We address the calculation of dynamical correlation functions for many fermion systems at zero temperature, using the auxiliary-field quantum Monte Carlo method. The two-dimensional Hubbard Hamiltonian is used as a model system. Although most of the calculations performed here are for cases where the sign problem is absent, the discussions are kept general for applications to physical problems when the sign problem does arise. We study the use of twisted boundary conditions to improve the extrapolation of the results to the thermodynamic limit. A strategy is proposed to drastically reduce finite size effects relying on a minimization among the twist angles. This approach is demonstrated by computing the charge gap at half filling. We obtain accurate results showing the scaling of the gap with the interaction strength $U$ in two dimensions, connecting to the scaling of the unrestricted Hartree-Fock method at small $U$ and Bethe ansatz exact result in one dimension at large $U$. An alternative algorithm is then proposed to compute dynamical Green functions and correlation functions which explicitly varies the number of particles during the random walks in the manifold of Slater determinants. In dilute systems, such as ultracold Fermi gases, this algorithm enables calculations with much more favorable complexity, with computational cost proportional to basis size or the number of lattice sites.

DOI: 10.1103/PhysRevB.94.085140

I. INTRODUCTION

The calculation of dynamical correlation functions of many-body quantum systems is a great challenge in theoretical condensed matter physics. Such functions provide a unique opportunity to explore the manifold of the excited states of a physical system. They often provide a much more direct connection to experimental measurements, giving access to crucial properties such as spectral functions, excitation spectra, and charge and spin gaps, to name a few.

With the advent of modern computational resources, quantum Monte Carlo (QMC) simulations [1–4] are becoming a very powerful tool for computations in quantum many-particle models and realistic systems. A vast array of total energy and related quantities have been computed. Equal-time correlation functions have also been studied fairly routinely by QMC, especially in model systems. To estimate dynamical properties from first principles using QMC has been more challenging. Important results have been obtained for bosonic strongly correlated systems [5–11]. Also excited states of low-dimensional bosonic systems have been recently studied [12,13]. For fermionic systems, in special situations where the sign problem is not present, accurate calculations have been performed [14–17]. A few attempts have been also made [14,18–21] in the more difficult situations where approximations are needed to deal with the sign problem [1,22–25].

In this paper we study the calculation of imaginary-time correlation functions and excited state information in interacting fermion systems at zero temperature. We formulate and discuss our methods in a general framework, keeping in mind applications in many-fermion systems where a sign problem arises and a constrained path approach is applied within an open-ended imaginary-time projection. Most of our illustrations and applications in the present paper, however, will be in systems where the sign problem is absent and exact results can be obtained. For these we use the two-dimensional Hubbard Hamiltonian with repulsive interaction at half filling. We employ the path-integral auxiliary-field quantum Monte Carlo (AFQMC) framework, but including a force-bias acceleration technique [26] in the Metropolis sampling and a technique to control Monte Carlo variance divergence [27].

In the first part of the paper, we implement a standard approach [15,28] of computing non-equal-time Green functions and correlation functions within AFQMC. Our focus is on the extrapolation of the results to the thermodynamic limit and testing the efficiency of different implementations in general many-fermion systems. We show that it is convenient to introduce twisted boundary conditions [29], and suggest a way to exploit the boundary conditions that dramatically reduces finite-size effects in the calculation of the charge gap. Accurate results for the gap are obtained in the repulsive Hubbard model even at weak interactions, which mimic many real materials where the gap might be very small compared to the energy scales, presenting challenges for numerical calculations. With our approach, the charge gap is determined even at $U/t$ as small as 0.5, far beyond the reach of previous unbiased many-body calculations.

In the second part, we propose an alternative algorithm to compute dynamical Green functions, density-density and spin-spin correlation functions which, in the dilute limit, dramatically reduces the complexity without affecting the numerical stability of the calculations or the accuracy of the results. The method relies on the explicit variation of the number of particles during the random walk in the manifold of Slater determinants. We show that, for example, the spectral function for one given momentum can be calculated with a computational complexity proportional to the number of lattice sites, which enables simulations on very large lattices. This method will have great advantages in systems such as atomic Fermi gases, which can be modeled by the attractive Hubbard model in the dilute limit [30], or real materials, where the calculations typically require [25] a basis size much larger than the number of electrons.
The remainder of this paper is organized as follows. In Sec. II we will briefly sketch the AFQMC methodology, and then describe our implementation of a stable algorithm to compute dynamical Green functions, both in the path-integral (for sign-problem-free systems) and open-ended random walk (for constrained path calculations when the sign problem is present). In Sec. III, we propose a strategy to control finite-size effects using twisted boundary conditions and present our results for the charge gap in the repulsive Hubbard model at half filling in the intermediate and weakly interacting regime. In Sec. IV we describe our alternative approach which, in the dilute limit, dramatically improves the complexity of the calculations. We conclude in Sec. V.

II. AUXILIARY-FIELD QUANTUM MONTE CARLO FORMALISM

We introduce the basic notations of the methodology using the Hubbard hamiltonian which, as mentioned, will be our model system:

\[ \hat{H} = -t \sum_{\langle i,j \rangle, \sigma = \uparrow, \downarrow} \epsilon_{i,\sigma} \hat{c}_{i,\sigma}^\dagger \hat{c}_{j,\sigma} + U \sum_i \left( \hat{n}_{i,\uparrow} - \frac{1}{2} \right) \left( \hat{n}_{i,\downarrow} - \frac{1}{2} \right), \]  

(1)

where the labels \( i, j \) run over the sites of a square lattice with \( N_s = L_x \times L_y = L^2 \) sites, the symbol \( \langle \cdot \rangle \) denotes, as usual, nearest neighbors, and \( \hat{n}_{i,\sigma} = \hat{c}_{i,\sigma}^\dagger \hat{c}_{i,\sigma} \) is the particle number density operator on site \( i \) for the given spin direction \( \sigma \). The Hubbard model \([31,32] \) is one of the most widely studied models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics. It is of fundamental theoretical importance in the context of magnetism and cuprate models in condensed matter physics.
Whenever $\mathcal{W}(\mathbf{X}) \geq 0$ for each auxiliary-field configuration, as it happens if $U < 0$ and $N_\uparrow = N_\downarrow$ or at half filling ($N_\uparrow = N_\downarrow$) when $U > 0$, the integral in (10) can be evaluated via Monte Carlo. We use an efficient Metropolis sampling of the paths, exploiting a force bias [2,26] that allows a high acceptance ratio in the updates of the path in the auxiliary-field configuration space, and eliminating the infinite variance problem [27] with a bridge link approach.

When a sign problem is present, a constrained path [24] or phase-free approximation [25] can be imposed to remove the exponentially growing noise (with system size or the length of the path) and restore the same computational scaling as in the sign-problem-free cases. The paths become open ended. The primary difference is that $\langle \phi_L \mid \hat{B}(\mathbf{x}(N_t)) \cdots \hat{B}(\mathbf{x}(1)) \hat{c}_j^\dagger \hat{c}_i \mid \phi_R \rangle$ is now produced by backpropagation (BP) [2], and a weight accompanies each path. The details of the backpropagation for static properties have been discussed elsewhere [35]. The modification to compute dynamical properties is minimal beyond what is necessary for the path-integral formalism, and we will comment on it as needed below.

**B. Dynamical properties**

Dynamical correlation functions in imaginary time at zero temperature have the general form

$$f(\tau) = \frac{\langle \Psi_0 | \hat{A} e^{-\tau(\hat{H} - E_0)} \hat{B} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle}, \quad (15)$$

where $\hat{A}$ and $\hat{B}$ can be destruction or creation operators, or one-body operators such as the particle density or the spin density or even more general operators.

Let us focus on the dynamical particles and holes Green functions in imaginary time, i.e., the matrices

$$G^{\sigma,\sigma_j}_{\alpha,\alpha_j}(\tau) = \frac{\langle \Psi_0 | \hat{c}_{\alpha,\sigma} e^{-\tau(\hat{H} - E_0)} \hat{c}_{\alpha_j,\sigma_j}^\dagger | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle}, \quad (16)$$

and

$$G^{h}_{\alpha,\alpha_j}(\tau) = \frac{\langle \Psi_0 | \hat{c}_{\alpha,\sigma} e^{-\tau(\hat{H} - E_0)} \hat{c}_{\alpha_j,\sigma_j} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle}. \quad (17)$$

When the Hamiltonian is spin independent as is the case in the Hubbard model, all the terms with $\sigma \neq \sigma'$ identically vanish. In order to keep the notations simple, we will work for $\sigma = \uparrow$ (the other case being analogous) and neglect the spin index. Dealing with translationally invariant systems, we will denote $G^{p,h}_{\sigma,\sigma_j}(\tau)$ by $G^{p,h}(R,\tau)$ with $R = (i-j)$. The Fourier transforms, i.e., the dynamical Green functions in momentum space, will be denoted by $\hat{G}^{p,h}(Q,\tau)$, where $Q = (Q_x, Q_y)$ is a wave vector of the reciprocal lattice: $Q_x = \frac{2\pi}{L_x} n_x$, with $n_x \in [0, \ldots, L_x-1]$, and similarly for $Q_y$.

The imaginary-time propagator between the operators $\hat{A}$ and $\hat{B}$ can again be expressed using Eq. (3). This can be thought of as inserting an extra segment to the path we keep: a number $N_\tau = \tau/\delta \tau$ of time-slices, say $\mathbf{x}(1), \ldots, \mathbf{x}(N_\tau)$. The static estimator Eq. (14) is replaced by a dynamical estimator which (for example, in the case of the particles Green function) can be cast in the form (see Ref. [14])

$$g(\mathbf{x}, \tau) = \frac{\langle \phi_L | \hat{c}_i \hat{B}(\mathbf{x}(N_t)) \cdots \hat{B}(\mathbf{x}(1)) \hat{c}_j^\dagger \mid \phi_R \rangle}{\langle \phi_L | \hat{B}(\mathbf{x}(N_t)) \cdots \hat{B}(\mathbf{x}(1)) \mid \phi_R \rangle}. \quad (18)$$

To keep the notation simple, we will write $\hat{B}$, instead of $\hat{B}(\mathbf{x}(i))$, from now on.

