THE JOURNAL OF CHEMICAL PHYSICS **126**, 194105 (2007)

A study of $H+H_2$ and several H-bonded molecules by phaseless auxiliary-field quantum Monte Carlo with plane wave and Gaussian basis sets

W. A. Al-Saidi,[a] Henry Krakauer, and Shiwei Zhang
Department of Physics, College of William and Mary, Williamsburg, Virginia 23187-8795

(Received 21 February 2007; accepted 4 April 2007; published online 21 May 2007)

The authors present phaseless auxiliary-field (AF) quantum Monte Carlo (QMC) calculations of the ground states of some hydrogen-bonded systems. These systems were selected to test and benchmark different aspects of the new phaseless AF QMC method. They include the transition state of $H+H_2$ near the equilibrium geometry and in the van der Walls limit, as well as the H_2O, OH, and H_2O_2 molecules. Most of these systems present significant challenges for traditional independent-particle electronic structure approaches, and many also have exact results available. The phaseless AF QMC method is used either with a plane wave basis with pseudopotentials or with all-electron Gaussian basis sets. For some systems, calculations are done with both to compare and characterize the performance of AF QMC under different basis sets and different Hubbard-Stratonovich decompositions. Excellent results are obtained using as input single Slater determinant wave functions taken from independent-particle calculations. Comparisons of the Gaussian based AF QMC results with exact full configuration interaction show that the errors from controlling the phase problem with the phaseless approximation are small. At the large basis-size limit, the AF QMC results using both types of basis sets are in good agreement with each other and with experimental values. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2735296]

I. INTRODUCTION

Quantum Monte Carlo (QMC) methods[1,2] offer a unique way to treat explicitly the many-electron problem. The many-body solution is obtained in a statistical sense by building stochastic ensembles that sample the wave function in some representation. This leads to computational costs that scale as a low power with the number of particles and with the basis size. Although in practice QMC methods are often not exact, they have shown considerably greater accuracy than traditional electronic structure approaches in a variety of systems. They are increasingly applied and are establishing themselves as a unique approach for studying both realistic materials and important model systems.

Recently, a new phaseless auxiliary-field QMC (AF QMC) method has been developed and applied for electronic structure calculations.[2,3] This method is formulated in a many-particle Hilbert space whose span is defined by a single-particle basis set. The freedom to choose the basis set can potentially result in increased efficiency. This can be very useful both for quantum chemistry applications and in calculations with model Hamiltonians. Further, it is straightforward in this method to exploit well-established techniques of independent-particle theories for the chosen basis set. The ability to use any single-particle basis is thus an attractive feature of the AF QMC method. On the other hand, the use of finite basis sets often requires monitoring the convergence of calculated properties and extrapolation of the results to the infinite basis-size limit.

Plane wave and Gaussian basis sets are the most widely used basis sets in electronic structure calculations. Plane waves are appealing because they form a complete orthonormal basis set, and convergence with respect to basis size is easily controlled. A single energy cutoff parameter E_{cut} controls the basis size by including all plane waves with wave vector $\mathbf{k}$ such that $\mathbf{k}^2/2<E_{\mathrm{cut}}$ (hartree a.u. are used throughout the paper). The infinite basis limit is approached by simply increasing E_{cut}.[4] Localized basis sets, by contrast, offer a compact and efficient representation of the system's wave functions. Moreover, the resulting sparsity of the Hamiltonians can be very useful in $\mathcal{O}(\mathrm{N})$ methods. Achieving basis set convergence, however, requires more care. For Gaussian basis sets, quantum chemists have compiled lists of basis sets of increasing quality for most of the elements.[5] Some of these basis sets have been designed for basis extrapolation not only in mean-field theories, but also in correlated calculations.[6,7]

The AF QMC method also provides a different route to controlling the Fermion sign problem.[2,8–10] The standard diffusion Monte Carlo (DMC) method[1,11,12] employs the fixed-node approximation[11] in real coordinate space. The AF QMC method uses random walks in a manifold of Slater determinants (in which antisymmetry is automatically imposed on each random walker). The Fermion sign/phase problem is controlled approximately according to the overlap of each random walker (Slater determinant) with a trial wave function. Applications of the phaseless AF QMC method to date, including second-row systems[2] and transition metal molecules[13] with plane wave basis sets, and first-row[3] and post-*d* (Ref. 14) molecular systems with Gaussian basis sets, indicate that this often reduces the reliance of the results on the quality of the trial wave function. For example, with

---

[a] Present address: Cornell Theory Center, Rhodes Hall, Cornell University, Ithaca, NY 14850.

single determinant trial wave functions, the calculated total energies at equilibrium geometries in molecules show typical systematic errors of no more than a few millihartrees compared to exact/experimental results. This is roughly comparable to that of coupled cluster with single and double excitations plus an approximate treatment of triple excitations [CCSD(T)]. For stretched bonds in H_2O (Ref. 3) as well as in N_2 and F_2,[15] the AF QMC method exhibits better overall accuracy and a more uniform behavior than CCSD(T) in mapping the potential energy curve.

The key features of the AF QMC method are thus its freedom of basis choice and control of the fermion sign/phase problem via a constraint in Slater determinant space. The motivation for this study is therefore twofold. First, we would like to further benchmark the plane wave AF QMC method in challenging conditions, with large basis sets and correspondingly many auxiliary fields. Here, we examine the transition state of the H_2+H system as well as several hydrogen-bonded molecules. These are relatively simple systems, which have been difficult for standard independent-electron methods and for which various results are available for comparison. Second, we are interested in comparing the performance of the AF QMC method using two very different basis sets, namely, plane wave basis sets, together with pseudopotentials, and all-electron Gaussian basis sets. For this, calculations are carried out with Gaussian basis sets for H_2 and the H_2+H transition state, and comparisons are made with the plane wave calculations. Additional Gaussian benchmark calculations are carried out in collinear H_2+H near the van der Waals minimum, which requires resolution of the energy on extremely small scales.

