Hybrid Auxiliary Field Quantum Monte Carlo for Molecular Systems

Yixiao Chen, Linfeng Zhang, Weinan E, and Roberto Car*

ABSTRACT: We propose a quantum Monte Carlo approach to solve the many-body Schrödinger equation for the electronic ground state. The method combines optimization from variational Monte Carlo and propagation from auxiliary field quantum Monte Carlo in a way that significantly alleviates the sign problem. In application to molecular systems, we obtain highly accurate results for configurations dominated by either dynamic or static electronic correlation.

1. INTRODUCTION

Calculating the ground state solution of the many-electron Schrödinger equation is one of the most fundamental problems in computational physics and chemistry, particularly so in the machine learning era in which solutions of the many-electron Schrödinger equation serve as the ultimate source of accurate data for a hierarchy of physical models across several scales. In principle, the exact solution can be found by methods like full configuration interaction (FCI), but their computational cost scales badly (in the case of FCI, exponentially) with system size. Consequently, their application is limited to very small systems. The search for a good approximation to the ground-state electronic problem remains very challenging after many decades of intense efforts.

Traditional quantum chemistry methods, such as configuration interaction, coupled cluster, and Møller–Plesset perturbation theory, approximate the solution by hierarchical expansions, which allow systematic improvement by raising the order of the expansions, at increasing computational costs. These methods work very well for systems dominated by dynamic correlation. However, the expansions on which they rely converge slowly in the presence of a strong static correlation.

Several promising methods have been proposed to deal with strongly correlated systems. Multireference methods often include a large number of determinants in the zeroth-order approximation. In addition, the a priori choice of the active space may affect significantly the results. Methods like density matrix renormalization group (DMRG) and heat-bath configuration interaction (HCI) introduce a cutoff (like the bond dimension in DMRG, and the number of determinants in HCI) and solve the truncated problem to avoid exponential scaling. Usually, these methods require extrapolation of the cutoff to the truncation-free limit, a procedure that is hampered by large computational cost for large cutoffs.

By introducing stochastic sampling, quantum Monte Carlo (QMC) methods are able to reach graceful (N³ to N⁴) scaling and treat dynamic and static correlation on equal footing. There are two main approaches in QMC. One scheme, Variational Monte Carlo (VMC), approximates the ground state by minimizing the energy expectation value

\[ E_0 = \min_\theta \mathbb{E}[\psi_\theta | H | \psi_\theta] \]

By introducing stochastic sampling, quantum Monte Carlo (QMC) methods are able to reach graceful (N³ to N⁴) scaling and treat dynamic and static correlation on equal footing. There are two main approaches in QMC. One scheme, Variational Monte Carlo (VMC), approximates the ground state by minimizing the energy expectation value

\[ E_0 = \min_\theta \mathbb{E}[\psi_\theta | H | \psi_\theta] \]

The method combines optimization from variational Monte Carlo and propagation from auxiliary field quantum Monte Carlo in a way that significantly alleviates the sign problem. In application to molecular systems, we obtain highly accurate results for configurations dominated by either dynamic or static electronic correlation.

By introducing stochastic sampling, quantum Monte Carlo (QMC) methods are able to reach graceful (N³ to N⁴) scaling and treat dynamic and static correlation on equal footing. There are two main approaches in QMC. One scheme, Variational Monte Carlo (VMC), approximates the ground state by minimizing the energy expectation value

\[ E_0 = \min_\theta \mathbb{E}[\psi_\theta | H | \psi_\theta] \]

The method combines optimization from variational Monte Carlo and propagation from auxiliary field quantum Monte Carlo in a way that significantly alleviates the sign problem. In application to molecular systems, we obtain highly accurate results for configurations dominated by either dynamic or static electronic correlation.

Special Issue: Machine Learning for Molecular Simulation

Received: January 9, 2023
Published: April 18, 2023
its second-quantized formulation and use of exponential projectors. Nevertheless, the sign problem is still present in AFQMC simulations, where it is usually eliminated by adopting a phaseless approximation.

In this work, we adopt an alternative strategy to overcome these limitations by variationally optimizing the imaginary time propagator so that it approaches rapidly the ground state before the sign problem manifests. Approaches of this kind have been developed for multiple applications, including the simulation of quantum models on a lattice and the preparation of quantum states for quantum computing applications. Recently, Sorella generalized this idea in the context of the AFQMC framework to study the phase diagram of the (zero temperature) Hubbard model. In this work, we propose a similar but more flexible ansatz to study the ground state of molecular systems. The ground state energy predicted in this way can be further refined by a short propagation using standard AFQMC to reach near-exact accuracy. With our approach, the sign problem is substantially mitigated. In addition, the method can be systematically improved by extending the projector ansatz and prolonging the propagation time. We call the method Hybrid Auxiliary Field Quantum Monte Carlo (HAFQMC) to emphasize the combination of the variational and projector approaches within the AFQMC framework. We test the method on various molecular systems including different spatial configurations of $H_2$, $N_2$, $C_6H_6$ (benzene), and $C_6H_4$ (cyclobutadiene) molecules. The studied molecular conformations are dominated by either static or dynamic correlation. We find that the method performs uniformly well and reaches near-exact accuracy for all the tested systems, without suffering from any serious sign problem.

