Theory and Applications of Quantum Monte Carlo

by

Michael John Deible

B.Sc.Ed. in Chemistry Education, Indiana University of Pennsylvania, 2008

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Physical Chemistry

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

THE DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Michael John Deible

It was defended on

29 May 2015

and approved by

Daniel Lambrecht, Assistant Professor, Department of Chemistry

Alexander Star, Associate Professor, Department of Chemistry

John Keith, Assistant Professor, Department of Chemical and Petroleum Engineering

Dissertation Advisor: Kenneth D. Jordan, Richard King Mellon Professor and Distinguished Professor of Computational Chemistry, Department of Chemistry

Theory and Applications of Quantum Monte Carlo

Michael J. Deible, PhD

University of Pittsburgh, 2015

With the development of peta-scale computers and exa-scale only a few years away, the quantum Monte Carlo (QMC) method, with favorable scaling and inherent parrallelizability, is poised to increase its impact on the electronic structure community. The most widely used variation of QMC is the diffusion Monte Carlo (DMC) method. The accuracy of the DMC method is only limited by the trial wave function that it employs. The effect of the trial wave function is studied here by initially developing correlation-consistent Gaussian basis sets for use in DMC calculations. These basis sets give a low variance in variance Monte Carlo calculations and improved convergence in DMC. The orbital type used in the trial wave function is then investigated, and it is shown that Brueckner orbitals result in a DMC energy comparable to a DMC energy with orbitals from density functional theory and significantly lower than orbitals from Hartree-Fock theory. Three large weakly interacting systems are then studied; a water-16 isomer, a methane clathrate, and a carbon dioxide clathrate. The DMC method is seen to be in good agreement with MP2 calculations and provides reliable benchmarks. Several strongly correlated systems are then studied. An H_4 model system that allows for a fine tuning of the multi-configurational character of the wave function shows when the accuracy of the DMC method with a single Slater-determinant trial function begins to deviate from multi-reference benchmarks. The weakly interacting face-to-face ethylene dimer is studied with and without a rotation around the π bond, which is used to increase the multi-configurational nature of the

wave function. This test shows that the effect of a multi-configurational wave function in weakly interacting systems causes DMC with a single Slater-determinant to be unable to achieve sub-chemical accuracy. The beryllium dimer is studied, and it is shown that a very large determinant expansion is required for DMC to predict a binding energy that is in close agreement with experiment. Finally, water interacting with increasingly large acenes is studied, as is the benzene and anthracene dimer. Deviations from benchmarks are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I wish to thank Professor Jordan for his unparalleled scientific guidance and patience. I hope to be able to carry on the values and commitment to success that I have learned from him. I would like to thank my committee members for their oversight. I would like to thank the SAM team for continued support in my computational endeavors. I would also like to thank the present and former Jordan group members for valuable discussions and guidance.

And of course, my deepest gratitude goes to Stephanie and Michael. This has all been done because of her, and its all been done for him.

1.0 INTRODUCTION

A fundamental problem in modern quantum mechanics is the inability to exactly solve the Schrödinger equation for a system with more than one electron. This has not limited the practical uses for *ab initio* methods, however, and development of new approximations along with the continued increase in computing power has made quantum mechanical applications routine.

One of the earliest methods developed to approximately solve the Schrödinger equation that is still in wide use today is the Hartree-Fock (HF) method. This basic but powerful approach leads to the development of molecular orbital theory that is familiar to every chemist. The HF method removes the problem of solving the Schrödinger equation for more than one electron by treating each electron in a field of the remaining electrons.[1–3] Thus, the solutions to the approximation are said to be uncorrelated, as each electron is moving independently of the other electrons. (Technically, the determinant introduces correlation of same-spin electrons, but the opposite spin electrons remain uncorrelated.) This lack of electron correlation results in a small fraction of the overall energy of a system but is incredibly important for practical applications. Capturing the remaining energy missing in HF theory has spurred the development of a wide array of post HF methods, where the HF result is used as a starting point. Configuration interaction (CI), Møller-Plesset second order perturbation theory (MP2), and Coupled Cluster theory (CC) are a few of the most common methods used to recover correlation energy missing

in the HF solution. Post HF methods are not without their own flaws, however. For example, the gold standard in quantum chemistry is the coupled cluster with singles, doubles, and perturbative triples (CCSD(T)), which formally scales as N^7 (where N is the number of electrons), which means that it can only be used for very small systems. Additionally, HF and most of the post-HF methods require a basis set, which can lead to basis set superposition error (BSSE) and limits the overall accuracy of the method.

An alternative to HF based methods is density functional theory (DFT).[4] This approach recasts the problem of solving the Schrödinger equation for many electrons into solving for the property of interest as a functional of the electron density. DFT has been widely used, particularly over the past quarter century, as functionals have improved and high quality results can be obtained for a wide range of systems. DFT offers several advantages over HF theory. The most important advantage is that DFT includes correlation effects in the calculation of the energy. The practical implementation of DFT rests with the exchange-correlation functional; in principle, an exact functional exists,[5] but in practice it is unknown and must be parameterized. Libraries of advanced functionals have been developed, and each functional can provide highly accurate results for particular systems.

HF and DFT both have successes and limitations. One major limitation of these methods are their inability to accurately calculate van der Waals interactions, which are dominant in weakly interacting systems such as water clusters. For DFT, remedies for this problem have been proposed in the form of dispersion corrections.[6,7] Various schemes for correcting DFT for dispersion are available but are not highly accurate for a range of systems. The post-HF methods described above can accurately predict van der Waals interactions for many systems, but the limitations of the basis set and high computational cost remain a road block to simulations on

systems with hundreds of electrons. An additional flaw for these methods is that they are incapable of accurately representing systems that have a degenerate ground state. Post HF methods like multi-configurational self-consistent field theory and configuration interaction are capable of building on the HF wavefunction to include the effects of degenerate ground states, but are computationally demanding and can only be applied to systems of several tens of electrons. DFT functionals that correctly represent simple degenerate systems have been proposed,[8] but their applicability to a wider range of systems remains unclear. It is evident that studying weakly interacting systems and systems with degenerate ground states is very challenging and requires an alternative approach to HF or DFT.

The diffusion Monte Carlo method[9] (DMC) has many advantages over DFT and HF. The formal scaling of N^3 means that it can be applied to very large systems, and DMC has been used to study systems as large as hundreds of atoms. The inherent parallelizability makes DMC methods uniquely suited to modern supercomputers, which are built with increasing numbers of processors. DMC is also less sensitive to the basis set than the traditional approaches, and is free of basis set superposition error.[10] It has been shown that DMC can achieve accuracy similar to CCSD(T)[11,12] and complete basis set limit MP2[13] for weakly interacting systems with non-degenerate ground states. An additional advantage of DMC over HF and DFT is that DMC is capable of using more than one Slater-determinant to represent the ground state, which means that it can be used to calculate the energy of multi-configurational systems.

Like HF and DFT, DMC is not without its flaws. As is discussed in section 1.4, the only uncontrolled approximation in DMC is the fixed-node error (FNA).[14] There are several approaches to improving the FNA.[15] The basis set and the type of orbitals used in the trial wave function can determine the accuracy of the FNA.[16] The use of a single determinant may not be

adequate in many cases.[17] Improving the nodes of the wave function by adding more determinants is a difficult task.[18] Generating a trial wave function with enough determinants to capture the multi-configurational nature of the ground state can be demanding for large systems. Evaluation of the derivative of the wave function with many determinants is a formidable task for DMC, and although algorithms[19] have been developed to expedite this process, it remains computationally intensive. Therefore, in general, all of the determinants from the trial wave function cannot be used, and a selection criteria must be applied to the determinants to make a trial wave function. This can be based on excitations or can be applied as a cutoff threshold to the CI coefficients. It is not always clear which, if any, method is the best for selecting determinants, and so the application of DMC to multi-configurational systems remains challenging.

DMC clearly has advantages over traditional quantum chemistry methods that have made it more popular in recent years, but problems remain. This thesis is dedicated to showing the advantages of DMC and elucidating the challenges. The rest of this thesis is structured as follows: the remaining sections of the introduction will give an overview of the methods mentioned above, namely HF, DFT, and quantum Monte Carlo, including variational and diffusion Monte Carlo. Chapter two will evaluate the effect of the trial wave function on QMC calculations. Chapter 3 will apply DMC calculations to studies of large weakly interacting systems. Chapter 4 will highlight several successes for multi-configurational systems. Chapters 5 and 6 will address a current area of difficulty; weakly interacting systems with a moderately degenerate ground state. Chapter 7 will present the conclusions.

1.1 ATOMIC ORBITALS AND SLATER DETERMINANTS

1.1.1 Atomic orbitals

For a spherical one-electron system, the wave function for an atomic orbital can be defined as a product of a radial and an angular function:

$$\psi_{nlm} = R_{nl}(r) Y_{lm}(\theta\phi) \tag{1.1}$$

where n, l, and m are the principle, azimuthal, and magnetic quantum numbers, respectively, and r, θ, and ϕ take their standard polar coordinate definitions. The angular portion of the wave function is a product of a function of θ, and ϕ:

$$\Phi_m(\varphi) = \frac{1}{\sqrt{(2\pi)}} \exp(i m \phi) \tag{1.2}$$

$$\Theta_{lm}(\theta) = \left[\frac{(2l+1)}{2} \frac{(l-|m|)!}{(l+|m|)!} \right]^{\frac{1}{2}} P_l^{|m|}(\cos\theta) \tag{1.3}$$

Equation 1.2 are the solutions to the Schrödinger equation for a particle on a ring. The term in square brackets in equation 1.3 is a normalizing factor and the $P_l^{|m|}(\cos\theta)$ term is a Legendre polynomial. The radial function is given by:

$$R_{nl}(r) = -\left[\left(\frac{2Z}{na_0} \right)^3 \frac{(n-l-1)!}{2n[(n+l)!]^3} \right]^{\frac{1}{2}} \exp\left(\frac{-\rho}{2} \right) \rho^l L_{n+1}^{2l+1}(\rho) \tag{1.4}$$

where the term in brackets is a normalizing factor and second term can be simplified by writing it in terms of the orbital exponent $\zeta = Z/n$. This simplifies the radial term to what is typically referred to as a Slater-type orbital (STO):

$$R_{nl}(r) = (2\zeta)^{n+\frac{1}{2}} [(2n)!]^{\frac{-1}{2}} r^{n-1} e^{-\zeta r} \tag{1.5}$$

Evaluating the atomic orbitals defined in equation 1.1 is straightforward when when the atomic orbitals are located on the same atom. However, if the orbitals are located on different nuclei, evaluating the radial portion defined in equation 1.5 can be very difficult. It is common to alleviate this problem by replacing the STO by a linear combination of Gaussian-type orbitals (GTO):

$$\psi = \sum_{i=1}^{L} d_i \left[x^a y^b z^c \exp\left(-\alpha r^2\right) \right] \tag{1.6}$$

where d_i is the contraction coefficient, and L is the length of the contraction, and a, b, and c are used to determine the order of the function and replace the angular portion from above.

The replacement of the Slater-type orbital with a Gaussian-type orbital makes the calculation more efficient, but introduces several limitations. First, the different behavior of each function at the origin means that there is no nuclear cusp when using a GTO, which will present a problem for quantum Monte Carlo simulations (section 1.4).[20] Second, the tail of the GTO wave function decays to zero much more quickly than a STO.

The functions used to represent the one electron orbitals of 1.5 and 1.6 are commonly referred to as a basis set. In a many-electron system, the minimum number of basis functions will be equal to the total number of electrons. For greater accuracy, more functions can be added to the basis set. The exponents and contraction coefficients are optimized for each atom, or each pseudopotential for each atom. Development of Gaussian basis sets for use in quantum Monte Carlo calculations is the focus of section 2.1.1.

Gaussian-type orbitals are the most commonly used basis sets in *ab initio* calculations of gas phase molecules. In periodic systems, the wave function takes a plane-wave form:

$$\psi_i^k = \sum_G a_{i,k+G} \exp\left(i\left(k+G\right)\cdot r\right) \tag{1.7}$$

where **G** is a reciprocal lattice vector, **k** is the wavevector, *a* are the orbital coefficients, and r is a positional vector. The plane-wave basis set offers the advantage of being systematically improved by increasing the kinetic energy cutoff $\left(\left(\hbar/2\mathrm{m}\right)|k+G|^2\right)$ to include more plane-waves. Additionally, the Fourier transform of the plane-waves makes them computationally efficient.

To completely describe the electron, the wavefunction must be a combination of the spatial component given above and a spin term. For an N electron system, the spatial term can be combined with a spin component, α or β, and a spin orbital can be defined as:

$$\chi_1(N)=\psi(N)\alpha(N)\text{and }\chi_2(N)=\psi(N)\beta(N) \tag{1.8}$$

In a many electron system, the molecular orbitals, represented by basis functions, can be recast as a linear combination of the atomic orbitals:

$$\psi=\sum_i^K c_i\psi_i \tag{1.9}$$

where the sum runs over the *K* basis functions and c_i are the atomic orbital coefficients.

1.1.2 The Slater Determinant

For an N electron system, the Hartree product is a wave function that is simply the product of each spin orbital for each electron:

$$\Psi^{HP}=\chi_i(1)\chi_i(2)\cdots\chi_i(N) \tag{1.10}$$

The square of the Hartree product gives the simultaneous probability of finding one electron in a region of space independent of the other electrons. The Hartree product violates the antisymmetry principle which requires each particle to be indistinguishable and that the wave function must change sign upon interchange of two electrons. This can be remedied by taking

the appropriate linear combination of permutations of spin orbitals. Consider a two-electron system with spin orbitals *i* and *j* and electrons 1 and 2. Clearly, the Hartree products, $\Psi_1^{HP} = \chi_i(1)\chi_j(2)$ and $\Psi_2^{HP} = \chi_i(2)\chi_j(1)$ are identical. However, the linear combination,

$$\Psi = \frac{1}{\sqrt{2}}\left[\chi_i(1)\chi_j(2) - \chi_j(1)\chi_i(2)\right] \tag{1.11}$$

ensures that the wave function changes sign upon interchange of the electrons. The $2^{-1/2}$ term is a normalizing factor, and the term in the square brackets is a Slater determinant of the matrix:

$$\Psi = \frac{1}{\sqrt{2}}\begin{bmatrix} \chi_i(1) & \chi_j(1) \\ \chi_i(2) & \chi_j(2) \end{bmatrix} \tag{1.12}$$

This form of the wavefunction ensures that the Pauli exclusion principle is enforced for any N-electron wavefunction.

1.2 HARTREE-FOCK THEORY

1.2.1 Self-Consistent Field

Hartree-Fock (HF) theory gives an approximate solution to the non-relativistic time-independent electronic Schrödinger equation for an *N* electron system by moving each electron in a potential created by the remaining electrons and the nuclei. The Hartree-Fock method calculates the energy of electron 1 through the equation

$$\left\{-\frac{1}{2}\nabla_1^2 - \sum_I \frac{Z_I}{r_{1\mathrm{I}}} + \sum_{j=1}^{N}\left[J_j(1) - K_j(1)\right]\right\}\chi_i(1) = \varepsilon_i\chi_i(1) \tag{1.13a}$$

$$J_j(1) = \int d\tau_2 \chi_j(2)\frac{1}{r_{12}}\chi_j(2) \tag{1.13b}$$

$$K_j(1)\chi_i(1)=\left[\int d\tau_2 \chi_j(2)\frac{1}{r_{12}}\chi_i(2)\right]\chi_j(1) \tag{1.13c}$$

where *i, j,* denotes an orbital, (1), (2) denotes an electron, *I* denotes a nucleus, and the sum over *j* is over all of the orbitals. The first term of equation 1.13a is the kinetic energy of electron 1, the second term is the interaction of electron 1 and nucleus I, the third term is the Coulombic interaction of an electron and the other electrons given by equation 1.13b, and the fourth term accounts for the exchange of electron 1 with the other electrons and is given in equation 1.13c. The term in braces on the left side of equation 1.13a is known as the Fock operator.

When the atomic orbitals are expressed as a linear combination of basis functions as outlined 1.1.1, the energy of the wave function can be calculated as a simple eigenvalue problem,

$$FC=SC\varepsilon \tag{1.14}$$

where F is the Fock matrix, C is the density matrix made of the coefficients of the orbitals (see equation 1.9), ε is the energy matrix, and S is the overlap matrix. This is known as the Roothan-Hall equation, and it gives a solution to many electron Schrödinger equation within the confines of Hartree-Fock theory. The solution is a wavefunction in the form of a Slater determinant, where the coefficient matrix has been optimized to produce the lowest energy. It is solved iteratively, with an initial guess supplied for the density matrix which is operated on by the Fock matrix such that an energy and a new density matrix is produced. This new value for C replaces the initial guess, and the equation is solved to self consistency.

In the absence of relativistic effects, the difference between the energy calculated in HF theory and the true ground state energy is known as the correlation energy. As mentioned, many post Hartree-Fock methods have been introduced to recover correlation energy and the two that are the most relevant to this thesis are discussed in greater detail in sections 1.2.2-1.2.3. In

general, the self consistent field (SCF) procedure described above scales as ~N^4, where N is the number of basis functions, with number of integrals being the most time consuming step.

1.2.2 Configuration Interaction

One method to add correlation to the Hartree-Fock solution is configuration interaction (CI), in which determinants with swapped occupied and virtual orbitals are added to the ground state wave function calculated with Hartree-Fock theory. In this method, the wave function takes the form:

$$\Psi = C_0\Psi_0 + C_1\Psi_1 + C_2\Psi_2 + \cdots \tag{1.15}$$

where C_n are expansion coefficients. Ψ_0 is the Hartree-Fock determinant, and Ψ_n are determinants that have one or more occupied orbitals swapped with an equal number of virtual orbitals. The Slater determinants will form an orthonormal set and so the overlap integrals become unity, and the CI wave function can be solved by standard diagonalization methods. If all of the possible combinations of electrons in all orbitals are added to the wave function, this method, then known as full CI, will give the exact energy for the system, limited only by the size of the basis set used to represent the atomic orbitals. For N electrons and K orbitals, the total number of determinants that can be created is $(2K!)/[N!(2K-N)!]$. Clearly, this problem is intractable for large values of N or K. The number of required determinants can be reduced by considering symmetry constraints for the wave function, but it still remains large for all but the smallest systems. Another method to reduce the number of determinants is to limit the excitations, ie only allowing single or single and double excitations. While this effectively reduces the number of determinants, it has the drawback of not being size-consistent.

1.2.3 Multi-Configuration Self Consistent Field

In the CI treatment discussed above, a Hartree-Fock calculation proceeds the CI calculation, and the orbital coefficients, c_i of equation 1.9, are held fixed while the determinant coefficients, C_i are optimized. In multi-configuration self-consistent field (MCSCF) calculations, both the orbital coefficients and the determinant coefficients are optimized, in a similar manner (though more complicated) to the Roothan-Hall equations. This is a computationally demanding task, but offers a distinct advantage over the two methods above: a full CI in the restricted space restores size-consistency to the calculation.

There are several ways of performing an MCSCF calculation, though many of them involve dividing the orbitals of a Hartree-Fock determinant into three separate spaces: an inactive space, where the lowest energy orbitals are doubly occupied in all determinants, a virtual space, where the highest energy orbitals are unoccupied in all determinants, and an active space, where the orbitals are of intermediate energy. If all excitations are allowed within the active space, ie a full CI in the active space, the calculation is considered a Complete Active Space SCF, or CASCF. Excitations within the active space can be further restricted to reduce the number of determinants. Clearly, a judicious choice of orbital partitioning is paramount to the success of MCSCF calculations.

1.3 DENSITY FUNCTIONAL THEORY

Hohenberg and Kohn proved that the ground state energy and all other ground-state properties are uniquely determined by the electron density.[5] This theorem guarantees that there exists a

functional for which the exact energy can be found, but does not state what the functional is. The functional is known exactly with the exception of two terms: the kinetic energy and the exchange-correlation. In order to alleviate the first problem and generate a set of solvable equations, Kohn and Sham[21] introduced the idea of using orbitals that are eigenfunctions of a one-electron Hamiltonian. Similar to the HF method discussed above, these orbitals are collected into a Slater-determinant and leads to a solvable equation similar in form to equation 1.13a:

$$\left\{-\frac{1}{2}\nabla_1^2-\left(\sum_{I=1}^{M}\frac{Z_I}{r_{1I}}\right)+\int\frac{\rho(r_2)}{r_{12}}dr_2+V_{XC}(r_1)\right\}\chi_i(r_1)=\epsilon_i\chi_i(r_1) \qquad (1.16)$$

where the first term is the kinetic energy and the second term is the interaction of of electron 1 with nucleus I, similar to equation 1.13a. However, in equation 1.16, the Coulomb operator is an integral over the charge density. These leads to a self-interaction error, where the interaction of electron 1 is affected by the net density, which includes the charge of electron 1. Additionally, there is no exact exchange term, which exactly cancels the self interaction error in the Hartree-Fock expression. This has a computational advantage, however: the Coulomb term can be related to the second derivative of the electric potential through Poisson's equation, and thus can be solved numerically on a grid, which is a much more efficient way to evaluate the integral. The additional V_{xc} potential is the exchange-correlation potential and is related to the exchange-correlation (xc) functional by $\frac{\delta E_{xc}[\rho(\boldsymbol{r})]}{\delta\rho(\boldsymbol{r})}$. To solve equation 1.16, it is necessary to define the xc functional. It is important to note that solving equation 1.16 makes DFT variational; however, parameterization of the xc functional may lead to a solution to equation 1.16 that is lower than the true ground state energy.

The most basic definition for xc functional is the local density approximation(LDA),[22] where the exchange-correlation functional is parameterized from the density of the homogeneous electron gas. In the LDA, E_{xc} takes the form:

$$E_{xc}[\rho(\boldsymbol{r})] = A\int \rho(\boldsymbol{r})^{\frac{4}{3}} dr \tag{1.17}$$

where A is a parameter derived from the homogeneous electron gas. While this is a simple approximation, it can give accurate results for many systems. Accuracy can be improved by going beyond the local density approximation and including the gradient of the density at a point, known as the generalized-gradient approximation (GGA). This can be extended to include the second derivative of the density, though in many cases this offers little improvement. Another path to increase the accuracy of DFT is to consider the exchange-correlation functional as a sum of an exchange functional and a correlation functional. Then, the exchange functional can be further considered as a sum of functionals, where a fraction of exact exchange from Hartree-Fock theory is mixed in with the exchange from LDA. These are referred to as hybrid functionals, and they can give very high accuracy for certain properties, such as geometric parameters, for many systems. A generic example is:

$$E_{xc} = E_{XC}^{LDA} + A\left(E_X^{HF} - E_X^{LDA}\right) + B\Delta E_X^{GC} + C\Delta E_C^{GC} \tag{1.18}$$

Where the A, B, and C, are parameters that can be adjusted based on fitting to a data set, E^{HF} , and ΔE_x and ΔE_C are gradient corrected (GC) exchange and correlation terms. Including exact exchange into the DFT functional makes the calculation much more computationally demanding.

1.4 QUANTUM MONTE CARLO

A third method to solve the Schrödinger equation for a many electron system is by use of Monte Carlo methods. Monte Carlo is a stochastic method for the integration of an equation. Monte Carlo is commonly applied as importance sampling Monte Carlo, introduced by Metropolis.[23] With importance sampling, a probability density function is used to accept or reject a Monte Carlo move and steer the sampling towards areas of greater importance. There are two common methods for applying Monte Carlo simulations to quantum systems; variational Monte Carlo (VMC) and diffusion Monte Carlo (DMC). The details are discussed below.

1.4.1 Variational Monte Carlo

Variational Monte Carlo (VMC) uses a trial wave function to compute molecular properties. The trial wave function is one (or many) Slater determinant of orbitals taken from a HF or DFT calculation as outlined above. The variational energy of any wave function is given by:

$$E_v = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{1.19}$$

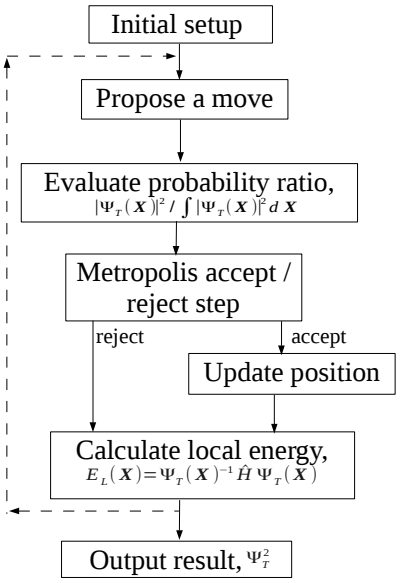
To evaluate the above energy using VMC, the Hamiltonian and overlap integrals are rewritten as a summation with the trial wave function acting as a probability distribution function:

$$E_v = \int \frac{\Psi_T^2}{\int \Psi_T^2} \frac{\hat{H}\,\Psi_T(\boldsymbol{r})}{\Psi_T(\boldsymbol{r})}\, d\boldsymbol{r} = \frac{1}{K} \sum_{k:\boldsymbol{r}_k \in \Psi_T^2}^{K} \frac{\hat{H}\,\Psi(\boldsymbol{r}_k)}{\Psi(\boldsymbol{r}_k)} \tag{1.20}$$

where the final term in the summation is E_L, the local energy. Samples are taken from a Gaussian distribution function from any point in $\boldsymbol{r}$ for K points. Figure 1 shows a simple flow chart that illustrates a VMC simulation.

VMC is a computationally efficient method, but the results are generally not an improvement over whatever method generated the trial wave function. The true power of VMC lies in using it to optimize parameters in the trial wave function to give an initial wave function for DMC with lower energy and variance.

Figure 1.1.1 - Typical VMC simulation



1.4.2 Optimization of correlation parameters in VMC

The trial wave function for use in a diffusion Monte Carlo simulation is generally written in the form:

$$\Psi_T = e^{\left(J_1^{\alpha} J_2^{\beta} J_3^{\gamma}\right)} \sum_{1}^{N} C_N \Psi_N \tag{1.21}$$

where J_x^n are Jastrow factors and C_N are expansion coefficients as defined in equation 1.16. The Jastrow factor can take many different functional forms. In general, J_1^α will be an electron-electron term that is a function of r_1-r_2 and the optimizable parameters α, J_2^β will be an electron-nucleus term that is a function of r_1-r_I and optimizable parameters β, and J_3^γ is an electron-electron-nucleus term that is a function of r_1-r_2, r_1-r_I, and r_2-r_I and optimizable parameters γ. Because the Jastrow factors are positive everywhere, they have no effect on the nodal surface. The addition of the Jastrow factors has two main goals, namely to reduce the energy and variance of the trial wave function and account for the electron-electron cusp. A lower variance and energy will lead to a faster convergence for the DMC calculation.[24] Historically, the electron-nucleus Jastrow factor had also been used to account for the nuclear cusp, but this is generally a poor method. In calculations with a pseudopotential (section 1.4.4), the nuclear cusp is zero, and in all electron calculations, there are several methods available[25] that are computationally more efficient and offer a similar reduction in variance and energy.

Historically, optimizing the parameters in the trial wave function has been done by choosing the variance of the wave function as the cost function to minimize:

$$\sigma_E^2(\alpha) = \frac{\int \Psi_T^2(\alpha)\left[E_L(\alpha) - E_V(\alpha)\right]^2 d\boldsymbol{r}}{\int \Psi_T^2(\alpha)\, d\boldsymbol{r}} \tag{1.22}$$

where α is used to indicate that the trial wave function is based on some set of parameters and E_v is the variational energy from equation 1.19. Minimizing the variance has been popular due to the stability of the algorithm. The local energy is independent of $\boldsymbol{r}$ for an eigenstate, so that eigenstates of $\hat{H}$ give the minimum value of zero variance for any set of configurations. This is not true for the variational energy, and makes energy minimization more difficult.

The minimum of the variance generally overlaps well with the minimum of the energy, but because the trial wave function cannot exactly represent an eigenstate, they are not the same minimum. Direct minimization of E_v may make more sense, as a DMC calculation attempts to determine the lowest energy of the system. There is another important advantage; when optimizing parameters that affect the nodal surface, such as determinant coefficients, energy minimization is essential; minimization of E_v does not reuse the same set of configurations with different parameters, and so any old configurations near a new nodal surface will not introduce difficulties in the optimization. Recently, progress has been made in developing stable algorithms[26] for optimization of parameters using the variational energy, or a linear combination of energy and variance, as a cost function.

The details for energy minimization can be found in references 26 and [28] but a brief overview is given here for completeness. The wave function Ψ_T^{α} can be Taylor-expanded as:

$$\Psi\left(\alpha^{(n+1)}\right)=\Psi\left(\alpha^{(n)}\right)+\sum_{i=1}^{p}\delta\alpha_i^{(n)}\frac{\partial\Psi}{\partial\alpha_i}\Big|_{\alpha^{(n)}}+O\left(\left[\delta\alpha^{(n)}\right]^2\right) \tag{1.23a}$$

$$\Psi\left(\alpha^{(n+1)}\right)=\sum_{i=0}^{p}a_i\varphi_i+O\left(\left[\delta\alpha^{(n)}\right]^2\right) \tag{1.23b}$$

where n+1 indicates a change in parameters by δ, and a_i, φ_i are the parameters and the derivative of the wave function with respect to the parameters, respectively. This form allows a_i to optimized by diagonalization and taking the vector of coefficients equal to the lowest eigenvalue.

1.4.3 Diffusion Monte Carlo

The time-dependent Schrödinger equation (TDSE) is

$$-i\frac{\partial \Phi}{\partial t}=\left(\frac{1}{2}\nabla_r^2+\left[V(r)-E_T\right]\right)\Phi \tag{1.24}$$

where the term in parenthesis is the Hamiltonian. The formal solution to the TDSE is

$$\Phi(r,t)=e^{-it\left(\hat{H}-E_T\right)}\Phi(r,t=0) \tag{1.25}$$

where E_T is an energy offset. This can be expanded in eigenfunctions of $\hat{H}$ that converge to the ground state when the imaginary time, $\tau=it$, becomes large:

$$\Phi(r,\tau)=e^{-\tau\left(\hat{H}-E_T\right)}\left(\sum_j c_j\Psi_j(\boldsymbol{r})\right) \tag{1.26a}$$

$$\lim_{\tau\to\infty}\Phi(r,\tau)=c_0\Psi_0(r) \tag{1.26b}$$

Thus, the TDSE can be propagated through imaginary time to give the exact ground state wavefunction. This can be solved by exploiting an isomorphism between the TDSE and a classic diffusion equation[29,30] modified by a first order rate term,

$$\frac{\partial \Psi}{\partial t}=D\nabla^2\Psi-k\Psi \tag{1.27}$$

where the diffusion constant D is ½ and is solved to give the kinetic energy of the TDSE, and the rate constant k is $V(\boldsymbol{r})-E_T$ is solved for the potential energy.

Diffusion is solved for using the random walk process established by Einstein for describing Brownian motion. A collection of points, **r**, called walkers, is sampled from some initial density ρ and take an independent random step η. The updated density function will then be:

$$\rho(\boldsymbol{r},\tau)=\int\rho(\boldsymbol{r}-\eta,\tau)\,g(\eta,\tau)\,d\eta \tag{1.28}$$

where $g(\eta,\tau)$ is a Gaussian distribution function. This result means that any initial density function can be used to solve for a solution to the diffusion equation using walkers taking random steps.

The first order kinetic term is solved by a branching process. Initially, each walker is assigned a unit weight that is updated at each Monte Carlo step according to:

$$w_k(\tau+\Delta\tau)=w_k(\tau)e^{(-V(\boldsymbol{r})-E_T)\Delta\tau} \tag{1.29}$$

Equation 1.29 will eventually diverge[31] due to a variance of the products of weights, and as such only a small number of walkers contribute to the average. A stochastic birth/death process is used to replicate walkers with a weight greater than unity and remove walkers with a weight less than one. This may introduce a population bias is too small of a population of walkers is used.