The rest of the paper is organized as follows. In the next section we outline the relevant formalism of the AF QMC method. Section III presents the plane wave and pseudopotential results and includes a study of the dissociation energy and the potential energy curve of H_2, the transition state of H_3, and the dissociation energies of several hydrogen-bonded molecules. In Sec. IV, we use a Gaussian basis to study the potential energy curves of H_2 and $H+H_2$, and compare some of these results with the AF QMC plane wave results. Finally, we conclude in Sec. V with a brief summary.

II. AF QMC METHOD

The auxiliary-field quantum Monte Carlo method has been described elsewhere.[2,3] Here, we outline the relevant formulas to facilitate the ensuing discussion. The method shares with other QMC methods its use of the imaginary-time propagator $e^{-\beta\hat{H}}$ to obtain the ground state $|\Psi_G\rangle$ of $\hat{H}$,

$$|\Psi_G\rangle \propto \lim_{\beta\to\infty} e^{-\beta\hat{H}}|\Psi_T\rangle. \tag{1}$$

The ground state is obtained by filtering out the excited state contributions in the trial wave function $|\Psi_T\rangle$, provided that $|\Psi_T\rangle$ has a nonzero overlap with $|\Psi_G\rangle$.

The many-body electronic Hamiltonian $\hat{H}$ can be written in any one-particle basis as

$$\hat{H} = \hat{H}_1 + \hat{H}_2,$$

$$\hat{H}_1 = \sum_{i,j,\sigma} T_{ij} c^\dagger_{i,\sigma} c_{j,\sigma}, \tag{2}$$

$$\hat{H}_2 = \frac{1}{2} \sum_{i,j,k,l,\sigma,\sigma'} V_{ijkl} c^\dagger_{i,\sigma} c^\dagger_{j,\sigma'} c_{k,\sigma'} c_{l,\sigma},$$

where $c^\dagger_{i,\sigma}$ and $c_{i,\sigma}$ are the corresponding creation and annihilation operators of an electron with spin σ in the ith orbital (the size of the single-particle basis is M). The one-electron and two-electron matrix elements (T_{ij} and V_{ijkl}) depend on the chosen basis and are assumed to be spin independent.

Equation (1) is realized iteratively with a small time step τ such that $\beta = N\tau$, and the $\beta\to\infty$ limit is realized by letting $N\to\infty$. In this case, the Trotter decomposition of the propagator $e^{-\tau\hat{H}}$ $[e^{-\tau\hat{H}} \doteq e^{-\tau\hat{H}_1/2} e^{-\tau\hat{H}_2} e^{-\tau\hat{H}_1/2} + \mathcal{O}(\tau^3)]$ leads to Trotter time-step errors, which can be removed by extrapolation, using separate calculations with different values of τ.

The central idea in the AF QMC method is the use of the Hubbard-Stratonovich (HS) transformation,[16]

$$e^{-\tau\hat{H}_2} = \prod_\alpha \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\sigma_\alpha e^{-(1/2)\sigma_\alpha^2} e^{\sqrt{\tau}\sigma_\alpha \sqrt{\zeta_\alpha}\hat{v}_\alpha} \right), \tag{3}$$

to map the many-body problem exemplified in $\hat{H}_2$ onto a linear combination of single-particle problems using only *one-body operators* $\hat{v}_\alpha$. The full many-body effect is recovered exactly through the interaction between the one-body operators $\{\hat{v}_\alpha\}$ and the external auxiliary fields $\{\sigma_\alpha\}$. This map relies on writing the two-body operator in a quadratic form, such as

$$\hat{H}_2 = -\frac{1}{2} \sum_\alpha \zeta_\alpha \hat{v}_\alpha^2, \tag{4}$$

with ζ_α a real number. This can always be done, as we illustrate below using first a plane wave basis, and then a general basis set.

In a plane wave basis set, the electron-electron interaction operator $\hat{H}_2$ can be written as

$$\begin{aligned}\hat{H}_2 &= \frac{1}{2\Omega} \sum_{\mathbf{k},\mathbf{k}',\sigma,\sigma'} \sum_{\mathbf{q}\neq 0} \frac{4\pi e^2}{\mathbf{q}^2} c^\dagger_{\mathbf{k}+\mathbf{q},\sigma} c^\dagger_{\mathbf{k}'-\mathbf{q},\sigma'} c_{\mathbf{k}',\sigma'} c_{\mathbf{k},\sigma} \\ &= \frac{1}{2\Omega} \sum_{\mathbf{q}>0} \frac{4\pi e^2}{\mathbf{q}^2} [\hat{\rho}(\mathbf{q})\hat{\rho}(-\mathbf{q}) + \text{H.c.}] + H_1'.\end{aligned} \tag{5}$$

Here, $c^\dagger_{\mathbf{k},\sigma}$ and $c_{\mathbf{k},\sigma}$ are the creation and annihilation operators of an electron with momentum $\mathbf{k}$ and spin σ. Ω is the supercell volume, $\mathbf{k}$ and $\mathbf{k}'$ are plane waves within the cutoff radius, and the $\mathbf{q}$ vectors satisfy $|\mathbf{k}+\mathbf{q}|^2/2 < E_{\text{cut}}$. $\hat{\rho}(\mathbf{q}) = \Sigma_{\mathbf{k},\sigma} c^\dagger_{\mathbf{k}+\mathbf{q},\sigma} c_{\mathbf{k},\sigma}$ is a Fourier component of the electron density operator, and H_1' is a one-body term which arises from the reordering of the creation and annihilation operators. For each wave vector $\mathbf{q}$, the two-body term in the final expression in Eq. (5) can be expressed in terms of squares of the one-body operators proportional to $\hat{\rho}(\mathbf{q}) + \hat{\rho}(-\mathbf{q})$ and $\hat{\rho}(\mathbf{q}) - \hat{\rho}(-\mathbf{q})$, which become the one-body operators $\hat{v}_\alpha$ in Eq. (4).