2. METHODS

We start by briefly introducing the standard AFQMC method. A more detailed discussion can be found, e.g., in ref 27. Consider the Hamiltonian of an electronic system written in second quantized form, with a given single-particle basis:

$$H = T + V = \sum_{\lambda} T_{\lambda} c_\lambda^\dagger c_\lambda + \frac{1}{2} \sum_{\lambda\mu} V_{\lambda\mu} c_\lambda^\dagger c_\mu^\dagger c_\mu c_\lambda$$

(1)

where $i, j, k, l$ are basis set indices, and $T_{\lambda}$ and $V_{\lambda\mu}$ are the corresponding matrix elements of the one- and two-body operators. The ground state $|\psi_0\rangle$ of this system can be obtained (up to a normalization constant) by projecting it out from an initial state $|\psi_i\rangle$ that has a nonzero overlap with the ground state, using imaginary-time propagation:

$$|\psi_i\rangle \rightarrow |\psi_i(t)\rangle = \lim_{t \rightarrow \infty} e^{-\Delta t \hat{H}}|\psi_i\rangle$$

(2)

In the framework of standard AFQMC, the propagator is split into $N$ small time steps, i.e., $e^{-\Delta t \hat{H}} = e^{-N \Delta t \hat{H}}$, where each step is approximated by the Trotter–Suzuki decomposition:

$$e^{-\Delta t \hat{H}} = e^{-\frac{\Delta t}{2} \hat{H}} e^{-\frac{\Delta t}{2} \hat{H}} + O(\Delta t^3)$$

(3)

The initial state $|\psi_i\rangle$ is assumed to be a single Slater determinant or a linear combination of Slater determinants. The action of the one-body term on $|\psi_i\rangle$ is easily evaluated as application of the one-body term on a Slater determinant yields another Slater determinant by virtue of the Thouless theorem. In order to handle the action of the two-body term, one expresses it as a sum of squares and then performs the Hubbard–Stratonovich transformation, by introducing a set of auxiliary fields $\{x_i\}$:

$$e^{-\Delta t \hat{V}} = \int \left( \prod_{\tau} \frac{dx_{\tau}}{\sqrt{2\pi}} e^{-\frac{1}{2} x_{\tau}^2} \right) \exp \left( \sqrt{-\Delta t} \sum_{\tau} x_{\tau} \hat{L}_{\gamma} \right)$$

(4)

Here, $\hat{L}_{\gamma} = \sum_{ij} \hat{L}_{ij} x_i x_j$ are one-body operators satisfying $\hat{V} = \frac{1}{\gamma} \sum_{\tau} \hat{L}_{\tau}$, that can be generated using, e.g., a modified Cholesky decomposition. The propagator is now mapped onto a high-dimensional integral of one-body terms interacting with the auxiliary fields. Such integrals can be evaluated using Monte Carlo (MC) techniques.

Although asymptotically exact, standard AFQMC is faced with the Fermionic sign problem, as the statistical error of the MC process grows exponentially with the propagation time $t$. In our approach, we limit the total propagation time to a finite value, and variationally optimize the propagator to approach the ground state. In order to improve flexibility and variational freedom, we allow the propagator to be different at different imaginary times, and we optimize the initial Slater determinant as well, leading to the following variational ansatz:

$$|\psi_0\rangle = \exp \left( \int_0^t -\hat{H}(s) ds \right) |\psi_i\rangle$$

(5)

where $\theta$ denotes the collection of all parameters in the ansatz, and we use hereafter the tilde symbol (~~) on operators and wave functions to indicate the presence of optimizeable parameters. In practice, we follow a similar procedure as in standard AFQMC in eq 3 and (4). Upon performing time discretization and the Hubbard-Stratonovich transformation, we arrive at the following expression for the ansatz:

$$|\psi_0\rangle = \exp \left( \prod_{\tau} \frac{dx_{\tau}}{\sqrt{2\pi}} e^{-\frac{1}{2} x_{\tau}^2} \right) \left( \prod_{i} \exp[-t_i \hat{L}_i] \right) \exp \left( \sqrt{-\Delta t} \sum_{\tau} x_{\tau} \hat{L}_{\tau} \right) |\psi_i\rangle$$

(6)

where $l$ is the index for discretized time steps, $t_i$ and $s_l$ are step sizes for one- and two-body terms respectively, and $\{x_{\tau}\}$ denotes the set of all auxiliary field variables. We take $t_i$ and $s_l$ to be optimizable as well. For simplicity, we let $L_{\tau}$ be the same at different time steps, and we use matrix elements of the operators directly as optimization parameters. In principle, the sign problem is not eliminated. In practice, when using a small number of projection steps, the method is free from the sign problem if the ansatz is optimized subject to the soft restraint described below. We found that this ansatz is expressive enough, so that a good accuracy can be reached within about three to five time steps. We also note that, thanks to the variational formalism, there is no need to worry about the Trotter error in eq 3: $t_i$ and $s_l$ can be much larger than those allowed in standard AFQMC. For the same reason, the exponential operator can be approximated by a low order
Taylor expansion such as $e^{\hat{A}} \approx 1 + \hat{A} + \frac{1}{2!} \hat{A}^2$ adopted in this paper to speed up the computation.

Similar approaches achieving good results have been used in multiple fields. The idea of using propagators as variational ansatze was first employed to calculate the ground state of the Hubbard model using the time step values as optimizable parameters. Sorella extended this ansatz to study the phase diagram of the Hubbard model, by optimizing the one-body operator $\hat{T}$ and the initial state in addition to the time step values. He still kept the interaction $\hat{V}$ fixed and the propagator in eq 5 independent of time. Due to the restricted character of this ansatz, an imaginary time extrapolation was required to reach the ground state. Similar approaches, optimizing the time steps for evolving multiple fixed operators, have been widely applied in quantum computing for state preparation, as well as in simulations on classical computers to study the Ising model in transverse field. The coupled cluster singles and doubles (CCSD) method can also be written as a special case of eq 6, in which only one large projection step is performed.

The ansatz proposed in eq 6 can be further extended. For instance, we may introduce neural networks (NN) that act on the auxiliary field term, leading to the following substitution:

$$\sum_{\hat{r}} x_{\hat{r}} \hat{L}_{\hat{r}} \rightarrow \sum_{\hat{r}} f_{\hat{r}}^{NN}(x_{\hat{r}}) \hat{L}_{\hat{r}}$$

where $x_{\hat{r}}$ is the set of auxiliary fields at time step $\hat{r}$, and $f_{\hat{r}}^{NN}$ is a vector-valued optimizable function parameterized by a neural network. The parameter of the neural network can be optimized simultaneously with other parameters in the ansatz. In this way, the Hubbard–Stratonovich transformation has been superseded, and the projection does not correspond any longer to a Hamiltonian propagation like in eq 5. However, the propagated state is still a superposition of Slater determinants.

