15 Auxiliary-Field Quantum Monte Carlo for Correlated Electron Systems

Shiwei Zhang
College of William & Mary
Department of Physics, Williamsburg, VA23185, USA

Contents

1 Introduction **2**

2 Formalism **4**
2.1 Ground-state projection 4
2.2 Slater determinant space 6
2.3 Hubbard-Stratonovich transformation 8
2.4 A simple example 10

3 Ground-State AFQMC Methods **12**
3.1 Free-projection AFQMC 12
3.2 Why and how does the sign problem occur? 13
3.3 The constrained path Monte Carlo method 16
3.4 The phaseless formalism for complex auxiliary-fields 18

4 Illustrative results **23**

5 Concluding remarks **25**

6 Acknowledgments **26**

A A few basics of Monte Carlo techniques **27**

E. Pavarini, E. Koch, and U. Schollwöck
Emergent Phenomena in Correlated Matter
Modeling and Simulation Vol. 3
Forschungszentrum Jülich, 2013, ISBN 978-3-89336-884-6
http://www.cond-mat.de/events/correl13

1 Introduction

Predicting materials properties requires robust and reliable calculations at the most fundamental level. Often the effects being studied or designed originate from electron correlations, and small errors in their treatment can result in crucial and qualitative differences in the properties. Accurate treatment of electron correlations and collective behaviors is one of the great challenges in modern science [1, 2]. Explicit solution of the many-body Schrödinger equation leads to rapidly growing computational cost as a function of system size: exponential, as in configuration interaction (CI) [3], or at the minimum, a high power, e.g., $\mathcal{O}(N^7)$ as in the coupled-cluster CCSD(T) [3], the preeminent quantum chemistry method.

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) [2, 4], 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. These methods have often been extremely effective in complex molecules and solids [4], and are the standard in electronic structure, widely applied in condensed matter, quantum chemistry, and materials science.

Despite the tremendous successes of DFT (most noticeably through its Car-Parrinello molecular dynamics implementation [5]), 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 [6–9], 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.

The development of alternatives to independent-particle theories is therefore of paramount fundamental and practical significance. Quantum Monte Carlo (QMC) methods [10–15] are among the most promising candidates for post-mean-field calculations. By using stochastic sampling rather than explicit integration over phase space, these methods have computational costs that scale gracefully (N^3-N^4) with system size and exceptional potential for parallel scaling [16].

For fermion systems, however, QMC methods suffer from the so-called "sign" problem [17–19]. In these systems, 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 Monte Carlo (MC) samples of phase space. In fact, as the temperature is lowered or the system size is increased, such cancellation becomes more and more complete. The net signal thus decays *exponentially*. The algebraic scaling is then lost, and the method breaks down. Clearly the impact of this problem on the study of correlated electron systems is extremely severe.

To date most applications of QMC methods to correlated electron systems have either lived with the sign problem or relied on some form of approximation to overcome the exponential

scaling. The former is seen more often in the community studying lattice models for strongly correlated systems where, understandably, it is more difficult to make approximations and more demanding of insensitivity of the many-body method to *a priori* input. However, this approach clearly has severe limitations in the systems that can be examined. The latter has been more prevalent in electronic structure, where without some control of the sign problem it is difficult to sustain signal in even the simplest realistic Hamiltonian. This approach loses exactness and the results can be biased by the approximation.

The approach we take in these lecture notes is in some sense to combine the two approaches. The sign/phase problem will be controlled approximately. 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 formulate constraints on the random-walk paths which lead to better approximations that 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) [20] for systems where there is a sign problem (for example, Hubbard-like models where the local interactions lead to auxiliary-fields that are real). For electronic systems where there is a phase problem (as the Coloumb interaction leads to complex fields), the methods have been referred to as phaseless or phase-free auxiliary-field QMC (AFQMC) [15, 21]. In both cases, the idea is to constrain the sign or phase of the overlap of the sampled Slater determinants with a trial wave function. It eliminates the sign or phase instability and restores low-power (typically to the third power of system size) computational scaling. Applications to a variety of systems have shown that the methods are very accurate, even with simple trial wave functions taken directly from mean-field calculations (see, e.g., Refs [22] and [23] references therein). As an example, in Ref. [23], an 8×128 lattice in a doped repulsive Hubbard model (with over 900 electrons) is studied to calculate the ground-state magnetic and charge-correlations. The dimension of the Hilbert space for this system is $\mathcal{O}(10^{600})$!

With these notes, we will give a pedagogical introduction to the AFQMC method, starting from the forms without any constraint [13, 14], then discussing the sign and phase problem [18, 19, 24], and ideas for constraining the paths in the path integral in AF space [15, 20, 25] to eliminate the problem. The chapter is not meant to be a comprehensive review. Rather, we focus on several basic algorithmic aspects, given the length restrictions of the chapter, to complement the lectures. We will cover ground-state calculations in Hubbard-like models [19, 20, 26] and in realistic materials [15, 21, 27]. We will not cover finite-temperature calculations, although corresponding methods have been developed [28, 29], nor will we discuss boson or mixture systems, although the algorithms share many common features and some of the papers on boson systems [29–31] contain details that can serve as useful references. Some of the more pedagogical materials in this chapter are based on Refs. [19] and [24]. The references cited are sources where we have drawn materials for the lectures, and point to possible additional reading; they are by no means a literature survey.

2 Formalism

2.1 Ground-state projection

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

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

where $\mathbf{r}_m$ is the real-space coordinate of the m^{th} fermion, $V_{\text{ext}}(\mathbf{r}_m)$ is the effect of the external potential on it, and $V_{\text{int}}(\mathbf{r}_m - \mathbf{r}_n)$ is the interaction between fermions m and n. 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. This will lead to a non-local function V_{ext} [27]. The total number of fermions, M, will be fixed in the calculations we discuss. For simplicity, we have suppressed spin-indices, but the spin will be made explicit when necessary. In that case, M_σ is 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 calculation, although it will be straightforward to generalize our discussions to treat other cases, for example, when there is spin-orbit coupling.

With any chosen one-particle basis, for example lattice sites (Hubbard model), plane-waves (as in solid state calculations) [27], or Gaussians (as in quantum chemistry) [21, 32], the Hamiltonian can be written 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 N is the size of the chosen one-particle basis, and $c_i^\dagger$ and c_i are the corresponding creation and annihilation operators. The one-body matrix elements, T_{ij}, contain the effect of both the kinetic energy and external potential, while the two-body matrix elements, V_{ijkl}, are from the interaction. The matrix elements are integrals expressed in terms of the basis functions and the potentials. We will assume that they have been evaluated and are known as we begin our many-body calculations. Examples are given below.

One of the simplest Hamiltonians of this form is the Hubbard model, which has played an important role in many-body physics and whose properties are still not fully understood:

$$\hat{H} = \hat{K} + \hat{V} = -t\sum_{\langle i,j\rangle\sigma}^{N} c_{i\sigma}^\dagger c_{j\sigma} + U\sum_{i}^{N} n_{i\uparrow} n_{i\downarrow}\,. \tag{3}$$

Here N is the number of lattice sites, $c_{i\sigma}^\dagger$ and $c_{j\sigma}$ are creation and annihilation operators of an electron of spin σ on the i-th lattice site, t is the nearest-neighbor hopping energy, $n_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$ is the density operator, and U is the interaction strength. Two parameters, the interaction U/t and the electron density $(M_\uparrow + M_\downarrow)/N$, determine the physics given the topology of the lattice.

Most ground-state QMC methods are based on

$$|\Psi_0\rangle \propto \lim_{\tau\to\infty} e^{-\tau\hat{H}}|\Psi_T\rangle; \tag{4}$$

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 $\langle\Psi_T|\Psi_0\rangle \neq 0$. In a numerical method, the limit can be obtained iteratively by

$$|\Psi^{(n+1)}\rangle = e^{-\Delta\tau\hat{H}}|\Psi^{(n)}\rangle, \tag{5}$$

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

$$\langle\hat{O}\rangle = \lim_{n\to\infty} \frac{\langle\Psi^{(n)}|\hat{O}|\Psi^{(n)}\rangle}{\langle\Psi^{(n)}|\Psi^{(n)}\rangle} . \tag{6}$$

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:

$$E_0 = \lim_{n\to\infty} \frac{\langle\Psi_T|\hat{H}|\Psi^{(n)}\rangle}{\langle\Psi_T|\Psi^{(n)}\rangle} . \tag{7}$$

QMC methods carry out the iteration in Eq. (5) by Monte Carlo sampling. The difference between different classes of methods amounts primarily to the space that is used to represent the wave function or density matrix and to carry out the integration. The auxiliary-field QMC (AFQMC) methods work in second quantized representation and in an auxiliary-field space, while Green's function Monte Carlo (GFMC) or diffusion Monte Carlo (DMC) works in first-quantized representation and in configuration space [10, 11].

Let us assume that $|\Psi_T\rangle$ is of the form of a single Slater determinant or a linear combination of Slater determinants. A Slater determinant is the form of a mean-field solution to $\hat{H}$ (see next section) expressed in terms of the basis functions chosen in Eq. (2). The operation of $e^{-\tau\hat{H}_1}$ on a Slater determinant is straightforward to calculate, and it simply yields another determinant. The ground-state projection would therefore turn into the propagation of a single Slater determinant if it were somehow possible to write the two-body propagator $e^{-\tau\hat{H}_2}$ as the exponential of a one-body operator. This is realized in mean-field methods. In the Hartree-Fock (HF) approximation $\hat{H}_2$ is replaced by one-body operators times expectations with respect to the current Slater determinant wave function, schematically:

$$c_i^\dagger c_j^\dagger c_k c_l \to c_i^\dagger c_l \langle c_j^\dagger c_k\rangle - c_i^\dagger c_k \langle c_j^\dagger c_l\rangle . \tag{8}$$

A decomposition that includes pairing is also possible, leading to a Hartree-Fock-Bogoliubov calculation. In the local density approximation (LDA) in DFT [4], $\hat{H}_2$ is replaced by the density operator in real-space times a functional of the local density calculated with respect to the current Slater determinant in the self-consistent process. In both these cases, an iterative procedure can be used, following Eq. (5), to project out the solution to the approximate Hamiltonians, as an imaginary-time evolution of a single Slater determinant [33].