The kinetic and potential energy terms of the TDSE can each be solved by their respective equations, and it is simple to combine the two by taking one diffusion step, and then one branching step. However, because these two operators do not commute, an error is introduced according to the Trotter-Suzuki formula (1.30a) known as time-step bias. This requires that a short time step is used, and iterative applications of this short time step are used to reach the large τ limit (1.30b-c).

$$e^{-(T+V)\tau}=e^{-\frac{1}{2}V\tau}e^{-T\tau}e^{-\frac{1}{2}V\tau}+O(\tau^3) \tag{1.30a}$$

$$\Phi(\boldsymbol{r},\tau)=\lim_{n\to\infty}\prod_n e^{-(T+V)\frac{\tau}{n}} \tag{1.30b}$$

$$\Phi(\boldsymbol{r},\tau)=\lim_{n\to\infty}\prod_n e^{-\frac{V}{2}\frac{\tau}{n}}e^{-T\frac{\tau}{n}}e^{-\frac{V}{2}\frac{\tau}{n}} \tag{1.30c}$$

In practice, time-step bias can be removed by calculating several different DMC energies at differing time steps and extrapolating to dτ=0.

The method outlined above can be used to solve the TDSE with Monte Carlo methods, albeit with a time-step error. This method is known as the simple-sampling[32] method, and a straightforward improvement can be made by introducing a trial wave function Ψ_T as an

importance sampling term. This gives a mixed density, $\rho=\Phi\Psi_T$, to sample, and multiplying equation 1.24 by Ψ_T yields an equation by which the density can evolve:

$$\frac{\partial}{\partial\tau}\rho(\boldsymbol{r},\tau)=\frac{1}{2}\nabla^2\rho-\frac{1}{2}\nabla\left(\frac{2\nabla\Psi_T(\boldsymbol{r})}{\Psi_T(\boldsymbol{r})}\rho\right)-(E_L(\boldsymbol{r})-E_T)\rho \tag{1.31}$$

where the first term is the diffusion equation, the second term is called the drift velocity, and the third term is the branching equation.

Importance sampling has several meaningful consequences. The density of walkers will be increased in areas where Ψ_T is large and reduced where it is small. Additionally, the branching term is now based on the local energy as opposed to the potential, which leads to a suppression of the branching process resulting in a more stable algorithm. Assuming a constant drift velocity between $\boldsymbol{r}$ and $\boldsymbol{r}'$ is equivalent to using normal ordered operators and introduces an error of $O(\tau^2)$. The branching process reduces the error in the distribution by approximately $O(\tau)$, and so the overall error due to the time step in the DMC calculation is $O(\tau)$.

The drift term can also be used to enforce the fixed-node approximation (FNA) by killing a walker that changes sign, or by rejecting the step that caused the sign change (although rejection typically gives smaller time-step errors.) The FNA is enforced by requiring the mixed density to have the same sign as the trial wave functions at all points in space. This approximation is required to enforce the antisymmetry of the wave function produced as a solution to the TDSE. Without it, DMC would propagate out all of the Fermionic states and result in a Bosonic ground state wave function. While little is known about the $3N$-1 dimensional nodal surface of an N electron system, it has been shown[14,33] that all of the nodal pockets of a wave function are symmetrically equivalent, and exploring one nodal pocket gives the same energy as any other nodal pocket. The fixed-node approximation is the only

uncontrollable error in the DMC calculation. Clearly then, it is important to minimize this error through a judicious choice of the trial wave function.

1.4.4 Pseudopotentials in DMC

DMC formally scales as $O(N^3)$, where N is the number of electrons. A step that is taken to reduce the computational cost and to increase efficiency of DMC calculations is the use of pseudopotentials. Pseudopotentials have been used in plane wave DFT calculations for many years to reduce the high kinetic energy oscillations close to the nuclei, and thus reduce the number of plane waves that are needed to accurately describe a system. Using pseudopotentials in QMC introduces another, albeit small, error. In general, pseudopotentials are angular momentum dependent, which makes them non-local. In DFT calculations, this does not introduce any errors, but in QMC calculations, where the local energy of a walker needs to be calculated, the non-local character of a pseudopotential needs to be included. To correct this, there are two schemes: the locality approximation[34] and the T-move scheme.[35] The locality approximation makes an effective Hamiltonian that splits the nonlocal portion of the pseudopotential into a potential on the trial wave function and a potential on the unknown wave function produced throughout the simulation. The nonlocal potential on the unknown wave is subsequently neglected. This makes the effective Hamiltonian non-variational. The locality approximation offers a small magnitude of localization error and small time step bias. In the T-move scheme, the non-local potential on the unknown wave function is reintroduced into the Hamiltonian, restoring the variational principle. This potential is reintroduced through the use of a second accept/reject step in the branching equation, where the non-local walk is rejected if the wave function changes sign and the sign-flip term (the potential from the rejected step) is added

to the local potential. The T-move scheme violates the detailed balance condition, but reduces energy instabilities in the random walk. The disadvantages of the T-move method are that it introduces larger time step bias than the locality approximation and can become more computationally demanding. Time step bias can be removed by extrapolation of the energy to zero time step. The magnitude of the non-local error is generally not a large portion of the final energy in either scheme, however, and the only uncontrolled error is largely due to the fixed-node approximation.

2.0 EFFECT OF THE TRIAL WAVE FUNCTION ON DMC CALCULATIONS

2.1 CORRELATION CONSISTENT GAUSSIAN BASIS SETS FOR H, B-NE WITH DIRAC-FOCK AREP PSEUDOPOTENTIALS: APPLICATIONS IN QUANTUM MONTE CARLO CALCULATIONS

This work has been published as J. Xu, M.J. Deible, K.A. Peterson, K.D. Jordan, "Correlation Consistent Gaussian Basis Sets for H, B-Ne with Dirac-Fock AREP Pseudopotentials: Applications in Quantum Monte Carlo Calculations," *J. Chem. Theory Comput.*, **2013**, 9(5), 2170. J.X. performed the water monomer and dimer calculations. M.J.D. performed the water-benzene calculations. K.A.P. designed the basis sets and performed the spectroscopic calculations. All authors contributed to the discussion.

2.1.1 Introduction

Quantum Monte Carlo methods,[29,30] 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[29,30] 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[36] 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 six larger than obtained with high cut-off plane-wave basis sets.[37,38] 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,[39,40] which we test in coupled cluster calculations on H_2, B_2, C_2, N_2, and F_2 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 non-local pseudopotentials in diffusion Monte Carlo calculations.

2.1.2 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 2.1.1 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),[41] augmented with diffuse *s*, *p*, and *d* functions from the aug-cc-pVDZ basis set[42,43] or by plane-wave basis sets with energy cutoffs of 60, 120, and 160 a.u. The geometry of the water monomer was taken from experiment,[44] with OH distances of 0.9572Å and an HOH angle of 104.52º. 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.[45] 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.[36] The Hartree-Fock and the LDA calculations with the augmented BFD basis set calculations were carried out using the Gaussian 03 package,[46] and the plane-wave DFT calculations were carried out using ABINIT.[47] The quantum Monte Carlo calculations were carried out using the CASINO code.[48]

The VMC calculations with the augmented BFD GTO basis set gave a variance of 1.25 a.u. compared with variances of 1.63, 0.34, and 0.26 a.u. obtained using plane-wave basis sets with cutoffs of 60, 120, and 160 a.u., respectively. The importance of going to very high energy cutoffs when using plane-wave basis sets in quantum Monte Carlo calculations has been noted previously in the literature.[49]

Table 2.1.1 - VMC energies and variances for the water monomer using the CASINO Dirac-Fock pseudopotential on all atoms.[a]

Basis set	VMC energy (a.u.)	Variance of the VMC energy (a.u.)
Augmented BFD	-17.161(3)	1.25
Plane-wave/BLIP (60 a.u.)[b]	-17.159(5)	1.63
Plane-wave/BLIP (120 a.u.)[b]	-17.191(2)	0.34
Plane-wave/BLIP (160 a.u.)[b]	-17.194(2)	0.26

[a]Hartree-Fock and LDA calculations were used to obtain the molecular orbitals for GTO and plane-wave basis sets, respectively.

[b]The plane-wave energy cutoff is given in parentheses.

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-pV5Z, and cc-pV6Z basis sets,[42] 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 a.u., respectively. Similar variances are obtained with the corresponding aug-cc-pVXZ basis sets basis sets.[42,43] 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 1*s* 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,[41] for which the BFD basis sets were developed, the variance from a VMC calculation on the water monomer using the augmented BFD basis set is less than 0.3 a.u.

2.1.3 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),[50–54] 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 1*s* orbitals by pseudopotentials. However, as noted above, the use of pseudopotentials to model the 1*s* 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 1*s* 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-pV5Z-CDF and are described in Tables S1-S5 in the supplemental 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 (ω_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.[55,56] For comparison, all-electron CCSD(T) calculations employing the Douglas-Kroll-Hess (DKH) scalar relativistic Hamiltonian[57,58] and the aug-cc-pV5Z-DK basis sets[42,43,59] within the frozen-core approximation were also carried out. For the atoms and open-shell molecules, the ROHF-UCCSD(T) method[60,61] was utilized. The calculated R_e, D_e, and ω_e results together with the corresponding experimental values are summarized in Table 2.1.2. For H_2, the CCSD(T) calculations with the CDF pseudopotential and aug-cc-pVQZ-CDF and aug-cc-pV5Z-CDF basis sets as well as the CCSD(T) calculations using the DKH Hamiltonian and the aug-cc-pV5Z-DK basis set give bond lengths, vibrational frequencies, and dissociation energies very close to the experimental values.

Table 2.1.2 - Calculated CCSD(T) spectroscopic constants for the ground states of H_2, B_2, C_2, N_2, O_2, and F_2 with the aug-cc-pV*n*Z-CDF basis sets.[a]

	Basis	E_e (E_h)	D_e (kcal/mol)	r_e (Å)	ω_e (cm^{-1})
H_2	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-pV5Z-CDF	-1.174295	109.37	0.7415	4403.2
	aug-cc-pV5Z-DK	-1.174263	109.35	0.7416	4403.1
	Expt[b]		109.49	0.7414	4401.2
B_2	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-pV5Z-CDF	-5.327859	63.64	1.5954	1039.3
	aug-cc-pV5Z-DK	-49.319978	64.83	1.5927	1048.6

	Expt[b]		67.65	1.5900	1051.3
C_2	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-pV5Z-CDF	-11.065867	142.96	1.2459	1846.5
	aug-cc-pV5Z-DK	-75.836466	144.27	1.2447	1858.0
	Expt[b]		147.8	1.2425	1854.7
N_2	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-pV5Z-CDF	-19.907062	223.06	1.1003	2345.0
	aug-cc-pV5Z-DK	-109.473998	225.46	1.0993	2358.5
	Expt[b]		228.4	1.0977	2358.6
O_2	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-pV5Z-CDF	-31.904714	117.19	1.2107	1587.7
	aug-cc-pV5Z-DK	-150.295526	118.89	1.2073	1598.9
	Expt[b]		120.6	1.2075	1580.2
F_2	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-pV5Z-CDF	-48.318677	37.37	1.4151	921.7
	aug-cc-pV5Z-DK	-199.558515	38.05	1.4114	926.5
	Expt[b]		39.0	1.4119	916.6

[a] H_2 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_2, O_2, and F_2 by using the experimental atomic splittings.

[b] Experimental results are from Ref. 62, except for the dissociation energy of C_2 which is taken from Ref. 63. The reference dissociation energy of B_2 is from high level theoretical calculations of Ref. 64.

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-pV5Z-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-pV5Z-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-pV5Z-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. 64.) 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-pV5Z-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.

2.1.4 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 2.1.3. (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 2.1.1 and 2.1.3, it is seen that the variance in the energy is reduced about threefold and the VMC energy is about 0.03 a.u. 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 a.u. cutoff. For the largest GTO basis set considered, aug-cc-pV5Z-CDF, the variance is only 0.22 a.u. 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-pV5Z-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 2.1.3 - VMC energies and variances for the water monomer[a]

Basis set	VMC energy (a.u.)	Variance (a.u.)
aug-cc-pVDZ-CDF[a]	-17.193(2)	0.42
aug-cc-pVTZ-CDF[a]	-17.197(1)	0.29
aug-cc-pVQZ-CDF[a]	-17.199(1)	0.23
aug-cc-pV5Z-CDF[a]	-17.200(1)	0.22
cc-pV5Z(*sp*)+2*d*-CDF[b]	-17.197(1)	0.25
cc-pV5Z(*sp*)+2*df*-CDF[b]	-17.198(1)	0.23

[a]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.

[b]Results obtained employing the CDF pseudopotentials on all atoms.

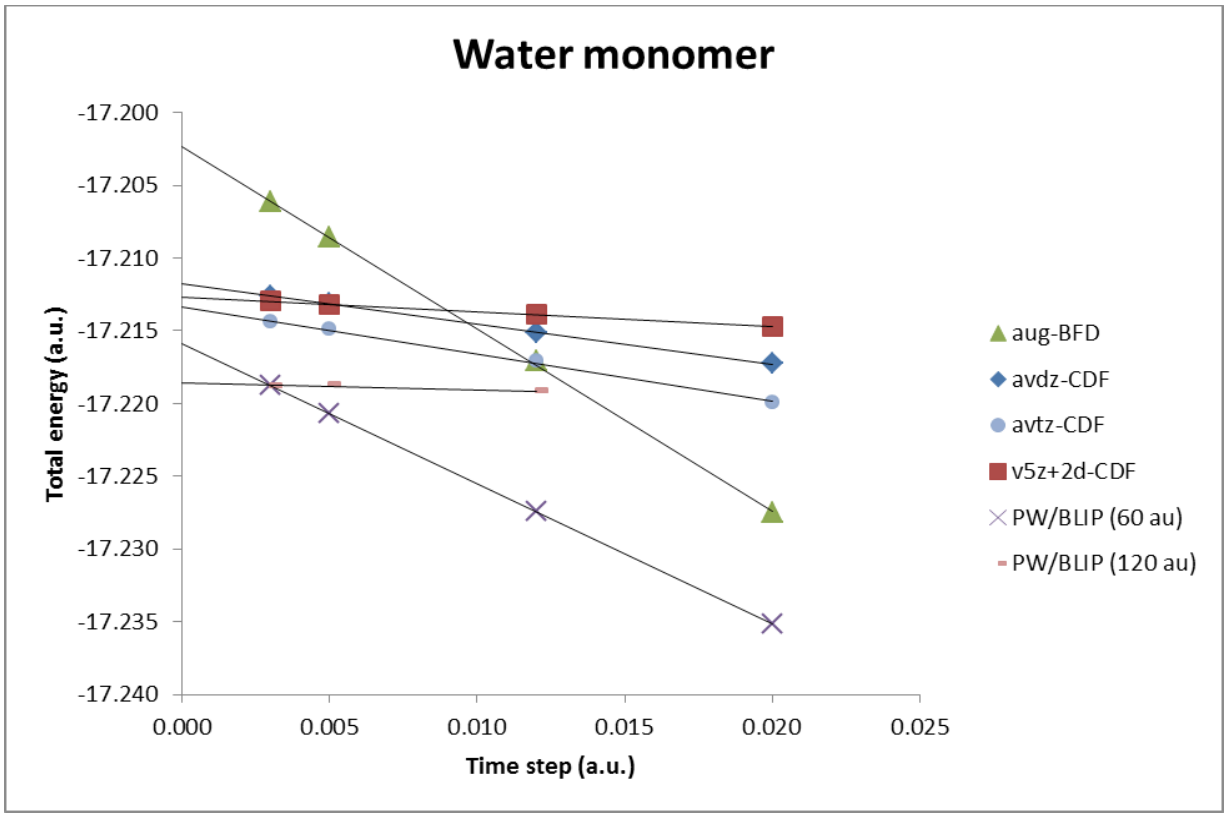
Table 2.1.3 also includes results obtained for two modified versions of the cc-pV5Z-CDF basis set denoted cc-pV5Z(*sp*)+2*d*-CDF and cc-pV5Z(*sp*)+2*df*-CDF, which differ from the full cc-pV5Z-CDF basis set by employing smaller sets of polarization functions, taken from the cc-pVTZ-CDF basis set. (The cc-pV5Z-CDF basis set is derived from aug-cc-pV5Z-CDF by deleting the most diffuse function of each angular momentum.) The corresponding basis sets for hydrogen differ from the full cc-pV5Z-CDF basis set by using as polarization functions just the 2*p* and 2*p*1*d* 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 2.1.1 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 a.u. 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 a.u., using the T-move procedure.[35,65]

Figure 2.1.1 - 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+2*d*-CDF refer to aug-cc-pVDZ-CDF, aug-cc-pVTZ-CDF, and cc-pV5Z(*sp*)+2*d*-CDF, respectively.



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. Secondly, DMC calculations using trial functions with the largest variances, namely those expanded in terms of the augmented BFD and the 60 a.u. cutoff plane-wave basis sets, display the steepest slopes. Thirdly, the DMC calculations with different trial

functions give different total energies in the t → 0 limit. The trial function using the augmented BFD basis set gives the highest DMC energy, -17.20207(9) a.u., and that represented in terms of the plane-wave basis set with 120 a.u. 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 a.u. 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_2O calculated using the aug-cc-pVTZ-CDF basis set with HF and LDA orbitals agree to within 0.0005 a.u.. We note also that the slope in the DMC energy vs. time step curve is reduced by about a factor of two 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 a.u. above the result obtained using the trial function expanded in the 120 a.u. 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 non-local 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[35] or locality approximation (LA),[34] 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,[34] 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.[35,66] 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 2.1.2 reports the DMC total energies of water dimer obtained using different basis sets and the two approaches for dealing with the non-local pseudopotential problem. The geometry used for the water dimer was obtained by optimization at the MP2/aug-cc-pV5Z level. Figure 2.3 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 non-zero 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 non-locality of the pseudopotentials was also noted by Gurtubay and Needs[67] 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 2.1.4. The water-benzene calculations will be discussed in Section 4C.

Figure 2.1.2 - 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 non-locality of the pseudopotentials.

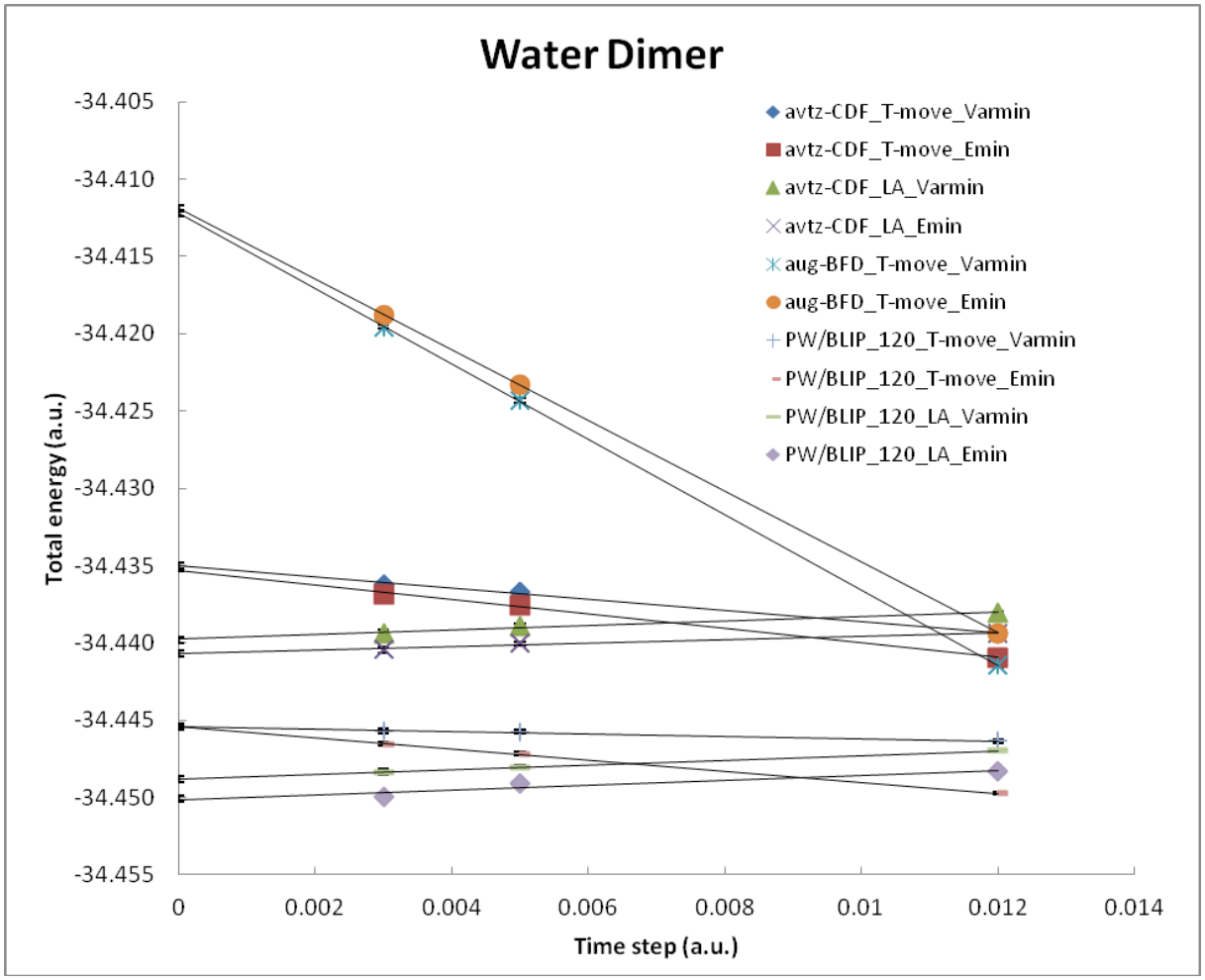
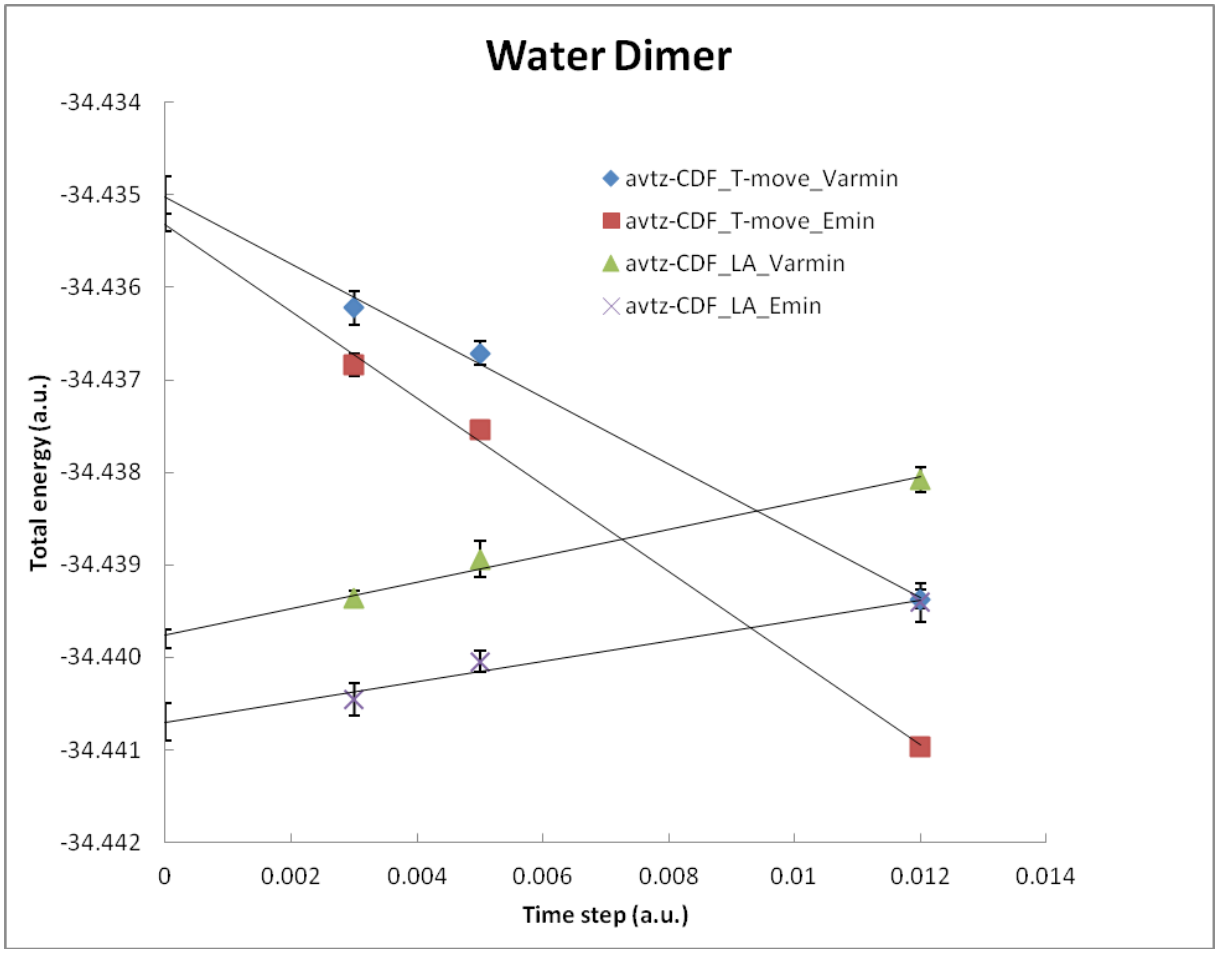


Figure 2.1.3 - Energies of the water dimer on an expanded scale, from DMC calculations using trial functions represented in terms of the aug-cc-pVTZ-CDF basis set.



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,[61] as well as from complete-basis-set limit CCSD(T)[62] calculations.[64] 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 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.

Table 2.1.4 - Calculated binding energy of water dimer.[a]

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[b]	-5.02±0.18
DMC/B3LYP[b]	-5.21±0.18
CCSD(T) CBS limit	-5.02±0.05
DMC/B3LYP[c]	-5.03±0.07
DMC/B3LYP[d]	-5.07±0.07

[a] Results obtained by subtracting twice the DMC energy of the monomer from the DMC energy of the dimer.

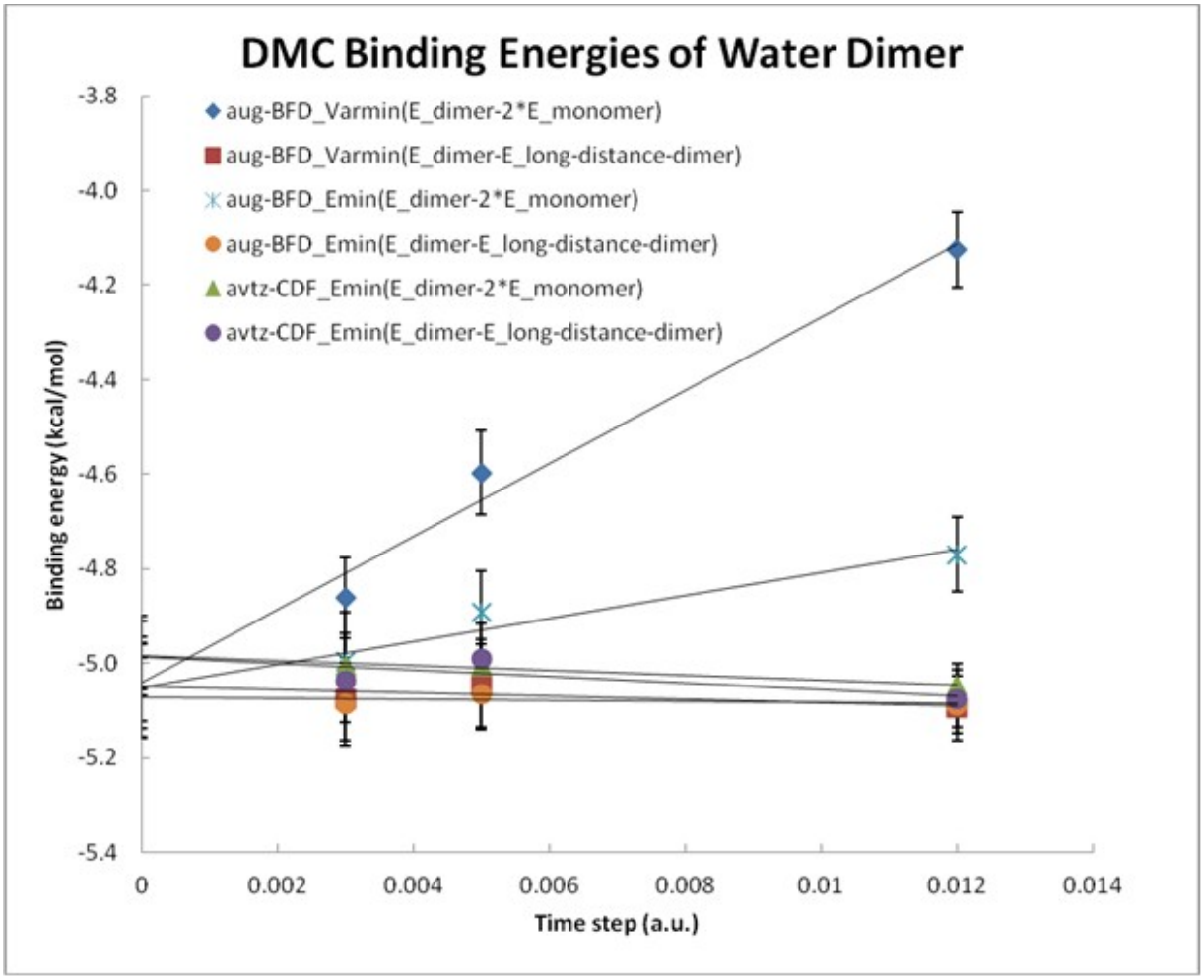
[b]From all-electron calculations, Reference 62

[c]From pseudopotential calculations using the locality approximation.

[d]From pseudopotential calculations using the T-move procedure, Reference 68

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 of water-benzene.[69]) 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.

Figure 2.1.4 - 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.



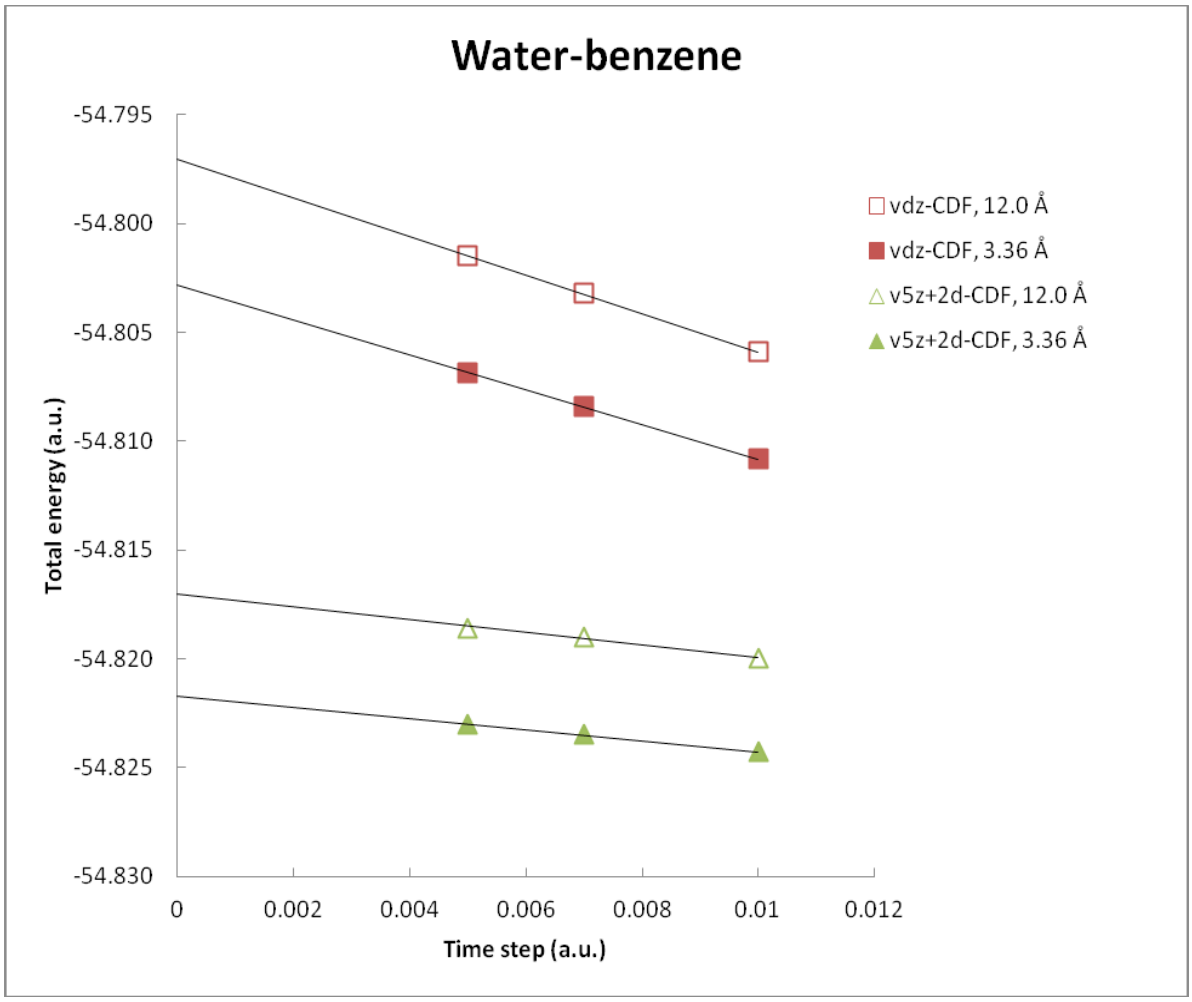
As shown in Figure 2.1.4, the extrapolated zero-time-step energies obtained using different trial functions and strategies for each basis set are essentially identical within the error bars. With the S1 procedure, there is a large time step bias for the binding energy with the aug-BFD basis set, while 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. 70. 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 the benzene ring, and using the S2 strategy described in Section 2.1.3. 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 a.u., 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-CFD and the cc-pV5Z(*sp*)+2*d*-CDF basis sets described in Section 2. The DMC energies for the short and long distances are reported in Fig. 2.1.5 from which it is seen that the time-step error is significantly reduced in going from the cc-pVDZ-CDF to the cc-pV5Z(*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-pV5Z(*sp*)+2*d*-CDF basis sets are -3.6(1) and -2.9(2) kcal/mol, respectively, with the later value being in good agreement with the -3.2 kcal/mol CCSD(T)-F12[71] result of Ref. 72. For comparison, we note that Ma and coworkers obtained a binding energy -2.9 kcal/mol using the DMC method together with a LDA trial function expanded in terms of planewave/BLYP functions, but employing a geometry slightly different from that used here.[69]

Figure 2.1.5 - DMC energies of the benzene water complex obtained using different basis sets for representing the orbitals in the trial functions. vdz-CDF and v5z+2*d*-CDF refer to cc-pVDZ-CDF and cc-pV5Z(*sp*)+2*d*-CDF basis sets, respectively. All results were obtained using the T-move procedure.



