6 Auxiliary-Field Quantum Monte Carlo at Zero- and Finite-Temperature

Shiwei Zhang

Center for Computational Quantum Physics

Flatiron Institute, New York, NY10010, USA

Department of Physics, College of William & Mary

Williamsburg, VA23185, USA

Contents

1 Introduction

Auxiliary-field quantum Monte Carlo (AFQMC) is increasingly becoming a general and powerful tool for studying interacting many-fermion systems, in different sub-areas including the study of correlated electron models, cold Fermi gas, electronic structure of solids, and quantum chemistry. My lecture will give an introduction to the advances, both in the zero-temperature or ground-state form and in the non-zero-temperature ($T > 0$ K) form, which have allowed AFQMC to become a systematic and scalable tool. The ground-state portion of the lecture will draw (or even copy!) from Refs. [1] and [2]; the $T > 0$ K portion is based on Refs. [3] and [4].

As we have seen repeatedly echoed through this school, the accurate treatment of interacting quantum systems is one of the grand challenges in modern science. Explicit solution of the many-body Schrödinger equation leads to rapidly growing computational cost as a function of system size (see, e.g., [5]). To circumvent the problem, most computational quantum mechanical studies of large, realistic systems rely on simpler independent-particle approaches based on density-functional theory (DFT) (see, e.g., [6, 7]), using an approximate energy functional to include many-body effects. These replace the electron-electron interaction by an effective potential, thereby reducing the problem to a set of one-electron equations. Methods based on DFT and through its Car-Parrinello molecular dynamics implementation [8] have been extremely effective in complex molecules and solids [6]. Such approaches are the standard in electronic structure, widely applied in condensed matter, quantum chemistry, and materials science.

Despite the tremendous successes of DFT, the treatment of electronic correlation is approximate. For strongly correlated systems (e.g., high-temperature superconductors, heavy-fermion metals, magnetic materials, optical lattices), where correlation effects from particle interaction crucially modify materials properties, the approximation can lead to qualitatively incorrect results. Even in moderately correlated systems when the method is qualitatively correct, the results are sometimes not sufficiently accurate. For example, in ferroelectric materials the usually acceptable 1% errors in DFT predictions of the equilibrium lattice constant can lead to almost full suppression of the ferroelectric order. Additional challenges are present to go beyond ground state and account for thermal as well as quantum fluctuations.

The development of alternatives to independent-particle theories is therefore of paramount fundamental and practical significance. To accurately capture the quantum many-body effects, the size of the Hilbert space involved often grows exponentially. Simulation methods utilizing Monte Carlo (MC) sampling [9–14] are, in principle, both non-perturbative and well-equipped to handle details and complexities in the external environment. They are a unique combination of accuracy, general applicability, favorable scaling (low-power) of computational cost with physical system size, and scalability on parallel computing platforms [15].

For fermion systems, however, a so-called sign problem [16–18] arises in varying forms in these MC simulation methods. The Pauli exclusion principle requires that the states be anti-symmetric under interchange of two particles. As a consequence, negative signs appear, which cause cancellations among contributions of the MC samples of the wave function or density matrix. In some formalisms, as we discuss below, a phase appears which leads to a continuous

degeneracy and more severe cancellations. As the temperature is lowered or the system size is increased, such cancellation becomes more and more complete. The net signal thus decays *exponentially* versus noise. The algebraic scaling is then lost, and the method breaks down.

In AFQMC, we cast the MC random walks in a space of over-complete Slater determinants, which significantly reduces the severity of the sign problem. In this space we study the properties of the paths and derive exact boundary conditions for the sign of their contributions in the ground-state wave function or $T > 0$ K density matrix. This then allows us to formulate approximate constraints on the random walk paths which are less sensitive to the details of the constraint. We then develop internal checks and constraint release methods to systematically improve the approach. These methods have come under the name of constrained path Monte Carlo (CPMC) [19] for systems where there is a sign problem (for example, Hubbard-like models where the auxiliary-fields are real due to the short-ranged interactions). For electronic systems where there is a phase problem (as the Coulomb interaction leads to complex auxiliary fields), the methods have been referred to as phaseless or phase-free AFQMC [14, 20, 21]. We will refer to the methods as AFQMC following more recent literature in *ab initio* electronic structure. When it is necessary to emphasize the constrained-path (CP) approximation, to distinguish from unconstrained zero-temperature (free-projection) or $T > 0$ K (sometimes referred to as determinantal MC) calculations, we will refer to them as CP-AFQMC.

2   Formalism

The Hamiltonian for any many-fermion system with two-body interactions (e.g., the electronic Hamiltonian under the Born-Oppenheimer approximation) can be written as

$$\hat{H} = \hat{H}_1 + \hat{H}_2 = -\frac{\hbar^2}{2m} \sum_{m=1}^{M} \nabla_m^2 + \sum_{m=1}^{M} V_{\text{ext}}(\mathbf{r}_m) + \sum_{m<n}^{M} V_{\text{int}}(\mathbf{r}_m - \mathbf{r}_n), \qquad (1)$$

where $\mathbf{r}_m$ is the real-space coordinate of the m-th fermion. The one-body part of the Hamiltonian, $\hat{H}_1$, consists of the kinetic energy of the electrons and the effect of the ionic (and any other external) potentials. (We have represented the external potential as local, although this does not have to be the case. For example, in plane-wave calculations we will use a norm-conserving pseudopotential, which will lead to a non-local function V_{ext}.) The two-body part of the Hamiltonian, $\hat{H}_2$, contains the electron-electron interaction terms. The total number of fermions, M, will be fixed in the ground-state calculations; for $T > 0$ K, a chemical potential term will be included and the number of fermions will fluctuate. For simplicity, we have suppressed the spin-index; when the spins need to be explicitly distinguished, M_σ will denote the number of electrons with spin σ ($\sigma = \uparrow$ or $\downarrow$). We assume that the interaction is spin-independent, so the total S_z, defined by $(M_\uparrow - M_\downarrow)$, is fixed in the ground-state calculation, although it will be straightforward to generalize our discussions to treat other cases, for example, when there is spin-orbit coupling (SOC) [22].

With any chosen one-particle basis, the Hamiltonian can be written in second quantization in the general form

$$\hat{H} = \hat{H}_1 + \hat{H}_2 = \sum_{i,j}^{N} T_{ij} c_i^\dagger c_j + \frac{1}{2} \sum_{i,j,k,l}^{N} V_{ijkl} c_i^\dagger c_j^\dagger c_k c_l \,, \tag{2}$$

where the one-particle basis, $\{|\chi_i\rangle\}$ with $i = 1, 2, \ldots, N$, can be lattice sites (Hubbard model), plane-waves (as in solid state calculations) [23], or Gaussians (as in quantum chemistry) [20, 24], etc. The operators $c_i^\dagger$ and c_i are creation and annihilation operators on $|\chi_i\rangle$, satisfying standard fermion commutation relations. The one-body matrix elements, T_{ij}, contain the effect of both the kinetic energy and external potential. For the $T > 0\,\mathrm{K}$ calculations we will discuss, a term $\mu\,\hat{n}$ containing the chemical potential μ and density operator is included [12,25,26]. The two-body matrix elements, V_{ijkl}, are from the interaction. The matrix elements are expressed in terms of the basis functions, for example,

$$V_{ijkl} = \int d\mathbf{r}_1 d\mathbf{r}_2 \chi_i^*(\mathbf{r}_1)\chi_j^*(\mathbf{r}_2) V_{\mathrm{int}}(\mathbf{r}_1 - \mathbf{r}_2)\chi_k(\mathbf{r}_2)\chi_l(\mathbf{r}_1) \,. \tag{3}$$

In quantum chemistry calculations, these are readily evaluated with standard Gaussian basis sets. In solid state calculations with plane-waves, the kinetic and electron-electron interaction terms have simple analytic expressions, while the electron-ion potential leads to terms which are provided by the pseudopotential generation. We will assume that all matrix elements in Eq. (2) have been evaluated and are known as we begin our many-body calculations.

2.1 Hubbard-Stratonovich transformation

The two-body part in the Hamiltonian in Eq. (2), $\hat{H}_2$, can be written in the following form

$$\hat{H}_2 = \frac{1}{2} \sum_{\gamma=1}^{N_\gamma} \lambda_\gamma \hat{v}_\gamma^2 \,, \tag{4}$$

where λ_γ is a constant, $\hat{v}_\gamma$ is a one-body operator similar to $\hat{H}_1$, and N_γ is an integer. There are various ways to achieve the decomposition in Eq. (4) for a general two-body term [27]. Below we outline the two most commonly applied cases in electronic structure: (**a**) *with plane-wave basis* and (**b**) for a more dense matrix V_{ijkl} resulting from a *general basis set* such as Gaussians. In a *plane-wave basis*, the two-body part is the Fourier transform of $1/|\mathbf{r}_m - \mathbf{r}_n|$ [23]

$$\hat{H}_2 \rightarrow \frac{1}{2\Omega} \sum_{i,j,k,l} \frac{4\pi}{|\mathbf{G_i} - \mathbf{G_k}|^2} \, c_i^\dagger c_j^\dagger c_l c_k \, \delta_{\mathbf{G_i} - \mathbf{G_k}, \mathbf{G_l} - \mathbf{G_j}} \, \delta_{\sigma_i, \sigma_k} \, \delta_{\sigma_j, \sigma_l} \,, \tag{5}$$

where $\{\mathbf{G_i}\}$ are planewave wave-vectors, Ω is the volume of the supercell, and σ denotes spin. Let us use $\mathbf{Q} \equiv \mathbf{G_i} - \mathbf{G_k}$, and define a density operator in momentum space

$$\hat{\rho}(\mathbf{Q}) \equiv \sum_{\mathbf{G}, \sigma} c_{\mathbf{G}+\mathbf{Q}, \sigma}^\dagger c_{\mathbf{G}, \sigma}, \tag{6}$$

where the sum is over all $\mathbf{G}$ vectors which allow both $\mathbf{G}$ and $\mathbf{G} + \mathbf{Q}$ to fall within the pre-defined kinetic energy cutoff, E_{cut}, in the planewave basis. The two-body term in Eq. (5) can then be manipulated into the form

$$\hat{H}_2 \to \sum_{\mathbf{Q} \neq \mathbf{0}} \frac{\pi}{\Omega Q^2} \left[\hat{\rho}^\dagger(\mathbf{Q}) \, \hat{\rho}(\mathbf{Q}) + \hat{\rho}(\mathbf{Q}) \, \hat{\rho}^\dagger(\mathbf{Q}) \right], \tag{7}$$

where the sum is over all $\mathbf{Q}$'s except $\mathbf{Q} = 0$, since in Eq. (5) the $\mathbf{G_i} = \mathbf{G_k}$ term is excluded due to charge neutrality, and we have invoked $\rho^\dagger(\mathbf{Q}) = \rho(-\mathbf{Q})$. By making linear combinations of $\left[(\rho^\dagger(\mathbf{Q}) + \rho(\mathbf{Q}) \right]$ and $\left[(\rho^\dagger(\mathbf{Q}) - \rho(\mathbf{Q}) \right]$ terms, we can then readily write the right-hand side in Eq. (7) in the desired square form of Eq. (4) [23].