In the application of this approach to molecules, we found that in most cases such an extension was unnecessary, because the original ansatz was expressive enough. However, we found that in some strongly correlated systems like the $N_2$ molecule in the dissociation limit, discussed in the next section, this approach was useful to overcome the limitation of a single reference initial state.

The optimization of the parameters follows the variational principle, similar to what is done in VMC. In particular, we seek to minimize the energy expectation value:

$$E_{\theta} = \langle \psi_{\theta} | \hat{H} | \psi_{\theta} \rangle = \int dx dx' \langle \psi_{\theta} | \hat{U}^{-1}(x') \hat{H} \hat{U}(x) | \psi_{\theta} \rangle$$

where $\hat{H}$ is the Hamiltonian of the system in eq 1. To perform Monte Carlo sampling, we separate sign and magnitude of the overlap amplitude $\langle \psi_{\theta} | \hat{U}^{-1}(x') \hat{U}(x) | \psi_{\theta} \rangle$ in the integrals in eq 8. Then the expectation value of the energy can be written as

$$E_{\theta} = \frac{\int dx dx' D(x, x') E(x, x') S(x, x')}{\int dx dx' D(x, x') S(x, x')} = \langle \mathcal{E} S \rangle_D$$

where $D(x, x') = I(\langle \psi_{\theta} | \hat{U}^{-1}(x') \hat{U}(x) | \psi_{\theta} \rangle)$ and $S(x, x') = \langle \psi_{\theta} | \hat{U}(x') \hat{U}(x) | \psi_{\theta} \rangle$ are the magnitude and phase of the overlap amplitude, respectively, and

$$\mathcal{E}(x, x') = \langle \psi_{\theta} | \hat{U}^{-1}(x') \hat{H} \hat{U}(x) | \psi_{\theta} \rangle$$

is the so-called local energy in AFQMC. The magnitude $D$ is non-negative, and can be viewed as an unnormalized probability density associated with the auxiliary field configurations. We use $I(\langle \psi_{\theta} | \hat{U}^{-1}(x') \hat{U}(x) | \psi_{\theta} \rangle)$ to denote expectation over that distribution, which can be calculated by sampling from $D$ using Markov chain Monte Carlo (MCMC) techniques.

To further eliminate the sign problem, we add a one-side restraint on the average sign $S_\theta = \langle I(\langle \psi_{\theta} | \hat{U}^{-1}(x') \hat{U}(x) | \psi_{\theta} \rangle) \rangle$ to prevent it from getting too close to zero during the optimization procedure. The final objective function of the optimization can be written as

$$\min_\theta [E_{\theta} + \lambda \langle \max(B - S_\theta, 0) \rangle^2]$$

where $\lambda$ and $B$ are hyperparameters that control the strength and location of the restraint. Unlike the constraint used in the phaseless approach, this restraint should not introduce any systematic bias for ansatze that are flexible enough. Indeed we found that the results are very insensitive to the choice of the restraint, and we take $\lambda = 1$ and $B = 0.7$ throughout the paper. During the optimization process, the average sign is almost always larger than 0.7, indicating that the sign problem is largely avoided.

The optimized wave function $|\psi_{\theta}\rangle$ can be viewed as a superposition of Slater determinants, and can be easily integrated into the standard AFQMC framework. After completing the optimization, we may use the so-called mixed estimator of the ground state energy to further reduce the sampling variance and speed up the computation:

$$E_M = \langle \psi_{\theta} | \hat{H} | \psi_{\theta} \rangle \langle \psi_{\theta} | \hat{H}^{-1} \hat{U}(x) | \psi_{\theta} \rangle = \frac{\int dx dx' \langle \psi_{\theta} | \hat{U}^{-1}(x') \hat{H} \hat{U}(x) | \psi_{\theta} \rangle \langle \psi_{\theta} | \hat{U}(x') \hat{H}^{-1} \hat{U}(x) | \psi_{\theta} \rangle}{\int dx \langle \psi_{\theta} | \hat{U}(x) | \psi_{\theta} \rangle}$$

where $|\psi_{\theta}\rangle$ is some trial wave function. Moreover, in cases in which the optimization does not fully converge to the ground state, we may improve the accuracy by further propagating the wave function with the original Hamiltonian according to the following expression:

$$\mathcal{E}_M(\tau) = \langle \psi_{\theta} | \hat{H} e^{-\tau \hat{H}} | \psi_{\theta} \rangle$$

which is similar to the constraint release approach in standard AFQMC studies of the Hubbard model and to the release-node diffusion Monte Carlo. These approaches are exact in the large $\tau$ limit but suffer from the sign problem. In practice, we found that in all the studied cases, a rather short propagation was sufficient to converge to satisfactory results, before the sign problem appeared.

We note that the variational optimization of our ansatz does not have the benefit of the zero-variance principle of standard VMC, due to the need of sampling auxiliary field configurations along the propagation. This process requires estimating the gradients of the propagators with respect to the optimization parameters, a task affected by statistical noise, the magnitude of which does not diminish as the optimization proceeds. At the early stages of the optimization, this is not a problem since the magnitude of the gradients is large. However, as the optimization approaches the ground state, the scale of the gradients is reduced, with a corresponding
increase of the signal-to-noise ratio and of the relative error of the estimated gradients. Thus, full convergence in the optimization would require averaging over many samples, a procedure that, for large systems, can be quite inefficient. On the other hand, free projection AFQMC does not require gradients, and only suffer from the vanishing signal-to-noise ratio due to the sign problem. When the propagation starts from a state close enough to the ground state, a short propagation is sufficient to reach convergence before the sign problem sets in. Therefore, it is convenient to combine the strength of optimization and propagation, as we do in our calculations, by using the optimization to get a solution sufficiently close to the ground state, and then using free projection to achieve full convergence.