An explicit HS transformation can be given for any general basis as follows (more efficient transformations may exist, however). The two-body interaction matrix V_{ijkl} is first expressed as a Hermitian supermatrix $\mathcal{V}_{\mu[i,l],\nu[k,j]}$ where $\mu,\nu=1,\ldots,M^2$. This is then expressed in terms of its eigenvalues $(-\lambda_\alpha)$ and eigenvectors X^α_μ: $\mathcal{V}_{\mu,\nu}=-\Sigma_\alpha\lambda_\alpha X^{*\alpha}_\mu X^\alpha_\nu$. The two-body operator $\hat{H}_2$ of Eq. (2) can be written as the sum of a one-body operator and a two-body operator $\hat{H}_2'$ such that the latter can be further expressed in terms of the eigenvectors of $\mathcal{V}_{\mu,\nu}$ as

$$\hat{H}_2'=-\frac{1}{4}\sum_\alpha \lambda_\alpha(\hat{\Lambda}_\alpha^\dagger\hat{\Lambda}_\alpha+\hat{\Lambda}_\alpha\hat{\Lambda}_\alpha^\dagger), \tag{6}$$

where the one-body operators $\hat{\Lambda}_\alpha$ are defined as

$$\hat{\Lambda}_\alpha=\sum_{i,l,\sigma} X^\alpha_{\mu[i,l]}c^\dagger_{i,\sigma}c_{l,\sigma}. \tag{7}$$

Similar to the plane wave basis, for each nonzero eigenvalue λ_α, there are two one-body operators $\hat{v}_\alpha$ proportional to $\hat{\Lambda}_\alpha+\hat{\Lambda}_\alpha^\dagger$ and $\hat{\Lambda}_\alpha-\hat{\Lambda}_\alpha^\dagger$. If the chosen basis set is real, then the HS transformation can be further simplified, and the number of auxiliary fields will be equal to only the number of nonzero eigenvalues λ_α.[3]

The phaseless AF QMC method[2] used in this paper controls the phase/sign problem[2,10] in an approximate manner. The method recasts the imaginary-time path integral as branching random walks in Slater determinant space.[10] It uses a trial wave function $|\Psi_T\rangle$ to construct a *complex* importance-sampling transformation and to constrain the paths of the random walks. The ground-state energy, computed with the so-called mixed estimator, is approximate and not variational in the phaseless method. The error depends on $|\Psi_T\rangle$, vanishing when $|\Psi_T\rangle$ is exact. This is the only uncontrolled error in the method, in that it cannot be easily and systematically eliminated. In applications to date, $|\Psi_T\rangle$ has been taken as a single Slater determinant directly from mean-field calculations, and the systematic error is shown to be quite small.[2,3,13,14]

Before presenting our results, we give here some general computational details about the QMC runs. As mentioned before, all of our energies were extrapolated to the zero Trotter time step using independent runs. We typically used four to five Trotter time steps τ between 0.01 and 0.05 hartree^{-1}. The number of walkers used is of the order of a hundred in the plane wave calculations and of a few hundreds with the Gaussian basis sets. For each time step τ, the AF QMC energy is obtained after an equilibration phase followed by a measurement phase of hundreds of blocks, each of several hartree^{-1} in projection length. Our results are obtained using single and parallel runs. To give an idea of computer time requirements, for example, each single point in the H_2 potential energy curve of Fig. 1 at a single Trotter time step τ was obtained using ≈ 100 hours/processor on 32 processors on the Xeon cluster at NCSA. This was needed in order to obtain a relatively smooth potential surface because of the high cutoff in the hydrogen pseudopotential used. For comparison, with the localized basis set, a single Trotter time-step study of H_3/aug-cc-pVTZ (shown in Table V) took an average of around 400 hours on a single ITANIUM processor.

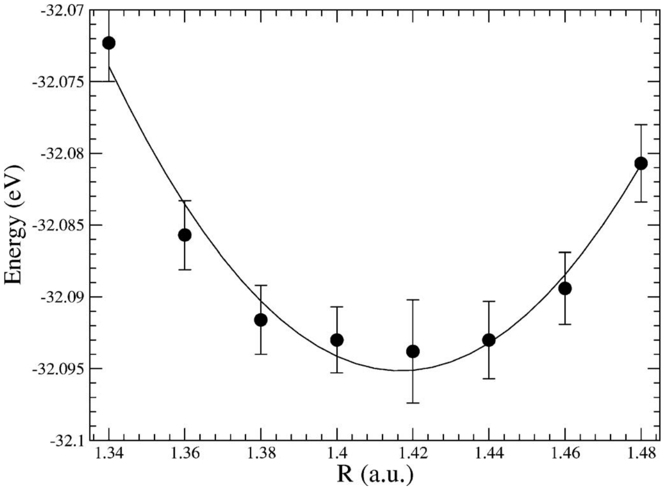

FIG. 1. The potential energy curve of H_2 as obtained by AF QMC with a plane wave basis and a hydrogen pseudopotential. We show also a Morse potential fit for the QMC data. The QMC equilibrium bond length from the fit is 1.416(4) bohr to be compared with the exact value of 1.400 83 bohr. The supercell used is $16\times12\times11$ bohr^3.

III. RESULTS USING PLANE WAVE BASIS SETS

Plane waves are more suited to periodic systems and require pseudopotentials to yield a tractable number of basis functions. However, isolated molecules can be studied with plane waves by employing periodic boundary conditions and large supercells, as in standard density functional theory (DFT) calculations. This is disadvantageous because one has to ensure that the supercells are large enough to control the spurious interactions between the periodic images of the molecule. For a given plane wave cutoff energy E_{cut}, the size of the plane wave basis increases in proportion to the volume of the supercell. Consequently, the computational cost for the isolated molecule tends to be higher than that using a localized basis, as we further discuss in Sec. IV. Although the plane wave basis calculations are expensive, they are valuable as they show the robustness and accuracy of the phaseless AF QMC method for extremely large basis sets (and correspondingly many auxiliary fields).

Here, we study H_2, H_3, and several other hydrogen-bonded molecules H_2O, OH, and H_2O_2. As is well known, first-row atoms such as oxygen are challenging since they have strong or "hard" pseudopotentials and require relatively large plane wave basis sets to achieve convergence. Even in hydrogen, where there are no core-electron states, pseudopotentials are usually used since they significantly reduce the plane wave basis size compared to treating the bare Coulomb potential of the proton. The hydrogen and oxygen pseudopotentials are generated by the OPIUM program,[19] using the neutral atoms as reference configurations. The cutoff radii used in the generation of the oxygen pseudopotentials are $r_c(s)=1.05$ and $r_c(p)=1.02$ bohr, where s and p correspond to $l=0$ and $l=1$ partial waves, respectively. For hydrogen $r_c(s)=0.66$ bohr was used. These relatively small r_c's are needed

TABLE I. Plane wave based calculations of the binding energy of H_2 vs supercell size. DFT/GGA and the phaseless AF QMC results are shown. All energies are in eV, and supercell dimensions are in a.u. For comparison, the all-electron GGA number is 4.568 eV (Ref. 17). Statistical errors are on the last digit and are shown in parentheses. The exact theoretical value is 4.746 eV (Ref. 18) and the experimental value is 4.75 eV (with zero-point energy removed).

Supercell	DFT/GGA	AF QMC
$11\times9\times7$	4.283	4.36(1)
$12\times10\times9$	4.444	4.57(1)
$14\times12\times11$	4.511	4.69(1)
$16\times12\times11$	4.512	4.70(1)
$22\times18\times14$	4.530	4.74(2)
∞	4.531	

for both atoms due to the short bond lengths in H_2O, and result in relatively hard pseudopotentials. Small r_c's, however, generally result in pseudopotentials with better transferability. In all of the studies shown below, the same pseudopotentials were used, even in molecules with larger bond lengths. The E_{cut} needed with these pseudopotentials is about 41 hartree. This E_{cut} was chosen such that the resulting plane wave basis convergence errors are less than a few meV in DFT calculations. A roughly similar plane wave basis convergence error is expected at the AF QMC level based on previous applications in TiO and other systems.[13,20,21] These convergence errors are much smaller than the QMC statistical error.

The quality of the pseudopotentials is further assessed by comparing the pseudopotential calculations with all-electron (AE) results using density functional methods, which test the pseudopotentials at least at the mean-field level. In all of the cases reported in this study, we found excellent agreement between AE and pseudopotential results, except in cases where the nonlinear core correction error[22,23] is important in DFT (all molecules containing oxygen), as we will discuss in Sec. III C.

As mentioned, AF QMC relies on a trial wave function to control the phase problem. In the plane wave calculations, we used a single Slater determinant from a plane wave based density functional calculation obtained with a generalized gradient approximation (GGA) functional,[24] with no further optimizations.

A. Singlet and triplet H_2 molecule

Table I summarizes results for the binding energy of the H_2 molecule, using DFT/GGA and AF QMC for several supercells, and compares these to exact results[18] and experiment. The experimental bondlength of H_2 was used in all the calculations. The binding energy is calculated as the difference in energy between the H atom (times two) and the molecule, each placed in the same supercell. The density functional binding energy obtained using the hydrogen pseudopotential converges, with respect to size effect, to 4.531 eV, which is in reasonable agreement with the all-electron (i.e., using the proton's bare Coulomb potential) binding energy of 4.568 eV obtained using NWCHEM (Ref. 17) and with the all-electron value of 4.540 eV reported in Ref. 25. The agreement between the pseudopotential and all-electron results is a reflection of the good transferability of the hydrogen pseudopotential. The AF QMC binding energy with the largest supercell is 4.74(2) eV, which is in excellent agreement with the experimental value of 4.75 eV (zero point energy removed) and the exact calculated value of 4.746 eV.[18]

TABLE II. Plane wave based AF QMC energies of the singlet ($^1\Sigma$) and triplet H_2 ($^3\Sigma$) molecule for two supercell sizes. The bond length was fixed at $R=1.42$ bohr in all cases. All energies are in eV. The exact calculated energy gap Δ is 10.495 eV (Ref. 18). Statistical errors are on the last digit and are shown in parenthesis.

Supercell	$^1\Sigma$	$^3\Sigma$	Δ
$11\times9\times7$	−32.59(1)	−22.329(4)	10.26(1)
$22\times18\times14$	−32.01(2)	−21.546(7)	10.46(2)

Figure 1 shows the H_2 AF QMC potential energy curve using a $16\times12\times11$ bohr^3 supercell. Finite-size effects, as in Table I, likely vary with the H_2 bond length and would affect the shape of the curve. Using a Morse potential fit, we obtained an estimated bond length of 1.416(4) bohr. (Using a second or fourth order polynomial fit leads to similar results; the fourth order fit the error bar is three times larger). For comparison, the exact equilibrium bond length of H_2 is 1.400 83 bohr,[18] and the DFT/GGA bond length is 1.4213 bohr.

The energy difference between the singlet and triplet H_2 spin states ($^1\Sigma$ and $^3\Sigma$, respectively) was also calculated. We note that for the singlet H_2 two-electron system, a HS transformation based on the magnetization[26] can be made to eliminate the sign problem and thus the need for the phaseless approximation. In this case the AF QMC calculations will become exact. This is not done here since our goal is to benchmark the general algorithm. The calculations for both singlet and triplet H_2 were at the experimental bond length of singlet H_2. Table II summarizes the results. The exact value obtained by Kolos and Roothaan is 10.495 eV,[18] with which the AF QMC value at the larger supercell size is in excellent agreement.

B. $H_2+H\rightarrow H+H_2$ transition state

The problem of calculating the transition state of H_3 is well benchmarked using a variety of methods.[27–31] The activation energy for the reaction $H_2+H\rightarrow H+H_2$ is defined as the difference between the energy of the H_3 saddle point and that of the well separated H atom and H_2 molecule.

Density functional methods are generally not very accurate in calculating the activation energy. For example, DFT with a local density approximation functional gives H_3 as a bound molecule with a binding energy of 0.087 eV at the symmetric configuration with $R_1=R_2=1.795$ bohr. DFT/GGA, on the other hand, gives a barrier height of 0.152 eV at the symmetric configuration $R=1.767$ bohr.[31] The experimental barrier height is 9.7 kcal/mol=0.42 eV.[32]

Using the AF QMC method, we studied the collinear H_3 system for three configurations with $R_1=R_2=1.600$, 1.757, and 1.900 bohr. Table III shows the calculated barrier

TABLE III. Symmetric collinear H_3 transition state energies using plane waves with pseudopotentials. Results are shown from density functional GGA [with pseudopotential (PSP) and without (AE) pseudopotentials], DMC, and the present AF QMC methods. [The "all-electron" GGA(AE) results are from well-converged large Gaussian basis set calculations.] The calculated results are for the linear H_3 molecule with $R_1=R_2=R$, for three values of R (in bohr). All energies are in eV. Statistical errors are on the last digit and are shown in parentheses.

R	GGA(AE)	GGA(PSP)	DMC (exact)	AF QMC
1.600	0.297	0.30	0.543 09(8)	0.54(3)
1.757	0.156	0.16	0.416 64(4)	0.43(3)
1.900	0.222	0.22	0.494 39(8)	0.48(4)

heights and compares these to results from DFT/GGA all-electron and pseudopotential calculations and to results from recent DMC calculations.[29] The AE and pseudopotential DFT/GGA results are in excellent agreement with each other, a further indication of the good quality of the H pseudopotential. The DMC calculations[29] are exact in this case through the use of a cancellation scheme,[33] which is very effective at eliminating the sign problem for small systems. The AF QMC values are in good agreement with the exact calculated results.

Plane wave based AF QMC calculations of the H_3 transition state are very expensive since the energy variations in the Born-Oppenheimer curve are quite small, as seen in Table III. To achieve the necessary accuracy, large supercells are needed, which results in large plane wave basis sets. The large basis sets lead to many thousands of AF's in Eq. (3). Moreover, a large number of AF's in general lead to a more severe phase problem and thus, potentially, to a more pronounced role for the phaseless approximation. The larger AF QMC statistical errors, compared to the highly optimized DMC results as well as to our Gaussian basis results in Sec. IV B, reflect the inefficiency of plane wave basis sets for isolated molecules. These calculations are valuable despite their computational cost, as they demonstrate the robustness of the method.

C. Hydrogen-bonded molecules

Complementing the study above of the H_3 system, where energy differences are small, we also examined three other hydrogen-bonded molecules: H_2O, OH, and H_2O_2, where the energy scales are large. Table IV compares the binding energies calculated using DFT/GGA (both pseudopotential and all electron), DMC,[34] and the present AF QMC method. (Results for the O_2 and O_3 molecules are included because they are pertinent to the discussion of pseudopotential errors below.) The experimental values,[35] with the zero point energy removed, are also shown. All of the calculations are performed at the experimental geometries of the molecules. The density functional all-electron binding energies in Table IV were obtained using the highly converged triple-zeta atomic natural orbital basis sets of Widmark *et al.*[36] They are in good agreement with published all-electron results. For example, the all-electron binding energy of H_2O is 10.147 eV and that of OH is 4.77 eV in Ref. 24. In Ref. 25, the binding energy of H_2O is 10.265 eV, and that of O_2 is 6.298 eV.[25]

In all of the molecules except H_2O, the DFT pseudopotential result seems to be in better agreement with the experimental value than the all-electron result. This is fortuitous and by no means suggests that the pseudopotential results are better than the all-electron values, since the pseudopotential results should reproduce the all-electron value obtained with the same theory. Any differences are, in fact, due to transferability errors of the pseudopotentials. At the density functional level, the molecular systems H_2O, OH, and H_2O_2 all need a nonlinear core correction (NLCC). The NLCC was introduced into DFT pseudopotential calculations by Louie *et al.*[22] It arises from the DFT-generated pseudopotential for oxygen, at the pseudopotential construction level in the descreening step, where the valence Hartree and nonlinear exchange-correlation terms are subtracted to obtain the ionic pseudopotential. The Hartree term is linear in the valence charge and can be subtracted exactly. This is not the case with the nonlinear exchange-correlation potential and will lead to errors especially when there is an overlap between the core and the valence charge densities. According to the NLCC correction scheme, this error can be largely rectified by retaining an approximate pseudocore charge density and by carrying it properly in the target (molecular or solid) calculations. This generally improves the transferability of the pseudopotentials.[22,23] The problem of NLCC is absent in effective core potentials generated using the Hartree-Fock method.

All of the molecules in Table IV suffer from the NLCC error which originates predominantly from the spin-polarized

TABLE IV. Calculated binding energies of H_2O, OH, H_2O_2, O_2, and O_3. Results are shown from density functional GGA [with (PSP) and without (AE) pseudopotentials], DMC, and the present AF QMC methods. Experimental results are also shown. DFT/GGA(PSP) and the present AF QMC results were calculated using plane wave basis sets with pseudopotentials. DFT/GGA(AE) is calculated using highly converged Gaussian basis sets. The DMC (Ref. 34) results were also obtained using pseudopotentials. The zero point energy is removed from the experimental data (Ref. 35). All energies are in eV. Statistical errors are on the last digit and are shown in parentheses.

	GGA(AE)	GGA(PSP)	DMC	AF QMC	Expt.
H_2O	10.19	9.82	10.10(8)	9.9(1)	10.09
OH	4.79	4.60	4.6(1)	4.7(1)	4.63
H_2O_2	12.26	11.66	11.4(1)	11.9(3)	11.65
O_2	6.22	5.72		5.2(1)	5.21
O_3	7.99	7.12		6.2(2)	5.82

oxygen atom, where the NLCC can be as large as 0.3 eV/atom within a GGA-PBE calculation.[23] For this reason, we have also included results for the O_2 and O_3 molecules. (The AF QMC value for O_2 is taken from Ref. 13.) As seen in the table, the binding energies of H_2O, OH, H_2O_2, O_2, and O_3 are smaller than the corresponding all-electron values of ≈0.37, 0.19, 0.60, 0.50, and 0.87 eV, respectively. These values are approximately proportional to the number of oxygen atoms in the corresponding molecule with a proportionality constant ≈0.3 eV, which agrees with the value reported in Ref. 23.

The pseudopotential is, of course, used differently in many-body AF QMC calculations. Despite the need for NLCC at the DFT level, the oxygen pseudopotential seems to be of good quality when used in AF QMC. In all the cases, the AF QMC results are in good agreement with DMC and with the experimental values. The largest discrepancy with experiment is ≈0.4(2) eV with O_3, and it is in opposite direction to the NLCC as done at the density functional level.

The need for the nonlinear core correction does not indicate a failure of the frozen-core approximation, but rather is a consequence of the nonlinear dependence of the spin-dependent exchange-correlation potential on the *total* spin density (valence+core) in the density functional theory. The QMC calculations depend only on the bare ionic pseudopotential and do not have this explicit dependence on the (frozen) core-electron spin densities. It is thus reasonable to expect the QMC results to be not as sensitive to this issue.

IV. RESULTS USING GAUSSIAN BASIS SETS

In this section, we present our studies using Gaussian basis sets. For comparison, some of the systems are repeated from the plane wave and pseudopotential studies in the previous section. Gaussian basis sets are, in general, more efficient for isolated molecules. For example, the calculations below on the van der Waals minimum in H_3 would be very difficult with the plane wave formalism because of the large supercells necessary and because of the high statistical accuracy required to distinguish the small energy scales. Also, all-electron calculations are feasible with a Gaussian basis, at least for lighter elements, so systematic errors due to the use of pseudopotentials can be avoided without incurring much additional cost.

A direct comparison with experimental results requires large, well-converged basis sets in the AF QMC calculations.[3,14] As mentioned, the convergence of Gaussian basis sets is not as straightforward to control as that of plane waves. For benchmarking the accuracy of the AF QMC method, however, we can also compare with other established correlated methods such as full configuration interaction (FCI) and CCSD(T) since all the methods operate on the same Hilbert space. FCI energies are the exact results for the Hilbert space thus defined. The FCI method has an exponential scaling with the number of particles and basis size, so it is only used with small systems. In this section, we study H_2 and H_3, which are challenging examples for mean-field methods, and compare the AF QMC results with exact results.

TABLE V. Symmetric collinear H_3 transition state total energies using aug-cc-pVDZ and aug-cc-pVTZ Gaussian basis sets (Ref. 6). We examined five configurations with $R_1=R_2=R$, and we report the unrestricted Hartree-Fock (UHF), full configuration interaction (FCI), and AF QMC total energies. Bond lengths are in bohr and energies are in hartrees. Statistical errors are on the last digit and are shown in parentheses.

R	UHF	FCI	AF QMC
aug-cc-pVDZ			
1.600	−1.595026	−1.642820	−1.64256(5)
1.700	−1.600252	−1.648186	−1.64775(5)
1.757	−1.601336	−1.649328	−1.64882(5)
1.800	−1.601406	−1.649433	−1.64898(6)
1.900	−1.599536	−1.647606	−1.64697(6)
aug-cc-pVTZ			
1.600	−1.599843	−1.652219	−1.65178(7)
1.700	−1.604162	−1.656405	−1.65586(7)
1.757	−1.604835	−1.657013	−1.65652(7)
1.800	−1.604638	−1.656770	−1.65624(8)
1.900	−1.602269	−1.654285	−1.65368(9)

The matrix elements which enter in the definition of the Hamiltonian of the system of Eq. (1) are calculated using NWCHEM.[3,17] The trial wave functions, which are used to control the phase problem, are mostly computed using unrestricted Hartree-Fock (UHF) methods, although we have also tested ones from density functional methods. In previous studies, we have rarely seen any difference in the AF QMC results between these two types of trial wave functions. This is the case for most of the systems in the present work, and only one set of results is reported. In H_3 near the van der Waals minimum, where extremely small energy scales need to be resolved, we find small differences (∼0.1 millihartree), and we report results from the separate trial wave functions. The FCI calculations were performed using MOLPRO.[37,38]

A. Bond length of H_2

We first study H_2 again, with a cc-pVTZ basis set which has 28 basis functions for the molecule. This is to be compared with the plane wave calculations which has about 5000–70 000 plane waves for the different supercells used. These H-bonded systems are especially favorable for localized basis sets. The AF QMC equilibrium bondlength $R=1.4025(6)$ bohr compares well to the corresponding FCI bond length of $R=1.402\,65$ bohr, with both methods using the cc-pVTZ basis. This is a substantially better estimate of the exact infinite basis result of $R_e=1.400\,83$ bohr (Ref. 18) than was obtained from the plane wave AF QMC results in Fig. 1. The remaining finite-basis error is much smaller than the statistical errors in the planewave calculations. (The small residual finite-basis error is mostly removed at the cc-pVQZ basis set level, with an equilibrium bond length of $R=1.401\,11$ bohr from FCI.)

B. $H_2+H\rightarrow H+H_2$ transition state

Table V presents calculated total energies of H_3 with aug-cc-pVDZ and aug-cc-pVTZ basis sets.[6] Results obtained

TABLE VI. H_3 total energies in the van der Walls limit. R_1 is fixed at 1.4 bohr, and R_2 is varied between 4 and 10 bohr. The aug-cc-pVTZ basis set is used. Energies are in hartrees. Statistical errors are on the last digit and are shown in parentheses.

R_2	FCI	AF QMC/UHF
4	−1.671577	−1.67160(9)
5	−1.672455	−1.67250(8)
6	−1.672535	−1.67263(6)
7	−1.672508	−1.67265(5)
10	−1.672462	−1.67257(6)

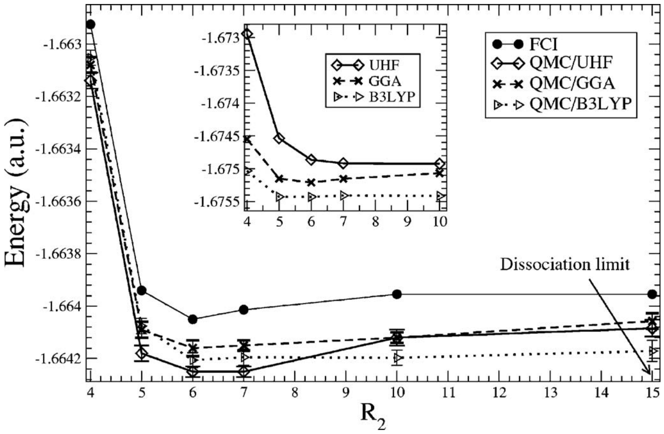

FIG. 2. The potential energy curve of H_3 in the van der Waals limit using aug-cc-pVDZ basis set. R_1 is set to 1.4 bohr, and R_2 is varied between 4 and 10 bohr. (The dissociation limit is shown at $R_2=15$ bohr in the figure). FCI results are compared with AF QMC results with three different trial wave functions, from UHF and DFT with GGA and B3LYP functionals, respectively. The inset shows the corresponding potential energy curves obtained from UHF, GGA, and B3LYP. (For clarity, the UHF and GGA energies are shifted by −0.047 and −0.154 hartrees in the inset, respectively.)

with UHF, FCI, and the present AF QMC methods are shown for five different geometries in the collinear H_3 system. The present TZ-basis FCI results were cross-checked with those in Ref. 30, which contains a detailed study of the Born-Oppenheimer potential energy curves for the $H+H_2$ system.

The AF QMC total energies are in excellent agreement, to within less than 1 mE_h, with the FCI energies. The AF QMC barrier heights with the aug-cc-pVDZ and aug-cc-pVTZ basis sets at $R=1.757$ bohr are 0.444(2) and 0.434(3) eV, respectively. The corresponding FCI results are 0.4309 and 0.4202 eV, respectively. Thus, the AF QMC results show a systematic error of ~0.015 eV in the barrier height. It is possible to resolve these small discrepancies because the basis sets are much more compact, with 25–75 Gaussian basis functions as opposed to approximately 10 000 plane waves in the calculations in Sec. III B. As a result, the statistical errors are smaller than in the plane wave calculations by a factor of 10, with only a small fraction of the computational time. Even with these relatively small basis sets, we see that the finite-basis errors are quite small here. In fact, the FCI barrier height with the aug-cc-pVTZ basis is in agreement with the experimental value of 0.42 eV.[32]

C. van der Waals minimum in collinear H_3

The van der Waals minimum of H_3 is studied by fixing $R_1=1.4$ bohr (the H_2 equilibrium bond length), while the distance R_2 between the third H atom and the closer of the two atoms in H_2 was varied between 4 and 10 bohr. The potential energy curve of this system exhibits a very shallow minimum of approximately 85 μE_h (Ref. 30) at $R_2 \sim 6$ bohr. Two different basis sets, aug-cc-pVDZ and aug-cc-pVTZ, were used. Table VI shows the aug-cc-pVTZ results, and Fig. 2 plots the aug-cc-pVDZ results.

As seen in Table VI, the AF QMC all-electron total energies are in excellent agreement with FCI, with a maximum discrepancy of about 0.14(5) mE_h. The AF QMC energies, which are calculated with the mixed estimator, are not variational, as is evident in the results from both basis sets compared to FCI. The AF QMC results in Table VI are obtained with an UHF trial wave function. In most of our molecular calculations with Gaussian basis sets, the UHF solution, which is the variationally optimal single Slater determinant, has been chosen as the trial wave function.[3,14] In the present case, the UHF method actually fails to give a van der Waals minimum, as can be seen from the inset in Fig. 2. It is reassuring that AF QMC correctly reproduces the minimum with UHF as a trial wave function.

The effects of using two other single Slater determinant trial wave functions were also tested. These were obtained from DFT GGA and B3LYP calculations with the aug-cc-pVDZ basis set. The corresponding results are also shown in Fig. 2. In the DFT calculations (shown in the inset in Fig. 2), both GGA and B3LYP predict the existence of a minimum, although B3LYP gives an unphysical small barrier at about $R_2 \approx 7$ bohr. The AF QMC results obtained with UHF, GGA, and B3LYP Slater determinants as trial wave functions differ somewhat, but are reasonably close to each other. With the GGA trial wave function, AF QMC “repairs” the well depth (possibly with a slight overcorrection). With the B3LYP trial wave function, AF QMC appears to underestimate the well depth, giving a well shape that is difficult to characterize because of the statistical errors and the extremely small energy scale of these features.

V. SUMMARY

We have presented a benchmark study of the phaseless AF QMC method in various H-bonded molecules. The auxiliary-field QMC method is a many-body approach formulated in a Hilbert space defined by a single-particle basis. The choice of a basis set is often of key importance, as it can affect the efficiency of the calculation. In the case of AF QMC, the basis set choice can also affect the systematic error because of the different HS transformation that can result. In this study, we employed plane wave basis sets with pseudopotentials and all-electron Gaussian basis sets to compare the performance of the AF QMC method. The plane wave HS decomposition was tailored to the plane wave representation, resulting in $\mathcal{O}(8M)$ auxiliary fields, where M is the number of plane waves. For the Gaussian basis sets, the generic HS decomposition described in Sec. II was used, resulting in $\mathcal{O}(M^2)$ auxiliary fields. Typical M values in this study were tens of thousands in the plane wave calculations and a hundred in the Gaussian calculations.

The plane wave calculations were carried out for H_2, H_2+H near the transition state, H_2O, OH, and H_2O_2. Nonlinear core corrections to the oxygen pseudopotential were discussed using additional calculations for the O_2 and O_3 molecules. DFT GGA pseudopotentials were employed. The trial wave functions were single Slater determinants obtained from DFT GGA with identical plane wave and pseudopotential parameters as in the AF QMC calculations. Hard pseudopotentials and large plane wave cutoffs were used to ensure basis-size convergence and the transferability of the pseudopotentials. Large supercells were employed to remove finite-size errors. To mimic typical systems in the solid state, no optimization was done to take advantage of the simplicity of these particular systems. The binding energies computed from AF QMC have statistical errors of 0.1–0.3 eV as a result. Within this accuracy, the AF QMC results are in excellent agreement with experimental values.