2.2 Slater determinant space

The building blocks of Slater determinants are single-particle basis states. These single-particle basis states can be plane waves, or lattice sites in the Hubbard model, Gaussians, or energy eigenstates in a mean-field potential. Often the space of single-particle basis states is truncated. Single-particle wave functions (orbitals) are formed with the basis states. Slater determinants are then built from the single-particle orbitals.
We define some notations that we will use throughout the discussion [24]:

- N: number of single-electron basis states. For example, N can be the number of lattice sites ($L \times L$) in the two-dimensional Hubbard model.
- $|\chi_i\rangle$: the i^{th} single-particle basis state ($i = 1, 2, \cdots, N$). For example, $|\chi_{\mathbf{G}}\rangle$ can be a plane wave basis state with $\chi_{\mathbf{G}}(\mathbf{r}) \propto e^{i\mathbf{G}\cdot\mathbf{r}}$, where $\mathbf{r}$ is a real-space coordinate.
- $c_i^\dagger$ and c_i: creation and annihilation operators for an electron in $|\chi_i\rangle$. $n_i \equiv c_i^\dagger c_i$ is the corresponding number operator.
- M: number of electrons (if we omit spin index, e.g., if the system is fully polarized). In the more general case, M_σ is the number of electrons with spin σ ($\sigma = \uparrow$ or $\downarrow$). Of course, the choice of N above must ensure that $M_\sigma \leq N$.
- φ_m: single-particle orbital (we include an index m for discussions below to distinguish different single-particle orbitals). A single-particle orbital φ_m, given in terms of the single-particle basis states $\{|\chi_i\rangle\}$ as $\sum_i \varphi_{i,m}|\chi_i\rangle = \sum_i c_i^\dagger \varphi_{i,m}|0\rangle$, can be conveniently expressed as an N-dimensional vector:

$$\begin{pmatrix} \varphi_{1,m} \\ \varphi_{2,m} \\ \vdots \\ \varphi_{N,m} \end{pmatrix}$$

Given M different single-particle orbitals, we form a many-body wave function from their anti-symmetrized product:

$$|\phi\rangle \equiv \hat{\varphi}_1^\dagger \hat{\varphi}_2^\dagger \cdots \hat{\varphi}_M^\dagger |0\rangle \tag{9}$$

where the operator

$$\hat{\varphi}_m^\dagger \equiv \sum_i c_i^\dagger \, \varphi_{i,m} \tag{10}$$

creates an electron in the m^{th} single-particle orbital $\{\varphi_{1,m}, \varphi_{2,m}, \cdots, \varphi_{N,m}\}$. The many-body state $|\phi\rangle$ in Eq. (9) can be conveniently 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. It is straightforward to keep the different columns orthonormalized. If the real-space coordinates of the electrons are $R = \{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_M\}$, the many-body state in Eq. (9) gives (see Sec. 2.4)

$$\langle R|\phi\rangle = \phi(R) = \det \begin{pmatrix} \varphi_1(\mathbf{r}_1) & \varphi_2(\mathbf{r}_1) & \cdots & \varphi_M(\mathbf{r}_1) \\ \varphi_1(\mathbf{r}_2) & \varphi_2(\mathbf{r}_2) & \cdots & \varphi_M(\mathbf{r}_2) \\ \vdots & \vdots & & \vdots \\ \varphi_1(\mathbf{r}_M) & \varphi_2(\mathbf{r}_M) & \cdots & \varphi_M(\mathbf{r}_M) \end{pmatrix},$$

where $\varphi_m(\mathbf{r}) = \sum_i \varphi_{i,m}\chi_i(\mathbf{r})$.
The many-body state $|\phi\rangle$ is known as a Slater determinant. One example of a Slater determinant is the HF solution $|\phi_{\rm HF}\rangle = \prod_\sigma |\phi^\sigma_{\rm HF}\rangle$, where $|\phi^\sigma_{\rm HF}\rangle$ is defined by a matrix $\Phi^\sigma_{\rm HF}$ whose columns are the M_σ lowest HF eigenstates. We can now add the following to our list of notations above:

- $|\phi\rangle$: a many-body wave function which can be written as a Slater determinant.
- Φ: an $N \times M$ matrix which represents the Slater determinant $|\phi\rangle$. Φ_{ij} will denote the matrix element of the matrix Φ in the $i^{\rm th}$ row and $j^{\rm th}$ column. For example, $\Phi_{ij} = \varphi_{i,j}$ above in Φ. Below when a Slater determinant $|\phi\rangle$ is referred to, it will often be helpful to think in terms of the matrix representation Φ operationally.
- $|\Psi\rangle$: a many-body wave function which is not necessarily a single Slater determinant, e.g., $|\Psi^{(n)}\rangle$ in Eq. (5).

Several properties of the Slater determinant are worth mentioning. For any two non-orthogonal Slater determinants, $|\phi\rangle$ and $|\phi'\rangle$, it can be shown that their overlap integral is

$$\langle\phi|\phi'\rangle = \det(\Phi^\dagger\Phi'). \tag{11}$$

The operation on any Slater determinant by any operator $\hat{B}$ of the form

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

simply leads to another Slater determinant [34], i.e.,

$$\hat{B}|\phi\rangle = \hat{\phi}_1'^\dagger \hat{\phi}_2'^\dagger \cdots \hat{\phi}_M'^\dagger |0\rangle \equiv |\phi'\rangle \tag{13}$$

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 $B \equiv e^U$ 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 by an $N \times M$ matrix.
We can define the expectation of an operator $\hat{O}$ with respect to a pair of non-orthogonal Slater determinants:

$$\overline{\langle\hat{O}\rangle} \equiv \frac{\langle\phi|\hat{O}|\phi'\rangle}{\langle\phi|\phi'\rangle}\,. \tag{14}$$

The "bar" distinguishes it from the true interacting many-body expectations in Eq. (6), which we wish to compute with QMC. The simplest example of Eq. (14) is the single-particle Green's function $G_{ij} \equiv \overline{\langle c_i c_j^\dagger \rangle}$. In the ground-state formalism,

$$G_{ij} \equiv \frac{\langle \phi | c_i c_j^\dagger | \phi' \rangle}{\langle \phi | \phi' \rangle} = \delta_{ij} - [\Phi'(\Phi^\dagger \Phi')^{-1} \Phi^\dagger]_{ij}. \tag{15}$$

Given the Green's function G, the general expectation defined in Eq. (14) can be computed for most operators of interest. This is an important property that will be used in our QMC calculations. 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. (14):

$$\overline{\langle \hat{O} \rangle} = \sum_{ijkl} O_{ijkl} (G'_{jk} G'_{il} - G'_{ik} G'_{jl}), \tag{16}$$

where the matrix G' is defined as $G' \equiv \mathbb{1} - G$.
Recently the use of a projected BCS wave function as a trial wave function in AFQMC has been formulated [35]. The overlap between a projected BCS wave function $|\Psi_{\rm BCS}\rangle$ and a Slater determinant $|\phi\rangle$, and the corresponding "Green's function" of Eq. (15) can be evaluated at a cost similar to that between two Slater determinants.

2.3 Hubbard-Stratonovich transformation

In order to carry out Eq. (5) in the Slater-determinant space we have introduced above, we write the many-body propagator $e^{-\Delta\tau \hat{H}}$ in single-particle form. With a small $\Delta\tau > 0$, the Trotter approximation can be used:

$$e^{-\Delta\tau \hat{H}} \approx e^{-\Delta\tau \hat{H}_1} e^{-\Delta\tau \hat{H}_2}, \tag{17}$$

which introduces a Trotter error. For actual implementation of the algorithms we discuss here, higher order Trotter break-ups are often used. The Trotter error can be further reduced with an extrapolation procedure after separate calculations have been done with different values of $\Delta\tau$. The $\hat{H}_1$ part of the propagator in (17) is the exponential of a one-body operator. The $\hat{H}_2$ part is not. It is, however, possible to rewrite $e^{-\Delta\tau \hat{H}_2}$ in this desired form through a so-called Hubbard-Stratonovich (HS) transformation [36]. For example, for the Hubbard model in Eq. (3), an exact, discrete HS transformation [37] exists for the repulsive on-site repulsion

$$e^{-\Delta\tau U n_{i\uparrow} n_{i\downarrow}} = e^{-\Delta\tau U (n_{i\uparrow} + n_{i\downarrow})/2} \sum_{x_i = \pm 1} \frac{1}{2} e^{\gamma x_i (n_{i\uparrow} - n_{i\downarrow})}, \tag{18}$$

where the constant γ is determined by $\cosh(\gamma) = \exp(\Delta\tau U/2)$. Similarly, for an attractive interaction $V n_{i\uparrow} n_{i\downarrow}$ with $V < 0$

$$e^{-\Delta\tau V n_{i\uparrow} n_{i\downarrow}} = e^{-\Delta\tau V (n_{i\uparrow} + n_{i\downarrow} - 1)/2} \sum_{x_i = \pm 1} \frac{1}{2} e^{\gamma x_i (n_{i\uparrow} + n_{i\downarrow} - 1)}, \tag{19}$$

where $\cosh(\gamma) = \exp(\Delta\tau |V|/2)$.

In general, if $\hat{H}_2$ can be written as a sum of squares of one-body operators, of the form

$$\hat{H}_2 = \frac{1}{2}\sum_{\gamma=1}^{N_\gamma} \lambda_\gamma \hat{v}_\gamma^2, \tag{20}$$

we can apply the HS transformation to each term (after Trotter break-up):

$$e^{-(\Delta\tau/2)\,\lambda\,\hat{v}^2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, e^{-x^2/2}\, e^{x\sqrt{-\Delta\tau\lambda}\,\hat{v}}, \tag{21}$$

where x is an auxiliary-field variable. The constant in front of $\hat{v}$ in the exponent on the right-hand side can be real or imaginary, depending on the sign of λ. The key is that the quadratic form (in $\hat{v}$) on the left is replaced by a linear one on the right. In Eq. (20), λ_γ is a constant and $\hat{v}_\gamma$ is a one-body operator similar to $\hat{H}_1$. Each γ leads to one auxiliary-field in Eq. (21).
We now show one way to manipulate a general $\hat{H}_2$ into the form in Eq. (20). Let us cast V_{ijkl} in Eq. (2) in the form of a Hermitian matrix by introducing two indices $\alpha = (i,l)$ and $\beta = (k,j)$: $\mathcal{V}_{\alpha\beta} = \mathcal{V}_{(i,l),(k,j)} = V_{ijkl}$. The Hermitian matrix $\mathcal{V}$ can then be diagonalized and written as $\mathcal{V} = R\Lambda R^\dagger$, where R is a matrix whose columns are the eigenvectors of $\mathcal{V}$ and Λ is a diagonal matrix containing the corresponding eigenvalues λ_α. That is

$$\mathcal{V}_{\alpha\beta} = \sum_\gamma R_{\alpha\gamma}\lambda_\gamma R^*_{\beta\gamma}. \tag{22}$$