2.1.5 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_2O, $(H_2O)_2$, 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.[73] 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.

2.1.6 Acknowledgements

This research was carried out with the support of the National Science Foundation (KDJ). 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.

2.2 EXPLORATION OF BRUECKNER ORBITAL TRIAL WAVE FUNCTIONS IN DMC CALCULATIONS

This work has been submitted to the Journal of Chemical Theory and Computation as M. J. Deible, K. D. Jordan, "Exploration of Brueckner Orbital Trial Wave Functions in DMC Calculations." M.J.D. performed the calculations. All authors contributed to the discussion.

2.2.1 Introduction

The diffusion Monte Carlo (DMC) method[9,74] is capable of giving exact electronic energies within the fixed-node approximation.[29,30] A trial wave function, generally taken to be a Slater determinant of orbitals from density-functional theory (DFT) or Hartree-Fock (HF) calculations, is used to enforce the fixed-node approximation. This condition is required to prevent the ground state wave function from collapsing on a Bosonic state. If the trial wave function were to exactly describe the nodal surface for exchange of electrons, the DMC method, if used in all-electron calculations and run for sufficiently large number of moves and corrected for time step bias, would give the exact non-relativistic ground state energy.[14] In general, trial functions employing a Slater determinant of DFT orbitals give lower total energies than do trial functions employing a Slater determinant of Hartree-Fock orbitals, which implies that the use of DFT orbitals provides a better description of the nodal surface.[16,69] In the present study, we explore the use of trial functions comprised of a Slater determinant of Brueckner orbitals (BO).[75,76] One might expect that such trial wave functions would give a nodal surfaces superior to a Slater determinant of DFT orbitals, as the Slater determinant of BOs is that with the maximum overlap with the exact wave function for the basis set employed.[77]

There has been considerable discussion in the literature concerning the similarity and differences between HF, DFT, and Brueckner orbitals. Scuseria[78] has shown that the DFT equations can be derived through approximations made to the Brueckner equations. Lindgren and Salmonson[79] have argued that DFT and Brueckner orbitals are closely related. Heßelmann and Jansen have shown that Brueckner orbitals offer an improvement over Hartree-Fock orbitals when calculating first-order intermolecular interaction energies,[80] whereas DFT orbitals may or may not give an improved description of the first-order Coulomb and exchange energies, with the performance depending on the functional used to generate the orbitals.[81] Jankowsi *et. al.*[82,83] calculated the distance between orbital subspaces, and based on this measure, concluded that DFT orbitals can differ appreciably from both HF and Brueckner orbitals. An alternative method for testing the quality of various types of orbitals is how well they describe the nodal surface for exchange of electrons which can be evaluated by assessing their performance when used as trial functions for DMC calculations. In this study, we investigate the performance of trial functions using HF, DFT, and Brueckner orbitals in DMC calculations on a series of diatomic molecules and on a bent CO_2^- ion to determine whether the use of Brueckner orbitals leads to lower DMC energies than obtained using trial functions comprised of DFT orbitals.

2.2.2 Methodology

The diatomic molecules studied include BeO,[62] N_2,[64] O_2,[64] F_2,[64] and CN,[64] with the equilibrium geometries being taken from the respective references. To obtain dissociation energies, calculations were also carried out on the atoms in their ground electronic states. For these test systems, both pseudopotential and all-electron calculations were carried out. In the all-electron calculations the orbitals were expanded in terms of the cc-pVTZ basis set.[84] In the

calculations using pseudopotentials for all species other than Be the Trail-Needs AREP pseduopotentials[85] and basis sets of Xu et al.[86] were used. Specifically, for N, F, and C, the *spd* portions of the triple-zeta basis set of Xu et al. were used and for oxygen the *sp* portion of the quintuple basis set was combined with the *d* functions from the triple-zeta basis set. For Be, the pseudopotential and valence triple-zeta basis set of Burkatzki, Filippi, and Dolg (BFD) were used.[41] The CO_2 calculations also made use of the Trail-Needs pseudopotentials[85] and an 8*s*8*p*3*d* contracted Gaussian basis set formed by adding to the 6*s*6*p* portion of the quintuple-zeta basis set of Xu et al.[86] two diffuse *s* and two diffuse *p* functions with exponents determined by dividing the exponents of the most diffuse primitive functions in the 6*s*6*p* set by three and nine. *d* functions with exponents of 0.5586, 0.2271, and 0.1024 and 1.2192, 0.4904, and 0.2053 were included for carbon and oxygen atoms, respectively, which are the *d* functions of the double and triple-zeta basis set of Xu et. al. and the *d* function of the aug-cc-pVDZ basis set of Xu et. al. scaled by a factor of 1.5. The hybrid Becke3LYP[87,88] functional was used for the DFT calculations on the all-electron systems, and both the Becke3LYP and PBE0[89] functionals were used for the pseudopotential calculations of the diatomics. The Brueckner orbitals were obtained from coupled-cluster calculations in which the orbitals are rotated so as to eliminate single excitations to all orders. For one system, N_2, we also considered a trial function based on PBE[90] orbitals and also employed the more flexible cc-pVQZ-*g* basis set for the all-electron calculations.[84] In addition to the diatomic test cases described above, we also considered CO_2 and CO_2^- with CO bond lengths of 1.215 Å and an OCO angle of 147°. This geometry was chosen because earlier studies have shown that the anion, while bound (i.e., lying energetically below the neutral at the same geometry) is non-valence correlation bound[91] and is not properly described with either the B3LYP, PBE, or PBE0 DFT functionals. All open-shell systems were

described using spin-unrestricted orbitals. The trial functions were generated using the Gaussian 09 code,[92] and the diffusion Monte Carlo calculations were carried out using the CASINO code.[48]

The correction scheme of Ma *et al*.[93] was used in the DMC calculations to account for electron-nuclear cusps in the all-electron calculations. The DMC calculations were preceded by variational Monte Carlo (VMC) calculations to optimize (*via* energy minimization) the parameters in the Jastrow factors[36] which include explicit electron-nuclear (e-n), electron-electron (e-e), and e-e-n terms. The all-electron DMC calculations were carried out for time steps of 0.001, 0.003, 0.005, and 0.007 a.u., and the resulting energies were extrapolated to zero time step by use of quadratic fits. The calculations were carried out with 40,000 walkers and for sufficient number of steps so as to reduce the statistical errors in the extrapolated DMC energies of the diatomics and their atomization energies to under 0.3 and 0.5 kcal/mol, respectively. The errors in the atomization energies are defined as the differences between the experimental values, corrected for vibrational zero-point energy (ZPE), and the corresponding DMC results. For N_2, O_2, F_2, and CN, the experimental atomization energy values and zero-point energies are taken from reference 94. For BeO, the experimental atomization energy and ZPE are taken from cccbdb.nist.gov.

The parameters of the Jastrow factors for the diatomics with pseudopotntials and for the CO_2 test system were optimized *via* variance minimization. Time steps of 0.005, 0.0075, and 0.01 a.u. were used in the DMC calculations on the diatomic species with pseudopotentials, and time steps of 0.0125, 0.005, and 0.003 a.u. were used for the CO_2 test system. Linear fits were used to extrapolate to zero time step. The T-move procedure was used in the pseudopotential calculations to correct for the localization error.[35]

2.2.3 Results

Table 2.2.1 - Total energies from DMC calculations[a] using Hartree-Fock, B3LYP, and Bruckner orbitals.

	Energy (a.u.)[b]			
Species	Hartree-Fock	B3LYP	Brueckner	Brueckner(cc)[c]
Beryllium	-14.6575(1)	-14.6572(1)	-14.6575(1)	-
Carbon	-37.8296(2)	-37.8301(2)	-37.8296(2)	-
Nitrogen	-54.5759(2)	-54.5765(2)	-54.5757(2)	-
Oxygen	-75.0512(2)	-75.0518(2)	-75.0518(3)	-
Fluorine	-99.7161(3)	-99.7169(3)	-99.7161(3)	-99.7163(2)
N2	-109.5007(6)	-109.5047(4)	-109.5042(4)	-
O2	-150.2808(5)	-150.2873(5)	-150.2856(5)	-
F2	-199.4740(5)	-199.4850(2)	-199.4812(5)	-199.4816(3)
BeO	-89.8823(4)	-89.8849(3)	-89.8845(3)	-
CN	-92.6668(5)	-92.6888(5)	-92.6876(5)	-

[a] Results extrapolated to dt=0 as described in the text.
[b] Statistical errors (one standard deviation) are given in parentheses.
[c] These results obtained using Brueckner orbitals obtained from coupled cluster calculations including core correlation.

The total energies from the all-electron DMC calculations on the diatomic species and associated atoms are summarized in Table 2.2.1. For the atoms, DMC calculations with the Hartree-Fock, B3LYP, and Brueckner orbitals give total energies that essentially agree to within statistical error. This is consistent with earlier studies[16] that found that for atoms DMC energies were largely insensitive to whether DFT or HF orbitals are used for the trial functions. The situation is quite different for the diatomics: going from HF orbitals to B3LYP orbitals leads to an energy decrease from 1.63 kcal/mol for BeO to 13.8 kcal/mol for CN. The DMC energies obtained using B3LYP and Brueckner orbitals agree to within one standard deviation for N_2 and BeO, but for O_2, F_2, and CN, significantly lower DMC energies are obtained when using B3LYP

orbitals in the trial function. It should be noted that the use of spin-unrestricted calculations introduces spin contamination in the wave function. This has a larger effect on the Hartree-Fock orbitals than it does for the B3LYP and Bruckner orbitals. The largest spin contamination is found for the CN molecule, which has S^2 expectation values of 1.158, 0.758, and 0.764 for the HF, B3LYP, and Brueckner orbital wave functions, respectively, compared to the exact value of 0.750. While using RHF rather than UHF orbitals can result in lower DMC energies for open-shell systems,[16,17] it would not change the qualitative result that lower DMC energies result when using B3LYP and Brueckner orbitals than when using HF orbitals (whether RHF or UHF), and that use of B3LYP orbitals generally give a lower DMC energy, thus superior nodal surface, than use of Brueckner orbitals. Our DMC calculations using Hartree-Fock and B3LYP orbitals for the atoms give slightly higher DMC energies than obtained by Per *et al*.[16] using the same trial wave functions. However, for the diatomics, Per *et al*'s DMC energies are lower than our energies by up to 2.4 kcal/mol. These differences are due to the greater flexibility of the atomic basis sets used to represent the orbitals in the study of Per *et al*. Indeed, for N_2, our DMC energy using the trial function of B3LYP orbitals is 1.7 kcal/mol lower in energy when using the cc-pVQZ-*g* rather than the cc-pVTZ basis set.

Table 2.2.2 – Errors in the atomization energies calculated using the DMC method with single determinant trial functions and different orbital choices.

	Atomization energy error (kcal/mol)[a]		
Species	Hartree-Fock	B3LYP	Brueckner
N2	-9.52(45)	-7.76(35)	-7.07(35)
O2	-8.26(40)	-4.94(40)	-6.00(49)
F2	-11.97(49)	-6.07(40)	-7.45(49)

BeO	3.47(29)	4.91(23)	4.47(27)
CN	-17.08(36)	-3.97(36)	-3.90(36)

[a] Statistical errors (one standard deviation) are given in parentheses.

Figure 2.2.1 – Errors in the dimer atomization energies for different trial functions.

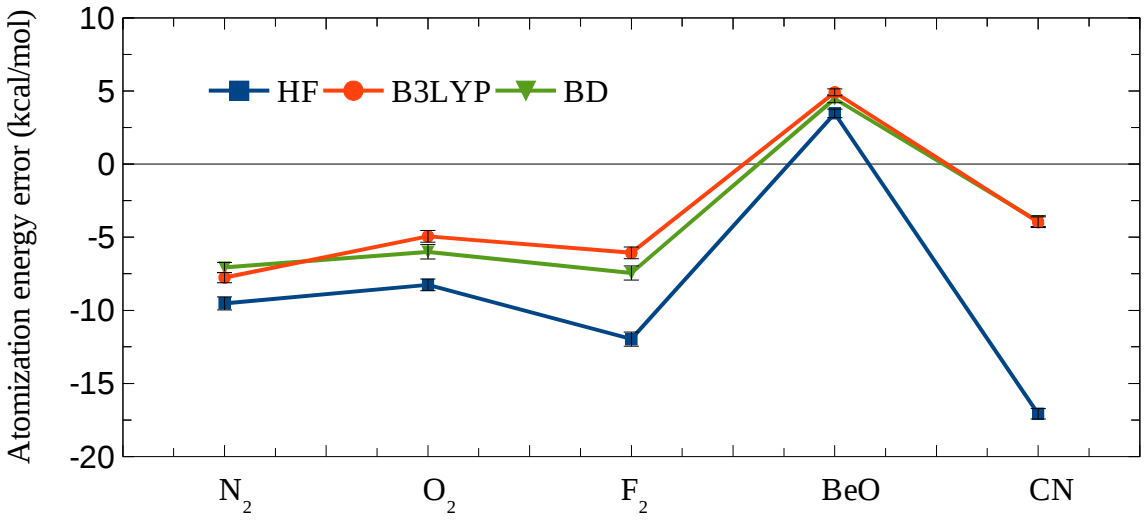


Table 2.2.2 and Figure 2.2.1 report the errors in the calculated atomization energies as previously defined. Consistent with the results discussed above for the total energies, the errors in the atomization energies tend to be smaller when calculated using either DFT or Brueckner orbitals than HF orbitals, with the errors obtained using the DFT orbitals being somewhat smaller. The exception is BeO, for which the DMC calculations using a Hartree-Fock determinant as the trial function give the smallest error in the atomization energy. As seen from Table 2.2.2, of the molecules considered, this is the only one for which the DMC calculations overestimate the atomization energy. The reason for this is well understood, namely, that the Be

atom has significant multiconfigurational character, which is important for describing its nodal surface in all-electron DMC calculations.[33] Thus, using a single determinant trial function results in too high a value of the DMC energy of the Be atom and an overestimation of the atomization energy of BeO.

In analyzing these results, it is important to note that the Brueckner orbitals were obtained from calculations using frozen 1*s* cores. As a check on whether relaxing this constraint significantly impacts the DMC energies in the case of F_2, we also carried out DMC calculations using Brueckner orbitals generated by correlating all electrons and using the cc-pCVTZ basis set[84,95] which includes functions for correlating the core. The error in the DMC value of the atomization energy using the Brueckner orbitals generated correlating all electrons is essentially identical to that using Brueckner orbitals generated in the frozen-core approximation.

For all diatomic molecules considered, the single-determinant based DMC calculations are unable to achieve chemical accuracy ($\pm$1 kcal/mol) in the atomization energies regardless of the orbitals used. Part of the error is due to the limitations of the basis set used to represent the orbitals, but most of the error is due to the inadequacy of single Slater determinant trial functions for calculating atomization energies which has been noted several times in the past, and reflects the inadequacy of a single determinant trial function for describing the nodal surface regardless of the choice of orbitals.[16,96,97] (In the case of N_2, the error in the atomization energy using B3LYP orbitals is reduced from -7.83 to -6.28 kcal/mol in going from the cc-pVTZ to the cc-pVQZ-*g* basis set. In contrast, the same atomization is obtained whether using PBE or B3LYP orbitals.) Indeed, it has been found that significantly improved results are obtained by use of full-valence CASSCF trial functions, particularly when optimizing the CI coefficients, orbitals,

and Jastrow factors simultaneously.[17] Also, it has been shown that chemical accuracy can be achieved by using large trial functions from CI calculations employing natural orbitals.[98]

Table 2.2.3 - Total energies from DMC calculations[a] using Hartree-Fock, Becke3LYP , PBE0, and Bruckner orbitals with a pseudopotential.

	Energy (a.u.)[b]			
Species	HF	B3LYP	PBE0	BD
N2	-19.8688(2)	-19.8715(2)	-19.8723(3)	-19.8721(3)
O2	-31.8668(3)	-31.8707(3)	-31.8715(3)	-31.8717(3)
F2	-48.2722(3)	-48.2768(3)	-48.2781(3)	-48.2788(3)
CN	-15.4266(2)	-15.4476(2)	-15.4480(2)	-15.4478(2)
BeO	-17.0119(3)	-17.0142(3)	-17.0146(3)	-17.0149(3)

[a] Results extrapolated to dt=0 as described in the text.
[b] Statistical errors (one standard deviation) are given in parentheses.

We now consider the results obtained for BeO, CN, O_2, F_2, and N_2 in the calculations employing pseudopotentials. The total energies obtained from DMC calculations using various types of orbitals in the single determinant trial functions are summarized in Table 2.2.3. From this table it is seen that the DMC calculations using trial functions employing Brueckner orbitals give energies for the N_2, O_2, and F_2 dimers lower than the Becke3LYP orbitals, but statistically similar results to those obtained using PBE0 orbitals. The total DMC energies for CN and BeO calculated using the three sets of orbitals agree to within statistical uncertainty. This indicates that when pseudopotentials are used to eliminate the 1*s* core orbitals, Brueckner orbitals are more effective at describing the nodal surfaces than are Becke3LYP orbitals and are equally effective as PBE0 orbitals.

Table 2.2.4 – DMC energies (a.u.) of CO_2 and CO_2^- at an OCO angle of 147°.[a,b]

Orbitals	Anion	Neutral	Energy difference (eV)
B3LYP	-37.6811(2)	-37.6691(2)	-0.327(8)
PBE0	-37.6809(3)	-37.6696(2)	-0.31(1)
PBE	-37.6781(3)	-37.6691(2)	-0.24(1)
BD	-37.6825(2)	-37.6688(2)	-0.378(8)
HF	-37.6745(2)	-37.6608(2)	-0.373(8)

[a] The CO bond lengths are set to 1.215 Å.
[b] Statistical errors (one standard deviation) are given in parentheses .

Table 2.2.4 summarizes the results of the DMC calculations on CO_2 and CO_2^-, as well as the energies of the various single determinant wave functions. It is seen from the table that for neutral CO_2, the DMC calculations using trial functions of DFT orbitals (B3LYP, PBE0, or PBE), give comparable or slightly lower energies than obtained using a trial function represented in terms of Brueckner orbitals. In contrast, for the anion, the DMC calculations using trial function in terms of Brueckner orbitals give appreciably lower energies than the calculations using trial functions represented in terms of DFT orbitals, with the energy difference being much more pronounced for the case of PBE than for the B3LYP or PBE0 orbitals.

2.2.4 Conclusion

In summary, we have demonstrated that DMC calculations using trial functions with a single Slater determinant of Brueckner orbitals gives atomization energies of a set of test diatomics much closer to experiment than DMC calculations employing Slater determinants of HF orbitals. When pseudopotentials are employed the DMC energies obtained using Brueckner orbitals are essentially identical to those employing PBE0 orbitals and give lower energies than

those using Becke3LYP orbitals. It is well known that DFT methods do not properly describe certain types of anions, and for these there can be a significant advantage to using Brueckner rather than DFT orbitals in the trial function. For example, for a bent CO_2^- test system, the DMC calculations using as a trial function Slater determinant of Brueckner orbitals give a lower energy than when employing DFT orbitals thereby establishing that the Slater determinant of Brueckner orbitals better describes the nodal surface in this case.

2.2.5 Acknowledgement

Acknowledgement: This research was supported by grant CHE136234 from the National Science Foundation. Computational resources were provided by the Center of Simulation and Modeling at the University of Pittsburgh, with hardware purchased under the NSF MRI award 1229064.

3.0 WEAKLY CORRELATED SYSTEMS

3.1 BENCHMARK STUDY OF THE INTERACTION ENERGY FOR AN $(H_2O)_{16}$ CLUSTER: QUANTUM MONTE CARLO AND COMPLETE BASIS SET LIMIT MP2 RESULTS

This work has been published as F. Wang, M.J. Deible, K.D. Jordan, "Benchmark Study of the Interaction Energy of an (H2O)16 Cluster: Quantum Monte Carlo and Complete Basis set Limit MP2 results," *J. Phys. Chem. A*, 2013, 117 (32), 7606. F.W. Performed the many-body and force-field calculations. M.J.D. Performed the QMC calculations. All authors contributed to the discussion.

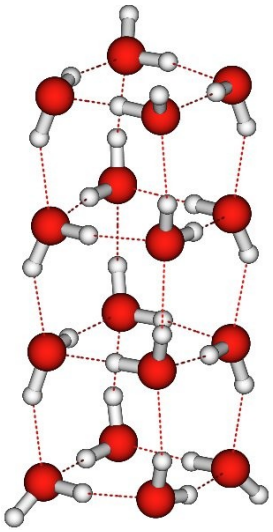
3.1.1 Introduction

In recent years, much attention has been devoted to the calculation of accurate interaction energies of water clusters.[99–102] The results of these studies have proven especially valuable in testing and refining force fields for describing water and for evaluating the performance of more approximate electronic-structure methods for describing hydrogen bonding. In this work, we demonstrate that the quantum Monte Carlo method is a viable method for predicting accurate interaction energies of $(H_2O)_n$ clusters for which large-basis set supermolecule CCSD(T)[103] calculations would be computationally prohibitive. As our test system, we choose

an $(H_2O)_{16}$ cluster that has been the subject of four recent theoretical studies.[73,104–106] For this cluster, we calculate the binding energy using the diffusion Monte Carlo (DMC) and complete-basis-set (CBS) limit MP2 methods. The resulting binding energies are compared with the *ab initio* results of Góra *et al*.[105] as well as with the results for several model potentials.

3.1.2 Test system and Methodology

Figure 3.1.1 - The 4444-a isomer of $(H_2O)_{16}$



The $(H_2O)_{16}$ cluster considered in this study is depicted in Figure 3.1.1. It is comprised of fused water cubes and is designated 4444-a using the nomenclature of Yoo *et al*.[104] Although such an arrangement of water molecules is not realized in any of the ices of water and is highly unlikely to be sampled in liquid water, it can be realized in appropriate diameter confining pores, *e.g*., in a (14, 0) carbon nanotube.[107] The geometry of the 4444-a cluster was optimized at the MP2/aug-cc-pVTZ[84] level by Yoo *et al*.,[104] and their geometry was employed in the present study as well as in the study of Góra *et al*.,[105] who estimated MP2 and CCSD(T) level interaction energies using an

N-body decomposition procedure.[108] The 4444-a cluster was also investigated by Wang and coworkers using their analytical WHBB potential[73] which also exploits *N*-body decomposition. Levenentz *et al.* employed the 4444a clusters in their study evaluating the performance of several density functional methods.[109]

The quantum Monte Carlo calculations of the interaction energy of 4444-a followed the usual two-step procedure of first doing a variational Monte Carlo (VMC) calculation followed by diffusion Monte Carlo calculations.[74] Specifically, for the geometry of interest, a B3LYP[110,111] calculation was carried out using a basis set formed by combining the s*p* functions from the aug-cc-pV5Z-CDF basis set[86] of Xu *et al*. augmented on oxygen with the *d* functions and on hydrogen with the *p* functions from the aug-cc-pVDZ-CDF basis set of these authors. In addition, the exponents of the most diffuse *d* functions on oxygen, and the most diffuse *p* function on hydrogen, were multiplied by a factor of 1.5 to minimize linear dependency problems. The Slater determinant of B3LYP orbitals was then multiplied by a three-term (electron-nuclear, electron-electron, and electron-electron-nuclear) Jastrow factor.[36] The parameters in the Jastrow factor were optimized by VMC minimization of the variance in the energy. The VMC wave function was then used as a trial function in the DMC calculations. The Dirac-Fock AREP pseudopotentials[85] were employed on the H and O atoms in each of the steps described above. The DMC calculations made use of the T-move procedure[35] to correct for errors due to non-locality of the pseudopotentials and were carried out for time steps of 0.0025, 0.005, and 0.0075 au. In order to calculate the binding energy of 4444-a, DMC calculations were also carried out on the water monomer using the MP2/aug-cc-pVTZ[104] optimized geometry. Sufficient numbers of moves were employed in the DMC calculations so that the statistical error in the extrapolated zero time-step binding energy was less than 1 kcal/mol.

In order to obtain the CBS-limit MP2-level interaction energy for the 4444-a cluster, Møller-Plesset perturbation[112] theory with density fitting[113,114] (DF-MP2)[115] calculations were carried out with the aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets,[43,84] with extrapolation to the CBS limit being accomplished using the approaches of Feller[116] and Helgaker *et al.*[117] for the Hartree-Fock and correlation contributions, respectively. These calculations did not apply a correction for basis set superposition error (BSSE). In addition, one- through three-body energies were calculated using the CCSD(T)-F12b method,[118] and one- through four-body energies were calculated using the DF-MP2-F12 method.[119] The VQZ-F12[120] basis set was used for the F12 calculations of the one- and two-body energies, and the VTZ-F12 basis set[120] was used in the F12 calculations of the three- and four-body energies. The *n*-body contributions were corrected for BSSE using the counterpoise method.[121] The MP2 and CCSD(T) calculations were carried out in the frozen-core approximation.

The B3LYP calculations for the generation of the trial functions were carried out using Gaussian 09,[92] the quantum Monte Carlo calculations were carried out using the CASINO[48] code. The MP2 and CCSD(T) calculations were performed using MOLPRO.[122]

3.1.3 Results

Table 3.1.1 - Binding energy (kcal/mol) of the 4444-a isomer of $(H_2O)_{16}$ obtained using different theoretical methods.

Method	Binding energy
DMC	-165.1(8)
MP2/CBS	-164.1
MP2 (Ref. 105)[a]	-161.6
CCSD(T) (Ref. 105)[b]	-162.8
CCSD(T) (Ref. 104)	-171.1
M06-2X (Ref. 123)	-172.0
M06-L (Ref. 123)	-164.5

[a] Estimated in Ref. 105 using one- and two-body CBS-limit MP2 energies and three- and four-body MP2 energies calculated with the aug-cc-pVDZ basis set

[b] Estimated in Ref. 105 as described in footnote *a* except that the CCSD(T) energies are used for the one- to three-body energies, and MP2 energies are used for the four-body energy.

Figure 3.1.2 - DMC results of the binding energy of 4444-a. Binding energies at each time step are shown as solid blocks with error bars. The dashed line shows the linear extrapolations.

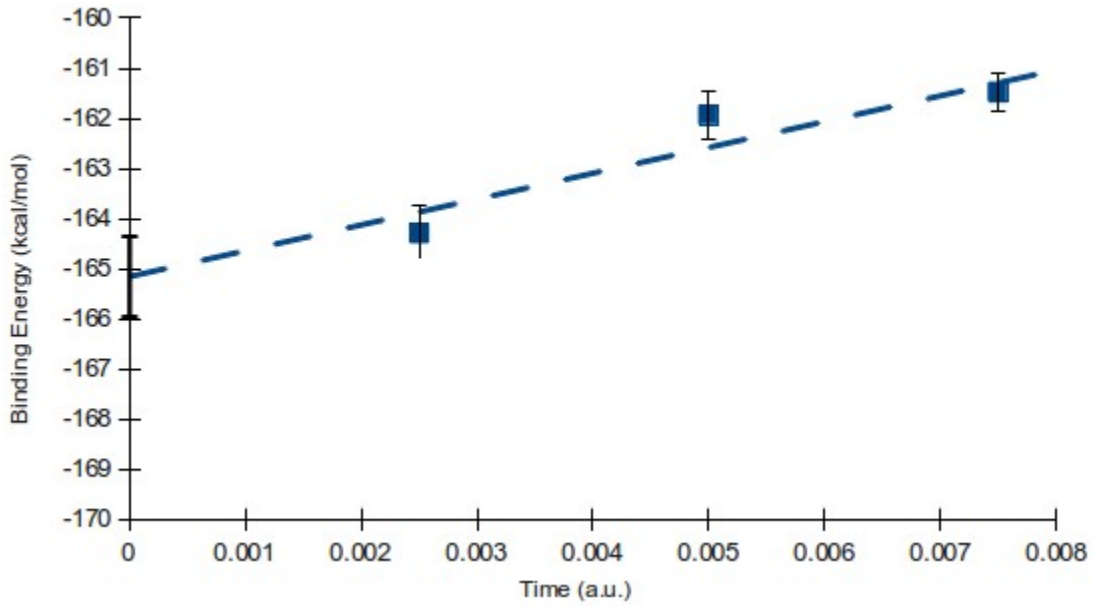


Figure 3.1.2 reports the results of the DMC calculations of the binding energy of 4444-a, and Table 3.1.1 summarizes the binding energies obtained using the different theoretical methods. The DMC value of the binding energy obtained from a linear fit to the results at the three time steps and extrapolation to zero time step is 165.1(8) kcal/mol, which is very close to our CBS-limit MP2 result of -164.1 kcal/mol. (A quadratic fit to the DMC data gives a binding energy greater in magnitude but with a much larger error bar.) Both the CBS-limit MP2 and DMC values of the binding energy are slightly larger in magnitude than the MP2 and CCSD(T) estimates (-161.6 and -162.8 kcal/mol, respectively) of Góra *et al.*[105] The CCSD(T) binding energy of Góra *et al.* was obtained by combining CBS-limit CCSD(T) one- and two-body energies with CCSD(T) three-body and MP2 four-body energies both calculated using the aug-

cc-pVDZ basis set. The MP2 interaction energy of Góra et al. was obtained in a similar manner, except that MP2 energies were used for the one- through four-body contributions.

Table 3.1.2 - Many-body interaction energies (in kcal/mol) of 4444-a.

	Ab initio Present Study			*Ab Initio* Ref. 105		Model potentials		
Contribution	HF[a]	MP2[b]	CCSD(T)[c]	MP2[d]	CCSD(T)[e]	TTM3-F	AMOEBA	WHBB
One-body	17.56	7.12	7.10	7.53	7.42	7.52	8.85	7.50
Two-body	-74.66	-133.95	-134.67	-134.37	-136.93	-147.83	-130.20	-134.00
Three-body	-34.41	-34.12	-32.71	-33.89	-32.42	-20.48	-36.63	-32.65
Four-body	-0.73	-0.89		-0.92		-0.83	-2.72	-0.83
Five-body	0.38					0.28	0.87	0.28
$N \geq 6$	-1.08					-0.02	-0.17	-0.02
Net	-92.94	-164.14		-161.65	-162.85	-161.36	-159.99	-159.74

[a] One- and two-body energies calculated using the VQZ-F12 basis set, and three-, four-, and five-body energies calculated using the VTZ-F12 basis set. The net interaction energy is at the HF/CBS level. The *N*-body energies include the couterpoise correction for BSSE.