3. RESULTS

To assess the accuracy of the scheme, we study the performance of the HAFQMC method on various molecular systems. We implemented the main algorithm in Python using the JAX57 and Flax58 libraries, with the electronic integrals obtained from PySCF59 and the modified Cholesky decomposition performed by a script from PAUXY.50 We used Hamiltonian Monte Carlo (HMC)51 to sample the auxiliary fields in the optimization procedure, in which we update all the auxiliary fields along an entire path in a single move. HMC is significantly more efficient than straightforward MCMC at sampling auxiliary field configurations, because it uses gradient information and can have a much higher acceptance rate. Direct Gaussian sampling is used in the mixed estimator after optimization. The adabief belief optimizer52 implemented in the Optax library53 was employed, and the mean-field subtraction technique54 was performed at every step during the optimization. In order to compare with the literature and maintain the computational time within reasonable limits, all the calculations were done in the cc-pVDZ basis set.55 Unless otherwise specified, we used canonical molecular orbitals from the restricted Hartree–Fock (RHF) solution as the single-particle basis in eq 1. For the number \( N_f \) of projection steps and \( N_p \) of auxiliary fields appearing in eq 6, we chose the values \( N_f = 3 \) and \( N_p = 100 \) in all our calculations with the exception of benzene, for which we used \( N_p = 200 \). Additional computation details for each tested system can be found in Appendix A.

As a first example, we test the method on a simple but paradigmatic system: the \( \text{H}_4 \) molecule in a rectangular configuration, also known as the Paldus test.56 The system is small enough to be solved exactly by full configuration interaction (FCI), yet it can be so strongly correlated that coupled cluster methods may fail.57 We examine both the stretching of a square \( \text{H}_4 \) and the deformation of a rectangular one.

First, we consider the potential energy curve for the uniformly stretched square \( \text{H}_4 \) molecule (see Figure 2). We notice that, by varying the stretching length \( d \), the ground state of the square \( \text{H}_4 \) molecule undergoes a crossing from triplet to singlet state. The potential energy curves of different spin configurations from FCI calculations are shown in the upper part of Figure 1. We initialized \( | \bar{\psi} \rangle \) with singlet and triplet spin configurations and examined our methods near the crossing point. Since \( \text{H}_4 \) is relatively small and the optimization is fully converged, we applied eq 9 to estimate the ground state energy directly from the final stage of the optimization. This approach guarantees that the estimated energy is variational (i.e., greater than or equal to the true ground state energy). The results can be found in Figure 1. For both spin configurations, HAFQMC is able to provide near-exact results, with errors in the total energies of less than 0.2 mEh for all configurations.

However, it would be more convenient to automatically determine the ground state energy corresponding to the lowest spin state, without the need for specifying the spin configuration in advance. To achieve this goal, we adopt the idea of the generalized Hartree–Fock (GHF) method that allows electrons to contain both spin components. We extend both \( \hat{U}(\mathbf{x}) \) and \( | \bar{\psi} \rangle \) in our ansatz to include off-diagonal terms between spin components, enabling hopping between spin configurations during propagation. We find that with this generalization, HAFQMC can successfully find the correct ground state energy, irrespective of the initial spin configuration. In addition, the generalized ansatz reduces the variance in the estimation of the energy. Therefore, we adopt exclusively this approach hereafter. We note that a similar scheme was applied in neural network based VMC,58 achieving good results.

Figure 2 shows the numerical results for the stretching curve of the square \( \text{H}_4 \) molecule calculated by the generalized HAFQMC approach, along with FCI and restricted/unrestricted CCSD(T) results for comparison. Since the system is restricted to singlet states, the RCCSD(T) results deviate significantly from those of FCI. UCCSD(T) agrees well with FCI near the equilibrium geometry, thanks to the correct assignment of the spin state, but deviates from FCI with errors of several milli-Hartrees when the molecule is partially dissociated. By contrast, HAFQMC exhibits uniformly accurate performance along the dissociation curve. The only relatively large error occurs at 2.5 \( a_0 \), which is the crossing location of the singlet and triplet states. This is likely due to the near
degeneracy of the two spin states, which makes it harder for the optimization procedure to find the lowest energy state. Evidence for that can be found in Figure 1 where the error is smaller because the spin configuration has been preassigned.

We further study the deformation of the rectangular H\textsubscript{4} system. We adopted the test case reported in ref 56, where the distance from the hydrogen atoms to their center of mass is fixed at 3.284 \(a_0\) (around twice the equilibrium distance) while the central angle \(\theta\) of the short side varies (see Figure 3). The system becomes strongly correlated when \(\theta\) approaches 90\textdegree. It is known that restricted coupled cluster methods predict an unphysical downward-facing cusp at \(\theta = 90\textdegree\), due to the crossing of two HF configurations with different symmetries.\textsuperscript{56} UCCSD(T) works better than RCCSD(T) and predicts a smooth curve, but still presents a relatively large error near the square geometry compared to configurations with small \(\theta\) angles. On the other hand, HAFQMC does not suffer from these limitations. As can be seen in Figure 3, HAFQMC calculations agree very well with the FCI results and the errors are less than 0.3 mE\textsubscript{h} for all configurations. Note that here we used meta-Lowdin localized orbitals\textsuperscript{59} for the single particle basis to speed up the optimization, taking advantage of the sparse geometry of the system.

Another well-studied strongly correlated system is bond stretching of the N\textsubscript{2} molecule. Due to the breaking of a triple bond, this system is challenging for traditional methods like coupled cluster. We benchmark HAFQMC for multiple bond lengths \(d\) by comparing its results with the near-exact DMRG energies from ref 60, as shown in Figure 4. We also report in the same figure UCCSD(T) and phaseless (ph-) AFQMC results from ref 61. Since the system is larger than H\textsubscript{4} we did not run the optimization to full convergence, but used the mixed estimator of eq 12 with a propagation time \(\tau = 2\) au to approach the ground state. For trial state \(|\psi_f\rangle\) we used the RHF solution. For better comparison, we also include results from free projection (fp-) AFQMC with \(\tau = 6\) starting from a CCSD initial state\textsuperscript{28} and still using the RHF trial state in eq 12.