With a *general basis* (e.g., Gaussians basis as standard in chemistry), the most straightforward way to decompose $\hat{H}_2$ is through a direct diagonalization [20, 28, 2]. However, this is computationally costly. A modified Cholesky decomposition leads to $\mathcal{O}(N)$ fields [24, 21]. This approach, which has been commonly used in AFQMC for molecular systems with Gaussian basis sets and for downfolded Hamiltonians [29], realizes the following

$$V_{ijkl} = V_{(i,l),(j,k)} = V_{\mu\nu} \doteq \sum_{\gamma=1}^{N_\gamma} L_\mu^\gamma L_\nu^\gamma, \tag{8}$$

where $\mu = (i, l)$ and $\nu = (j, k)$ are composite indices introduced for convenience. The process is carried out recursively using a modified Cholesky algorithm [30, 31]. Recently an alternative with density-fitting has also been used [32]. The 4-index matrix elements can be represented to a pre-defined accuracy tolerance δ (which for molecular calculations is typical of δ range between 10^{-4} and 10^{-6} in atomic units [21]), such that

$$\hat{H}_2 \to \frac{1}{2} \sum_{\gamma=1}^{N_{\text{CD}}} \left(\sum_{il} L_{\mu(i,l)}^\gamma c_i^\dagger c_l \right) \left(\sum_{jk} L_{\nu(j,k)}^\gamma c_j^\dagger c_k \right) + \mathcal{O}(\delta). \tag{9}$$

Hence the form in Eq. (4) is realized, with $\hat{v}_\gamma = \sum_{il} L_{\mu(i,l)}^\gamma c_i^\dagger c_l$. The process decomposing the two-body interaction is illustrated in Fig. 1.
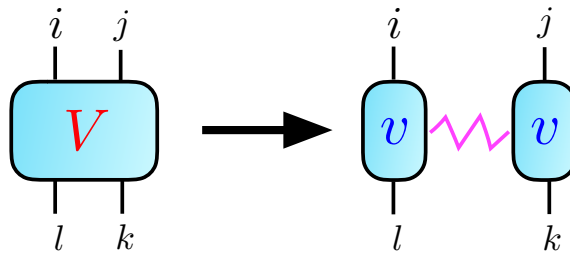

Fig. 1: *Schematic illustration of the decoupling of the two-body interaction, either via Cholesky decomposition, the planewave factorization, or density fitting. The number of auxiliary index γ (magenta wiggly line) controls the number of auxiliary fields.*

We can then apply the Hubbard-Stratonovich (HS) transformation to each term in Eq. (4)

$$e^{-\frac{\Delta\tau}{2}\lambda\hat{v}^2} = \int_{-\infty}^{\infty} dx \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} e^{x\sqrt{-\Delta\tau\lambda}\,\hat{v}}, \tag{10}$$

where x is an auxiliary-field variable. The key idea is that the quadratic form (in $\hat{v}$) on the left is replaced by a linear one on the right. If we denote the collection of auxiliary fields by $\mathbf{x}$ and combine one-body terms from $\hat{H}_1$ and $\hat{H}_2$, we obtain the following compact representation of the outcome of the HS transformation

$$e^{-\Delta\tau\hat{H}} = \int d\mathbf{x}\, p(\mathbf{x})\hat{B}(\mathbf{x}), \tag{11}$$

where $p(\mathbf{x})$ is a probability density function (PDF), for example, a multi-dimensional Gaussian. The propagator $\hat{B}(\mathbf{x})$ in Eq. (11) has a special form, namely, a product of operators of the type

$$\hat{B} = \exp\Big(\sum_{ij} c_i^\dagger U_{ij} c_j\Big), \tag{12}$$

with U_{ij} depending on the auxiliary field. The matrix representation of $\hat{B}(\mathbf{x})$ will be denoted by $B(\mathbf{x})$. See Appendix B for more details.

Note that the constant in front of $\hat{v}$ in the exponent on the right-hand side of Eq. (10) can be real or imaginary. So the matrix elements of $B(\mathbf{x})$ can become complex, for example when λ in Eq. (10) is positive, which occurs in both of the forms discussed above. Sometimes we will refer to this situation as having complex auxiliary fields, but it should be understood that this is interchangeable with $\hat{B}(\mathbf{x})$ being complex, and the relevant point is whether the Slater determinant develops complex matrix elements which evolve stochastically.

In essence, the HS transformation replaces the two-body interaction by one-body interactions with a set of random external auxiliary fields. In other words, it converts an interacting system into many *non-interacting* systems living in fluctuating external auxiliary-fields. The sum over all configurations of auxiliary fields recovers the interaction.

Different forms of the HS transformation can affect the performance of the AFQMC method. For example, it is useful to subtract a mean-field "background" from the two-body term prior to the decomposition [33, 34, 20]. The idea is that using the HS to decompose any constant shifts in the two-body interaction will necessarily result in more statistical noise. In fact, it has been shown [35, 21] that the mean-field background subtraction can not only impact the statistical accuracy, but also lead to different quality of approximations under the CP methods that we discuss in the next section.

2.2 Ground-state projection

Most ground-state quantum MC (QMC) methods are based on iterative projection

$$|\Psi_0\rangle \propto \lim_{\tau\to\infty} e^{-\tau\hat{H}}|\Psi_T\rangle; \tag{13}$$

that is, the ground state $|\Psi_0\rangle$ of a many-body Hamiltonian $\hat{H}$ can be projected from any known trial state $|\Psi_T\rangle$ that satisfies $\langle \Psi_T|\Psi_0\rangle \neq 0$. This can be achieved by numerical iteration

$$|\Psi^{(n+1)}\rangle = e^{-\Delta\tau\hat{H}}|\Psi^{(n)}\rangle, \tag{14}$$

where $|\Psi^{(0)}\rangle = |\Psi_T\rangle$. The ground-state expectation $\langle \hat{O}\rangle$ of a physical observable $\hat{O}$ is given by

$$\langle \hat{O}\rangle = \lim_{n\to\infty} \frac{\langle \Psi^{(n)}|\hat{O}|\Psi^{(n)}\rangle}{\langle \Psi^{(n)}|\Psi^{(n)}\rangle}. \tag{15}$$

For example, the ground-state energy can be obtained by letting $\hat{O} = \hat{H}$. A so-called mixed estimator exists, however, which is exact for the energy (or any other $\hat{O}$ that commutes with $\hat{H}$) and can lead to considerable simplifications in practice

$$E_0 = \lim_{n\to\infty} \frac{\langle \Psi_T|\hat{H}|\Psi^{(n)}\rangle}{\langle \Psi_T|\Psi^{(n)}\rangle}. \tag{16}$$

QMC methods carry out the iteration in Eq. (14) by Monte Carlo (MC) sampling. At the simplest conceptual level, one can understand the difference between different classes of QMC methods as what space is used to represent the wave function or density matrix and to carry out the integration. The AFQMC methods work in second quantized representation and in a non-orthogonal space of Slater determinants. Ground-state AFQMC represents the many-body wave function stochastically in the form

$$|\Psi^{(n)}\rangle = \sum_\phi \alpha_\phi|\phi\rangle\,, \tag{17}$$

where $|\phi\rangle$ is a Slater determinant

$$|\phi\rangle \equiv \hat{\varphi}_1^\dagger \hat{\varphi}_2^\dagger \cdots \hat{\varphi}_M^\dagger|0\rangle\,. \tag{18}$$

The Slater determinants evolve with n via rotations of the orbitals, as do their coefficients, which are represented by the weights in the MC sampling. The $T > 0$ AFQMC formalism, as we will see later, is closely related and works in the same space.

From Thouless' theorem, the operation of $e^{-\tau\hat{H}_1}$ on a Slater determinant simply yields another determinant. Thus for an independent-particle Hamiltonian, where $\hat{H}_2$ is replaced by a one-body operator, the ground-state projection would therefore turn into the propagation of a single Slater determinant. In DFT, for example, under the local density approximation (LDA), $\hat{H}_2$ is replaced by $\hat{H}_{\text{LDA}} = \hat{H}_1 + \hat{V}_{xc}$, where $\hat{V}_{xc}$ contains the density operator in real-space, with matrix elements given by the exchange-correlation functional which is computed with the local density from the current Slater determinant in the self-consistent process. An iterative procedure can be carried out, following Eq. (14), to project the solution using the approximate Hamiltonians, as an imaginary-time evolution of a single Slater determinant [36]. This is illustrated by the blue line in Fig. 2. Note that this procedure is formally very similar to time-dependent DFT (TDDFT), except for the distinction of imaginary versus real time.

With Eq. (11), we can now turn the many-body projection into a linear combination of iterative projections, mimicking the evolution of an ensemble of the corresponding non-interacting systems subjected to fluctuating external (auxiliary) fields. For simplicity, we will take $|\Psi^{(0)}\rangle$ (i.e., $|\Psi_T\rangle$) as a single Slater determinant. Using the mixed estimator in Eq. (16), we can visualize the calculation of the energy as carrying out the projection of the ket $|\Psi^{(n)}\rangle$ by an open-ended random walk. Aside from technical steps such as importance sampling, population control, numerical stabilization, etc. to make this process practical and more efficient [19, 2], we could think of the calculation at the conceptual level as: (i) start a population of walkers $\{|\phi_k^{(0)}\rangle\}$ with $k = 1, \ldots, N_w$ labeling the walkers, (ii) sample an $\mathbf{x}$ from $p(\mathbf{x})$ for each walker, (iii) propagate it by $B(\mathbf{x})$, (iv) sweep through the population to advance to $n=1$, and repeat steps (ii) and (iii) to iterate n. The ideas are illustrated in Fig. 2. [Note that, if we use a linear combination of Slater determinants of the form of Eq. (17) for $|\Psi^{(0)}\rangle$, we can sample the determinants to initialize the population in (i).] In Appendix A we include a brief review of MC and random walks.