The two-body operator $\hat{V}$ can therefore be written as

$$\begin{aligned}
\hat{H}_2 &= \sum_{ijkl} V_{ijkl}\, c_i^\dagger c_l c_j^\dagger c_k - \sum_{ijkl} V_{ijkl}\, c_i^\dagger c_k\, \delta_{jl} \\
&= \sum_\gamma \lambda_\gamma \Big(\sum_{il} R_{(i,l)\gamma} c_i^\dagger c_l \Big) \Big(\sum_{jk} R^*_{(k,j)\gamma} c_j^\dagger c_k \Big) - \sum_{ik} \Big(\sum_j V_{ijkj} \Big)\, c_i^\dagger c_k.
\end{aligned}$$

Noting that $\hat{H}_2$ is Hermitian, we can put the above in a more symmetric form

$$\hat{H}_2 = \frac{1}{2}\sum_\gamma \lambda_\gamma \{\hat{\rho}_\gamma, \hat{\rho}_\gamma^\dagger\} + \hat{\rho}_0, \tag{23}$$

where the one-body operators are defined as $\hat{\rho}_\gamma \equiv \sum_{il} R_{(i,l)\gamma} c_i^\dagger c_l$ and $\hat{\rho}_0 \equiv -\sum_{ik}[\sum_j (V_{ijkj} + V_{jijk})/2] c_i^\dagger c_k$. Since

$$\{\hat{\rho}_\gamma, \hat{\rho}_\gamma^\dagger\} = \frac{1}{2}[(\hat{\rho}_\gamma + \hat{\rho}_\gamma^\dagger)^2 - (\hat{\rho}_\gamma - \hat{\rho}_\gamma^\dagger)^2] \equiv \hat{v}_\gamma^2 - \hat{v'_\gamma}^2, \tag{24}$$

we have succeeded in writing $\hat{H}_2$ in the desired form.
The way to decompose $\hat{H}_2$ above leads to approximately $2N^2$ auxiliary fields. Often the interaction simplifies $\hat{H}_2$ and the number of auxiliary fields can be much reduced. For example, straightforward applications to molecular systems using Gaussian basis sets employed the above decomposition, except that the matrix elements are all real [21, 38] (and terms with λ below a

small threshold can be discarded). A modified Cholesky decomposition has been used which leads to $\mathcal{O}(N)$ fields [32]. In a plane-wave basis, the two-body part is the Fourier transform of $1/|\mathbf{r}_m - \mathbf{r}_n|$ which leads to $1/\mathbf{Q}^2$ [27].

Different forms of the HS transformation exist [39], and can affect the performance of the QMC method. For example, subtracting a mean-field "background" from the two-body term prior to the decomposition improves the calculations [21, 31, 40]. Indeed experience shows that they can not only impact the statistical accuracy, but also lead to different quality of approximations under the constrained-path methods that we discuss below. Therefore, although we discuss the algorithms under the generic form of Eq. (25), we emphasize that it is worthwhile to explore different forms of the HS transformation in an actual implementation [41].

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:

$$e^{-\Delta\tau\hat{H}} = \int d\mathbf{x}\, p(\mathbf{x})\hat{B}(\mathbf{x}), \tag{25}$$

where $p(\mathbf{x})$ is a probability density function (PDF), for example, a multi-dimensional Gaussian. The propagator $\hat{B}(\mathbf{x})$ in Eq. (25) has a *special form*, namely, it is a product of operators of the type in Eq. (12), with U_{ij} depending on the auxiliary field. The matrix representation of $\hat{B}(\mathbf{x})$ will be denoted by $B(\mathbf{x})$.

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.

2.4 A simple example

We now study in some details the ground-state of the one-dimensional Hubbard model with 4 sites and with open boundary conditions We can think of the lattice as a crude representation of particles in a one-dimensional box, with only four equally-spaced grid points to discretize the continuous space inside the box and the kinetic energy written in terms of finite difference. We will consider $N_\uparrow = 2$ and $N_\downarrow = 1$. This can be thought of as a chain of 4 hydrogen atoms, with one of them ionized to H^+, in a minimal basis. We label the four sites 1 through 4.

First let us examine the trivial case of free electrons, i.e. $U = 0$. We can write down the *1-electron* Hamiltonian matrix, based on Eq. (3), which is of dimension 4×4:

$$H = \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \tag{26}$$

The eigenstates of H can be obtained by direct diagonalization. With these eigenstates, we immediately obtain the ground-state wave function $|\psi_0\rangle$ of the 3-electron system from the Pauli

exclusion principle:

$$|\psi_0\rangle = \begin{pmatrix} 0.3717 & -0.6015 \\ 0.6015 & -0.3717 \\ 0.6015 & 0.3717 \\ 0.3717 & 0.6015 \end{pmatrix} \otimes \begin{pmatrix} 0.3717 \\ 0.6015 \\ 0.6015 \\ 0.3717 \end{pmatrix},$$

where the first matrix contains two single-particle orbitals (two columns) for the two $\uparrow$ electrons and the second matrix contains one single-electron orbital for the one $\downarrow$ electron. Each single-electron orbital is an eigenvector of H. Note the second and third lowest single-electron states are degenerate, causing the 3-electron system to be open shell and $|\psi_0\rangle$ to be degenerate. We have simply picked one particular linear combination for the second $\uparrow$-electron in $|\psi_0\rangle$ above.
An object of the form of $|\psi_0\rangle$ is of course nothing more than a Slater determinant. For example, the amplitude of the configuration $|R\rangle = |\downarrow\uparrow\, 0\, \uparrow\rangle$, i.e., two $\uparrow$ electrons on sites 2 and 4 and the one $\downarrow$ electron on site 1, is given by

$$\langle R|\psi_0\rangle = \det \begin{pmatrix} 0.6015 & -0.3717 \\ 0.3717 & 0.6015 \end{pmatrix} \cdot \det \begin{pmatrix} 0.3717 \end{pmatrix}.$$

That is, more formally,

$$|R\rangle = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

and $\langle R|\psi_0\rangle = \det(R^\dagger \cdot \Psi_0)$ is a number, where R and Ψ_0 denote the matrices corresponding to $|R\rangle$ and $|\psi_0\rangle$, respectively.
For this non-interacting system, an alternative (albeit indirect and indeed circular) way of obtaining $|\psi_0\rangle$ is by the power method. From the eigenvalues and eigenvectors of $\hat{K}$, we can easily construct the matrix for $e^{-\Delta\tau\hat{K}}$, which has the same structure as H in Eq. (26) above (i.e., a 4×4 matrix $e^{-\Delta\tau H}$). Denote this matrix by B_K. With an arbitrarily chosen initial Slater determinant $|\psi^{(0)}\rangle$ (with non-zero overlap with $|\psi_0\rangle$), we can then repeatedly apply $e^{-\Delta\tau\hat{H}}$ to carry out the iterative process in Eq. (5), which means multiplying both the $4 \times 2\uparrow$ matrix and the $4 \times 1\downarrow$ matrix by B_K. The process will lead to $|\psi_0\rangle$ as $n \to \infty$.
Note that the solution remains a single Slater determinant during the iteration, as discussed at the end of Sec. 2.1. We could obtain the self-consistent solution of a mean-field equation via this procedure, which looks like a (imaginary-)time-dependent HF or DFT calculation. The orbitals in the determinant (in this example the two $\uparrow$-spin orbitals) must be periodically re-orthonormalized to avoid the loss of numerical precision. This procedure is often done with a modified Gram-Schmidt, and is standard in mean-field calculations. Incidentally, the tendency for the orbitals to collapse to the lowest energy single-particle orbital is intimately related to the sign problem. In any simulation without the use of Slater determinant (second quantization) as its random walker, for example in standard DMC, this is the driving force for the sign problem.

Now let us consider the many-body case, by turning on the interaction U. The first approach of directly diagonalizing H is the method of exact diagonalization. The size of the matrix is now the size of the Hilbert space (not just N), which in this example is 24. Thus the computational cost of this approach grows exponentially with system size. The power method of Eq. (5), on the other hand, can still apply if we can write $e^{-\Delta\tau\hat{H}}$ in some one-electron form. The HS transformation does just that. Assuming $\Delta\tau$ is small and applying the Trotter break-up, we have

$$e^{-\Delta\tau\hat{H}} \Rightarrow \sum_{\mathbf{x}} p(\mathbf{x}) \begin{pmatrix} e^{\gamma x_1} & 0 & 0 & 0 \\ 0 & e^{\gamma x_2} & 0 & 0 \\ 0 & 0 & e^{\gamma x_3} & 0 \\ 0 & 0 & 0 & e^{\gamma x_4} \end{pmatrix} \cdot B_K \otimes \begin{pmatrix} e^{-\gamma x_1} & 0 & 0 & 0 \\ 0 & e^{-\gamma x_2} & 0 & 0 \\ 0 & 0 & e^{-\gamma x_3} & 0 \\ 0 & 0 & 0 & e^{-\gamma x_4} \end{pmatrix} \cdot B_K,$$

where $\mathbf{x} = \{x_1, x_2, x_3, x_4\}$. This is just Eq. (25). Note that $B(\mathbf{x})$ has an $\uparrow$ and a $\downarrow$ component, each of which is a 4×4 matrix. Applying each $B(\mathbf{x})$ to a Slater determinant means precisely the same as in the non-interacting case (with $B_K \otimes B_K$). In other words, $B(\mathbf{x})$ operating on any Slater determinant $|\phi\rangle$ simply involves matrix multiplications for the $\uparrow$ and $\downarrow$ components separately, leading to another Slater determinant $|\phi'\rangle$, as in Eq. (13). Starting from a single Slater determinant, e.g., $|\Psi^{(0)}\rangle = |\psi_0\rangle$, we would end up with a multi-determinant representation of $|\Psi^{(1)}\rangle$, and so on. We will maintain a constant population as described below.

3 Ground-State AFQMC Methods

In this section, we discuss the ground-state AFQMC method. The CPMC and its generalization to the phase problem are free of the decay of the average sign. The methods are approximate, relying on what we will generally refer to as the constrained path approximation.