By comparing the prediction of our method to DMRG results for the potential energy curve of N\textsubscript{2}, we found that HAFQMC is more accurate than coupled cluster and phaseless
AFQMC, even though the latter made use of better trial wave functions from multiconfiguration self-consistent field (MCSCF) calculations. As it can be seen in Figure 4, HAFQMC predict almost exact energies for geometries near equilibrium. The error is slightly larger in the strongly correlated regime associated with bond lengths in the range of 3.6–4.2 \(a_0\). But, even in this regime, the error is less than 1.6 mE\(_0\). We suggest that the larger error for \(d \sim 3.6–4.2\) \(a_0\) should originate from the RHF trial wave function, which is inappropriate at these separation distances. This conjecture is supported by the finding that the free projection AFQMC with RHF trial wave function converges very slowly in this regime, with errors of nearly 10 mE\(_0\) in spite of the much longer propagation time adopted in these calculations. Figure 7 in Appendix B further shows the prediction error of fp-AFQMC for different propagation times \(\tau\) at multiple bond lengths \(d\), providing more detailed evidence for the slow convergence in the dissociating regime, consistent with strong correlation character.

It is worth noting that the HAFQMC method can be further improved in the dissociating regime by introducing a neural network in the ansatz, as described in eq 7. We employed a three-layer fully connected NN having a width equal to the number of auxiliary fields and using the GELU activation function and skip connections. In this way, the largest error can be reduced to around 0.5 mE\(_0\) without changing the projection time, as shown in Figure 4. In addition, it is straightforward to combine HAFQMC with more sophisticated trial wave functions, such as those from MCSCF or HCI calculations, which should perform much better than RHF. We leave this to future work.

We then study the behavior of HAFQMC on larger molecular systems, starting from the well-benchmarked benzene molecule with the cc-pVDZ basis. We use the same geometry and frozen carbon 1s orbitals as in the published benchmark study. Similar to N\(_{2}\), we performed the variational optimization with a fixed number of steps and then further propagated the state to approach the ground state using the mixed estimator for the energy in eq 12. The propagation converges very quickly (\(\tau \approx 2\) au) to the ground state energy of about 231.5856 E\(_0\), with an estimated statistical uncertainty less than 1 mE\(_0\), as shown in Figure 5, which also reports the average sign along the propagation. After an initial oscillation, the energy reaches a steady value (within error bars), suggesting that convergence has been achieved. Note that we did completely independent propagations restarting from the initial state for different projection times, instead of using the same trajectory. Therefore, issues like insufficient mixing in the Monte Carlo process do not exist. The calculated result from HAFQMC is consistent with previous free projection AFQMC calculations using heat-bath configuration interaction (HCI) trial wave functions, and agrees also within chemical accuracy with several highly accurate methods such as DMRG and semistochastic HCI (SHCI) in the benchmark reported in ref 64. The slight differences between the different methods in Figure 5 and ref 64 may come from different causes: the Trotter error or the statistical error in the AFQMC propagation, or the extrapolation error with methods like DMRG.

We note that, for the equilibrium configuration of the benzene molecule, coupled cluster methods work well, and so does standard fp-AFQMC prior to the onset of the sign problem. This is to be expected since the system is mainly affected by dynamic correlation. We can see in Figure 5 that with a sufficiently long propagation, fp-AFQMC with RHF trial wave function is capable of reaching near exact ground state energy in a reasonable time, but the error bar is inevitably larger when the propagation time is longer. We can also see in Figure 5 that HAFQMC introduces only a small decrease in the average sign value, compared to standard fp-AFQMC. This is not a problem since the average sign is far from zero at convergence. The different average signs associated with different methods in Figure 5 are mainly due to the different initial states adopted for the propagation, as illustrated by the fact that, after a short initial relaxation period, the average signs decay linearly with the same rate for the three different methods reported in the figure.

Our last example is the cyclobutadiene (C\(_4\)H\(_5\)) molecule, for which we study the equilibrium configuration and the transition state of the automerization reaction. At the transition state, the system is strongly correlated, posing a challenge to common electronic structure methods. In our calculations, we used the geometries of ref 66 and conducted calculations as in benzene’s case, freezing the carbon 1s orbitals and using the mixed estimator for the energy. The HAFQMC estimated barrier height for the reaction is 12.6 mE\(_0\) (7.9 kcal/mol) with a statistical uncertainty of about 0.4 mE\(_0\). This calculated barrier height is compatible with experimental measurements and aligns well with other high-accuracy computational results reported in ref 28.
Figure 6 shows the evolution of the estimated ground state energies at equilibrium and transition-state geometries from HAFQMC calculations during imaginary time propagation, and compares them with free projection AFQMC results using different trial and initial states. At the equilibrium geometry, the correlation is mainly dynamic, similar to benzene, and the standard fp-AFQMC converges with a long projection time and large statistical errors. However, at the transition state geometry, free projection with RHF trial state performs badly due to the multireference nature of the electronic state. On the other hand, HAFQMC converges very well with nearly the same rates for both equilibrium and transition state configurations. We also find that HAFQMC converges to slightly lower energies than those found with the free projection AFQMC calculations reported in ref 28. These findings provide further evidence of the ability of HAFQMC to recover highly accurate ground state energies of strongly correlated molecular systems without a priori information on the multiconfiguration state.

4. CONCLUSION

In this paper, we presented a quantum Monte Carlo approach that we called HAFQMC, which combines VMC and AFQMC, to solve the many-body Schrödinger equation in electronic structure problems. The method helps to circumvent the sign problem by variationally finding an imaginary time propagator that approaches the ground state in a short projection time. The variationally optimized state can be further refined by applying to it a short propagation within standard AFQMC. When applied to the variationally optimized states, AFQMC converged rapidly, before manifestations of the sign problem, in all the cases considered. The variational ansatz can be made more flexible in strongly correlated situations by including in it a neural network structure. The method scales gracefully ($O(N^3)$) with system size. Thanks to the flexible ansatz, HAFQMC does not need extrapolations, either in imaginary time like in ref 36, or in DMRG bond dimensions like in refs 60, 64. An additional advantage of the approach is that it requires little prior knowledge of the system under consideration. It is not necessary to assign a priori the spin configuration, the active space, or a high-quality initial wave function. Highly accurate ground state energies can be achieved already with RHF wave functions. Benchmark tests showed uniform accuracy for various systems, ranging from weakly to strongly correlated, without ever incurring in serious sign problems.

The method can be smoothly integrated into existing AFQMC frameworks. Several techniques developed in the field can be used to further improve the method, such as using high-quality initial and trial wave functions to speedup convergence,38 or utilizing tensor decompositions to reduce the computational cost of calculations on large systems.38 In the current implementation, the matrix elements of the operators in the chosen basis are treated as optimizable parameters. Thus, sparser representations of the operators that encode prior knowledge should improve the efficiency of the optimization procedure and its transferability to different molecular geometries. Other potential improvements could be achieved by using second-order optimizers.69 We would like to explore these possibilities in future work.

Additional computational details relative to the main text are provided here.

APPENDIX A: COMPUTATIONAL DETAILS

The cutoff used in the modified Cholesky decomposition is $10^{-6}$. For a given basis, the matrix elements of the Hamiltonian and the components of the initial RHF wave functions are treated as optimizable parameters. Initially, each optimizable parameter takes the value dictated by the exact Hamiltonian of the problem and its RHF solution, multiplied by a symmetry breaking Gaussian random variable of mean 1 and standard deviation 0.1.

The optimization is conducted with the adabelief optimizer, using the default settings and a learning rate decay of $(1 + n/n_0)^{-1}$, where $n$ is the number of optimization steps. The gradients of $E_\theta$ in eq 9 with respect to the optimizable parameters are calculated with the backward automatic differentiation technique and clipped to make the absolute value of every element be less than 1. The estimation of the energy and the gradients during the optimization is done with a Hamiltonian Monte Carlo (HMC) sampler with symplectic integration step size equal to 0.1 and length equal to 1 for every trial move. The sampler is burned in for 100 steps in the Markov chain before the optimization begins.

As the size and number of parameters is different in each system, the corresponding number of optimization samples varies, being 10000 for $H_4$ and $N_6$, and equal to 1000 for benzene and cyclobutadiene. The number of chains in HMC is 1000 for $H_4$ and $N_6$, 500 for cyclobutadiene, and 100 for benzene due to memory limits. The starting learning rate differs accordingly, being $3 \times 10^{-3}$ for $H_4$ and $N_6$, and $1 \times 10^{-4}$ for the other two molecules. The total number of optimization steps was 40 000 for $H_4$ and $N_6$, 80 000 for cyclobutadiene, and 60 000 for benzene.
Free projection AFQMC following optimization was carried out using $\Delta \tau = 0.1$ au for $N_2$ and $\Delta \tau = 0.05$ au for benzene and cyclobutadiene, as in ref 28. The energy was estimated using $10^6$ samples from direct Gaussian sampling in all cases, except in the neural network construction adopted for the $N_2$ system, where $2 \times 10^7$ HMC samples were used.

**APPENDIX B: SUPPLEMENTARY RESULTS**

In this section we report supplementary results for the $N_2$ molecule, showing the difficulties of fp-AFQMC with this molecule. These difficulties are evident in Figure 7 that plots the errors relative to DMRG$^{60}$ of the ground state energy obtained with fp-AFQMC, using initial CCSD wave function and RHF trial in the mixed estimator, at different bond distances $d$. The convergence is particularly bad in the dissociating regime, when $d \sim 3.6-4.2\alpha_0$, due to a rapid emergence of the sign problem.

**AUTHOR INFORMATION**

**Corresponding Author**

Roberto Car — Department of Chemistry, Department of Physics, Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, New Jersey 08544, United States; Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08544, United States; orcid.org/0000-0001-5243-2647; Email: rcar@princeton.edu

**Authors**

Yixiao Chen — Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08544, United States; orcid.org/0000-0001-8201-5887

Linfeng Zhang — AI for Science Institute, Beijing 100080, People's Republic of China; DP Technology, Beijing 100080, People's Republic of China; orcid.org/0000-0002-8470-8546

Weinan E — Center for Machine Learning Research, School of Mathematical Sciences, Peking University, Beijing 100084, People’s Republic of China; AI for Science Institute, Beijing 100080, People’s Republic of China

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c00038

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We thank Han Wang, Lin Lin, Lei Wang, and Yuzhi Liu for fruitful discussions. We also thank Shiwei Zhang, Hao Shi, and Yuanyao He for inspiring discussions at the early stage of this work. Y.C. and R.C. were supported by the Computational Chemical Sciences Center: Chemistry in Solution and at Interfaces (CSI) funded by DOE Award DE-SC0019394. Y.C. and W.E. were also supported by a gift from iFlytek to Princeton University. All calculations were performed using the Princeton Research Computing resources at Princeton University.

**REFERENCES**

(1) E, W.; Han, J.; Zhang, L. Machine-learning-assisted modeling. Phys. Today 2021, 74, 36–41.

(2) Knowles, P. J.; Handy, N. C. A new determinant-based full configuration interaction method. Chemical physics letters 1984, 111, 315–321.

(3) Löwdin, P.-O. Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction. Phys. Rev. 1955, 97, 1474.

(4) Čiček, J. On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods. J. Chem. Phys. 1966, 45, 4256–4266.

(5) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. Physical review 1934, 46, 618.