[b] One- and two-body energies are at DF-MP2-F12/VQZ-F12 level, three- and four-body energies are at DF-MP2-F12/VTZ-F12 level, and the net interaction energy is at the DF-MP2/CBS level using a supermolecular calculation. The five- and higher-body interaction energy obtained from the DF-MP2 calculations is -2.30 kcal/mol. The *N*-body energies include the counterpoise correction for BSSE.

[c] One- and two-body energies calculated using the CCSD(T)-F12/VQZ-F12 method, three-body energy calculated using the CCSD(T)-F12/VTZ-F12 method. The *N*-body energies include the counterpoise correction for BSSE.

[d] One- and two-body energies from CBS-limit MP2 calculations, with the three- and four-body energies being calculated at the MP2/aug-cc-pVDZ level.

[e] One- and two-body energies from CBS-limit CCSD(T) calculations, three-body and four-body energies from CCSD(T)/pVDZ the MP2/aug-cc-pVDZ calculations, respectively.

[f] Using the WHBB water model (fifth-order fit for the three-body energies with the 5-6 cutoff) of Ref. 124.

Table 3.1.2 reports the *N*-body contributions to the interaction energy of 4444-a calculated using different theoretical methods. For the one-, two-, three-, and four-body energies, our DF-MP2 results are close to the MP2 results of Góra *et al*.[105] The sum of our one-, two-, three-, and four-body DF-MP2 energies is -161.84 kcal/mol compared to our CBS-limit DF-MP2 value of -164.1 kcal/mol for the net interaction energy. This leads us to conclude that at the MP2 level of theory, the five- and higher-body interactions contribute about -2.3 kcal/mol to the net interaction energy of 4444-a. Table 3.1.2 also lists the Hartree-Fock interaction energies through five-body contributions. The MP2 and Hartree-Fock methods give similar values of the three- and four-body interaction energies as expected based on previous studies of water clusters.[125–127] At the Hartree-Fock level of theory, the five-body energy is only 0.4 kcal/mol and the six- and higher-body energies combine to -1.1 kcal/mol. Assuming that the value of the five-body energy at the MP2 level is close to the Hartree-Fock result, this would imply that the $n \geq 6$ body interactions contribute -2.7 kcal/mol to the net interaction energy in the supermolecule MP2 calculations.

Our CCSD(T)-F12b value for the two-body energy is -134.67 kcal/mol, 2.26 kcal/mol smaller in magnitude than the CCSD(T) result of Ref. 105 which was obtained by extrapolating the results obtained using the aug-cc-pVTZ and aug-cc-pVQZ basis sets. To understand the source of this discrepancy we examined the sensitivity of the MP2 and CCSD(T) values of the binding energy of the water dimer to the procedure used to extrapolate to the complete basis set limit. Specifically, we did two-point extrapolations using the results using the aug-cc-pVTZ and aug-cc-pVQZ basis sets and three-point extrapolations using the results of the aug-cc-pVTZ, aug-cc-pVQZ, and aug-cc-pV5Z basis sets. The former extrapolation procedure is that used in Ref. 105. In addition, the dimer binding energy was also calculated using the MP2-F12/VQZ-

F12 and CCSD(T)-F12b/VQZ-F12 methods. The three-point extrapolated binding energies at both the MP2 and CCSD(T) levels are close to their MP2-F12 and CCSD(T)-F12b counterparts, leading us to conclude that these results are indeed close to their CBS-limit values. Whereas essentially the same value of the binding energy is obtained from the two- and three-point extrapolation methods when using the MP2 method, the dimer binding obtained using the two-point extrapolation procedure is about 0.05 kcal/mol larger in magnitude than the value obtained using the three-point extrapolation procedure when using the CCSD(T) method. Thus, it appears that the ~2.3 kcal/mol difference between the CCSD(T) two-body energy of 4444-a obtained in this study and that reported in Ref. 105 is mainly the result of the inadequacy of using only the aug-cc-pVTZ and aug-cc-pVQZ basis sets in the extrapolation to the CBS limit in Ref. 105

Reference [104] reported the total energy of 4444a at the CCSD(T)/aug-cc-pVTZ level of theory, and, to obtain the corresponding interaction energy we subtracted 16 times the energy of the monomer obtained at the same level of theory. The resulting CCSD(T) interaction energy is much larger (~7 kcal/mol) in magnitude than that obtained in Ref. 105 and in the present study. This is a consequence of the sizable BSSE due to the use of the aug-cc-pVTZ basis set. Table 3.1.1 also includes the results of DFT calculations using the M06-2X and M06-L functionals taken from Ref. 123, using the jun-cc-pVTZ basis set.[128] The authors of that study concluded that the M06-2X functional performed the best for calculating binding energies of water clusters. However, this conclusion was based on comparison with the results of CCSD(T) calculations of Ref. 104, which, as noted above, considerably overbinds due to BSSE. In fact, it is the M06-L functional that gives a binding energy closest to our DMC result.

Table 3.2.2 also includes results obtained using the AMOEBA,[129] TTM3-F,[130] and WHBB[124] water models. It is seen that while the AMOEBA model underestimates the magnitude

of the two-body interaction energy of 4444-a by nearly 2 kcal/mol, it overestimates the magnitude of the three- and four-body interaction energies by about 4.2 and 2.8 kcal/mol, respectively. This is consistent with an earlier observation that the AMOEBA model overestimates polarization.[131] The underestimation of the net two-body interaction energy in the AMOEBA model is due, in part, to its neglect of charge-transfer. The TTM3-F model overestimates the magnitude of the two-body energy by 13 kcal/mol and underestimates the magnitude of the three-body energy by a comparable amount. The TTM3-F and AMOEBA models give negligible contributions for the six- and higher-body interaction energies, although our MP2 calculations indicate these are sizable (~-2.7 kcal/mol).

We now turn to the results for the WHBB model which employs the Partridge-Schwenke[132] one-body potential and two- and three-body potentials fit to CCSD(T)/aug-cc-pVTZ energies at a large number of geometries. The $N \geq 4$ body contributions in the WHBB model are described using the TTM3-F force field. The WHBB model (fifth-order fit for the three-body energies with the 5-6 cutoff)[124] gives a net interaction of energy of 4444-a about 4.4 kcal/mol smaller in magnitude than our supermolecular DF-MP2 and DMC (linear extrapolation) values. Comparison of the individual *N*-body contributions from the WHBB model with those from CCSD(T) calculations (Table 2.1.2) reveals that 1.2 kcal/mol of discrepancy of the net interaction energies obtained with these two approaches derives from the one- through three-body contributions which should be better described in our study due to the use of the F12 procedure with the F12-VQZ basis set. The remaining ~3 kcal/mol discrepancy between the net interaction energy from the WHBB model and the net interaction energies from our MP2 and DMC calculations is due to fourth- and higher-order interactions that are not recovered by the TTM3-F model used by the WHBB potential to describe these contributions. Calculating the full

four-body energy of 4444-a at the CCSD(T)-F12b/VTZ-F12 method would be very demanding computationally. However, we did carry out CCSD(T)-F12b/VTZ-F12 calculations of ten tetramers extracted from 4444-a, and, for these tetramers, the values of the four-body interaction energies tend to be more negative when calculated with the CCSD(T)-F12b/VTZ-F12 method than with the MP2-F12/VTZ-F12 method, with the largest difference being 0.005 kcal/mol. Based on these results, we conclude that the CBS-limit CCSD(T) four-body energy of 4444-a could be a few tenths of a kcal/mol larger in magnitude than the corresponding MP2 result and that five- and higher-body contributions are the major cause of the difference between the net interaction energies from the WHBB model and our DMC result.

Figure 3.1.3 - Trimer extracted from 4444-a, used for energy decomposition analysis.

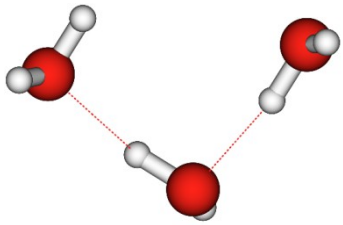


Table 3.1.3 - Contributions to the three-body energy (in kcal/mol) for the water trimer shown in Figure 3.1.3.

Contributions	Energy[a]
Exchange	-0.25
Induction[b]	-1.39
Charge-transfer[c]	-0.25
Polarization[c]	-0.89
δ(HF)	-0.25
Dispersion	0.13
NET	-1.51

[a] Calculated with the three-body SAPT(DFT) method.[45] [b]The induction and dispersion energies include the exchange-induction and exchange-dispersion, respectively. Also, the δ(HF) contribution of -0.25 kcal/mol has been absorbed into the induction term. [c]The three-body polarization energy was obtained by subtracting the ALMO-EDA three-body charge-transfer energy from the three-body induction energy.

In interpreting these results, we note that the TTM3-F and AMOEBA models, like most other polarizable force fields for water, describe the three- and higher-body interactions derive solely in terms of polarization. In fact, charge-transfer, exchange, and dispersion interactions, as well as various cross terms between these, all contribute to the three- and higher-body interaction energies. To some extent, exchange-polarization can be accommodated in an effective manner *via* the Thole-type[133] damping used in the TTM3-F and AMOEBA force fields. Table 3.1.3 reports the exchange induction, polarization, charge-transfer, and dispersion contributions to the three-body energy for a water trimer (shown in Figure 3.1.3) that was extracted from 4444-a. The three-body induction, exchange, and dispersion (including exchange-dispersion) contributions calculated using the three-body SAPT(DFT)[134] method are -1.14, -0.25, and 0.12 kcal/mol, respectively. Here the induction contribution includes both polarization and charge transfer, as well as their cross terms with exchange. Using an ALMO-EDA[135] analysis, the polarization and charge-transfer components of the induction energy are estimated to be -0.89

and -0.25 kcal/mol, respectively. Although the dissection of induction into polarization and charge-transfer components is sensitive to the procedure used, it is clear that the charge-transfer contribution to the three-body energy of this trimer is sizable. Based on the results of the SAPT and ALMO-EDA calculations, we conclude that polarization accounts for only about 75% of the net three-body energy of the selected trimer. Thus, it is not surprising that force field models that treat only the polarization part of the $N \geq 3$ body interactions are inadequate for quantitatively describing the many-body interactions in the 4444-a cluster. Indeed, it would be fortuitous if a force field including only polarization-type many-body interactions were to quantitatively reproduce the CCSD(T) value of the three-body energy.

The fact that the six- and higher-body interaction energies from the TTM3-F and AMEOBA models are only -0.02 and -0.17 kcal/mol, respectively, while the corresponding result from the Hartree-Fock calculations is -1.08 kcal/mol indicates that the $N \geq 6$-body exchange and/or charge transfer interactions, which are treated explicitly in the simple polarizable force field models, are significant in 4444a. As noted above, the N = 6-16-body interactions combined appear to be about twice as important in the MP2 than in the Hartree-Fock calculations, which raises the possibility that cross terms involving dispersion also play a non-negligible role.

In summary, for an $(H_2O)_{16}$ cluster that has been the subject of four other recent theoretical studies, we find that CBS-limit MP2 calculations give a binding energy of about -164.1 kcal/mol, 2.5 kcal/mol larger in magnitude than the N = 1~4 -body MP2 result of Góra et al.[105] Our DMC calculations give an interaction energy of $\approx$ -165 kcal/mol, as compared to the -162.8 kcal/mol N = 1-4 -body CCSD(T) result of Góra et al. Our calculations give the surprising result that the $N \geq 5$-body contributions are significant, being $\approx$ -2.3 kcal/mol at

the MP2 level. Based on the results of our DMC calculations, it appears that the $N \geq 5$ body interactions may be even more important when high-order correlation effects are included. The *ab initio*-based WHBB model of Bowman and coworkers[124] gives an interaction energy of the 4444-a isomer of $(H_2O)_{16}$, about 4.4 kcal/mol smaller in magnitude than the interaction energies obtained from our MP2 and DMC calculations. We conclude that the $N \geq 5$-body interactions are more important than indicated by the WHBB model.

Finally, we note that converging the DMC calculations reported in this study were carried out using 128 cores on a local computer cluster. Given the $O(N^3)$ scaling and high parallelization of the DMC method, it is clear that the DMC approach can be used to obtain accurate interaction energies of much larger water clusters for which large basis set supermolecule CCSD(T) calculations would not be feasible.

3.1.4 Acknowledgements

This research was carried out with the support of NSF grant CHE1111235. We also acknowledge a DOE INCITE award of computer time. We thank Drs. Wang and Bowman for valuable discussions about their WHBB model.

3.2 THEORETICAL STUDY OF THE BINDING ENERGY OF A METHANE MOLECULE IN A $(H_2O)_{20}$ DODECAHEDRAL CAGE

This work has been published as M.J. Deible, O. Tuguldur, K.D. Jordan, "Theoretical Study of the Binding Energy of a Methane Molecule in a (H2O)20 Dodecahedral Cage," J. Phys. Chem B, 2014, 118 (28),8257. MJD performed the DMC and many body calculations. OT performed the three-body SAPT calculations. All authors contributed to the discussion.

3.2.1 Introduction

It is estimated that there are about 10^{16} Kg of methane trapped in methane hydrate clathrate deposits on the ocean floor and in the permafrost.[136] As a result, methane hydrate has attracted considerable attention as a possible source of natural gas and because of the environmental consequences of its decomposition; the later concern derives from the fact that CH_4 is a potent greenhouse gas.

The most common form of methane hydrate crystal has a type I hydrate structure, with the unit cell consisting of two 5^{12} and six $5^{12}6^2$ water cages,[137,138] with a methane molecule in the center of each cage. The 5^{12} cage has a dodecahedral structure, while the 24 molecule $5^{12}6^2$ cage has 12 pentagonal faces and two opposing hexagonal faces. Numerous computational studies have been carried out on the properties of methane hydrate crystal (for example, see reference 139 and references therein) as well as on the $CH_4@(H_2O)_{20}$ gas-phase cluster with a methane encapsulated in an $(H_2O)_{20}$ dodecaheral cage.[140–143] The isolated $CH_4@(H_2O)_{20}$ system has been studied using a wide range of electronic-structure methods, with various dispersion-corrected density functional theory (DFT) methods giving binding energies between -4 and -7 kcal/mol.[140–143] The complete-basis-set (CBS) limit MP2 binding energy has been estimated to be -6.1 kcal/mol.[140] However, the extrapolation to the CBS limit in Ref. 140 was done using energies obtained with only the aug-cc-pVDZ and aug-cc-pVTZ basis sets,[84] a strategy which is

known to be inadequate.[144] In the force field studies of crystalline methane hydrate, it has generally been assumed that three- and higher-body interactions are not important for describing the interaction of the methane molecule with the water cage, although, in a paper from our group, it was reported that inclusion of polarization effects significantly impacts the thermal conductivity.[145]

The lack of agreement of the various theoretical results for the stability of a methane molecule in the $(H_2O)_{20}$ cage and the paucity of information on the role of three-body interactions on the binding of the methane in the water cage has motivated us to undertake diffusion Monte Carlo and near CBS-limit MP2 and MP2C[119,146] calculations of the binding energy as well as to calculate the two- and three-body contributions to the methane-$(H_2O)_{20}$ binding energy at various levels of theory, including CCSD(T)-F12,[118,147] DF-MP2-F12,[119] DF-MP2C-F12,[119,146] and symmetry-adapted perturbation theory (SAPT).[148–150]

3.2.2 Computational details

The dodecahedral water cage has 30026 symmetry distinct isomers with different arrangements of the protons.[151] In the present study, we employ the lowest energy isomer identified by Kirov et. al[152] which corresponds to structure 15 in a study by Wales and Hodges.[153] The geometry of the empty water cage and of the isolated methane molecule were optimized using second-order Møller-Plesset theory[112] with density fitting (DF-MP2)[115] together with the aug-cc-pVDZ basis set. The methane molecule was then placed in the water cage with the carbon located at the cage's center, and the orientation of the methane was optimized at the M06-2X[154]/aug-cc-pVDZ level of theory, keeping all other degrees of freedom frozen. The resulting geometry parameters are similar to those of earlier *ab initio* studies of $CH_4@(H_2O)_{20}$,[140] and are reported in the supporting information. These geometries were used for all subsequent calculations. The optimized structure of $CH_4@(H_2O)_{20}$ is shown in Figure 3.2.1.

Net interaction energies were calculated using:

$$E_{\text{int}} = E_{CH_4@(H_2O)_{20}} - E_{(H_2O)_{20}} - E_{CH_4} \qquad (3.2.1)$$

The DFT calculations made use of the BLYP,[110,155] M06-2X,[154] PBE,[90] and PBE0[156] functionals together with the aug-cc-pVTZ basis set and density fitting. The PBE0 functional is a hybrid functional with 25% exact exchange, and M06-2X is a hybrid meta functional with 54% exact exchange. The BLYP, PBE, and PBE0 calculations were carried out with and without the D3 dispersion correction of Grimme et al.[157] In addition, supermolecule calculations were carried out using the DF-MP2 and DF-MP2C-F12[146,158] methods, where DF refers to the use of density fitting,[113,114] and F12[118,119,147] refers to an explicitly correlated method that give energies that would otherwise require much larger Gaussian basis sets. The dispersion energy at the MP2 level can be shown to be equivalent to the use of uncoupled Hartree-Fock (HF) monomer polarizabilities in the Casimir-Polder expression.[146] The MP2C method replaces the uncoupled HF polarizabilites in the MP2 contribution to the dispersion energy with coupled Kohn Sham polarizabilities and, thus, can yield accurate interaction energies for systems for which the MP2 method fares poorly.[146] The DF-MP2C-F12 calculations were carried out using the aug-cc-pVTZ basis set.[84] For the DF-MP2 calculations, complete basis-set-limit results were obtained by extrapolating the energies from calculations using the aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets.[84] The methods of Feller[116] and Helgaker et al.[117] were used for extrapolation of the Hartree-Fock and correlation contributions, respectively. The interaction energies calculated using the various DFT and wavefunction methods listed above were corrected for basis set superposition error (BSSE) using the counterpoise method.[121] These calculations were carried out using the MOLPRO[122] code.

The diffusion Monte Carlo (DMC) method was also used to obtain an accurate value of the net binding energy. For the trial wave functions, Slater determinants of B3LYP[110,111] orbitals were generated using the Trail-Needs pseudopotential[85] on all atoms and the valence triple-zeta basis sets of Xu et al.[86] without the *f* functions or supplemental diffuse functions. These basis sets were designed for use with the Trail-Needs pseudopotentials. The Slater determinants were combined with Jastrow factors[36] with electron-electron, electron-nucleus, and electron-electron-nucleus terms, optimized *via* the variational Monte Carlo (VMC) procedure with variance minimization. The trial functions impose the fixed-node approximation, which should cause a negligible error in the binding energy. The T-

move[35] scheme was used to account for the non-locality of the pseudopotentials. Time step bias was removed by use of three time steps of 0.0025 a.u., 0.005 a.u., and 0.0075 a.u. for extrapolation to zero time step. The B3LYP calculations for the generation of the trial function were carried out with Gaussian 09,[92] and the DMC calculations were carried out with the CASINO[48] code.

The two- and three-body energies were calculated using:

$$\Delta E_2^j = \left[E(m,j) - E(m) - E(j) \right] \tag{3.2.2}$$

and

$$\Delta E_3^{(j,k)} = E(m,j,k) - E(m,j) - E(m,k) - E(j,k) + E(m) + E(j) + E(k), \tag{3.2.3}$$

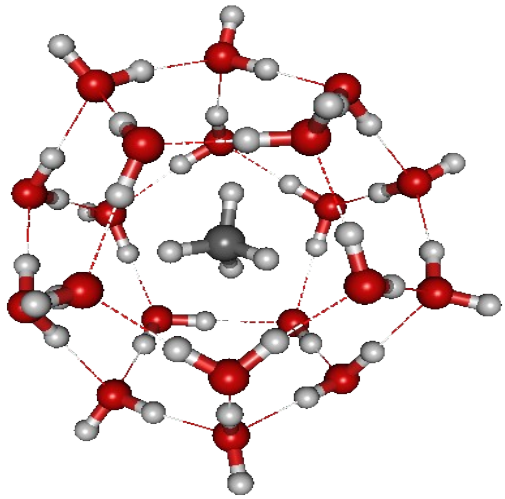
where *m* refers to the methane molecule, and *j* and *k* refer to water monomers. Because each monomer was held to a rigid geometry, there is no one-body contribution. The two- and three-body contributions to the binding energy were calculated at the CCSD(T)-F12b[118] and DF-MP2-F12[119] level with the VTZ-F12[120] basis set as well as with each of the density functional methods considered using the aug-cc-pVTZ basis set. The two- and three-body energies were corrected for BSSE using the counterpoise method. The two-body contribution was also calculated using the DF-MP2c-F12 method together with the VTZ-F12 basis set.

The two- and three-body energies were dissected into physical contributions by use of the DF-DFT-SAPT[148] and HF-based SAPT[149,150] methods, respectively. The DF-DFT-SAPT and HF-based SAPT calculations were carried out using the aug-cc-pVTZ basis set and aug-cc-pVDZ basis set respectively and, by design, are free of BSSE. The two-body SAPT calculations dissect the net two-body energy into electrostatics, exchange, induction, exchange-induction, dispersion, and exchange-induction contributions. There is also a so-called δHF contribution which we combine with induction and exchange induction to obtain an estimate of the net induction. For the two-body DF-DFT-SAPT calculations, the PBE0 functional[156] was used with an asymptotic correction for the ionization potential. The adiabatic local density approximation (ALDA)[159] kernel was used in the calculation of the response functions employed to evaluate the dispersion contribution. The three-body SAPT energies include exchange,

induction, exchange-induction, dispersion, and exchange-dispersion contributions. There are both second- and third-order contributions to the three-body induction contribution, and, again, there is a δHF correction which we incorporate in the net induction.

In calculating the net two-body SAPT contributions, two different strategies were pursued, one following the usual approach which involves summing the contributions for each methane-water dimer in the $CH_4@(H_2O)_{20}$ complex and the second treating the $(H_2O)_{20}$ cluster as a single molecule. For the first strategy, experimental IPs for methane and H_2O were used for the asymptotic correction. For the second strategy, the experimental IP of methane was again used, but the Hartree-Fock Koopmans' theorem[160] estimate of the IP was used in the asymptotic correction for the $(H_2O)_{20}$ cluster. The three-body SAPT energy was calculated by considering all methane-$(H_2O)_2$ trimers extracted from the $CH_4@(H_2O)_{20}$ system.

Figure 3.2.1 – Geometry of the methane hydrate structure studied. Red atoms are oxygen, white are hydrogen, and gray is carbon. (Color online.)



3.2.3 Results and discussion

Supermolecule interaction energies.

Figure 3.2.2 – Time step extrapolation of the DMC interaction energy. Interaction energy is solved for as in equation (1) at each time step (solid blocks, with error bars), and a linear fit is used to extrapolate to zero time step (dashed line).

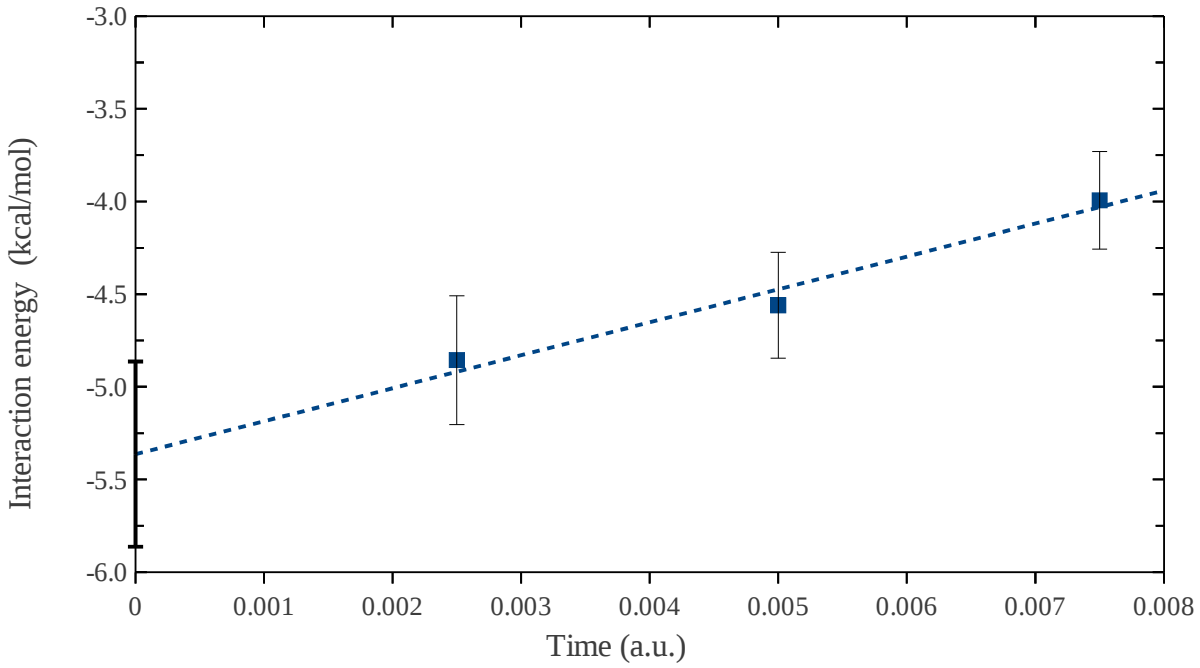


Table 3.2.1 – Energy (kcal/mol) for binding of a methane molecule in a $(H_2O)_{20}$ cage with the structure given in Figure 3.2.1.

Method[a]	Interaction energy
DF-HF[b]	4.13
PBE	1.31
PBE-D3	-6.61
PBE0	1.07
PBE0-D3	-6.89
BLYP	5.92
BLYP-D3	-6.72
M06-2X	-5.70
DF-MP2[b]	-5.04
DF-MP2C-F12	-4.60
DMC[c]	-5.3(5)

[a.] The DFT and DF-MP2C-F12 calculations used the aug-cc-pVTZ basis set and the binding energies include the counterpoise correction for BSSE. [b.] Extrapolated to the complete-basis-set limit as described in the text. [c.] The DMC calculations were carried out as described in the text.

Table 3.2.1 summarizes the net methane-$(H_2O)_{20}$ binding energies obtained at the various levels of theory. The DMC calculations give a binding energy of -5.3 $\pm$ 0.5 kcal/mol, which, to within statistical uncertainty, agrees with our CBS MP2 binding energy of -5.04 kcal/mol. (The data used in the extrapolation of the DMC results to zero time step are shown in Figure 3.2.2.) We note that our methane binding energy for the $CH_4@(H_2O)_{20}$ cluster system is also in excellent agreement with a recent DMC result of 5.6 $\pm$ 0.3 kcal/mol for crystalline methane hydrate.[161] Our best estimate of binding energy, derived from the *N*-body decomposition discussed below, is -5.2 kcal/mol. The close agreement of the MP2 result with the DMC and best estimate values is, in part, fortuitous as the MP2 method has errors of about +0.8 and -0.9 kcal/mol in the two- and three-body interactions, respectively. The MP2C-F12 method with the aug-cc-pVTZ basis set gives a net binding energy of -4.6 kcal/mol, but this result may be slightly underestimated in magnitude due to the basis set employed.

As noted in the Introduction, Kumar and Sathyamurthy[140] have reported a CBS MP2 value of -6.1 kcal/mol for the binding of a methane molecule in an $(H_2O)_{20}$ dodecahedral cage. This is significantly more attractive than our CBS-limit MP2 value of -5.04 kcal/mol. Much of the difference between these two CBS-limit MP2 results is likely due to differences in the geometries used in the two studies. (The key geometrical parameters are reported in the supporting information.) However, part of the difference between the two CBS-limit MP2 results could be a consequence of the different strategies used to extrapolate to the CBS limit in the two studies, with the extrapolation procedure used in the present study being expected to give more accurate results.

As seen from Table 3.2.1, of the DFT functionals considered, only the M06-2X functional gives a binding energy close to our best estimate -5.2 kcal/mol value. The PBE, PBE0, and BLYP functionals fail to give a bound complex, while with the inclusion of the D3 dispersion correction they overbind the complex by about 1.5 kcal/mol. It should be noted that the D3 corrections did not include a three-body Axilrod-Teller[162] contribution, which is repulsive.

Table 3.2.2 – Two body interaction energies (kcal/mol).

Method[a]	**Interaction energy**
DF-HF	3.85
PBE	-3.88
PBE-D3	-11.81
PBE0	-2.36
PBE0-D3	-10.28
BLYP	6.31
BLYP-D3	-6.34
M06-2X	-5.22
DF-MP2-F12	-4.95
DF-MP2c-F12	-5.54
CCSD(T)-F12b	-5.85

[a.]The HF and DFT calculations were carried out using the aug-cc-pVTZ basis set, while the DF-MP2C-F12, DF-MP2-F12, and CCSD(T)-F12b calculations were carried out using the VTZ-f12 basis set. All results are corrected for BSSE with the counterpoise method.

Table 3.2.3 – DF-DFT-SAPT energy (kcal/mol) decomposition of the two-body interaction energy of $CH_4@(H_2O)_{20}$.

Contribution [a]	treating each H_2O-methane pair separately	treating the $(H_2O)_{20}$ as a single molecule
$E_{es}^{(1)}$	-3.15	-3.08
$E_{exch}^{(1)}$	10.04	9.74
$E_{ind}^{(2)}$	-3.15	-2.29
$E_{exch-ind}^{(2)}$	1.97	2.24
$E_{\delta \mathrm{HF}}$	-0.38	-0.30
E_{ind}^{net}	-1.57	-0.36
$E_{disp}^{(2)}$	-12.64	-11.60
$E_{exch-disp}^{(2)}$	1.25	1.40
E_{disp}^{net}	-11.39	-10.19
E_{SAPT}	-5.88	-3.88

[a] The DF-DFT-SAPT calculations were carried out using the aug-cc-pVTZ basis set.

Two-body interaction energies.

Table 3.2.2 lists, for the various theoretical methods considered, the two-body contributions to the methane-$(H_2O)_{20}$ interaction energy. The CCSD(T)-F12 value for the two-body interaction energy, which is expected to be the most accurate result, is -5.85 kcal/mol. In comparison, the DF-MP2-F12 and DF-MP2C-F12 methods give two-body interaction energies of -4.95 and -5.54 kcal/mol, respectively. Thus, it is seen that the MP2 method significantly (by 0.9 kcal/mol) underestimates the magnitude of the two-body interaction energy, while the MP2C method fares much better.

The only DFT methods that give two-body contributions within 0.6 kcal/mol of the CCSD(T)-F12 result are BLYP-D3 and M06-2X which give two-body contributions of -6.34 and -5.22 kcal/mol, respectively. Both the PBE and PBE0 functionals give a bound $CH_4@(H_2O)_{20}$ complex at the two-body level, albeit underestimating the binding. In contrast, at the HF level of theory, the two-body contribution is repulsive by 3.85 kcal/mol. While some of the binding with the PBE and PBE0 functionals at the two-

body level could be due to their recovering short-range (ie., overlap dependent) intermonomer correlation effects, much of the binding found with these two functionals is due to their underestimating exchange-repulsion.[163] Not surprisingly, the PBE-D3 and PBE0-D3 methods give far too attractive two-body contributions to the binding energy.

The DFT-SAPT analysis of the two-body contribution to the binding energy is reported in Table 3.2.3, from which it is seen that the electrostatics, exchange-repulsion, induction, and exchange-induction contributions to the two-body energy are -3.15, 10.04, -3.15, and 1.97 kcal/mol, respectively. There is also a small δHF contribution of -0.38 kcal/mol. The net induction contribution, defined as $E_{ind}^{net} = E_{ind\delta HF} + E$, to the two-body interaction energy is -1.57 kcal/mol. Thus, two-body induction is surprisingly important in the interaction of the methane molecule with the $(H_2O)_{20}$ cage.