We now expand on the formalism a bit more to make a more explicit connection with the $T > 0$ formalism that we will discuss in Sec. 2.3. There we will look more deeply into the origin of the sign problem. By making the formal connection, the discussion on the sign problem at finite-T can be straightforwardly connected to the ground-state situation here. For computing
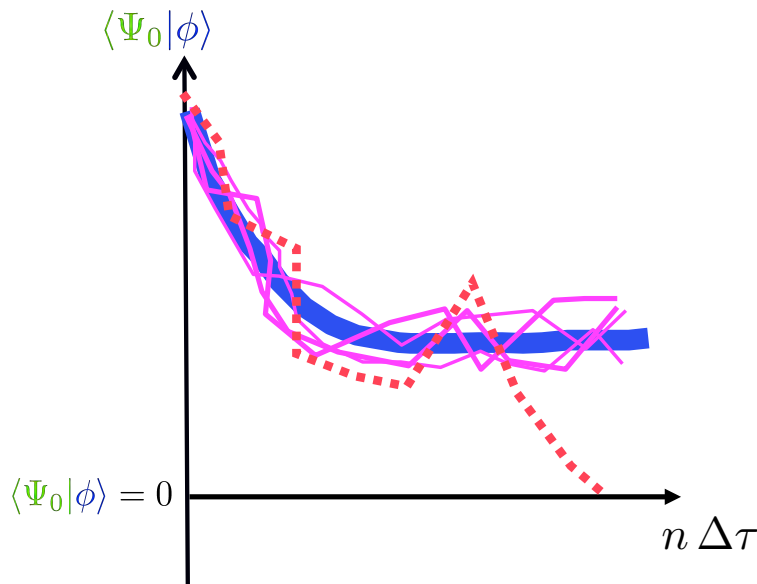

Fig. 2: *Illustration of the iterative imaginary-time projection to the ground state. The overlap of the Slater determinants with a test wave function (e.g., the exact ground state $|\Psi_0\rangle$) is plotted vs. imaginary time $n\Delta\tau$. The thick blue line indicates a projection using $e^{-\Delta\tau \hat{H}_{\mathrm{LDA}}(\phi^{(n)})}$ which converges to the LDA ground state (or a local minimum). The wiggly magenta lines indicate an AFQMC projection which captures the many-body effect beyond LDA as a stochastic linear superposition. The propagator is obtained by expanding the two-body part of the $\hat{H}$, namely $\hat{H}_2 - \hat{V}_{xc}$, by a Hubbard-Stratonovich transformation as discussed in Sec. 2.1. The dotted redline indicates a path which can lead to a sign problem (Sec. 3.1).*

the ground-state expectation $\langle \hat{O} \rangle$ in Eq. (15), the denominator is

$$
\begin{aligned}
\langle \psi^{(0)} | e^{-n\Delta\tau\hat{H}} \, e^{-n\Delta\tau\hat{H}} | \psi^{(0)} \rangle &= \int \langle \psi^{(0)} | \prod_{l=1}^{2n} d\mathbf{x}^{(l)} p(\mathbf{x}^{(l)}) \, \hat{B}(\mathbf{x}^{(l)}) | \psi^{(0)} \rangle \\
&= \int \prod_l d\mathbf{x}^{(l)} p(\mathbf{x}^{(l)}) \, \det\left([\Psi^{(0)}]^\dagger \prod_l B(\mathbf{x}^{(l)}) \Psi^{(0)} \right). \quad (19)
\end{aligned}
$$

In the early forms of ground-state AFQMC calculations [13, 37] (which in the presence of a sign problem are transient estimates that have exponential scaling), a value of n is first chosen and fixed throughout the calculation. If we use X to denote the collection of the auxiliary-fields $X = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(2n)}\}$ and $D(X)$ to represent the integrand in Eq. (19), we can write an estimator of the expectation value of Eq. (15) as

$$
\langle \hat{O} \rangle = \frac{\int \overline{\langle \hat{O} \rangle} \, D(X) \, dX}{\int D(X) \, dX} = \frac{\int \overline{\langle \hat{O} \rangle} \, |D(X)| \, s(X) \, dX}{\int |D(X)| \, s(X) \, dX}, \quad (20)
$$

where

$$
s(X) \equiv D(X)/|D(X)| \quad (21)
$$

measures the "sign" of $D(X)$, and $\langle s \rangle = \langle s(X) \rangle_{|D|}$ gives an indication of the severity of the sign problem, as further discussed in Sec. 3.1 under the $T > 0$ formalism. The non-interacting expectation for a given X is $\overline{\langle \hat{O} \rangle} \equiv \langle \hat{O} \rangle_{\phi_L \phi_R}$ as defined in Eq. (55) in Appendix B, where

$$
\begin{aligned}
\langle \phi_L | &= \langle \psi^{(0)} | \, \hat{B}(\mathbf{x}^{(2n)}) \hat{B}(\mathbf{x}^{(2n-1)}) \cdots \hat{B}(\mathbf{x}^{(n+1)}) \\
| \phi_R \rangle &= \hat{B}(\mathbf{x}^{(n)}) \hat{B}(\mathbf{x}^{(n-1)}) \cdots \hat{B}(\mathbf{x}^{(1)}) | \psi^{(0)} \rangle,
\end{aligned}
$$

which are both Slater determinants. In Appendix B, basic properties of Slater determinants and the computation of expectation values are reviewed.

$D(X)$ as well as $\langle \phi_L |$ and $| \phi_R \rangle$ are completely determined by the path X in auxiliary-field space. The expectation in Eq. (20) is therefore in the form of Eq. (49), with $f(X) = |D(X)|$ and $g(X) = \overline{\langle \hat{O} \rangle}$. The important point is that, for each X, $|D(X)|$ is a number and $g(X)$ can be evaluated using Eqs. (56) and (57). Often the Metropolis Monte Carlo algorithm [38] is used to sample auxiliary-fields X from $|D(X)|$. Any $\langle \hat{O} \rangle$ can then be computed following the procedure described by Eq. (48) in Appendix A.

2.3   Finite-temperature, grand-canonical ensemble calculations

At a temperature $T > 0\,\mathrm{K}$, the expectation value of any physical observable

$$
\langle \hat{O} \rangle \equiv \frac{\mathrm{Tr}(\hat{O} \, e^{-\beta\hat{H}})}{\mathrm{Tr}(e^{-\beta\hat{H}})}, \quad (22)
$$

which is a weighted average in terms of the partition function Z in the denominator

$$
Z \equiv \mathrm{Tr}(e^{-\beta\hat{H}}) = \mathrm{Tr}[\underbrace{e^{-\Delta\tau\hat{H}} \cdots e^{-\Delta\tau\hat{H}} e^{-\Delta\tau\hat{H}}}_{L}], \quad (23)
$$

where $\beta = 1/kT$ is the inverse temperature, $\Delta\tau = \beta/L$ is the time-step, and L is the number of "time slices." Substituting Eq. (11) into Eq. (23) gives

$$Z = \sum_X \int dX P(X) \operatorname{Tr}[B(\mathbf{x}_L) \cdots B(\mathbf{x}_2) B(\mathbf{x}_1)], \qquad (24)$$

where $X \equiv \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L\}$ denotes a complete *path* in auxiliary-field space, and $P(X) = \prod_{l=1}^{L} p(\mathbf{x}_l)$. Note that, similar to the $T = 0\,\mathrm{K}$ case, a Trotter error can be present in Eq. (11), which is controllable (by, e.g., extrapolating results with different values of $\Delta\tau$). The trace over fermion degrees of freedom can be performed analytically [12, 39], which yields

$$Z = \int dX P(X) \det[I + B(\mathbf{x}_L) \cdots B(\mathbf{x}_2) B(\mathbf{x}_1)], \qquad (25)$$

where I is the $N \times N$ unit matrix. Note that the trace in Eq. (24), which is over numbers of particles and initial states, is now replaced by the fermion determinant

$$D(X) \equiv P(X) \det[I + B(\mathbf{x}_L) \cdots B(\mathbf{x}_2) B(\mathbf{x}_1)], \qquad (26)$$

which can be readily computed for each path X. The sum over all paths is evaluated by standard Monte Carlo techniques. Relating L to $2n$ in the ground-state discussions around Eq. (20), we can now connect these two classes of methods. Below we will rely on the finite-T form in understanding the origin of the sign problem. This will serve as a common framework for thinking about the sign problem, and then subsequently the phase problem, by analyzing the nature and behavior of the paths in X space.

The symptom of the sign problem is that $D(X)$ is not always positive. In practice, the MC samples of X are drawn from the probability distribution defined by $|D(X)|$. As β increases, $D(X)$ approaches an antisymmetric function and its average sign,

$$\langle s \rangle = \sum_X D(X) / \sum_X |D(X)|, \qquad (27)$$

vanishes exponentially, as illustrated in Fig. 3. In the limit of $T \to 0$ or $\beta \to \infty$, the distribution becomes fully antisymmetric. (In ground-state calculations discussed in Sec. 2.2, the situation depicted in Fig. 3 corresponds to the equilibration phase at shorter projection times τ. When the ground state is reached after equilibration, there will be no "green" part left.)
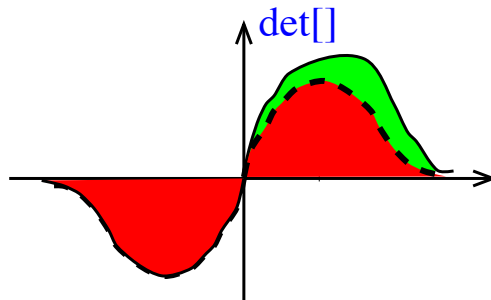

Fig. 3: *Schematic illustration of the sign problem. The horizontal axis denotes an abstraction of the many-dimensional auxiliary-field paths X. The sign problem occurs because the contributing part (shaded area) is exponentially small compared to what is sampled, namely $|D(X)|$. The origin of the symptoms shown here is explained in Sec. 3.1.*

3   Constraining the paths in AFQMC

The sign/phase problem is a generic feature, and is only absent in special cases where the single-particle propagator $\hat{B}(\mathbf{x})$ satisfies particular symmetries (see, for example, Ref. [40]). In these cases, $D(x)$ is real and non-negative, and $s(X)$ remains strictly 1. The symmetries are affected by the choice of the basis and the form of the decomposition, so it is possible to "engineer away" the sign problem in some cases. However, these are still limited to very special classes of Hamiltonians. In general, a sign problem arises if $\hat{B}(\mathbf{x})$ is real, and a phase problem arises if $\hat{B}(\mathbf{x})$ is complex.

For real $\hat{B}(\mathbf{x})$ (e.g. Hubbard-type of short-range repulsive interactions with the spin decomposition), the sign problem occurs because of the fundamental symmetry between the fermion ground-state $|\Psi_0\rangle$ and its negative $-|\Psi_0\rangle$ [18, 41]. In $T = 0$ K calculations, for any ensemble of Slater determinants $\{|\phi\rangle\}$ which gives a MC representation of the ground-state wave function, as in Eq. (17), this symmetry implies that there exists another ensemble $\{-|\phi\rangle\}$ which is also a correct representation. In other words, the Slater determinant space can be divided into two degenerate halves ($+$ and $-$) whose bounding surface $\mathcal{N}$ is defined by $\langle\Psi_0|\phi\rangle = 0$. This dividing surface is unknown. (In the cases with special symmetry mentioned above, the two sides separated by the surface are both positive.) In the illustration in Fig. 2, the surface is the horizontal axis; the "$-$" ensemble is given by the mirror images of the paths shown, i.e., by reflecting them with respect to the horizontal axis.

3.1   An exact boundary condition

To gain further insight on the origin of the sign problem, we conduct a thought experiment in which we generate all the complete paths X by L successive steps, from $\mathbf{x}_1$ to $\mathbf{x}_L$ [25]. We use the $T > 0$ K formalism, whose understanding will provide a straightforward connection to the ground-state method. We consider the contribution in Z of an individual *partial path* $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_l\}$ at step l

$$\mathcal{P}_l(\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_l\}) \equiv \mathrm{Tr}[\, \underbrace{\mathcal{B}\mathcal{B}\cdots\mathcal{B}}_{L-l}\; B(\mathbf{x}_l)\cdots B(\mathbf{x}_2)B(\mathbf{x}_1)], \qquad (28)$$

where $\mathcal{B} \equiv e^{-\Delta\tau H}$, which in general is not a single-particle propagator. In particular, we consider the case when $\mathcal{P}_l = 0$. This means that, after the remaining $L - l$ steps are finished, the collective contribution from *all* complete paths that result from the partial path will be precisely zero. In other words, complete paths whose first l elements are $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_l\}$ do not contribute in Z; the sum over all possible configurations of $\{\mathbf{x}_{l+1}, \mathbf{x}_{l+2}, \cdots, \mathbf{x}_L\}$ simply reproduces the $\mathcal{B}$'s in (28), leading to zero by definition.

Thus, in our thought experiment any partial path that reaches the axis in Fig. 4 immediately turns into noise, regardless of what it does at future l's. A complete path which is in contact with the axis at any point belongs to the "antisymmetric" part of $D(X)$ in Fig. 3, whose contributions cancel. The "noise" paths, which become an increasingly larger portion of all paths as β increases, are the origin of the sign problem.

Since $\mathcal{P}_0$ is positive and $\mathcal{P}_l$ changes continuously with l at the limit $\Delta\tau \to 0$, a complete path contributes **iff** it stays entirely *above* the axis in Fig. 4. Thus, in our thought experiment, imposition of the boundary condition (BC)

$$\mathcal{P}_1(\{\mathbf{x}_1\}) > 0, \quad \mathcal{P}_2(\{\mathbf{x}_1, \mathbf{x}_2\}) > 0, \quad \cdots, \quad \mathcal{P}_L(\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_L\}) > 0 \qquad (29)$$

will ensure all contributing complete paths to be selected while eliminating all noise paths. *The axis acts as an infinitely absorbing boundary.* A partial path is terminated and discarded as soon as it reaches the boundary. By discarding a path, we eliminate all of its future paths, including the ones that would eventually make positive contributions. The BC makes the distribution of complete paths vanish at the axis, which accomplishes complete cancellation of the negative and the corresponding positive contributions in the antisymmetric part of $D(X)$. Calculation of Z from our thought experiment remains exact.

3.2 Implementing the boundary condition and importance sampling

3.2.1 Implementation of the boundary condition (BC) at finite $\Delta\tau$

In actual simulations $\Delta\tau$ is finite and paths are defined only at a discrete set of imaginary times. The BC on the underlying continuous paths is the same, namely that the probability distribution must vanish at the axis in Fig. 4.

In Fig. 5, we illustrate how the BC is imposed under the discrete representation. The "contact" point is likely to be between time slices and not well defined, i.e., $\mathcal{P}_l$ may be zero at a non-integer value of l. To the lowest order, we can terminate a partial path when its $\mathcal{P}_l$ first turns negative. That is, we still impose Eq. (29) in our thought experiment to generate paths. In Fig. 5, this means terminating the path at $l = n$ (point B) and thereby discarding all its future paths (represented by the dashed lines 'BS...' and 'BT...').

We actually use a higher order approximation, by terminating at either $l = n - 1$ or $l = n$, i.e., either point B or point A. The probability for terminating at A is chosen such that it approaches 1 smoothly as $\mathcal{P}_{n-1} \to 0$, for example, $p_A = 1/[1 + \mathcal{P}_{n-1}/|\mathcal{P}_n|]$. If A is chosen, all future
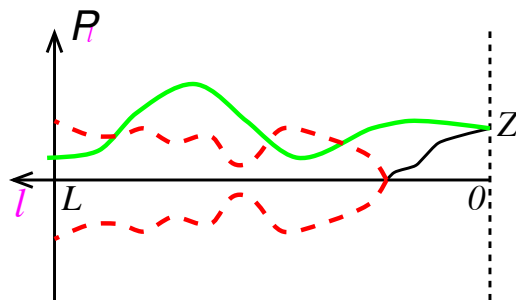

Fig. 4: *Schematic illustration the boundary condition to control the sign problem. $\mathcal{P}_l$ (Eq. (28)) is shown as a function of the length of the partial path, l, for several paths. When $\mathcal{P}_l$ becomes 0, ensuing paths (dashed lines) collectively contribute zero. Only complete paths with $\mathcal{P}_l > 0$ for all l (solid line) contribute in Z.*

paths from A are discarded (represented by 'AR...' and 'AB...'); otherwise we terminate at B as above. This is in the spirit of the so-called mirror correction (see, e.g., Ref. [42, 19]).

It is important to note that, in both approaches, the finite-$\Delta\tau$ error in imposing the BC vanishes as $\Delta\tau \to 0$. The first method, terminating at B, makes a fluctuating error of $\mathcal{O}(\Delta\tau)$ in the location of the absorbing boundary. The method we actually use ensures that the location of the boundary is correct to the next order, and is a second order algorithm.

3.2.2 Approximating the boundary condition

In practice $\mathcal{B}$ is not known. We replace it by a known trial propagator B_T. The BC now yields approximate results, which become exact if B_T is exact. If B_T is in the form of a single-particle propagator, we can evaluate Eq. (28) and combine it with Eq. (29) to obtain the approximate BC: *For each l,*

$$\mathcal{P}_l^T(\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_l\}) = \det[I + (\prod_{m=1}^{L-l} B_T)\, B(\mathbf{x}_l)\cdots B(\mathbf{x}_1)] > 0. \tag{30}$$

Different forms of $\mathcal{B}$ are possible, and of course it is valuable and important to think about improving $\mathcal{B}$. Recently self-consistent constraints formulated in ground-state AFQMC [43] have been generalized to the $T > 0$ method [4].

3.2.3 Importance sampling algorithm—automatic imposition of the constraint *and* "nudge" in the random walks

(1) Automatic imposition of the constraint. To implement the constraint in a path-integral framework of Eq. (20) at $T = 0$ K and Eq. (25) at $T > 0$ K would have severe difficulties with ergodicity. We wish to generate MC samples of X which satisfy the conditions in (30). The most natural way to accomplish this is to incorporate the boundary conditions on the path as an additional acceptance condition [44]. However, the BC is *non-local*; it also breaks translational invariance in imaginary time, since the condition selects an $l = 1$ to start. Updating the auxiliary-fields at a particular time l can cause violation of the constraint in the future (at
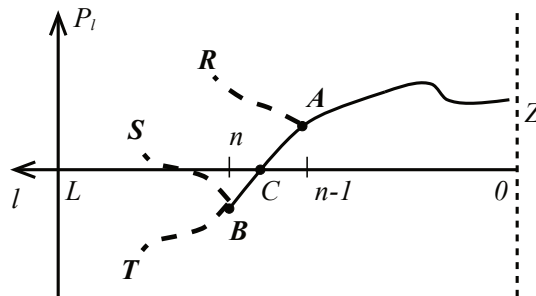

Fig. 5: *Imposition of the boundary condition at finite $\Delta\tau$. Paths are discrete. The point of contact, C (see Fig. 4), must be approximated, either by B (low order algorithm) or by interpolation between B and A (higher order).*

larger l) or in the past (when sweeping the path backwards in l). Without a scheme to propose paths that incorporates information on future contributions, it is difficult to find complete paths which satisfy all the constraints, especially as β increases. Global updating is difficult, because the number of green paths is exponentially small compared to the total number of possible paths, as illustrated in Fig. 4. The formulation of the open-ended, branching random walk approach [19, 25] in imaginary-time solves this problem. An additional advantage is that it is straightforward to project to longer imaginary-time in order to approach the ground state for $T = 0\,\mathrm{K}$. Moreover, when we carry out constraint release [35], the formalism will rely on the open-ended random walk.

(2) Nudging the random walk. The goal of the branching random walk algorithm is to generate MC samples of X which *both* satisfy the conditions in (30) *and* are distributed according to $D(X)$. The basic idea is to carry out the thought experiment stochastically. We construct an algorithm which builds directly into the sampling process both the constraints and some knowledge of the projected contribution. In terms of the *trial* projected partial contributions $\mathcal{P}_l^T \equiv \mathcal{P}_l^T(\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_l\})$ in the $T > 0\,\mathrm{K}$ method, the fermion determinant $D(X)$ can be written as

$$D(X) = \frac{\mathcal{P}_L^T}{\mathcal{P}_{L-1}^T} \frac{\mathcal{P}_{L-1}^T}{\mathcal{P}_{L-2}^T} \cdots \frac{\mathcal{P}_2^T}{\mathcal{P}_1^T} \frac{\mathcal{P}_1^T}{\mathcal{P}_0^T} \, \mathcal{P}_0^T. \tag{31}$$

As illustrated in Fig. 6, we construct the path X by a random walk of L steps, corresponding to stochastic representations of the L ratios in Eq. (31). At the l-th step, we sample $\mathbf{x}_l$ from the conditional probability density function defined by $(\mathcal{P}_l^T/\mathcal{P}_{l-1}^T)$. This allows us to select $\mathbf{x}_l$ according to the best estimate of its projected contribution in Z. In other words, we sample from a probability distribution function of the contributing paths only (solid lines in Fig. 4), instead of *all* possible paths. Note that the probability distribution for $\mathbf{x}_l$ vanishes smoothly as $\mathcal{P}_l^T$ approaches zero, naturally imposing the BC in Eq. (30) as anticipated in part (1) of this section. Ref. [4] contains a more detailed outline of the procedure.

(3) Connecting the importance sampling algorithms between $T = 0$ and $T > 0\,K$ methods. In the above we have used the $T > 0\,\mathrm{K}$ framework to describe the basic idea of importance sampling. At each step, the importance-sampling transformation uses $\mathcal{P}_l^T/\mathcal{P}_{l-1}^T$ to modify the probability distribution from which $\mathbf{x}_l$ is sampled. The ground-state version, conceptually, uses $\langle \Psi_T | \phi^{(n)} \rangle / \langle \Psi_T | \phi^{(n-1)} \rangle$. In Sec. 3.3, we show how this is actually realized when the auxiliary-fields are continuous (complex), by the use of a force bias. To connect back to the $T > 0\,\mathrm{K}$ form, all we need is to invoke the formal equivalence

$$\frac{\langle \Psi_T | \phi'(\mathbf{x}) \rangle}{\langle \Psi_T | \phi \rangle} \Longleftrightarrow \frac{\det[I + (\prod_{m=1}^{L-l} B_T)\, B(\mathbf{x}) B(\mathbf{x}_{l-1}) \cdots B(\mathbf{x}_1)]}{\det[I + (\prod_{m=1}^{L-l+1} B_T)\, B(\mathbf{x}_{l-1}) \cdots B(\mathbf{x}_1)]} \, . \tag{32}$$

On the right of Eq. (32), the finite-T path segment $\{\mathbf{x}_1, \cdots, \mathbf{x}_{l-1}\}$ has been generated already, and the step in question is l, where we wish to generate $\mathbf{x}_l$, denoted by $\mathbf{x}$.

3.3 The phaseless formalism for complex auxiliary-fields

When the many-body Hamiltonian leads to a decomposition with $\lambda > 0$ in Eq. (4), the resulting HS transformation will have complex auxiliary-fields. This is the case for the electron-electron repulsion. (As mentioned earlier, when we refer loosely to having complex auxiliary-fields, what we really mean is that the propagator resulting from the two-body Hamiltonian is complex. Incidentally, it is always possible to have real auxiliary-fields, for example by making a negative shift to the positive potential, but that simply leads to many fluctuating fields to recover a constant background, and a much more severe sign problem [14, 35].) In this situation a phase problem arises, as illustrated in Fig. 7. The random walks now extend into a complex plane, and the boundary condition discussed in Sec. 3.1 must be generalized. This generalization is not straightforward, since the orbitals (or one-particle propagators) contain random phases which are entangled with the phases in the coefficients in front of the determinants. We next describe the formalism to deal with the problem [14], using the $T = 0\,\mathrm{K}$ framework. See Eq. (32) in the previous section for how to "translate" this to $T > 0\,\mathrm{K}$.

With a continuous auxiliary-field, importance sampling is achieved with a force bias [14, 45]. To sketch a derivation we write the full many-body propagator as

$$\int \left(\frac{1}{\sqrt{2\pi}} \right)^{N_\gamma} e^{-\mathbf{x}^2/2} B(\mathbf{x})\, d\mathbf{x}, \tag{33}$$

where for formal convenience we will assume that we can combine products of one-body propagators, as in Eq. (12), into a joint form by summing the one-body operators in the exponent, or vice versa, possibly incurring additional Trotter errors which will be dealt with separately. For
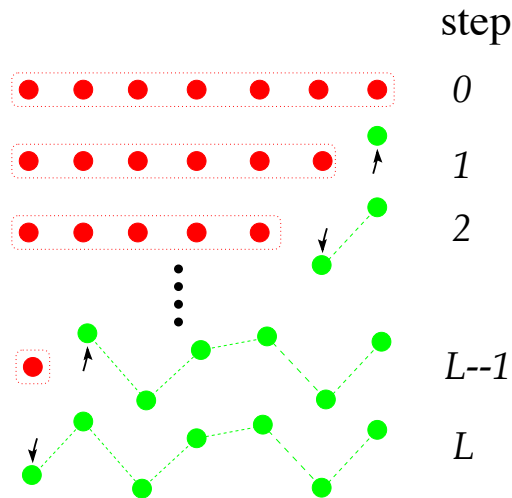

Fig. 6: *Illustration of the sampling procedure in the algorithm. Circles represent auxiliary-fields $\mathbf{x}_l$. A row shows the field configuration at the corresponding step number shown on the right. Within each row, the imaginary-time index l increases as we move to the left, i.e., the first circle is $\mathbf{x}_1$ and the last $\mathbf{x}_L$. Red circles indicate fields which are not "activated" yet, i.e., B_T is still in place of B. Green circles indicate fields that have been sampled, with the arrow indicating the one being sampled in the current step.*

compactness we will also omit the normalization factor, $(1/\sqrt{2\pi})^{N_\gamma}$, of the Gaussian probability density function from here on.

We introduce a shift in the integral in Eq. (33), which leads to an alternative propagator

$$\int e^{-\mathbf{x}^2/2}\, e^{\mathbf{x}\cdot\bar{\mathbf{x}}-\bar{\mathbf{x}}^2/2}\, B(\mathbf{x}-\bar{\mathbf{x}})\, d\mathbf{x}. \tag{34}$$

The new propagator is exact for any choice of the shift $\bar{\mathbf{x}}$, which can be complex in general.

We recall that the random walk is supposed to sample the coefficient α_ϕ in Eq. (17)

$$|\Psi_0\rangle \doteq \sum_{\{\phi\}} w_\phi |\phi\rangle . \tag{35}$$

The sum in Eq. (35) is over the population of walkers after equilibration and is over the Monte Carlo samples, typically much smaller than the sum in Eq. (17). The weight of each walker $|\phi\rangle$, w_ϕ, can be thought of as 1 (all walkers with equal weight); it is allowed to fluctuate only for efficiency considerations.

Using the idea of importance sampling, we seek to replace Eq. (35) by the following to sample Eq. (17):

$$|\Psi_0\rangle = \sum_\phi w_\phi \frac{|\phi\rangle}{\langle\Psi_T|\phi\rangle} , \tag{36}$$

where any overall phase of the walker $|\phi\rangle$ is cancelled in the numerator and denominator on the right-hand side [14]. This implies a modification to the propagator in Eq. (34):

$$\int \langle\Psi_T|\phi'(\mathbf{x})\rangle e^{-\mathbf{x}^2/2}\, e^{\mathbf{x}\bar{\mathbf{x}}-\bar{\mathbf{x}}^2/2}\, B(\mathbf{x}-\bar{\mathbf{x}}) \frac{1}{\langle\Psi_T|\phi\rangle} d\mathbf{x}, \tag{37}$$

where $|\phi'(\mathbf{x})\rangle = B(\mathbf{x}-\bar{\mathbf{x}})|\phi\rangle$ and the trial wave function $|\Psi_T\rangle$ represents the best guess to $|\Psi_0\rangle$. Now the weight of the waker under importance sampling becomes

$$w_{\phi'(\mathbf{x})} = w_\phi \frac{\langle\Psi_T|\phi'(\mathbf{x})\rangle}{\langle\Psi_T|\phi\rangle} e^{\mathbf{x}\cdot\bar{\mathbf{x}}-\bar{\mathbf{x}}^2/2} . \tag{38}$$

We can minimize the fluctuation of the factor on the right-hand side with respect to $\mathbf{x}$, by evaluating the ratio $\langle\Psi_T|\phi'(\mathbf{x})\rangle/\langle\Psi_T|\phi\rangle$ in Eqs. (37) and (38), or correspondingly, the ratio on the right-hand side of Eq. (32) for $T > 0$ K. Expanding the propagators in $\Delta\tau$, and rearranging terms [1, 21], we obtain the optimal choice for the force bias

$$\bar{\mathbf{x}} = \bar{\mathbf{v}} \equiv -\frac{\langle\Psi_T|\hat{\mathbf{v}}|\phi\rangle}{\langle\Psi_T|\phi\rangle} \sim \mathcal{O}(\sqrt{\Delta\tau}) \tag{39}$$

for ground state and a similar expectation value $\bar{\mathbf{v}}$ evaluated at time-step l for $T > 0$. The weight factor in Eq. (38) can then be manipulated into a more compact form

$$w_{\phi'(\mathbf{x})} = w_\phi \exp\big(-\Delta\tau E_L(\phi)\big) , \tag{40}$$

where E_L is the local energy of the Slater determinant

$$E_L(\phi) \equiv \frac{\langle\Psi_T|\hat{H}|\phi\rangle}{\langle\Psi_T|\phi\rangle}. \tag{41}$$
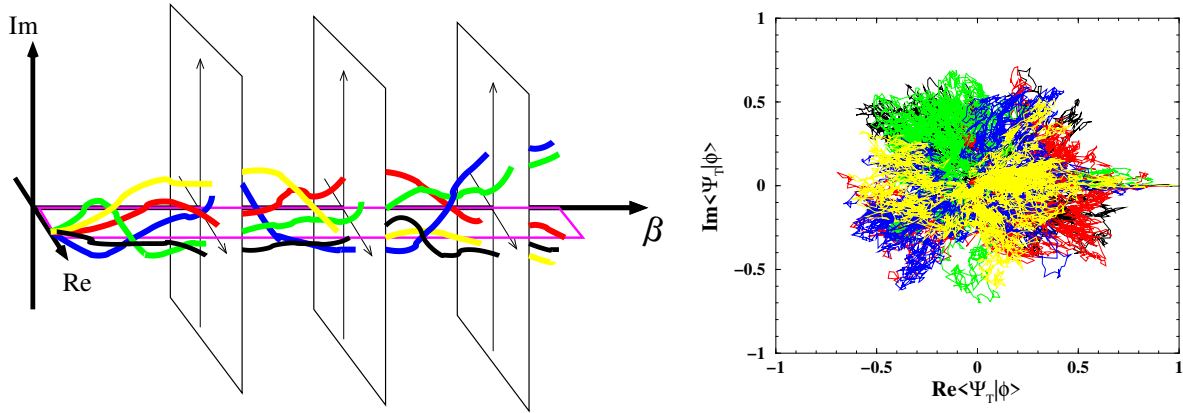
Fig. 7: *Schematic illustration of the phase problem and the constraint to control it, using the ground-state formalism. The left panel shows, as a function of projection time $\beta \equiv n\Delta\tau$, trajectories of 5 walkers characterized by the real (Re) and imaginary (Im) parts of their overlap with the ground-state wave function. The right panel shows the walker distribution integrated over imaginary time, i.e., the different frames in the left panel stacked together along β. The phase problem occurs because the initial phase "coherence" of the random walkers rapidly deteriorates with β, as they become uniformly distributed in the Re-Im-plane. The idea of the phase constraint [14] is to apply a gauge transformation such that confining the random walk in the single magenta plane (left) is a good approximation.*

We now have the full propagator of Eq. (37)

$$\int e^{-\mathbf{x}^2/2}\, B(\mathbf{x} - \bar{\mathbf{v}}) \exp\big(-\Delta\tau E_L(\phi)\big)\, d\mathbf{x}\,. \tag{42}$$

Projection with Eq. (42) will in principle lead to the ground-state wave function in the form of Eq. (36). The weight of the walker is determined by E_L, which is independent of any phase factor of the determinant. We will refer to this as the local energy form of AFQMC. As an alternative, referred to as the hybrid form, we could evaluate the weight of each walker directly according to Eq. (38) after the propagation. (In this form $\bar{\mathbf{x}}$ can in principle be anything, but of course poor choices will lead to larger fluctuations in the weights.)

In the limit of an exact $|\Psi_T\rangle$, E_L is a *real* constant, and the weight of each walker remains real. The mixed estimate for the energy from Eq. (16) is phaseless

$$E_0^{\mathrm{c}} = \frac{\sum_\phi w_\phi E_L(\phi)}{\sum_\phi w_\phi}\,. \tag{43}$$

With a general $|\Psi_T\rangle$ which is not exact, a natural approximation is to replace E_L in Eq. (42) by its real part, $\mathrm{Re}\, E_L$. The same replacement is then necessary in Eq. (43).

In early tests in the electron gas and in simple solids, we found that the phase of E_L could be carried for very long imaginary-times (and then resetting) without causing any noticeable effect on the phase problem. Keeping the phase also did not appear to affect the total energy computed from the mixed estimator. For computing observables in the ground state, back-propagation (BP) [19, 45] is needed. It was found that restoring the phases of E_L in the BP paths helped

improve the computed observables and correlations [46]. More tests for $T > 0$ are necessary but it is likely that keeping the phase will be preferable (especially because the paths are much shorter than in ground-state calculations). At $T > 0$, it was also found that restoring time-translation symmetry after the compete path has been sampled improves the results [4].

This formalism is all that is needed to handle the sign problem in the case of a *real* $\hat{\mathbf{v}}$. For any $\hat{\mathbf{v}}$ the shift $\bar{\mathbf{x}}$ diverges as a walker approaches the origin in the complex plane shown in the right panel of Fig. 7, i.e., as $\langle \Psi_T | \phi' \rangle \to 0$. The effect of the divergence is to move the walker away from the origin. With a *real* $\hat{\mathbf{v}}$, $\Delta\theta = 0$ and the random walkers move only on the real axis. If they are initialized to have positive overlaps with $|\Psi_T\rangle$, $\bar{\mathbf{x}}$ will ensure that the overlaps remain positive throughout the random walk. Thus in this case our formalism above reduces to the CP methods ground state [42, 19] and finite-T [25].

For a general case with a complex $\hat{\mathbf{v}}$, however, the phaseless formalism alone is not sufficient to remove the phase problem. To illustrate this we consider the phase of $\langle \Psi_T | \phi'(\mathbf{x} - \bar{\mathbf{x}}) \rangle / \langle \Psi_T | \phi \rangle$, or the equivalent form for $T > 0\,\mathrm{K}$ given by Eq. (32). This phase, which we shall denote by $\Delta\theta$, is in general non-zero: $\Delta\theta \sim \mathcal{O}(-\mathbf{x}\,\mathrm{Im}(\bar{\mathbf{x}}))$. The walkers will thus undergo a random walk in the complex plane defined by $\langle \Psi_T | \phi' \rangle$. At large β they will therefore populate the complex plane symmetrically, independent of their initial positions. This is illustrated in the right panel of Fig. 7, which shows $\langle \Psi_T | \phi \rangle$ for three-dimensional jellium with two electrons at $r_s = 10$ for a total projection time of $\beta = 250$. The distribution of the walkers is seen to be symmetric about the phase angle, and any signal that the walkers are all real initially (and $\langle \Psi_T | \phi^{(0)} \rangle = 1$) is lost in the statistical noise.

In other words, for a complex $\hat{\mathbf{v}}$, the random walk is "rotationally invariant" in the complex plane, and the divergence of $\bar{\mathbf{x}}$ is not enough to prevent the build-up of a finite density at the origin. Near the origin the local energy E_L diverges, which causes diverging fluctuations in the weights of walkers. To address this we make an additional approximation. We project the random walk to "one-dimension" and multiply the weight of each walker in each step by $\cos(\Delta\theta)$

$$w_{\phi'} = w_{\phi'} \max\{0, \cos(\Delta\theta)\} \tag{44}$$

in addition to Eq. (40) (or Eq. (38) if the hybrid form is used). This is only a good approximation in the presence of the similarity transformation that we have already preformed in what we have been calling importance sampling. There is a subtle but fundamental distinction between the formalism we have introduced and "traditional" importance sampling. In the latter, only the sampling efficiency is affected, not the expectation, because the transformation involves only real and non-negative functions. Here, in contrast, the functions, as given in Eq. (32), are *complex* and determine a gauge choice for our random walks. The proper choice of the force bias ensures that the leading order in the overall phase of $|\phi\rangle$ in the propagator in Eq. (37) is eliminated.

Several alternatives to Eq. (44) were tested [14, 47, 34]. One that seemed to work as well was $\exp\left(-(\mathrm{Im}(\bar{\mathbf{x}}))^2/2\right)$, which is the same as $\cos(\Delta\theta)$ in the limit of small values of $\Delta\theta$. Another was to impose $\mathrm{Re}\,\langle \Psi_T | \phi' \rangle > 0$, which gave similar results, but with somewhat larger variance.

4   Overview of applications, further reading, and outlook

Due to space limitations, we will not include any examples of applications here. The AFQMC method has been applied to lattice models, ultracold atoms, solids, and molecular systems.

In lattice models, most of the applications involve "only" a sign problem, because of the short-range nature of the interaction, although in multi-band models there can be tradeoffs between decompositions which lead to a sign or a phase problem [48]. A large body of results exist, including recent benchmark results [49]. Systems of $\mathcal{O}(1000)$ electrons have been treated quite routinely. The AFQMC method has demonstrated excellent capabilities and accuracy, illustrating its potential as a general many-body computational paradigm. A key recent development [43] is to use the density or density matrix computed from AFQMC as a feedback into a mean-field calculation. The trial wave function or density matrix obtained from the mean-field is then fed back into the AFQMC as a constraint, and a self-consistent constraining condition is achieved. This has lead to further improvement in the accuracy and robustness of the calculation [43,50]. This development is also seeing more applications in molecules and solids.

A related area involves ultracold atoms, where many valuable calculations have been performed with AFQMC. In addition to the physics advances these calculations have lead to, this has proved a fertile test ground for methodological developments, including better sampling methods [51,52], computation of excitations and dynamical correlations [53], the use of BCS [or antisymmetrized geminal power (AGP)] trial wave functions [54,55] and projection/path-integral in Hartree-Fock-Bogoliubov (HFB) space [56], treatment of spin-orbit coupling (SOC) [22], Bose systems and Fermi-Bose mixtures [45,57], and achieving linear scaling in lattice or basis size at $T > 0\,\mathrm{K}$ [58]. Many of these developments have direct applications in correlated systems of molecules and solids.

For molecular systems, a recent review article [21] describes in more detail the application of AFQMC in quantum chemistry. The formulation of AFQMC with Gaussian basis sets [20,24] has been extremely valuable. Direct comparisons can be made with high-level QC results, which have provided valuable benchmark information and have been crucial in developing the method. Many calculations have been performed using AFQMC as a "post-DFT" or "post-HF" approach for molecules and solids by simply feeding in the solution from standard DFT or HF. Several other systematic benchmarks have appeared, for example on the G1 set [59], on a set of 44 $3d$ transition-metal containing diatomics [60], on singlet-triplet gaps in organic biradicals [61], etc. Major recent methodological advances include the computation of observables [46], geometry optimization [62], and speedups using GPUs [32], low-rank tensor decomposition [63], embedding [64] and localization/downfolding [65], and correlated sampling [66], etc.

For solids, calculations have been done using planewaves and pseudopotentials including recent implementation of multiple-projector ones [67], with downfolding [29], with Gaussian-type orbitals [68], etc. A benchmark study [69] was recently carried out involving a large set of modern many-body methods. AFQMC was found to be comparable in accuracy to CCSD(T), the gold standard in chemistry [70,71], near equilibrium geometry. For bond breaking, AFQMC was able to maintain systematic accuracy. The AFQMC method can also be used to study

excited states. Excited states distinguished by different symmetry from the ground state can be computed in a manner similar to the ground state. For other excited states, prevention of collapse into the ground state and control of the fermion sign/phase problem are accomplished by a constraint using an excited state trial wave function [72]. An additional orthogonalization constraint is formulated to use virtual orbitals in solids for band structure calculations [73].

The AFQMC method is a computational framework for many-body calculations which combines a field-theoretic description with stochastic sampling. AFQMC has shown strong promise with its scalability (with system size and with parallel computing platforms), capability (total energy computation and beyond), and accuracy. The method is just coming into form, and rapid advances in algorithmic development and in applications are on-going. The literature is growing and I only listed a portion of it above. In addition, there is a pedagogical code for lattice models written in Matlab [74] that will be very useful for getting into the method.

We have discussed both ground-state and $T > 0$ auxiliary-field-based methods for correlated electron systems. We have outlined the formalism in a rather general way which allows for a systematic understanding of the origin of the sign/phase problem as well as the underlying theory for a scalable method capable of handling large molecules and bulk systems.

Often in the study of correlated models in condensed matter or cold atoms in optical lattices, the comment "But there is a sign problem" is made, to imply "so QMC calculations cannot be used here." I hope that these lectures and the large number of results in the literature with AFQMC will change this mindset. Calculations indeed can be done in systems where a sign problem is present—often with some of the most accurate results that are presently possible.

The AFQMC method has low-polynomial (cubic) scaling with system size, similar to DFT calculations. The structure of the open-ended random walk makes it ideally suited for modern high-performance computing platforms, with exceptional capacity for parallel scaling [15]. The rapid growth of high-performance computing resources will thus provide a strong boost to the application of AFQMC in the study of molecules and solids. The connection with independent-electron calculations, as we have highlighted, makes it straightforward to build AFQMC on top of DFT or HF codes, and take advantage of the many existing technical machineries developed over the past few decades in materials modeling. It also gives AFQMC much versatility as a general many-body method, offering, for example, straightforward treatment of heavier elements and spin-orbit coupling, computation of dynamical correlations, and the capability to embed AFQMC naturally and seamlessly in a calculation at the independent-particle level.

The development of AFQMC is entering an exciting new phase. We expect the method and the concept discussed here to see many applications, and to significantly enhance the capability of quantum simulations in interacting fermion systems. A large number of possible directions can be pursued, including many opportunities for algorithmic improvements and speedups. These will be spurred forward and stimulated by growth in applications, which we hope will in turn allow more rapid realization of a general many-body computational framework for correlated quantum materials.

Acknowledgments

Appendices

A   A few basics of Monte Carlo techniques

We list a few key elements from standard Monte Carlo (MC) techniques which are important to our discussions on QMC. For an introduction on MC methods, see, e.g., Ref. [38].

MC methods are often used to compute many-dimensional integrals of the form

$$G = \frac{\int_{\Omega_0} f(\mathbf{x})g(\mathbf{x})d\mathbf{x}}{\int_{\Omega_0} f(\mathbf{x})d\mathbf{x}}, \tag{45}$$

where $\mathbf{x}$ is a vector in a many-dimensional space and Ω_0 is a domain in this space. We will assume that $f(\mathbf{x}) \geq 0$ on Ω_0 and that it is normalizable, i.e., the denominator is finite. A familiar example of the integral in Eq. (45) comes from classical statistical physics, where $f(\mathbf{x})$ is the Boltzmann distribution.

To compute G by MC, we *sample* $\mathbf{x}$ from a probability density function (PDF) proportional to $f(\mathbf{x})$: $\bar{f}(\mathbf{x}) \equiv f(\mathbf{x})/\int_{\Omega_0} f(\mathbf{x})d\mathbf{x}$. This means to generate a sequence $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i, \ldots\}$ so that the probability that any $\mathbf{x}_i$ is in the sub-domain $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$ is

$$\mathrm{Prob}\{\mathbf{x}_i \in (\mathbf{x}, \mathbf{x} + d\mathbf{x})\} = \bar{f}(\mathbf{x})d\mathbf{x}. \tag{46}$$

There are different techniques to sample a many-dimensional function $f(\mathbf{x})$. The most general and perhaps most commonly used technique to sample $f(\mathbf{x})$ (i.e., the PDF $\bar{f}(\mathbf{x})$) is the Metropolis algorithm, which creates a Markov chain random walk [38] in $\mathbf{x}$-space whose equilibrium distribution is the desired function. We will also use a branching random walk, in which case there can be a weight w_i associated with each sampled $\mathbf{x}_i$. (In Metropolis, $w_i = 1$.) The MC samples provide a formal representation of f

$$f(\mathbf{x}) \propto \sum_i w_i \, \delta(\mathbf{x} - \mathbf{x}_i). \tag{47}$$

Given $\mathcal{M}$ independent samples from $f(\mathbf{x})$, the integral in Eq. (45) is estimated by

$$G_{\mathcal{M}} = \sum_{i=1}^{\mathcal{M}} w_i \, g(\mathbf{x}_i) / \sum_{i=1}^{\mathcal{M}} w_i \tag{48}$$

The error in the estimate decays algebraically with the number of samples: $|G_{\mathcal{M}} - G| \propto 1/\sqrt{\mathcal{M}}$. Using the results above, we can compute

$$G' = \frac{\int_{\Omega_0} f(\mathbf{x})g(\mathbf{x})h(\mathbf{x})d\mathbf{x}}{\int_{\Omega_0} f(\mathbf{x})h(\mathbf{x})d\mathbf{x}}, \tag{49}$$

if the function $h(\mathbf{x})$ is such that both the numerator and denominator exist. Formally

$$G'_{\mathcal{M}} = \frac{\sum_{i=1}^{\mathcal{M}} w_i g(\mathbf{x}_i) h(\mathbf{x}_i)}{\sum_{i=1}^{\mathcal{M}} w_i h(\mathbf{x}_i)}, \tag{50}$$

although, as we will see, difficulties arise when $h(\mathbf{x})$ can change sign and is rapidly oscillating. Integral equations are another main area of applications of MC methods. For example [38], the integral equation

$$\Psi'(\mathbf{x}) = \int_{\Omega_0} K(\mathbf{x}, \mathbf{y})\, w(\mathbf{y})\, \Psi(\mathbf{y}) d\mathbf{y}, \tag{51}$$

can be viewed in terms of a *random walk* if it has the following properties: $\Psi(\mathbf{y})$ and $\Psi'(\mathbf{x})$ can be viewed as PDF's (in the sense of f in Eq. (45)), $w(\mathbf{y}) \geq 0$, and $K(\mathbf{x}, \mathbf{y})$ is a PDF for $\mathbf{x}$ conditional on $\mathbf{y}$. Then, given an ensemble $\{\mathbf{y}_i\}$ sampling $\Psi(\mathbf{y})$, the following two steps will allow us to generate an ensemble that samples $\Psi'(\mathbf{x})$. First an absorption/branching process is applied to each $\mathbf{y}_i$ according to $w(\mathbf{y}_i)$. For example, we can make $\mathrm{int}(w(\mathbf{y}_i) + \xi)$ copies of $\mathbf{y}_i$, where ξ is a uniform random number on $(0, 1)$. Second we randomly walk each new $\mathbf{y}_j$ to an $\mathbf{x}_j$ by sampling the PDF $K(\mathbf{x}_j, \mathbf{y}_j)$. The resulting $\{\mathbf{x}_j\}$ are MC samples of $\Psi'(\mathbf{x})$. An example of this is the one-dimensional integral equation

$$\Psi(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-(x-y)^2} \sqrt{2} e^{-y^2/2} \Psi(y) dy, \tag{52}$$

which has a solution $\Psi(x) = e^{-x^2/2}$. The random walks, starting from an arbitrary distribution, will iteratively converge to a distribution sampling $\Psi(x)$.

B   Properties of non-orthogonal Slater determinants

In Eq. (18), the operator $\hat{\varphi}_m^\dagger \equiv \sum_i c_i^\dagger \varphi_{i,m}$, with m taking an integer value among $1, 2, \ldots, M$, creates an electron in a single-particle orbital φ_m: $\hat{\varphi}_m^\dagger |0\rangle = \sum_i \varphi_{i,m} |\chi_i\rangle$. The content of the orbital can thus be conveniently expressed as an N-dimensional vector $\{\varphi_{1,m}, \varphi_{2,m}, \ldots, \varphi_{N,m}\}$. The Slater determinant $|\phi\rangle$ in Eq. (18) can then be expressed as an $N \times M$ matrix

$$
\Phi \equiv \begin{pmatrix}
\varphi_{1,1} & \varphi_{1,2} & \cdots & \varphi_{1,M} \\
\varphi_{2,1} & \varphi_{2,2} & \cdots & \varphi_{2,M} \\
\vdots & \vdots & & \vdots \\
\varphi_{N,1} & \varphi_{N,2} & \cdots & \varphi_{N,M}
\end{pmatrix} .
$$

Each column of this matrix represents a single-particle orbital that is completely specified by its N-dimensional vector. For convenience, we will think of the different columns as all properly orthonormalized, which is straightforward to achieve by, for example, modified Gram-Schmidt (see e.g., [2,21]). For example the occupied manifold in a DFT solution forms a "wave function" which is a Slater determinant.

A key property of Slater determinants is the *Thouless Theorem*: any one-particle operator $\hat{B}$ of the form in Eq. (12), when acting on a Slater determinant, simply leads to another Slater determinant [75], i.e.,

$$
\hat{B}|\phi\rangle = \hat{\phi}_1'^\dagger \hat{\phi}_2'^\dagger \cdots \hat{\phi}_M'^\dagger |0\rangle \equiv |\phi'\rangle \tag{53}
$$

with $\hat{\phi}_m'^\dagger = \sum_j c_j^\dagger \Phi_{jm}'$ and $\Phi' \equiv e^U \Phi$, where U is a square matrix whose elements are given by U_{ij} and $B \equiv \exp(U)$ is therefore an $N \times N$ square matrix as well. In other words, the operation of $\hat{B}$ on $|\phi\rangle$ simply involves multiplying an $N \times N$ matrix to the $N \times M$ matrix representing the Slater determinant.

In standard quantum chemistry (QC) methods, the many-body ground-state wave function is also represented by a sum of Slater determinants. However, there is a key difference between it and the AFQMC representation. In QC methods, the different Slater determinants are orthogonal, because they are obtained by excitations based on a fixed set of orbitals (for instance, the occupied and virtual orbitals from a reference calculation such as Hartree-Fock). In AFQMC, the Slater determinants are non-orthogonal because they are generated by "rotating" (only) the occupied orbitals through the operations in Eq. (53).

Several properties of non-orthogonal Slater determinants are worth mentioning. The overlap between two of them is given by

$$
\langle \phi | \phi' \rangle = \det(\Phi^\dagger \Phi'). \tag{54}
$$

We can define the expectation of an operator $\hat{O}$ with respect to a pair of non-orthogonal Slater determinants

$$
\langle \hat{O} \rangle_{\phi\phi'} \equiv \frac{\langle \phi | \hat{O} | \phi' \rangle}{\langle \phi | \phi' \rangle}, \tag{55}
$$

for instance single-particle Green function $G_{ij} \equiv \langle c_i c_j^\dagger \rangle_{\phi\phi'}$

$$G_{ij} \equiv \frac{\langle \phi | c_i c_j^\dagger | \phi' \rangle}{\langle \phi | \phi' \rangle} = \delta_{ij} - [\Phi'(\Phi^\dagger \Phi')^{-1} \Phi^\dagger]_{ij}. \tag{56}$$

Given the Green function matrix G, the general expectation defined in Eq. (55) can be computed for most operators of interest. For example, we can calculate the expectation of a general two-body operator, $\hat{O} = \sum_{ijkl} O_{ijkl} c_i^\dagger c_j^\dagger c_k c_l$, under the definition of Eq. (55)

$$\langle \hat{O} \rangle_{\phi\phi'} = \sum_{ijkl} O_{ijkl} \big(G'_{jk} G'_{il} - G'_{ik} G'_{jl} \big), \tag{57}$$

where the matrix G' is defined as $G' \equiv I - G$.

There are several generalizations of the formalism we have discussed which extend the capability and/or accuracy of the AFQMC framework. These can be thought of as generalizing one or both of the Slater determinants in Eqs. (54), (55), and (56). From the viewpoint of AFQMC, the "bra" in these equations represents the trial wave function, and the "ket" represents the random walker:

- The first generalization is to replace $\langle \phi |$ by a projected Bardeen-Cooper-Schrieffer (BCS) wave function, that is, to use a projected BCS as a trial wave function, which can be advantageous for systems with pairing order. The corresponding overlap, Green functions, and two-body mixed expectations have been worked out [54].

- The second is to have both $\langle \phi |$ and $| \phi' \rangle$ in generalized HF (GHF) form, which is necessary to treat systems with SOC. The required modification to the formalism outlined above is given by [22].

- The third generalization is to have both sides in Hartree-Fock-Bogoliubov (HFB) form, for example, to treat Hamiltonians with pairing fields. This will also be useful when using AFQMC as an impurity solver in which the embedding induces pairing order. The corresponding AFQMC formalism has been described [56].

