightarrow s^*$
	$s_\uparrow \leftrightarrow 1/s_\downarrow$	
$\alpha = 1$	$s \leftrightarrow se^{2\pi i/N_t}$	$s \leftrightarrow -s$
	$s_p \leftrightarrow s_h^*$	$s_p \leftrightarrow s_h^*$
	$s_p \leftrightarrow 1/s_h$	

of how large N_t must be to ensure positivity to future work. We focus on N_t even.

Knowledge of eigenvalues of S can help us discover eigenvectors of M . Solving the eigenvalue equation using a known eigenvalue s , $(S - s\mathbb{1})v = 0$ yields the eigenvector v of S . Since $S = M - \mathbb{1}$, S and M share all their eigenvectors. However, S may be much better conditioned than M and thus much easier to invert. It may also provide a numerical speedup to find an additional eigenvector for the cost of a single solve if the associated eigenvalue comes for free. We are investigating the acceleration of the inversion of M using knowledge of the relationships between eigenvalues in Table I.

E. Extreme limits

In the general case, it is hard to analytically extract features of the probability weight function. However, in particular limits, additional symmetries emerge and yield additional information about the weight.

First, when U/κ gets very small, for fixed $U\beta$, the interaction is effectively turned off and F gets close to $\mathbb{1}_{N_t}$. Then the path integral nearly factorizes into N_t copies of the noninteracting path integral,

$$\lim_{U/\kappa \rightarrow 0} \mathcal{Z}_{N_t} = (\mathcal{Z}_1)^{N_t} \quad (54)$$

where the right-hand side is N_t powers of the $N_t = 1$ problem with the same $\tilde{U}$. We call this limit the weak-coupling limit.

In the opposite limit—the limit of no hopping, $\kappa \rightarrow 0$ with fixed $U\beta$, the partition function on N_x sites factorizes into N_x copies of the one-site the partition function,

$$\lim_{\kappa \rightarrow 0} \mathcal{Z}_{N_x} = (\mathcal{Z}_1)^{N_x}. \quad (55)$$

where now the right-hand side is N_x powers of the one-site problem. We call this limit the strong-coupling limit, $U/\kappa \rightarrow \infty$ (with $U\beta$ held fixed implicit).

In the general case we have the charge conjugation symmetry discussed in Sec. IID 1,

$$\phi \rightarrow -\phi \quad (56)$$

In the two limits we can make independent negations of ϕ ,

$$\phi_{x,t} \rightarrow \begin{cases} \text{sign}_t \phi_{x,t} & (\text{weak coupling}) \\ \text{sign}_x \phi_{x,t} & (\text{strong coupling}) \end{cases}. \quad (57)$$

In the general case, we have invariance under temporal shifts and time reversal, as discussed in Secs. IID 7 and IID 8,

$$\phi_t \rightarrow \phi_{t+\tau}, \quad \phi_{+t} \rightarrow \phi_{-t}. \quad (58)$$

In the two limits, we again can make a larger set of transformations. In the weak-coupling limit, the symmetry is enhanced and we can arbitrarily permute the time slices,

$$\phi_t \rightarrow \phi_{\mathcal{T}t} \quad (\text{weak coupling}), \quad (59)$$

where $\mathcal{T}$ is a permutation. In the strong-coupling limit, we can independently perform these operations on each thread of spatial sites

$$\phi_{xt} \rightarrow \phi_{x(t+\tau_x)} \quad \phi_{xt} \rightarrow \phi_{x(\text{sign}_x t)}. \quad (\text{strong coupling}) \quad (60)$$

In the general case, we have the spatial symmetries discussed in Sec. IID 9,

$$\phi_{xt} \rightarrow \phi_{(\mathcal{P}x)t} \quad (\mathcal{P} \text{ a lattice symmetry}), \quad (61)$$

where the same operation $\mathcal{P}$ is a symmetry of the lattice and is applied to every time slice. In the weak-coupling limit, we can apply a different spatial transformation on each time slice and in the strong-coupling limit, we can arbitrarily permute the threads of spatial sites,

$$\phi_{xt} \rightarrow \begin{cases} \phi_{(\mathcal{P}_t x)t} & (\text{weak coupling, } \mathcal{P} \text{ a lattice symmetry}) \\ \phi_{(\mathcal{P}_x)t} & (\text{strong coupling, } \mathcal{P} \text{ any permutation}) \end{cases}. \quad (62)$$

These operations will provide inspiration for proposal machines which give large field transformations that are still accepted often enough, helping overcome ergodicity problems HMC may encounter. This strategy is discussed in Sec. IVC.

III. ERGODICITY PROBLEMS

We now turn to the issue of ergodicity, which is required for an accurate, unbiased Markov-Chain Monte Carlo (MCMC) algorithm. An *ergodicity problem* arises when the algorithm for updating the state of the Markov chain is unable to visit the neighborhood of every field configuration. In this case, we can introduce bias and find inaccurate results. We can further delineate between *in-principle* or *formal* ergodicity problems and *in-practice* ergodicity problems. In a formal ergodicity problem there are regions of configuration space that the update algorithm cannot find, by any means. An in-practice problem might arise when the update algorithm *could* explore the whole space but is unlikely to find important regions of configuration space in the finite amount of time you are willing to run your computer. We emphasize that in this context ergodicity is a property of an algorithm, and not of physics itself.

As hybrid Monte Carlo (HMC) is an MCMC algorithm, it relies on the previous state to propose a new state which is subjected to the Metropolis-Hastings accept/reject step. This final step is essential for maintaining detailed balance and ultimately corrects (via the ensemble average) for any numerical

errors in the evolution of the state [1]. The proposed state is obtained by integrating the equations of motion (EoMs) derived from an artificial Hamiltonian $\mathcal{H}$ in a newly introduced time direction. In the case of the Hubbard model and $\alpha = 0$, this artificial Hamiltonian is the Legendre transform of the action (10)

$$\mathcal{H}[\pi, \phi] = \frac{1}{2} \sum_{x,t} \left(\pi_{xt}^2 + \frac{1}{U} \phi_{xt}^2 \right) - \ln \det M[\phi] - \ln \det M[-\phi], \quad (63)$$

where π are newly introduced momenta conjugate to the field variables ϕ . For $\alpha = 1$ given by (11), replace $\phi \rightarrow i\phi$ in the fermion matrices. Details of this method can be found in the pioneering Ref. [1].

Areas where the integrand of the partition function [(10) and (11)] has zero weight, for example when $\det M(\phi) = 0$, are represented by infinitely tall potential barriers in the Hamiltonian (63) which repel the state during the integration of the EoMs. This, in general, is a wanted feature, since such locations in configuration space contribute nothing to the partition function and should thus be avoided. Problems arise, however, when such barriers separate regions that *do* contribute to the partition function. If there are manifolds in configuration space of codimension 1 ($N_x N_t - 1$ dimensional, when there is only one degree of freedom per site) the configuration space is partitioned and HMC trajectories cannot propagate to these different sectors, thereby violating ergodicity. This is an in-principle problem.

It may be that there are extended codimension-1 manifolds in configuration space that terminate on boundaries, in which case it is in principle possible for HMC to find a sequence of updates that visits states on both sides of the manifold. Whether this is a problem in practice depends on the model of interest; if these manifolds are very big they might take a long time to circumnavigate. The manifolds in the cases discussed here are boundary-free.

If there are zero-weight manifolds of higher codimension, the configuration space isn't partitioned and HMC can always explore the whole space, though there could still be a problem in practice.

Reference [48] pointed out that the factorization (47) and reality of $f[i\phi]$ implies a formal ergodicity problem for the exponential discretization when $\alpha = 1$. Since $f[i\phi]$ is real and takes both positive and negative values, by the intermediate value theorem, there must be zeros separating the two regions, as discussed at the end of Sec. IID 10. As first suggested in Ref. [48], there exist codimension-1 manifolds in ϕ where the determinant is zero. We can understand this result by seeing that Q in (50) cannot be changed by a continuous change in f unless f passes through zero. Were the manifolds smaller in dimension, HMC would find its way around, without having to go through the barriers, though there could nevertheless have been an issue in practice, as it might take a long time to circumnavigate the zeros. Formally, an infinitely precise HMC integrator cannot penetrate these barriers and is therefore not ergodic.

When considering the diagonal discretization in (14), the factorization (47) does not naturally emerge. As discussed in Sec. IID 10, this factorization was the result of the exact chi-

ral symmetry found in the exponential discretization. When considering the spin basis ($\alpha = 0$) with these discretizations the determinant is still real, and may (but need not) still have both negative and positive values and the intermediate zeros may obstruct the exploration of configuration space by HMC.

For the particle/hole basis using the diagonal discretization in (14), the situation is quite different. The factorization still does not naturally emerge. If one insists on writing it that way, the function $f[i\phi]$ is complex and thus we can avoid the conclusion of the intermediate value theorem that would force HMC to go through a zero in order to change the sign of f . The function $f[i\phi]$ may still have zeros, but there no longer need be codimension-1 zero manifolds and thus the space is not partitioned into regions that trap HMC trajectories. We provide numerical examples of this in the following sections.

We stress that our arguments here do not prove unequivocally that the particle/hole basis using the diagonal discretization (14) does not suffer from any formal ergodicity issues, only that it does not suffer from those identified in Ref. [40]. However, we are not aware of any other formal ergodicity issues this discretization may have. As one nears the continuum limit, since the two discretizations must agree, very tall potential barriers can rise between the exceptional configurations. This raises the possibility of an *in practice* problem; how difficult it is to overcome depends on the exact example and how fine a temporal discretization one uses. This is later demonstrated for a simple problem in Figs. 5 and 6. When we propose a general solution to the formal ergodicity problem of the exponential $\alpha = 1$ case in Sec. IV C, such a solution can also resolve the in-practice problem in the diagonal $\alpha = 1$ case that emerges near the continuum limit.

In the remainder of this section, we discuss the dynamics of the eigenvalues and the role they play in formal ergodicity problems, and then present numerical examples that support our discussion above, detailing in-principle and in-practice problems. We investigate small systems where we can perform direct comparisons with exact solutions. Though the systems are small in dimension, they capture all the relevant aspects of ergodicity (or lack thereof) that are present in larger simulations, and provide the added benefit that these aspects can be visualized.

In all our HMC simulations, unless otherwise stated, we always target an acceptance rate $\geq 80\%$ by adjusting the accuracy of our numerical integration of the EoMs. As previously stated, the error in our integration is corrected by the accept/reject step. A more accurate integration corresponds to a higher acceptance rate, but the configuration space probed by each trajectory is diminished. Conversely, a less accurate integrator allows HMC to probe more configuration space at the expense of a lower acceptance rate. Since these examples are so small, we produce an ensemble of fields $\{\phi\}$ by thermalizing for a few thousand trajectories and generating 10 000 to 100 000 HMC trajectories per simulation. We compute correlators only on every 10th configuration to reduce autocorrelations. Our uncertainties are given by the standard deviation of bootstrap samples of the particular quantity in question.

The HMC ensembles and correlator data are available online in Ref. [50]. This data was generated using Isle [51], a new library currently in development for HMC calculations in the Hubbard model.

A. Exceptional configurations and zero eigenvalues

An *exceptional configuration* is one with zero weight in the path integral. The zero weight implies an infinite potential in the classical EoMs used in HMC, and as a trajectory nears an exceptional configuration the force diverges. In the case of the Hubbard model these zeros arise from a vanishing fermion determinant, which in turn corresponds to a vanishing eigenvalue of the fermion matrix.

For $\alpha = 0$, the fact that $\det M$ can be positive and negative, and is always real, implies that manifolds of exceptional configurations partition the configuration space; there is a formal ergodicity problem. We observe that the frequency with which exceptional configurations are encountered increases as one approaches the continuum limit $N_t \rightarrow \infty$. This can be understood by considering the noninteracting eigenvalues in (33). Two factors drive the increased frequency. First, δ , which vanishes with increasing N_t , controls how close eigenvalues are to the origin. In the exponential discretization, the radii go to 1 with vanishing δ . In the diagonal discretization, the rings' centers converge on 1 with vanishing δ . The other issue is that the rings of eigenvalues become increasingly dense with N_t , as the Matsubara frequencies come closer together. This observation also holds for the interacting case with typical auxiliary field configurations.

The frequency of exceptional configurations also depends on the spatial lattice. Again considering the noninteracting limit and assuming the hopping term H_0/κ has a vanishing energy eigenvalue $\epsilon = 0$. In the exponential case there is at least one ring of eigenvalues with unit radius which puts the eigenvalues close to the origin; in the diagonal case one of the rings is centered exactly on 1.

Moreover, as the infinite-volume limit is taken the eigenvalues become more dense. For regular lattices with n nearest neighbors, the nearest-neighbor hopping Hamiltonian H_0/κ has eigenvalues ϵ bounded by $-n \leq \epsilon \leq n$. The number of eigenvalues of the noninteracting Hamiltonian is given by the number of sites; going towards the infinite volume limit means more and more rings will appear and can get close to the origin. This observation, again, is borne out of the interacting case with typical auxiliary field configurations. Of course, with a nonzero field configuration, the eigenvalues move and the observations are only qualitatively true—the eigenvalues are no longer exactly controlled by the noninteracting energies, for example.

Even when the lattice is large, some geometries may be more favorable than others. The square lattice with periodic boundary conditions always has a zero eigenvalue; the honeycomb lattice only has a zero eigenvalue when the lattice dimensions are congruent to 0 (mod 3).

In Fig. 1, we show the spectra of fermion matrix eigenvalues in the complex plane. The interacting eigenvalues for the example exponential $\alpha = 0$ configuration lie on rings centered on $1 + 0i$ at angles determined by the Matsubara frequencies. Were this generically true for all ϕ , the determinant would always be positive and there would be no in-principle ergodicity problem. However, the eigenvalues are only constrained to obey the partnerships in the exponential $\alpha = 0$ portion of Table I.

Another way to satisfy those relationships emerges when two rings of eigenvalues have the same radius. Then, as ϕ changes, the two rings can counter-rotate by the same angle, still obeying the constraints of Table I. When the two counter-rotated rings of eigenvalues eventually meet they can then part, moving radially. Since the counter-rotation is always by the same angle, when the rings part the eigenvalues always lie on the rays determined by the Matsubara frequencies, or exactly halfway between those frequencies—putting eigenvalues on the real axis. If one ring crosses $r = 1$, the determinant changes sign.

We conjecture that once the eigenvalues are off of these rays the only way to maintain all the symmetry properties of the fermion matrix's spectrum is to remain locked to the radius where they first collided, and then once apart the symmetry properties cannot be maintained unless the eigenvalues are locked to the rays. We have not seen examples where more than two rings all have the same radius and perform an even more complicated dance, though such a dance may be possible.

To illustrate the above statements, we track eigenvalues of M from a configuration with positive determinant to a configuration with negative determinant in Fig. 2. Animations of the processes depicted in the figure are available in the supplementary material [52]. In the animation of the $\alpha = 0$ exponential case, one can see the inner ring crosses $r = 1$ between $\tau = 0.8430$ and 0.8431 and the determinant flips sign.

We generated a single configuration ϕ^- with a real, negative determinant for all four bases and discretization choices. We show the spectrum of $M[\tau\phi^-]$ where the fictitious time τ runs from 0 (purple) to 1 (yellow), so that the spectrum goes from the noninteracting spectrum, which has a positive determinant, to the spectrum of ϕ^- with negative determinant. The two leftmost panels show the eigenvalue trajectories for the exponential $\alpha = 0$ case, and f , related to the determinant by (47). While it looks like the eigenvalues meet and rotate around the point 1 at a radius $r = 1$, this is only approximately true.

The other cases are similar to the exponential $\alpha = 0$ case, though the constraints on the eigenvalues are different. In the exponential $\alpha = 1$ case, the eigenvalues enjoy conjugate reciprocity. If one eigenvalue crosses the origin, another eigenvalue must cross in the opposite direction, preserving the sign of the determinant. When instead two rings meet at $r = 1$, the eigenvalues are their own conjugate reciprocals. Then, the rings can rotate oppositely, though the angles they rotate by need not be equal as there is no complex-conjugation constraint as in the $\alpha = 0$ case. The determinant changes sign while the two rings rotate. When the two rings again meet, they can part radially, preserving the conjugate reciprocity constraint.

Eigenvalues in the $\alpha = 0$ diagonal case are no longer locked to the rays, because the eigenvalues do not come in perfect rings. However, the complex-conjugate pairing still means a real eigenvalue must cross the origin to flip the sign of the determinant.