(6) Roos, B. O.; Taylor, P. R.; Sigbahn, P. E. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. Chem. Phys. 1980, 48, 157–173.

(7) Werner, H.-J.; Knowles, P. J. An efficient internally contracted multiconfiguration-reference configuration interaction method. J. Chem. Phys. 1988, 89, 5803–5814.

(8) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-order perturbation theory with a CASSCF reference function. J. Phys. Chem. 1990, 94, 5483–5488.

(9) White, S. R. Density matrix formulation for quantum renormalization groups. Physical review letters 1992, 69, 2863.

(10) White, S. R.; Martin, R. L. Ab initio quantum chemistry using the density matrix renormalization group. J. Chem. Phys. 1999, 110, 4127–4130.

(11) Holmes, A. A.; Tubman, N. M.; Umrigar, C. Heat-bath configuration interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling. J. Chem. Theory Comput. 2016, 12, 3674–3680.

(12) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory. J. Chem. Theory Comput. 2017, 13, 1595–1604.

(13) Ceperley, D.; Chester, G. V.; Kalos, M. H. Monte Carlo simulation of a many-fermion study. Phys. Rev. B 1977, 16, 3081.

(14) Ceperley, D. M.; Alder, B. J. Ground state of the electron gas by a stochastic method. Physical review letters 1980, 45, 566.

(15) Becca, F.; Sorella, S. Quantum Monte Carlo approaches for correlated systems; Cambridge University Press, 2017.
(16) Han, J.; Zhang, L.; E. W. Solving many-electron Schrödinger equation using deep neural networks. J. Comput. Phys. 2019, 399, 108929.

(17) Hermann, J.; Schätzle, Z.; Noé, F. Deep-neural-network solution of the electronic Schrödinger equation. Nat. Chem. 2020, 12 (10), 891–897.

(18) Pfau, D.; Spencer, J. S.; Matthews, A. G.; Foulkes, W. M. C. Ab initio solution of the many-electron Schrödinger equation with deep neural networks. Physical Review Research 2020, 2, 033429.

(19) Kalos, M. H.; Levesque, D.; Verlet, L. Helium at zero temperature with hard-sphere and other forces. Phys. Rev. A 1974, 9, 2178.

(20) Sugiyama, G.; Koonin, S. Auxiliary field Monte-Carlo for quantum many-body ground states. Annals of Physics 1986, 168, 1–26.

(21) Sorella, S.; Baroni, S.; Car, R.; Parrinello, M. A novel technique for the simulation of interacting fermion systems. EPL (Europhysics Letters) 1989, 8, 663.

(22) Booth, G. H.; Thom, A. J.; Alavi, A. Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space. J. Chem. Phys. 2009, 131, 054106.

(23) Troyer, M.; Wiese, U.-J. Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations. Physical review letters 2005, 94, 170201.

(24) Anderson, J. B. A random-walk simulation of the Schrödinger equation: H+ 3. J. Chem. Phys. 1975, 63, 1499–1503.

(25) Zhang, S.; Carlson, J.; Gubernatis, J. E. Constrained path Monte Carlo method for fermion ground states. Phys. Rev. B 1997, 55, 7464.

(26) Zhang, S.; Krakauer, H. Quantum Monte Carlo method using phase-free random walks with Slater determinants. Physical review letters 2003, 90, 136401.

(27) Motta, M.; Zhang, S. Ab initio computations of molecular systems by the auxiliary-field quantum Monte Carlo method. Wiley Interdisciplinary Reviews: Computational Molecular Science 2018, 8, No. e1364.

(28) Mahajan, A.; Sharma, S. Taming the Sign Problem in Auxiliary-Field Quantum Monte Carlo Using Accurate Wave Functions. J. Chem. Theory Comput. 2021, 17, 4786–4798.

(29) Yanagisawa, T.; Koike, S.; Yamaji, K. Off-diagonal wave function Monte Carlo Studies of Hubbard model I. J. Phys. Soc. Jpn. 1998, 67, 3867–3874.

(30) Eichenberger, D.; Baeriswyl, D. Superconductivity and antiferromagnetism in the two-dimensional Hubbard model: A variational study. Phys. Rev. B 2007, 76, 180504.

(31) Yanagisawa, T. Crossover from Weakly to Strongly Correlated Regions in the Two-dimensional Hubbard Model—Off-diagonal Wave Function Monte Carlo Studies of Hubbard Model II—. J. Phys. Soc. Jpn. 2016, 85, 114707.

(32) Beach, M. J.; Melko, R. G.; Grover, T.; Hsieh, T. H. Making trottlers sprint: A variational imaginary time ansatz for quantum many-body systems. Phys. Rev. B 2019, 100, 094434.

(33) Farhi, E.; Goldstone, J.; Gutmann, S. A quantum approximate optimization algorithm. arXiv preprint arXiv:1411.4028 2014. DOI: 10.48550/arXiv.1411.4028.

(34) Wecker, D.; Hastings, M. B.; Troyer, M. Progress towards practical quantum variational algorithms. Phys. Rev. A 2015, 92, 042303.

(35) Hadfield, S.; Wang, Z.; O’Gorman, B.; Rieffel, E. G.; Venturelli, D.; Biswas, R. From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. Algorithms 2019, 12, 34.

(36) Sorella, S. The phase diagram of the Hubbard model by Variational Auxiliary Field quantum Monte Carlo. arXiv preprint arXiv:2101.07045 2021. DOI: 10.48550/arXiv.2101.07045.

(37) Trotter, H. F. On the product of semi-groups of operators. Proceedings of the American Mathematical Society 1959, 10, 545–551.

(38) Suzuki, M. Relationship between d-dimensional quantal spin systems and (d+1)-dimensional using systems: Equivalence, critical exponents and systematic approximations of the partition function and spin correlations. Progress of theoretical physics 1976, 56, 1454–1469.

(39) Thouless, D. J. Stability conditions and nuclear rotations in the Hartree-Fock theory. Nucl. Phys. 1960, 21, 225–232.

(40) Hubbard, J. Calculation of partition functions. Phys. Rev. Lett. 1959, 3, 77.

(41) Beebe, N. H.; Linderberg, J. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. Int. J. Quantum Chem. 1977, 12, 683–705.

(42) Koch, H.; Sánchez de Merás, A.; Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. J. Chem. Phys. 2003, 118, 9481–9484.

(43) Andrei, C.; De Freitas, N.; Doucet, A.; Jordan, M. I. An introduction to MCMC for machine learning. Machine learning 2003, 50, 5–43.

(44) Sorella, S. Linearized auxiliary fields Monte Carlo technique: Efficient sampling of the fermion sign. Phys. Rev. B 2011, 84, 241110.

(45) Shi, H.; Zhang, S. Symmetry in auxiliary-field quantum Monte Carlo calculations. Phys. Rev. B 2013, 88, 125132.

(46) Ceperley, D.; Alder, B. Quantum Monte Carlo for molecules: Green’s function and nodal release. J. Chem. Phys. 1981, 81, 5833–5844.

(47) Frostig, R.; Johnson, M. J.; Leary, C. Compiling machine learning programs via high-level tracing. Syst. Mach. Learning 2018, 23–24.

(48) Heek, J.; Levskaya, A.; Oliver, A.; Ritter, M.; Rondepiere, B.; Steiner, A.; van Zee, M. Flax: A neural network library and ecosystem for JAX. 2020; https://github.com/google/flax.

(49) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; et al. PySCF: The Python-based simulations of chemistry framework. Wiley Interdisciplinary Reviews: Computational Molecular Science 2018, 8 (1), No. e1340.

(50) Malone, F.; Lee, J.; Frost, J. M. PAUXY: A collection of Python implementations of auxiliary field quantum Monte Carlo algorithms. 2017; https://github.com/pauxy-qmc/pauxy.

(51) Betancourt, M. A conceptual introduction to Hamiltonian Monte Carlo. arXiv preprint arXiv:1701.02434 2017. DOI: 10.48550/arXiv.1701.02434.

(52) Zhuang, J.; Tang, T.; Ding, Y.; Tatikonda, S. C.; Dvornek, N.; Papademetris, X.; Duncan, J. Adalbelief optimizer: Adapting stepsizes to the belief in observed gradients. Advances in neural information processing systems 2020, 33, 18795–18809.

(53) Hessel, M.; Budden, D.; Viola, F.; Rosca, M.; Sezener, E.; Hennigan, T. Optax: composable gradient transformation and optimization, in JAX. 2020; http://github.com/deepmind/optax.

(54) Dunning, T. H., Jr Gaussian basis functions for use in molecular calculations. I. Contraction of (9s5p) atomic basis sets for the first-row atoms. J. Chem. Phys. 1970, 53, 2823–2833.

(55) Jankowski, K.; Paldus, J. Applicability of coupled-pair theories to quasidegenerate electronic states: A model study. Int. J. Quantum Chem. 1980, 18, 1243–1269.

(56) Van Voorhis, T.; Head-Gordon, M. Benchmark variational coupled cluster doubles results. J. Chem. Phys. 2000, 113, 8873–8879.

(57) Jiménez-Hoyos, C. A.; Henderson, T. M.; Scuseria, G. E. Generalized hamtter–fock description of molecular dissociation. J. Chem. Theory Comput. 2011, 7, 2667–2674.

(58) Lin, J.; Goldshläger, G.; Lin, L. Explicitly antisymmetrized neural network layers for variational Monte Carlo simulation. arXiv preprint arXiv:2112.03491 2021. DOI: 10.48550/arXiv.2112.03491.

(59) Löwdin, P.-O. On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. J. Chem. Phys. 1950, 18, 365–375.

(60) Chan, G. K.-L.; Källay, M.; Gauss, J. State-of-the-art density matrix renormalization group and coupled cluster theory studies of the nitrogen binding curve. J. Chem. Phys. 2004, 121, 6110–6116.

(61) Al-Saidi, W. A.; Zhang, S.; Krakauer, H. Bond breaking with auxiliary-field quantum Monte Carlo. J. Chem. Phys. 2007, 127, 144101.
(62) Hendrycks, D.; Gimpel, K. Gaussian error linear units (GELUs). arXiv preprint arXiv:1606.08415 2016. DOI: 10.48550/arXiv.1606.08415.

(63) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; IEEE: New York, 2016; pp 770–778.

(64) Eriksen, J. J.; Anderson, T. A.; Deustua, J. E.; Ghanem, K.; Hait, D.; Hoffmann, M. R.; Lee, S.; Levine, D. S.; Magoulas, I.; Shen, J.; et al. The ground state electronic energy of benzene. Journal of physical chemistry letters 2020, 11, 8922–8929.

(65) Lee, J.; Malone, F. D.; Reichman, D. R. The performance of phaseless auxiliary-field quantum Monte Carlo on the ground state electronic energy of benzene. J. Chem. Phys. 2020, 153, 126101.

(66) Lyakh, D. I.; Lotrich, V. F.; Bartlett, R. J. The ‘tailored’ CCSD (T) description of the automerization of cyclobutadiene. Chem. Phys. Lett. 2011, 501, 166–171.

(67) Whitman, D. W.; Carpenter, B. K. Limits on the activation parameters for automerization of cyclobutadiene-1, 2-d2. J. Am. Chem. Soc. 1982, 104, 6473–6474.

(68) Motta, M.; Shee, J.; Zhang, S.; Chan, G. K.-L. Efficient ab initio auxiliary-field quantum Monte Carlo calculations in Gaussian bases via low-rank tensor decomposition. J. Chem. Theory Comput. 2019, 15, 3510–3521.

(69) Martens, J.; Grosse, R. Optimizing neural networks with kronecker-factored approximate curvature. International conference on machine learning; JMLR, 2015; pp 2408–2417.