The sum of the DFT-SAPT interactions considered thus far is 5.71 kcal/mol, which is 1.86 kcal/mol more repulsive than the HF value of the two-body interaction energy. Thus, electron correlation effects significantly destabilize the electrostatics + exchange-repulsion + induction contribution to the two-body binding energy of $CH_4@(H_2O)_{20}$. The DFT-SAPT calculations give two-body dispersion and exchange-dispersion contributions of -12.64 and 1.25 kcal/mol, respectively. Adding these two contributions to the non-dispersion contributions discussed above, gives a net two-body contribution energy of -5.88 kcal/mol, nearly identical to the CCSD(T)-F12b result.

We also carried out DFT-SAPT calculations treating the entire $(H_2O)_{20}$ cage as a single molecule. The DFT-SAPT calculations treating the $(H_2O)_{20}$ as a single molecule give essentially the same electrostatics energy, and a value of the exchange-repulsion energies only 0.3 kcal/mol smaller than that obtained by treating the water molecules individually. On the other hand, the net induction and dispersion contributions are each about 1.2 kcal/mol less stabilizing in the former approach. The different induction and dispersion contributions obtained from the two types of "two-body" DFT-SAPT analysis can be understood in terms of the fact that treating $(H_2O)_{20}$ as a single molecule incorporates some contributions that would be considered many-body in a procedure where one builds up the cluster one

molecule at a time. We return to the issue of the similar exchange-repulsion interactions obtained using the two SAPT approaches described above after considering the three-body interaction energies.

Three-body interaction energies

Table 3.2.4 – Three-body contributions to the binding energy (kcal/mol) of a methane molecule in the $(H_2O)_{20}$ cage.

Method[a]	**30 H-bonded dimer pairs**	**160 dimer pairs without H-bonds**	**Total 3-body interaction energy**
HF	-1.48	1.20	-0.28
PBE	3.17	4.36	7.53
PBE0	1.50	2.95	4.45
BLYP	-2.37	-0.41	-2.79
M06-2X	-1.23	2.87	1.64
DF-MP2-F12	-0.96	1.21	0.251
CCSD(T)-F12b	-0.42	1.43	1.01
HF-SAPT energy decomposition			
$E_{exch}^{(1)}$	-1.31	-0.01	-1.32
$E_{ind}^{(2)}$	-0.05	0.85	0.80
$E_{ind}^{(3)}$	-0.41	0.01	-0.40
$E_{exch-ind}^{(2)}$	-0.02	0.06	0.04
$E_{\delta \mathrm{HF}}$	0.29	0.33	0.62
E_{ind}^{net}	-0.19	1.25	1.06
$E_{disp}^{(3)}$	0.41	0.22	0.63
$E_{exch-disp}^{(2)}$	0.71	0.13	0.84
E_{disp}^{net}	1.12	0.35	1.47
HF-SAPT	-0.37	1.57	1.21

[a]The HF-SAPT calculations were carried out using the aug-cc-pVDZ basis set. The DF-MP2-F12 and CCSD(T)-F12b calculations were carried out using the VTZ-f12 basis set and the DFT calculations were carried out using the aug-cc-pVTZ basis set. These results include the counterpoise correction for BSSE.

The three-body interaction energies are summarized in Table 3.2.4. The net three-body interaction energy is calculated to be -0.28, 0.25 and 1.01 kcal/mol at the HF, DF-MP2-F12, and CCSD(T)-F12b levels of theory, respectively. The CCSD(T)-F12b value of the net three-body energy is 0.76 kcal/mol more repulsive than the corresponding MP2-F12 value, consistent with the importance of the Axilrod-Teller three-body dispersion contribution which appears at third-order perturbation theory. The various DFT methods give values of the three-body interaction energies ranging from -2.79 to 7.53 kcal/mol, with only the M06-2X functional giving a three-body interaction energy within 1 kcal/mol of the CCSD(T) result.

Table 3.2.4 also summarizes the results of the three-body SAPT calculations on the $CH_4@(H_2O)_{20}$ system. The exchange, induction, and exchange-induction three-body contributions are -1.3, 0.4, and 0.0 kcal/mol, respectively, while the δHF contribution is 0.6 kcal/mol. Combining the induction, exchange-induction, and *δHF* contributions, we obtain a net three-body induction contribution of 1.1 kcal/mol for the binding of the methane molecule in the $(H_2O)_{20}$ cage. Recalling that the net two-body induction contribution was about -1.6 kcal/mol, we see that the combined two- plus three-body induction contribution is only -0.5 kcal/mol. The three-body dispersion and exchange-dispersion contributions are 0.6 and 0.8 kcal/mol, respectively. The overall three-body contribution to the methane binding obtained using the SAPT method is 1.3 kcal/mol in reasonable agreement with to the CCSD(T) result of 0.9 kcal/mol. If we simply add the SAPT three-body dispersion and exchange-dispersion contribution to the DFT three-body energies, the PBE and PBE0 results would be even further removed from the CCSD(T)-F12b result, while the BLYP result for the three-body interaction energy would still differ from the CCSD(T)-F12b result by 2.4 kcal/mol. It is clear that some deficiency other than the neglect of long-range dispersion interactions is responsible for the large errors in the PBE, PBE0, and BLYP three-body contribution of the binding of CH_4 in the $(H_2O)_{20}$.

It is instructive to further decompose the three-body contributions into two parts, that due to the thirty trimers with the two water monomers H-bonded to one another, referred to as set A, and that due to

the 160 trimers without H-bonding between the two water monomers, referred to as set B. From Table 3.2.3, it is seen that while the CCSD(T) three-body energy is -0.4 kcal/mol for set A trimers, it is 1.5 kcal/mol for set B. The corresponding MP2-F12 results are -1.0 and 1.2 kcal/mol. None of the DFT methods considered closely reproduces the CCSD(T) values of the three-body interaction energy of either the A- or B-type trimers. Table 3.2.4 also reports the three-body SAPT contributions for the two types of trimers. From the Table, we see that the three-body exchange contribution derives almost entirely from the set-A trimers, while the three-body induction (including the δHF term) is dominated by the set B trimers. The three-body dispersion is dominated by the A-type trimers.

It was noted in the previous section that the differences in the SAPT values of the two-body induction, exchange, and dispersion contributions to the methane binding energy as calculated treating each monomer separately and treating the $(H_2O)_{20}$ cage as a single molecule are 1.24, -0.30, and 1.18 kcal/mol, respectively. From Table 3.2.4, it is seen that for induction and dispersion these differences are comparable to the corresponding three-body contributions (calculated using all methane-$(H_2O)_2$ trimers). This is consistent with the fact that the DFT-SAPT "two-body" calculations, treating the entire $(H_2O)_{20}$ as a single monomer include a subset of the n $\geq$ three-body interactions as evaluated treating the water monomers as separate molecules. On the other hand, the three-body exchange-repulsion contribution of -1.32 kcal/mol, is about 1.0 kcal/mol larger in magnitude than the difference of the two-body exchange contributions calculated using the two strategies described above. Examination of the various contributions to the three-body exchange energy reveals that about half of this discrepancy is due to the three-body exchange contributions that are not recovered in the "two-body" SAPT calculations treating the $(H_2O)_{20}$ as a single monomer.

Table 3.2.5 – Energy (kcal/mol) of the n-body decomposition.

		N-body contributions		
Method	**Full**	**2**	**3**	***N*≥4**
HF	4.13	3.85	-0.28	0.56
PBE	1.31	-3.88	7.53	-2.34
PBE0	1.07	-2.36	4.45	-1.02
BLYP	5.92	6.31	-2.79	2.40
M06-2X	-5.70	-5.22	1.64	-2.12
DF-MP2-F12	-5.04[a]	-4.95	0.25	-0.34
CCSD(T)-F12b	(-5.3)[b]	-5.85	1.01	(-0.46)[c]

[a.] CBS-limit DF-MP2 result

[b.] DMC result

[c.] Estimated using the DMC value of -5.3 for the full interaction energy.

Table 3.2.5 summarizes the two-, three-, and higher-body contributions for the binding of a methane molecule in the $(H_2O)_{20}$ cage. At the HF and DF-MP2 levels of theory, the $N \geq 4$ contributions are only 0.6 and -0.3 kcal/mol, respectively. CCSD(T) calculations for the entire complex with the basis sets used here would be computationally prohibitive. However, if we use the DMC result for the net binding energy we can obtain an estimate of the CCSD(T)-F12 higher-body contribution to the binding energy. Using this strategy, we obtain a value of -0.5 kcal/mol for the $N \geq 4$ contribution to the binding energy. However, this is subject to a ±0.5 kcal/mol statistical uncertainty due to the uncertainty in the DMC value of the net binding energy. In contrast to the small $N \geq 4$-body contribution obtained using the wave function methods, the PBE, M06-2X, and BLYP density functional methods give higher-body contributions of -2.3, -2.1, and 2.4 kcal/mol, respectively. The PBE0 method, on the other hand, gives a higher-body interaction energy of -0.9 kcal/mol, consistent with our expectation that self-interaction is primarily responsible for the large overestimation of the three-body energy by the PBE method. The MP2 and CCSD(T) results presented in Table 3.2.5 can be combined to obtain an improved estimate of the net binding energy. In particular, by adding the differences of the CCSD(T) and MP2 values of the two- and

three-body interaction energies to the CBS-limit DF-MP2 value of the net binding energy we obtain a value of -5.2 kcal/mol, which is nearly identical to the DMC result.

3.2.4 Conclusion

The binding energy of a methane molecule in a $(H_2O)_{20}$ dodecahedral cage was calculated using a variety of electronic structure methods. Diffusion Monte Carlo calculations give a binding energy of -5.3 ±0.5 kcal/mol, in excellent agreement with our best estimate value of -5.2 kcal/mol, obtained by correcting the CBS-limit MP2 result with the CCSD(T)-F12b - MP2-F12 differences for the two- and three- body binding energies. Of the density functional methods tested, only M06-2X gives a binding energy within 1 kcal/mol of our best estimate value. The PBE-D3, PBE0-D3, and BLYP-D3 methods overbind the methane molecule by 1.5-1.8 kcal/mol.

A SAPT analysis reveals that exchange, induction, and dispersion all make important contributions to the three-body interaction energy. Thus, for force field methods to accurately describe the interaction of a methane in an $(H_2O)_{20}$ cage, it will be necessary to include explicit terms for three-body exchange, induction, and dispersion. However, because the net three-body exchange contribution is negative and the three-body induction and dispersion contributions are positive and the three terms are roughly comparable in magnitude, a force field with only induction or dispersion, for three-body interactions could fortuitously give a three-body energy close to the *ab initio* result. We also find that none of the density functional methods considered fare well at predicting the two-, three-, and higher-body contributions to the binding energy. It is clear from comparison of the DFT and SAPT results for the two- and three-body contributions to the binding energies that the DFT methods have shortcomings other than those associated with the neglect of long-range dispersion interactions. Strikingly, with the PBE functional, the three-body contribution to the binding energy of the $CH_4@(H_2O)_{20}$ is too large by a factor of seven. To a large extent this is a result of self-interaction error in the DFT methods. The failure of standard DFT methods to accurately describe the terms in the *N*-body expansion of water clusters and

ice has been noted in other recent studies.[73,125] In addition, in a very recent study Cox and co-workers reported that none of the density functional methods that they examined performed well for the methane hydrate crystal.[161]

3.2.5 Acknowledgements

This research was carried out with the support of NSF grant CHE1111235. An award of computer time was provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357. We thank Dr. Mike Gillan for valuable discussions.

3.3 THEORETICAL STUDY OF CARBON DIOXIDE HYDRATE

3.3.1 Introduction

Clathrate hydrates can form at low temperature and moderate pressure, where a ice-like solid is formed with an encapsulated guest molecule in an appropriate water cage. The non-covalent interactions of the guest gas molecule and the host water cage are strong enough to affect the structure of the cage and the stability of the clathrate.[164,165] This has lead to the proposal of mitigating the greenhouse effect through carbon dioxide sequestration by clathrate formation.[166] In a similar vein, the replacement of the methane in methane clathrates with carbon dioxide would simultaneously release valuable natural gas and store carbon dioxide.[167] Therefore, it is relevant to understand the nature of the interaction of a CO_2 molecule encapsulated in a water dodecahedral cage.

Previously, this system has been investigated with molecular dynamics[168] and force fields or the interaction energy has been studied with density functional theory (DFT).[169] Kumar and Sathyamurthy[140] have used a high level of theory, namely Møller-Plesset second order perturbation theory (MP2) at the complete basis set limit (CBS), to estimate the binding energy of the CO_2 in the $(H_2O)_{20}$ cage. This study predicted an interaction energy of a CO_2 encapsulated in a $(H_2O)_{20}$ cage of -9.18 kcal/mol. However, this is based on an extrapolation that only considers the double- and triple-zeta basis set, which is known to be inadequate.[144] It should also be pointed out that the authors gave an estimate of the interaction energy of -6.01 kcal/mol

when an aug-cc-pvdz basis set is used and corrected for basis set superposition error. Additionally, various DFT estimates have ranged from 3.47 to -13.38 kcal/mol for the interaction energy. Clearly, a more accurate picture of the interaction energy for this important system is required in order to provide reliable benchmarks for DFT and force fields used in molecular dynamics. To this end, we have undertaken diffusion Monte Carlo calculations of the supermolecular interaction energy of CO_2 in the dodecahedral $(H_2O)_{20}$ cage as a benchmark, and compared it to MP2 calculations at the complete basis set limit using aug-cc-pvdz, aug-cc-pvtz, and aug-cc-pvqz basis sets for the extrapolation. Additionally, we have used MP2 and the "gold standard" coupled cluster with singles, doubles, and perturbative triples with an explicitly correlated method in a many body decomposition procedure to determine the role of two- and three- body interactions to the net binding energy of the CO_2 in the dodecahedral water cage. Symmetry adapted perturbation theory is used at the two- and three-body level to lend insight into the contributions to the interaction energy.

3.3.2 Computational details

The geometry of the water cage is the same used in a previous study of methane clathrate.[170] This is the lowest energy water dodecahedron identified by Kirov et. al.[152] The geometry of the CO_2 clathrate was found by placing the carbon atom of CO_2 at the center of mass of the cage and fixing the carbon-oxygen bond length at the experimental value of 1.162 Å. The orientation of the the CO_2 was optimized at the M06-2X[154] level of theory with the aug-cc-pvdz basis set.[171] This geometry was used for all subsequent calculations.

The DFT calculations made use of the BLYP,[88,155] M06-2X, and PBE[172] functionals with the aug-cc-pvtz basis set and were corrected for basis set superposition error (BSSE) using the

counterpoise correction.[121] For the supermolecular calculations, DF-MP2 calculations were carried out and extrapolated to the complete basis set limit with the methods of Feller[116] and Helgaker et. al.[117] DF refers to the use of density fitting.[113,114] To evaluate the two- and three-body interaction energies, the CCSD(T)-F12b[118] method was used along with the DF-MP2-F12 method, where F12 refers to an explicitly correlated method.[118,119,147] At the *N*-body level, the DF-MP2-F12 method and MP2C-F12 method, in which the uncoupled polarizability of the MP2 method is replaced by the coupled polarizability from a time-dependent DFT calculation, and CCSD(T)-F12b calculations all make use of the vtz-f12[120] basis set and are corrected for BSSE. The above calculations were carried out with the MOLPRO code.[122]

Symmetry adapted perturbation theory based on density functional theory with density fitting, DF-DFT-SAPT,[148] was used at the two-body level with the aug-cc-pvtz basis set and is free of BSSE. The PBE0[173] functional was used with an asymptotic correction for the ionization potentials of the monomers. The experimental IPs for water and carbon dioxide were taken from http://cccbdb.nist.gov. The adiabatic local density approximation kernel[159] was used in the calculation of the response functions to estimate the dispersion contribution. The three-body SAPT[174] is based on Hartree-Fock and is carried out with the SAPT program;[149,150] the two-body SAPT is carried out with the MOLPRO code.

The diffusion Monte Carlo (DMC) calculations used trial wave functions composed of a single Slater determinant of B3LYP[87,88] orbitals and a three term Jastrow factor[36] with parameters optimized via variance minimization. The trial functions were generated in the Gaussian09B[92] software package and used the pseudopotential of Trail and Needs[85] along with the corresponding triple-zeta basis set of Xu et al. The basis sets consisted of the quintuple-zeta *s* and *p* functions and the triple-zeta *d* functions without the diffuse augmented functions, as

described previously.[13] The localization of the pseudopotential was treated with the T-move scheme.[35] Time step bias was removed via extrapolation to zero time step using a linear fit to the 0.0025, 0.005, and 0.0075 a.u. time steps. The quantum Monte Carlo calculations were carried out using the CASINO[48] code.

3.3.3 Results and discussion

Table 3.3.1 - Interaction energy in kcal/mol for the CO_2 in the dodecahedral water cage using the supermolecular and two-body schemes.

supermolecule	
DF-HF[a]	6.26
DF-MP2[b]	-6.52
M06-2X[a]	-5.95
BLYP[a]	8.12
DMC	-5.4(4)
2-body	
DF-HF[c]	6.12
MP2-F12[c]	-6.05
MP2C-F12[c]	-5.58
CCSD(T)-F12b[c]	-6.41
BLYP[a]	8.21
M06-2X[a]	-5.27
DF-DFT-SAPT[a]	
$E^1_{Electrostatic}$	-6.06
$E^1_{Exchange}$	16.33
$E^2_{Induction}$	-6.60
$E^2_{Exchange\text{-}Induction}$	4.90
δHF	-0.67
Net induction	-2.37
$E^2_{Dispersion}$	-15.87
$E^2_{Exchange\text{-}Dispersion}$	1.72
Net dispersion	-14.16
DF-DFT-SAPT	-5.81

[a.] aug-cc-pvtz basis set.

[b.] Complete basis set limit.

[c.] vtz-f12 basis set.

Table 3.3.2 - Three-body interaction energy in kcal/mol for the CO_2 in the dodecahedral water cage.

3-body			
	H-Bonded dimer	Not H-Bonded dimer	Total
HF[a]	-1.84	1.43	-0.41
MP2-F12[a]	-1.19	1.56	0.37
CCSD(T)-F12b[a]	-0.56	1.80	1.24
BLYP[b]	-2.49	-0.11	-2.60
M06-2X[b]	-2.11	3.45	1.34
HF-SAPT[c]	-0.21	2.05	1.39
HF-SAPT[c]			
$E^1_{Exchange}$	-1.79	0.02	-1.76
$E^2_{Induction}$	-0.17	1.16	0.99
$E^3_{Induction}$	-0.46	0.09	-0.37
$E^2_{Exchagne\text{-}Induction}$	-0.01	0.02	0.02
δHF	0.55	0.26	0.35
Net induction	-0.08	1.53	0.99
$E^3_{Dispersion}$	0.75	0.35	1.10
$E^2_{Dispersion\text{-}Exchange}$	0.91	0.15	1.06
Net dispersion	1.66	0.49	2.16
HF-SAPT	-0.21	2.05	1.39

[a.] vtz-f12 basis set.

[b.] aug-cc-pvtz basis set.

[c.] aug-cc-pvdz basis set.

The results of the supermolecular calculations are given in Table 3.3.1. The DMC benchmark energy -5.4(4) kcal/mol is slightly lower than the complete basis set limit MP2 value

of -6.52 kcal/mol. This deviation will be discussed below, in the context of the *N*-body decomposition. Neither the HF nor BLYP method predict a repulsive interaction between the CO_2 and the water cage. On the other hand the M06-2X method gives a binding energy -5.95 which is close the DMC benchmark value.

The two-body interaction energy is also given in Table 3.3.1. Here, the CCSD(T)-F12b value of -6.41 kcal/mol is taken as the benchmark. The HF and BLYP methods give two-body energies nearly equal to the full supermolecule results. The M06-2X and the MP2-F12 method underestimates the two-body energy by 1.14 and 0.36 kcal/mol, respectively. Turning to the DF-DFT-SAPT results, the net induction energy is surprisingly important to the stabilization of the cluster, contributing -2.37 kcal/mol to the interaction energy. The sum of the induction, electrostatic, and exchange two-body contributions is 7.9 kcal/mol, which 1.8 kcal/mol more repulsive than the corresponding DF-HF result but very close to the BLYP result. The net DF-DFT-SAPT interaction energy is -5.81, 0.6 kcal/mol smaller in magnitude than the CCCSD(T)-F12b result. This may be a result of the dispersion contribution being to small in magnitude. This is corroborated by the MP2C-F12 method, which is 0.83 kcal/mol smaller in magnitude than the CCSD(T)-F12b result and 0.47 kcal/mol smaller in magnitude than the MP2-F12 two-body energy.

The three-body interaction energies is broken down into two groups, those that involve water dimer pairs that are hydrogen bonded and those that do not. These results are given in Table 3.3.2. Adding the CCSD(T)-F12b energy for the two- and three-body decomposition gives a binding energy for the supermolecular system of -5.17 kcal/mol, in excellent agreement with the DMC estimate of -5.4(4) kcal/mol. The DF-MP2-F12 method gives a three-body energy of 0.37 kcal/mol as compared to the CCSD(T)-F12b result of 1.24 kcal/mol. This is largely due to

the MP2 method failing to recover the three-body Axilrod-Teller[162] type dispersion effects. If we add to the supermolecular MP2-CBS interaction energy a correction derived from the difference of the MP2-F12 method and CCSD(T)-F12b at both the two- and three-body level, we get an estimate of -6.01 kcal/mol for the net binding energy, in reasonable agreement with the DMC result of -5.4(4) kcal/mol. The BLYP functional over binds both types of trimers. When considering the interaction energy of all 190 water dimers and the CO_2, the M06-2X results look promising giving a three-body energy within 0.1 kcal/mol of the CCSD(T)-F12b result. However, upon inspection of the interaction energy by water dimer type, it is clear that this result is due to a cancellation of errors, as the trimers with hydrogen bonded water monomers are too strongly bonded, while the interaction in the trimers without the hydrogen bonded water momoners are predicted to be too repulsive. Turning to the HF-SAPT results, there is good agreement for the trimers with hydrogen bonded and non-hydrogen bonded water monomers. The exchange and dispersion energy are dominated by the trimers with hydrogen bonded dimers, but the trimers with non-bonded water dimers are the largest contributer to the induction energy. It is striking to note that, even at the three-body level, the induction contributes about ~1 kcal/mol to the net interaction, and the dispersion and exchange contribute 2.16 and -1.76 kcal/mol, respectively.

In summary, the diffusion Monte Carlo method gives a binding energy of -5.4(4) kcal/mol for a CO_2 in the $(H_2O)_{20}$ cage. The *N*-Body decomposition scheme was used to further analyze the interactions. CCSD(T)-F12b two- plus three-body energies result in a net binding energy of -5.17 kcal/mol, in close agreement with the DMC result. None of the three DFT methods considered properly describes the individual *N*-body contributions. The SAPT method was used at both the two- and three-body level to determine the contributions to the binding

energy. This shows that at the two-body level, exchange and dispersion have large contributions of 16.33 and -14.16 kcal/mol, respectively. The electrostatics and induction are also important to the net interaction energy but give slightly smaller contributions of -6.06 and -2.37 kcal/mol, respectively. At the three-body HF-SAPT level, the energy contributions are much smaller in magnitude but similar trends remain. The dispersion and exchange give the largest contributions to the energy of -1.76 and 2.16 kcal/mol, respectively, while the induction contribution is only 0.99 kcal/mol.

4.0 STRONGLY CORRELATED SYSTEMS

The DMC method is can make use of a trial wave function comprised of many Slater determinants to give a better description of the nodal surface. Recent algorithm improvements have made the evaluation of many determinant trial wave functions more computationally efficient.[19,175,176] Still, important questions remain about the implementation of multideterminant trial wave functions in DMC. As is stated in section 1.2.2, there is a well know size consistency problem in truncated CI calculations, and it is unclear what the consequences of this are on a DMC calculation. A full CI calculation, which will not have a size consistency error, in a limited number of orbitals will produce $(2K!)/[N!(2K-N)!]$ determinants for N electrons and K orbitals. Even with improved algorithms offering computational efficiency, DMC simulations on systems with more than a few atoms cannot include all of the determinants of the trial wave function. One possibility for selecting a subset of determinants to keep from the trial wave function is to apply a threshold to the CI coefficients of the trial wave function. The implications for the nodal surface of retaining some determinants via a CI cutoff and rejecting others remains an active area of research.

There have been several successful DMC calculations on small systems that use trial wave functions comprised of many determinants. There is no hard definition for when a system requires a multi-configurational trial wave function, though as a general trend it is seen that adding more determinants results in a lower DMC energy.[17,19,98,177,178] In this chapter, we will

investigate the effect of the multi-reference trial functions on the DMC energy for simple test systems and compare it to other benchmark methods. In the first subsection, two systems will be studied; each has a ground state that can be easily tuned from being well described by a single Slater determinant to requiring many determinants. The square H_4 system will give an indication of how well a single reference trial wave function performs for DMC, and the interaction energy of the ethylene dimer will show how well a single reference describes weakly interacting systems as the degeneracy is increased. In the second section, a prototypical multi-reference system, the interaction energy of beryllium dimer, is investigated with DMC. This system has been studied several times with quantum Monte Carlo.[15,179,180] We obtain the closest agreement with experimental results for the binding energy achieved with DMC to date.

4.1 H_4 AND THE TWISTED ETHYLENE DIMER

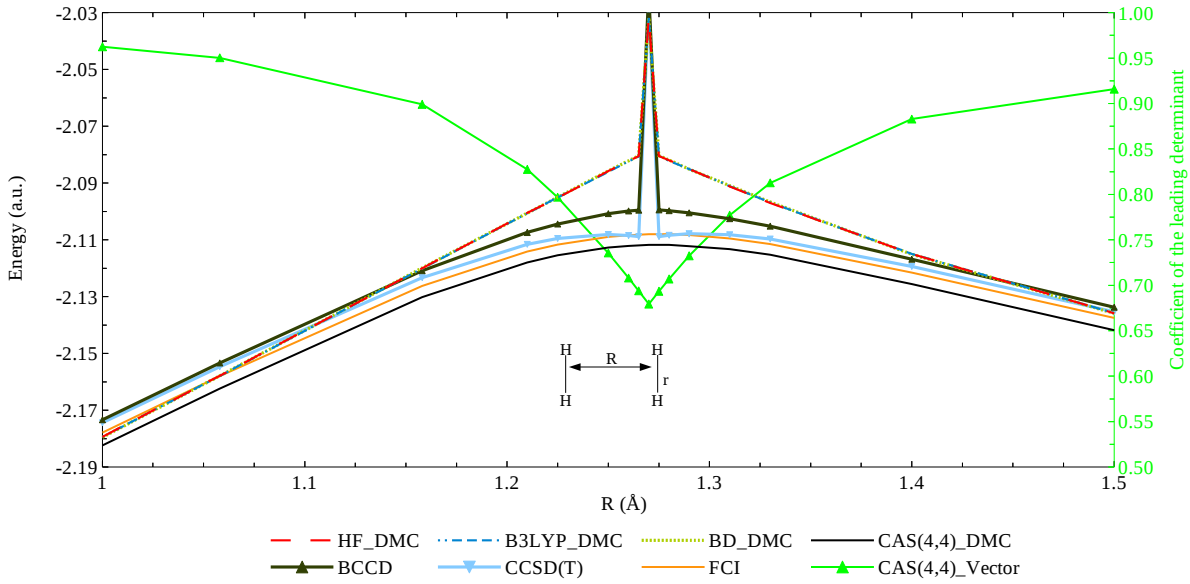
4.1.1 Introduction

The majority of DMC calculations employ a single-determinant trial wave functions. It has been shown [17,96–98] that this approximation cannot reach chemical accuracy for a large test set of atomization energies of polyatomic molecules, regardless of orbital choice, but that a trial wave function of many determinants can reach chemical accuracy.[17,98] However, for some systems dominated by weak interactions, a single-determinant can achieve sub-chemical accuracy.[181] This raises the important question of whether this is always true for weakly interacting systems.

An initial test presented here will use a system of two H_2 molecules as a function of their separation to tune the degeneracy of the wave function to benchmark DMC with a single or

multi-reference trial wave function against standard methods. To determine the effect of a degenerate ground state on weakly interacting systems, a second test system will be considered: stacked ethylene dimers rotated around the π-bond to tune the degeneracy.

Figure 4.1.1 – Energy of the square hydrogen system studied with several methods. The inset shows the geometry. The H_2 bond length (r) is set to 1.27 Å.



4.1.2 Computational details

For the square hydrogen system, the geometrical parameters are given in the inset of Figure 4.1. The distance R between the molecules was varied to tune the degeneracy of the ground state. The trial wave functions were generated with the Gaussian 09 code. The pseudopotential of Trail and Needs[85] was used together with the basis set of Xu et al.[86] that was formed by taking the

quintuple-zeta *s* functions and the triple-zeta *p* functions. The T-move[35] scheme with an asymmetric branching factor is used for the DMC calculations. All calculations employed e-e, e-n, and e-e-n Jastrow factors optimized by minimization of the energy. For the multi-determinant calculations, determinant coefficients were optimized simultaneously with the parameters of the Jastrow factors[36] via energy minimization. All QMC calculations were carried out with the CASINO code.[48] The CCSD(T) and full CI (FCI) calculations were carried out using the MOLPRO[122] code with the cc-pVTZ Dunning basis set.[84]

For the ethylene dimer, the geometry of the monomer is taken from the S-22[182] set and replicated in the face-to-face configuration to form the dimer. The TN pseudopotential was used, and the basis set uses the quintuple-zeta *s* and *p* functions and triple-zeta *d* functions of Xu et al. for carbon, and the same basis set for hydrogen described above. The T-move scheme with a symmetric branching factor[66] is used for the DMC calculations. The trial wave functions were generated with the Gaussian 09 code. All calculations employed e-e, e-n, and e-e-n Jastrow factors optimized by minimization of the energy. For the multi-determinant calculations, determinant coefficients were optimized simultaneously with the Jastrow factors *via* energy minimization. CASSCF calculations used an active space comprised of the π electrons and π^* orbitals; thus, a CAS(4,4) was used for the dimer calculations, and a CAS(2,2) was used for the monomer calculations. All QMC calculations were carried out with the CASINO code. The time-steps used were 0.01, 0.025, 0.05, and 0.1, and were extrapolated to zero time step by a quadratic fit. The CCSD(T)-F12a, MP2C-F12, and MP2-F12 calculations were carried out using the MOLPRO code, were corrected for basis set superposition error (BSSE) with the counterpoise correction,[183] and used a vtz-f12 basis set.[120] DF-DFT-SAPT was also carried out in MOLPRO and used the aug-cc-pVTZ basis set. An asymptotic correction is applied in the

DF-DFT-SAPT calculation, which is found by subtracting the DFT HOMO energy from the experimental ionization potential. For the ethylene with zero rotation around the π bond, the experimental ionization potential is taken from http://cccbdb.nist.gov/. In lieu of an experimental value for the ethylene with an 80^{O} rotation about the π bond, the ionization potential is estimated from Koopmans theorem.[160] In all calculations, the binding energy was calculated by subtracting twice the energy of the monomer from the dimer.

4.1.3 Results and discussion - H_4

The energies of H_4 from DMC calculations with several trial wave functions are compared to CCSD(T) and FCI in Figure 4.1.1. Over the entire range of R, the DMC calculation with a single-determinant of HF, BD, or B3LYP orbitals result in similar energies, showing no preference for orbital type. The CCSD(T) curve is roughly the same distance from the FCI over the entire range of R, with the exception of the square structure at R = r = 1.27 Å. The method that gives the lowest energy over the entire range is DMC with a CAS(4,4) trial wave function. This is lower in energy than the FCI and CCSD(T) curve due to the sensitivity to the limited basis set for FCI and CCSD(T). The FCI and CAS(4,4) calculations also give a smooth curve. For the H_4 system, the ground-state at the R = r is well described by two determinants. Thus, an unrestricted Hartree-Fock calculation and a CAS(2,2) calculation will give a smooth curve over the entire range of R, albeit higher in energy than the DMC with a CAS(4,4) trial wave function or the FCI calculation. For the remainder of the discussion of the square H_4 model system, unrestricted trial wave functions will not be considered.

At a distance of R = 1.0 Å, the DMC single-determinant calculations give a lower energy than both the FCI and CCSD(T) results, largely due to the sensitivity of the FCI and CCSD(T)

results to the basis set truncation. This result demonstrates the advantages of the DMC method. For non-degenerate ground-states, DMC can give very accurate estimates of the energy.