In the $\alpha = 1$ diagonal case, the eigenvalues need not meet at all to traverse from a real and positive determinant to a real and negative determinant. We show the movement of

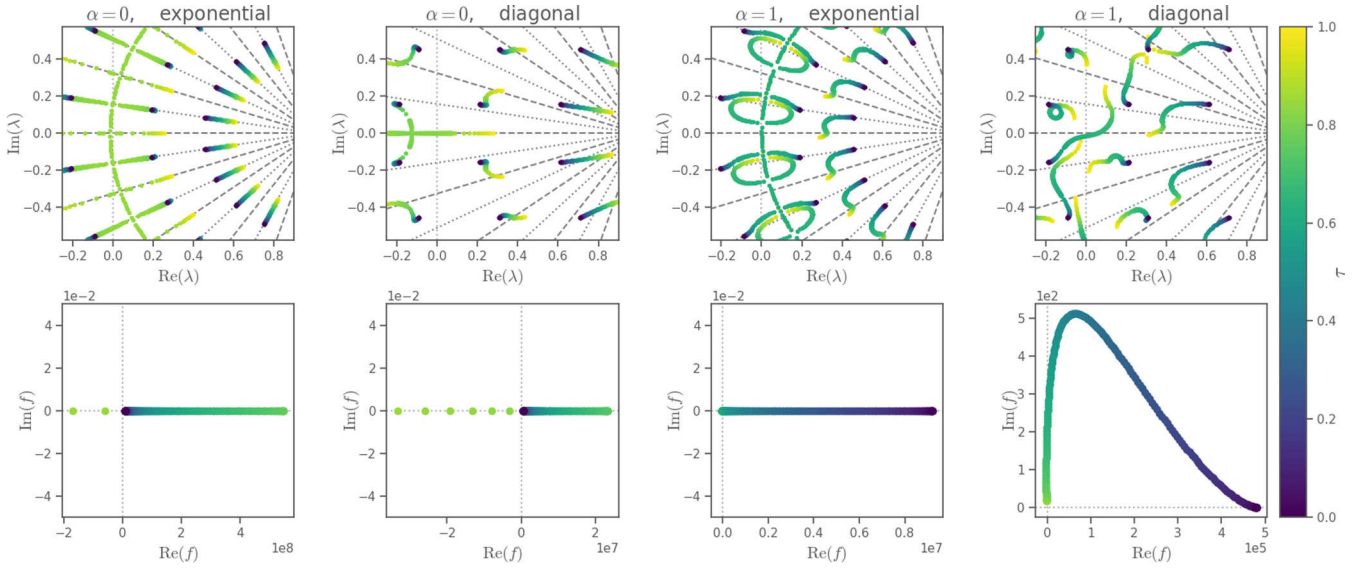


FIG. 2. Eigenvalues λ and f as in (47) for $\phi = \tau\phi^-$, where $\det M[\phi^-] < 0$ in all cases. The fictional time τ runs from 0 (purple), so that ϕ is the noninteracting case and the determinant is positive, to 1 (yellow) so that the determinant is negative; intermediate values of τ are shown in the color bar on the right. The eigenvalue trajectories are described in the text. The eigenvalue locations were computed on an $N_t = 20$ four-site honeycomb lattice with $\kappa\beta = 4$. See Ref. [52] for animations of these eigenvalue dances.

eigenvalues and f (defined by (47) even for the diagonal case where such a factorization does not naturally emerge) for each case in Fig. 2.

B. The one-site problem

Strictly speaking, the one-site problem is not bipartite. However, since there is no hopping in this case, the tight-binding Hamiltonian H_0 can be ignored ($\kappa = 0$) and the exponential, linear, and diagonal discretizations (12), (13), and (14) are all equivalent

$$M[\phi]_{x't',xt} = \delta_{x',x}\delta_{t',t} - e^{\phi_{xt}}\delta_{t',t+1}. \quad (64)$$

The determinant of the fermion matrix in this case can be expressed in closed form:

$$\det M[\phi] = 2 \cosh\left(\frac{\Phi}{2}\right) e^{\Phi/2}, \quad (65)$$

$$\det M[i\phi] = 2 \cos\left(\frac{\Phi}{2}\right) e^{i\Phi/2}, \quad (66)$$

where $\Phi = \sum_t^N \phi_t$ as in (31). We see that

$$f[\phi] = 2 \cosh(\Phi/2), \quad f[i\phi] = 2 \cos(\Phi/2), \quad (67)$$

that f is even, and that $f[\phi]$ does not go through zero, while $f[i\phi]$ does. Thus we expect that the particle/hole basis has infinite barriers in the artificial Hamiltonian, and formally has ergodicity issues, while the spin basis has no formal ergodicity issue.

However, as we shall show shortly, calculations in the spin basis exhibit a bimodal distribution that becomes increasingly separated for large $U\beta$. Though calculations in this basis do not formally suffer from ergodicity issues as $\det M[\phi]$ never vanishes for any ϕ , in practice the separation of the modes for large $U\beta$ essentially separates two regions that are extremely unlikely to be connected via HMC, regardless of the accuracy

of the integration of the EoMs. This presents an *in practice* ergodicity issue.

Using (65) with (10), one finds that the weight of a field configuration ϕ is given, up to an overall normalization, by

$$W[\phi] = \det M[\phi] \det M[-\phi] e^{-\frac{1}{2\beta} \sum_t \phi_t^2} \\ = 4 \cosh^2\left(\frac{\Phi}{2}\right) e^{-\frac{1}{2\beta} \sum_t \phi_t^2} \quad (68)$$

$$= 4 \cosh^2\left(\frac{\Phi}{2}\right) e^{-\frac{1}{2\beta} \Phi^2} \exp\left(-\frac{1}{4U\beta} \sum_{t_1, t_2} (\phi_{t_1} - \phi_{t_2})^2\right). \quad (69)$$

Note that after completing the squares in the exponent, we have completely exposed the Φ dependence of the probability weight. The other factor,

$$\exp\left(-\frac{1}{4U\beta} \sum_{t_1, t_2} (\phi_{t_1} - \phi_{t_2})^2\right), \quad (70)$$

can be shown to be independent of Φ , by directly differentiating with respect to Φ and using $\partial\Phi/\partial\phi_t = 1$ and $\partial\phi_i/\partial\phi_j = \delta_{ij}$ for simplification. Thus the distribution of Φ is bimodal, as seen in the factor $4 \cosh^2(\frac{\Phi}{2}) e^{-\frac{1}{2\beta} \Phi^2}$ in (69). This distribution is strongly peaked about $\Phi \approx \pm U\beta$, implying that the peaks of the modes are separated by a distance $2U\beta$, and is exponentially small in between. This analysis also shows that the modes become further separated when either U is increased (strong-coupling limit), or β is increased (zero temperature limit), or both.

For the $\alpha = 1$ case, using (66) with (11) gives

$$W[\phi] = \det M[i\phi] \det M[-i\phi] e^{-\frac{1}{2\beta} \sum_t \phi_t^2} \\ = 4 \cos^2\left(\frac{\Phi}{2}\right) e^{-\frac{1}{2\beta} \Phi^2} e^{-\frac{1}{4U\beta} \sum_{t_1, t_2} (\phi_{t_1} - \phi_{t_2})^2}. \quad (71)$$

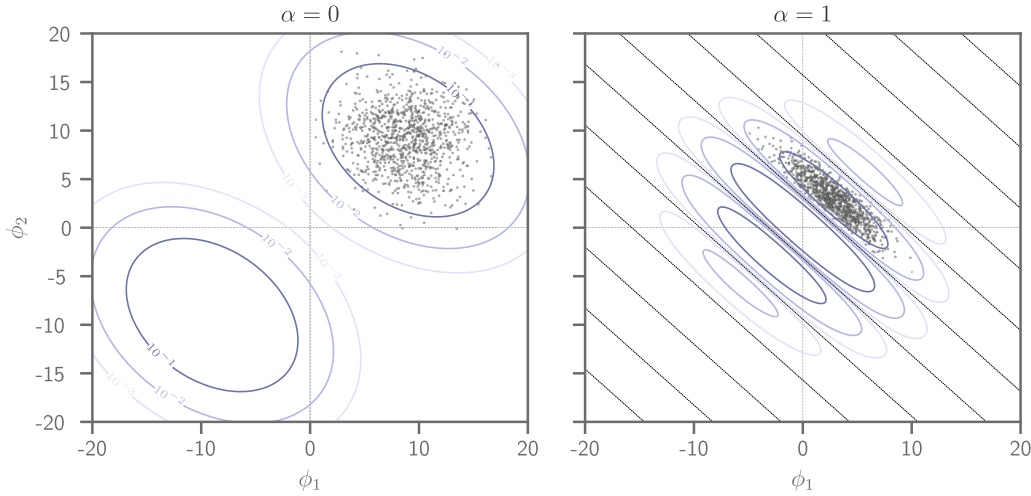


FIG. 3. Contours of constant probability, (69) (left, $\alpha = 0$) and (71) (right, $\alpha = 1$) with $U\beta = 18$ and $N_t = 2$ for the one-site problem, where the subscript on the Hubbard-Stratonovich fields ϕ indicates the time slice. The dark contours represent high probability density, which decreases as the contours get lighter. The Φ axis runs from bottom left to top right. (Right) The diagonal (black) lines correspond to locations where the probability weight (71) is exactly zero, creating barriers to HMC; their regular spacing is a consequence of the periodicity of the determinant, discussed in Sec. IID 3. We show every tenth configuration of 10 000 configurations (generated with a very precise integrator) as points. It is evident that the HMC algorithm was trapped in a region of high probability density in the $\alpha = 0$ case by the wide separation with low probability density and in the $\alpha = 1$ case by the lines of zero weight.

The last factor is the same as before and is Φ -independent, and thus the manifold of zeros is indeed codimension 1 so that a precise HMC integrator cannot circumnavigate the zeros and a formal ergodicity problem arises. The distribution of Φ is determined by $4 \cos^2(\frac{\Phi}{2}) e^{-\frac{1}{2U\beta}\Phi^2}$, a product of a Gaussian of width $\sqrt{U\beta}$ centered at $\Phi = 0$ and a simple, periodic $\cos^2(\frac{\Phi}{2})$ function. The zeros of $\cos(\frac{\Phi}{2})$ dictate the zeros of the kernel, and we find zeros at $\Phi = (2n + 1)\pi$ for all integers n , independent of the value $U\beta$. This function is multimodal, but the modes remain close together, even as U or β are taken large. Rather than separating two important modes, taking the low-temperature or strong-coupling limit broadens the Gaussian and increases the number of important modes. We can analytically determine the probability distribution for Φ ,

$$W[\Phi] = \frac{e^{-\frac{\Phi^2}{2U\beta} - \frac{U\beta}{4}} \cosh^2\left(\frac{\Phi}{2}\right)}{\sqrt{2\pi U\beta} \cosh\left(\frac{U\beta}{4}\right)} \quad (\alpha = 0), \quad (72)$$

$$W[i\Phi] = \frac{e^{-\frac{\Phi^2}{2U\beta} + \frac{U\beta}{4}} \cos^2\left(\frac{\Phi}{2}\right)}{\sqrt{2\pi U\beta} \cosh\left(\frac{U\beta}{4}\right)} \quad (\alpha = 1). \quad (73)$$

Note that these expressions depend only on the product $U\beta$, and in particular are independent of N_t .

As a visual aid, we consider the $N_t = 2$ case. Here there are only two Hubbard-Stratonovich degrees of freedom, ϕ_1 and ϕ_2 . We plot contours of the kernels for the two bases [(69) and (71)] in the case when $U\beta = 18$, so that the modes in the $\alpha = 0$ case are well-separated, in Fig. 3 and also show the codimension-1 manifolds that produce a formal ergodicity problem in the case of $\alpha = 1$. Note that such lines are absent in the $\alpha = 0$ case. Nevertheless, in both cases we will get a biased result—the $\alpha = 0$ case has an in-practice problem caused by the isolation of modes by a region of very small probability density. For both cases, we generated 10 000 configurations

using a precise HMC integrator (with 20 leapfrog steps for a unit-length molecular dynamics trajectory, yielding a near 100% acceptance rate), shown as points. The Φ axis in these plots runs along the diagonal from bottom left to top right, and the numerical distribution generated by HMC is clearly not symmetric around $\Phi = 0$, a feature clear in the analytic expressions (72) and (73) and guaranteed by the fact that f is even (47).

We can also study problems with substantially larger N_t , or equivalently, finer discretizations. One observable that is of interest is the correlation function $C_{ij}(\tau)$, where i, j refer to spatial locations. For the one-site problem, we are restricted to $i = j = 1$. In the continuum limit, we can compute the correlation function exactly

$$C_{11}(\tau) \equiv \langle a(\tau) a^\dagger(0) \rangle = \left\langle \sum_t (M_{11}^{-1})_{t+\tau,t} \right\rangle = \frac{\cosh(U(\beta - 2\tau)/4)}{2 \cosh(U\beta/4)}, \quad (74)$$

which, being a continuum-limit quantity, is independent of δ . We can estimate C as an ensemble average over configurations generated by HMC by

$$C_{ij}(\tau) \approx \frac{1}{N_c} \sum_{\phi \in \{\phi\}} \sum_t M_{ij}^{-1} \left[\begin{cases} \phi & \text{if } \alpha = 0 \\ i\phi & \text{if } \alpha = 1 \end{cases} \right]_{t+\tau,t}, \quad (75)$$

where N_c is the number of generated Hubbard-Stratonovich field configurations in the ensemble $\{\phi\}$. Lack of ergodicity in our sampling of ϕ will give disagreement between simulated and exact correlators. To demonstrate this, we consider in Fig. 4 as case using a precise MD integrator (acceptance rate >99%) with extreme $U/\kappa = 10$ and $\beta\kappa = 6$, so that when $\alpha = 0$ the modes are widely separated and when $\alpha = 1$ many modes contribute. The top row corresponds to $\alpha = 0$, whereas the bottom row is $\alpha = 1$. The left column shows the histogram

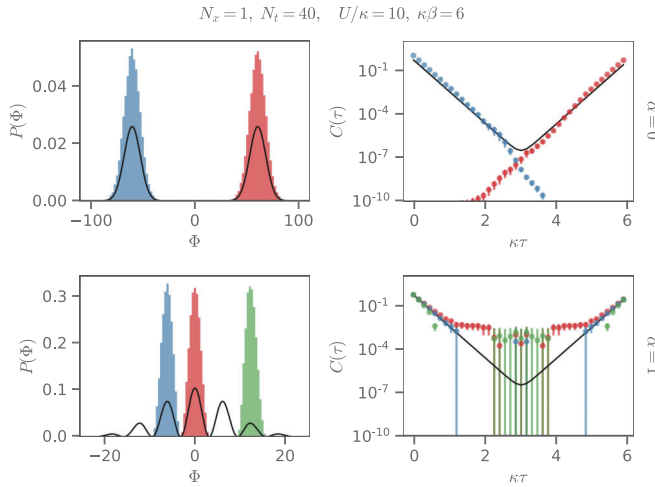


FIG. 4. Histograms of Φ and correlators for the one-site problem with very fine integration. Monte Carlo evolution was started in different lobes and the trajectories are stuck in those lobes, shown by the different colors. All ensembles consist of 50 k thermalized configurations. Single-particle correlators to the right are color coded to match the histograms in the left column. Points where $C(\tau) \leq 0$ are not shown because of the logarithmic scale. The black lines show the exact results. The histograms are all normalized but, because HMC was trapped, the numerically generated histograms are much taller and thinner than the exact results.

of $\Phi = \sum_t \phi_t$ for different runs, while the right column shows the calculated correlators. In all plots, the black line is the exact result. Different HMC runs are differentiated by color. For the $\alpha = 0$ case, it is clear that the HMC trajectories are trapped in one of the two modes (top left panel), despite there being no regions with $\det M = 0$. The corresponding correlators calculated with these fields are color-matched and shown in the top right panel. Clearly the sampling of fields is grossly biased, and this is reflected in the large disagreement between simulated and exact correlators. For the $\alpha = 1$ case, three different runs (red, blue, and green) were performed

with different starting points for Φ , each separated by a point where $\det M = 0$. From the histogram (lower left panel), it is clear that the runs are trapped within their respective sectors, and their corresponding correlators (color matched with the histograms) each differ from the exact result. Presumably, the correct linear combination of these correlators [with relative weights given by (73)] would give the exact correlator. Of course, in a more complicated system we do not know such weights *a priori*, and therefore would not know how to combine such correlators to produce the correct result.

These examples, though extreme, demonstrate how both *formal* (for $\alpha = 1$) and *in practice* (for $\alpha = 0$) ergodicity problems can arise in HMC simulations, even when simulations are of the same physical system. The choice of basis greatly influences the behavior of the sampled field configurations, and in turn can drastically impact calculated observables such as the correlator.

C. The two-site problem

We now consider the two-site problem to demonstrate how different discretizations can lead to ergodicity issues. To keep the presentation reasonable, we restrict our analysis to the exponential and diagonal discretizations given by (12) and (14), respectively, using the $\alpha = 0$ and 1 bases.⁷ In (14), the term h is given by

$$h_{x',x} = \tilde{\kappa} \delta_{\langle x',x \rangle} = \begin{pmatrix} 0 & \tilde{\kappa} \\ \tilde{\kappa} & 0 \end{pmatrix}, \quad (76)$$

while the matrix in (12) is its exponential, given by

$$[e^h]_{x',x} = \begin{pmatrix} \cosh \tilde{\kappa} & \sinh \tilde{\kappa} \\ \sinh \tilde{\kappa} & \cosh \tilde{\kappa} \end{pmatrix}. \quad (77)$$

For extreme simplicity, we turn to the problem of two sites on a single time slice where, as in the previous section, there are two degrees of freedom, ϕ_1 and ϕ_2 , the label now indicates the spatial site. Adopting the factorization shown in (47) of the determinant of the fermion matrix, we have that

$$f[\phi] = \begin{cases} 2[\cosh(\Phi/2) + \cosh(\frac{\phi_1 - \phi_2}{2}) \cosh(\tilde{\kappa})] & \text{exponential discretization} \\ 2[\cosh(\Phi/2) + \cosh(\frac{\phi_1 - \phi_2}{2}) - \frac{\tilde{\kappa}^2}{2} e^{-\Phi/2}] & \text{diagonal discretization} \end{cases}, \quad (78)$$

where again $\Phi = \phi_1 + \phi_2$. Close inspection of the equations above shows that both $f[\phi]$ and $f[i\phi]$ are always real in the exponential discretization. In the diagonal discretization, on the other hand, only $f[\phi]$ is real and $f[i\phi]$ is in general complex.

The probability density

$$W[\phi] \propto f[\phi]f[-\phi]e^{-\frac{\phi_1^2 + \phi_2^2}{2U\beta}} \quad (79)$$

is positive semidefinite for both discretizations with $\alpha = 1$ because $f[i\phi]f[-i\phi] = |f[i\phi]|^2$. In the exponential $\alpha = 0$ case it is positive definite as well because $f[\phi]$ is. Positive (semi)-definiteness is however not guaranteed in the diagonal discretization with $\alpha = 0$. For large enough values of $\tilde{\kappa}$, the product $f[\phi]f[-\phi]$ can become negative, which means

that (79) cannot be interpreted as a probability distribution. Normally this is not an issue; we can increase N_t and positivity is eventually assured. However, for this particular example, we must enforce an additional constraint to maintain positivity, namely, $\tilde{\kappa}^2 \leq 1$.

Figure 5 shows probability contours of this single-time slice problem in the case when $\tilde{U} = U\beta = 18$ and $\tilde{\kappa} = \kappa\beta = 1$ for these different discretizations and bases as well as

⁷In simple cases, we have found that the linear discretization (13) exhibits the same behavior as that of the diagonal discretization (14) but have not explored the linear case as extensively. The lack of conjugate reciprocity suggests there ought to be no formal ergodicity problem when $\alpha = 1$.

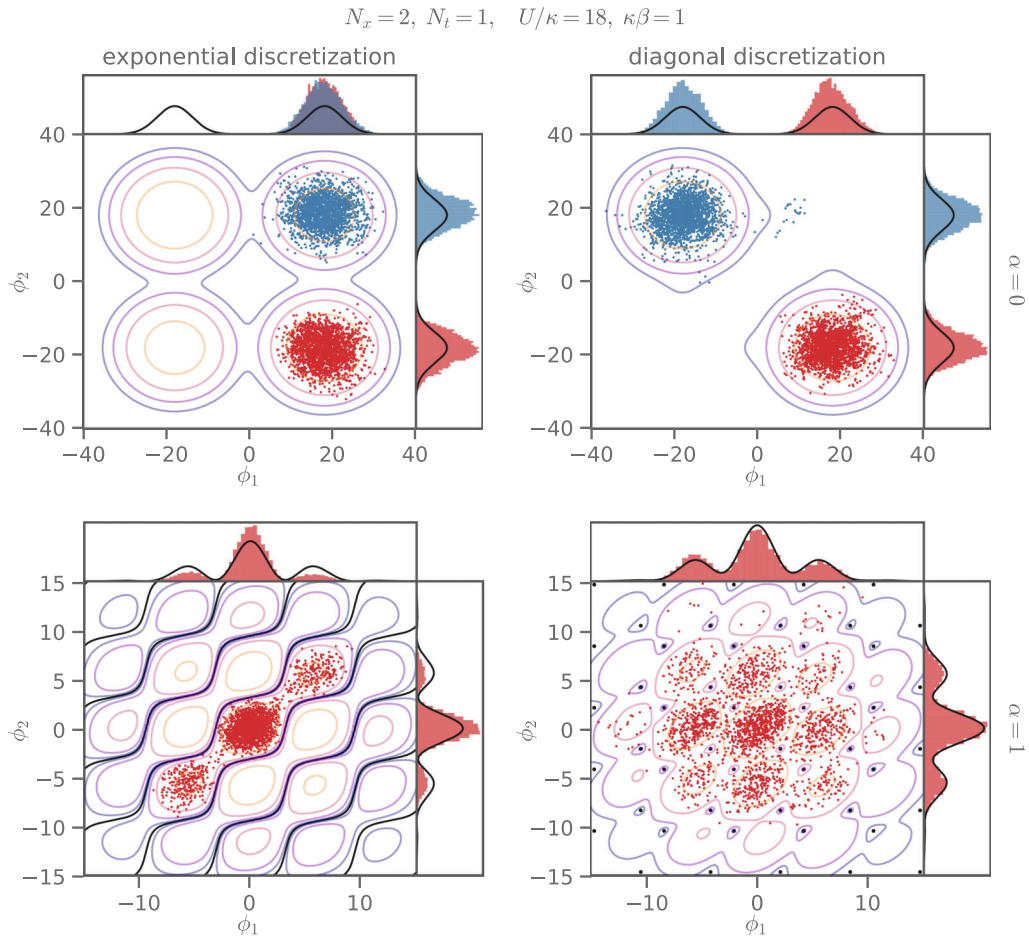


FIG. 5. Contours of constant probability weight (79) using the exponential discretization (left column) and the diagonal discretization (right column) with $\alpha = 0$ (top row) and $\alpha = 1$ (bottom row) for the two-site problem with $U\beta = 18$, $\kappa\beta = 1$, and $N_t = 1$. Contour colors range from orange (light) for large to blue (dark) for small values in arbitrary units. In the bottom panels, the black diagonal squiggly lines and regularly spaced black dots correspond to locations where (79) is exactly zero; their regular spacing is a consequence of the periodicity of the determinant, discussed in Sec. IID 3. Crossing a black line changes the sign of the fermion determinant. The colored points show 10 k HMC trajectories. For $\alpha = 0$, evolution was started in different modes of the field space, shown in different colors. For $\alpha = 1$, evolution was started at the origin with a small random deviation. The outer histograms show the marginal distributions in ϕ_1 and ϕ_2 where the black lines are the exact results obtained from (79).

field configurations generated by HMC with a fine integrator (acceptance rate $> 99\%$).

Regardless of discretization, $\alpha = 0$ exhibits no field configurations of zero weight when $\tilde{\kappa} \leq 1$, so there is no formal ergodicity problem. HMC nevertheless gets trapped in single lobes of the probability distribution, and there is a problem in practice. In the case of the diagonal discretization, the two modes are symmetric about the $\phi_1 = \phi_2$ line, and the histogram of Φ will *not* show evidence of this bimodal structure as Φ is peaked about zero in both modes in the same way.

With $\alpha = 1$ there are field configurations with zero weight. With the exponential discretization given by (12) there are entire lines of zero weight, separating the field space into different sectors, and giving rise to a formal ergodicity problem. With a fine integrator, HMC gets trapped between the infinite barriers.

The $\alpha = 1$ case with the diagonal discretization given by (14) is particularly interesting since the probability density vanishes only at isolated points, not lines. Regions of relevant weights are no longer separated by infinite barriers, and

therefore ergodicity is formally preserved. Evidence of this is seen in the distribution of field configurations generated by HMC, shown as red points. The HMC algorithm successfully reached into nearby basins that would have been unreachable in the case of the other discretization. The diagonal discretization is clearly less restrictive in the $\alpha = 1$ case, compared to the exponential discretization.

The preceding example has only little bearing on a realistic calculation due to the fact that it represents a single time slice calculation. Before considering a larger N_t case, we point out that we can mimic a finer time discretization by reducing κ . This can also be viewed as taking the strong-coupling $U/\kappa \gg 1$ limit. Figure 6 shows the contours in the case where κ has been reduced by an order of magnitude compared to Fig. 5. Note that all contours within their respective α bases are nearly the same for both discretizations, giving credence to our claim that the discretizations become equivalent in the continuum limit. This is despite the fact that for $\alpha = 1$ the allowed values where $\det M[i\phi] = 0$ (black lines and dots of bottom row) are topologically different.

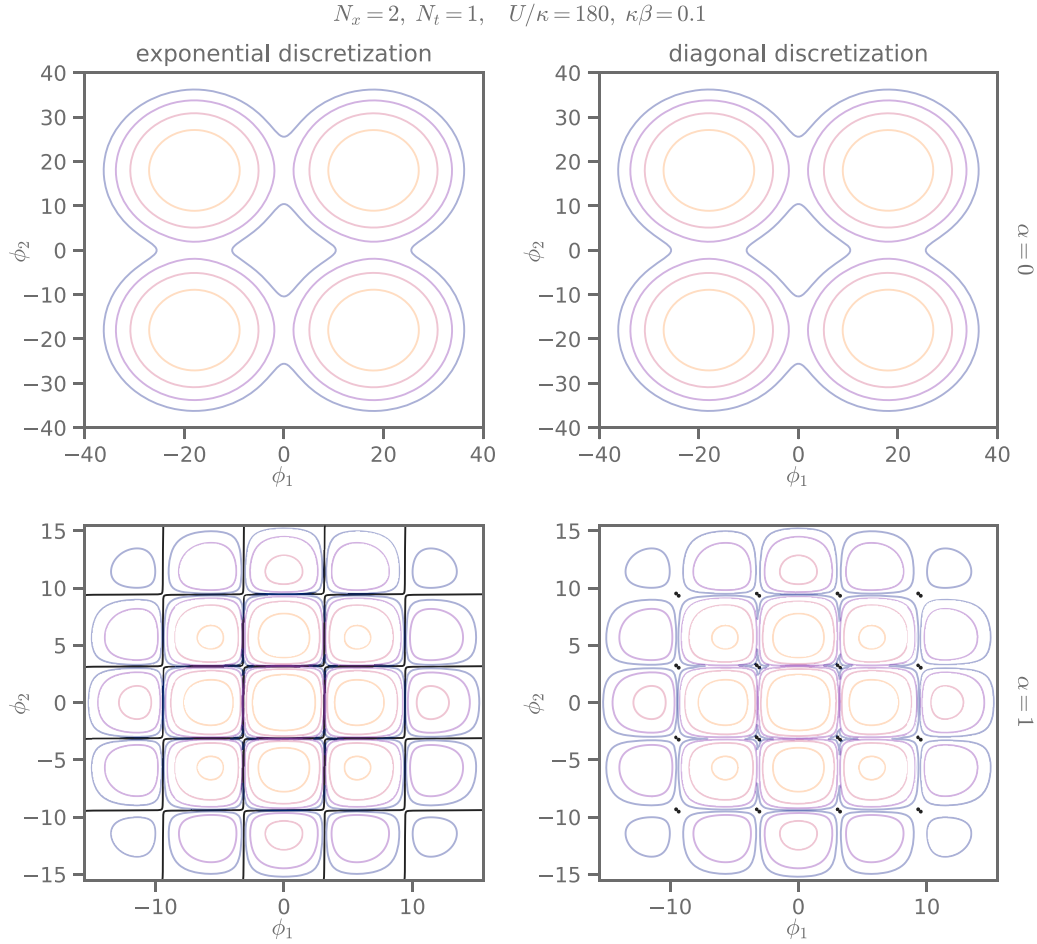


FIG. 6. Analogous contours to Fig. 5 but using $\kappa = 0.1$ to mimic approaching the continuum limit. The exponential and diagonal discretizations now have very similar probability contours. For the $\alpha = 1$ case, the black lines of zero weight of the exponential discretization pinch and nearly close, while the zero-weight configurations remain isolated in the diagonal discretization.

We now consider $N_t = 40$ again with extreme $U/\kappa = 10$ and $\beta\kappa = 6$. In Fig. 7, we show the distribution of $\Phi_x \equiv \sum_t \phi_{xt}$ as in (31) from different HMC simulations using the different bases and discretizations. We again use a high precision integrator with acceptance rate $> 99\%$. The left column represents the exponential discretization and the right the diagonal one. The top row has $\alpha = 0$, the bottom $\alpha = 1$. Histograms of Φ_1 and Φ_2 are also shown in the figures. Recall that in the large U/κ limit, the hopping term becomes negligible compared to the on-site interaction and the two-site problem factors into the product of two one-site problems, as discussed in Sec. II E. One expects then that the distributions of Φ_x are given by (72) and (73), which are shown as black lines in the marginal histograms. The fact that our simulated distributions qualitatively agree with these distributions is due to the fact that $U\beta = 60$ is in the strong-coupling regime. For the $\alpha = 0$ case, it is clear that there is a multimodal distribution and HMC trajectories are separated into these modes, even though there are no configurations with $\det M = 0$, which is similar to the one-site case. HMC samplings are again grossly biased. For the exponential $\alpha = 1$ case (bottom left panel), trajectories are also biased because of the separation of regions by $\det M = 0$. The sampling of fields for the diagonal $\alpha = 1$ case (bottom

right) on the other hand is relatively symmetric and seemingly unbiased.

For the two-site system, there are two linearly independent correlators. If we label one site A , and the other B , then the two correlators are

$$C_{\pm}(\tau) = \frac{1}{2}(C_{AA}(\tau) + C_{BB}(\tau) \pm [C_{AB}(\tau) + C_{BA}(\tau)]), \quad (80)$$

where $C_{ij}(\tau)$ represents the correlator of a quasiparticle starting at site i and propagating to site j and is estimated by (75). In the strong-coupling limit, these two correlators approach the 1-site correlator solution of (74). In Fig. 8, we show the corresponding calculated correlators from the field distributions in Fig. 7, arranged and colored in the same way. As expected, the correlators for the $\alpha = 0$ case agree very poorly with the exact result given by the black lines due to the biased sampling of fields. In the exponential $\alpha = 1$ case, the red points sampled from the diagonal band in Fig. 7 show only a small deviation from the exact result in regions with small noise. If a different band is sampled however, deviations can be larger as demonstrated by the blue points which show a clear deviation from the black line. The impact of an ergodicity problem on correlators depends on the contributing states as demonstrated in the next section. On the other hand, correlators calculated in the diagonal $\alpha = 1$ case have better

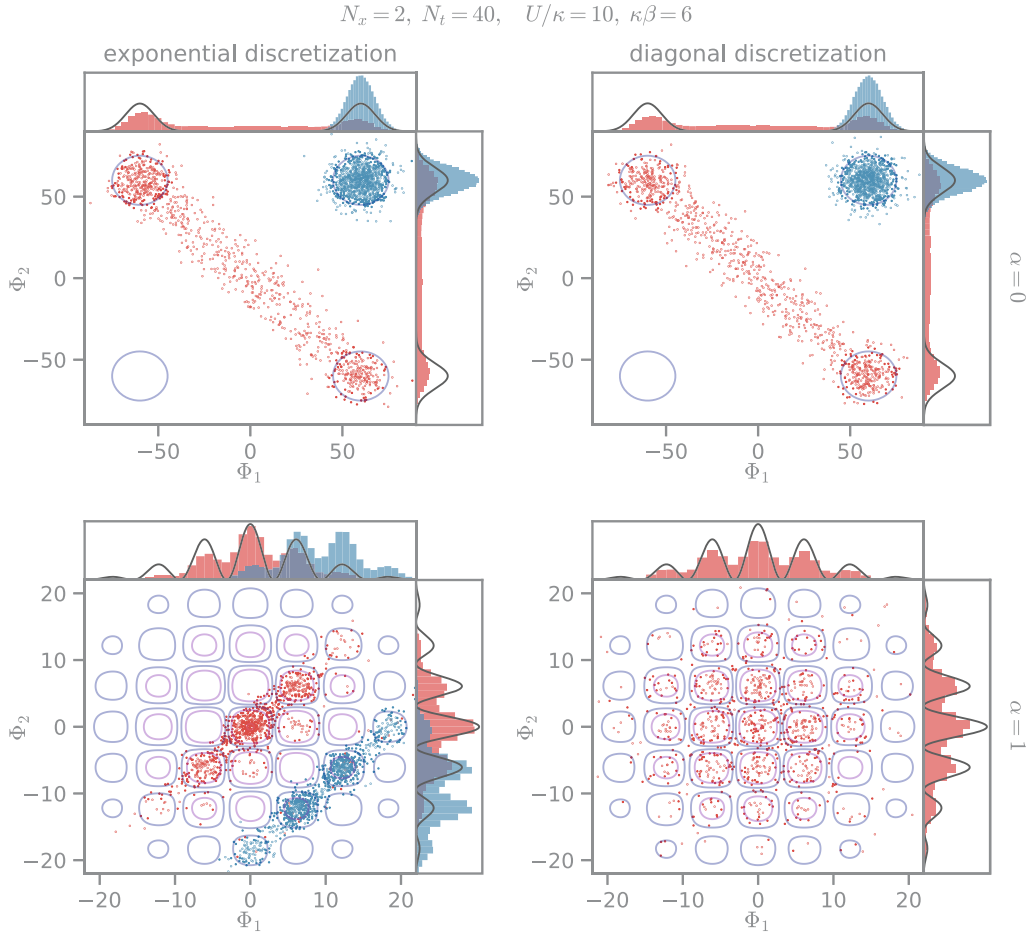


FIG. 7. HMC history of $\Phi_x = \sum_i \phi_{xi}$ for the two-site problem with 100 k trajectories with $U\beta = 60$, $\kappa\beta = 6$, and $N_t = 40$. Only every 100th trajectory of each ensemble is shown. Several HMC streams for each case were produced by starting in different modes in configuration space as indicated by color. Only for $\alpha = 1$, diagonal there is no difference, where the run is started, so only one stream is shown. The black lines in the marginal distributions for Φ_1 and Φ_2 are the exact *one-site* distributions which are recovered in the strong-coupling limit. The contours show the product of the one-site distributions and are *not* exact results for this case.

agreement with the exact results, particularly for early and late times, due to a less biased sampling of fields.

These examples give further evidence of how the choice of bases can impact the sampled fields, as discussed in the previous section. In addition to this, however, is the fact that different discretizations can also lead to disparate sampling of fields, and ultimately impact the fidelity of observable calculations. In Appendix D, we use these same ensembles to calculate other observables and similarly find that the exponential discretization can suffer from ergodicity problems that are absent in the diagonal case.

D. The four-site problem

Up to this point only the exponential $\alpha = 1$ discretization has exhibited cases where $\det M$ is always real and can be negative. With four and more sites, the exponential $\alpha = 0$ discretization also exhibits $\det M < 0$ cases, as was originally pointed out in Ref. [40]. It would be interesting to understand why the two site problem is protected from negative determinants in the other cases.

Reference [40] argued that the initial starting point of the HMC evolution can lead to drastically different results due to

the separation of $\det M < 0$ and $\det M > 0$ in the exponential $\alpha = 0$ case. The authors provide an explicit example of the equal-site $\langle C_{ii}(\tau) \rangle$ correlator calculated on a 4×4 square lattice, showing a clear dependence of the correlator determined from HMC runs that originated from either a $\det M > 0$ or a $\det M < 0$ configuration (Fig. 1 of Ref. [40]). Ergodicity is clearly violated in these extreme cases. We reproduce these results in Appendix B. However, we find that the diagonal $\alpha = 1$ seems not to suffer from this problem (see Fig. 17 of Appendix B).

To substantiate our claim, we consider instead the 2×2 square lattice Hubbard model (four sites) and repeat the exercise that was done for the 4×4 case of Ref. [40]. The added benefit here is that we can compare directly to exact solutions obtained via direct diagonalization. We do exactly this by considering correlators in momentum space,

$$C_{\pm}(\mathbf{k}, \tau) = \frac{1}{2} \sum_x e^{ik \cdot x} C_{\pm}(x, \tau), \quad (81)$$

where the sum is over unit cell locations (each unit cell containing one *A* site and one *B* site) and $C_{\pm}(x, \tau)$ is given by Eq. (80) but now with explicit unit cell location *x* in its

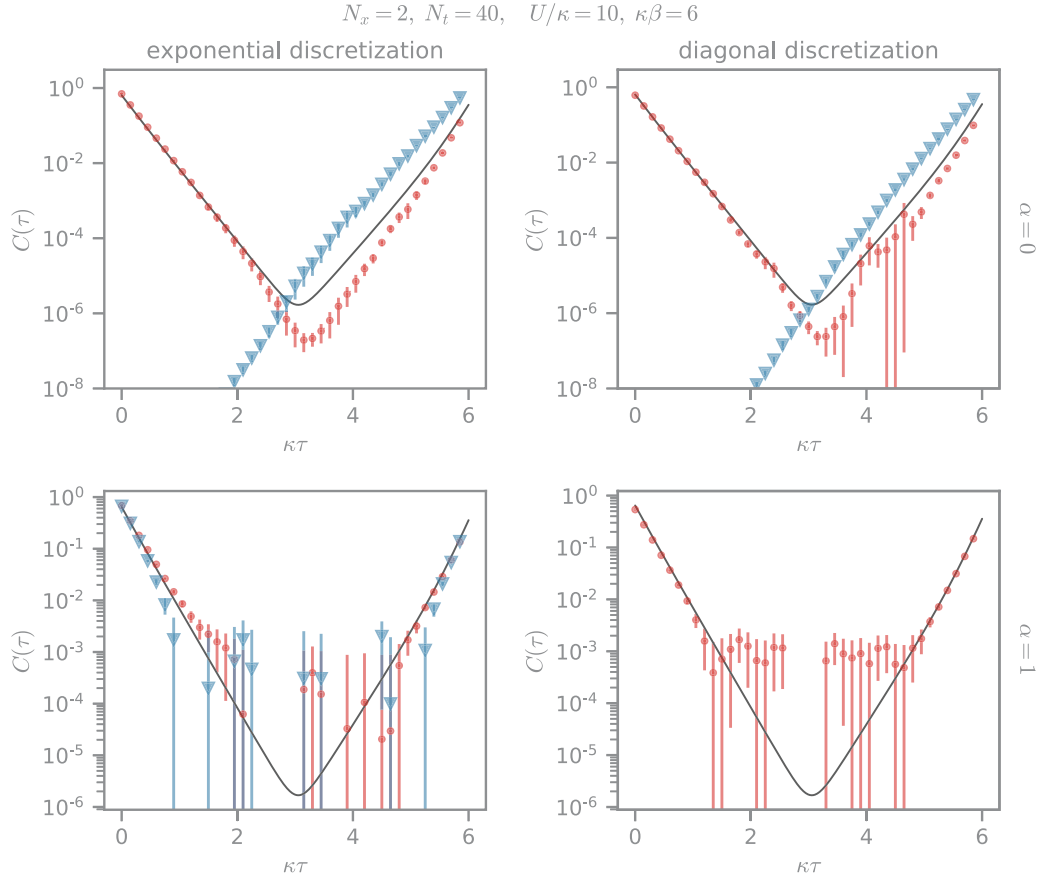


FIG. 8. Correlators from Eq. (80) color coded to match the ensembles shown in Fig. 7. Points where $C(\tau) \leq 0$ are not shown because of the logarithmic scale. The black lines show exact results from direct diagonalization of the two-site Hubbard model. For clarity, the figure shows only C_+ , C_- behaves in the same way.

argument. As there are two allowed momenta for this system, $ak = (0, 0)$ and $(\pi/2, \pi/2)$, there are in principle a total of four possible correlators. However, at $ak = (\pi/2, \pi/2)$, one has $C_+ = C_-$, and thus we have only three distinct correlators.

In Fig. 9, we show these three correlators for the four different discretization schemes and bases. For both $\alpha = 0$ cases, we start the HMC evolution from a $\det M[\phi] < 0$ configuration while for the $\alpha = 1$, exponential discretization the evolution starts at a configuration such that $0 > e^{-i\Phi/2} \det M[\phi] \in \mathbb{R}$ in accordance with Eq. (51). For the remaining diagonal $\alpha = 1$ case, no such criterion can be formulated since the determinant is complex and does not factorize as in the exponential case. In all cases, we use a very precise MD integrator. It is clear from the disagreement with the exact result that the HMC trajectories are *not* properly sampling all the important regions of configuration space. It is worth noting that not all correlators are affected by the lack of ergodicity in the same way. For the exponential $\alpha = 1$ case, the higher-energy correlators (green and blue) are very precise, only the ground-state correlator shows a strong deviation.

When we consider the histograms of the MC histories of $\det M$ for the $\alpha = 0$ and exponential $\alpha = 1$ cases, as shown in Fig. 10, we find that $\det M$ is confined to the negative region only. Ergodicity is indeed violated. The lone exception is the diagonal $\alpha = 1$ case whose correlators agree very well with the exact result (bottom right panel of Fig. 9). The corre-

sponding histogram of the MC history is shown in the bottom right panel of Figure 10. In this case, since $\det M$ is complex, the histogram is shown as a density on the complex plane. In this case, the zero of $\det M$ can be easily circumnavigated and there does not exist an ergodicity problem.

IV. OVERCOMING ERGODICITY ISSUES

Reference [40] already showed that one may avoid ergodicity issues by complexifying the auxilliary field (taking an intermediate value of α). In this section, we will examine a variety of other solutions.

A. Coarse molecular dynamics integration

When the molecular dynamics integrator is not very precise, the Markov chain can hop over the barriers that separate neighboring basins in configuration space in the particle/hole basis. This, for all practical purposes, avoids the ergodicity issues as long as the integrator takes sufficiently coarse steps. We have found that targeting an acceptance rate of around 70% allows the integrator to readily explore the areas in field space that would be separated by an impenetrable barrier in the case of a very precise integrator.

We emphasize that the numerical “errors” introduced by using an imprecise molecular dynamics integrator do not

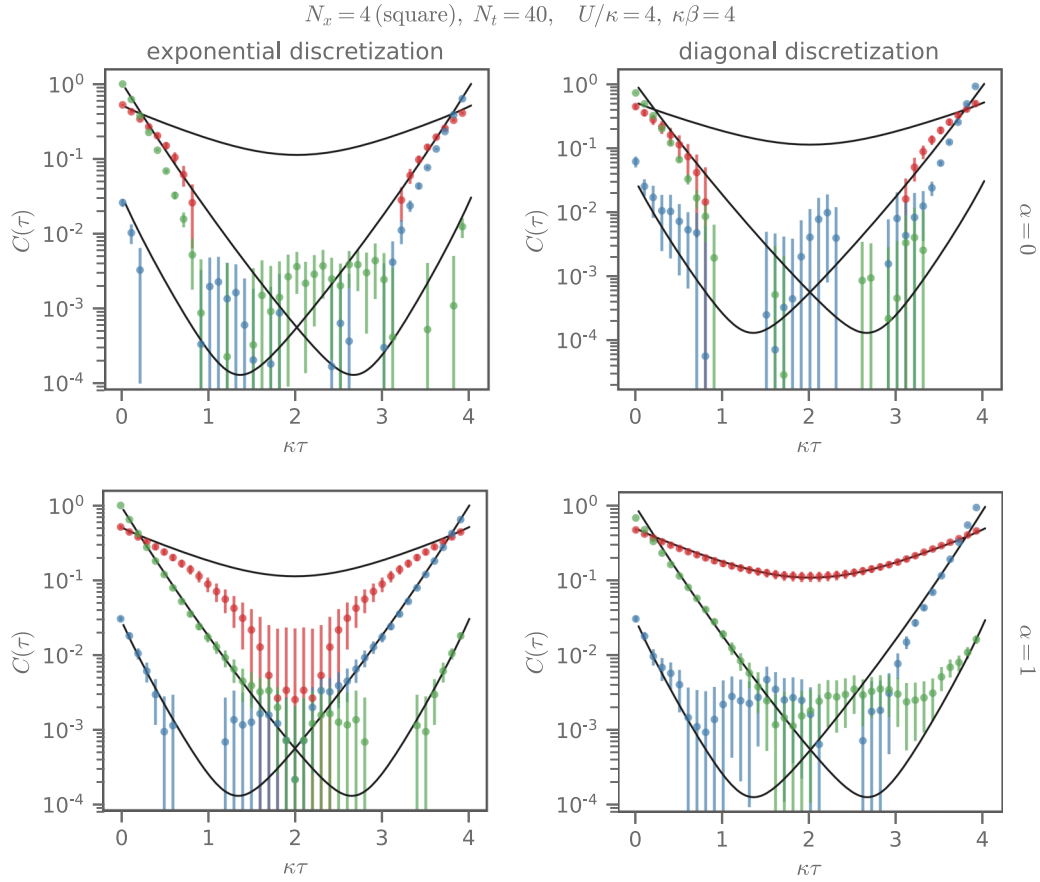


FIG. 9. Dependence of correlators on initial HMC configuration. The red correlators are calculated at $ak = (\pi/2, \pi/2)$ which for this case means that $C_+ = C_-$. The green and blue correlators (lower, off center minima) are calculated at $ak = (0, 0)$. For $\alpha = 0$ the initial configuration was chosen such that $\det M[\phi] < 0$. For $\alpha = 1$, exponential discretization the starting configuration satisfies $0 > e^{-i\Phi/2} \det M[\phi] \in \mathbb{R}$. For $\alpha = 1$, diagonal no particular starting criterion was set. All plots show data from 2400 trajectories. The starting criteria stayed fulfilled for all of those configurations.

invalidate the stochastic algorithm. In other words, HMC with a coarse integrator is still a valid proposal generator from the perspective of the Metropolis-Hastings algorithm. As long as the accept/reject step is maintained and the integration is reversible, the algorithm still faithfully samples the distribution dictated by the action in the path integral.

In the spin basis, the basins of Φ for the one-site problem grow farther apart when increasing $U\beta$. Between the two modes the probability is never zero, but is nevertheless exponentially small. Unless an *extremely* coarse integrator with a long trajectory length is used, it seems unlikely that molecular dynamics can cross such a wide exponentially small region.

With a fixed molecular dynamics time between trajectories, there is a limit to how coarse the integration may be made. With either a very or moderately coarse integrator combined with a long molecular dynamics time between configurations, we expect the acceptance rate to plummet. There thus may be an in-practice problem of rejecting too frequently, not evolving the configuration enough, or an explosion of computational cost.

In Fig. 11, we see that coarse integration allows the $\alpha = 1$ case to jump over the exact zeros, whereas the fine integration

was trapped in the initial mode, as seen in Fig. 4. In contrast, we can see that coarse integration does not alleviate the in-practice problem seen in the $\alpha = 0$ case because the modes are so widely separated.

A similar improvement can be observed in two-site problem when using a coarse integrator with $\approx 60\%$ acceptance rate. Figure 12 shows that configurations in this case are more evenly distributed for $\alpha = 1$, especially in the exponential discretization; compare with Fig. 7. For $\alpha = 0$, however, there is no change compared to the fine integrator as expected.

The correlators in Fig. 13 show no significant improvement over those in Fig. 8 for a fine integrator. It is however no longer possible to construct an ensemble that is stuck in a specific region of configuration space and leads to a systematically biased correlator in the exponential $\alpha = 1$ case, cf. Figs. 7 and 8. Noise in the systematically wrong medium τ range has increased for $\alpha = 0$, particularly strongly in the diagonal discretization.

Once many sites are included, it may be infeasible to rely on a coarse integrator for crossing zeros, so it is worthwhile to consider additional techniques to aid ergodic exploration of the configuration space [40,48].

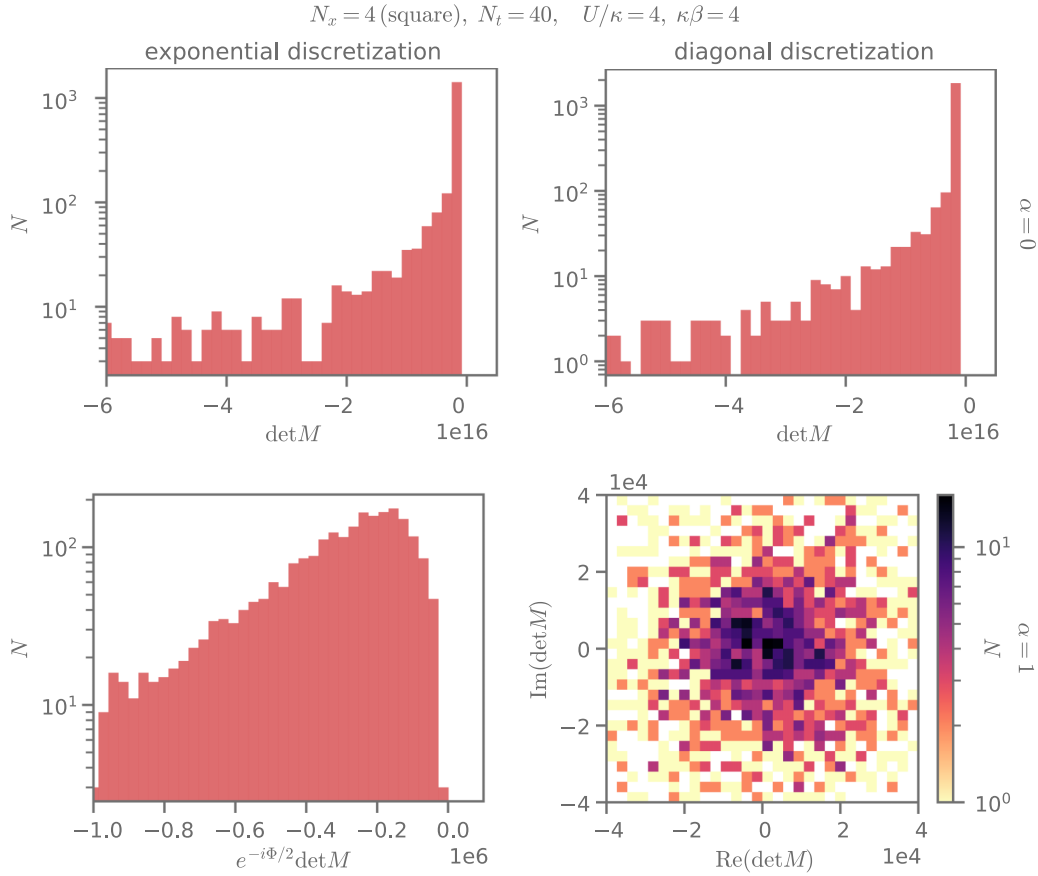


FIG. 10. Histograms of $\det M$ for the $\alpha = 0$ basis (top row) and diagonal $\alpha = 1$ case (bottom right), and $e^{-i\Phi/2} \det M[\phi]$ for the exponential $\alpha = 1$ case (bottom left) for the ensembles used for Fig. 9. Except for the diagonal $\alpha = 1$ case, $\det M$ is complex and the domain of the histogram is the complex plane.

B. Symmetrization

For the one-site problem in the spin basis, the in-practice ergodicity problem arises entirely because of the large separation of the modes in Φ in (72) and a possible solution immediately presents itself. If the configuration ϕ corresponds to Φ , the field configuration $-\phi$ corresponds to $-\Phi$.

Then, we can perform Monte Carlo on equivalence classes of ϕ —for each step in the Markov chain, we receive two (entirely correlated) field configurations, ϕ and $-\phi$, perform measurements on both configurations and average the result. Since their weights are exactly equal, we know they appear equally often in the original method of simply sampling individual field configurations—we simply ensure the different Φ lobes are sampled equally without the need to cross over the wide, exponentially small barrier.

This approach is not restricted to the one-site problem. Recall that, as discussed in Sec. IID 1, $W[\phi] = W[-\phi]$ independent of basis and α at half-filling on a bipartite lattice.

In Fig. 11, the $\alpha = 0$ case has coarse HMC integration that is trapped in one of the two modes. The purple correlator in the top-right panel is the average of the correlation function measured on ϕ and $-\phi$ for each configuration. The improvement is obvious.

Looking at the basins and boundaries in the lower-left panels of Figs. 5 and 7 makes it clear that this symmetrization

is not a complete solution for the exponential $\alpha = 1$ case. For example, if HMC winds up in a basin that does not include $\phi = 0$, no symmetrization will ever get it there. For the $\alpha = 0$ cases in those figures, it is clear that this symmetrization still fails to join different regions of nonzero probability.

Generalizing from just the charge conjugation operation $\phi \rightarrow -\phi$, we can also immediately symmetrize in terms of the other operational symmetries discussed in Sec. IID: temporal shifts, time reversal, and the spatial symmetries of the lattice. Put another way, if observables are formulated that are invariant under these operations, the ergodicity problem is reduced from the full configuration space to only the configuration space with these symmetries modded out.

This symmetrization may be performed stochastically, as well. For example, one often finds that in lattice QCD literature the chiral condensate is evaluated on stochastic noise sources or that correlation functions are measured from sources on different randomly chosen sites in the lattice for each field configuration.

Unfortunately, this is not a cure-all. Since the determinant is invariant under many of these symmetries, it cannot repair the in-principle problem we might encounter in, for example, the exponential $\alpha = 0$ many-site problem where we know, from (47), a formal problem that requires flipping the determinant's sign can arise.

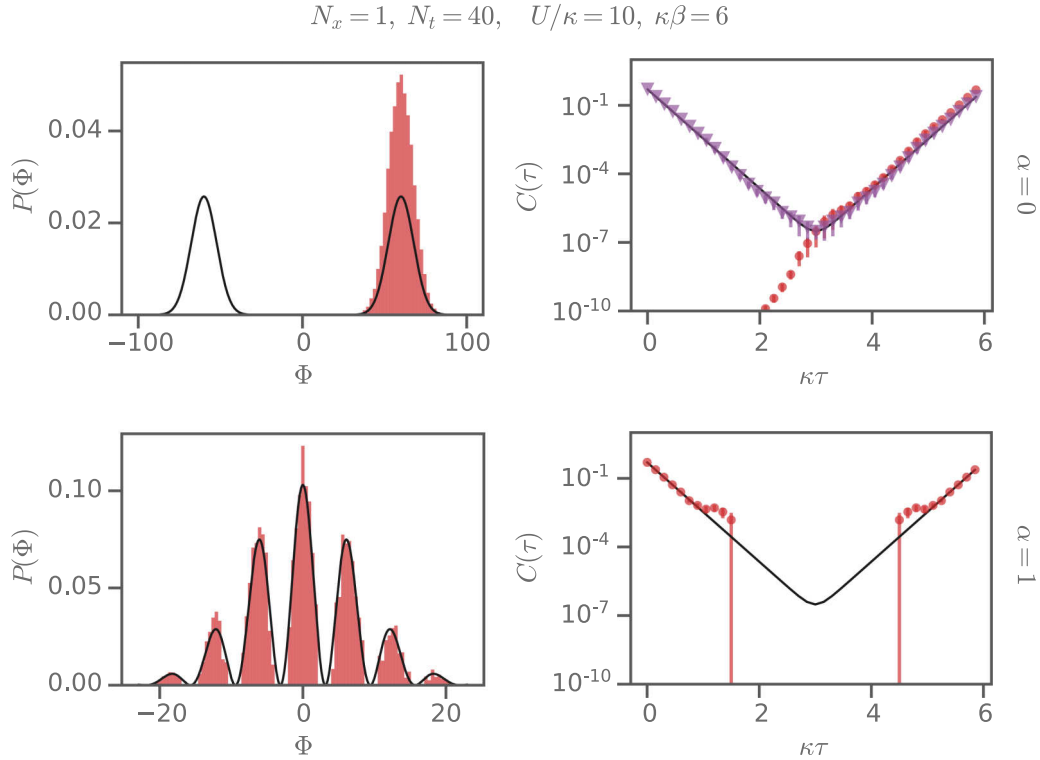


FIG. 11. Same as Fig. 4 but with coarser integration with an acceptance rate of about 60%. For $\alpha = 0$, Monte Carlo evolution was started in the right hand lobe and remained stuck there. Both ensembles consist of 50k thermalized configurations. The purple correlator is $C_{\nabla}(\phi) = (C_{\bullet}(\phi) + C_{\bullet}(-\phi))/2$, the result of the symmetrization discussed in Sec. IV B.

C. Large jumps

Rather than restricting Markov chain updates to HMC only, one can use a mix of proposal generators followed by the Metropolis accept/reject step. Those proposal machines must be statistically balanced so that the first criterion discussed in Sec. II A is fulfilled. In addition, selecting which proposer to use based on the current field configuration is not allowed. One approach is that of tempered transitions [53,54] which have found recent application in, for example, the study of the 0 + 1-dimensional Thirring model [55].

Another approach is to interleave large mode jumps [56] with normal HMC. By making a large change to the field configuration that is not produced by integrating equations of motion, the barriers that repel HMC trajectories, or the uncrossable valleys separating modes, can be bypassed. However, as shown by the inefficacy of simple random updates, it can be a challenge to make a proposal with a large change to the field configuration such that the new configuration contributes meaningfully to the partition function and the proposal is thus likely to be accepted.

By taking advantage of some of the features of the problem, we can propose large jumps to new regions in configuration space where the weight of the new field configuration is non-negligible. In the general case, it is hard to analytically extract features of the probability weight function. In the previous section, we saw that we could use symmetries to better cover the configuration space.

In fact, the symmetries that emerge in the strong or weak-coupling limits can also be used to generate proposals.

However, while a true symmetry operation is guaranteed to produce an accepted configuration, these other operations need not be accepted by the Metropolis step. The symmetries that emerge in the strong and weak-coupling regime were reviewed in Sec. II E.

It may be possible to craft proposals based on knowledge of the probability distribution in the weak-coupling limit. However, the failure of perturbation theory for problems of interest suggests these proposals will hardly ever be accepted in a calculation. We therefore leave the detailed description of these proposals to other work and focus on the approximate factorization of (55).

In the strong-coupling limit, we can flip the sign of ϕ on a thread of spatial sites, temporally shift or time-reverse threads individually, and arbitrarily permute the spatial threads rather than remain restricted to symmetries of the lattice. Proposals adjusting ϕ according to those manipulations are likely to be accepted if the system of interest is sufficiently close to the appropriate limit.

1. Spin basis

As an example, in the strong-coupling limit we can take advantage of our knowledge of the probability distribution for the one-site field configurations. For convenience, we reproduce (72) here:

$$W_1[\phi] = \frac{e^{-\frac{\phi^2}{2U\beta} - \frac{U\beta}{4}} \cosh^2\left(\frac{\phi}{2}\right)}{\sqrt{2\pi U\beta} \cosh\left(\frac{U\beta}{4}\right)} \quad (\alpha = 0), \quad (82)$$

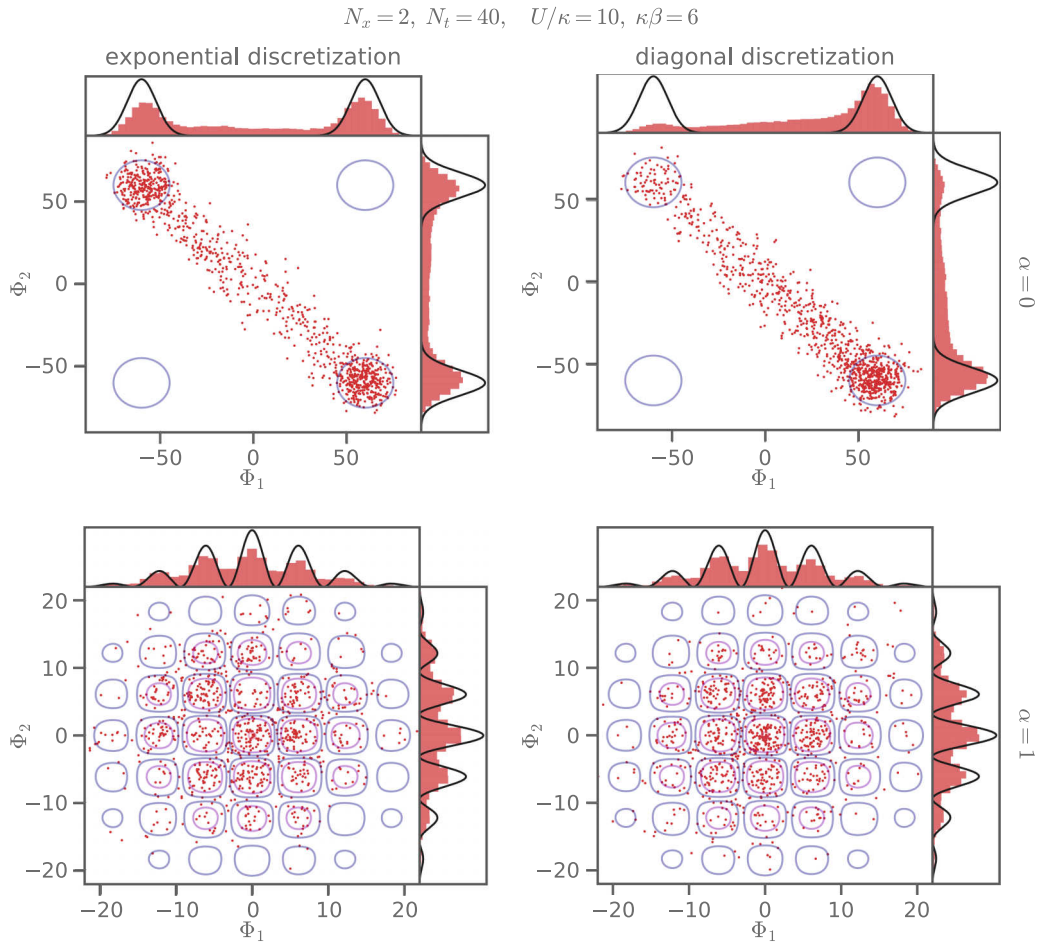


FIG. 12. HMC history of Φ_x for the two-site problem with 100 k trajectories and a coarse integrator. Only every 100th trajectory of each ensemble is shown. The black lines in the marginal distributions for Φ_1 and Φ_2 are the exact *one-site* distributions which are recovered in the strong-coupling limit. The contours show the product of the one-site distributions and are *not* exact results for this case.

where we attach a 1 subscript to emphasize it is for the one-site problem.

The one-site probability distribution is symmetric under $\Phi \rightarrow -\Phi$. In the strong-coupling limit, the probability distribution for a problem with N_x spatial sites is simply the product of the one-site distribution for each site. The probability is concentrated on the corners of an N_x -dimensional hypercube whose 2^{N_x} vertices have value $\Phi_x \simeq \pm U\beta$.

That distribution is invariant if we send $\Phi \rightarrow -\Phi$ independently for each single spatial site across all time,

$$\Phi_x \rightarrow \text{sign}_x \Phi_x, \quad (83)$$

where $\Phi_x = \sum_t \phi_{x,t}$, the auxiliary field summed across all time but only on a single site, as in (31) (and so $\Phi = \sum_x \Phi_x$). We can pick any subset of the spatial lattice, and negate all the auxiliary fields for all time for those sites.

For the one-site problem with $\alpha = 0$ in the strong-coupling limit, this operation allows us to propose a new configuration with the same weight but in another mode that will always be accepted. Away from strong coupling the guarantee of equal weight no longer holds as the factorization of the path integral fails. Assuming the probability weight does not deform too much, we can visit all the modes with proposals of this type.

Regardless of $U\beta$, this eliminates the in-practice sampling problem one might encounter due to widely separated modes, but does not resolve the possible formal problems.

We demonstrate the success of this method in Figs. 14 and 15. They show results from a run using a coarse integrator with an acceptance rate of $\approx 60\%$ augmented by random sign flips based on (83) every 100 trajectories. The sign is chosen for each spatial lattice site separately and with equal probability for either sign. Evidently, these jumps allow sampling the far apart lobes for $\alpha = 0$ which is not possible with just the fine or coarse integrator (compare to Figs. 7 and 12).

The correlators in Fig. 15 are similar to those in Fig. 8. The most notable difference is that for $\alpha = 0$ the minimum is now in the correct place whereas it was shifted to the right for the fine integrator without sign flips. Another unknown systematic remains however.

2. Particle-hole basis

For the one-site problem with $\alpha = 1$ in the strong-coupling limit, the sign-flipping operation in (83) yields a proposed configuration in a different domain from where HMC might be currently stuck. However, as mentioned in Sec. IV B, this operation helps to symmetrize the distribution around zero,

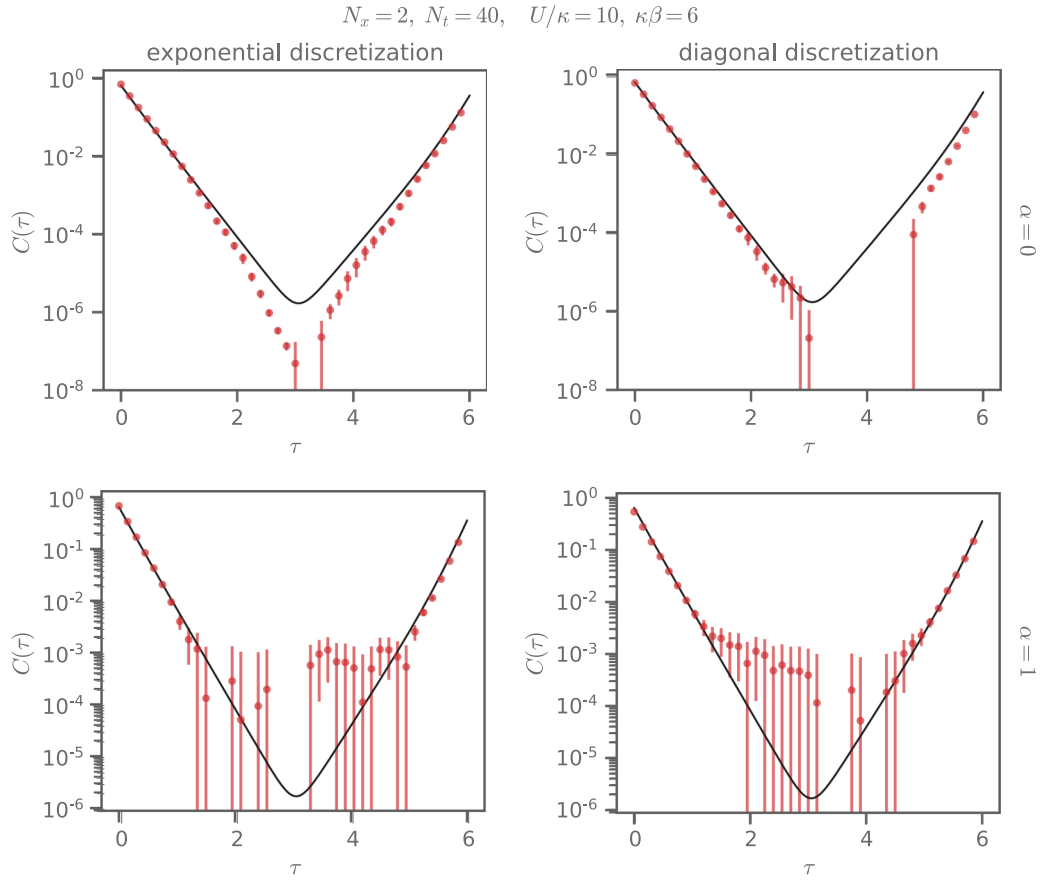


FIG. 13. Correlators from Eq. (80) for the ensembles shown in Fig. 12. Points where $C(\tau) \leq 0$ are not shown because of the logarithmic scale. The black lines show exact results from direct diagonalization of the two-site Hubbard model. For clarity, the figure shows only C_+ , C_- behaves in the same way.

but it does not allow us to access all the modes and thus does not offer a complete solution.

We nonetheless use those sign flips in Figs. 14 and 15. In this simple case this operation (paired with a coarse MD integrator) is sufficient to sample the relevant region of configuration space properly. This has no discernible impact on the correlators however which is likely caused by the large noise for medium τ .

When $\alpha = 1$ the one-site result is (73)

$$W_1[i\phi] = \frac{e^{-\frac{\Phi_x^2}{2U\beta} + \frac{U\beta}{4}} \cos^2\left(\frac{\Phi}{2}\right)}{\sqrt{2\pi U\beta} \cosh\left(\frac{U\beta}{4}\right)} \quad (\alpha = 1). \quad (84)$$

The full probability distribution in the strong-coupling limit is simply the product of this distribution for each site, and is concentrated around the origin $\Phi_x = 0$.

In this case, we can also propose a new configuration by increasing or decreasing all the auxiliary fields on a single spatial thread by $2\pi/N_t$,

$$\phi_{x,t} \rightarrow \phi_{x,t} \pm \frac{2\pi}{N_t} \delta_{x,x_0} \quad (85)$$

which changes Φ_{x_0} by 2π , putting it on the other side of an exact zero and thus into a different mode. Such a proposed

update will be accepted according to the ratio of probabilities⁸

$$\begin{aligned} & \frac{W[i(\phi_{x,t} \pm \frac{2\pi}{N_t} \delta_{x,x_0})]}{W[i\phi]} \\ &= (\text{ratios of determinants that cancel at strong coupling}) \\ & \times e^{-\frac{2\pi}{U\beta}(\pi \pm \Phi_{x_0})}, \end{aligned} \quad (86)$$

meaning that if the proposal drives Φ_{x_0} towards zero it will always be accepted and will occasionally be accepted if it drives Φ_{x_0} away from zero. This update will allow us to access all the modes of Φ_{x_0} without encumbrance from exact zeros in the probability distribution. In the strong-coupling limit, we may make this proposed update to each thread of spatial sites independently. Away from strong coupling, this proposal may nevertheless prove beneficial, assuming the probability weight doesn't deform too much. In fact, the 2π change in Φ_{x_0} need not be evenly spread across all the time slices, at the cost of reducing the likelihood according to a change in (70).

In a similar spirit, we can use the field-shift transformation (27), shifting each time slice by its own θ_t , and as long as the shifts sum to 0 modulo 2π , according to (30), we can make

⁸This identity holds in general and not only in the strong-coupling limit.

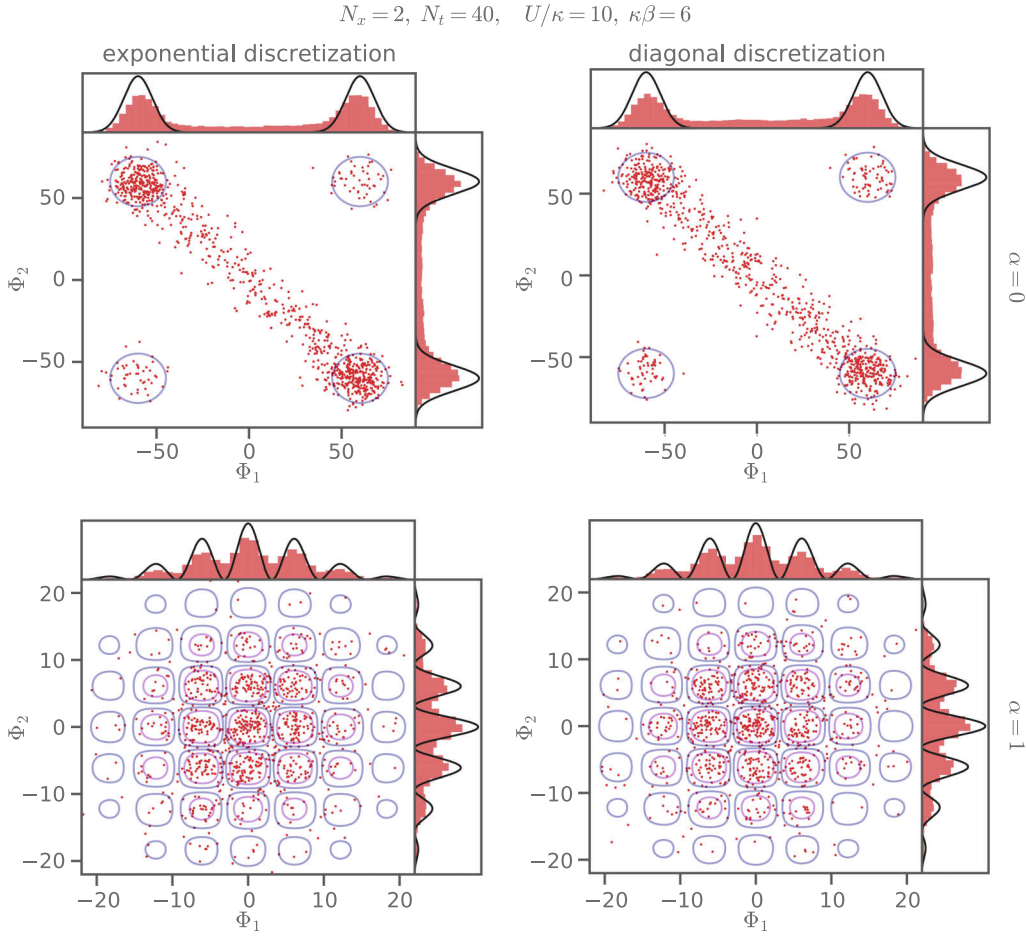


FIG. 14. HMC history of $\Phi_x = \sum_t \phi_{xt}$ for the two-site problem with 100 k trajectories and a sign flip every 100 trajectories according to (83). Only every 100th trajectory of each ensemble is shown. The black lines in the marginal distributions for Φ_1 and Φ_2 are the exact *one-site* distributions which are recovered in the strong-coupling limit. The contours show the product of the one-site distributions and are *not* exact results for this case.

the cancellation of the determinant exact,

$$\begin{aligned} & \frac{W[i(\phi_{x,t} + \theta_t)]}{W[i\phi]} \\ &= (\text{ratios of determinants that exactly cancel}) \\ & \times e^{-\frac{1}{2U} \sum_t 2\phi_t \theta_t + N_x \theta_t^2}. \end{aligned} \quad (87)$$

If we pick a constant $\theta_t = \pm 2\pi/N_t$ we can simplify further,

$$\frac{W[i(\phi_{x,t} \pm 2\pi/N_t)]}{W[i\phi]} = e^{-\frac{2\pi}{U\beta} (\pi N_x \pm \Phi)}. \quad (88)$$

The average acceptance rate is analyzed in Appendix C. In Fig. 3, such a coordinated jump is a jump in the diagonal Φ direction, orthogonal to the lines of zero determinant.

Most generally, we know that when $\alpha = 1$ the determinants are 2π periodic in each field variable. For a $\pm 2\pi$ change on site x_0 at time t_0 the ratio of the weights is simply given by the ratio of the Gaussian factors,

$$\begin{aligned} & \frac{W[i(\phi_{x,t} \pm 2\pi \delta_{x,x_0} \delta_{t,t_0})]}{W[i\phi]} \\ &= (\text{ratios of determinants that exactly cancel}) \\ & \times e^{-\frac{2\pi}{U} (\pi \pm \phi_{x_0,t_0})}. \end{aligned} \quad (89)$$

We can independently propose and accept or reject a change by $\pm 2\pi$ on each site, knowing the determinant will remain invariant. Again, if this drives the auxiliary field on a site towards zero it will be accepted and it will occasionally be accepted when the field is driven away from zero. This proposal is entirely local and extremely speedy, not requiring an evaluation of the determinant. Moreover, in the exponential $\alpha = 1$ case, each accepted change hops over a zero manifold, as can be seen from (53). This alleviates the formal ergodicity problem arising from conjugate reciprocity. In the diagonal $\alpha = 1$ case, while there is no formal ergodicity problem, interleaving such a proposal can help in practice, especially as barriers are raised as one approaches the continuum limit.

Reconsidering Fig. 3, it is clear that an evenly distributed change of Φ by 2π is more likely to be accepted than a 2π shift in either auxiliary field variable alone, because a coordinated jump tends to keep you close to the middle of the Gaussian. As discussed in Appendix C, the acceptance rate of coordinated jumps remains finite in the continuum limit, while the acceptance rate for jumps in individual field components vanishes in the continuum limit. Whether such coordinated changes are guaranteed to cross every ergodicity barrier remains obscure to us, while it seems apparent that independent changes provide such assurance.

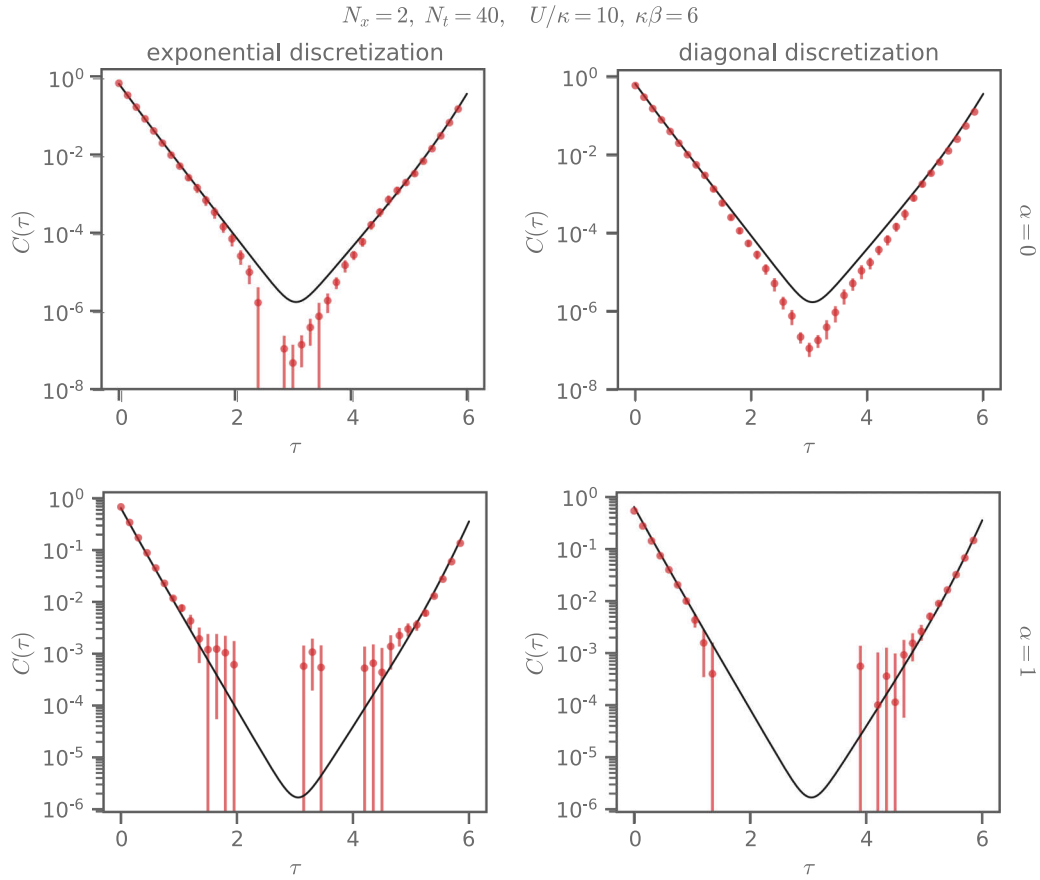


FIG. 15. Correlators from Eq. (80) from the ensembles with sign flips shown in Fig. 14. Points where $C(\tau) \leq 0$ are not shown because of the logarithmic scale. The black lines show exact results from direct diagonalization of the one-site Hubbard model. For clarity, the figure shows only C_+ , C_- behaves in the same way.

In Fig. 16, we show an example run in the two site system with the exponential discretization. The integrator is as fine as in Figs. 7 and 8 but every 100 trajectories two jumps are performed. First, $N_t N_x$ single site jumps by 2π are performed

on random sites and each is accepted or rejected according to (89). Then, all field variables are shifted by $\pm 2\pi/N_t$ in a coordinated jump and the change is accepted or rejected according to (88). The result shows a similar improvement

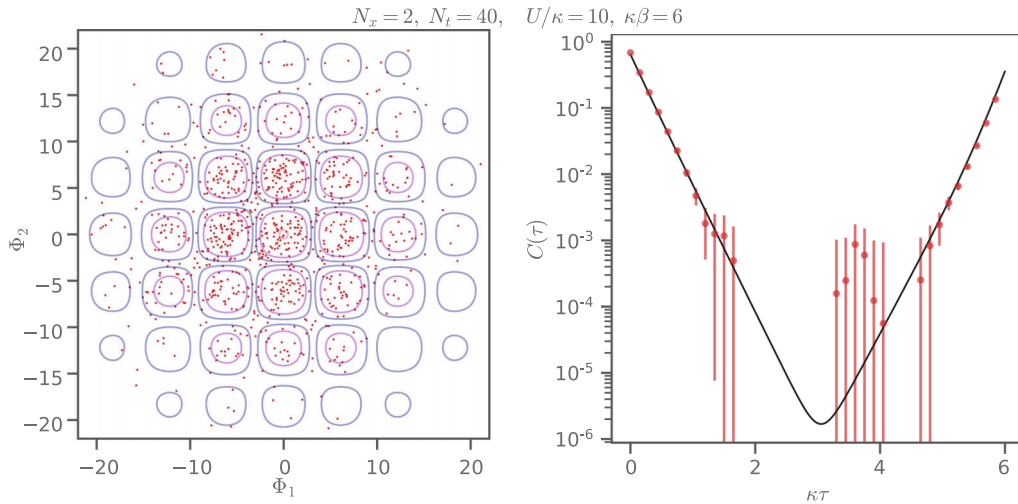


FIG. 16. HMC trajectories and correlator for the exponential $\alpha = 1$ case with the same parameters as Figs. 7 and 8, but large jumps as in Eqs. (88) and (89) every 100 trajectories. Only every 100 th trajectory is shown in the left panel and the ensemble consists of 100k trajectories in total.

to the sign flips in Figs. 14 and 15. The histogram shows marked improvement, showing no hint of trapping, and the correlation function is no longer incompatible with the exact result, especially visible around $\kappa\tau \sim 1.5 - 2$.

V. CONCLUSION

The issue of ergodicity is of great significance for the validity of any stochastic calculation which relies on an MCMC process. Throughout this work, we have emphasized that ergodicity (or the lack thereof) has a *formal* as well as a *practical* meaning. In the former case, regions in the configuration space exist which are *formally* disconnected due to appearance of boundaries of codimension 1, at which $\det M[\phi] = 0$. In the latter case, such boundaries do not exist, yet MCMC cannot sample all relevant regions of configuration space in a practical amount of computing time. In other words, thermalization and decorrelation would be exponentially slow. In such cases, we speak of an *in practice* ergodicity issue. We have investigated both situations, in terms of the Hubbard model in 2D. In particular, we have shown that the exponential discretization in the particle-hole basis exhibits a *formal* ergodicity problem arising from conjugate reciprocity, while the diagonal discretization in the particle-hole basis does not.⁹ Furthermore, for the case of $\alpha = 0$ (spin basis) at large $U\beta$, an *in practice* ergodicity problem also appears, as the field distributions fragment into widely separated multimodal lobes. HMC algorithms become trapped in one of these lobes and require an exponential simulation time to traverse to any other lobe, despite there being no boundaries where the fermion determinant changes sign.

We note that ergodicity issues have been encountered numerous times in MCMC simulations of lattice gauge theories. In cases where the problem is *formal*, it is often observed that it is sufficient (at least for small systems) to reduce the accuracy of the MD integrator so that the $\det M[\phi] = 0$ barriers can be traversed. We have observed such behavior for the $\alpha = 1$ exponential discretization which has $\det M[\phi] = 0$ boundaries, which nevertheless could be traversed with sufficiently coarse MD integration. However, this comes at the price of reducing the acceptance rate of the HMC algorithm, and for cases where U and β are very large, ergodicity can no longer be restored by such brute force methods. For the $\alpha = 0$ basis (both exponential and diagonal), the separation of the field distributions into widely spaced lobes at large $U\beta$ presents a more serious challenge for HMC, despite it being an *in practice* rather than *formal* issue. For such cases, we have proposed a new algorithm which takes advantage of the symmetries of the action and the fermion matrix M . Effectively, we augment the standard HMC algorithm with large, carefully crafted jumps between regions of large probability. The accept/reject step of the algorithm ensures that

each region is sampled with the correct relative probability. We have found that such jumps fix this particular type of *in practice* ergodicity problem.

Our studies were concentrated on histograms of the HS field and single particle correlators. We stress that this does not present a full proof that the algorithm is ergodic. It is, of course, still possible that there is a problem caused by a barrier that has escaped notice. It is therefore useful to carefully verify that observables match known results that might be calculated. If the algorithm is ergodic one expects a faithful ensemble and that every observable match its true value.

In the $\alpha = 0$ basis, we have observed the appearance of configurations with $\det M[\phi] \leq 0$ for lattices with 4 or more sites, as was also found in Ref. [40] for the 16-site problem. We observe that the frequency with which negative determinants occur increases dramatically with system size, which we have attributed to an accumulation of states with nearly zero energy. This is highly detrimental to the HMC algorithm and its evolution, as the frequency with which exceptional configurations are encountered also increases.

Our studies show how the *reduction* of symmetries removed formal ergodicity issues. Missing from this discussion are the implications of reduced symmetry for observables. The reduction of symmetries had little to no effect on the two-point correlators we have concentrated on (which give access to quasi-single-particle energies). We note that this is by no means a general statement. Still, a similar behavior for the spin-density wave (SDW) order parameter was found in Ref. [40]. On the other hand, in Ref. [58] the diagonal discretization led to chiral symmetry violations that impacted the spin-density wave order parameter, which can only be restored in the continuum limit. Thus the restoration of ergodicity in this case made the calculation of the SDW order parameter more difficult, and emphasizes the need for a robust continuum limit extrapolation. For additional evidence that the diagonal discretization compromises on exact symmetries in favor of better ergodic properties, the reader is pointed to Appendix D.

In the examples we have considered here, we have taken values of U and β which are deliberately chosen to be large (such as $U\beta \simeq 60$), in order to ensure that ergodicity issues are not unduly suppressed. However, we note that the $\alpha = 1$ diagonal discretization was found to suffer from neither *formal* nor *practical* ergodicity issues, even for extreme values of U and β . When comparisons with exact results are feasible (such as for one-site, two-site, and four-site square and hexagonal lattices), we have found that this discretization performed equally well as the other ones (after application of the remedies mentioned above). For cases with large U and β , the diagonal $\alpha = 1$ discretization was found to be superior, as it requires no modification to the HMC algorithm, nor monitoring for exceptional configurations. For the 16-site calculation, we were able to reproduce published results from the BSS algorithm and never encountered any exceptional configurations. For these reasons, we conclude that the diagonal discretization with $\alpha = 1$ is optimal for our purposes, even though it only recovers the exact chiral symmetry of the exponential discretization in the continuum limit.

⁹In Ref. [57], the authors use a fermion operator not considered here, where transport in time alternates between a linearized hopping operator and interaction through the auxiliary field. They claim that the simulations are free from ergodicity problems because the determinant is complex, in line with our detailed argument.

ACKNOWLEDGMENTS

We thank Stefan Krieg, Johann Ostmeier, Carsten Urbach, and Enrico Rinaldi for insightful discussions. We are indebted to Stefan Beyl for his assistance in comparing our results with the BSS formalism. C.K. gratefully acknowledges funding through the Alexander von Humboldt Foundation through a Feodor Lynen Research Fellowship. This work was done in part through financial support from the Deutsche Forschungsgemeinschaft (Sino-German CRC 110). The authors gratefully acknowledge the computing time granted through JARA-HPC on the supercomputer JURECA [59] at Forschungszentrum Jülich.

APPENDIX A: DETERMINANTS

1. Cyclic lower block bidiagonal matrices

In this section, we derive the central equations for the determinants of the fermion matrices. We look at matrices of the form

$$M_{x't',xt} = D_{x',x}\delta_{t',t} - (T_{t'})_{x',x}\mathcal{B}_{t'}\delta_{t',t+1}, \quad (\text{A1})$$

where

$$\mathcal{B}_t \equiv \begin{cases} +1, & 0 < t < N_t \\ -1, & t = 0 \end{cases} \quad (\text{A2})$$

encodes antiperiodic boundary conditions in time and is factored out to simplify representing $M = M^e$ or $M = M^d$. In time-major layout, M is a cyclic lower block bidiagonal $N_t \times N_t$ matrix with blocks of size $N_x \times N_x$. Written in matrix form this is

$$M = \begin{pmatrix} D & & & T_0 \\ -T_1 & D & & \\ & -T_2 & D & \\ & & \ddots & \ddots \\ & & & -T_{N_t-1} & D \end{pmatrix}. \quad (\text{A3})$$

We can compute the determinant of M by means of an LU decomposition in terms of the matrices D and T_t . This decomposition can be performed by hand thanks to the sparsity of M . We use the following ansatz which is an adaptation of the ansatz presented in Ref. [60]:

$$L = \begin{pmatrix} 1 & & & & \\ l_0 & 1 & & & \\ & l_1 & \ddots & & \\ & & \ddots & 1 & \\ & & & l_{n-3} & 1 \\ & & & & l_{n-2} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} d_0 & & & v_0 \\ & d_1 & & v_1 \\ & & d_2 & \vdots \\ & & & \ddots & v_{n-3} \\ & & & & d_{n-2} & v_{n-2} \\ & & & & & d_{n-1} \end{pmatrix}, \quad (\text{A4})$$

Multiplying out $LU = M$ and solving the straightforward recursive equations gives

$$d_i = D \quad \text{for} \quad 0 \leq i < N_t - 1, \quad (\text{A5})$$

$$d_{N_t-1} = D(1 + A), \quad (\text{A6})$$

$$v_i = DA_{0,i}, \quad (\text{A7})$$

$$l_i = -T_{i+1}D^{-1}. \quad (\text{A8})$$

Here we have used

$$A_{t,t'} \equiv D^{-1}T_{t'}D^{-1}T_{t-1} \cdots D^{-1}T_t, \quad (\text{A9})$$

$$A \equiv A_{0,N_t-1}. \quad (\text{A10})$$

The determinant can be computed from the determinants of the blocks on the diagonals using

$$\begin{aligned} \det M &= \det L \det U = \left(\prod_{i=0}^{N_t-1} \det 1 \right) \left(\prod_{i=0}^{N_t-1} \det d_i \right) \\ &= (\det D)^{N_t-1} \det D(1 + A) \\ &= (\det D)^{N_t} \det (1 + D^{-1}T_{N_t-1}D^{-1}T_{N_t-2} \cdots D^{-1}T_1D^{-1}T_0). \end{aligned} \quad (\text{A11})$$

This result can alternatively be written as

$$\det M = \det (T_{N_t-1}T_{N_t-2} \cdots T_0) \det (1 + T_0^{-1}DT_1^{-1}D \cdots T_{N_t-2}^{-1}DT_{N_t-1}^{-1}D) \quad (\text{A12})$$

$$= \det (T_{N_t-1}T_{N_t-2} \cdots T_0) \det (1 + A^{-1}), \quad (\text{A13})$$

which can be beneficial depending on the forms of D and T . Expression (A12) can be obtained from (A11) by factoring in one D and factoring out one T at a time.

2. Fermion determinants

The determinant of M^e (12) can be calculated from (A11) by inserting

$$D \mapsto \mathbb{1}, \quad T_{t'} \mapsto e^h F_{t'}, \quad (\text{A14})$$

with

$$F_{t'}[\phi]_{x',x} \equiv e^{\phi_{x(t'-1)}} \delta_{x',x}. \quad (\text{A15})$$

Thus

$$\det M^e[\phi, h] = \det (1 + e^h F_{N_t-1}[\phi] e^h F_{N_t-2}[\phi] \cdots e^h F_1[\phi] e^h F_0[\phi]) \quad (\text{A16})$$

$$= \det(1 + B[\phi, h]). \quad (\text{A17})$$

We can proceed in a similar way for the matrix in the diagonal discretization M^d (14). This time we insert the following into the alternate form (A12):

$$D \mapsto \mathbb{1} - h, \quad T_{t'} \mapsto F_{t'}, \quad (\text{A18})$$

with the same F as before (A15). Thus

$$\det M^d[\phi, h] = e^\Phi \det (1 + F_0^{-1}[\phi](\mathbb{1} - h)F_1^{-1}[\phi](\mathbb{1} - h) \cdots F_{N_t-2}^{-1}[\phi](\mathbb{1} - h)F_{N_t-1}^{-1}[\phi](\mathbb{1} - h)) \quad (\text{A19})$$

$$= e^\Phi \det(1 + A^{-1}[\phi, h]). \quad (\text{A20})$$

At this point, a note on numerical stability is in order. The calculation of determinants of dense $N_x \times N_x$ matrices via a standard LU decomposition should not be a problem. Especially so since we are ultimately interested in $\ln \det M$ which is more stable for large matrices. The spatial matrices A and B are however constructed from a product of $2N_t$ matrices. Such a product can incur large round-off errors when the involved matrices have significantly different scales. Both $\mathbb{1} - h$ and e^h have elements $\lesssim 1$ and present no problem.

For the spin basis, the matrices $F \sim e^\phi$ on the other hand have elements of widely varying size which can be significantly larger than one. We have observed instabilities of the action and therefore HMC evolution as well as the solver for systems of linear equations based on Eqs. (A16) and (A19).

For the particle/hole basis, the problematic matrices are replaced by $F \sim e^{i\phi}$ whose elements are on the unit circle. All matrices that contribute to A and B are therefore of the same order and floating point errors are greatly reduced. Our tests show that both action and solver are precise to 12 or more digits for lattices of size $N_x \approx 100$ and greater.

We use algorithms based on (A16) and (A19) for $\alpha = 1$. For $\alpha = 0$, we use dense LU based algorithms for the full space-time matrices M . Other algorithms have been used to solve this problem, including a stabilization of this algorithm using singular value decompositions [61] and a different solver based on Schur complements [62].

APPENDIX B: SIXTEEN-SITE PROBLEM

For completeness we reproduce Fig. 1 of Ref. [40] which outlines the susceptibility of correlators to the starting configuration in case HMC is not ergodic. The correlators examined

in this case are the equal-site correlators,

$$C(\tau) = \frac{1}{N_x} \sum_i \langle C_{ii}(\tau) \rangle,$$

where the sum is over all lattice sites. We show three different HMC runs in Fig. 17 as well as the result obtained from the Blankenbecler-Scalapino-Sugar method (BSS) taken from Ref. [40]. The latter is claimed to be free of any ergodicity problems and thus provides a benchmark for our HMC results.

One HMC run is in the exponential discretization in the spin basis and starts with a configuration ϕ_0 such that

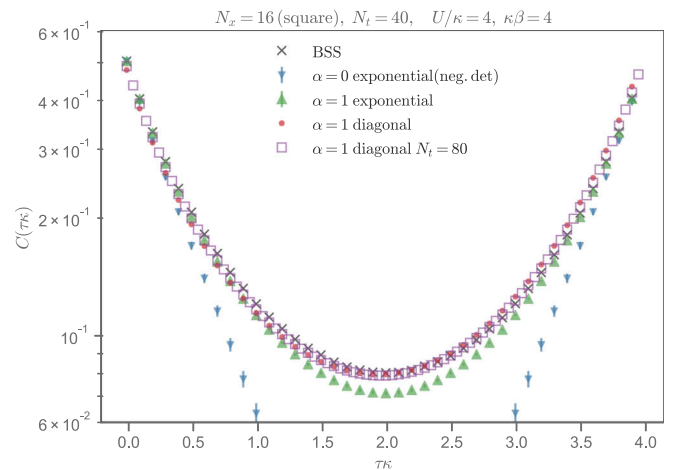


FIG. 17. Comparison of correlators on a 4×4 square lattice. The data points for BSS are taken from Ref. [40]. The exponential $\alpha = 0$ HMC run has negative $\det M[\phi]$ for all configurations. Started from a positive determinant this case produces results consistent with BSS and $\alpha = 1$.

$\det M[\phi_0] < 0$. The MD integrator is very fine (acceptance rate $>98\%$) in this particular example, to ensure no barriers are accidentally jumped by coarse integration. This run clearly deviates from the BSS result and matches the (green) correlator shown in Fig. 1 of Ref. [40] (shown as black $\times$ s in Fig. 17). Because of the very fine MD integration, HMC gets trapped in a part of the integration domain where the determinant is negative. For a coarser integrator, HMC is able to traverse to $\det M > 0$ regions and in this case we find good agreement with the BSS points, though we do not show these results here.

The second HMC run is again in the exponential discretization but in the particle-hole basis, again with a very fine MD integration to ensure that the ergodicity barriers are impenetrable. While matching the BSS results at early times, in the middle of the temporal extent it differs markedly from the BSS result.

The third HMC run is in the diagonal discretization in the particle/hole basis. Since the determinant is complex in this case, there is no particular starting criterion to be chosen and we stress that we do not find any dependence on the initial configuration. The correlator agrees well with the BSS result at intermediate times; at early times there is a systematic discrepancy; recall that different discretizations need only agree in the continuum.

To emphasize this point, we also show a third HMC run in the diagonal discretization in the particle/hole basis, but with twice as many time slices. On each common time slice, the finer discretization is closer to the BSS result than the coarser.

Also note that at early times the trend towards the BSS result is upward, while at later time it is downward. Averaging the correlator and its time-reversed partner results in a correlator indistinguishable from the BSS result. In future work, we will detail methods for reducing discretization artifacts of correlator measurements. We emphasize that one *still must take a controlled continuum limit*, even if results are formally closer to the continuum.

APPENDIX C: JUMP ACCEPTANCE

In this section, we analyze the acceptance rate of the jumps proposed in Sec. IV C that take advantage of the periodicity of the determinant, a coordinated jump uniform across all sites

$$\phi_{x,t} \rightarrow \phi_{x,t} + \frac{2\pi}{N_t} \quad (C1)$$

and an individual jump

$$\phi_{x,t} \rightarrow \phi_{x,t} + 2\pi\delta_{x,x_0}\delta_{t,t_0} \quad (C2)$$

in the field component on site (x_0, t_0) .

For the coordinated proposal, the acceptance rate is given by (88)

$$\frac{W[i(\phi_{x,t} \pm 2\pi/N_t)]}{W[i\phi]} = e^{-\frac{2\pi}{\tilde{U}}(\pi N_x \pm \Phi)}. \quad (C3)$$

To calculate the average acceptance rate, we should average that weight change in such a way that Φ is generally representative of a field configuration. Were the action entirely controlled by the Gaussian part, ϕ would be distributed normally with mean 0 and variance $\sigma^2 = \tilde{U}$. Then, summing $N_x N_t$ such random variables, we expect Φ to be drawn from a zero-mean

Gaussian with variance $N_x N_t \tilde{U} = N_x U \beta$. So, we find that the acceptance rate is independent of N_t , but decreases markedly with spatial volume.

When implementing the ergodicity jumps of individual field components it is natural to ask: should each site's jump be accepted or rejected independently? Or should they be amalgamated into one proposal and considered as a whole? Here we will argue that it is better to consider each decision independently.

The change in the weight (89) for a 2π shift on site at location x_0 and time t_0 is given by

$$\begin{aligned} & \frac{W[i(\phi_{x,t} \pm 2\pi\delta_{x,x_0}\delta_{t,t_0})]}{W[i\phi]} \\ &= (\text{ratios of determinants that exactly cancel}) \\ & \times \exp\left(-\frac{2\pi}{\tilde{U}}(\pi \pm \phi_{x_0,t_0})\right), \end{aligned} \quad (C4)$$

while if we did many 2π jumps, the change in the action would be summed, so that the acceptance would be controlled by

$$\exp\left(-\frac{2\pi^2}{\tilde{U}} \sum_x \left(1 \pm_x \frac{\phi}{\pi}\right)\right), \quad (C5)$$

where x summarizes the space and time coordinates and the subscript on $\pm_x$ is there to remind us that the sign choice is x -dependent and can be zero for some sites.

To calculate the average acceptance rate we should again average that weight change in such a way supposes we start in a very likely configuration. As above, we assume each ϕ is a normal, random variable with variance $\tilde{U}$.

If, in a single proposal, each site is given a probability p of undergoing a jump, on average we have to sum the action changes from $pN_x N_t$ jumps. We would like to find the best possible p , deciding whether we should completely amalgamate the proposals into one, consider each individually, or something in between. We can estimate

$$\sum_x \left(1 \pm_x \frac{\phi}{\pi}\right) = pN_x N_t + \frac{\varphi}{\pi}, \quad (C6)$$

where φ is drawn from a zero-mean Gaussian with variance $pN_x N_t \tilde{U}$. On average φ is 0 and we get an acceptance probability of

$$\exp\left(-\frac{2\pi^2}{U\beta/N_t} pN_x N_t\right). \quad (C7)$$

That is, the acceptance rate is reduced with increasing $pN_x N_t$. The best choice is to set $p = 1/N_x N_t$ —that is, to accept or reject a jump in one auxiliary field at a time. Still, in that case, the acceptance rate exponentially decreases with N_t , so that these jumps are much less important as one takes the continuum limit, though they increase in importance with $U\beta$. Moreover, when making proposals of jumps in individual components, there is no suppression with the spatial volume N_x , unlike the case of coordinated jumps.

APPENDIX D: OTHER OBSERVABLES

In this Appendix, we consider other observables. In particular, we consider correlations between local bilinear operators.

Letting x denote a spatial index, we define the bilinear operators

$$\begin{aligned} S_x^0 &= \frac{1}{2}[a_x a_x^\dagger - b_x b_x^\dagger + 1], & S_x^1 &= \frac{1}{2}(-1)^x [b_x^\dagger a_x^\dagger + a_x b_x], \\ S_x^2 &= \frac{i}{2}(-1)^x [b_x^\dagger a_x^\dagger - a_x b_x], & S_x^3 &= \frac{1}{2}[a_x a_x^\dagger + b_x b_x^\dagger - 1], \end{aligned} \quad (\text{D1})$$

raising and lowering operators,

$$S_x^\pm = S_x^1 \pm i S_x^2 \quad (\text{D2})$$

and the charge-density operator $\rho = 1 - 2S^0$ so that at half-filling, the expectation value $\langle \rho \rangle$ vanishes. The correlation functions

$$C_{xy}^{ij}(\tau) = \frac{1}{N_t} \sum_t \langle S_{x,t+\tau}^i S_{y,t}^j \rangle \quad (\text{D3})$$

carry two spatial indices x and y which can be projected to definite irreps of the lattice. The two irreps we will consider here are the spatially uniform irrep and its doubler, the alternating irrep. Let A and B be the two sublattices and $\hat{A}$ and $\hat{B}$ be idempotent orthogonal projection operators that sum to the identity. Denoting the uniform irrep $+$ and the alternating, staggered irrep $-$, projection to those irreps requires performing the sums

$$+: N_x^{-1/2} \left(\sum_{x \in A} + \sum_{x \in B} \right) = N_x^{-1/2} \sum_x (\hat{A} + \hat{B}), \quad (\text{D4})$$

$$-: N_x^{-1/2} \left(\sum_{x \in A} - \sum_{x \in B} \right) = N_x^{-1/2} \sum_x (\hat{A} - \hat{B}), \quad (\text{D5})$$

where A and B indicate the set of sites belonging to the two different sublattices. Once projected to definite irrep, the correlation functions are diagonal. We denote irrep-projected correlators with the $+$ and $-$ subscripts, rather than spatial subscripts.

We can study these correlators themselves or boil them down into order parameters. One possible charge density wave (CDW) order parameter is given by¹⁰

$$\text{CDW} = V^{-1} \lim_{\tau \rightarrow 0} C_{--}^{\rho\rho}(\tau), \quad (\text{D6})$$

where the ρ index indicates that we should use the charge operator, rather than one of the spin operators,¹¹ and we can also define the (extensive) antiferromagnetic structure factor [65]

$$S_{\text{AF}} = \lim_{\tau \rightarrow 0} C_{--}^{33}(\tau) \quad (\text{D7})$$

and the antiferromagnetic susceptibility [65]

$$\chi_{\text{AF}} = V^{-1} \lim_{\tau \rightarrow 0} C_{--}^{+-}(\tau). \quad (\text{D8})$$

With this formalism we can also characterize spatially uniform order parameters, such as the mean squared magnetization [49]

$$\langle m_i^2 \rangle = V^{-1} \lim_{\tau \rightarrow 0} C_{++}^{ii}(\tau). \quad (\text{D9})$$

The antiferromagnetic order parameters in Ref. [49] can similarly be extracted from same-time information in the correlators. There, the authors explain that the true indication of a staggered order is a spin (or charge) separation between the two sublattices, but that in a finite volume this order parameter will exactly vanish unless a bias is introduced, simulated with, and ultimately taken to zero. Rather than simply squaring this difference, which corresponds to considering the above correlators projected to the alternating irrep on both indices at vanishing temporal separation, they define additional quadratic observables. The total-per-sublattice order parameters $\langle q^2 \rangle$ and $\langle S_i^2 \rangle$ in Ref. [49] are given by

$$\begin{aligned} \langle q^2 \rangle &= V^{-1} \lim_{\tau \rightarrow 0} (\hat{A} C^{\rho\rho} \hat{A} + \hat{B} C^{\rho\rho} \hat{B})_{--} \\ \langle S_i^2 \rangle &= V^{-1} \lim_{\tau \rightarrow 0} (\hat{A} C^{ii} \hat{A} + \hat{B} C^{ii} \hat{B})_{--}, \end{aligned} \quad (\text{D10})$$

which differ from those parameters defined above by the absence of cross-terms $\hat{A} \hat{B}$ and $\hat{B} \hat{A}$. Note that the signs these cross terms carry are what distinguish the $+$ irrep from the $-$ irrep, in the sense that dropping them from the $++$ correlators yields the same numerical data as dropping them from the $--$ correlators,

$$(\hat{A} C^{\rho\rho} \hat{A} + \hat{B} C^{\rho\rho} \hat{B})_{--} = (\hat{A} C^{\rho\rho} \hat{A} + \hat{B} C^{\rho\rho} \hat{B})_{++}, \quad (\text{D11})$$

for example. Hence, once the cross terms are dropped, the distinction between staggered and uniform order is lost. In order to distinguish a uniform order from a staggered order, it makes physical sense to consider

$$\left(\sum_{\substack{x \in A \\ y \in B}} + \sum_{\substack{x \in B \\ y \in A}} \right) C_{xy}^{\rho\rho}(\tau) \quad \text{or} \quad \sum_{\langle x,y \rangle} C_{xy}^{\rho\rho}(\tau) \quad (\text{D12})$$

where the sum can be taken over $x \in A$ and $y \in B$ (and vice versa) or, if contributions from wide spatial separations contribute too much noise, over nearest neighbors $\langle x, y \rangle$. In the two-site example considered below we simply do irrep projection.

The uniform irreps yield correlations between the operators summed over space. These operators commute with the Hamiltonian and are therefore conserved. Commuting with the Hamiltonian also implies the time independence of the spectral decomposition of the respective correlator. We may still calculate the correlation function as a function of time and average over time to get an estimator for the same-time correlator,

$$\langle \rho_+ \rho_+ \rangle = \frac{1}{N_t} \sum_\tau C_{++}^{\rho\rho}(\tau), \quad (\text{D13})$$

for example.

¹⁰At equal time this definition matches the definition in Ref. [38].

¹¹This definition matches, for example, Eqs. (2) and (3) of Ref. [63], which was also studied in Ref. [64].

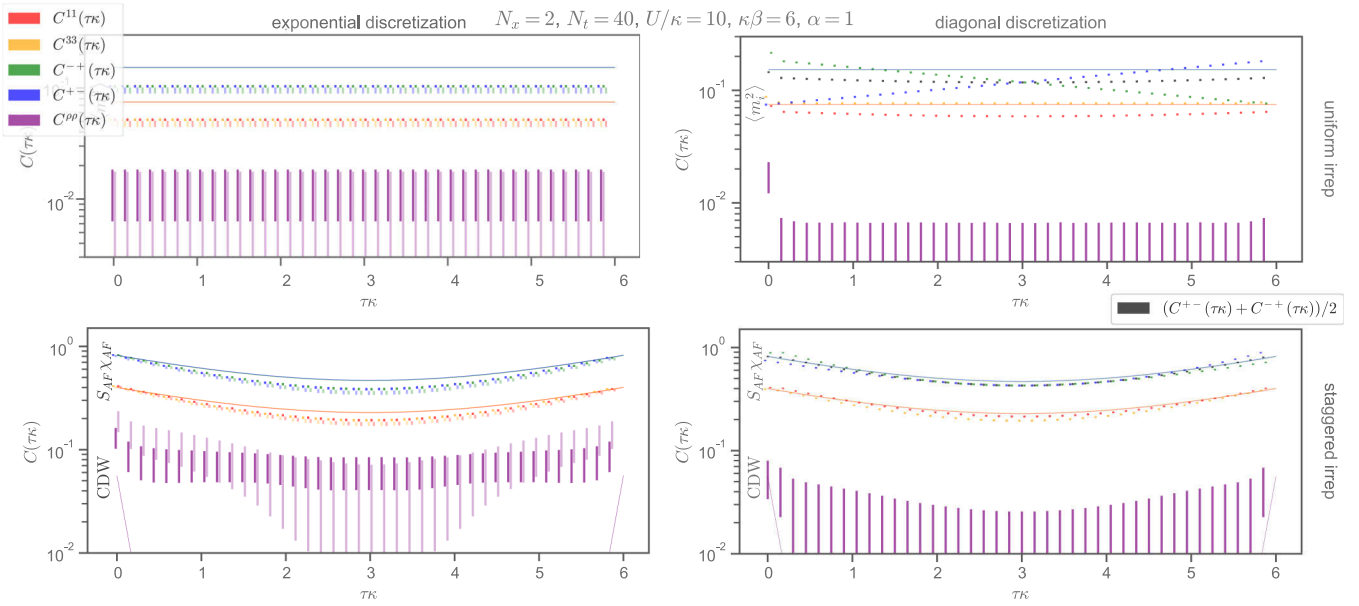


FIG. 18. A variety of correlation functions (shown in different colors) of local bilinear operators as a function of Euclidean time, for the exponential and diagonal discretizations, projected to the uniform and staggered irreps on both spatial indices. Exact results are shown as thin solid lines, while numerical results are shown as markerless error bars, and the respective equal-time order parameters are indicated in text, though the volume factors are omitted for clarity. In the left panel, the dark and light results correspond to different ensembles, as described in the text. For additional details and discussion, see the text.

1. Numerical results

In Fig. 18, we reuse the ensembles of field configurations from the two-site problem of Sec. III C (in particular, the $\alpha = 1$ ensembles used in Figs. 7 and 8, produced with a very fine molecular dynamics integration), measuring a variety of correlation functions, which are shown in different colors. We show the exponential and diagonal discretizations in the left and right columns, respectively, and the uniform irrep and the staggered irrep in the top and bottom rows, respectively. The results from an exact diagonalization are shown as thin lines, while the numerical results are shown as error bars without markers. In the exponential case, we show two data sets—the darker corresponds to the red ensemble of the mentioned figures and the lighter to the blue ensemble. For visibility, C^{11} and C^{33} are slightly offset horizontally, as are C^{+-} and C^{+} .

As mentioned, the uniform irrep has time-independent correlators. The exact uniform charge-charge correlator is not shown because it is approximately 5.8×10^{-12} . The darker exponential ensemble is clearly incompatible with this result, while the lighter is entirely uncertain $0.007(10)$; the diagonal ensemble is similarly uncertain, $0.001(5)$, except for the first time slice, which is incompatible with the exact result.

Additionally, in the diagonal case we see that C^{+-} and C^{+} are not time independent (we average them, shown as black points). We expect this behavior to vanish in the continuum limit, though we emphasize that according to (D13) the time average is a good estimator of the true value. The reason for this nonconstant behavior is presumably related to the violations of chiral symmetry in the diagonal case. We see the further effects of chiral symmetry breaking— C^{11} and C^{33} differ, as do C^{+-} and C^{+} , while those pairs are the same in the exponential case, as in Ref. [58]. Further, in the uniform

irrep we see that the diagonal correlators are not perfectly equal from time slice to time slice (the first time slice, in particular, differs markedly), in contrast to the exponential case.

This figure also emphasizes the need for a controlled continuum limit, even in the exponential case where, remarkably, the S_{AF} and χ_{AF} order parameters are very near to their continuum values that come from direct diagonalization, though the corresponding intermediate-time behaviors differ from the exact result. Absent a continuum limit, a formal demonstration that the equal-time correlator is improved is needed to justify not performing a controlled limit. Interestingly, the diagonal C_{++}^{33} correlator gets very close to the continuum (though the C_{++}^{11} differs—it should converge to the exact answer in the continuum limit). At intermediate times, the diagonal discretization seems to give results closer to the exact result; this observation is in line with what was found in Appendix B. Nevertheless, we stress that in either discretization a continuum limit is required for reliable, systematically controlled results.

The CDW order parameter, which is known to be susceptible to ergodicity problems when using the exponential discretization [49], can be seen to be incompatible with the exact result using that discretization, even in this small example—and the two different ensembles give different values, indicating a dependence on the ensemble, and therefore, sensitivity to a violation of ergodicity. In contrast, the CDW order parameter comes out correct for the diagonal discretization, bolstering our claim that the diagonal discretization can be advantageous for large-scale simulations. The exact $C_{--}^{\rho\rho}$ correlator falls extremely quickly for these values of $U\beta$ and $\kappa\beta$ —and neither exponential ensemble gives much indication of decay, while the diagonal discretization is statistically compatible with zero after the third time slice.

- [1] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid Monte Carlo, *Phys. Lett. B* **195**, 216 (1987).
- [2] A. D. Kennedy, Algorithms for dynamical fermions, in *Perspectives in Lattice QCD* (World Scientific, Singapore, 2008), pp. 15–81.
- [3] Y. Saad, *Iterative Methods for Sparse Linear Systems* (SIAM, Philadelphia, PA, 2003), Vol. 82.
- [4] C. Bernard *et al.*, Panel discussion on the cost of dynamical quark simulations, Lattice field theory, *Nucl. Phys. B: Proc. Suppl.* **106–107**, 199 (2002).
- [5] K. Jansen, Lattice QCD: A critical status report, *PoS (Lattice 2008)* 010.
- [6] C. Urbach, K. Jansen, A. Shindler, and U. Wenger, HMC algorithm with multiple time scale integration and mass pre-conditioning, *Comput. Phys. Commun.* **174**, 87 (2006).
- [7] K. Orginos, Innovations in lattice QCD algorithms, *J. Phys.: Conf. Ser.* **46**, 132 (2006).
- [8] M. A. Clark, The rational hybrid Monte Carlo algorithm, *PoS (Lattice 2006)* 004.
- [9] H. B. Nielsen and M. Ninomiya, Absence of neutrinos on a lattice: (I). Proof by homotopy theory, *Nucl. Phys. B* **185**, 20 (1981).
- [10] H. B. Nielsen and M. Ninomiya, Absence of neutrinos on a lattice: (II). Intuitive topological proof, *Nucl. Phys. B* **193**, 173 (1981).
- [11] R. Narayanan and H. Neuberger, Chiral determinant as an overlap of two vacua, *Nucl. Phys. B* **412**, 574 (1994).
- [12] R. Narayanan and H. Neuberger, A Construction of lattice chiral gauge theories, *Nucl. Phys. B* **443**, 305 (1995).
- [13] H. Neuberger, Exactly massless quarks on the lattice, *Phys. Lett. B* **417**, 141 (1998).
- [14] H. Neuberger, More about exactly massless quarks on the lattice, *Phys. Lett. B* **427**, 353 (1998).
- [15] D. B. Kaplan, A Method for simulating chiral fermions on the lattice, *Phys. Lett. B* **288**, 342 (1992).
- [16] Y. Shamir, Chiral fermions from lattice boundaries, *Nucl. Phys. B* **406**, 90 (1993).
- [17] V. Furman and Y. Shamir, Axial symmetries in lattice QCD with Kaplan fermions, *Nucl. Phys. B* **439**, 54 (1995).
- [18] K. Symanzik, Concerning the continuum limit in some lattice theories, *J. Phys. Colloq.* **43**, C3-254 (1982).
- [19] K. Symanzik, Continuum Limit and Improved Action in Lattice Theories, in *Topical Symposium on High-Energy Physics Tokyo, Japan* (World Scientific, Singapore, 1982), Vol. 463, pp. 17–28.
- [20] M. Creutz, Global Monte Carlo algorithms for many-fermion systems, *Phys. Rev. D* **38**, 1228 (1988).
- [21] J. Hubbard, Electron correlations in narrow energy bands, *Proc. R. Soc. London A* **276**, 238 (1963).
- [22] J. Hubbard, Electron correlations in narrow energy bands. II. The degenerate band case, *Proc. R. Soc. London A* **277**, 237 (1964).
- [23] J. Hubbard, Electron correlations in narrow energy bands. III. An improved solution, *Proc. R. Soc. London A* **281**, 401 (1964).
- [24] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, Monte Carlo calculations of coupled boson-fermion systems. I, *Phys. Rev. D* **24**, 2278 (1981).
- [25] J. E. Hirsch, Discrete Hubbard-Stratonovich transformation for fermion lattice models, *Phys. Rev. B* **28**, 4059 (1983).
- [26] S. R. White, D. J. Scalapino, R. L. Sugar, E. Y. Loh, J. E. Gubernatis, and R. T. Scalettar, Numerical study of the two-dimensional Hubbard model, *Phys. Rev. B* **40**, 506 (1989).
- [27] M. Bercx, F. Goth, J. S. Hofmann, and F. F. Assaad, The ALF (Algorithms for Lattice Fermions) project release 1.0. Documentation for the auxiliary field quantum Monte Carlo code, *SciPost Phys.* **3**, 013 (2017).
- [28] R. T. Scalettar, D. J. Scalapino, R. L. Sugar, and D. Toussaint, Hybrid molecular-dynamics algorithm for the numerical simulation of many-electron systems, *Phys. Rev. B* **36**, 8632 (1987).
- [29] J. E. Hirsch, Stable monte carlo algorithm for fermion lattice systems at low temperatures, *Phys. Rev. B* **38**, 12023 (1988).
- [30] Z. Fodor, S. D. Katz, and K. K. Szabo, Dynamical overlap fermions, results with hybrid Monte Carlo algorithm, *J. High Energy Phys.* **08** (2004) 003.
- [31] Z. Fodor, S. D. Katz, and K. K. Szabo, Dynamical overlap fermions, results with HMC algorithm, *Nucl. Phys. Proc. Suppl.* **140**, 704 (2005).
- [32] N. Cundy, S. Krieg, G. Arnold, A. Frommer, Th. Lippert, and K. Schilling, Numerical methods for the QCD overlap operator IV: Hybrid Monte Carlo, *Comput. Phys. Commun.* **180**, 26 (2009).
- [33] T. O. Wehling, E. Şaşıoğlu, C. Friedrich, A. I. Lichtenstein, M. I. Katsnelson, and S. Blügel, Strength of Effective Coulomb Interactions in Graphene and Graphite, *Phys. Rev. Lett.* **106**, 236805 (2011).
- [34] H.-K. Tang, E. Laksono, J. N. B. Rodrigues, P. Sengupta, F. F. Assaad, and S. Adam, Interaction-Driven Metal-Insulator Transition in Strained Graphene, *Phys. Rev. Lett.* **115**, 186602 (2015).
- [35] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, The electronic properties of graphene, *Rev. Mod. Phys.* **81**, 109 (2009).
- [36] V. N. Kotov, B. Uchoa, V. M. Pereira, F. Guinea, and A. H. Castro Neto, Electron-electron interactions in graphene: Current status and perspectives, *Rev. Mod. Phys.* **84**, 1067 (2012).
- [37] T. Paiva, R. T. Scalettar, W. Zheng, R. R. P. Singh, and J. Oitmaa, Ground-state and finite-temperature signatures of quantum phase transitions in the half-filled hubbard model on a honeycomb lattice, *Phys. Rev. B* **72**, 085123 (2005).
- [38] Z. Y. Meng, T. C. Lang, S. Wessel, F. F. Assaad, and A. Muramatsu, Quantum spin liquid emerging in two-dimensional correlated Dirac fermions, *Nature (London)* **464**, 847 (2010).
- [39] Y. Otsuka, S. Yunoki, and S. Sorella, Universal Quantum Criticality in the Metal-Insulator Transition of Two-Dimensional Interacting Dirac Electrons, *Phys. Rev. X* **6**, 011029 (2016).
- [40] 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).
- [41] A. J. Heeger, S. Kivelson, J. R. Schrieffer, and W. P. Su, Solitons in conducting polymers, *Rev. Mod. Phys.* **60**, 781 (1988).
- [42] S. Krieg, T. Luu, J. Ostmeier, P. Papaphilippou, and C. Urbach, Accelerating hybrid Monte Carlo simulations of the Hubbard model on the hexagonal lattice, *Comput. Phys. Commun.* **236**, 15 (2018).
- [43] R. Brower, C. Rebbi, and D. Schaich, Hybrid Monte Carlo simulation on the graphene hexagonal lattice, *PoS (Lattice 2011)* 056.

- [44] 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).
- [45] 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).
- [46] T. Luu and T. A. Lähde, Quantum Monte Carlo calculations for carbon nanotubes, *Phys. Rev. B* **93**, 155106 (2016).
- [47] E. Berkowitz, C. Körber, S. Krieg, P. Labus, T. A. Lähde, and T. Luu, Extracting the single-particle gap in carbon nanotubes with lattice quantum Monte Carlo, *EPJ Web Conf.* **175**, 03009 (2018).
- [48] M. V. Ulybyshev and S. N. Valgushev, Path integral representation for the Hubbard model with reduced number of Lefschetz thimbles, [arXiv:1712.02188](https://arxiv.org/abs/1712.02188).
- [49] 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).
- [50] J.-L. Wynen, E. Berkowitz, C. Körber, T. A. Lähde, and T. Luu, Avoiding Ergodicity Problems in Lattice Discretizations of the Hubbard Model - Data (Version 2018-12-19) [Data set] (2018), doi:[10.5281/zenodo.2414853](https://doi.org/10.5281/zenodo.2414853).
- [51] J.-L. Wynen, E. Berkowitz, and C. Körber, jl-wynen/isle: Release 0.1 (Version v0.1). Zenodo (2018), doi:[10.5281/zenodo.2413854](https://doi.org/10.5281/zenodo.2413854).
- [52] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevB.100.075141> for animations of eigenvalue motions corresponding to Fig. 2.
- [53] R. M. Neal, Sampling from multimodal distributions using tempered transitions, *Stat. Comput.* **6**, 353 (1996).
- [54] M. J. Betancourt, Adiabatic Monte Carlo, [arXiv:1405.3489](https://arxiv.org/abs/1405.3489).
- [55] A. Alexandru, G. Basar, P. F. Bedaque, and N. C. Warrington, Tempered transitions between thimbles, *Phys. Rev. D* **96**, 034513 (2017).
- [56] J. Besag, P. Green, D. Higdon, and K. Mengersen, Bayesian computation and stochastic systems, *Stat. Sci.* **10**, 3 (1995).
- [57] M. Ulybyshev, C. Winterowd, and S. Zafeiropoulos, Collective charge excitations and the metal-insulator transition in the square lattice Hubbard-Coulomb model, *Phys. Rev. B* **96**, 205115 (2017).
- [58] P. Buividovich, D. Smith, M. Ulybyshev, and L. von Smekal, Competing order in the fermionic Hubbard model on the hexagonal graphene lattice, *PoS (Lattice 2016)* 244.
- [59] Jülich Supercomputing Centre, JURECA: Modular supercomputer at Jülich Supercomputing Centre, *J. Large-Scale Res. Facilit.* **4**, A132 (2018).
- [60] T. Živković, LU-decomposition of tridiagonal and cyclic tridiagonal matrices, *Osječki matematički list* **12**, 153 (2012).
- [61] A. Bulgac, J. E. Drut, and P. Magierski, Quantum Monte Carlo simulations of the BCS-BEC crossover at finite temperature, *Phys. Rev. A* **78**, 023625 (2008).
- [62] M. Ulybyshev, N. Kintscher, K. Kahl, and P. Buividovich, Schur complement solver for quantum Monte-Carlo simulations of strongly interacting fermions, *Comput. Phys. Commun.* **236**, 118 (2019).
- [63] Y.-X. Zhang, W.-T. Chiu, N. C. Costa, G. G. Batrouni, and R. T. Scalettar, Charge Order in the Holstein Model on a Honeycomb Lattice, *Phys. Rev. Lett.* **122**, 077602 (2019).
- [64] C. Chen, X. Y. Xu, Z. Y. Meng, and M. Hohenadler, Charge-Density-Wave Transitions of Dirac Fermions Coupled to Phonons, *Phys. Rev. Lett.* **122**, 077601 (2019).
- [65] T. C. Lang, Quantum Monte Carlo methods and strongly correlated electrons on honeycomb structures, PhD thesis, Universität Würzburg (2010).