Gaussian basis AF QMC calculations were carried out on H_2, the transition state of H_2+H, as well as the van der Waals minimum in linear H_2+H. These calculations are within the framework of standard quantum chemistry many-body approaches using the full Hamiltonian without pseudopotentials. UHF single Slater determinants were used as the trial wave function. For various geometries, the absolute total energies from AF QMC agree with FCI to well within 1 mE_h. The calculated equilibrium bond lengths and potential energy curves are also in excellent agreement with FCI. In H_2+H, AF QMC correctly recovers the van der Waals well with an UHF trial wave function which in itself predicts no binding.

Comparing plane wave and Gaussian basis set AF QMC results, we can conclude the following. In the Gaussian basis calculations, as evident from FCI comparisons, errors due to controlling the phase problem in the phaseless approximation are well within 1 mE_h in the absolute energies. Achieving the infinite basis limit is more straightforward using plane wave based AF QMC, but statistical errors are larger for the isolated molecules studied due to the need for large supercells. Within statistical errors, however, the AF QMC results using both types of basis sets were in agreement. This indicates that errors due to the use of pseudopotentials with plane wave basis sets were smaller than the statistical errors. Finally, within statistical errors, the performance of the phaseless AF QMC method did not appear to be sensitive to the type of HS decomposition used, despite drastic differences in the basis size and the number of auxiliary fields.

ACKNOWLEDGMENTS

The authors would like to thank E. J. Walter for many useful discussions. This work is supported by ONR (N000140110365 and N000140510055), NSF (DMR-0535529), and ARO (Grant No. 48752PH). Computations were carried out in part at the Center for Piezoelectrics by Design, the SciClone Cluster at the College of William and Mary, and NCSA at UIUC.

[1] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **71**, 33 (2001).
[2] S. Zhang and H. Krakauer, Phys. Rev. Lett. **90**, 136401 (2003).
[3] W. A. Al-Saidi, S. Zhang, and H. Krakauer, J. Chem. Phys. **124**, 224101 (2006).
[4] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).
[5] A compilation of basis sets is present in the Extensible Computational Chemistry Environment Basis Set Database (http://www.emsl. pnl. gov/forms/basisform. html).
[6] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[7] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **98**, 1358 (1993).
[8] D. M. Ceperley and B. J. Alder, J. Chem. Phys. **81**, 5833 (1984); J. B. Anderson, in *Quantum Monte Carlo: Atoms, Molecules, Clusters, Liquids and Solids*, Reviews in Computational Chemistry Vol. 13, edited by K. B. Lipkowitz and D. B. Boyd (Wiley, NY, 1999).
[9] S. Zhang and M. H. Kalos, Phys. Rev. Lett. **71**, 2159 (1993).
[10] S. Zhang, J. Carlson, and J. E. Gubernatis, Phys. Rev. B **55**, 7464 (1997).
[11] J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975).
[12] J. W. Moskowitz, K. E. Schmidt, M. A. Lee, and M. H. Kalos, J. Chem. Phys. **77**, 349 (1982); P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, *ibid.* **77**, 5593 (1982).
[13] W. A. Al-Saidi, H. Krakauer, and S. Zhang, Phys. Rev. B **73**, 075103 (2006).
[14] W. A. Al-Saidi, H. Krakauer, and S. Zhang, J. Chem. Phys. **125**, 154110 (2006).
[15] W. A. Al-Saidi, H. Krakauer, and S. Zhang, (unpublished).
[16] R. L. Stratonovich, Sov. Phys. Dokl. **2**, 416 (1958); J. Hubbard, Phys. Rev. Lett. **3**, 77 (1959).
[17] T. P. Straatsma, E. Aprá, T. L. Windus *et al.*, NWCHEM, Version 4.6, a computational chemistry package for parallel computers, Pacific Northwest National Laboratory, Richland, WA 99352-0999, 2004.
[18] W. Kolos and C. C. J. Roothaan, Rev. Mod. Phys. **32**, 219 (1960).
[19] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, Phys. Rev. B **41**, R1227 (1990).
[20] S. Zhang, H. Krakauer, W. Al-Saidi, and M. Suewattana, Comput. Phys. Commun. **169**, 394 (2005).
[21] M. Suewattana, W. Purwanto, S. Zhang, H. Krakauer, and E. J. Walter (unpublished).
[22] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).
[23] D. Porezag, M. R. Pederson, and A. Y. Liu, Phys. Rev. B **60**, 14132 (1999).
[24] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[25] D. C. Patton, D. V. Porezag, and M. R. Pederson, Phys. Rev. B **55**, 7454 (1996).
[26] P. L. Silvestrelli, S. Baroni, and R. Car, Phys. Rev. Lett. **71**, 1148 (1993).
[27] P. Siegbahn and B. Liu, J. Chem. Phys. **68**, 2457 (1978).
[28] D. L. Diedrich and J. B. Anderson, J. Chem. Phys. **100**, 8089 (1994).
[29] K. E. Riley and J. B. Anderson, J. Chem. Phys. **118**, 3437 (2002).
[30] S. L. Mielke, B. C. Garrett, and K. A. Peterson, J. Chem. Phys. **116**, 4142 (2001).
[31] D. Porezag and M. R. Pederson, J. Chem. Phys. **102**, 9345 (1995).
[32] W. R. Schulz and D. J. Le Roy, J. Chem. Phys. **42**, 3869 (1965).
[33] J. B. Anderson, C. A. Traynor, and B. M. Boghosian, J. Chem. Phys. **95**, 7418 (1991).
[34] J. C. Grossman, J. Chem. Phys. **117**, 1434 (2002).
[35] D. Feller and K. A. Peterson, J. Chem. Phys. **110**, 8384 (2002).
[36] P. O. Widmark, P. A. Malmqvist, and B. Roos, Theor. Chim. Acta **77**, 291 (1990).
[37] H.-J. Werner, P. J. Knowles, J. Almolf *et al.*, MOLPRO, Version 2002.6, a package of *ab initio* programs, Universitat Bielefeld, Bielefeld, Germany, University of Sussex, Falmer, Brighton, England, 1996.
[38] P. J. Knowles and N. C. Handy, Chem. Phys. Lett. **111**, 315 (1984); Comput. Phys. Commun. **54**, 75 (1989).