As the distance between the H_2 molecules increases and becomes closer to a square geometry, the multiconfigurational character of the ground-state increases, as shown by the plot of the leading CAS(4,4) vector. The single-determinant DMC results cross the other benchmark results at R = 1.06 Å. At this distance, the leading CAS coefficient is 0.95. This indicates the limit for DMC calculations with a single determinant to achieve accuracy similar to FCI and CCSD(T). As R continues to increase, the departure of the single-determinant DMC calculations from the CCSD(T) and FCI results is striking. At a R = 1.158, the leading CAS coefficient becomes 0.90, and the energy is noticeably higher with all of the single-determinant DMC calculations than CCSD(T), FCI, and DMC with a CAS(4,4) trial wave function. At R = 1.21 Angstrom, the leading CAS coefficient becomes 0.83, and the DMC single-determinant methods fail to accurately represent the ground-state compared to the other benchmark methods. It is striking to note that the Brueckner coupled cluster method performs similar to the CCSD(T) method at R = 1.21 Angstrom, but that the DMC with Brueckner orbitals deviates strongly from the benchmark calculations. Clearly, the coupled cluster method is recovering more correlation energy than the DMC method can when the nodal surface is poorly described by a single-determinant.

At the square geometry of R = r = 1.27 Å, only the FCI and DMC with a CAS(4,4) trial wave function are able to accurately describe the ground-state of the H_4 system. At this geometry, the leading CAS coefficient has fallen to 0.68. The single-determinant DMC calculations out perform the CCSD(T) calculations at this point due to the sensitivity of CCSD(T) to the truncation of the basis set.

4.1.4 Results and discussion – Ethylene dimer

The monomer is twisted around the π bond to tune the separation of the highest occupied molecular orbital (HOMO) and lowest occupied molecular orbital (LUMO); at 90 degrees rotation, the HOMO and LUMO become degenerate and the ground state is a singlet diradical. A CAS(2,2) calculation with the two electrons in the π orbital and the π^* orbital was carried out at every five degrees of rotation from zero to ninety degrees on the monomer, where the active space is the. At eighty degrees rotation, the leading CAS coefficient is 0.864. With the results demonstrated for H_4 in the previous section, this appears to be an ideal test of the near degeneracy effects on weakly interacting systems. This degeneracy should not impact the CCSD(T) results and the single-determinant based DMC results become questionable. DF-DFT-SAPT was used to calculate the interaction energy of two monomers each rotated 80 degrees around the π bond. This gave a minimum energy at a separation of 3.50 Å. This distance was subsequently set as the dimer separation for the dimer with monomers rotated zero degrees around the π bond. The geometry of the each set of dimers is shown in figure 4.1.2.

Figure 4.1.2 - Dimers with each monomer twisted 0 (A) and 80^O (B) around the π-bond, with a separation of 3.50 Å.

(A) (B)

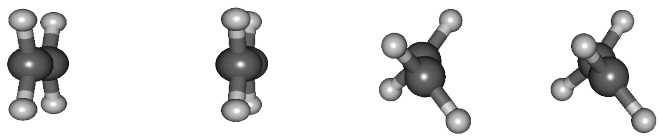


Table 4.1.1 – Interaction energy for the ethylene dimer at 0° and 80° twist angles. All energies are in kcal/mol. The number in parenthesis after the DMC energies indicates the error bar. (One standard deviation)

	Zero degree rotation	80 Degree rotation
Leading coefficient for monomer	0.978	0.864
Leading coefficient for dimer	0.956	0.744
MP2-F12	0.698	-1.028
MP2C-F12	1.030	-0.341
CCSD(T)-F12a	0.988	-0.593
	DF-DFT-SAPT	
$E^1_{Electrostatic}$	-0.492	-2.002
$E^1_{Exchange}$	3.939	4.806
$E^2_{Induction}$	-1.429	-2.336
$E^2_{Exchange-Induction}$	1.365	2.407
$E^2_{Dispersion}$	-3.069	-3.803
$E^2_{Dispersion-Exchange}$	0.594	1.040
SAPT Interaction	0.911	-0.532
DMC/HF	1.3(1)	-0.2**(1)**
DMC/B3LYP	1.2**(1)**	-0.13(9)
DMC/BD	1.2**(1)**	-0.18**(9)**
DMC/CAS	0.9**(1)**	-0.46(9)

Table 4.1.1 gives the interaction energy for each set of dimers using various methods. Given the results shown above for the H_4 test system, it is reasonable to assume that the CCSD(T)-F12a method can be taken to be the benchmark for the interaction of this system, as CCSD(T) was shown above to be accurate for systems where the leading CAS coefficient is as small as 0.864.

The DF-DFT-SAPT result gives close agreement to the CCSD(T)-F12a result for both dimers, over binding by only 0.07 kcal/mol for dimer A and under binidng by 0.07 kcal/mol for dimer B. Comparing dimers A and B, the electrostatic contribution increases by ~1.5 kcal/mol in magnitude for the twisted dimer. The exchange contribution is 0.9 kcal/mol more attractive for

dimer B, but this is offset by the induction term being more repulsive by 0.9 kcal/mol. The dispersion energy is more attractive for dimer B, due to the smaller HOMO/LUMO gaps of the monomers.

For the dimer comprised of monomers with no rotation around their π bonds, MP2-F12 gives an interaction energy that is to attractive, which is expected for MP2 applied to dispersion-bound π systems.[92] The MP2C-F12 method corrects this over binding, and gives excellent agreement with the CCSD(T)-F12a result. The under binding of the DMC method when using a HF trial wave function is more than two standard deviations from the CCSD(T)-F12a result. The DMC result when using a B3LYP or Brueckner trial wave function is more reasonable, and is 0.2 kcal/mol from the CCSD(T)-F12a result. The superiority of a correlated set of orbitals, with DFT orbitals generally performing the best, is a result that has been pointed out by other researchers.[184] However, a simple CAS trial wave function, which correlates only the π electrons and the π orbitals, corrects the nodal surface enough to give excellent agreement with the CCSD(T)-F12a result.

For dimer B the trends are similar. The MP2-F12 method again over binds, and the MP2C-F12 method gives an interaction energy close to the CCSD(T)-F12a result though still 0.25 kcal/mol too small in magnitude. The DMC calculations with single-determinant trial wave functions are underestimating the interaction energy. The CAS trial wave function allows for a much better description of the nodal surface, and this results in sub-chemical accuracy in comparison to the CCSD(T)-F12a result.

The DMC method with a single determinant trial wave function under binds for each set of dimers. However, the error for the twisted monomers is more egregious. This should not be surprising, given that the leading CAS coefficient for the dimer is lower than it is for the

monomer. This fact aside, it is reassuring to note that a small CAS(4,4) calculation on the dimer and CAS(2,2) calculation on the monomers, taking into account only the π electrons and π* orbitals, allows for a superior nodal surface that gives an excellent agreement with the CCSD(T)-F12a result.

This system is also demonstrative of a larger problem in using multi-determinant trial wave functions. For the ethylene dimer, the active space of four electrons in four orbitals generates only twelve non-zero determinants, and the two electrons in two orbitals makes only two non-zero determinants for the monomer. Table 4.1.2 gives the determinants for structure A at the equilibrium distance, a "long distance," where the monomers are separated by 20 Å, and the determinants for the monomer. Twelve determinants is a compact determinant expansion to use as a trial wave function in DMC; however, for a larger system where the number of determinants grows, a cutoff is generally applied as a selection criteria for which determinants to keep in the trial wave function. As can be seen from Table 4.1.2, applying a threshold of 0.01 to the magnitude of determinant coefficients would keep every determinant for the monomer but truncate the expansion for the equilibrium structure at only eight determinants, resulting in a possibly unbalanced description of the nodal surface. Additionally, if the binding energy were calculated as the energy of the equilibrium structure minus the energy of the monomers at long distance, applying a threshold of 0.001 would result in twelve determinants for the equilibrium structure and only ten determinants for the long distance structure, again resulting in an unbalanced description of the nodal surface for one structure.

Table 4.1.2 – Determinants and coefficients from a CAS(4,4) calculation on structure A, where long refers a 20 A separation between the monomers, short refers to the equilibrium structure, and monomer is for a CAS(2,2) calculation on an isolated ethylene monomer. a and b refer to the electron occupation of each orbital, 1 shows that both electrons are in the same orbital, and 0 shows that no electrons occupy the orbital.

	Coefficient		
Determinant	Monomers at long distance	Equilibrium structure	Configuration
1	0.957	0.955	1100
2	0.102	0.101	baba
3	0.102	0.101	abab
4	0.102	0.105	baab
5	0.102	0.105	abba
6	-0.102	-0.101	0110
7	-0.102	-0.109	1001
8	-0.102	-0.113	1010
9	-0.102	-0.090	0101
10	0.043	0.044	0011
11	0.000	-0.005	bbaa
12	0.000	-0.005	aabb

	Monomer	
Determinant	Coefficient	Configuration
1	0.978	10
2	-0.209	01

4.1.5 Conclusions

The DMC method is commonly employed with a single determinant trial wave function, and has been shown to give lower DMC energies when the determinant is formed using correlated orbitals, than when HF orbitals are used. It has been shown here that this is a very good

approximation when the system being studied has very little multi-determinant character. However, a simple model system H_4 test demonstrates that as the frontier orbitals become more nearly degenerate, the quality of the nodal surface given by a single determinant, regardless of the orbitals employed, decreases. Taking the twisted ethylene as an example, this ground-state multi-configurational nature has been shown to impact the interaction energy of weakly interacting systems. Additionally, the difficulty of using multi-determinant trial wave functions due to the CI threshold has been discussed.

4.2 QAUNTUM MONTE CARLO CALCULATION OF THE BINDING ENERGY OF THE BERYLLIUM DIMER

This work has been submitted to the Journal of Chemical Physics as M.J. Deible, M. Kessler, and K.D. Jordan, "Quantum Monte Carlo Calculation of the Binding Energy of the Beryllium Dimer." M.J.D. Performed the single-determinant and several of the multi-determinant calculations. M.K. performed several of the multi-determinant calculations. All authors contributed to the discussion.

4.2.1 Introduction

The beryllium dimer has been the subject of numerous experimental and theoretical studies.[27,179,180,185–209] In 1984, Bondybey and English, using ro-vibrational data from near the bottom of the ground state $^1\Sigma_g^+$ potential of Be_2, deduced a value of 790 $\pm$ 30 cm^{-1} for the binding energy (here defined from the potential energy minimum, i.e., neglecting vibrational zero-point energy).[188,210,211] Based on rotational structure in the v=0 level, Bondybey and England determined a bond length of 2.45 Å. More

recently, Merritt and coworkers experimentally observed eleven vibrational levels of Be_2, allowing them to obtain a more refined estimate of 929.74 cm^{-1} for the well depth.[186] This was subsequently revised to 934.9 cm^{-1} upon further analysis of the experimental data.[185] Over the past few years, several electronic structure calculations have been reported that obtained well depths close to the recent experimental value.[179,202–207,212–214] The keys to the successful calculations are the use of large, flexible basis sets and the recovery of a large portion of the correlation energy including contributions from the 1*s* core orbitals. To illustrate the difficulty of calculating an accurate binding energy of Be_2, we note that a complete basis set limit CCSD(T) calculation including correlation of the 1*s* core electrons underestimates the binding energy by 224 cm^{-1}.[208] Moreover, basis functions beyond those included in the aug-cc-pVCQZ basis set[215–217] contribute 79 cm^{-1} to the CCSD(T) value of the binding energy.[208]

In this study, we apply the diffusion Monte Carlo (DMC) method[9,29,30,74] to the Be dimer. The DMC method is capable of giving the exact ground state energy under the constraint of the fixed-node approximation,[10,12,33,218,219] which is required to maintain the fermionic nature of the wave function. The constraint is imposed by use of a trial function generally taken to be a single Slater determinant of Hartree-Fock or density functional theory (DFT) orbitals. If the nodal surface of the trial wave function were exact, then the DMC method, if run for a sufficient number of steps, and extrapolated to zero time step, would give the exact ground state energy. It is generally assumed[10] that for weakly interacting dimers the errors introduced by the use of single determinant trial functions to impose the fixed nodes largely cancel when the interaction energy is calculated by subtracting twice the energy of the two monomers from that of the dimer, and this has been confirmed for systems such as the water dimer and the methane dimer.[11] However, it is not clear that this will be the case for weakly interacting species for which static correlation effects are important. The Be dimer is thus a particularly interesting test system, as the ground state wavefunction of Be has considerable $2s^2 \rightarrow 2p^2$ character. Indeed, all-electron DMC calculations on Be using a CAS(2,4) complete active space trial function allowing for $2s^2 \rightarrow 2p^2$ mixing give a significantly lower total energy than do DMC calculations using a single Slater determinant trial

function.[179,220] [10] However, DMC calculations using a CAS(4,8) trial function for the dimer and a CAS(2,4) trial function for the atom considerably underestimates the binding of the dimer.[221] Harkless and Irikura[179] used a truncated CAS(4,8) space and Anderson and Goddardg[180] used a GVB trial function and each reported DMC values of the binding energy of Be_2 in good agreement with experiment. As will be discussed later in the manuscript, the good agreement of the binding energy obtained from these two DMC studies and experiment is likely fortuitous. In the present study, we calculate the binding energy of Be_2 using the DMC method in conjunction with more flexible multiconfigurational trial functions than were employed in earlier studies.

4.2.2 Computational Details

The experimental value of the equilibrium bond length, 2.4536 Å,[186] was used for all calculations on the beryllium dimer. In the first set of calculations, single determinant trial functions were considered, with the orbitals being obtained from the HF approximation and from several DFT methods including the local density approximation (LDA), the PBE[90] and BLYP[110,222] generalized gradient approximation (GGA) functionals, and the Becke3LYP,[110,111] PBE0,[156] and BHandHLYP hybrid functionals,[92] which contain 20, 25, and 50% exact exchange, respectively. In addition, a trial function comprised of a single Slater determinant of Brueckner orbitals determined from Brueckner coupled cluster singles plus doubles (CCSD) calculations[223,224] was considered. The cc-pVQZ 5*s*4*p*3*d*2*f* contracted Gaussian-type orbital basis set[84] was used to represent the orbitals in the single Slater determinant trial functions. Both cc-pVQZ-*fg* and cc-pVQZ-*g* basis sets were used in generating the multiconfigurational trial functions. Here –*fg* indicates that both the *f* and *g* functions were omitted from the basis set, while –*g* indicates that only the *g* functions were omitted.

DMC calculations were also carried out using multiconfigurational trial functions generated from CAS and configuration interaction (CI) calculations. For the beryllium dimer, both CAS(4,8) and CAS(4,16) trial functions were considered. The CAS(4,8) wave function allows all arrangements of the

four valence electrons in the space of the molecular orbitals (MOs) derived from the 2*s* and 2*p* atomic orbitals (AOs). The CAS(4,16) wavefunction expands the active space to include the π_g, π_u, σ_g, σ_u molecular orbitals derived from the 3*s* and 3*p* atomic orbitals and has 816 configuration state functions (CSFs). The DMC calculations were carried out retaining all CSFs with coefficients greater than 0.001, 0.0025, 0.005, and 0.01 in magnitude, and these results were used to extrapolate the energies to the value for the full configuration space. The extrapolation is shown in Figure 4.2.2. With the 0.001 coefficient threshold 341 CSFs are retained from the CAS(4,16) space.

CI trial functions were generated by carrying out configuration interaction calculations, allowing for up to four electron excitations from the valence space into the full virtual space and employing CAS(4,8) orbitals. Natural orbitals were then generated and used to carry out a subsequent CI calculation allowing up to quadruple excitations in the space of all natural orbitals with occupations greater than 0.0001 in the first CI calculation (again keeping the $1\sigma_g$ and $1\sigma_u$ orbitals frozen). Thresholds of 0.01, 0.005, 0.0025, and 0.001 were applied to the resulting CI expansion and gave 484 CSFs for the dimer at the 0.001 cutoff (out of a total of 4500 CSFs). For calculating the binding energy, a single plus double excitation CI (SDCI) calculation was carried out on the atom using CAS(2,8) orbitals and followed by a subsequent SDCI calculation using natural orbitals with occupations greater than 0.0001.

Each of the trial functions was combined with a Jastrow factor[36] with electron-electron, electron-nucleus, and electron-electron-nucleus terms. Variational Monte Carlo (VMC) calculations were used to optimize the Jastrow factors *via* energy minimization. For the multiconfigurational trial functions, the coefficients of the CSFs were optimized simultaneously with the parameters in the Jastrow function. The resulting trial functions, including the Jastrow factors, were then [98]used to carry out DMC simulations using 40,000-50,000 walkers at a single time step of 0.001 a.u. The correction scheme of Ma *et al*.[225] was used to account for the electron-nuclear cusps. For one set of DMC calculations using the CAS(4,16) trial function, time steps of 0.0005, 0.003, and 0.005 a.u. were also used, allowing extrapolation of the energies to the zero time-step limit. This extrapolation is shown in Figure 4.2.3.

The single determinant trial functions were generated using Gaussian09[92] and the multiconfigurational trial functions were generated using GAMESS.[226] The quantum Monte Carlo calculations were carried out using the CASINO[48] and QMCPack[227] codes for the single determinant and multideterminent trial functions, respectively. QMCPack was used for the latter calculations due to its implementation of an efficient algorithm for handling multideterminent trial functions.

4.2.3 Results

Table 4.2.1 - Total energies of Be and Be_2 and the Be_2 dissociation energy computed with DMC using various trial functions.

Trial function[a]	Total energy (a.u.)		D_e (cm^{-1})
	Be	Be_2	
HF/QZ	-14.65730(4)	-29.31789(6)	724(21)
LDA/QZ	-14.65721(4)	-29.31977(7)	1174(25)
PBE/QZ	-14.65731(5)	-29.31960(8)	1094(26)
BLYP/QZ	-14.65725(4)	-29.31956(8)	1113(26)
B3LYP/QZ	-14.65727(3)	-29.31946(8)	1079(23)
PBE0/QZ	-14.65728(3)	-29.31907(8)	992(21)
BHandH/QZ	-14.65726(5)	-29.31891(7)	966(26)
BD/QZ	-14.65718(4)	-29.31872(7)	955(24)
CAS(4,8)/QZ-*fg*[b]	-14.667228(9)	-29.33707(3)	573(8)
CAS(4,16) /QZ-*fg*[b]	-14.66730(1)	-29.33832(3)	819(8)
Ext. CAS(4,16)/QZ-*fg*[b]	-14.66730(1)	-29.33841(2)	838(7)
CAS(4,16)/QZ-*g*[b]	-14.66727(2)	-29.33838(3)	845(8)
Ext. CAS(4,16)/QZ-*g*[b]	-14.66727(2)	-29.33844(2)	857(9)
CI/QZ-*g*[b]	-14.667250(9)	-29.33848(2)	873(6)
Ext. CI/QZ-*g*[b]	-14.667250(9)	-29.33864(2)	908(6)
Experimental[c]	-14.667356	-29.33897	934.9

[a]TZ and QZ refer to the cc-pVTZ and cc-pVQZ basis sets, respectively. The "-*g*" and –*fg*" indicate, respectively, that the *g* functions, and *f* and *g* functions were omitted from the basis sets. Ext. refers to CAS and CI results extrapolated to the full configuration space for the active orbital list as described in the text.

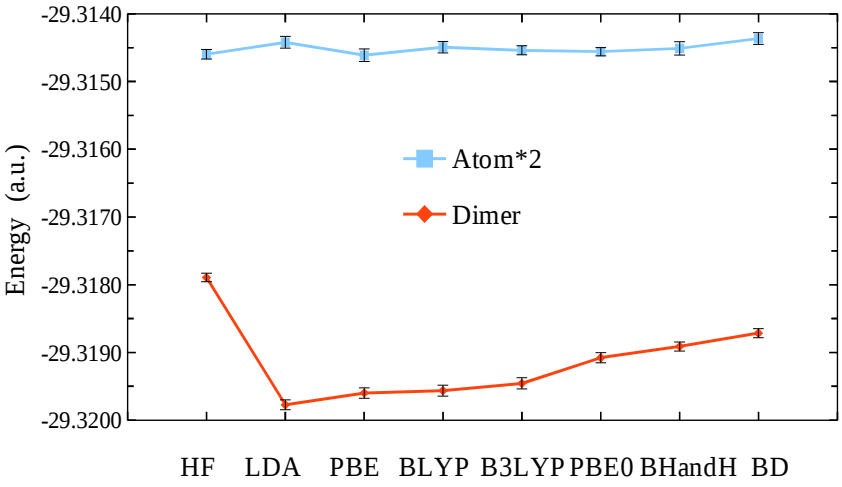
[b]0.001 threshold on CI coefficients

[c]The experimental D_e value for Be_2 is from Ref. 185. The non-relativistic energy of the Be atom is from Ref. 228, and the energy for Be_2 subtracting the experimental value of the dimer binding energy from twice the energy of Be.

The results of the DMC calculations at the 0.001 a.u. time step are reported in Table 4.2.1. With the HF trial function, the DMC calculations give a binding energy of 724 cm^{-1}, significantly smaller than the experimental value of 935 cm^{-1}. On the other hand, the DMC calculations using trial functions

employing LDA or GGA orbitals considerably overestimate the binding energy of Be_2. Significantly improved agreement with experiment is obtained when using orbitals from hybrid functionals containing a component of exact exchange or from Brueckner CCSD calculations. Specifically, the DMC calculations using PBE0, BHandH, and Brueckner orbitals result in binding energies of 992, 966, and 955 cm^{-1}, respectively. The result obtained using Brueckner orbitals, in particular, is in excellent agreement with experiment. In contrast, we note that Toulouse *et al.*,[26] obtained a binding energy of 618 cm^{-1} when using single determinant trial functions but optimizing the orbitals and basis functions of the atom and dimer in the VMC optimization steps. For both Be and Be_2, regardless of the orbitals used, the DMC calculations using single determinant trial functions give energies considerably above the exact energies of these species, suggesting that the good agreement with experiment of the binding energies obtained using trial functions based on a single determinant of hybrid DFT or Brueckner orbitals is likely fortuitous. Support for this conjecture is provided by Fig. 4.2.1 from which it is seen that the calculations that give the binding energies closest to experiment do so because they give a higher energy for the dimer.

Figure 4.2.1- DMC energy of twice the beryllium atom and the dimer for several single-determinant trial wave functions.



As expected, based on earlier studies,[179] [221] DMC calculations using valence-space CAS trial functions give significantly lower energies for the Be atom and dimer than do the DMC calculations using the trial functions based on single Slater determinants. However, the DMC calculations using the CAS(4,8) trial function for the dimer and CAS(2,4) for the atom give a binding energy of only 573 cm^{-1}, which is even smaller than that obtained using HF trial functions. This indicates that use of valence space CAS trial functions does not result in a balanced treatment of the nodal surfaces of the atom and molecule. Most of error must be due to the inadequacy of the CAS(4,8) space in describing the nodal surfaces of the dimer since the DMC calculations on the atom using the CAS(2,4) trial function give an energy very closest to the current best estimate[228] of the energy of Be (-14.667228 *vs.* -14.667356 a.u.). Expanding the CAS space to include also the MOs derived from the 3*s* and 3*p* AOs, giving CAS(2,8) and CAS(4,16) for the atom and dimer, respectively, lowers the DMC energies of the atom and dimer, by 10 and 300 cm^{-1}, respectively, and results in a dimer binding energy of 845 cm^{-1}, at the 0.001 coefficient threshold and using the cc-pVQZ-*g* basis set. The corresponding binding energy obtained using the cc-pVQZ-*fg* basis set is 819 cm^{-1}, indicating that the nodal surface of Be_2 is slightly improved by including *f* functions in the basis set. Extrapolating these results along the sequence of coefficient cutoffs gives binding energies of 838 and 857 cm^{-1} for trial functions expanded in terms of the cc-pVQZ-*fg* and cc-pVQZ-*g* basis sets, respectively (see Fig. 4.2.2). The extrapolation to zero time step of the DMC/CAS(4,16) results obtained with the 0.001 coefficient threshold and using the cc-pVQZ-*g* basis set gives a DMC binding energy of 849 cm^{-1} (see Fig. 4.2.3), *vs.* the 845 cm^{-1} value obtained with the 0.001 a.u. time step. Thus, we conclude that the error due to the use of the finite time step is inconsequential.

Figure 4.2.2- Extrapolation of the DMC energy of the beryllium dimer to zero CI coefficient in the calculations using the CI/cc-pVQZ-g trial function. The dashed red line is a linear fit to the DMC energies (blue squares).

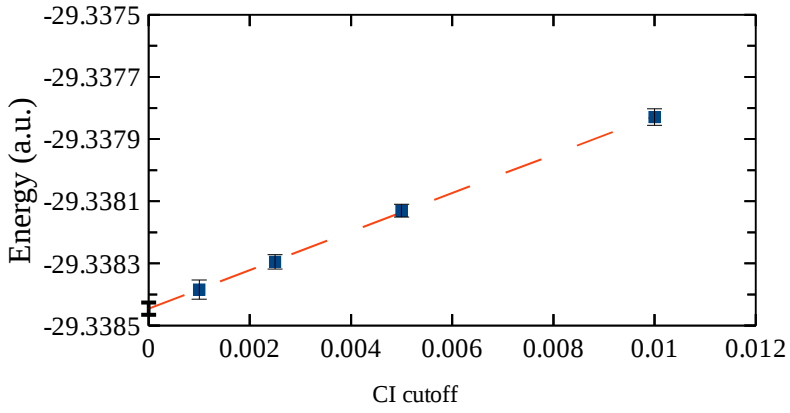
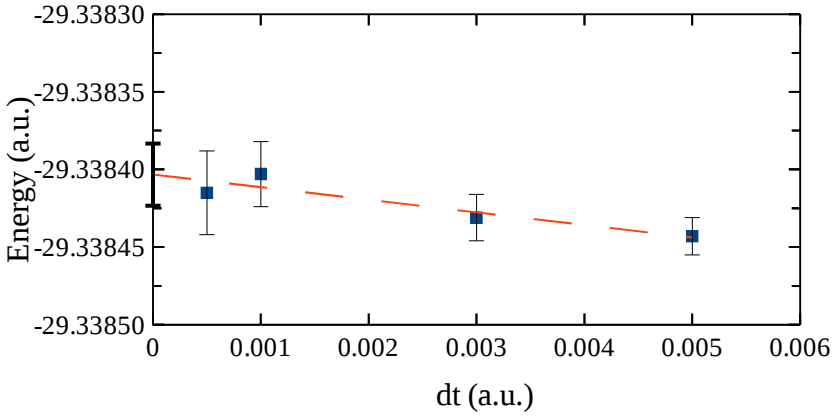


Figure 4.2.3- Extrapolation to zero time step for DMC on the CI natural orbitals for Be dimer at equilibrium bond length using the CI trial wave function. A linear fit was used for the extrapolation. Results obtained for the CAS(4,16) trial function.



The DMC calculations using the CI trial function with the 0.001 coefficient cutoff and cc-pVQZ-*g* basis set yielded a dimer binding energy of 873 cm^{-1}, while the corresponding result extrapolated using the different coefficient thresholds is 908 cm^{-1}, which is only 27 cm^{-1} smaller than the experimental value of the binding energy. These results demonstrate that correlation effects involving configurations outside the CAS(8,16) space are important for describing the nodal surface of Be_2.

It should be noted that the SDTQ CI calculations using the cc-pVQZ-*g* basis set and freezing the 1*s* orbitals give a binding energy of only 601 cm^{-1}, which is 334 cm^{-1} lower than the experimental value. About 70 cm^{-1} of the error in this result is due to the neglect of the correlation effects involving the core 1*s* orbitals.[229] while the remaining error is due to correlation effects that are not captured due to the basis set truncation. This underscores one of the major advantages of the DMC method, namely, that it achieves convergence with much smaller basis sets (for the trial functions) than required for traditional quantum chemistry methods.

4.2.4 Conclusions

In conclusion, the binding energy of the beryllium dimer has been calculated using the diffusion Monte Carlo method in conjunction with a wide variety of trial wave functions. Even DMC calculations with a trial wave function as large as CAS(4,16) considerably underestimate the binding energy of the beryllium dimer. CI trial functions allowing excitations from the valence space into the entire virtual space give a binding energy within 27 cm^{-1} of the experimental value. It is possible that this small remaining discrepancy from experiment is due to the neglect of excitations from the 1*s* orbitals in the trial functions used for the DMC calculations. Although DMC calculations using small configurational spaces that give binding energies close to experiment have been reported for Be_2, they also give energies for the atom and dimer that are appreciably higher than those obtained using the CI trial functions employed here. Thus, the good agreement of the binding energy of Be_2 with the experimental value obtained with

small configuration trial functions could be fortuitous. We believe our findings are relevant for a wide range of other dimers, e.g., the benzene dimer, where there is appreciable configuration mixing in the wavefunctions of the monomers. In particular, achieving well converged binding energies for such systems is likely to require the use of multiconfigurtional trial functions allowing for high-order excitations as well as excitations outside the valence space.

4.2.5 Acknowledgements

This research was supported by grant CHE 136234 from the National Science Foundation. The calculations were done on computers in the University of Pittsburgh's Center for Simulation and Modeling. We acknowledge valuable discussions with M. Morales.

5.0 DIFFUSION MONTE CARLO CALCULATIONS OF THE WATER ACENE INTERACTION ENERGY

5.1.1 Introduction

The interaction of molecules with acenes has attracted considerable interest for a variety of reasons, including the use of such systems as models for understanding molecular adsorption on graphene and graphite and for testing theoretical approaches for describing weak interactions. Of particular interest is the magnitude of the interaction of a water molecule with the graphene surface. This question has been addressed in numerous theoretical studies, with most of these concluding that the binding energy of a water molecule to a graphene sheet is about -3.1 kcal/mol.[181] However, diffusion Monte Carlo (DMC)[229–231] and random-phase approximation (RPA)[29,30,74] calculations give significantly smaller (in magnitude) binding energies (-1.6 and -2.3, respectively.[232–234] In studies using extrapolation of the results of calculations of water-acene systems to obtain the water-graphene limit, water-benzene and water-coronene systems play an important role. Based on the highest level calculation available for these systems, the binding energies of water-benzene and water-coronene are about -3.2 and -3.05 to -3.35 kcal/mol, respectively.[235] In the case of water-benzene, the theoretical estimates are in excellent agreement with the values deduced from experiment, while there is no experimental value for the binding energy of the water-coronene system.

There are multiple challenges in extending electronic structure calculations to the larger acenes needed to realistically model water interacting with graphene. Foremost among these is the fact that dispersion interactions play a major role in the binding.[229,231,236] As a result, traditional generalized gradient or hybrid density functional theory (DFT) methods are not suitable. This problem is partially overcome by the use of dispersion-corrected DFT approaches. However, several such approaches were recently tested for water-coronene and none were found to give a quantitatively accurate description of the interaction potential.[231] The MP2 method[231] does include dispersion interactions, but can overestimate their importance. Although this problem can be solved by use of the CCSD(T) method,[112,237] this approach (as traditionally formulated) is computationally prohibitive for large acenes. In addition to the challenges posed by dispersion interactions, traditional quantum chemistry methods using Gaussian-type orbitals are plagued by near-linear dependency and basis set superposition error (BSSE)[238–240] problems when applied to molecules interacting with large acenes. Two of the most promising methods for characterizing the interaction of water and other molecules with acenes are DFT-based symmetry-adapted perturbation theory with density fitting[241,183] and the MP2C method of Hesselmann.[148,242] There are two implementations of the former – the DF-DFT-SAPT approach of Hesselmann[148,184] and co-workers and the DF-SAPT (DFT) method of Szalewicz and co-workers.[148] These methods display O(N^5) scaling, where N is the number of electrons, and thus are computationally attractive compared to CCSD(T). However, they still suffer from near-linear dependency problems when flexible basis sets containing diffuse functions are employed. Moreover, the MP2, MP2C, and other methods that involve perturbative corrections to the Hartree-Fock wavefunction might not be appropriate for large acenes due to their small HOMO/LUMO gaps.

An alternative approach for calculating interaction energies, which is free of the problems described above, is the diffusion Monte Carlo (DMC) method.[242,243] This method has already been applied to several weakly interacting systems, including water clusters,[29,30] and the water-benzene dimer,[13,38,73,86] and, as noted above, the water-graphene system.[235] In the usual fixed-node implementation, DMC calculations make use of a trial function – generally a Slater determinant comprised of Hartree-Fock (HF) or DFT orbitals multiplied by Jastrow factors[69,86] to describe the electron-electron, electron-nuclear, and electron-electron-nuclear correlations. DMC calculations afford the advantages of being relatively insensitive to the basis set used for the trial function and having a relatively low, ~O(N^3) scaling, albeit with a large prefactor. The large prefactor is "compensated" by the fact that the DMC calculations are highly parallel.