In order to calculate Eq. (18) for a given configuration of auxiliary fields, we use the manipulations presented in Ref. [15]. We introduce the equal-time Green function matrix

$$G^{p}(n,n) = \frac{\langle \phi_L | \hat{B}_N \cdots \hat{B}_n \hat{c}_j^\dagger \hat{B}_{n-1} \cdots \hat{B}_1 | \phi_R \rangle}{\langle \phi_L | \hat{B}_N \cdots \hat{B}_1 | \phi_R \rangle}. \quad (19)$$

and the time displaced one

$$G^{p}(n,m) = \frac{\langle \phi_L | \hat{B}_N \cdots \hat{B}_{n-1} \hat{c}_j \hat{B}_m \cdots \hat{B}_1 | \phi_R \rangle}{\langle \phi_L | \hat{B}_N \cdots \hat{B}_1 | \phi_R \rangle}. \quad (20)$$

The central result is provided by the following:

$$G^{p}(n,m) = G^{p}(n,n-1)G^{p}(n-1,n-2) \cdots G^{p}(m+1,m) \quad (21)$$

and

$$G^{p}(l,l-1) = G^{p}(l,l)B_l, \quad (22)$$

where $B_l \equiv B(\mathbf{x}(l))$ is the matrix of Eq. (7). These identities are straightforward to show using the commutators between the propagator and the creation or annihilation operators [36,37]. Similar relations can be easily obtained for the holes

$$G^{h}(n,m) = G^{h}(m+1,m)G^{h}(m+2,m+1) \cdots G^{h}(n,n-1) \quad (23)$$

and

$$G^{h}(l,l-1) = (B_l)^{-1}G^{h}(l,l). \quad (24)$$

The composite properties above allow one to build dynamical Green functions from shorter segments. One can break the original imaginary time interval into segments of arbitrary length, say $(n,n-k)$, which turns the matrix in Eq. (22) into the product of the $k$ matrices $B_l$ inside the segment. Multiplying the Green functions, which are physical properties, instead of composing the $B_l$ matrices, provides more accurate and numerically stable results for the dynamical correlations even for large values of the imaginary time. The length of the segments can be chosen for the best balance between numerical stability (requiring short segments) and computational time (less for longer segments). The complexity of this approach scales with the lattice size as $N^3$, since it requires multiplications of matrices of size $N_x \times N_y$.

Although we have discussed the formalism in the path-integral AFQMC framework, it can be easily modified for a constrained path approach of open-ended imaginary time projection, which becomes necessary when the sign problem is present. This is sketched in Fig. 1. When backpropagation begins, one records the initial population and their weights, which play the role of $\langle |\phi_R\rangle$. As the random walk proceeds
The central object in many-body theory. From the imaginary-time particle spectral weight for a correlated system, and is a spectral function provides insight into the nature of the single photoemission and inverse photoemission spectroscopy. The Green function. It can be experimentally measured from is proportional to the imaginary part of the time-ordered correlation function.

Here \( \Psi_{n}^{N,\pm1} \) are the eigenstates of the Hamiltonian operator with \( N_p \pm 1 \) particles corresponding to the energies \( E_{n}^{N,\pm1} \), while \( \mu \) is the chemical potential. The spectral function is proportional to the imaginary part of the time-ordered Green function. It can be experimentally measured from photoemission and inverse photoemission spectroscopy. The spectral function provides insight into the nature of the single particle spectral weight for a correlated system, and is a central object in many-body theory. From the imaginary-time correlation functions, a Fredholm integral equation has to be solved to determine the spectral function; in practice,

\[
\int_{-\infty}^{+\infty} e^{-\tau\omega} A(Q,\omega) d\omega = \left\langle \Psi_{0} \left| \hat{c}_{\sigma} e^{-\tau(H-E_{\mu})} \right| \right. Q,\omega > \mu \rangle \left. \Psi_{0} \right) \right| \Psi_{0} \rangle (26)
\]

and

\[
\int_{-\infty}^{+\infty} e^{-\tau\omega} A(Q,\omega) d\omega = \left\langle \Psi_{0} \left| \hat{c}_{\sigma} e^{-\tau(H-E_{\mu})} \right| \right. Q,\omega < \mu \rangle \left. \Psi_{0} \right) \right| \Psi_{0} \rangle (27)
\]

The charge gap \( \Delta \) is typically inferred from the behavior of the zero distance real-space Green function for large imaginary time. In general,

\[
\Delta = \frac{1}{2}(\epsilon_{p} + \epsilon_{h}) (28)
\]

with

\[
\epsilon_{p,h} = -\lim_{\tau \to +\infty} \frac{\log \left( G^{p,h}(R = 0,\tau) \right)}{\tau} = E(N_p \pm 1) - E(N_p) , (29)
\]

where \( E(N_p) \) is the ground state energy for \( N \) particles, while \( E(N_p \pm 1) \) correspond to the lowest energy eigenstates of the \( N_p \pm 1 \) systems having nonzero overlap with the state obtained by adding/removing a particle in any momentum state to the \( N_p \)-particle ground state. At half filling, particle-hole symmetry [14] allows us to simplify the above definition:

\[
\Delta = -\lim_{\tau \to +\infty} \frac{\log \left( \tilde{G}^{p}(R = 0,\tau) \right)}{\tau} - \mu. (30)
\]

Since the Hamiltonian defined in Eq. (1) is particle-hole symmetric, the chemical potential \( \mu \) is zero at half filling.

We could also compute the gap in momentum space:

\[
\Delta = \min_{Q} \left( -\lim_{\tau \to +\infty} \frac{\log \left( \tilde{G}^{p}(Q,\tau) \right)}{\tau} \right). (31)
\]

Equation (31) provides an intuitive physical meaning of the charge gap: \( \Delta \) is related to the minimum energy among the unoccupied states, which can be probed via inverse photoemission spectroscopy (see, for example, Ref. [38] and references therein), using a collimated beam of electrons directed at the sample. (At half filling, particle-hole symmetry makes this energy coincide with the minimum energy needed to extract a photoelectron in a direct spectroscopy experiment when the sample is illuminated via electromagnetic radiation.)

Computing the dynamical Green function in momentum space is more convenient than in real space, at least for smaller values of \( U/t \). In this regime \( \tilde{G}^{p}(R = 0,\tau) \) contains a linear combination of exponentials while \( \tilde{G}^{p}(Q,\tau) \) has a simpler structure which can be handled more easily via analytic continuation methodologies. It is straightforward to see this in the limiting case of \( U = 0 \), when \( \tilde{G}^{p}(Q,\tau) \) is a single exponential:

\[
\tilde{G}^{p}_{0}(Q,\tau) = \tilde{n}_{0}(Q) e^{-\tau\epsilon(Q)}, (32)
\]

where \( \tilde{n}_{0}(Q) = 1 - n_{0}(Q) \), with \( n_{0}(Q) \) being the Fermi distribution. In contrast,

\[
G^{p}_{0}(R = 0,\tau) = \frac{1}{N_{F}} \sum_{Q} \tilde{n}_{0}(Q) e^{-\tau\epsilon(Q)}, (33)
\]

in which the minimum gap \[\epsilon(Q_{F}) = 0 \text{ in this case, } Q_{F} \text{ belonging to the Fermi surface} \] provides the charge gap at
sufficiently large $\tau$ but many other exponentials can persist for significant $\tau$ values, especially as the system size grows. In fact we could construct a linear combination to target in order to optimize convergence and statistical accuracy in the computation of the gap:

$$\tilde{G}_B^p(\tau) \propto \sum_{Q \in B} \tilde{G}_Q^p(Q, \tau).$$

One simple definition for the momentum domain in the sum-
mation could be $B = \{ Q : |\pi(Q) - \bar{\pi}_B(Q)| > \text{const} \}$, where $\pi(Q) = 1 - n(Q)$ is defined with respect to the momentum distribution of the interacting system, $n(Q)$. The value of constant can be tuned.

Even more generally, one could use any single particle orbital $|\mu\rangle$ to create a quasiparticle excitation. The real and momentum space shown above are simply two special cases. In the formalism presented above, suitable linear combinations of the dynamical Green functions would be required. This possibility can be particularly useful in the alternative approach we propose in Sec. IV, where any single-particle orbital (for example, a natural orbital obtained from the many-body calculation) can be propagated along with the ground-state random walker with little additional cost.

In Fig. 2 we show an example of the calculation of imaginary-time Green functions at half filling with $U/t = 0.5$, in both real space and momentum space. It is evident that the two reach the same slope in logarithmic scale at large imaginary time but that $\tilde{G}_Q^p(Q, \tau)$ has a much simpler structure, allowing us to accurately calculate the slope without the need of reaching very large imaginary times. This is important since the relative statistical uncertainty increases exponentially, as shown in the inset.

The charge gap can also be estimated with an addition/removal technique, calculating directly $\varepsilon_p$ and $\varepsilon_h$ in

Eq. (28). This approach has the advantage that it does not require the evaluation of dynamical correlation functions. It also has several disadvantages. With three separate calculations (or two, if at half filling) of $N_p$ and $N_p \pm 1$ particles, it involves the difference between extensive quantities, which can give rise to large statistical uncertainties for large systems. Second, while not present in Hubbard-like models, the addition/removal of an electron in a supercell in real solids tends to create a significant additional finite-size effects which requires larger supercells or better correction schemes in the many-body calculation. Moreover, in the Hubbard model at half filling, the systems with $N_p \pm 1$ particles both have a sign problem, while the (half-filled) $N_p$ system does not [14].

We have performed addition/removal calculations of the gap to help check the robustness of the imaginary-time Green function approach. These calculations were performed with the constrained path Monte Carlo (CPMC) method [2,24]. We used trial wave functions obtained from the generalized Hartree-Fock (GHF), which were found to improve the results near half filling [34,39]. For the $N_p - 1$ calculation, we use the same GHF trial wave function as for half filling, simply omitting the extra orbital in the minority-spin sector, while the next virtual GHF orbital is used in the $(N_p + 1)$ case. This was found, by comparison with exact diagonalization results, to give exact results on the gap to within statistical errors.

In Fig. 3 the gaps computed from the imaginary-time Green function and from addition/removal are compared for a variety of systems ranging from $4 \times 4$ to $16 \times 16$ periodic supercells. Excellent agreement is seen between the two approaches. It is also evident that the gap converges rather slowly with supercell size. Especially at smaller values of $U$, very large lattices are needed and a fit in $1/L$ is difficult and can be unreliable. We discuss how to obtain more robust estimates at the thermodynamic limit in the next section.

III. GAPS AT THE THERMODYNAMIC LIMIT

A. Approaches to reduce the finite-size effects

To access bulk properties it is crucial to be able to extrapolate the results to the thermodynamic limit. We perform simulations up to lattices containing $24 \times 24$ sites, i.e., 576
et et al. (2016)

FIG. 4. Charge gap measured from UHF calculations at $U = 0.5$, as a function of $1/L$. The straight line indicates the UHF gap value at the thermodynamic limit, 0.0044272.

FIG. 5. Finite-size effects in computing the gap, the use of twist boundary conditions, and special twist values. The top panel shows the noninteracting gap and the exact many-body gap as a function of the twist parameters. The bottom panel shows the corrected gaps and identifies the minimum. The system is a $14 \times 14$ lattice at $U/t = 0.5$.

We can further improve convergence to the thermodynamic limit by removing the one-body finite-size effect [40,41]. We can correct the values of the computed gap for a given $\theta$ by the finite-size noninteracting gap, to eliminate or reduce the effects arising purely from the shift of the Fermi sphere and the shell structure. The upper panel of Fig. 5 illustrates this effect, where a strong correlation is evident between the true many-body gap $\Delta(\theta)$ and the noninteracting gap $\Delta^0(\theta)$. The simple correction $\tilde{\Delta}(\theta) = \Delta(\theta) / \Delta^0(\theta)$, where $\Delta^0 = 0$ is the noninteracting gap at the thermodynamic limit, gives one order of magnitude reduction in the fluctuations of the gap values. This is expected at small $U$. We find that, although the dependence on the twist parameter is weaker at, say, $U = 4$, the correlation is also present at larger $U$. The gap result as a function of system size after the one-body finite-size correction is shown in Fig. 6. We see that the twist-averaged result [averaging $\tilde{\Delta}(\theta)$ over the twist $\theta$, in this case nearly 50 quasirandom twist parameters] has much smaller error bars, because of the reduction in the fluctuations between twist parameters mentioned above, and displays better convergence and a quadratic scaling in $1/L$.

We propose a third step for accelerating convergence, involving a different way to use the twisted boundary conditions than the more standard procedure above. After one-body finite-size correction, we seek the minimum gap among all the twist angles:

$$
\Delta_{\text{min}} = \min_{\theta} (\tilde{\Delta}(\theta)) .
$$

This allows one to better sample for the minimum in Eq. (31). We stress that the difference between Eq. (36), averaging over $\Delta(\theta)$, and even $\Delta(\theta)$ itself vanishes in the thermodynamic limit. However, taking the minimum among the postcorrection twist results enables access to estimations of the values of Eq. (31) for a much larger set of points, which improves the estimation of the minimum. As we see in Fig. 6, this procedure leads to a further improvement over the TA results, with the computed gap becoming essentially flat for $L \gtrsim 12$. Most of our results in the next section are obtained with this procedure of finding the minimum corrected gap. The statistical uncertainties on the minimum corrected gap are estimated simply as the QMC errors; we have checked that...
this estimation is reliable by performing, in some situations, several independent calculations and computing the variance of $\Delta_{\text{min}}$.

We have examined the location of $\Delta_{\text{min}}$. In the bottom panel of Fig. 5 we have marked the twist angle which yielded the minimum. We find that the corrected gap $\Delta(\theta)$ reaches its minimum when the noninteracting gap is maximum, that is around $\theta_{\text{min}} = (0,1/2)$ or symmetry-related points. This observation holds for all the cases we have studied, including a variety of lattice sizes $L$ at multiple values of $U$ (0.5, 1, and 4). We rationalize the observation as follows: when a particle is added at half filling, the system accommodates it by creating a spin wave. The value $\theta_{\text{min}} = (0,1/2)$ allows the maximum wavelength for this excitation, thus allowing the minimum energy. In the next section, we apply this special $\theta_{\text{min}}$ to obtain the gap value in many other $U$ values between 0.5 and 4. Of course the particular value of $\theta_{\text{min}}$ will depend on the system, but any insight towards identifying its value or narrowing its range will help reduce the computational cost.

**B. Results on gaps and the spectral function for the Hubbard model at half filling**

In this section, we present our results of the charge gap in the repulsive Hubbard model at half filling, systematically as a function of the interaction, as well as the spectral function at $U = 4t$. These calculations are similar to prior efforts that exist to study such quantities, using both the Lanczos method for small lattices [42,43] and QMC for larger lattices [14,16,17]. Our calculations reach larger system sizes and apply the approaches discussed above to systematically reach the thermodynamic limit.

In Fig. 6 we show the final results obtained for $U = 0.5t$. Lattice sizes up to $24 \times 24$ were studied using multiple quasirandom twist angles. We find a charge gap of $\Delta = 0.00027(4)$ at the thermodynamic limit. This very small value is clearly impossible to determine using conventional calculations with PBC or even twist-averaged boundary conditions (TABC). That the gap value is small but nonzero is significant, confirming that the ground state of the Hubbard model is insulating at small finite $U$.

In Fig. 7 we show the results of the same calculation for $U/t = 1$. Qualitatively the behavior is the same as for $U = 0.5$. The statistical uncertainties are larger, since $U = 0.5$ is so small that the dynamical Green functions are very similar to the ones of the noninteracting system, which makes the finite-size correction in Eq. (36) especially effective in reducing the statistical uncertainty. Our estimation of the charge gap at $U/t = 1$ is $\Delta = 0.0034(4)$.

We present results for $U/t = 4$ in Fig. 8. In this case, the twist-averaged results display a nearly linear dependence on $1/L$, consistent with results from a previous QMC study [16]. As in the other cases, our estimator from Eq. (36) becomes flat also in this case. Our estimate of the charge gap at $U/t = 4$ is $\Delta = 0.65(2)$.

Next, we map out a detailed $\Delta$ vs $U$ curve by studying a variety of interaction strengths. Having established the $\theta_{\text{min}}$ value from the systematic searches at the $U$ values studied above, we now use the special twist value $\theta_{\text{min}}$ for each additional system, computing the gap at a sufficiently large $L$. We have verified in a few systems by calculations at multiple $L$’s that convergence to the thermodynamic limit has been reached. The results are shown in Fig. 9. We find that, at small $U$, the gap behaves as $\Delta(U) = \alpha \exp(-\beta \sqrt{U})$, as predicted by UHF [14], with renormalized parameters $\alpha$ and $\beta$. We find $\alpha = 0.23(4)$ and $\beta = 4.3(2)$, in contrast with the UHF
predictions $\alpha = 32$ and $\beta = 2\pi$. (The actual UHF results are seen to, not surprisingly, severely overestimate the gap.) On the other hand, at large $U/t$, the gap appears to follow the same behavior predicted in one dimension from the Bethe ansatz [44]:

$$\Delta_{1D}(U) = \frac{U}{2} - 2 + 4 \int_0^{+\infty} d\omega \frac{J_1(\omega)}{\omega[1 + \exp(\omega U/2)]}, \quad (37)$$

where $J_1(\omega)$ is a Bessel function.

Finally, in Fig. 10 we show a calculation of the full spectral function $A(Q,\omega)$ defined in Eq. (25). The horizontal axis shows $Q$ along a path in the Brillouin zone, indicated by the end points of the straight-line segments. The spectral function was obtained from analytic continuation of the imaginary-time Green functions, as discussed in Eqs. (26) and (27).

The analytic continuation has been performed using the genetic inversion via falsification of theories (GIFT) method [5]. As mentioned, the spectral function can be measured experimentally via photoemission experiments, and provides a map of the single particle states of the system. From the plot, where the noninteracting spectral function $A^0(Q,\omega) = \delta(\omega - \varepsilon(Q))$ is also shown, the effect of the interactions is evident, with the opening of a gap at the Fermi surface, as well as the broadening and renormalization of the dispersion relation.

**IV. ALTERNATIVE METHOD**

In this section we propose an alternative method which will enable faster computations of the Green functions in a larger number of situations than the method of Sec. II B. In particular, the alternative method allows us to calculate single matrix elements of the Green function with a number of operations scaling linearly with lattice sites (or basis size). In systems such as a dilute Fermi gas and ab initio calculation of realistic systems [2,25], the lattice or basis size is significantly larger than the number of particles, $N_s \gg N_p$, so that it is advantageous to be able to calculate the Green functions (and certain correlation functions) with computational cost of $O(N_s N_p^2)$ versus $O(N_s^2)$. Even if the calculation of the full Green function matrix would still require the latter, generally we are interested in in a subset of them, not all the elements. Moreover, as we will show below, the method we are proposing allows one to extend the calculation in a straightforward way to two-body dynamical correlations without affecting the complexity.

**A. Particle excitations**

We consider the estimator in Eq. (18) in a more general sense:

$$g_{\nu,\mu}(\mathbf{X},\tau) = \frac{\langle \phi_L | \hat{c}_\nu \hat{B}_{N_s} \cdots \hat{B}_1 \hat{c}_\mu | \phi_R \rangle}{\langle \phi_L | \hat{B}_{N_s} \cdots \hat{B}_1 | \phi_R \rangle}, \quad (38)$$
\[ |\phi_R^{N_p+1}\rangle = \hat{c}_\mu^\dagger |\phi_R^N\rangle, \]

where \( |\mu\rangle \) is the orbital after projecting out the linear dependence on \(|u_i\rangle\):

\[ |\tilde{\mu}\rangle = \frac{\mu - \sum_{u_i} \langle u_i | \mu \rangle |u_i\rangle}{\| \mu - \sum_{u_i} \langle u_i | \mu \rangle |u_i\rangle \|}, \]

with \( D_{N_p+1} = (\tilde{\mu} | \mu \rangle \). Now, if a propagator \( \hat{B} \) is applied, we have

\[ \hat{B} |\phi_R^N\rangle = |\phi_R^N\rangle \frac{D}{D}, \]

with

\[ |\phi_R^N\rangle = \hat{c}_{u_i^1} \cdots \hat{c}_{u_{N_p^p}}^\dagger |0\rangle, \]

where the orbitals \( |u_i^1\rangle, \ldots, |u_{N_p^p}\rangle \) form an orthonormal set and \( D \) is the factor arising from the GS decomposition, and

\[ \hat{B} |\phi_R^{N_p+1}\rangle = \left| \phi_R^{N_p+1}\right\rangle \frac{D}{D} D_{N_p+1} D_{N_p+1}, \]

with

\[ |\phi_R^{N_p+1}\rangle = \hat{c}_\mu^\dagger |\phi_R^N\rangle = \hat{c}_\mu^\dagger \hat{c}_{u_i^1} \cdots \hat{c}_{u_{N_p^p}}^\dagger |0\rangle, \]

where, as before, \( |\tilde{\mu}\rangle \) is the orbital after projecting out the linear dependence on \(|u_i^1\rangle\), and \( D_{N_p+1} = (\tilde{\mu} | \hat{B} \tilde{\mu} \rangle \). This procedure can be iterated for the product of \( \hat{B} \) in Eq. (38), propagating the single-particle orbitals and accumulating the weights \( D' \). As in regular AFQMC for static observables, the GS reorthonormalization need not be applied at every iteration, only with a frequency sufficient to ensure numerical stability [24]. The evaluation of the Green function element requires the calculation of

\[ g_{\nu\mu}(X, \tau) = \frac{\langle \phi_L^N | \hat{c}_\nu^\dagger | \phi_R^{N_p+1}\rangle}{\langle \phi_L^N | \phi_R^N\rangle}. \]
Sec. II B and the one computed with the present method. The perfect agreement shows the robustness of both approaches. As mentioned, the savings from the present method occur when a small number of Green function elements are targeted. When the whole Green function is needed, the approach becomes more computationally expensive. Also, the advantage relies on $N_p \ll N_c$, so that at half filling as in Fig. 11, for example, there is no advantage over the previous method. On the other hand, in the dilute limit the reduction in computational cost and potentially in memory requirement as well, since the propagated determinants are of size $O(N_c N_p)$ vs $O(N_c^2)$ for $G$ can be dramatic, such as in systems of cold atoms [26] and especially with spin-orbit coupling present [45] which further increases the basis size.

B. Hole excitations

Suppose now we wish to compute the estimator

$$h_{\nu,\mu}(X,\tau) = \frac{\langle \phi_L | \hat{c}_\mu | \hat{B}_N \cdots \hat{B}_1 \hat{c}_\nu | \phi_R \rangle}{\langle \phi_L | \hat{B}_N \cdots \hat{B}_1 | \phi_R \rangle}. \quad (54)$$

We again consider $\hat{c}_\mu$ acting on $|\phi_R\rangle$:

$$|\phi_R^{N_p-1} \rangle = \hat{c}_\mu |\phi_R\rangle = \hat{c}_\mu \hat{c}_{u_{N_p}} \cdots \hat{c}_{u_1} |0\rangle,$$  \quad (55)

which is no longer a single Slater determinant, but a linear combination of $N_p$ Slater determinants. However, because these determinants all have the same structure, each containing $(N_p - 1)$ orbitals taken from $|\phi_R\rangle$, their propagation by a propagator $\hat{B}$ can be handled together in a convenient way.

We rewrite Eq. (55) as

$$|\tilde{\mu}\rangle = \sum_{\sigma=1}^{N_p} |\mu\rangle |u_\sigma\rangle |u_\sigma\rangle.$$

We then consider two-body correlation functions $\tilde{\Phi}_{\alpha,\beta}$ as in Eqs (45) and (46), we can rewrite

$$\hat{B} |\tilde{\phi}_R^{N_p-1} \rangle = \hat{c}_\tilde{\mu} \hat{c}_{\tilde{u}_{N_p}} \cdots \hat{c}_{\tilde{u}_1} |0\rangle,$$  \quad (66)

where $|\tilde{\mu}\rangle$ evolves under the action of $(\hat{B}^{-1})$, since it corresponds to a destruction operator.

The application of $\hat{\tilde{B}}$ on the orbitals $|\tilde{\mu}\rangle$ evolves under the action of $(\hat{B}^{-1})$, since it corresponds to a destruction operator.

After reorthonormalization of $\hat{\tilde{B}}|\tilde{\phi}_R\rangle$ as in Eqs (45) and (46), we can rewrite

$$\hat{\tilde{B}} |\tilde{\phi}_R^{N_p-1} \rangle = |\tilde{\phi}_R^{N_p-1} \rangle \mathcal{D},$$

with

$$|\tilde{\phi}_R^{N_p-1} \rangle = \hat{c}_\tilde{\mu} |\tilde{\phi}_R^N \rangle = \hat{c}_\tilde{\mu} \hat{c}_{\tilde{u}_{N_p}} \cdots \hat{c}_{\tilde{u}_1} |0\rangle,$$  \quad (67)

where $|\tilde{\mu}\rangle$ is the projection of the orbital $(\hat{B}^{-1})$ onto the linear space of $|\mu\rangle$, as in Eq. (57). The evaluation of the hole Green function in Eq. (54) now becomes

$$h_{\nu,\mu}(X,\tau) = \frac{\langle \phi_L | \tilde{\phi}_R^{N_p-1} | \phi_R \rangle}{\langle \phi_L | \phi_R \rangle} = \frac{\langle \phi_L | \tilde{\phi}_R^{N_p-1} | \phi_R \rangle}{\langle \phi_L | \phi_R \rangle}. \quad (68)$$

Simple algebraic manipulations in the matrix representation of Eqs. (51) and (52) allow us to express the matrix element on the right as

$$\sum_{\alpha,\beta=1}^{N_p} |\tilde{\mu}\rangle |u'_\alpha\rangle (u'_{\alpha} |v\rangle) |\Phi_L\rangle^* |\Phi_R^{N_p-1}\rangle^{-1}. \quad (69)$$

C. Two-body correlation functions

Suppose we wish to compute a density-density or a spin-spin correlation function. In both cases, we have to compute

$$n(X,\tau) = \frac{\langle \phi_L | \hat{n}_{\nu,\mu} B \cdots B \hat{\tilde{B}} \hat{n}_{\nu,\mu} | \phi_R \rangle}{\langle \phi_L | \hat{B}_N \cdots \hat{B}_1 | \phi_R \rangle}, \quad (69)$$

where $\hat{n}_{\nu,\mu}$ is the fermion density operator. In this case, we will write explicitly the spin index.

We will exploit the following exact property:

$$\hat{n}_{\nu,\mu} = \frac{\hat{\tilde{B}}_{\nu,\mu} - 1}{e - 1}, \quad (66)$$

where $\hat{\tilde{B}}_{\nu,\mu}$ is defined in Eq. (65). The application of the one-body propagator $e^{i\tilde{B}}$ on the orbitals can be carried out straightforwardly. Now, the estimator Eq. (63) can be broken into two pieces:

$$n(X,\tau) = \frac{1}{e - 1} [n_1(X,\tau) - n_2(X,\tau)], \quad (67)$$

which can be conveniently expressed as

$$n_1(X,\tau) = \frac{\langle \phi_L | \hat{\tilde{B}} \hat{n}_{\nu,\mu} | \phi_R \rangle}{\langle \phi_L | \hat{B} | \phi_R \rangle} \frac{\langle \phi_L | \hat{B} | \phi_R \rangle}{\langle \phi_L | \hat{B} | \phi_R \rangle} \frac{\langle \phi_L | \hat{B} | \phi_R \rangle}{\langle \phi_L | \hat{B} | \phi_R \rangle} \frac{\langle \phi_L | \hat{B} | \phi_R \rangle}{\langle \phi_L | \hat{B} | \phi_R \rangle}. \quad (68)$$
V. CONCLUSIONS

We study the calculation of dynamical correlation functions in imaginary time using auxiliary-field quantum Monte Carlo. The use of twisted-boundary conditions is systematically explored. One-body finite-size corrections help improve the convergence to the thermodynamic limit. We study the role of special twists which correspond to the minimum corrected gap, and show that this dramatically reduces finite-size effects. In the repulsive Hubbard model at half filling, the charge gaps and spectral functions are computed for different interaction strengths. Much higher accuracy was reached than previously possible, especially for small gap values. We propose an alternative approach to compute the imaginary-time Green functions by explicitly varying the number of particles in the QMC random walk. This method has several advantages, including a much more favorable computational cost in “dilute” systems where the size of the basis is significantly larger than the number of fermions (from cubic to linear in lattice/basis size).

ACKNOWLEDGMENTS

We thank Huy Nguyen for his contributions at early stages of this work. This work was supported by the NSF (Grant No. DMR-1409510), E.V. and S.Z. were also supported by the Simons Foundation. Computing was carried out at the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation Grant No. ACI-1053575, and the computational facilities at William and Mary.