References

[1] S. Zhang, in W. Andreoni and S. Yip (eds.): *Handbook of Materials Modeling* (Springer, Cham, 2018)

[2] S. Zhang, in E. Pavarini, E. Koch, and U. Schollwöck (eds.): *Emergent Phenomena in Correlated Matter*, Modeling and Simulation, Vol. 3 (Forschungszentrum Jülich, 2013)

[3] S. Zhang, Comput. Phys. Commun. **127**, 150 (2000)

[4] Y.Y. He, M. Qin, H. Shi, Z.Y. Lu, and S. Zhang, Phys. Rev. B **99**, 045108 (2019)

[5] A. Szabo and N. Ostlund: *Modern quantum chemistry* (McGraw-Hill, New York, 1989)

[6] W. Kohn, Rev. Mod. Phys. **71**, 1253 (1999) and references therein

[7] R.M. Martin: *Electronic Structure: Basic theory and practical methods* (Cambridge University Press, 2004)

[8] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985)

[9] M.H. Kalos, D. Levesque, and L. Verlet, Phys. Rev. A **9**, 2178 (1974)

[10] W.M.C. Foulkes, L. Mitas, R.J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001) and references therein

[11] D.M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995) and references therein

[12] R. Blankenbecler, D.J. Scalapino, and R.L. Sugar, Phys. Rev. D **24**, 2278 (1981)

[13] G. Sugiyama and S.E. Koonin, Ann. Phys. (NY) **168**, 1 (1986)

[14] S. Zhang and H. Krakauer, Phys. Rev. Lett. **90**, 136401 (2003)

[15] K.P. Esler, J. Kim, D.M. Ceperley, W. Purwanto, E.J. Walter, H. Krakauer, S. Zhang, P.R.C. Kent, R.G. Hennig, C. Umrigar, M. Bajdich, J. Kolorenc, L. Mitas, and A. Srinivasan, J. Phys.: Conf. Series **125**, 012057 (2008)

[16] K.E. Schmidt and M.H. Kalos, in K. Binder (ed.): *Applications of the Monte Carlo Method in Statistical Physics* (Springer, Heidelberg, 1984)

[17] E.Y. Loh Jr., J.E. Gubernatis, R.T. Scalettar, S.R. White, D.J. Scalapino, and R. Sugar, Phys. Rev. B **41**, 9301 (1990)

[18] S. Zhang, in M.P. Nightingale and C.J. Umrigar (eds.): *Quantum Monte Carlo Methods in Physics and Chemistry* (Kluwer Academic Publishers, 1999)

[19] S. Zhang, J. Carlson, and J.E. Gubernatis, Phys. Rev. B **55**, 7464 (1997)

[20] W.A. Al-Saidi, S. Zhang, and H. Krakauer, J. Chem. Phys. **124**, 224101 (2006)

[21] M. Motta and S. Zhang, Wiley Interdiscip. Rev. Comput. Mol. Sci. **8**, e1364 (2018)

[22] P. Rosenberg, H. Shi, and S. Zhang, J. Phys. Chem. Solids **128**, 161 (2019)

[23] M. Suewattana, W. Purwanto, S. Zhang, H. Krakauer, and E.J. Walter,
     Phys. Rev. B **75**, 245123 (2007)

[24] W. Purwanto, H. Krakauer, Y. Virgus, and S. Zhang, J. Chem. Phys. **135**, 164105 (2011)

[25] S. Zhang, Phys. Rev. Lett. **83**, 2777 (1999)

[26] Y. Liu, M. Cho, and B. Rubenstein, J. Chem. Theory and Comput. **14**, 4722 (2018)

[27] J.W. Negele and H. Orland: *Quantum Many-Particle Systems*
     (Perseus Books, Reading, MA, 1998)

[28] W.A. Al-Saidi, H. Krakauer, and S. Zhang, J. Chem. Phys. **126**, 194105 (2007)

[29] F. Ma, W. Purwanto, S. Zhang, and H. Krakauer, Phys. Rev. Lett. **114**, 226401 (2015)

[30] H. Koch, A.S. de Merás, and T.B. Pedersen, J. Chem. Phys. **118**, 9481 (2003)

[31] F. Aquilante, L. De Vico, N. Ferre, G. Ghigo, P.Å Malmqvist, P. Neogrady, T. Pedersen,
     M. Pitonak, M. Reiher, B. Roos, L. Serrano-Andres, M. Urban, V. Veryazov, and R. Lindh,
     J. Comp. Chem. **31**, 224 (2010)

[32] J. Shee, E.J. Arthur, S. Zhang, D.R. Reichman, and R.A. Friesner,
     J. Chem. Theory Comput. **14**, 4109 (2018)

[33] R. Baer, M. Head-Gordon, and D. Neuhauser, J. Chem. Phys. **109**, 6219 (1998)

[34] W. Purwanto and S. Zhang, Phys. Rev. A **72**, 053610 (2005)

[35] H. Shi and S. Zhang, Phys. Rev. B **88**, 125132 (2013)

[36] S. Zhang and D.M. Ceperley, Phys. Rev. Lett. **100**, 236404 (2008)

[37] S. Sorella, S. Baroni, R. Car, and M. Parrinello, Europhys. Lett. **8**, 663 (1989)

[38] M.H. Kalos and P.A. Whitlock: *Monte Carlo methods*, Vol I (Wiley, 1986)

[39] J.E. Hirsch, Phys. Rev. B **31**, 4403 (1985)

[40] Z.C. Wei, C. Wu, Y. Li, S. Zhang, and T. Xiang, Phys. Rev. Lett. **116**, 250601 (2016)

[41] S. Zhang and M.H. Kalos, Phys. Rev. Lett. **67**, 3074 (1991)

[42] S. Zhang, J. Carlson, and J.E. Gubernatis, Phys. Rev. Lett. **74**, 3652 (1995)

[43] M. Qin, H. Shi, and S. Zhang, Phys. Rev. B **94**, 235119 (2016)

[44] S.B. Fahy and D.R. Hamann, Phys. Rev. Lett. **65**, 3437 (1990)

[45] W. Purwanto and S. Zhang, Phys. Rev. E **70**, 056702 (2004)

[46] M. Motta and S. Zhang, J. Chem. Theory Comput. **13**, 5367 (2017)

[47] S. Zhang, H. Krakauer, W.A. Al-Saidi, and M. Suewattana,
     Comput. Phys. Commun. **169**, 394 (2005)

[48] H. Hao, B.M. Rubenstein, and H. Shi, Phys. Rev. B **99**, 235142 (2019)

[49] J.P.F. LeBlanc, A.E. Antipov, F. Becca, I.W. Bulik, G.K.L. Chan, C.M. Chung, Y. Deng,
     M. Ferrero, T.M. Henderson, C.A. Jiménez-Hoyos, E. Kozik, X.W. Liu, A.J. Millis,
     N.V. Prokof'ev, M. Qin, G.E. Scuseria, H. Shi, B.V. Svistunov, L.F. Tocchio, I.S. Tupitsyn,
     S.R. White, S. Zhang, B.X. Zheng, Z. Zhu, and E. Gull, Phys. Rev. X **5**, 041041 (2015)

[50] B.X. Zheng, C.M. Chung, P. Corboz, G. Ehlers, M.P. Qin, R.M. Noack, H. Shi, S.R. White,
     S. Zhang, and G.K.L. Chan, Science **358**, 1155 (2017)

[51] H. Shi, S. Chiesa, and S. Zhang, Phys. Rev. A **92**, 033603 (2015)

[52] H. Shi and S. Zhang, Phys. Rev. E **93**, 033303 (2016)

[53] E. Vitali, H. Shi, M. Qin, and S. Zhang, Phys. Rev. B **94**, 085140 (2016)

[54] J. Carlson, S. Gandolfi, K.E. Schmidt, and S. Zhang, Phys. Rev. A **84**, 061602 (2011)

[55] E. Vitali, P. Rosenberg, and S. Zhang, arXiv:1905.05012

[56] H. Shi and S. Zhang, Phys. Rev. B **95**, 045144 (2017)

[57] B.M. Rubenstein, S. Zhang, and D.R. Reichman, Phys. Rev. A **86**, 053606 (2012)

[58] Y.Y. He, H. Shi, and S. Zhang, arXiv e-prints arXiv:1906.02247 (2019)

[59] E.J. Landinez Borda, J. Gomez, and M.A. Morales, J. Chem. Phys. **150**, 074105 (2019)

[60] J. Shee, B. Rudshteyn, E.J. Arthur, S. Zhang, D.R. Reichman, and R.A. Friesner,
     J. Chem. Theory Comput. **15**, 2346 (2019)

[61] J. Shee, E.J. Arthur, S. Zhang, D.R. Reichman, and R.A. Friesner, arXiv:1905.13316

[62] M. Motta and S. Zhang, J. Chem. Phys. **148**, 181101 (2018)

[63] M. Motta, J. Shee, S. Zhang, and G. Kin-Lic Chan, arXiv:1810.01549

[64] Y. Virgus, W. Purwanto, H. Krakauer, and S. Zhang, Phys. Rev. Lett. **113**, 175502 (2014)

[65] B. Eskridge, H. Krakauer, and S. Zhang, J. Chem. Theory Comput. **15**, 3949 (2019)

[66] J. Shee, S. Zhang, D.R. Reichman, and R.A. Friesner,
J. Chem. Theory Comput. **13**, 2667 (2017)

[67] F. Ma, S. Zhang, and H. Krakauer, Phys. Rev. B **95**, 165103 (2017)

[68] S. Zhang, F.D. Malone, and M.A. Morales, J. Chem. Phys. **149**, 164102 (2018)

[69] M. Motta, D.M. Ceperley, G.K.L. Chan, J.A. Gomez, E. Gull, S. Guo, C.A. Jiménez-Hoyos, T.N. Lan, J. Li, F. Ma, A.J. Millis, N.V. Prokof'ev, U. Ray, G.E. Scuseria, S. Sorella, E.M. Stoudenmire, Q. Sun, I.S. Tupitsyn, S.R. White, D. Zgid, and S. Zhang, Phys. Rev. X **7**, 031059 (2017)

[70] R.J. Bartlett and M. Musiał, Rev. Mod. Phys. **79**, 291 (2007)

[71] T.D. Crawford and H.F. Schaefer III, Rev. Comput. Chem. **14**, 33 (2000)

[72] W. Purwanto, S. Zhang, and H. Krakauer, J. Chem. Phys. **130**, 094107 (2009)

[73] F. Ma, S. Zhang, and H. Krakauer, New J. Phys. **15**, 093017 (2013)

[74] H. Nguyen, H. Shi, J. Xu, and S. Zhang, Comput. Phys. Commun. **185**, 3344 (2014)

[75] D.R. Hamann and S.B. Fahy, Phys. Rev. B **41**, 11352 (1990)