In this paper, we use the DMC method with a B3LYP trial wave function to calculate the interaction energies of the water-benzene, water-anthracene, water-triphenylene, and water-coreonene complexes. The binding energies calculated with the DMC method will be compared to other benchmark methods. CCSD(T)-F12b calculations for benzene and anthracene will establish the benchmark for these smaller system. The close agreement of MP2C-F12 and DF-DFT-SAPT to CCSD(T)-F12b for the small acenes will give confidence to using the MP2C and DF-DFT-SAPT methods as benchmarks for the triphenylene and coronene complexes.

5.1.2 Computational details

The structure of the water-coronene complex considered in this study is shown in Figure 5.1.1. The geometries employed are the same as those in used in Ref. 70. For the acenes, the experimental C-C bond length, 1.420 Å, and C-C-C angles, 120°, for graphite are employed.[231] The acene C-H bond lengths and C-C-H angles were taken to be 1.09 Å and 120°,

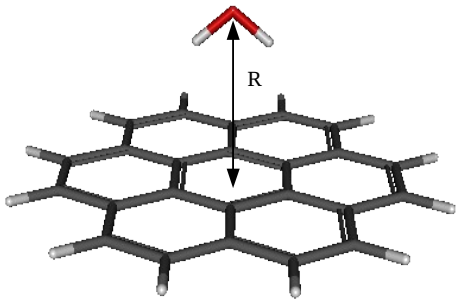
respectively. The experimental gas-phase geometry is employed for the water monomer, i.e., the O-H bond lengths are 0.9572 Å and the H-O-H angle is 104.52°.[244] The water molecule was placed above and perpendicular to the middle of the central ring, with both OH groups pointing towards the acene, and an oxygen-ring distance of 3.36 Å as shown in Figure 5.1.1 This distance came from an optimization of the geometry of water-coronene using the DF-DFT-SAPT method.[231] The experimentally observed water-benzene complex has a minimum energy structure with a "tilted" water molecule with one OH group pointed toward the center of the ring.[231] However, the energy difference between the one H-down minimum and the two H-down saddle point structure is less than 0.2 kcal/mol.[245]

The DMC calculations were carried out using single-determinant trial wave functions obtained from HF or B3LYP[110,111] calculations, combined with three-term Jastrow factors[36] to describe the electron-electron, electron-nuclear, and electron-electron-nuclear correlations. The parameters in the Jastrow factors were obtained by minimization of the variance of the local energy. The Dirac-Fock pseudopotentials of Trail and Needs[36] were employed on all atoms and the basis sets of Xu et. al[86] that have been designed for use with these pseudopotentials. The basis set for the oxygen and carbon atoms consist of the *s* and *p* functions of the quintuple-zeta basis set together with the two *d* functions from the double-zeta basis set of Xu et al. Additionally, the most diffuse *d* functions were scaled by a factor of 1.5 in order to avoid linear dependency in the trial wave function. For the hydrogen atoms, the quintuple-zeta set of *s* functions was combined with the double-zeta *p* functions, and the most diffuse *p* was scaled by a factor of 1.5. The pseudopotential localization error in the DMC calculations was treated with the T-move scheme[85] and used a symmetric branching factor.[35] Time steps of 0.01, 0.025, 0.05, and 0.1 a.u. were used in the DMC calculations, and all binding energies were extrapolated to

zero time step by use of quadratic fits. Too calculate the binding energy, two structures were considered; the equilibrium structure described above, and a long-distance structure, where the water molecule is moved 15 Å away from the acene. The DMC simulations were run using a target population of 60,000 walkers for up to 30,000 Monte Carlo steps. The trial wave functions were generated using the Gaussian09B code.[70] All QMC calculations were carried out with the CASINO[92] code.

The DF-MP2C-F12[119,146]calculations used the aug-cc-pvtz[84] basis set and are corrected for basis set superposition error (BSSE) using the counterpoise method.[121] The DF-DFT-SAPT calculations on the water-benzene, water-anthracene, and water-triphenylene complexes were carried out using the aug-cc-pvtz basis set and are free of BSSE by definition. These calculations were carried out using the MOLPRO[122] code as outlined in Ref. 231 . For water-coronene, the DFT-SAPT results are taken from reference 231 and used a truncated aug-cc-pvtz basis set described therein. The CCSD(T)-F12b results for benzene and anthracene are also taken from reference 231 and use the vtz-f12 basis set.

Figure 5.1.1 – Structure of the water-coronene system studied.



5.1.3 Results and discussion

Table 5.1 – Binding energy of a water molecule to an acene. All energies are given in kcal/mol.

	B3LYP/DMC	CCSD(T)-F12b	DF-MP2C-F12	DF-DFT-SAPT
Benzene	-2.9(1)	-3.17	-3.24	-3.20
Anthracene	-3.4(2)	-3.37	-3.34	-3.34
Triphenylene	-2.8(2)	-	-3.21	-3.15
Coronene	-2.6(3)	-	-3.13	-3.05

The results of the DMC, CCSD(T)-F12b, MP2C, and DF-DFT-SAPT calculations are given in Table 5.1.1 For all systems considered, the CCSD(T)-F12b, DF-MP2C-F12, and DF-DFT-SAPT methods give essentially the same binding energies; CCSD(T)-F12b calculations were carried out only for the two smallest systems. This gives confidence in using the DF-MP2C-F12 method as the benchmark result for the larger acenes interacting with the water. The vibrational zero-point energy correction to the dissociation energy of water-benzene has been estimated to be about 1.0 kcal/mol.[69,86] Applying this correction to binding energies from DMC calculations gives a D_0 value of -1.8(1) kcal/mol, which are in close agreement with the experimental value of -2.44(9) and -2.25(28) kcal/mol. [236,246]

For water-anthracene, DMC value of the binding energy obtained using the B3LYP trial wave function is in excellent agreement with the three reference values, giving a binding energy of -3.4(2) kcal/mol, compared to the CCSD(T)-F12b value of -3.37 kcal/mol.

For water-triphenylene, the DMC calculation with a B3LYP trial wave function underestimates by 0.3 kcal/mol the binding energy, using the DF-MP2C-F12 method is taken as the benchmark.

In the case of water-coronene, the DMC calculations with a B3LYP trial wave function gives a binding energy of -2.6(3) kcal/mol, which is beyond one standard deviation from the DF-

MP2C-F12 result of -3.13 kcal/mol. The DMC results is underestimating the binding energy by 0.5 kcal/mol.

It has been shown for linear acenes that the ground-state wavefunction becomes more multi-configurational as more rings are added.[212,247] The leading coefficient of a CAS(14,14) calculation on anthracene, using all of the π and π* orbitals as the active space, is 0.86[211] which indicates a significant multi-configurational character. For triphenylene and coronene, a CAS(14,14) active space results in a leading coefficient of 0.90 and 0.83, respectively, on the leading coefficient. However, a CAS(14,14) active space does not include all of the π and π* orbitals for triphenylene or coronene, and it is likely that the coefficient of the leading configuration is smaller in magnitude than the above estimates.

Table 5.1.2 shows the DF-DFT-SAPT energy decomposition for each acene interacting with a water molecule. The electrostatic contribution to the energy decreases as the acene grows in size, but is compensated for by an increase in the dispersion energy. The exchange energy decreases going from benzene to anthracene, but remains constant from anthracene to coronene. The induction energy and higher-order induction terms, accounted for in the δHF term, remains largely constant across all sizes of the acenes. These results are consistent with those from earlier work,[70,231] all though these two studies did not include triphenylene. The electrostatic term decreases by 0.43 kcal/mol going from anthracene to triphenylene, and only by 0.13 kcal/mol when going from triphenylene to coronene.

Table 5.1.2 – DF-DFT-SAPT interaction energy decomposition, in kcal/mol.

	Benzene	Anthracene	Triphenylene	Coronene
E^1_{es}	-2.83	-2.29	-1.86	-1.73
E^1_{Ex}	3.25	2.85	2.76	2.79
E^2_{Ind}	-1.29	-1.22	-1.27	-1.29
$E^2_{Ind\text{-}Ex}$	0.83	0.76	0.78	0.80
δHF	-0.26	-0.21	-0.20	-0.20
Net induction	-0.72	-0.67	-0.69	-0.69
E^2_{Disp}	-3.38	-3.66	-3.77	-3.83
$E^2_{Disp\text{-}Ex}$	0.46	0.43	0.42	0.42
Net dispersion	-2.92	-3.23	-3.35	-3.41
SAPT Interaction	-3.22	-3.34	-3.15	-3.04

5.1.4 Conclusions

In this study, it has been shown that the DMC method is capable of predicting accurate binding energies for water and small acenes if a B3LYP trial function is used. The increasing size of the acene does not change the interaction energy for the benchmark calculations of CCSD(T)-F12a, DF-DFT-SAPT, and DF-MP2C-F12. At first sight it may seem surprising that the binding energy of water on acenes remains constant despite the size of the acene. However, the dispersion energy grows with the acene size and is roughly compensated by the decreasing electrostatic contribution,[70,231] which is demonstrated by the energy decomposition afforded by the DF-DFT-SAPT procedure.

It was noted in the Introduction that a recent DMC calculation gave a value of -1.6 kcal/mol for the binding of a water molecule to the graphene surface.[70,248] This value is about 1.4 kcal/mol smaller in magnitude than obtained in other recent theoretical studies of water-graphene.[229–231] The DMC calculations of reference 235 were carried out using periodic boundary

conditions, a single k point, and with a supercell containing 50 carbon atoms. Based on the results of RPA calculations, the authors of reference 235 established that the error due to the use of the Γ only for k-point sampling is about 0.5 kcal/mol. Applying this correction to their DMC result gave a binding energy of -2.1 kcal/mol, which is still appreciably smaller in magnitude than other recent estimates of the binding energy of water-graphene. Possible remaining errors include the interactions between water molecules in adjacent cells and the inadequacy of a single determinant wavefunction for describing the nodal surfaces. Given the spacing between the water molecules, the error in the binding energy due to water-water interactions should be less than 0.2 kcal/mol. This suggests that the use of a single determinant trial function introduces an error of about 0.9 kcal/mol in the DMC value of the interaction energy for water-graphene.

6.0 DIFFUSION MONTE CARLO CALCULATIONS OF THE BENZENE AND ANTHRACENE DIMERS

6.1.1 Introduction

Polycyclic aromatic hydrocarbons (PAHs) are molecules with fused benzene rings that are hydrogen terminated. These are important molecules in many fields; they are known to be carcinogenic;[249] in materials science, they are constituent molecules for many organic semiconductors due to their high conductivity.[250] Also, large PAHs can serve as a model system for graphene.[251]

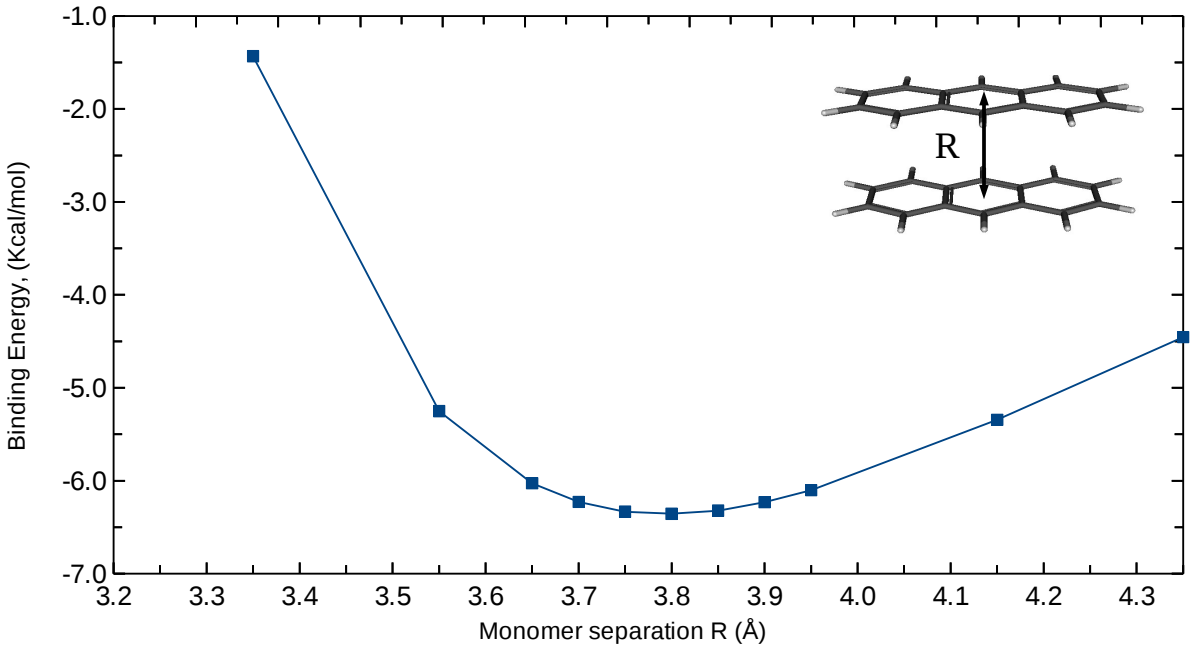
The interaction between two PAH molecules is dominated by dispersion forces. Traditional DFT methods without dispersion corrections do no adequately describe the interaction between PAHs. Additionally, it is well known that second-order Møller-Plesset perturbation theory (MP2), overestimates the magnitude of the dispersion interaction in the "stacked" (i.e. face-to-face) arrangement of the benzene dimer. The CCSD(T) method does properly describe the dispersion interaction in such systems but is computationally prohibitive for the interaction between two large PAH molecules. An alternative method is diffusion Monte Carlo (DMC).[74] DMC has several advantages over the above methods for studying pi-stacked systems. DMC is free of basis set superposition error (BSSE),[10] has very favorable scaling with system size, and is capable of sub-chemical accuracy for weakly interacting systems.[12]

The benzene dimer is the most commonly studied[251–255] (also see two recent reviews[256,257] and references therein) system to model interacting PAHs. Several studies of the anthracene dimer have appeared,[258–263] attempting to increase the size of the acene for which reliable benchmarks can be obtained.

Experimental results[264,265] for the anthracene dimer indicate two structurally different isomers for the dimer, though the structures were not resolved and the binding energy was not measured. For the benzene dimer, two minimum, the tilted T-shape dimer and parallel-displaced dimer are only separated by 0.1 kcal/mol with the T-shape dimer being lower in energy.[266] This is consistent with experimental results, which can resolve each structure dependent on the carrier gas.[267]

Here, the benzene dimer and the anthracene dimer are considered in the face-to-face (AA) stacking conformation. This is not the global minimum for either structure; for benzene, the titled T-shaped dimer is the global minimum,[266] and the crossed dimer (where a monomer in the AA conformation is rotated ninety degrees about the principle axis) is the minimum for the anthracene dimer.[261] (It should be noted that only four geometries were tested for the anthracene dimer.) However, the face-to-face dimer arrangements of acene dimers are particularly valuable for benchmarking methods for describing strong dispersion interactions. The DF-DFT-SAPT[148] method is used to better understand the components of the interaction energy. The MP2C[184] method of Heßelmann, which replaces the uncoupled polarizability in the MP2 expression with the coupled polarizability from a time-dependent DFT calculation, and the However, using DMC with a single determinant trial wave function may not always be the best method for studying systems where the monomers are strongly correlated.

Figure 6.1.1 – Binding energy curve of the anthracene dimer calculated using the DF-MP2C-F12 method with a vdz-f12 basis set. The inset shows the binding energy minimum. The geometry and description of R are also shown.



6.1.2 Computational details

The geometry for the benzene dimer is taken from reference 253, where the carbon-carbon bond length and carbon-hydrogen bond length are set at 1.3915 and 1.08 Å, respectively, and the distance between the rings is 3.9 Å. The geometry of the anthracene monomer was optimized at the MP2 level with the Dunning aug-cc-pVTZ basis set[84] and no geometrical constraints. The equilibrium distance for the anthracene dimer, shown in Figure 6.1.1, was determined by calculating an interaction curve at the DF-MP2C-F12 level of theory with the VDZ-F12 basis set[120] using the monomer optimized geometry for both monomers in the dimer. The DF-DFT-

SAPT calculations used the aug-cc-pVTZ basis set. The above calculations were performed with the MOLPRO code.[122]

The trial wave functions for the DMC calculations were generated in the MOLPRO code, and the DMC calculations were performed with the CASINO[48] code. The trial wave functions consisted of a three-term Jastrow factor[36] with parameters that were optimized via variance minimization in a Variance Monte Carlo (VMC) calculation. The Trail-Needs pseudopotential[85] and a basis set of Xu *et al.* was used for the trial wave function. The basis set consists of the quintuple-zeta *s* and *p* functions for each carbon and the triple-zeta *d* functions. For hydrogen, the quintuple-zeta *s* functions were used with the triple-zeta *p* functions. Orbitals from both Hartree-Fock and B3LYP[87,88] calculations are used for the trial wave function. The binding energy is calculated as the energy of the equilibrium structure minus the energy of the two molecules at 15.0 Å separation (10 Å for benzene). The localization of the pseudopotential was treated beyond the locality approximation.[35] A symmetric[66] branching factor was used in the DMC calculation. Time steps of 0.1, 0.05, 0.025, and 0.01 a.u. are used and a quadratic fit to the binding energy is used to extrapolate to zero time step.

Complete active space (CAS) calculations were carried out with the GAMESS[226] program package, and the QMCPACK[227] code was used for the DMC calculations on the benzene dimer with a multi-determinant trial wave function. The trial wave functions used the pseudopotential and corresponding basis sets of Burkatzki, Filippi, and Dolg (BFD).[41] The basis set for carbon is constructed from the quintuple-zeta *s* and *p* functions, and the triple-zeta *d* functions. For hydrogen, the basis set consisted of the quintuple-zeta *s* functions and the triple-zeta *p* functions. The Jastrow factor and CI coefficients were simultaneously optimized in a VMC calculation via energy minimization. For this calculation, the binding energy is calculated as the equilibrium

structure minus twice the monomer. This is a more efficient method than using the long-distance structure, as it reduces the number of determinants to be evaluated. A time step of 0.01(a.u.) is used for this calculation.

6.1.3 Results and discussion

Table 6.1.1 – SAPT energy components and binding energy (kcal/mol) for the benzene and anthracene dimer.

	Benzene	Anthracene
E^1_{Es}	0.142	-0.844
E^1_{Ex}	2.892	8.272
E^2_{Ind}	-0.995	-2.921
$E^2_{Ex\text{-}Ind}$	0.873	2.698
Net induction	-0.122	-0.222
E^2_{Disp}	-5.165	-15.392
$E^2_{Disp\text{-}Ex}$	0.600	1.761
Net dispersion	-4.565	-13.632
DF-DFT-SAPT	-1.653	-6.426
MP2C-F12	-1.670	-6.530
CCSD(T) CBS[a]	-1.660	-
B3LYP/DMC	-0.7(2)	-5.5(3)

[a.] Reference 253

The results of the DF-DFT-SAPT, MP2C-F12, and DMC calculations are summarized in table 6.1.1. The MP2C-F12 and DF-DFT-SAPT methods are in excellent agreement. Additionally, for the benzene dimer, the MP2C-F12 and DF-DFT-SAPT methods are in very close agreement with the complete-basis-set limit CCSD(T) calculations of Sherril and coworkers.[253] As mentioned in the introduction, the dispersion energy is the main contribution to the interaction energy.

The DMC method consistently underestimates the binding energy of these systems with respect to the MP2C-F12 and DF-DFT-SAPT benchmark methods. It has been shown[181] that the

the B3LYP trial function is in general more suitable for high accuracy DMC calculations of weakly interacting systems. For the benzene and anthracene dimers, this does not appear to be the case, as it underestimates the binding energy by nearly one kcal/mol in both cases. The DMC calculation of the binding energy of the benzene dimer that uses the B3LYP trial wave function is in very close agreement with a previous DMC calculation[268] of the benzene dimer; however, it was noted in that study, as it is here, that this estimate is more repulsive compared to other high level calculations.

One possible explanation for why the DMC method predicts an underestimate of the binding energy of the acene dimers is due to the multi-configurational nature of the benzene and anthracene monomers. For each monomer, a complete active space SCF (CAS-SCF) calculation is possible for an active space that consists only of the π electrons and the π* orbitals. This CAS calculation with a cc-pVDZ basis set gives a leading CAS coefficient of 0.94 and 0.86 for benzene and anthracene, respectively. For the benzene dimer, a CAS(12,12) calculation will include the π electrons and the π* orbitals of each monomer and results in a leading CAS coefficient of 0.88; clearly, the CASSCF calculation including the π electrons and the π* orbitals of each monomer for the anthracene dimer is intractable, as it would be require a CAS(28,28) calculation. Because the CAS(12,12) calculation produces more than 200,000 determinants, a trial wave function for a DMC calculation requires a threshold to be applied to the CI coefficients. This threshold is chosen to be 0.005, which retains 2139 determinants and gives a normalized CI vector of 0.9906. For the monomer, the CAS(6,6) trial wave function is used, and at the same CI threshold of 0.005, 117 determinants are retained and give a normalized CI vector of 0.9999. The binding energy that results from this calculation is 0.23(17) kcal/mol, qualitatively incorrect and clearly an underestimate compared to the benchmark methods given

in table 6.1. This is largely a result of the large threshold applied to the CI coefficient. However, a CI threshold of 0.0025 results in 7478 determinants for the trial wave function. Clearly, this problem very quickly becomes computationally demanding. However, the concerns raised in section 4.1.1 about the multi-determinant trial wave function for ethylene dimer are relevant here. Indeed, the lack of agreement between the normalized CI vector in the dimer and monomer indicate that the arbitrary CI threshold of 0.005 is not having a similar effect on the nodal surface of the monomer as it is on the dimer.

In conclusion, the interaction energy of the face-to-face benzene dimer and anthracene dimer were studied. It was shown that for the benzene dimer, the DF-DFT-SAPT and MP2C-F12 methods are in very good agreement with complete-basis-set limit CCSD(T) calculations. For the anthracene dimer, the DF-DFT-SAPT and MP2C-F12 methods agree very well in predicting a binding energy of -6.5 kcal/mol. The DMC method was also used to study the interaction energy of these two dimers. It is seen that for two different single-determinant trial wave functions, the DMC method underestimates the magnitude of the binding energy. This may be due to the multi-configurational nature of the monomers, but a DMC calculation that retains only a small number of determinants in the trial wave function did not improve the binding energy.

7.0 CONCLUSIONS

In my dissertation research, I carried out several electronic structure calculations in support of diffusion Monte Carlo studies. I studied large systems of environmental and economical importance as well as small model systems that help to further develop the successful application of diffusion Monte Carlo.

Newly designed correlation consistent Gaussian basis sets have been combined with the Trail-Needs pseudopotentials and tested in DMC calculations. It is shown these basis sets reduce the variance in VMC calculations and speed up convergence in DMC calculations. It is further shown that an expanded *s* and *p* space is required to accurately describe weakly interacting systems.

I also explored the utility of a trial wave function composed of Brueckner orbitals. It is shown that in all-electron calculations, Brueckner orbitals do not out perform B3LYP orbitals, but that for calculations with a pseudopotential, Brueckner performs as well as PBE0 and better than B3LYP orbitals for a test set of diatomic molecules. Additionally, I demonstrated that a bent CO_2 anion is better describe by a trial wave function comprised of Brueckner orbitals than a trial wave function of B3LYP or PBE orbitals. Both DFT and Brueckner orbitals result in a lower variational energy than Hartree-Fock orbitals for the diatomic test set and the CO_2 anion test.

I also studied large, weakly interacting clusters. For the water-16 cluster, DMC was used to corroborate the complete-basis-set limit MP2 results. These results are further compared to a many-body energy estimate, where the one-, two-, three-, and four-body energies were calculated using high level *ab initio* techniques and compared to several force-fields. It was concluded that the five and higher-body interactions contribute nearly two kcal/mol to the net binding energy of the water-16 cluster.

The DMC method was also used to study a methane molecule encapsulated in a dodecahedral water cage. A many-body energy estimate was again made, considering the two- and three-body interactions of each water from the cage interacting with the encapsulated methane. The DMC calculation of the interaction energy of the supermolecular allowed for a refined estimate of the interaction energy from the many-body decomposition. Additionally, the two- and three-body SAPT method was used to break down the contributions to the interaction energy. It was found that even at the three-body level, induction, exchange, and dispersion all contribute approximately one kcal/mol to the interaction energy.

I used diffusion Monte Carlo to determine a benchmark interaction energy for a carbon dioxide clathrate. The *N*-body decomposition revealed that the DMC results are in good agreement with the binding energy calculated at the two- and three-body level with the CCSD(T)-F12b method. The SAPT method was again used to determine the contributions to the interaction energy, and it was shown that the dispersion, exchange, and induction are all contributing to the binding energy at both the two- and three-body level.

I also carried out all-electron calculations on the beryllium atom and dimer. This is a model system for strongly correlated weakly interacting systems that is known to be a challenging problem due to the ground state degeneracy of the atom. I used single determinant

and multi-determinant trial wave functions. I showed that Hartree-Fock underestimates the binding energy, and DFT overestimates the binding. Mixing a percent of exact exchange into the functional used to generate the DFT orbitals leads to a more accurate description of the binding energy, all though this is shown to be a fortuitous result. Several multi-determinant trial wave functions with different active spaces were considered. It was shown that a very large virtual space was required to generate a trial wave function that was able to accurately reproduce the experimental binding energy.

To gauge the effectiveness of DMC as the trial wave function became more multi-configurational, a model H_4 system was studied. It was shown that a single-determinant trial wave function does not give a good description of the nodal surface as the multi-configurational nature of the ground increases. Using full CI as a benchmark, it is shown that as the leading coefficient from a CAS(4,4) calculation increases to approximately 0.85, the single-determinant trial wavefunctions are not adequately describing the nodal surface.

To determine the effect of a multi-reference ground state on weakly interacting systems, stacked ethelyne dimers were studied. As the dimers are twisted around the π bond, the ground-state wave function becomes more multi-configurational. At a twist angle of eight degrees, the monomer has a leading CAS coefficient of 0.86. It is shown that for this dimer, a CAS trial wave function is required for the DMC binding energy to agree with the CCSD(T) becnchmark value.

Diffusion Monte Carlo studies were also carried out on the a series of acenes interacting with a water molecule. It is shown that the binding energy predicted by DMC calculations with a single determinant trial wave function decreases as the acene grows larger, in contrast to high level benchmark *ab initio* calculations such as MP2C-F12 and DF-DFT-SAPT. A single

determinant of B3LYP orbitals was used to in calculating the binding energy. It is estimated that a multi-reference trial wave function would provide a better description of the nodal surface.

Finally, DMC calculations were carried out for the benzene dimer and anthracene dimer. It is again shown that a single-determinant is not providing an accurate description of the nodal surface. A multi-configuration trial wave function for the benzene dimer indicated that a very large determinant expansion will be necessary to accurately describe the binding energy of large acenes.

QMC methods such as VMC and DMC are becoming more widely recognized as a valuable tool in finding a solution to the Schroedinger equation for atoms, molecules, and solids. The flexibility of the trial wave function, high accuracy, and low scaling make it an ideal method to apply to a wide range of systems. Additionally, the inherent parallel nature of the method makes it well suited to take advantage of today's modern computer architectures.

Besides the quantum Monte Carlo methods of VMC and DMC described here, other QMC methods are being developed. Reptation Monte Carlo is a method used to calculate unbiased expectation values of operators that do not commute with the Hamiltonian. It has been successfully used to compute dipole moments of transition metal oxides that are difficult to characterize with traditional DMC. Full CI QMC[269] is another recently developed method that uses the time-dependent Schröedinger equation to project an answer as walkers evolve in determinant space, as opposed to the coordinate space typically used in DMC. It has been used to study the homogeneous electron gas,[270] cohesive energies of bulk solids,[271] and excitation energies of ethene and butadiene.[272]

New trial wave functions are being studied in DMC. Aside from the single- and multi-determinant wave functions used in this dissertation, generalized valence bond (GVB)[273] and

anti-symmetrized geminal powers[274] are being used. Further developments in the area of trial wave functions are expected to bring higher accuracy and greater computational efficiency to diffusion Monte Carlo.

BIBLIOGRAPHY

(1) Leach, A. *Molecular Modeling: Principles and Applications*; 2nd ed.; Pearson Education Limited: England, 2001.

(2) Atkins, P.; Friedman, R. *Molecular Quantum Mechanics*; Fifth.; Oxford University Press: Oxford, 2005.

(3) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; 2nd ed.; Dover Pulications, Inc.: New York, N.Y., 1989.

(4) Argaman, N.; Makov, G. Density Functional Theory: An Introduction. *Am. J. Phys.* **2000**, *68*, 69.

(5) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, 864–871.

(6) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(7) Tkatchenko, A.; Scheffler, M. Accurate Molecular van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 6–9.

(8) Becke, A. D. Density Functionals for Static, Dynamical, and Strong Correlation. *J. Chem. Phys.* **2013**, *138*, 074109.

(9) Austin, B. M.; Zubarev, D. Y.; Lester, W. a. Quantum Monte Carlo and Related Approaches. *Chem. Rev.* **2012**, *112*, 263–288.

(10) Mella, M.; Anderson, J. B. Intermolecular Forces and Fixed-Node Diffusion Monte Carlo: A Brute Force Test of Accuracies for He[sub 2] and He–LiH. *J. Chem. Phys.* **2003**, *119*, 8225.

(11) Dubecky, M.; Derian, R.; Jurecka, P.; Mitas, L.; Hobza, P.; Otyepka, M. Quantum Monte Carlo for Noncovalent Interactions: An Efficient Protocol Attaining Benchmark Accuracy. *Phys. Chem. Chem. Phys.* **2014**, 20915–20923.

(12) Dubecký, M.; Jurečka, P.; Derian, R.; Hobza, P.; Otyepka, M.; Mitas, L. Quantum Monte Carlo Methods Describe Noncovalent Interactions with Subchemical Accuracy. *Journal of Chemical Theory and Computation*, 2013, *9*, 4287–4292.

(13) Wang, F.-F.; Deible, M. J.; Jordan, K. D. Benchmark Study of the Interaction Energy for an (H2O)16 Cluster: Quantum Monte Carlo and Complete Basis Set Limit MP2 Results. *J. Phys. Chem. A* **2013**, *117*, 7606–7611.

(14) Ceperley, D. M. Fermion Nodes. *J. Stat. Phys.* **1991**, *63*, 1237–1267.

(15) Umrigar, C. J.; Toulouse, J.; Filippi, C.; Sorella, S.; Hennig, R. G. Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions. *Phys. Rev. Lett.* **2007**, *98*, 1–4.

(16) Per, M. C.; Walker, K. a.; Russo, S. P. How Important Is Orbital Choice in Single-Determinant Diffusion Quantum Monte Carlo Calculations? *J. Chem. Theory Comput.* **2012**, *8*, 2255–2259.

(17) Petruzielo, F. R.; Toulouse, J.; Umrigar, C. J. Approaching Chemical Accuracy with Quantum Monte Carlo. *J. Chem. Phys.* **2012**, *136*, 124116.

(18) B. L. Hammond, W. A. Lester, Jr., and P. J. R. *Monte Carlo Methods in Ab Initio Quantum Chemistry*; Lester, William, J., Ed.; World Scientific: Singapore, 1994.

(19) Clark, B. K.; Morales, M. a; McMinis, J.; Kim, J.; Scuseria, G. E. Computing the Energy of a Water Molecule Using Multideterminants: A Simple, Efficient Algorithm. *J. Chem. Phys.* **2011**, *135*, 244105.

(20) Kato, T. On the Eigenfunctions of Many-Particle Systems in Quantum Mechanics. *Commun. Pure Appl. Math.* **1957**, *10*, 151–177.

(21) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(22) Ceperley, D.; Alder, B. J. Ground State of the Electron Gas by a Stochastic Method. *Phys. Rev. Lett.* **1980**, *45*.

(23) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087.

(24) Ma, a.; Drummond, N. D.; Towler, M. D.; Needs, R. J. All-Electron Quantum Monte Carlo Calculations for the Noble Gas Atoms He to Xe. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2005**, *71*, 1–5.

(25) Ma, A.; Towler, M. Scheme for Adding Electron–nucleus Cusps to Gaussian Orbitals. *J. Chem.* … **2005**, *122*, 224322.

(26) Toulouse, J.; Umrigar, C. J. Optimization of Quantum Monte Carlo Wave Functions by Energy Minimization. *J. Chem. Phys.* **2007**, *126*, 084102.

(27) Toulouse, J.; Umrigar, C. J. Full Optimization of Jastrow-Slater Wave Functions with Application to the First-Row Atoms and Homonuclear Diatomic Molecules. *J. Chem. Phys.* **2008**, *128*, 174101.

(28) Umrigar, C.; Filippi, C. Energy and Variance Optimization of Many-Body Wave Functions. *Phys. Rev. Lett.* **2005**, *94*, 1–4.

J. Chem. Phys. **1975**, *63*, 1499.

(30) Anderson, J. B. Quantum Chemistry by Random Walk. H 2P, H+3 D3h 1A′1, H2 3Σ+u, H4 1Σ+g, Be 1S. *J. Chem. Phys.* **1976**, *65*, 4121.

(31) Assaraf, R.; Caffarel, M.; Khelif, A. Diffusion Monte Carlo Methods with a Fixed Number of Walkers. *Phys. Rev. E* **2000**, *61*, 4566–4575.

(32) Umrigar, C. J.; Nightingale, M. P.; Runge, K. J. A Diffusion Monte Carlo Algorithm with Very Small Time-Step Errors. *J. Chem. Phys.* **1993**, *99*, 2865.

(33) Bressanini, D. Implications of the Two Nodal Domains Conjecture for Ground State Fermionic Wave Functions. *Phys. Rev. B - Condens. Matter Mater. Phys.* **2012**, *86*, 115120.

J. Chem. Phys. **1991**, *95*, 3467.

(35) Casula, M. Beyond the Locality Approximation in the Standard Diffusion Monte Carlo Method. *Phys. Rev. B* **2006**, *74*, 1–4.

(36) Drummond, N. D.; Towler, M. D.; Needs, R. J. Jastrow Correlation Factor for Atoms, Molecules, and Solids. *Phys. Rev. B* **2004**, *70*, 1–11.

(37) Xu, J.; Al-Saidi, W. a; Jordan, K. D. Unpublished Results.

(38) 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.

(39) 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.

(40) Trail, J. R.; Needs, R. J. Norm-Conserving Hartree-Fock Pseudopotentials and Their Asymptotic Behavior. *J. Chem. Phys.* **2005**, *122*, 14112.

(41) Burkatzki, M.; Filippi, C.; Dolg, M. Energy-Consistent Pseudopotentials for Quantum Monte Carlo Calculations. *J. Chem. Phys.* **2007**, *126*, 234105.

(42) Jr, T. D. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007.

(43) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796.

(44) Benedict, W. S.; Gailar, N.; Plyler, E. K. Rotation-Vibration Spectra of Deuterated Water Vapor. *J. Chem. Phys.* **1956**, *24*, 1139.

(45) Alfè, D.; Gillan, M. Efficient Localized Basis Set for Quantum Monte Carlo Calculations on Condensed Matter. *Phys. Rev. B* **2004**, *70*, 161101.

(46) Frisch, M. J.; Trucks, G. W.; Et., A. Gaussian 03.

(47) Gonze, X.; Amadon, B.; Anglade, P.-M.; Beuken, J.-M.; Bottin, F.; Boulanger, P.; Bruneval, F.; Caliste, D.; Caracas, R.; Côté, M.; et al. ABINIT: First-Principles Approach to Material and Nanosystem Properties. *Comput. Phys. Commun.* **2009**, *180*, 2582–2615.

(48) Needs, R. J.; Towler, M. D.; Drummond, N. D.; López Ríos, P. Continuum Variational and Diffusion Quantum Monte Carlo Calculations. *J. Phys. Condens. Matter* **2010**, *22*, 023201.

(49) Alfè, D.; Alfredsson, M.; Brodholt, J.; Gillan, M. J.; Towler, M. D.; Needs, R. J. Quantum Monte Carlo Calculations of the Structural Properties and the B1-B2 Phase Transition of MgO. *Phys. Rev. B* **2005**, *72*, 014114.

(50) 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.

(51) 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.

(52) 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.

(53) 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.

(54) 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.

(55) 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.

(56) 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.

(57) 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.

(58) Douglas, M.; Kroll, N. M. Quantum Electrodynamical Corrections to the Fine Structure of Helium. *Ann. Phys. (N. Y).* **1974**, *82*, 89–155.

(59) De Jong, W. A.; Harrison, R. J.; Dixon, D. A. Parallel Douglas–Kroll Energy and Gradients in NWChem: Estimating Scalar Relativistic Effects Using Douglas–Kroll Contracted Basis Sets. *J. Chem. Phys.* **2001**, *114*, 48.

(60) 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.

(61) 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.

(62) Huber, K. P.; Herzberg, G. *MOLECULAR SPECTA and MOLECULAR STRUCTURE IV. CONSTANTS OF DIATOMIC MOLECULES*; Van Nostrand Reinhold Company: New York, N.Y., 1979.

(63) Urdahl, R. S.; Bao, Y.; Jackson, W. M. An Experimental Determination of the Heat of Formation of C2 and the C□H Bond Dissociation Energy in C2H. *Chem. Phys. Lett.* **1991**, *178*, 425–428.

(64) 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.

(65) Casula, M.; Filippi, C.; Sorella, S. Diffusion Monte Carlo Method with Lattice Regularization. *Phys. Rev. Lett.* **2005**, *95*, 100201.

(66) Casula, M.; Moroni, S.; Sorella, S.; Filippi, C. Size-Consistent Variational Approaches to Non-Local Pseudopotentials: Standard and Lattice Regularized Diffusion Monte Carlo Methods Revisited. 1–10.

(67) Gurtubay, I. G.; Needs, R. J. Dissociation Energy of the Water Dimer from Quantum Monte Carlo Calculations. *J. Chem. Phys.* **2007**, *127*, 124306.

(68) 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.

(69) Ma, J.; Alfè, D.; Michaelides, A.; Wang, E. The Water-Benzene Interaction: Insight from Electronic Structure Theories. *J. Chem. Phys.* **2009**, *130*, 154303.

(70) 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.

(71) Knizia, G.; Adler, T. B.; Werner, H.-J. Simplified CCSD(T)-F12 Methods: Theory and Benchmarks. *J. Chem. Phys.* **2009**, *130*, 054104.

(72) Klopper, W.; M. van Duijneveldt-van de Rijdt, J. G. C.; van Duijneveldt, F. B. Computational Determination of Equilibrium Geometry and Dissociation Energy of the Water Dimer. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2227–2234.

J. Chem. Phys. **2012**, *136*, 244105.

(74) Foulkes, W. M. C.; Mitas, L.; Needs, R. J.; Rajagopal, G. Quantum Monte Carlo Simulations of Solids. *Rev. Mod. Phys.* **2001**, *73*, 33–83.

(75) Brueckner, K. A. Nuclear Saturation and Two-Body Forces. II. Tensor Forces. *Phys. Rev.* **1954**, *96*, 508–516.

(76) Nesbet, R. K. Brueckner's Theory and the Method of Superposition. *Phys. Rev.* **1958**, *109*, 1632–1638.

(77) Larsson, S.; Smith Jr., V. H. Analysis of the S 2 Ground State of Lithium in Terms of Natural and Best Overlap (Brueckner) Spin Orbitals with Implications for the Fermi Contact Term. *Phys. Rev.* **1969**, *178*, 137–152.

(78) Scuseria, G. On the Connections between Brueckner–coupled-cluster, Density-dependent Hartree–Fock, and Density Functional Theory. *Int. J. Quantum Chem.* **1995**, *55*, 165–171.

(79) Lindgren, I.; Salomonson, S. Brueckner Orbitals and Density-Functional Theory. *Int. J. Quantum Chem.* **2002**, *294*, 1–22.

(80) Hesselmann, A.; Jansen, G. First-Order Intermolecular Interaction Energies from Coupled-Cluster Brueckner Orbitals. *J. Chem. Phys.* **2000**, *112*, 6949–6952.

(81) Heßelmann, A.; Jansen, G. First-Order Intermolecular Interaction Energies from Kohn–Sham Orbitals. *Chem. Phys. Lett.* **2002**, *357*, 464–470.

(82) Jankowski, K.; Nowakowski, K.; Wasilewski, J. A Comparative Study of Kohn–Sham, Brueckner and Hartree–Fock Orbitals. *Chem. Phys. Lett.* **2004**, *389*, 393–399.

(83) Wasilewski, J.; Nowakowski, K.; Jankowski, K. On the Presumptive Similarity of Kohn–Sham and Brueckner Orbitals. *Struct. Chem.* **2004**, *15*, 437.

(84) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007.

(85) J.R. Trail and R.J. Needs, J. Chem. Phys. 122, 174109 (2005); J.R. Trail and R.J. Needs, J. Chem. Phys. 122, 014112 (2005), See Also Www.vallico.net/casinoqmc/pplib/.

(86) Xu, J.; Deible, M. J.; Peterson, K. A.; Jordan, K. D. Correlation Consistent Gaussian Basis Sets for H , B − Ne with Dirac − Fock AREP Pseudopotentials : Applications in Quantum Monte Carlo Calculations. *J. Chem. Theory Comput.* **2013**, 2170–2178.

(87) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648.

(88) Lee, C.; Yang, W.; Parr, R. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(89) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158.

(90) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple - The PBE Functional. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(91) Voora, V.; Sommerfeld, T.; Jordan, K. D. Unpublished Results.

(92) Frish, M. J.; Trucks, G. W.; Shlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. Gaussian 09, 2009.

(93) Ma, a; Towler, M. D.; Drummond, N. D.; Needs, R. J. Scheme for Adding Electron-Nucleus Cusps to Gaussian Orbitals. *J. Chem. Phys.* **2005**, *122*, 224322.

(94) Feller, D.; Peterson, K. a. Re-Examination of Atomization Energies for the Gaussian-2 Set of Molecules. *J. Chem. Phys.* **1999**, *110*, 8384–8396.

(95) Woon, D. E.; Dunning Jr., T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. V. Core-Valence Basis Sets for Boron through Neon. *J. Chem. Phys.* **1995**, *103*, 4572.

(96) Nemec, N.; Towler, M. D.; Needs, R. J. Benchmark All-Electron Ab Initio Quantum Monte Carlo Calculations for Small Molecules. *J. Chem. Phys.* **2010**, *132*, 034111.

(97) Grossman, J. C. Benchmark Quantum Monte Carlo Calculations. *J. Chem. Phys.* **2002**, *117*, 1434–1440.

(98) Morales, M. a.; McMinis, J.; Clark, B. K.; Kim, J.; Scuseria, G. E. Multideterminant Wave Functions in Quantum Monte Carlo. *J. Chem. Theory Comput.* **2012**, *8*, 2181–2188.

(99) Xantheas, S. S.; Burnham, C. J.; Harrison, R. J. Development of Transferable Interaction Models for Water. II. Accurate Energetics of the First Few Water Clusters from First Principles. *J. Chem. Phys.* **2002**, *116*, 1493.

(100) Santra, B.; Michaelides, A.; Fuchs, M.; Tkatchenko, A.; Filippi, C.; Scheffler, M. On the Accuracy of Density-Functional Theory Exchange-Correlation Functionals for H Bonds in Small Water Clusters. II. The Water Hexamer and van Der Waals Interactions. *J. Chem. Phys.* **2008**, *129*, 194111.

(101) Bates, D. M.; Tschumper, G. S. CCSD(T) Complete Basis Set Limit Relative Energies for Low-Lying Water Hexamer Structures. *J. Phys. Chem. A* **2009**, *113*, 3555–3559.

(102) Góra, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K. Interaction Energies of Large Clusters from Many-Body Expansion. *J. Chem. Phys.* **2011**, *135*, 224102.

(103) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. Quadratic Configuration Interaction. A General Technique for Determining Electron Correlation Energies. *J. Chem. Phys.* **1987**, *87*, 5968.

J. Phys. Chem. Lett. **2010**, *1*, 3122–3127.

(105) Góra, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K. Interaction Energies of Large Clusters from Many-Body Expansion. *J. Chem. Phys.* **2011**, *135*, 224102.

(106) Santra, B.; Michaelides, A.; Fuchs, M.; Tkatchenko, A.; Filippi, C.; Scheffler, M. On the Accuracy of Density-Functional Theory Exchange-Correlation Functionals for H Bonds in Small Water Clusters. II. The Water Hexamer and van Der Waals Interactions. *Journal of Chemical Physics*, 2008.

(107) Koga, K.; Gao, G. T.; Tanaka, H.; Zeng, X. C. Formation of Ordered Ice Nanotubes inside Carbon Nanotubes. *Nature* **2001**, *412*, 802–805.

(108) Hankins, D. Water Molecule Interactions. *J. Chem. Phys.* **1970**, *53*, 4544.

(109) Leverentz, H. R.; Qi, H. W.; Truhlar, D. G. Assessing the Accuracy of Density Functional and Semiempirical Wave Function Methods for Water Nanoparticles : Comparing Binding and Relative Energies of (H 2 O) 16 and (H 2 O) 17 to CCSD (T) Results.

(110) Lee, C.; Yang, W.; Parr, R. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical review. B, Condensed matter*, 1988, *37*, 785–789.

(111) Becke, a; Becke, a. Density Functional Thermochemistry III The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(112) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

(113) Schrader, D. M.; Prager, S. Use of Electrostatic Variation Principles in Molecular Energy Calculations. *J. Chem. Phys.* **1962**, *37*, 1456.

(114) Whitten, J. L. Coulombic Potential Energy Integrals and Approximations. *J. Chem. Phys.* **1973**, *58*, 4496.

(115) Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast Linear Scaling Second-Order Møller-Plesset Perturbation Theory (MP2) Using Local and Density Fitting Approximations. *J. Chem. Phys.* **2003**, *118*, 8149.

(116) Feller, D. The Use of Systematic Sequences of Wave Functions for Estimating the Complete Basis Set, Full Configuration Interaction Limit in Water. *J. Chem. Phys.* **1993**, *98*, 7059.

(117) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-Set Convergence of Correlated Calculations on Water. *J. Chem. Phys.* **1997**, *106*, 9639.

(118) Adler, T. B.; Knizia, G.; Werner, H.-J. A Simple and Efficient CCSD(T)-F12 Approximation. *J. Chem. Phys.* **2007**, *127*, 221106.

(119) Werner, H.-J.; Adler, T. B.; Manby, F. R. General Orbital Invariant MP2-F12 Theory. *J. Chem. Phys.* **2007**, *126*, 164102.

(120) Peterson, K. a.; Adler, T. B.; Werner, H. J. Systematically Convergent Basis Sets for Explicitly Correlated Wavefunctions: The Atoms H, He, B-Ne, and Al-Ar. *J. Chem. Phys.* **2008**, *128*.

(121) Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553–566.

(123) Leverentz, H. R.; Qi, H. W.; Truhlar, D. G. Assessing the Accuracy of Density Functional and Semiempirical Wave Function Methods for Water Nanoparticles: Comparing Binding and Relative Energies of (H 2 O) 16 and (H 2 O) 17 to CCSD(T) Results. *J. Chem. Theory Comput.* **2013**, *9*, 995–1006.

(124) Wang, Y.; Huang, X.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. Flexible, Ab Initio Potential, and Dipole Moment Surfaces for Water. I. Tests and Applications for Clusters up to the 22-Mer. *J. Chem. Phys.* **2011**, *134*, 094509.

(125) Wang, F.-F.; Jenness, G.; Al-Saidi, W. a; Jordan, K. D. Assessment of the Performance of Common Density Functional Methods for Describing the Interaction Energies of (H2O)6 Clusters. *J. Chem. Phys.* **2010**, *132*, 134303.

(126) Habitz, P.; Bagus, P.; Siegbahn, P.; Clementi, E. Electronic Correlation Contribution to the Three-Body Potentials for Water Trimers. *Int. J. Quantum Chem.* **1983**, *23*, 1803–1806.

J. Chem. Phys. **1991**, *94*, 2873.

(128) Papajak, E.; Truhlar, D. G. Convergent Partially Augmented Basis Sets for Post-Hartree−Fock Calculations of Molecular Properties and Reaction Barrier Heights. *J. Chem. Theory Comput.* **2011**, *7*, 10–18.

(129) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(130) Fanourgakis, G. S.; Xantheas, S. S. Development of Transferable Interaction Potentials for Water. V. Extension of the Flexible, Polarizable, Thole-Type Model Potential (TTM3-F,

v. 3.0) to Describe the Vibrational Spectra of Water Clusters and Liquid Water. *J. Chem. Phys.* **2008**, *128*, 074506.

(131) Kumar, R.; Wang, F.-F.; Jenness, G. R.; Jordan, K. D. A Second Generation Distributed Point Polarizable Water Model. *J. Chem. Phys.* **2010**, *132*, 014309.

(132) Partridge, H.; Schwenke, D. W. The Determination of an Accurate Isotope Dependent Potential Energy Surface for Water from Extensive Ab Initio Calculations and Experimental Data. *J. Chem. Phys.* **1997**, *106*, 4618.

(133) Thole, B. T. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem. Phys.* **1981**, *59*, 341–350.

(134) Podeszwa, R.; Szalewicz, K. Three-Body Symmetry-Adapted Perturbation Theory Based on Kohn-Sham Description of the Monomers. *J. Chem. Phys.* **2007**, *126*, 194101.

(135) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. Unravelling the Origin of Intermolecular Interactions Using Absolutely Localized Molecular Orbitals. *J. Phys. Chem. A* **2007**, *111*, 8753–8765.

(136) Buffett, B. A. Clathrate Hydrates. *Annu. Rev. Earth Planet. Sci.* **2000**, *28*, 477–507.

(137) Fleyfel, F.; Devlin, J. P. FT-IR Spectra of 90 K Films of Simple, Mixed, and Double Clathrate Hydrates of Trimethylene Oxide, Methyl Chloride, Carbon Dioxide, Tetrahydrofuran, and Ethylene Oxide Containing Decoupled Water-d2. *J. Phys. Chem.* **1988**, *92*, 631–635.

(138) Fleyfel, F.; Devlin, J. P. Carbon Dioxide Clathrate Hydrate Epitaxial Growth: Spectroscopic Evidence for Formation of the Simple Type-II Carbon Dioxide Hydrate. *J. Phys. Chem.* **1991**, *95*, 3811–3815.

(139) Baghel, V. S.; Kumar, R.; Roy, S. Heat Transfer Calculations for Decomposition of Structure I Methane Hydrates by Molecular Dynamics Simulation. *J. Phys. Chem. C* **2013**, *117*, 12172–12182.

(140) Kumar, P.; Sathyamurthy, N. Theoretical Studies of Host-Guest Interaction in Gas Hydrates. *J. Phys. Chem. A* **2011**, *115*, 14276–14281.

(141) Khan, A. Theoretical Studies of CH[sub 4](H[sub 2]O)[sub 20], (H[sub 2]O)[sub 21], (H[sub 2]O)[sub 20], and Fused Dodecahedral and Tetrakaidecahedral Structures: How Do Natural Gas Hydrates Form? *J. Chem. Phys.* **1999**, *110*, 11884.

(142) Ramya, K. R.; Venkatnathan, A. Stability and Reactivity of Methane Clathrate Hydrates: Insights from Density Functional Theory. *J. Phys. Chem. A* **2012**, *116*, 7742–7745.

(143) Liu, Y.; Zhao, J.; Li, F.; Chen, Z. Appropriate Description of Intermolecular Interactions in the Methane Hydrates: An Assessment of DFT Methods. *J. Comput. Chem.* **2013**, *34*, 121–131.

(144) Bakowies, D. Accurate Extrapolation of Electron Correlation Energies from Small Basis Sets. *J. Chem. Phys.* **2007**, *127*, 164109.

(145) Jiang, H.; Myshakin, E. M.; Jordan, K. D.; Warzinski, R. P. Molecular Dynamics Simulations of the Thermal Conductivity of Methane Hydrate. *J. Phys. Chem. B* **2008**, *112*, 10207–10216.

(146) Pitoňák, M.; Hesselmann, A. Accurate Intermolecular Interaction Energies from a Combination of MP2 and TDDFT Response Theory. **2010**, *6*, 168–178.

(147) Werner, H.-J.; Knizia, G.; Manby, F. R. Explicitly Correlated Coupled Cluster Methods with Pair-Specific Geminals. *Mol. Phys.* **2011**, *109*, 407–417.

(148) Heßelmann, a.; Jansen, G.; Schütz, M. Density-Functional Theory-Symmetry-Adapted Intermolecular Perturbation Theory with Density Fitting: A New Efficient Method to Study Intermolecular Interaction Energies. *J. Chem. Phys.* **2005**, *122*, 0–17.

(149) Lotrich, V. F.; Szalewicz, K. Symmetry-Adapted Perturbation Theory of Three-Body Nonadditivity of Intermolecular Interaction Energy. *J. Chem. Phys.* **1997**, *106*, 9668.

(150) Lotrich, V. F.; Szalewicz, K. Perturbation Theory of Three-Body Exchange Nonadditivity and Application to Helium Trimer. *J. Chem. Phys.* **2000**, *112*, 112.

(151) McDonald, S.; Ojamäe, L.; Singer, S. J. Graph Theoretical Generation and Analysis of Hydrogen-Bonded Structures with Applications to the Neutral and Protonated Water Cube and Dodecahedral Clusters. *J. Phys. Chem. A* **1998**, *102*, 2824–2832.

(152) Kirov, M. V.; Fanourgakis, G. S.; Xantheas, S. S. Identifying the Most Stable Networks in Polyhedral Water Clusters. *Chem. Phys. Lett.* **2008**, *461*, 180–188.

(153) Wales, D. J.; Hodges, M. P. Global Minima of Water Clusters (H2O)n, n≤21, Described by an Empirical Potential. *Chem. Phys. Lett.* **1998**, *286*, 65–72.

(154) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2007**, *120*, 215–241.

(155) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(156) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158.

(157) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(158) Werner, H.-J.; Adler, T. B.; Manby, F. R. General Orbital Invariant MP2-F12 Theory. *J. Chem. Phys.* **2007**, *126*, 164102.

(159) Zangwill, A.; Soven, P. Resonant Photoemission in Barium and Cerium. *Phys. Rev. Lett.* **1980**, *45*, 204–207.

(160) Koopmans, T. Über Die Zuordnung von Wellenfunktionen Und Eigenwerten Zu Den Einzelnen Elektronen Eines Atoms. *Physica* **1934**, *1*, 104–113.

(161) Cox, S. J.; Towler, M. D.; Alfè, D.; Michaelides, A. Benchmarking the Performance of Density Functional Theory and Point Charge Force Fields in Their Description of sI Methane Hydrate against Diffusion Monte Carlo. *J. Chem. Phys.* **2014**, *140*, 174703.

(162) Axilrod, B. M.; Teller, E. Interaction of the van Der Waals Type Between Three Atoms. *J. Chem. Phys.* **1943**, *11*, 299.

(163) Klimeš, J.; Bowler, D. R.; Michaelides, A. Chemical Accuracy for the van Der Waals Density Functional. *J. Phys. Condens. Matter* **2010**, *22*, 022201.

(164) Koh, C. A.; Sloan, E. D. Natural Gas Hydrates: Recent Advances and Challenges in Energy and Environmental Applications. *AIChE J.* **2007**, *53*, 1636–1643.

(165) Okano, Y.; Yasuoka, K. Free-Energy Calculation of Structure-H Hydrates. *J. Chem. Phys.* **2006**, *124*, 024510.

(166) Srivastava, H. K.; Sastry, G. N. Viability of Clathrate Hydrates as CO2 Capturing Agents: A Theoretical Study. *J. Phys. Chem. A* **2011**, *115*, 7633–7637.

(167) Alavi, S.; Woo, T. K. How Much Carbon Dioxide Can Be Stored in the Structure H Clathrate Hydrates?: A Molecular Dynamics Study. *J. Chem. Phys.* **2007**, *126*, 044703.

(168) Fleischer, E. B.; Janda, K. C. Prediction of Clathrate Structure Type and Guest Position by Molecular Mechanics. *J. Phys. Chem. A* **2013**, *117*, 4001–4010.

(169) Kumar, P.; Mishra, B. K.; Sathyamurthy, N. Density Functional Theoretic Studies of Host–guest Interaction in Gas Hydrates. *Comput. Theor. Chem.* **2014**, *1029*, 26–32.

(170) Deible, M. J.; Tuguldur, O.; Jordan, K. D. Theoretical Study of the Binding Energy of a Methane Molecule in a (H2O)20 Dodecahedral Cage. *J. Phys. Chem. B* **2014**, 8257–8263.

(171) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations . I . The Atoms Boron through Neon and Hydrogen. **1989**, 1007–1023.

(172) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple - The PBE Functional. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(173) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158.

(174) R. Bukowski, W. Cencek, P. Jankowski, M. Jeziorska, B. Jeziorski, S. A. Kucharski, V. F. Lotrich, A. J. Misquitta, R. Moszynski, K. Patkowski, R. Podeszwa, S. Rybak, K. Szalewicz, H. L. Williams, R. J. Wheatley, P. E. S. Wormer, and P. S. Z. SAPT2008: An Ab Initio Program for Many-Body Symmetry-Adapted Perturbation Theory Calculations of Intermolecular Interaction Energies.

(175) Weerasinghe, G. L.; Ríos, P. L.; Needs, R. J. Compression Algorithm for Multideterminant Wave Functions. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2014**, *89*, 1–7.

(176) Nukala, P. K. V. V; Kent, P. R. C. A Fast and Efficient Algorithm for Slater Determinant Updates in Quantum Monte Carlo Simulations. *J. Chem. Phys.* **2009**, *130*, 204105.

(177) Filippi, C.; Umrigar, C. J. Multiconfiguration Wave Functions for Quantum Monte Carlo Calculations of First-Row Diatomic Molecules. *J. Chem. Phys.* **1996**, *105*, 213.

(178) Zimmerman, P. M.; Toulouse, J.; Zhang, Z.; Musgrave, C. B.; Umrigar, C. J. Excited States of Methylene from Quantum Monte Carlo. *J. Chem. Phys.* **2009**, *131*, 124103.

(179) Harkless, J. A. W.; Irikura, K. K. Multi-Determinant Trial Functions in the Determination of the Dissociation Energy of the Beryllium Dimer : Quantum Monte Caro Study. *Int. J. Quantum Chem.* **2006**, *106*, 2373–2378.

(180) Anderson, A. G.; Goddard, W. a. Generalized Valence Bond Wave Functions in Quantum Monte Carlo. *J. Chem. Phys.* **2010**, *132*, 164110.

(181) Dubecký, M.; Derian, R.; Jurečka, P.; Mitas, L.; Hobza, P.; Otyepka, M. Quantum Monte Carlo for Noncovalent Interactions: Analysis of Protocols and Simplified Scheme Attaining Benchmark Accuracy. **2014**, 1–7.

(182) Jurecka, P.; Sponer, J.; Cerný, J.; Hobza, P. Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(183) Boys, S. F.; Handy, N. C. A Condition to Remove the Indeterminacy in Interelectronic Correlation Functions. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1969**, *309*, 209–220.

(184) Hesselmann, A. Improved Supermolecular Second Order Møller-Plesset Intermolecular Interaction Energies Using Time-Dependent Density Functional Response Theory. *J. Chem. Phys.* **2008**, *128*, 144112.

(185) Meshkov, V. V; Stolyarov, A. V; Heaven, M. C.; Haugen, C.; LeRoy, R. J. Direct-Potential-Fit Analyses Yield Improved Empirical Potentials for the Ground X (1)Σ(+)(g) State of Be2. *J. Chem. Phys.* **2014**, *140*, 064315.

(186) Merritt, J. M.; Bondybey, V. E.; Heaven, M. C. Beryllium Dimer--Caught in the Act of Bonding. *Science* **2009**, *324*, 1548–1551.

(187) Bondbey, V. E. Laser-Induced Fluorescence and Bonding of Metal Dimers. *Science* **1985**, *227*, 125–131.

(188) Bondybey, V. E. Laser-Induced Fluorescence a Bonding of Metal Dimers. *Science (80-.).* **1985**, *227*, 125–131.

(189) Bondybey, V. E. ELECTRONIC STRUCTURE AND BONDING OF Be2. *Chem. Phys. Lett.* **1984**, *109*, 436–441.

(190) Bender, C. F.; Davidson, E. R. Theoretical Calculation of the Potential Curves of the Be2 Molecule. *J. Chem. Phys.* **1967**, *47*, 4972–4978.

(191) Robb, M. A.; Wilson, S. Comparison within the Algebraic Approximation of Many-Body Perturbation Theory and Configuration Interaction for the Beryllium Dimer. *Mol. Phys.* **2011**, *40*, 1333–1340.

(192) Blomberg, M. R. A.; Siegbahn, P. E. M.; Roos, B. O. The Ground-State Potential Curve of the Beryllium Dimer. *Int. J. Quantum Chem.* **2009**, *18*, 229–247.

(193) Liu, B.; McLean, A. D. Ab Initio Potential Curve for Be2(1Σg+) from the Interacting Correlated Fragments Method. *J. Chem. Phys.* **1980**, *72*, 3418.

(194) Roeggen, I. ; J. A. Interatomic Potential for the X1Σ+g State of Be2. **1996**, *60*, 453–466.

(195) Martin, J. M. L. The Ground-State Spectroscopic Constants of Be 2 Revisited. **1999**.

(196) Stärck, J.; Meyer, W. The Ground State Potential of the Beryllium Dimer. *Chem. Phys. Lett.* **1996**, *4*.

(197) Füsti-Molnár, L.; Szalay, P. G. New Versions of Approximately Extensive Corrected Multireference Configuration Interaction Methods †. *J. Phys. Chem.* **1996**, *100*, 6288–6297.

(198) Gdanitz, R. J. Accurately Solving the Electronic Schrödinger Equation of Atoms and Molecules Using Explicitly Correlated (r12-)MR-CI. *Chem. Phys. Lett.* **1999**, *312*, 578–584.

(199) Harrison, R. J.; Handy, N. C. Full CI Results for Be2 and (H2)2 in Large Basis Sets. *Chem. Phys. Lett.* **1983**, *98*, 97–101.

(200) Bauschlicher, C. W.; Langhoff, S. R.; Partridge, H. Theoretical Study of the BeLi, BeNa, MgLi, MgNa, and AlBe Molecules and Their Negative Ions. *J. Chem. Phys.* **1992**, *96*, 1240.

(201) Røeggen, I.; Veseth, L. Interatomic Potential for the X1 G + State of Be2, Revisited. *Int. J. Quantum Chem.* **2005**, *101*, 201–210.

(202) Helal, W.; Evangelisti, S.; Leininger, T.; Monari, A. A FCI Benchmark on Beryllium Dimer: The Lowest Singlet and Triplet States. *Chem. Phys. Lett.* **2013**, *568-569*, 49–54.

(203) Sheng, X. W.; Kuang, X. Y.; Li, P.; Tang, K. T. Analyzing and Modeling the Interaction Potential of the Ground-State Beryllium Dimer. *Phys. Rev. A* **2013**, *88*, 022517.

(204) Patkowski, K.; Spirko, V.; Szalewicz, K. On the Elusive Twelfth Vibrational State of Beryllium Dimer. *Science (80-.).* **2009**, *326*, 1382.

(205) Patkowski, K.; Podeszwa, R.; Szalewicz, K. Interactions in Diatomic Dimers Involving Closed-Shell Metals. *J. Phys. Chem. A* **2007**, *111*, 12822–12838.

(206) Sharma, S.; Yanai, T.; Booth, G. H.; Umrigar, C. J.; Chan, G. K.-L. Spectroscopic Accuracy Directly from Quantum Chemistry: Application to Ground and Excited States of Beryllium Dimer. *J. Chem. Phys.* **2014**, *140*, 104112.

(207) Lesiuk, M.; Przybytek, M.; Musiał, M.; Jeziorski, B.; Moszynski, R. Reexamination of the Calculation of Two-Center, Two-Electron Integrals over Slater-Type Orbitals. III. Case Study of the Beryllium Dimer. *Phys. Rev. A* **2015**, *91*, 012510.

(208) Koput, J. The Ground-State Potential Energy Function of a Beryllium Dimer Determined Using the Single-Reference Coupled-Cluster Approach. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20311–20317.

(209) