

# Correlation Consistent Gaussian Basis Sets for H, B–Ne with Dirac–Fock AREP Pseudopotentials: Applications in Quantum Monte Carlo Calculations

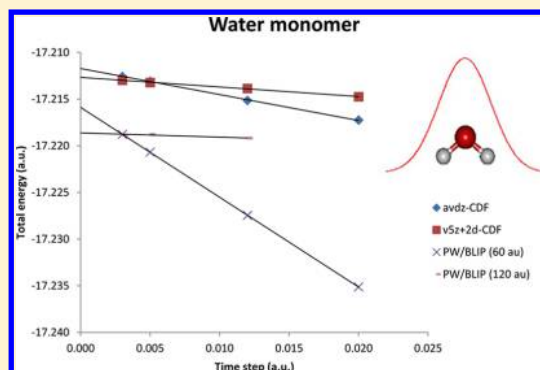
Jiawei Xu,<sup>†</sup> Michael J. Deible,<sup>†</sup> Kirk A. Peterson,<sup>‡</sup> and Kenneth D. Jordan<sup>\*,†</sup>

<sup>†</sup>Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15218, United States

<sup>‡</sup>Department of Chemistry, Washington State University, Pullman, Washington 99164, United States

## S Supporting Information

**ABSTRACT:** In this paper, we introduce correlation consistent Gaussian-type orbital basis sets for the H and B–Ne atoms for use with the CASINO Dirac–Fock AREP pseudopotentials. These basis sets are tested in coupled cluster calculations on H<sub>2</sub>, B<sub>2</sub>, C<sub>2</sub>, N<sub>2</sub>, O<sub>2</sub>, and F<sub>2</sub> as well as in quantum Monte Carlo calculations on the water monomer and dimer and the water–benzene complex, where they are found to give low variances in variational Monte Carlo calculations and to lead to reduced time step errors and improved convergence in diffusion Monte Carlo calculations compared to the use of nonoptimized basis sets. The use of basis sets with a large number of contracted *s* and *p* primitives is found to be especially important for the convergence of the energy in the diffusion Monte Carlo calculations.



## 1. INTRODUCTION

Quantum Monte Carlo methods,<sup>1,2</sup> because of their cubic scaling with the number of atoms, hold considerable promise for providing accurate interaction energies of molecular clusters and solids. Most quantum Monte Carlo electronic structure calculations make use of the fixed-node approximation<sup>1,2</sup> to enforce fermionic behavior on the wave function. The fixed nodal surface is enforced by a trial function, generally taken to be a single Slater determinant of Hartree–Fock or DFT orbitals. In practice, the trial function also contains Jastrow factors<sup>3</sup> to describe short-range electron–electron and electron–nuclei interactions, with the parameters in the Jastrow factors being optimized by the use of the variational Monte Carlo (VMC) procedure. The VMC step is generally followed by diffusion Monte Carlo (DMC) calculations where most of the computational effort is spent. The orbitals in the trial functions are most frequently represented in terms of plane-wave functions or Gaussian-type orbitals (GTOs). In the former case, use of pseudopotentials is essential to avoid the prohibitively high plane-wave cutoffs that would be required for all-electron calculations. Even when using GTO basis sets, it is advantageous to use pseudopotentials in quantum Monte Carlo calculations as this greatly reduces the computational effort to achieve small statistical errors.

In exploratory applications of quantum Monte Carlo methods employing pseudopotentials and trial functions expressed in terms of GTOs, we observed surprisingly large variances of the VMC energies. In some cases the variances were as much as a factor of 6 larger than obtained with high cutoff plane-wave basis sets.<sup>4,5</sup> This naturally raises concern

about the impact of such trial functions on the interaction energies obtained from subsequent DMC calculations. This concern led us to design for H and B–Ne correlation consistent GTO basis sets for use with the CASINO Dirac–Fock average relativistic (AREP) pseudopotentials,<sup>6,7</sup> which we test in coupled cluster calculations on H<sub>2</sub>, B<sub>2</sub>, C<sub>2</sub>, N<sub>2</sub>, O<sub>2</sub>, and F<sub>2</sub> and in quantum Monte Carlo calculations on the water monomer and dimer as well as on the water–benzene complex. In addition, we examine the performance of two methods that have been designed for dealing with the problems associated with using nonlocal pseudopotentials in diffusion Monte Carlo calculations.

## 2. PERFORMANCE OF GTO BASIS SETS IN VMC CALCULATIONS

To illustrate the nature of the large variance problem when using certain GTO basis set/pseudopotential combinations, we summarize in Table 1 the energies and variances from VMC calculations on the water molecule using the CASINO Dirac–Fock (CDF) pseudopotential on the H and O atoms, with the molecular orbitals in the trial functions being represented either by the valence double- $\zeta$  plus polarization function basis set of Burkatzki, Filippi, and Dolg (BFD),<sup>8</sup> augmented with diffuse *s*, *p*, and *d* functions from the aug-cc-pVDZ basis set,<sup>9,10</sup> or by plane-wave basis sets with energy cutoffs of 60, 120, and 160 au. The geometry of the water monomer was taken from experiment,<sup>11</sup> with OH distances of 0.9572 Å and an HOH

Received: November 9, 2012

Published: March 18, 2013

**Table 1. VMC Energies and Variances for the Water Monomer Using the CASINO Dirac-Fock Pseudopotential on All Atoms<sup>a</sup>**

basis set	VMC energy (au)	variance of the VMC energy (au)
augmented BFD	−17.161(3)	1.25
plane-wave/BLIP (60 au) <sup>b</sup>	−17.159(5)	1.63
plane-wave/BLIP (120 au) <sup>b</sup>	−17.191(2)	0.34
plane-wave/BLIP (160 au) <sup>b</sup>	−17.194(2)	0.26

<sup>a</sup>Hartree–Fock and LDA calculations were used to obtain the molecular orbitals for GTO and plane-wave basis sets, respectively.

<sup>b</sup>The plane-wave energy cutoff is given in parentheses.

angle of 104.52°. In the calculations using the GTOs, the trial wave functions were taken from Hartree–Fock calculations, and in the calculations with plane-wave (PW) basis sets, the orbitals for the trial function were taken from local density

approximation (LDA) density functional theory calculations, with the orbitals being converted to BLIP-type spline functions.<sup>12</sup> In separate calculations using the augmented BFD basis set, we confirmed that the energy and variance from the VMC calculations are nearly the same whether using trial functions expanded in terms of Hartree–Fock or LDA orbitals. Three-term (i.e., e-e, e-n, and e-e-n) Jastrow factors were employed, the parameters in which were optimized so as to minimize the variance of the energy.<sup>3</sup> The Hartree–Fock and the LDA calculations with the augmented BFD basis set calculations were carried out using the Gaussian 03 package,<sup>13</sup> and the plane-wave DFT calculations were carried out using ABINIT.<sup>14</sup> The quantum Monte Carlo calculations were carried out using the CASINO code.<sup>15</sup>

The VMC calculations with the augmented BFD GTO basis set gave a variance of 1.25 au compared with variances of 1.63, 0.34, and 0.26 au obtained using plane-wave basis sets with cutoffs of 60, 120, and 160 au, respectively. The importance of

**Table 2. Calculated CCSD(T) Spectroscopic Constants for the Ground States of H<sub>2</sub>, B<sub>2</sub>, C<sub>2</sub>, N<sub>2</sub>, O<sub>2</sub>, and F<sub>2</sub> with the aug-cc-pVnZ-CDF Basis Sets<sup>a</sup>**

	basis	$E_e (E_h)$	$D_e$ (kcal/mol)	$r_e$ (Å)	$\omega_e$ (cm <sup>−1</sup> )
H <sub>2</sub>	aug-cc-pVDZ-CDF	−1.167484	105.09	0.7566	4383.2
	aug-cc-pVTZ-CDF	−1.173075	108.60	0.7429	4408.2
	aug-cc-pVQZ-CDF	−1.174023	109.20	0.7418	4402.1
	aug-cc-pVSZ-CDF	−1.174295	109.37	0.7415	4403.2
	aug-cc-pVSZ-DK	−1.174263	109.35	0.7416	4403.1
	expt <sup>b</sup>		109.49	0.7414	4401.2
B <sub>2</sub>	aug-cc-pVDZ-CDF	−5.296648	55.48	1.6378	995.0
	aug-cc-pVTZ-CDF	−5.321179	61.78	1.6029	1029.3
	aug-cc-pVQZ-CDF	−5.326409	63.26	1.5965	1036.5
	aug-cc-pVSZ-CDF	−5.327859	63.64	1.5954	1039.3
	aug-cc-pVSZ-DK	−49.319978	64.83	1.5927	1048.6
	expt <sup>b</sup>		67.65	1.5900	1051.3
C <sub>2</sub>	aug-cc-pVDZ-CDF	−10.998941	126.15	1.2761	1803.8
	aug-cc-pVTZ-CDF	−11.049833	138.33	1.2520	1828.8
	aug-cc-pVQZ-CDF	−11.062220	141.96	1.2470	1842.9
	aug-cc-pVSZ-CDF	−11.065867	142.96	1.2459	1846.5
	aug-cc-pVSZ-DK	−75.836466	144.27	1.2447	1858.0
	expt <sup>b</sup>		147.8	1.2425	1854.7
N <sub>2</sub>	aug-cc-pVDZ-CDF	−19.798467	196.94	1.1228	2268.4
	aug-cc-pVTZ-CDF	−19.879169	215.16	1.1048	2325.0
	aug-cc-pVQZ-CDF	−19.900415	221.22	1.1012	2340.8
	aug-cc-pVSZ-CDF	−19.907062	223.06	1.1003	2345.0
	aug-cc-pVSZ-DK	−109.473998	225.46	1.0993	2358.5
	expt <sup>b</sup>		228.4	1.0977	2358.6
O <sub>2</sub>	aug-cc-pVDZ-CDF	−31.765110	104.53	1.2257	1551.6
	aug-cc-pVTZ-CDF	−31.866955	113.16	1.2165	1568.5
	aug-cc-pVQZ-CDF	−31.895146	116.21	1.2118	1583.0
	aug-cc-pVSZ-CDF	−31.904714	117.19	1.2107	1587.7
	aug-cc-pVSZ-DK	−150.295526	118.89	1.2073	1598.9
	expt <sup>b</sup>		120.6	1.2075	1580.2
F <sub>2</sub>	aug-cc-pVDZ-CDF	−48.139300	30.31	1.4483	844.1
	aug-cc-pVTZ-CDF	−48.267510	35.72	1.4226	906.2
	aug-cc-pVQZ-CDF	−48.305208	36.92	1.4171	917.2
	aug-cc-pVSZ-CDF	−48.318677	37.37	1.4151	921.7
	aug-cc-pVSZ-DK	−199.558515	38.05	1.4114	926.5
	expt <sup>b</sup>		39.0	1.4119	916.6

<sup>a</sup>H<sub>2</sub> is calculated at the CCSD (i.e., FCI) level. Atomic spin–orbit effects have been removed from the experimental dissociation energies in the cases of C<sub>2</sub>, O<sub>2</sub>, and F<sub>2</sub> by using the experimental atomic splittings. <sup>b</sup>Experimental results are from ref 28, except for the dissociation energy of C<sub>2</sub> which is taken from ref 29. The reference dissociation energy of B<sub>2</sub> is from high level theoretical calculations of ref 30.

going to very high energy cutoffs when using plane-wave basis sets in quantum Monte Carlo calculations has been noted previously in the literature.<sup>16</sup>

Adding higher angular momentum functions to the augmented BFD GTO basis set had little effect on the variance. This led us to examine the variances obtained in all-electron VMC calculations using Dunning's cc-pVDZ, cc-pVTZ, cc-pVQZ, cc-pVSZ, and cc-pV6Z basis sets,<sup>9</sup> omitting *g* and higher angular momentum functions, as these are not supported by the CASINO code. The variances of the VMC energy of the water monomer for the above sequence of basis sets are 3.6, 2.4, 1.5, 1.1, and 0.8 au, respectively. Similar variances are obtained with the corresponding aug-cc-pVXZ basis sets.<sup>9,10</sup> The variance in the complete basis set limit is necessarily larger in all-electron than in pseudopotential calculations, due to the large contribution of the 1s electrons to the total energies in the former. These results suggest that the large variance found in the VMC calculations with the augmented BFD basis set is due to this basis set being far from optimal for use with the CASINO Dirac–Fock pseudopotential. Indeed, when used with BFD pseudopotentials,<sup>8</sup> for which the BFD basis sets were developed, the variance from a VMC calculation of the water monomer using the augmented BFD basis set is less than 0.3 au.

### 3. AUG-CC-PVXZ-TYPE BASIS SETS FOR USE WITH THE CASINO DIRAC-FOCK PSEUDOPOTENTIALS

Although high quality aug-cc-pVXZ-type basis sets for use with pseudopotentials have been developed for heavier elements (aug-cc-pVXZ-PP),<sup>17–21</sup> such basis sets have not been developed for B–Ne, primarily due to the fact that with traditional quantum chemistry methods, there is little computational advantage to replacing the 1s orbitals by pseudopotentials. However, as noted above, the use of pseudopotentials to model the 1s electrons of B–Ne is more advantageous in quantum Monte Carlo calculations. With this in mind, we have designed a series of correlation consistent basis sets for boron, carbon, oxygen, nitrogen, fluorine, and neon with the core 1s electrons described by CASINO Dirac–Fock (CDF) pseudopotentials. In addition, to facilitate comparison with calculations employing trial functions expressed in terms of plane-wave basis sets, we also developed analogous basis sets for use with the CASINO Dirac–Fock pseudopotential for hydrogen. These basis sets are designated aug-cc-pVDZ-CDF, aug-cc-pVTZ-CDF, aug-cc-pVQZ-CDF, and aug-cc-pVSZ-CDF and are described in Tables S1–S5 in the Supporting Information.

As a test of the aug-cc-pVXZ-CDF basis sets, we have optimized the bond lengths ( $R_e$ ) and calculated the dissociation energies ( $D_e$ ) and harmonic vibrational frequencies ( $\omega_e$ ) of the electronic ground states of  $H_2$ ,  $B_2$ ,  $C_2$ ,  $N_2$ ,  $O_2$ , and  $F_2$  using the CCSD(T) method.<sup>22,23</sup> For comparison, all-electron CCSD(T) calculations employing the Douglas–Kroll–Hess (DKH) scalar relativistic Hamiltonian<sup>24</sup> and the aug-cc-pVSZ-DK basis sets<sup>9,10,25</sup> within the frozen-core approximation were also carried out. For the atoms and open-shell molecules, the ROHF-UCCSD(T) method<sup>26,27</sup> was utilized. The calculated  $R_e$ ,  $D_e$ , and  $\omega_e$  results together with the corresponding experimental values are summarized in Table 2. For  $H_2$ , the CCSD(T) calculations with the CDF pseudopotential and aug-cc-pVQZ-CDF and aug-cc-pVSZ-CDF basis sets as well as the CCSD(T) calculations using the DKH Hamiltonian and the aug-cc-pVSZ-DK basis set give bond lengths, vibrational

frequencies, and dissociation energies very close to the experimental values.

For the  $B_2$ ,  $C_2$ ,  $N_2$ ,  $O_2$ , and  $F_2$  diatomics, the bond lengths from the CCSD(T)/aug-cc-pVXZ-CDF optimizations systematically shorten as the basis set is enlarged, with the CCSD(T)/aug-cc-pVSZ-CDF bond lengths being only 0.0026–0.0036 Å longer than the experimental values. For all five diatomics, the difference of the calculated bond length from experiment is greater in the pseudopotential than in the all-electron calculations. The trends in the harmonic frequencies are consistent with those in the calculated bond lengths, with the values of the calculated frequencies increasing with increasing flexibility of the basis set. The largest errors in the harmonic frequencies calculated at the CCSD(T)/aug-cc-pVSZ-CDF level are  $-12$  and  $-13$   $cm^{-1}$  for  $B_2$  and  $N_2$ , respectively. For all five dimers, the vibrational frequencies from the pseudopotential calculations are about 10  $cm^{-1}$  smaller than those from the all-electron calculations. The results for both the frequencies and the bond lengths are indicative of small errors caused by the CDF pseudopotentials.

As expected, the CCSD(T) values of the dissociation energies smoothly increase along the sequence of aug-cc-pVXZ-CDF basis sets with the CCSD(T)/aug-cc-pVSZ-CDF values of the dissociation energies appearing to be converged to within 1 kcal/mol of the complete basis set limit (CBS) values for  $B_2$ ,  $C_2$ ,  $O_2$ , and  $F_2$  and to within 2 kcal/mol for  $N_2$ . (There is not an accurate experimental  $D_e$  value for  $B_2$ , and as a reference we have used instead the near full-CI result from ref 30.) This is consistent with the well-known trends in dissociation energies calculated using the all-electron aug-cc-pVXZ series of basis sets. Compared to experiment, the errors in the CCSD(T)/aug-cc-pVSZ-CDF binding energies range from 2.6 to 4.5 kcal/mol, with the largest error being for  $N_2$ . The errors in the dissociation energies of the corresponding all-electron calculations are 1.2–2.9 kcal/mol smaller, again indicating that small errors have been introduced by the CDF pseudopotentials. The errors in the dissociation energies from the all-electron calculations have several origins including basis set incompleteness, the neglect of core–valence correlation, and the neglect of correlation effects not recovered at the CCSD(T) level. Having established the suitability of the aug-cc-pVXZ-CDF basis sets for describing the bonding in the diatomic species, we now turn to their performance in quantum Monte Carlo calculations.

### 4. APPLICATION OF THE AUG-CC-PVXZ-CDF BASIS SETS IN QMC CALCULATIONS

**A. VMC Results for the Water Monomer.** The energies and variances from VMC calculations on the water monomer with trial functions expanded in terms of the aug-cc-pVXZ-CDF basis sets for the O atom and the standard aug-cc-pVXZ basis sets for the H atoms are reported in Table 3. These calculations used the CASINO Dirac–Fock pseudopotentials on the O atoms only. Test calculations revealed that nearly the same energies and variances result when the CDF pseudopotential and the aug-cc-pVXZ-CDF basis set are also employed on the H atoms. For this reason, unless noted otherwise, in the remainder of the paper in presenting results using GTO basis sets for the trial functions, the CASINO Dirac–Fock pseudopotential and aug-cc-pVXZ-CDF basis sets are used only for the non-hydrogen atoms. From comparison of the results in Tables 1 and 3, it is seen that the variance in the energy is reduced about threefold and the VMC energy is about

**Table 3. VMC Energies and Variances for the Water Monomer<sup>a</sup>**

basis set	VMC energy (au)	variance (au)
aug-cc-pVDZ-CDF <sup>a</sup>	-17.193(2)	0.42
aug-cc-pVTZ-CDF <sup>a</sup>	-17.197(1)	0.29
aug-cc-pVQZ-CDF <sup>a</sup>	-17.199(1)	0.23
aug-cc-pVSZ-CDF <sup>a</sup>	-17.200(1)	0.22
cc-pVSZ( <i>sp</i> )+2 <i>d</i> -CDF <sup>b</sup>	-17.197(1)	0.25
cc-pVSZ( <i>sp</i> )+2 <i>df</i> -CDF <sup>b</sup>	-17.198(1)	0.23

<sup>a</sup>Results obtained employing the CDF pseudopotential on the O atom only, with the corresponding aug-cc-pVXZ basis set being used on the H atoms. <sup>b</sup>Results obtained employing the CDF pseudopotentials on all atoms.

0.03 au lower with the aug-cc-pVDZ-CDF basis set than when using the augmented BFD basis set for the O atom in the representation of the trial function. With the aug-cc-pVTZ-CDF basis set, the variance is further reduced to 0.29 au, which is very close to the value obtained with the plane-wave basis set with the 160 au cutoff. For the largest GTO basis set considered, aug-cc-pVSZ-CDF, the variance is only 0.22 au. As for the all-electron calculations, the results for the larger GTO basis sets were obtained without the *g* and higher angular momentum functions from the O basis set and the *f* and higher angular momentum functions from the H basis set. However, given the fact that the energies and variances obtained with the aug-cc-pVTZ-CDF and aug-cc-pVQZ-CDF and aug-cc-pVSZ-CDF basis sets are very close to those obtained with large cutoff plane-wave calculations, we conclude that the higher angular momentum functions are relatively unimportant for the VMC calculations.

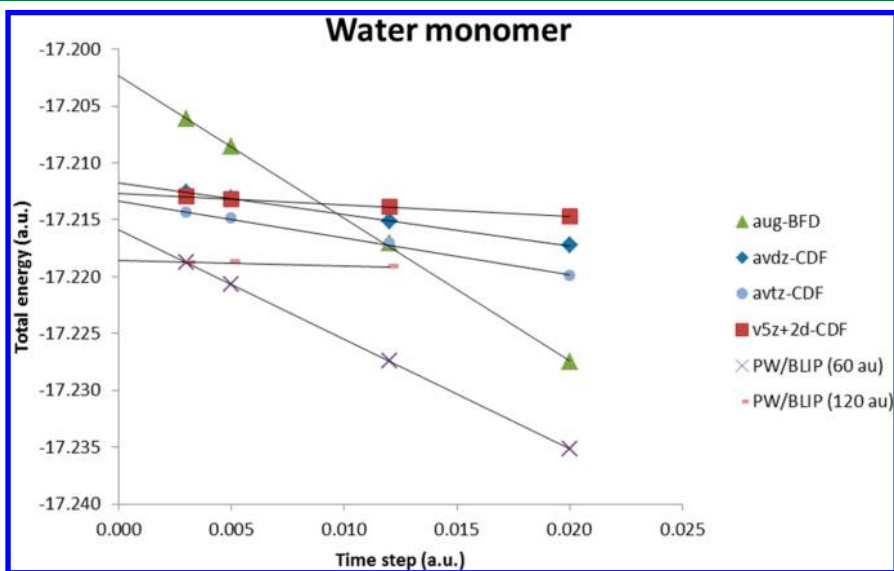
Table 3 also includes results obtained for two modified versions of the cc-pVSZ-CDF basis set denoted cc-pVSZ-(*sp*)+2*d*-CDF and cc-pVSZ(*sp*)+2*df*-CDF, which differ from the full cc-pVSZ-CDF basis set by employing smaller sets of polarization functions, taken from the cc-pVTZ-CDF basis set. (The cc-pVSZ-CDF basis set is derived from aug-cc-pVSZ-CDF by deleting the most diffuse function of each angular

momentum.) The corresponding basis sets for hydrogen differ from the full cc-pVSZ-CDF basis set by using as polarization functions just the 2*p* and 2*p1d* functions from the cc-pVTZ-CDF basis set. Of particular interest is that neither the VMC energy nor the variance is significantly impacted by the inclusion of the *f* function in the oxygen basis set and the *d* function in the hydrogen basis set. We consider these basis sets in more detail later in the manuscript.

### B. DMC Calculations: Water Monomer and Dimer.

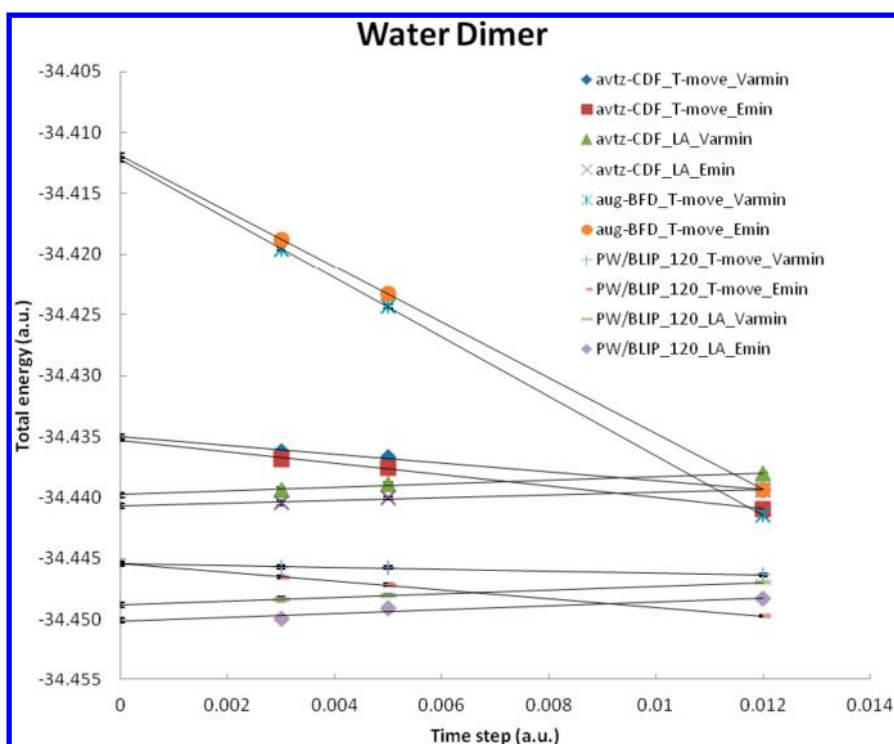
Figure 1 reports the results of DMC calculations on the water monomer with trial functions expanded in terms of the augmented BFD, the aug-cc-pVDZ-CDF, and aug-cc-pVTZ-CDF GTO basis sets, as well as in terms of the 60 and 120 au cutoff plane-wave basis sets. The DMC calculations were run using 10 000 walkers, about 35 000 Monte Carlo steps, and for time steps of 0.003, 0.005, 0.012, and 0.02 au, using the T-move procedure.<sup>31,32</sup>

Several trends are apparent from the data in this figure. First, in each set of calculations the energies from the various time steps are well represented by linear fits, facilitating extrapolation to zero time step. Second, DMC calculations using trial functions with the largest variances, namely, those expanded in terms of the augmented BFD and the 60 au cutoff plane-wave basis sets, display the steepest slopes. Third, the DMC calculations with different trial functions give different total energies in the  $t \rightarrow 0$  limit. The trial function using the augmented BFD basis set gives the highest DMC energy, -17.20207(9) au, and that represented in terms of the plane-wave basis set with 120 au cutoff gives the lowest DMC energy, -17.2186(2). The calculations using the aug-cc-pVDZ-CDF and aug-cc-pVTZ-CDF basis sets give extrapolated DMC energies of -17.2117(1) and -17.21341(9), respectively. It should be noted that most of the difference in the DMC energies obtained using trial functions expanded in terms of the 120 au plane wave and aug-cc-pVTZ-CDF basis sets is not due to the use of different orbitals (LDA vs HF) in the two calculations as the DMC energies of H<sub>2</sub>O calculated using the aug-cc-pVTZ-CDF basis set with HF and LDA orbitals agree to within 0.0005 au. We note also that the slope in the DMC

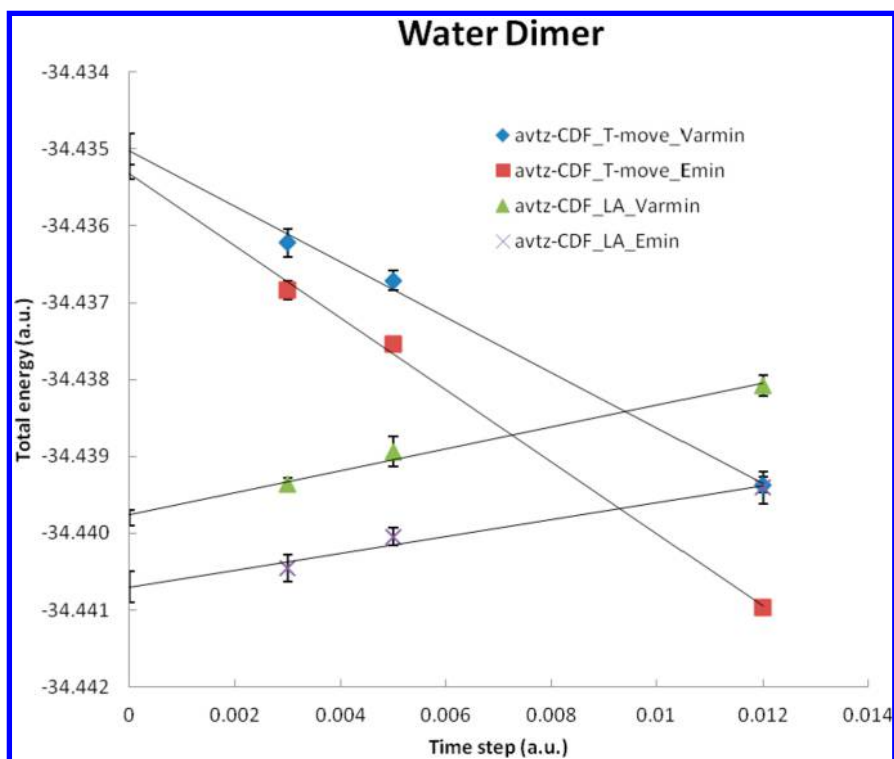


**Figure 1.** DMC energy of the water molecule with trial functions expanded in different basis sets. Results obtained using the CASINO Dirac–Fock pseudopotential and the T-move procedure. avdz-CDF, avtz-CDF, and v5z+2d-CDF refer to aug-cc-pVDZ-CDF, aug-cc-pVTZ-CDF, and cc-pVSZ(*sp*)+2*d*-CDF, respectively.





**Figure 2.** DMC energies of the water dimer obtained using different basis sets for representing the orbitals in the trial function, two choices of the Jastrow factors, and two strategies for dealing with nonlocality of the pseudopotentials.



**Figure 3.** Energies of the water dimer on an expanded scale, from DMC calculations using trial functions represented in terms of the aug-cc-pVTZ-CDF basis set.

energy vs time step curve is reduced by about a factor of 2 in going from the cc-pVTZ-CDF to the cc-pV5Z(*sp*)+2d-CDF basis set. In other words, the time step error is significantly reduced by use of a trial function represented by a basis set with a large number of *s* and *p* primitive functions. The major reason

the latter calculation gives an energy about 0.006 au above the result obtained using the trial function expanded in the 120 au cutoff plane-wave basis set appears to be due to the omission of the diffuse “aug” functions in the GTO basis set.

In DMC calculations, the use of a nonlocal pseudopotential is incompatible with the fixed-node boundary condition. In this study we examine the sensitivity of the total energies to the strategy, T-move<sup>32</sup> or locality approximation (LA),<sup>33</sup> used to deal with this problem. Each scheme has its advantages and disadvantages. The LA method is believed to have smaller time step bias but to have more stability problems,<sup>33</sup> while the T-move procedure has the advantage of being variational and is generally more stable, but requires smaller time steps, especially for large systems.<sup>34,35</sup> In addition, we examine the sensitivity of the DMC energies to the choice of Jastrow factor. Two different choices of the Jastrow factors are used, one from variance minimizations (Varmin) and the other from energy minimizations (Emin). Although in all-electron calculations DMC energies should not depend on the choice of the Jastrow factor, there can be a small sensitivity of the energy to the Jastrow factor in pseudopotential based DMC calculations.

Figure 2 reports the DMC total energies of water dimer obtained using different basis sets and the two approaches for dealing with the nonlocal pseudopotential problem. The geometry used for the water dimer was obtained by optimization at the MP2/aug-cc-pV5Z level. Figure 3 shows on an expanded scale the DMC results obtained using the aug-cc-pVTZ-CDF basis set. Most significantly, it is seen that the DMC energy is more sensitive to the choice of Jastrow factor in the LA approach than in the T-move approach. With the LA procedure, when used with the aug-cc-pVTZ-CDF basis set, the two choices of Jastrow factors lead to a difference of 0.5 kcal/mol in the total DMC energies in the zero time step limit. The difference is slightly greater when using the trial function represented in the plane-wave basis set. With the T-move scheme, the  $t \rightarrow 0$  DMC energies obtained using the two Jastrow factors agree to within the error bars, even though the energy differences are significant at nonzero time steps. We were unable to perform a stable DMC calculation with the LA approach using the trial function represented in terms of the augmented BFD basis set, although such calculations ran smoothly with the aug-cc-pVTZ-CDF basis set. The better convergence of the T-move procedure for dealing with the nonlocality of the pseudopotentials was also noted by Gurtubay and Needs<sup>36</sup> in their quantum Monte Carlo study of the water monomer and dimer.

Of course, the performance of the various calculations for predicting energy differences is of more chemical interest than the total energies. For this reason we have also calculated the binding energy of the water dimer and of the water–benzene complex. In the calculations of the binding energy of the water dimer, the geometries of the water monomer and dimer were taken from MP2/aug-cc-pV5Z level optimizations. The binding energy of the water dimer for each method was calculated by subtracting twice the energy of the monomer from the energy of the dimer, with the results being reported in Table 4. The water–benzene calculations will be discussed in Section 4C.

The binding energies of water dimer obtained from the various DMC calculations agree to within the error bars. They are also in good agreement with the binding energies from prior all-electron and pseudopotential DMC calculations,<sup>36,37</sup> as well as from complete-basis-set limit CCSD(T)<sup>23</sup> calculations.<sup>38</sup> Somewhat surprisingly, even though the trial functions employing orbitals expanded in terms of the augmented BFD basis set have much larger variances in the VMC step and much larger time step biases in the DMC step, they give, to within the statistical errors, binding energies in agreement with those from

Table 4. Calculated Binding Energy of Water Dimer<sup>a</sup>

methods	binding energy (kcal/mol)
DMC/avtz_T-move_Varmin	−5.15 ± 0.18
DMC/avtz_T-move_Emin	−5.06 ± 0.08
DMC/avtz_LA_Varmin	−5.23 ± 0.15
DMC/avtz_LA_Emin	−5.21 ± 0.15
DMC/BFD_T-move_Varmin	−5.00 ± 0.15
DMC/BFD_T-move_Emin	−5.06 ± 0.15
DMC/PL/BLIP_120_T-move_Varmin	−5.15 ± 0.18
DMC/PL/BLIP_120_T-move_Emin	−5.16 ± 0.09
DMC/PL/BLIP_120_LA_Varmin	−5.16 ± 0.18
DMC/PL/BLIP_120_LA_Emin	−5.03 ± 0.14
DMC/HF <sup>28,b</sup>	−5.02 ± 0.18
DMC/B3LYP <sup>28,b</sup>	−5.21 ± 0.18
CCSD(T) CBS limit <sup>30</sup>	−5.02 ± 0.05
DMC/B3LYP <sup>27,c</sup>	−5.03 ± 0.07
DMC/B3LYP <sup>27,d</sup>	−5.07 ± 0.07

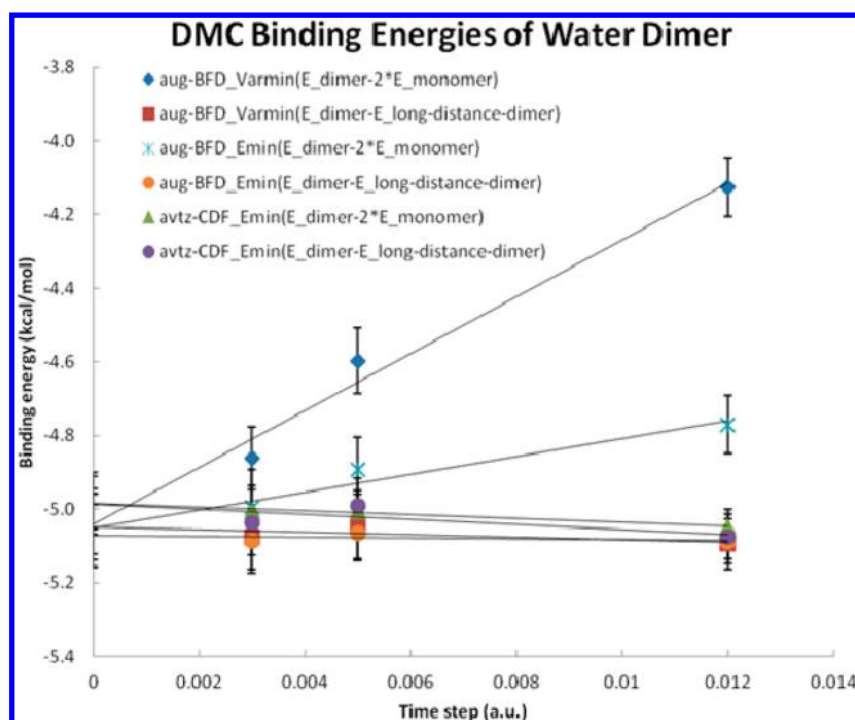
<sup>a</sup>Results obtained by subtracting twice the DMC energy of the monomer from the DMC energy of the dimer. <sup>b</sup>From all-electron calculations. <sup>c</sup>From pseudopotential calculations using the locality approximation. <sup>d</sup>From pseudopotential calculations using the T-move procedure.

the other calculations. However, it is not clear whether this would also be the case for more complex systems for which use of trial functions with large variance may prove more problematical for DMC energy differences.

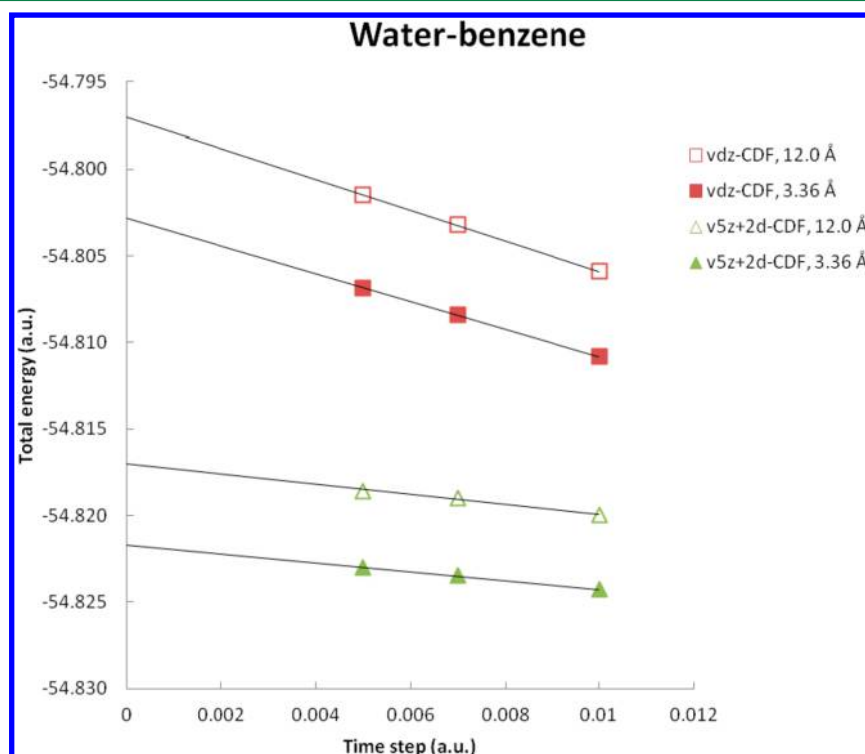
Up to this point all dimer binding energies were calculated by subtracting the energy of the water dimer from the sum of energies of two isolated water monomers (strategy S1). We also considered an alternative approach (designated S2), where the energy of the water dimer at large separation is used in place of the sum of the energies of the two monomers. (The S2 approach was employed previously by Ma and co-workers in their study of water–benzene.<sup>39</sup>) In the S1 method the energies of the individual species are extrapolated to the zero time step limit, and these extrapolated results are used to calculate the zero-time-step binding energies. In the S2 method, the zero time step binding energies were obtained by extrapolating the binding energies at different time steps. Here, we are interested in determining whether the errors due to the finite time step bias largely cancel in the S2 strategy.

As shown in Figure 4, the extrapolated zero-time-step energies obtained using different trial functions and strategies for each basis set are essentially identical within the error bars. With the S1 procedure, there is a large time step bias for the binding energy with the aug-BFD basis set, while the time step bias is much smaller when using the aug-cc-pVTZ-CDF basis set. However, when using the S2 procedure to calculate the binding energy, most of the bias due to the use of finite time steps is removed when using the aug-BFD basis set together with the CASINO Dirac–Fock pseudopotential. This is most encouraging because it indicates that with the S2 procedure one can use longer time steps and, perhaps also, that only one or two time steps may suffice, depending on the system of interest, greatly reducing the computational effort.

**C. DMC Results for the Water–Benzene Complex.** As a second test system, we have calculated the binding energy of the water–benzene complex using the geometry of ref 40. The binding energy of the complex was calculated using as the reference a water–benzene dimer with a 12 Å separation between the O atom of the water molecule and the center of



**Figure 4.** Binding energy of the water dimer calculated using the DMC method with different trial functions and strategies for calculating the reference energy of two monomers. All results obtained using the T-move procedure.



**Figure 5.** DMC energies of the benzene–water complex obtained using different basis sets for representing the orbitals in the trial functions. vdz-CDF and v5z+2d-CDF refer to cc-pVDZ-CDF and cc-pV5Z(*sp*)+2d-CDF basis sets, respectively. All results were obtained using the T-move procedure.

the benzene ring, and using the S2 strategy described in Section 3. The DMC calculations were run using 30 000 walkers, about 90 000 Monte Carlo steps, and for time steps of 0.005, 0.007, and 0.01 au, using the T-move procedure. The CASINO Dirac–Fock pseudopotentials were used on all atoms. The trial

functions were expanded in terms of HF orbitals calculated with the cc-pVDZ-CDF and the cc-pV5Z(*sp*)+2d-CDF basis sets described in Section 2. The DMC energies for the short and long distances are reported in Figure 5 from which it is seen that the time-step error is significantly reduced in going from

the cc-pVDZ-CDF to the cc-pVSZ(*sp*)+2*d*-CDF basis set for the expansion of the orbitals in the trial function. The DMC values of the binding energy obtained with the cc-pVDZ-CDF and cc-pVSZ(*sp*)+2*d*-CDF basis sets are  $-3.6(1)$  and  $-2.9(2)$  kcal/mol, respectively, with the latter value being in good agreement with the  $-3.2$  kcal/mol CCSD(T)-F12<sup>41</sup> result of ref 38. For comparison, we note that Ma and co-workers obtained a binding energy  $-2.9$  kcal/mol using the DMC method together with a LDA trial function expanded in terms of plane/wave/BLIP functions but employing a geometry slightly different from that used here.<sup>39</sup>

## 5. CONCLUSIONS

In this study, we reported correlation consistent basis sets for H and B–Ne for use with Casino Dirac–Fock pseudopotentials and showed that these perform well in coupled cluster calculations. The H, O, and C basis sets were tested in quantum Monte Carlo calculations on H<sub>2</sub>O, (H<sub>2</sub>O)<sub>2</sub>, and water–benzene. Although it is common practice in traditional quantum chemistry calculations employing pseudopotentials to use basis sets with relatively small numbers of primitive GTOs, our results indicate that to reduce time step errors and to improve convergence of DMC calculations, it is desirable to use large contracted sets of primitive *s* and *p* functions. We also found that the DMC energies obtained from the T-move procedure are less sensitive to the choice of Jastrow factor than are the corresponding values from the locality approximation. We note that the DMC method has recently been applied to obtain accurate interaction energies of water clusters up to the hexamer.<sup>42</sup> The availability of the aug-cc-pVXZ-CDF basis sets reported in the present study should facilitate quantum Monte Carlo calculations on significantly larger systems.

## ■ ASSOCIATED CONTENT

### Supporting Information

The basis sets and the procedure used for their optimization are reported. This information is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [jordan@pitt.edu](mailto:jordan@pitt.edu).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research was carried out with the support of the National Science Foundation (K.D.J.). We thank Professor Dario Alfè for helpful discussions. The calculations were carried out on computers in the University of Pittsburgh's Center for Simulation and Modeling.

## ■ REFERENCES

- (1) Anderson, J. B. Random-walk simulation of the Schrödinger equation. Hydrogen ion (H<sub>3</sub><sup>+</sup>). *J. Chem. Phys.* **1975**, *63*, 1499–1503.
- (2) Anderson, J. B. Quantum chemistry by random walk. H <sup>2</sup>P, H<sub>3</sub><sup>+</sup> D<sub>3h</sub> <sup>1</sup>A<sub>1</sub>, H<sub>2</sub> <sup>3</sup>Σ<sub>g</sub><sup>+</sup>, H<sub>4</sub> <sup>1</sup>Σ<sub>g</sub><sup>+</sup>, Be <sup>1</sup>S. *J. Chem. Phys.* **1976**, *65*, 4121–4127.
- (3) Drummond, N. D.; Towler, M. D.; Needs, R. J. Jastrow correlation factor for atoms, molecules, and solids. *Phys. Rev. B* **2004**, *70*, 235119–235129.
- (4) Xu, J.; Al-Saidi, W. A.; Jordan, K. D. Unpublished results.
- (5) Xu, J.; Jordan, K. D. Application of the diffusion Monte Carlo method to the binding of excess electrons to water clusters. *J. Phys. Chem. A* **2010**, *114*, 1364–1366.
- (6) Trail, J. R.; Needs, R. J. Norm-conserving Hartree-Fock pseudopotentials and their asymptotic behaviour. *J. Chem. Phys.* **2005**, *122*, 014112.
- (7) Trail, J. R.; Needs, R. J. Smooth relativistic Hartree-Fock pseudopotentials for H to Ba and Lu to Hg. *J. Chem. Phys.* **2005**, *122*, 174109.
- (8) Burkatzki, M.; Filippi, C.; Dolg, M. Energy-consistent pseudopotentials for quantum Monte Carlo calculations. *J. Chem. Phys.* **2007**, *126*, 234105.
- (9) Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007–1023.
- (10) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (11) Benedict, W. S.; Gailar, N.; Plyler, E. K. Rotation vibration spectra of deuterated water vapor. *J. Chem. Phys.* **1956**, *24*, 1139.
- (12) Alfè, D.; Gillan, M. J. An efficient localized basis set for quantum Monte Carlo calculations on condensed matter. *Phys. Rev. B* **2004**, *70*, 161101.
- (13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2003.
- (14) Gonze, X.; Amadon, B.; Anglade, P. M.; Beuken, J. M.; Bottin, F.; Boulanger, P.; Bruneval, F.; Caliste, D.; Caracas, R.; Deutsch, T.; Genovese, L.; Ghosez, P.; Giantomassi, M.; Goedecker, S.; Hamann, D. R.; Hermet, P.; Jollet, F.; Jomard, G.; Leroux, S.; Mancini, M.; Mazevet, S.; Oliveira, M. J. T.; Onida, G.; Pouillon, Y.; Rangel, T.; Rignanese, G. M.; Sangalli, D.; Shaltaf, R.; Torrent, M.; Verstraete, M. J.; Zerah, G.; Zwanziger, J. W. ABINIT: First-principles approach to material and nanosystem properties. *Comput. Phys. Commun.* **2009**, *180*, 2582–2615.
- (15) Needs, R. J.; Towler, M. D.; Drummond, N. D.; Rios, P. L. Continuum variational and diffusion quantum Monte Carlo calculations. *J. Phys.: Condens. Matter* **2010**, *22*, 023201.
- (16) Alfè, D.; Alfredsson, M.; Brodbolt, J.; Gillan, M. J.; Towler, M. D.; Needs, R. J. *Phys. Rev. B* **2005**, *72*, 0124114.
- (17) Peterson, K. A. Systematically convergent basis sets with relativistic pseudopotentials. I. Correlation consistent basis sets for the post-d group 13–15 elements. *J. Chem. Phys.* **2003**, *119*, 11099–11112.
- (18) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16–18 elements. *J. Chem. Phys.* **2003**, *119*, 11113–11123.
- (19) Peterson, K. A.; Shepler, B. C.; Figgen, D.; Stoll, H. On the spectroscopic and thermochemical properties of ClO, BrO, IO, and their anions. *J. Phys. Chem. A* **2006**, *110*, 13877–13883.
- (20) Peterson, K. A.; Figgen, D.; Dolg, M.; Stoll, H. Energy-consistent relativistic pseudopotentials and correlation consistent basis sets for the 4d elements Y–Pd. *J. Chem. Phys.* **2007**, *126*, 124101.



- (21) Figgen, D.; Peterson, K. A.; Dolg, M.; Stoll, H. Energy-consistent pseudopotentials and correlation consistent basis sets for the 5d elements Hf–Pt. *J. Chem. Phys.* **2009**, *130*, 164108.
- (22) Hampel, C.; Peterson, K. A.; Werner, H.-J. A comparison of the efficiency and accuracy of the quadratic configuration interaction (QCISD), coupled cluster (CCSD), and Brueckner coupled cluster (BCCD) methods. *Chem. Phys. Lett.* **1992**, *190*, 1–12.
- (23) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- (24) Douglas, M.; Kroll, N. M. Quantum electrodynamical corrections to the fine structure of helium. *Ann. Phys.* **1974**, *82*, 89–155. Hess, B. A. Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. *Phys. Rev. A* **1986**, *33*, 3742–3748.
- (25) de Jong, W. A.; Harrison, R. J.; Dixon, D. A. Parallel Douglas–Kroll energy and gradients in NW Chem: Estimating scalar relativistic effects using Douglas–Kroll contracted basis sets. *J. Chem. Phys.* **2001**, *114*, 48–53.
- (26) Knowles, P. J.; Hampel, C.; Werner, H.-J. Coupled cluster theory for high spin, open shell reference wave functions. *J. Chem. Phys.* **1993**, *99*, 5219–5227.
- (27) Scuseria, G. E. The open-shell restricted Hartree-Fock singles and doubles coupled-cluster method including triple excitations CCSD (T): Application to  $C_3^+$ . *Chem. Phys. Lett.* **1991**, *176*, 27–35.
- (28) Huber, K. P.; Herzberg, G. *Molecular Spectra and Molecular Structure IV. Constants of Diatomic Molecules*; Van Nostrand: Princeton, 1979.
- (29) Urdahl, R. S.; Bao, Y.; Jackson, W. M. An experimental determination of the heat of formation of  $C_2$  and the C–H bond dissociation energy in  $C_2H$ . *Chem. Phys. Lett.* **1991**, *178*, 425–428.
- (30) Feller, D.; Peterson, K. A.; Dixon, D. A. A survey of factors contributing to accurate theoretical predictions of atomization energies and molecular structures. *J. Chem. Phys.* **2008**, *129*, 204105.
- (31) Casula, M.; Filippi, C.; Sorella, S. Diffusion Monte Carlo method with lattice regularization. *Phys. Rev. Lett.* **2005**, *95*, 100201.
- (32) Casula, M. Beyond the locality approximation in the standard diffusion Monte Carlo method. *Phys. Rev. B* **2006**, *74*, 161102.
- (33) Mitas, L.; Shirley, E. L.; Ceperley, D. M. Nonlocal pseudopotentials and diffusion Monte Carlo. *J. Chem. Phys.* **1991**, *95*, 3467–3475.
- (34) Needs, R. J.; Towler, M. D.; Drummond, N. D.; Ros, P. L. Continuum variational and diffusion quantum Monte Carlo calculations. *J. Phys.: Condens. Matter* **2010**, *22*, 023201.
- (35) Casula, M.; Moroni, S.; Sorella, S.; Filippi, C. Size-consistent variational approaches to nonlocal pseudopotentials: Standard and lattice regularized diffusion Monte Carlo methods revisited. *J. Chem. Phys.* **2010**, *132*, 154113.
- (36) Gurtubay, I. G.; Needs, R. J. Dissociation energy of the water dimer from quantum Monte Carlo calculations. *J. Chem. Phys.* **2007**, *127*, 124306.
- (37) Benedek, N. A.; Snook, I. K.; Towler, M. D.; Needs, R. J. Quantum Monte Carlo calculations of the dissociation energy of the water dimer. *J. Chem. Phys.* **2006**, *125*, 104302.
- (38) Klopper, W.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. Computational determination of equilibrium geometry and dissociation energy of the water dimer. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2227.
- (39) Ma, J.; Alfè, D.; Michaelides, A.; Wang, E. The water-benzene interaction: insight from electronic structure theories. *J. Chem. Phys.* **2009**, *130*, 154303.
- (40) Jenness, G. R.; Karalti, O.; Al-Saidi, W. A.; Jordan, K. D. Evaluation of theoretical approaches for describing the interaction of water with linear acenes. *J. Phys. Chem. A* **2011**, *115*, 5955–5964.
- (41) Knizia, G.; Adler, T. B.; Werner, H.-J. Simplified CCSD(T)-F12 methods: Theory and benchmarks. *J. Chem. Phys.* **2009**, *130*, 054104.
- (42) Gillan, J. J.; Manby, F. R.; Towler, M. D.; Alfè, D. Assessing the accuracy of quantum Monte Carlo and density functional theory for energetic of small water clusters. *J. Chem. Phys.* **2012**, *136*, 244105.