3.1 Free-projection AFQMC

In this section we briefly describe the ground-state AFQMC method without any constraints. We will rely on the machinery established in the previous section. Our goal is to illustrate the essential ideas, in a way which will facilitate our discussion of the sign problem and help introduce the framework for the constrained path Monte Carlo methods. We will not go into details such as how to efficiently sample the auxiliary fields or how to stabilize the matrices. They are described in the literature. We write the usual path-integral and Metropolis form explicitly here to show that it is the same as the open-ended random walk approach we take in CPMC, as far as understanding the sign problem is concerned.

We want to compute ground-state expectation values $\langle\hat{O}\rangle$ using (6) and (25). The denominator in (6) is

$$\begin{aligned} \langle\psi^{(0)}|e^{-n\Delta\tau\hat{H}}\, e^{-n\Delta\tau\hat{H}}|\psi^{(0)}\rangle &= \int \langle\psi^{(0)}|\Big[\prod_{l=1}^{2n} d\mathbf{x}^{(l)} p(\mathbf{x}^{(l)})\hat{B}(\mathbf{x}^{(l)})\Big]|\psi^{(0)}\rangle \\ &= \int \Big[\prod_{l} d\mathbf{x}^{(l)} p(\mathbf{x}^{(l)})\Big] \det\Big([\Psi^{(0)}]^\dagger \prod_{l} B(\mathbf{x}^{(l)})\Psi^{(0)}\Big). \quad (27) \end{aligned}$$

In the standard ground-state AFQMC method [13], 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. (27), we can write the expectation value of Eq. (6) as

$$\langle \hat{O} \rangle = \frac{\int \overline{\langle \hat{O} \rangle}\, D(X)\, dX}{\int D(X)\, dX} = \frac{\int \overline{\langle \hat{O} \rangle}\, \big|D(X)\big|\, s(X)\, dX}{\int \big|D(X)\big|\, s(X)\, dX}, \tag{28}$$

where

$$s(X) \equiv D(X)/\big|D(X)\big| \tag{29}$$

measures the "sign" of $D(X)$. The non-interacting expectation value for a given X is that defined in Eq. (14)

$$\overline{\langle \hat{O} \rangle} \equiv \frac{\langle \phi_L | \hat{O} | \phi_R \rangle}{\langle \phi_L | \phi_R \rangle} \tag{30}$$

with

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

which are both Slater determinants.
$D(X)$ 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. (28) is therefore in the form of Eq. (62), 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 Eqn. (15) and (16). Often the Metropolis Monte Carlo algorithm [42] 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. (61) in the appendix.
We will carry out free-projection calculations with an open-ended random walk [20, 40] similar to the CP approach, instead of using Metropolis sampling as outlined above [14, 43] which keeps an entire path of a *fixed length* $2n$ as the object to sample. (An example of free-projection is shown in Fig. 2 in Sec. 3.3.) The open-ended random walk framework does not have any ergodicity issues, and it is straightforward to project to longer imaginary-time in order to approach the ground state. This approach also allows a natural implementation of importance sampling and the constraint, as we discuss below.

3.2 Why and how does the sign problem occur?

The sign problem occurs because of the fundamental symmetry between the fermion ground-state $|\Psi_0\rangle$ and its negative $-|\Psi_0\rangle$ [19, 44]. For any ensemble of Slater determinants $\{|\phi\rangle\}$ which gives a Monte Carlo representation of the ground-state wave function, 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 surface is in general *unknown*.

In some special cases symmetry prohibits any crossing of $\mathcal{N}$ in the random walk. The AFQMC calculation is then free of the sign problem. An example [45] is the half-filled ($M_\uparrow + M_\downarrow = N$) repulsive Hubbard model, which via particle-hole symmetry can be mapped to an attractive model with no spin polarization: $M_\uparrow = M_\downarrow$. It is easy to see this in the latter case: if we choose an initial population of walkers of the form $|\phi_\uparrow\rangle \otimes |\phi_\downarrow\rangle$ with $|\phi_\uparrow\rangle = |\phi_\downarrow\rangle$, the structure will be preserved by the propagation. (In general when there is a twist to the boundary condition, the symmetry between $\uparrow$- and $\downarrow$-spins should be of opposite momenta [35].) The overlap of each walker with a wave function $|\Psi_0\rangle$ which observes the same symmetry will be a square and thus always non-negative. In more general cases, however, walkers do cross $\mathcal{N}$ in their propagation by $e^{-\Delta\tau\hat{H}}$. The sign problem then invariably occurs.

In Fig. 1, we illustrate the space of Slater determinants by a one-dimensional (horizontal) line. The "node" $\mathcal{N}$ is the (red) dot in the middle. Imaginary time (or n) is in the vertical direction and increases as the arrow suggests. That is, as the walker moves in the horizontal line, we stretch out continuously "snapshots" of its position along the vertical direction. Now we follow an initial Slater determinant. With no loss of generality, we assume it has a positive overlap with $|\Psi_0\rangle$. At time $n = 0$ it is indicated by the (green) dot on the right. As the random walk evolves, the walker can reach the node, which is the (red) vertical line. At the instant it lands on $\mathcal{N}$, the walker will make no further contribution to the representation of the ground state, since

$$\langle\Psi_0|\phi\rangle = 0 \;\Rightarrow\; \langle\Psi_0|e^{-\beta\hat{H}}|\phi\rangle = 0 \text{ for any } \beta. \tag{31}$$

Paths that result from such a walker have equal probability of being in either half of the Slater determinant space. A few of these possible paths are shown by dashed lines. Computed analytically, they would cancel and not make any contribution in the ground-state wave function, as indicated by their symmetric placement with respect to the node line. But since the random walk has no knowledge of $\mathcal{N}$, these paths continue to be sampled (randomly) in the random walk and become MC noise. Only paths that are completely confined to the right-hand side, as shown by the solid (green) line, will lead to contributions to the ground state, but the relative number of such confined paths decreases exponentially with n. Asymptotically in n, the MC representation of the ground-state wave function consists of an *equal* mixture of the $+$ and $-$ walkers, regardless of where the random walks originated. The Monte Carlo signal is therefore lost. The decay of the signal-to-noise ratio, i.e. the decay of the average sign of $\langle\Psi_T|\phi\rangle$, occurs at an exponential rate with imaginary time.

To eliminate the decay of the signal-to-noise ratio, we impose the constrained-path approximation. Fahy and Hamann first used [25] such a constraint in the framework of the standard AFQMC method. However, the non-local nature of such a constraint proved difficult to implement efficiently in the "path-integral-like" scheme. Here, with the open-ended random walk formalism, the constraint only needs to be imposed one time-step at a time and is extremely simple to implement. It requires that each random walker at each step have a positive overlap with the trial wave function $|\Psi_T\rangle$:

$$\langle\Psi_T|\phi_k^{(n)}\rangle > 0. \tag{32}$$

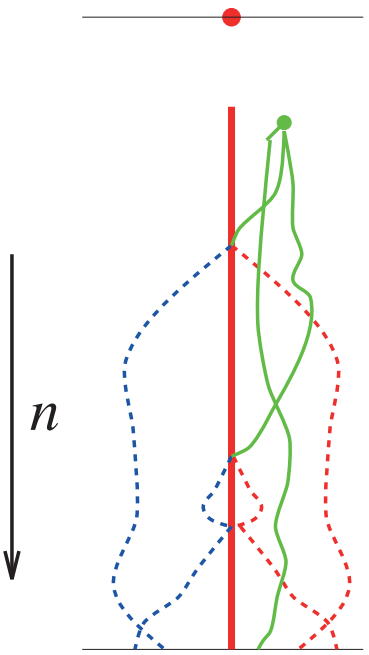

Fig. 1: *Schematic illustration of the sign problem [19]. The top line represents Slater determinant space; the dot represents the "node" $\mathcal{N}$, where a determinant is orthogonal to the ground state $|\Psi_0\rangle$. As the projection continues (increasing n), Slater determinants undergo random walks, tracing out "paths" as shown. When a walker reaches $\mathcal{N}$, its future paths will* collectively *cancel in their contribution to $|\Psi_0\rangle$, indicated by the symmetric distribution of dashed paths about the nodal line. The Monte Carlo sampling, with no knowledge of this cancellation, continues to sample such paths randomly. The relative number of paths with real contributions (solid paths) decreases* exponentially *as n increases.*

This yields an approximate solution to the ground-state wave function, $|\Psi_0^c\rangle = \sum_\phi |\phi\rangle$, in which all Slater determinants $|\phi\rangle$ satisfy (32).

From (31), it follows that the constrained path approximation becomes *exact* for an exact trial wave function $|\Psi_T\rangle = |\Psi_0\rangle$. The overall normalization of walkers remains a constant on the average, since the loss of walkers at $\mathcal{N}$ is compensated by the branching of walkers elsewhere; that is, the eigenvalue problem with $\mathcal{N}$ as boundary has a stable solution.

To implement the constrained-path approximation in the random walk, we define the importance function (see next section) by:

$$O_T(\phi) \equiv \max\{\langle\Psi_T|\phi\rangle, 0\}. \tag{33}$$

This prevents walkers from crossing the trial nodal surface $\mathcal{N}$ and entering the "$-$" half-space as defined by $|\Psi_T\rangle$. A walker k in the n^{th}-step is eliminated (assigned weight zero) for the $(n+1)^{\text{st}}$-step when its $O_T(\phi_k^{(n+1)})$ becomes zero. In the limit $\Delta\tau \to 0$, Eq. (33) ensures that the walker distribution vanishes smoothly at $\mathcal{N}$ and the constrained-path approximation is properly imposed. With a finite $\Delta\tau$, the procedure eliminates walker k when $O_T(\phi_k^{(n+1)}) \leq 0$. The error that results is a form of Trotter error [24] which can be further reduced using what is known as a mirror correction [20, 24].

3.3 The constrained path Monte Carlo method

The constrained path Monte Carlo (CPMC) method [19, 20, 46] incorporates the idea discussed above into a natural and efficient algorithm. The approach applies to Hubbard-like Hamiltonians where the auxiliary-field is real. Even in the case where a twist angle is applied to the boundary condition and the hopping matrix elements are complex, a simple generalization of this condition is sufficient [46]. Using (25), we write Eq. (5) as

$$|\Psi^{(n+1)}\rangle = \int d\mathbf{x}\, p(\mathbf{x})\hat{B}(\mathbf{x})|\Psi^{(n)}\rangle. \tag{34}$$

In the random walk realization of this iteration, we represent the wave function at each stage by a finite ensemble of Slater determinants, i.e.,

$$|\Psi^{(n)}\rangle \propto \sum_k w_k^{(n)}|\phi_k^{(n)}\rangle, \tag{35}$$

where k labels the Slater determinants and an overall normalization factor of the wave function has been omitted. A weight factor $w_k^{(n)}$ is introduced for each walker, even though in Eq. (34) the kernel p is normalized. This is because single-particle orbitals in a Slater determinants cease to be orthogonal to each other as a result of propagation by $\hat{B}$. When they are re-orthogonalized [19, 20], an overall factor appears, which we will view as the w term in the integral equation Eq. (64) of the appendix.

The structure of the random walk now resembles that of Eq. (64). For each random walker we sample an auxiliary-field configuration $\mathbf{x}$ from the PDF $p(\mathbf{x})$ and propagate the walker to a new one via $\Phi_k^{(n+1)} = B(\mathbf{x})\Phi_k^{(n)}$. If necessary, a re-orthogonalization procedure is applied to $\Phi_k^{(n)}$ prior to the propagation: $\Phi_k^{(n)} = [\Phi_k^{(n)}]'R$, where R is an $M \times M$ upper-triangular matrix. $[\Phi_k^{(n)}]'$ is then used in the propagation instead; the weight of the new walker is $w_k^{(n+1)} = w_k^{(n)}\det(R)$.

The simple random walk procedure is sufficient and convenient for thinking about many conceptual issues. It is, however, not efficient as a practical algorithm in most cases of interest, because the sampling of $\mathbf{x}$ is completely random with no regard to the potential contribution to $D(X)$. The idea of importance sampling is to iterate a modified equation with a modified wave function, without changing the underlying eigenvalue problem of (34). Specifically, for each Slater determinant $|\phi\rangle$, we define an importance function as given in Eq. (33) that estimates its overlap with the ground-state wave function. We can rewrite Eq. (34) as

$$|\tilde{\Psi}^{(n+1)}\rangle = \int d\mathbf{x}\tilde{p}(\mathbf{x})\hat{B}(\mathbf{x})|\tilde{\Psi}^{(n)}\rangle, \tag{36}$$

where the modified "PDF" is

$$\tilde{p}(\mathbf{x}) = \frac{O_T(\phi^{(n+1)})}{O_T(\phi^{(n)})}\, p(\mathbf{x})\,. \tag{37}$$

With the new kernel $\tilde{p}$, the probability distribution for $\mathbf{x}$ vanishes smoothly as $O_T(\hat{B}(\mathbf{x})\phi^{(n)})$ approaches zero, and the constraint is naturally imposed. As expected, $\tilde{p}(\mathbf{x})$ is a function of both the current and the next position in Slater-determinant space. Further, it modifies $p(\mathbf{x})$

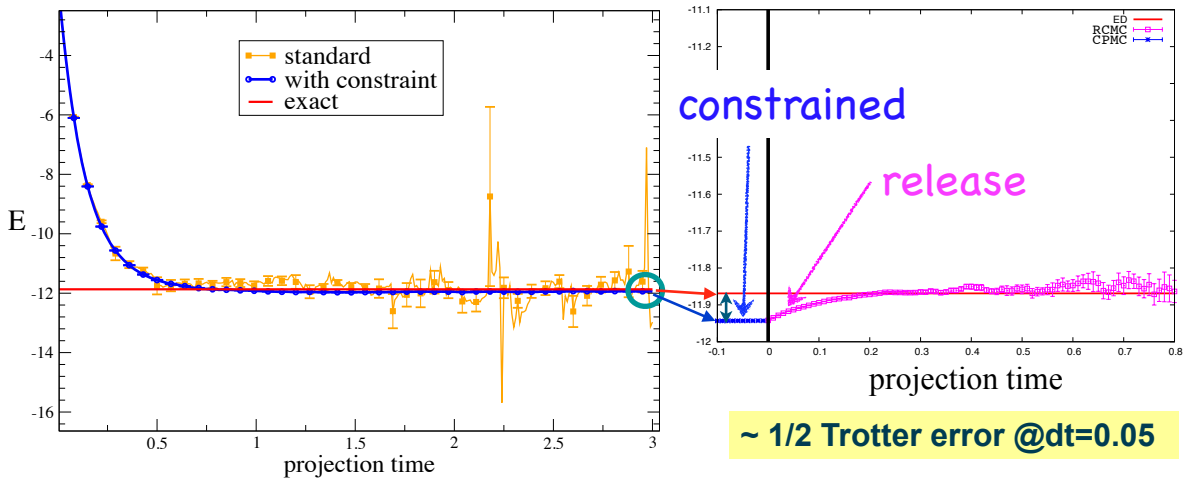

Fig. 2: *Controlling the sign problem with a* systematically improvable *approximation. The total energy is plotted vs. imaginary-time for a* 4×4 *Hubbard model with* $N_\uparrow = N_\downarrow = 7$ *electrons (i.e., density* $n = 0.875$*) and* $U/t = 8$*, which is in the regime important for magnetism and superconductivity, and where the sign problem is severe. The* **left panel** *shows the sign problem in free-projection [15, 22] (yellow), results with constraint [20, 46] (blue), and exact diagonalization (magenta). The* **right panel** *illustrates constraint release [41]. To the left of the vertical bar, a magnified view is shown of the exact and constrained path results. The error from the constraint is roughly half of the Trotter error with the typically recommended value of* $\Delta\tau = 0.05$*. Releasing the constraint removes the small bias and converges to the exact answer, with growing computational cost, as seen on the right.*

such that the probability is increased when $\mathbf{x}$ leads to a determinant with larger overlap and is decreased otherwise. It is trivially verified that Eqs. (34) and (36) are identical. If $|\Psi_T\rangle = |\Psi_0\rangle$, the normalization $\int \tilde{p}(\mathbf{x})d\mathbf{x}$ becomes a constant, which means that the weights of walkers will remain a constant and the random walk will have no fluctuations, as should be the case with perfect importance sampling.

In the random walk, the ensemble of walkers $\{\, |\phi_k^{(n)}\rangle \,\}$ now represents the modified wave function: $|\tilde{\Psi}^{(n)}\rangle \propto \sum_k w_k^{(n)} |\phi_k^{(n)}\rangle$. The true wave function is then given formally by

$$|\Psi^{(n)}\rangle \propto \sum_k w_k^{(n)} \frac{|\phi_k^{(n)}\rangle}{O_T(\phi_k^{(n)})}, \tag{38}$$

although in actual measurements it is $|\tilde{\Psi}^{(n)}\rangle$ that is needed and division by O_T does not appear. The random walk process is similar to that discussed for Eq. (34), but with $p(\mathbf{x})$ replaced by $\tilde{p}(\mathbf{x})$. The latter is in general not a normalized PDF, and we denote the normalization constant for walker k by $N(\phi_k^{(n)})$ and rewrite the iterative relation as

$$|\phi_k^{(n+1)}\rangle \leftarrow N(\phi_k^{(n)}) \int d\mathbf{x} \frac{\tilde{p}(\mathbf{x})}{N(\phi_k^{(n)})} B(\mathbf{x}) |\phi_k^{(n)}\rangle. \tag{39}$$

Thus, for each walker $|\phi_k^{(n)}\rangle$, one step of the random walk consists of:

1. sampling an $\mathbf{x}$ from the PDF $\tilde{p}(\mathbf{x})/N(\phi_k^{(n)})$. With discrete, Ising-like auxiliary-fields for a Hubbard interaction, the sampling is achieved by a heatbath-like algorithm, sweeping through each field x_i [20]

2. constructing the corresponding $B(\mathbf{x})$ and then propagating the walker $\Phi_k^{(n)}$ to generate a new walker

3. assigning a weight $w_k^{(n+1)} = w_k^{(n)} N(\phi_k^{(n)})$ to the new walker.

In contrast with the primitive algorithm in Eq. (34), the weight of a walker does not need to be modified here when the re-orthogonalization procedure is applied. This is because the upper-triangular matrix R only contributes to the overlap O_T, which is already represented by the walker weight. After each re-orthonormalization [24, 47], R can simply be discarded [20, 24].
To calculate the expectation value of an observable which does not commute with the Hamiltonian, the back-propogation technique [20, 30] is used. The idea is to create two coupled populations to represent the bra and ket in Eq. (6), respectively. Because the population in the random walk is importance-sampled, two independent populations which are uncoupled would lead to large fluctuations in the estimator after the importance functions have been "undone." In back-propagation, we choose an iteration n and store the entire population $\{\, |\phi_k^{(n)}\rangle \,\}$. As the random walk proceeds from n, we keep track of the following two items for each new walker: (1) the sampled auxiliary-field values that led to the new walker from its parent walker and (2) an integer label that identifies the parent. After an additional m iterations, we carry out the back-propagation: For each walker l in the $(n+m)^{\text{th}}$ (current) population, we initiate a determinant $\langle\psi_T|$ and act on it with the corresponding propagators, but taken in reverse order. The m successive propagators are constructed from the items stored between steps m and $m+l$, with $\exp(-\Delta\tau\hat{H}_1/2)$ inserted where necessary. The resulting determinants $\langle\bar{\phi}_l^{(m)}|$ are combined with its parent from iteration n to compute $\langle\mathcal{O}\rangle_{\text{BP}}$, in a way similar to the mixed estimator. The weights are given correctly by $w_l^{(n+m)}$, due to importance sampling [30]. The weight works similarly to that in forward walking in DMC. However, the trial wave function can be propagated backwards in AFQMC to obtain a population sampling the bra, from which the entire single-particle Green's function can be computed with Eq. (15). Thus AFQMC can compute both diagonal and off-diagonal observables.

3.4 The phaseless formalism for complex auxiliary-fields

When the many-body Hamiltonian leads to a decomposition with $\lambda > 0$ in Eq. (20), the resulting HS transformation will have complex auxiliary-fields. This is the case for the electron-electron repulsion. (We will sometimes refer loosely to having complex auxiliary-fields. It should be understood that this means 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 [15,41].) In this situation

a phase problem arises, as illustrated in Fig. 3. We next describe the formalism to deal with the problem.

With a continuous auxiliary-field, the importance sampling in step 1 in Sec. 3.3 is achieved with a force bias [15, 30]. To sketch a derivation we write the two-body propagator as

$$\int e^{-\mathbf{x}^2/2}\, e^{\mathbf{x}\cdot\hat{\mathbf{v}}}\, d\mathbf{x}\,, \tag{40}$$

where $\hat{\mathbf{v}}$ is a vector denoting the collection of one-body operators labeled by γ in Eq. (20), and the constant $\sqrt{-\Delta\tau\lambda_\gamma}$, which is complex if λ_γ is positive from in Eq. (21), has been absorbed into the definition of $\hat{\mathbf{v}}$. We introduce a shift in the integral in Eq. (40), which leads to an alternative propagator

$$\int e^{-\mathbf{x}^2/2}\, e^{\mathbf{x}\cdot\bar{\mathbf{x}}-\bar{\mathbf{x}}^2/2}\, e^{(\mathbf{x}-\bar{\mathbf{x}})\cdot\hat{\mathbf{v}}}\, d\mathbf{x}\,. \tag{41}$$

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

In free projection, application of the propagator in Eq. (41) to a Slater determinant $|\phi\rangle$ would be similar to that discussed in the previous section. We would sample $\mathbf{x}$ from the Gaussian, propagate $|\phi\rangle$ to a new $|\phi'(\mathbf{x})\rangle$ by

$$|\phi'(\mathbf{x})\rangle = e^{(\mathbf{x}-\bar{\mathbf{x}})\cdot\hat{\mathbf{v}}}|\phi\rangle, \tag{42}$$

and include the extra exponential with $\bar{\mathbf{x}}$ in the weight of the walker.

In the iterative projection, the ground-state wave function that emerges after the equilibration phase can be written in the following form

$$|\Psi_0\rangle = \int f(\phi)|\phi\rangle d\phi\,, \tag{43}$$

where we assume a criterion has been chosen and is followed consistently to separate the determinant $|\phi\rangle$ from its coefficient (e.g., orthonormalization). Because of the over-completeness of the basis space, the function $f(\phi)$ may be different in different steps of the iteration, but in each step it is uniquely determined. In free projection the function $f(\phi)$ is represented by the weights of the MC samples, and the MC representation of the ground-state wave function is given by Eq. (35). Omitting the explicit walker index, we have

$$|\Psi_0\rangle = \sum_\phi w_\phi|\phi\rangle, \tag{44}$$

where the sum runs over all walkers and w_ϕ is the weight of walker $|\phi\rangle$. In general, the weight w_ϕ can be complex (as can the orbitals), which is of course the cause for the phase problem.

Using the idea of importance sampling, we further modify the propagator in Eq. (41)

$$\int \langle\Psi_T|\phi'(\mathbf{x})\rangle\, e^{-\mathbf{x}^2/2}\, e^{\mathbf{x}\bar{\mathbf{x}}-\bar{\mathbf{x}}^2/2}\, e^{(\mathbf{x}-\bar{\mathbf{x}})\cdot\hat{\mathbf{v}}}\, \frac{1}{\langle\Psi_T|\phi\rangle}\, d\mathbf{x}\,, \tag{45}$$

where, as before, the trial wave function $|\Psi_T\rangle$ represents the best guess to $|\Psi_0\rangle$ and is the one used in the mixed estimate for observables. The new propagator means that the walker weights

now represent, instead of $f(\phi)$ in Eq. (43), a modified function $f(\phi)\langle\Psi_T|\phi\rangle$. In contrast with Eq. (44), the MC representation of the ground-state wave function with importance sampling is now given by Eq. (38)

$$|\Psi_0\rangle = \sum_\phi w_\phi \frac{|\phi\rangle}{\langle\Psi_T|\phi\rangle}\,, \tag{46}$$

where any overall phase of the walker is removed.
We now consider the ratio $\langle\Psi_T|\phi'(\mathbf{x})\rangle/\langle\Psi_T|\phi\rangle$ in Eq. (45). Defining

$$\bar{\mathbf{v}} \equiv -\frac{\langle\Psi_T|\hat{\mathbf{v}}|\phi\rangle}{\langle\Psi_T|\phi\rangle} \sim \mathcal{O}(\sqrt{\Delta\tau}) \tag{47}$$

and

$$\overline{\mathbf{v}^2} \equiv \frac{\langle\Psi_T|\hat{\mathbf{v}}^2|\phi\rangle}{\langle\Psi_T|\phi\rangle} \sim \mathcal{O}(\Delta\tau), \tag{48}$$

we can evaluate the ratio by expanding the propagator [48] in Eq. (42) to $\mathcal{O}(\tau)$. The overall weight factor in Eq. (45) can now be written in the form of an exponential [15, 30]

$$\frac{\langle\Psi_T|\phi'(\mathbf{x})\rangle}{\langle\Psi_T|\phi\rangle} e^{\mathbf{x}\cdot\bar{\mathbf{x}}-\bar{\mathbf{x}}^2/2} \doteq \exp[-(\mathbf{x}-\bar{\mathbf{x}})\cdot\bar{\mathbf{v}} + \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^2\overline{\mathbf{v}^2} - \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^2\bar{\mathbf{v}}^2 + \mathbf{x}\cdot\bar{\mathbf{x}} - \bar{\mathbf{x}}^2/2]. \tag{49}$$

The optimal choice of the shift $\bar{\mathbf{x}}$, which minimizes the fluctuation of Eq. (49) with respect to $\mathbf{x}$, is therefore $\bar{\mathbf{x}} = \bar{\mathbf{v}}$. With this choice, the weight factor in Eq. (49) is $\exp[\mathbf{x}^2\overline{\mathbf{v}^2}/2 + (1-\mathbf{x}^2)\bar{\mathbf{v}}^2/2]$, which can be further reduced to $\exp(\overline{\mathbf{v}^2}/2)$ if we set $\mathbf{x}^2 \to 1$. Setting $\mathbf{x}^2$ to its average value (with respect to the Gaussian PDF) is justified if $\bar{\mathbf{v}}$ and $\overline{\mathbf{v}^2}$ remain approximately constant within one imaginary-time step, which holds in the limit $\Delta\tau \to 0$. (An exception to this is when the walker $|\phi\rangle$ is in the vicinity — say, within a few time steps — of "the origin" as defined by $\langle\Psi_T|\phi\rangle = 0$. We discuss this special case, together with the characteristics of the origin and its effect on the method below.) The modified propagator in Eq. (45) can then be written approximately as

$$\int e^{-\mathbf{x}^2/2}\, e^{(\mathbf{x}-\bar{\mathbf{v}})\cdot\hat{\mathbf{v}}}\, e^{\overline{\mathbf{v}^2}/2}\, d\mathbf{x}\,. \tag{50}$$

We have only included the $\hat{H}_2$ terms in the discussion from Eq. (40) up to now. Re-introducing $\hat{H}_1$ into the propagator, we obtain the complete propagator for the Hamiltonian

$$\int e^{-\mathbf{x}^2/2} \exp\left[-\frac{\Delta\tau\hat{H}_1}{2}\right] \exp[(\mathbf{x}-\bar{\mathbf{v}})\cdot\hat{\mathbf{v}}] \exp\left[-\frac{\Delta\tau\hat{H}_1}{2}\right] \exp[-\Delta\tau E_L(\phi)]\, d\mathbf{x}\,, \tag{51}$$

where the term E_L has a similar form to the so-called local energy in real-space DMC

$$E_L(\phi) \equiv \frac{\langle\Psi_T|\hat{H}|\phi\rangle}{\langle\Psi_T|\phi\rangle}\,. \tag{52}$$

Projection with Eq. (51) will in principle lead to the ground-state wave function in the form of Eq. (38). The weight of the walker is determined by E_L, which is independent of any phase factor of the determinant.

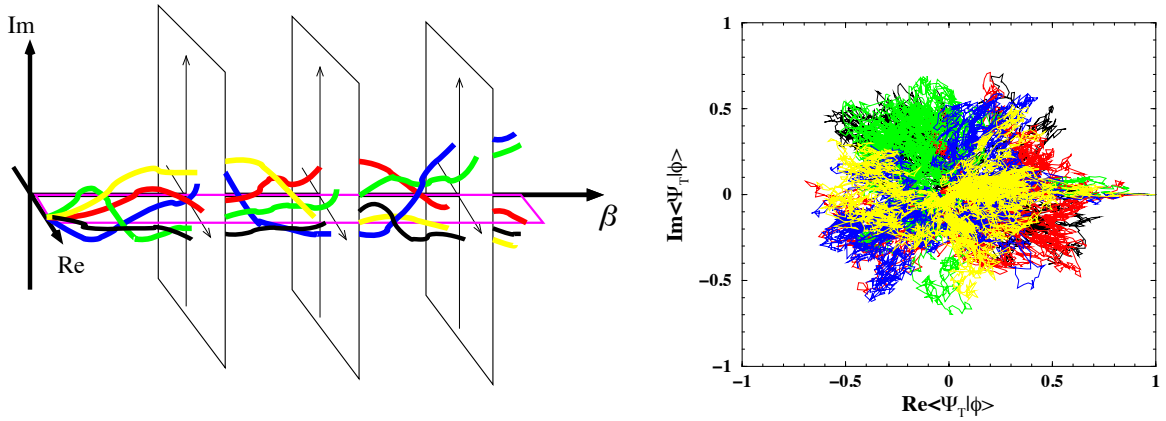

Fig. 3: *Schematic illustration of the phase problem and constraints to control it. The* **left panel** *shows, as a function of projection time $\beta \equiv n\Delta\tau$, trajectories of 5 walkers characterized by the real (Re) and imaginary (Im) parts of their overlap with the ground-state wave function. 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 [15] is to apply a gauge transformation such that confining the random walk in the single magenta plane (left) is a good approximation.*

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. (7) is phaseless

$$E_0^{\rm c} = \frac{\sum_\phi w_\phi E_L(\phi)}{\sum_\phi w_\phi}\,. \tag{53}$$

With a general $|\Psi_T\rangle$ which is not exact, a natural approximation is to replace E_L in Eq. (51) by its real part, ${\rm Re}E_L$. The same replacement is then necessary in Eq. (53).
With these replacements we arrive at a phaseless formalism for the random walks. We can summarize each step in this formalism as follows. For each random walker $|\phi\rangle$,

(a) sample $\mathbf{x}$ and propagate the walker to $|\phi'\rangle$

$$|\phi\rangle \to |\phi'\rangle = \exp\left[-\frac{\Delta\tau\hat{H}_1}{2}\right]\,\exp\left[(\mathbf{x}-\bar{\mathbf{v}})\cdot\hat{\mathbf{v}}\right]\,\exp[-\frac{\Delta\tau\hat{H}_1}{2}\right]\,|\phi\rangle, \tag{54}$$

(b) update the weight of the walker

$$w_\phi \to w_{\phi'} = w_\phi \exp\Big[-\Delta\tau\cdot{\rm Re}\Big(E_L(\phi')+E_L(\phi)\Big)/2\Big]. \tag{55}$$

Walkers so generated represent the ground-state wave function with importance sampling, in the sense of Eq. (38). We have also tested a slightly different formalism based on Eq. (45), which we shall refer to as the *hybrid formalism* [15, 30]. The same choice for $\bar{\mathbf{x}}$ is used: $\bar{\mathbf{x}} = \bar{\mathbf{v}}$. The walker is propagated in the same way as in **(a)** above, with the shifted or force-biased propagator

in Eq. (51). In **(b)**, however, we explicitly calculate the weight factor on the left-hand side of Eq. (49) instead of using the local energy:

(b′) update the weight of the walker

$$w_\phi \to w_{\phi'} = \left| \frac{\langle \Psi_T | \phi' \rangle}{\langle \Psi_T | \phi \rangle} e^{\mathbf{x}\cdot\bar{\mathbf{x}} - \bar{\mathbf{x}}^2/2} \right| \tag{56}$$

Conceptually the hybrid form is the same as that of the full local energy. However, it is in principle correct for any choice of $\bar{\mathbf{x}}$, while in the local-energy form only $\bar{\mathbf{x}} = \bar{\mathbf{v}}$ is correct. If calculating E_L is computationally costly (for example a brute-force calculation of $\hat{H}_2$ in a general representation in Eq. (2) costs $\mathcal{O}(N^4)$), the hybrid form can be a more efficient alternative. The Trotter errors for finite $\Delta\tau$ can be different in the two approaches [22].

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. 3, 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 CPMC. 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 the nature of the problem we consider the phase of $\langle \Psi_T | \phi'(\mathbf{x} - \bar{\mathbf{x}}) \rangle / \langle \Psi_T | \phi \rangle$, which we shall denote by $\Delta\theta$ and which is in general non-zero: $\Delta\theta \sim \mathcal{O}(-\mathbf{x}\,\mathrm{Im}(\bar{\mathbf{x}}))$. This means that the walkers will 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. 3, 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 all real initially started with $\langle \Psi_T | \phi^{(0)} \rangle = 1$) 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 $\cos(\Delta\theta)$

$$w_{\phi'} = w_{\phi'} \max\{0, \cos(\Delta\theta)\}\,, \tag{57}$$

in addition to step (**b**) [or (**b′**)] above. For this to be a good approximation, it is important that the importance sampling transformation has been done to eliminate the leading order in the overall phase of $|\phi\rangle$ in the propagator in Eq. (49).

Several alternative schemes have been tested [15, 31, 49] in place of the Eq. (57). One that seemed to work just as well was to project with the factor $\exp\{-[\mathrm{Im}(\bar{\mathbf{x}})]^2/2\}$, 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.

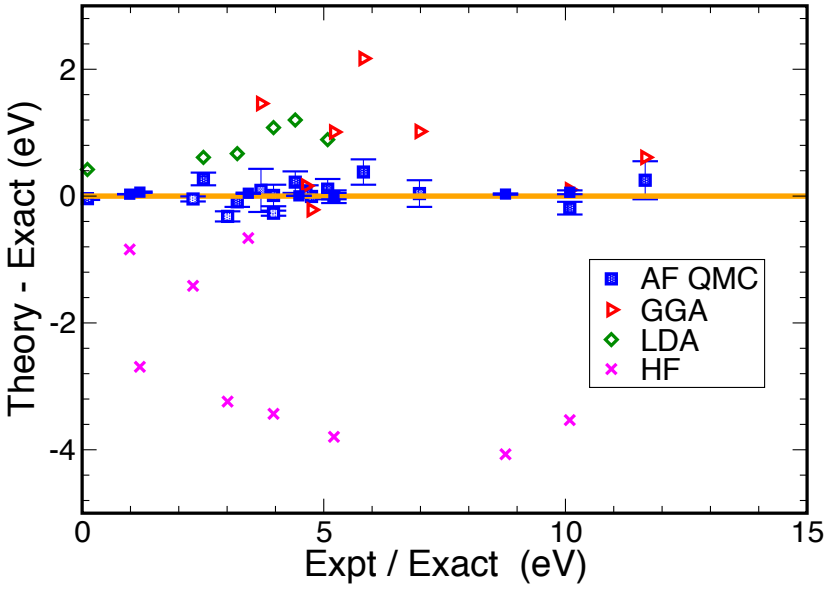

Fig. 4: *Calculated binding energies [16, 27, 49, 50] of molecules compared with experimental values. The discrepancy between theory and experiment is plotted. The AFQMC is fed a trial wave function to start, which is taken directly from DFT [with either LDA or the generalized-gradient approximation (GGA) functionals] or HF. The corresponding DFT or HF results are also shown. As can be readily observed, the AFQMC results are in excellent agreement with experiment and significantly improve upon the values from DFT and HF.*

4 Illustrative results

The AFQMC method, including the CPMC for real auxiliary-fields and the phaseless formalism for complex fields, has been applied to lattice models, to solids (using plane-wave basis and pseudo potentials), to molecular systems (using Gaussian basis sets), and to down-folded model Hamiltonians of real materials (using DFT orbitals as basis sets). It has also been applied in nuclear physics to treat shell models and in neutron matter calculations. Here we briefly mention a few examples from correlated-electron systems to provide an idea of what can already be done and the many opportunities for improvements and for breakthroughs.

Accuracy, or predictive power, is a key requirement of a successful paradigm for strongly correlated systems. The formulation of AFQMC with Gaussian basis sets and quantum chemistry applications have been crucial in establishing the method as a general approach that can now be applied to correlated materials, since direct comparisons can be made with high-level quantum chemistry results. Benchmark calculations to date have included close to 100 systems, from simple molecules [21,49] and solids [22,51,52] to transition metal systems [53] to energy storage problems [32].

Figure 4 illustrates the results on molecules using both planewave plus pseudopotentials and Gaussian basis sets. In these calculations we have operated largely in an automated mode, inputting only the DFT or HF solutions. The method demonstrated excellent accuracy, being consistently able to correct errors in the mean-field trial wave function. In molecules, the accuracy of the phaseless AFQMC is found to be comparable to CCSD(T), the gold standard in

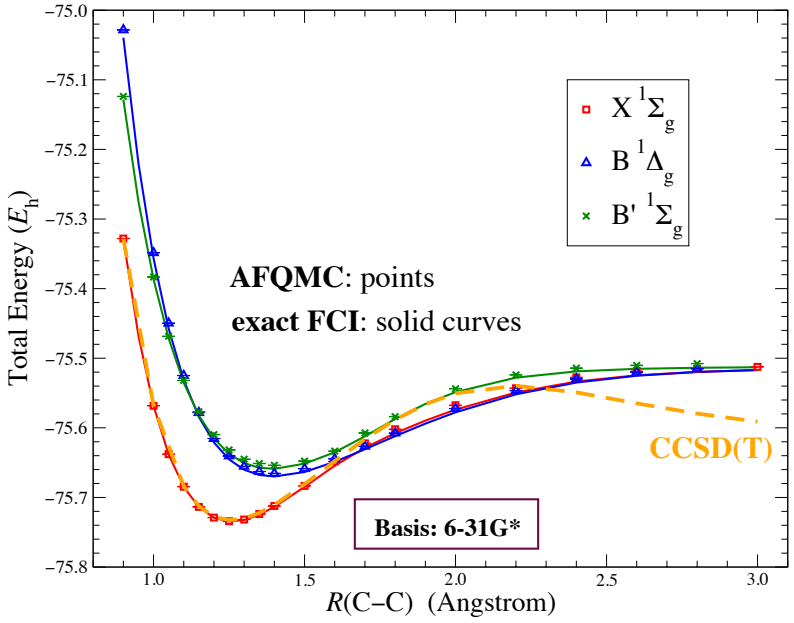

Fig. 5: *Accurate calculation [58] of excited states. The seemingly simple C_2 molecule is a significant challenge to many-body methods [59, 60]. Potential energy curves calculated by our QMC method (symbols with error bars) are shown for the three lowest lying singlet states in C_2, compared with exact results [59] (solid lines) in a small basis set. RCCSD(T), the preeminent many-body quantum chemistry method, does not work well (shown in orange lines).*

chemistry [54, 55], near equilibrium geometry.

Bond stretching in molecules and solids mimics increased levels of multi-reference characters and particle correlation. With single Slater determinants from unrestricted HF as trial wave function, the AFQMC method generally gives better overall accuracy and a more uniform behavior than CCSD(T) in mapping the potential-energy curve [56–58], as can be seen for the example of the C_2 ground state in Fig. 5.

The AFQMC method can be used to study excited states. Prevention of collapse into the ground state and control of the fermion sign/phase problem are accomplished by the constraint using an excited state trial wave function [58]. An additional orthogonalization constraint is formulated to use virtual orbitals in solids for band structure calculations [52]. Using the C_2 molecule as a test case, we calculated the potential energy curves of the ground and two low-lying singlet excited states in Fig. 5. The trial wave function was obtained by truncating complete active space wave functions, with no further optimization. The phaseless AFQMC results using a small basis set were in good agreement with exact calculations, while those using large basis sets were in excellent agreement with experimental spectroscopic constants [58]. Applications in solids have included the calculation of the fundamental band gap in wurtzite ZnO [52].

Figure 6 illustrates an application to determine [23, 46] magnetic correlations and collective modes in the ground state of the two-dimensional repulsive Hubbard model. Using CPMC and simulating large rectangular supercells (over 1000 lattice sites), we studied the magnetic and charge properties as a function of density. At intermediate interaction strengths, an incommensurate spin density wave (SDW) state is seen, with antiferromagnetic order and essentially ho-

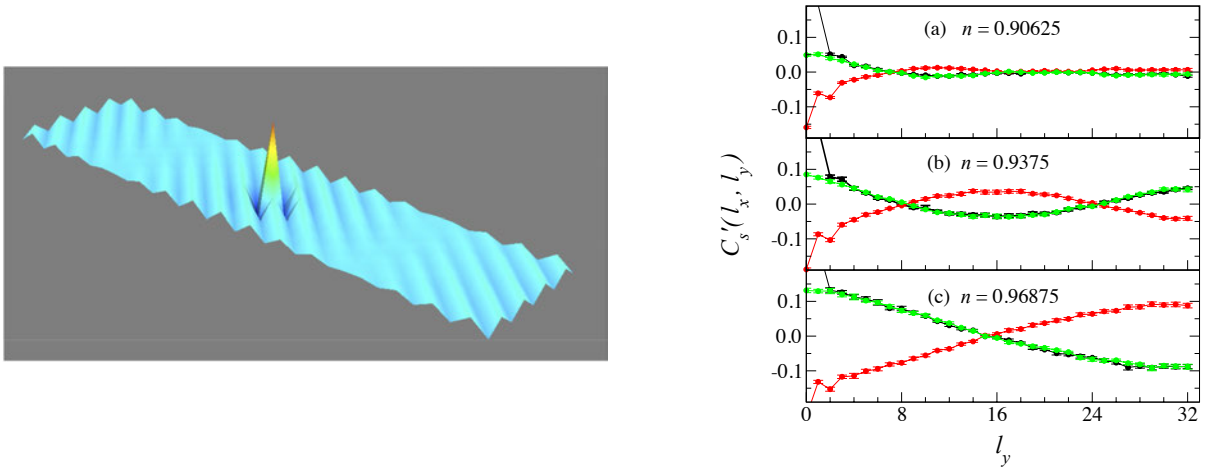

Fig. 6: *Spin-density wave (SDW) states with long wavelength modulation in the 2-D Hubbard model [23]. The* **left panel** *is a 3-D plot of the spin-spin correlation function in the ground state of* 8×32 *lattice with* 240 *electrons (*$n = 0.9375$*) and* $U/t = 4$*. Anti-ferromagnetic order can be seen from the diagonal "lines." Phase shifts occur at the nodal lines, across which there is a mismatch of the diagonal lines. The* **right panel** *examines the doping dependence. The staggered spin-spin correlation function is plotted at three different densities. Calculations are done using a* free-electron *trial wave function. The long wavelength modulation and nodal lines (two colors crossing) are clear. As doping is increased, the wavelength of the modulating wave decreases, as does the amplitude of the SDW.*

mogeneous charge correlations. The wavelength of the collective mode decreases with doping $(1 - (M_\uparrow + M_\downarrow)/N)$, as does its magnitude. The SDW order vanishes beyond a critical doping. As the interaction is increased, the holes go from a wavelike (delocalized) to a particlelike (localized) state, and charge ordering develops which eventually evolves into stripe states [61]. This work [23] advanced our understanding of the magnetic properties and provided predictions for optical lattice experiments.

5 Concluding remarks

We have discussed ground-state auxiliary-field-based methods for correlated-electron systems. A formalism has been outlined that allows for a systematic understanding of the origin of the sign/phase problem, and a platform on which to build methods whose computational cost scales with system size at a low power, and whose accuracy reaches or surpasses the best general many-body approaches. The AFQMC methods can be used to study both model systems and realistic Hamiltonians. Applications have included atoms, molecules, solids, and correlated-electron models, recently including down-folded Hamiltonians.

The original form of the AFQMC method allows essentially exact calculations of ground-state and finite-temperature equilibrium properties of interacting many-fermion systems. Their effectiveness, however, was severely limited by the sign problem, which prevents calculations at large system sizes or low temperatures. 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." It is hoped that the methods we have discussed will change this paradigm. Calculations indeed can be done in systems where a sign problem is present, and often accurate results are within reach by these calculations.
As a QMC method, this framework automatically accounts for particle permutations and allows easy computations of both diagonal and off-diagonal expectations, as well as imaginary-time correlations. It has much versatility as a general method, offering, for example, the possibility to treat heavier elements and spin-orbit coupling, and the potential to embed the AFQMC naturally and seamlessly in a calculation at the mean-field level. While much work is needed to study various forms of HS transformations and trial functions $|\psi_T\rangle$, and to understand the subtleties of the method and the constraints because of the non-orthogonal and over-complete nature of the Slater determinant space involved, we expect the method and the concept brought forth here to see many applications, and to significantly enhance the applicability of quantum simulations in interacting fermion systems. There are many opportunities both for further development and improvement of this framework, and for applications in a variety of problems.

6 Acknowledgments

I am indebted to numerous colleagues and students, whose interactions and collaborations I have immensely enjoyed and benefitted from. Length restriction does not allow the full list here, but I would like to thank W. Al-Saidi, J. Carlson, C.-C. Chang, H. Krakauer, F. Ma, W. Purwanto, and H. Shi. The work covered in these notes would not have been possible without their contributions. Support from the National Science Foundation (NSF), the Department of Energy (DOE), and the Army Research Office (ARO) is gratefully acknowledged. Computing was done on the Oak Ridge Leadership Computing Facilities via an INCITE award, on Blue Waters at UIUC supported by NSF PRAC, and on CPD at William & Mary.

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. [42].
MC methods are often used to compute many-dimensional integrals of the form

$$G = \frac{\int_{\Omega_0} f(\mathbf{x}) g(\mathbf{x}) d\mathbf{x}}{\int_{\Omega_0} f(\mathbf{x}) d\mathbf{x}}\,, \tag{58}$$

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. (58) 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 $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_i, \cdots\}$ so that the probability that any $\mathbf{x}_i$ is in the sub-domain $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$ is

$$\mathrm{Prob}\{\mathbf{x}_i \in (\mathbf{x}, \mathbf{x} + d\mathbf{x})\} = \bar{f}(\mathbf{x}) d\mathbf{x}\,. \tag{59}$$

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 [42] 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

$$f(\mathbf{x}) \propto \sum_i w_i \delta(\mathbf{x} - \mathbf{x}_i)\,. \tag{60}$$

Given $\mathcal{M}$ independent samples from $f(\mathbf{x})$, the integral in Eq. (58) is estimated by

$$G_{\mathcal{M}} = \sum_{i=1}^{\mathcal{M}} w_i g(\mathbf{x}_i) / \sum_{i=1}^{\mathcal{M}} w_i \tag{61}$$

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

$$G' = \frac{\int_{\Omega_0} f(\mathbf{x}) g(\mathbf{x}) h(\mathbf{x}) d\mathbf{x}}{\int_{\Omega_0} f(\mathbf{x}) h(\mathbf{x}) d\mathbf{x}}, \tag{62}$$

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

$$G'_{\mathcal{M}} = \frac{\sum_{i=1}^{\mathcal{M}} w_i g(\mathbf{x}_i) h(\mathbf{x}_i)}{\sum_{i=1}^{\mathcal{M}} w_i h(\mathbf{x}_i)}, \tag{63}$$

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 [42], the integral equation

$$\Psi'(\mathbf{x}) = \int_{\Omega_0} K(\mathbf{x},\mathbf{y})\ w(\mathbf{y})\ \Psi(\mathbf{y}) d\mathbf{y}, \tag{64}$$

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. (58)), $w(\mathbf{y}) \geq 0$, 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

$$\Psi(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-(x-y)^2}\ \sqrt{2} e^{-y^2/2}\ \Psi(y) dy, \tag{65}$$

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 $\Psi(x)$.

References

[1] See, e.g., N.-P. Ong and R. Bhatt (Ed.):
More is different: Fifty years of condensed matter physics (Princeton Univ. Press, 2001)

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

[3] See, e.g., A. Szabo and N.S. Ostlund: *Modern quantum chemistry*
(McGraw-Hill, New York, 1989)

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

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

[6] S.Y. Savrasov, G. Kotliar, and E. Abrahams, Nature **410**, 793 (2001)

[7] A.I. Lichtenstein, M.I. Katsnelson, and G. Kotliar, Phys. Rev. Lett. **87**, 067205 (2001)

[8] K. Held, A.K. McMahan, and R.T. Scalettar, Phys. Rev. Lett. **87**, 276404 (2001)

[9] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998)

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

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

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

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

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

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

[16] 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, Journal of Physics: Conference Series **125**, 012057 (2008)

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

[18] 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)

[19] S. Zhang: *Constrained path Monte Carlo for fermions* (cond-mat/9909090)
in M.P. Nightingale and C.J. Umrigar (eds.):
Quantum Monte Carlo Methods in Physics and Chemistry (Kluwer Dordrecht, 1999)

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

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

[22] W. Purwanto, H. Krakauer, and S. Zhang, Phys. Rev. B **80**, 214116 (2009)

[23] C.-C. Chang and S. Zhang, Phys. Rev. Lett. **104**, 116402 (2010)

[24] S. Zhang: in D. Sénéchal, A.-M. Tremblay, and C. Bourbonnais (eds.): *Theoretical Methods for Strongly Correlated Electrons* (Springer, 2003), and references therein; `http://physics.wm.edu/~shiwei`

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

[26] J. Carlson, J.E. Gubernatis, G. Ortiz, and S. Zhang, Phys. Rev. B **59**, 12788 (1999)

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

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

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

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

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

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

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

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

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

[36] R.L. Stratonovich, Sov. Phys. Dokl. **2**, 416 (1958)

[37] J.E. Hirsch, Phys. Rev. B **28**, 4059 (1983)

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

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

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

[41] H. Shi and S. Zhang, Phys. Rev. B, arXiv:1307.2147 (2013)

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

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

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

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

[46] C.-C. Chang and S. Zhang, Phys. Rev. B **78**, 165101 (2008)

[47] S.R. White, D.J. Scalapino, R.L. Sugar, E.Y. Loh, J.E. Gubernatis, and R.T. Scalettar, Phys. Rev. B **40**, 506 (1989)

[48] J.W. Moskowitz, K.E. Schmidt, M.A. Lee, and M.H. Kalos, J. Chem. Phys. **77**, 349 (1982)

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

[50] W.A. Al-Saidi, H. Krakauer, and S. Zhang, Phys. Rev. B **73**, 075103 (2006)

[51] H. Kwee, S. Zhang, and H. Krakauer, Phys. Rev. Lett. **100**, 126404 (2008)

[52] F. Ma, S. Zhang, and H. Krakauer, New J. of Phys. (arXiv:1211.4635) (2013)

[53] Y. Virgus, W. Purwanto, H. Krakauer, and S. Zhang, Phys. Rev. B **86**, 241406 (2012)

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

[55] T.D. Crawford and H.F. Schaefer III, Reviews in Computational Chemistry **14**, 33 (2000)

[56] W.A. Al-Saidi, S. Zhang, and H. Krakauer, J. Chem. Phys. **127**, 144101 (2007)

[57] W. Purwanto, W.A. Al-Saidi, H. Krakauer, and S. Zhang, J. Chem. Phys. **128**, 114309 (2008)

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

[59] M.L. Abrams and C.D. Sherrill, J. Chem. Phys. **121**, 9211 (2004)

[60] C.J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R.G. Hennig, Phys. Rev. Lett. **98**, 110201 (2007)

[61] S.R. White and D.J. Scalapino, Phys. Rev. Lett. **91**, 136403 (2003)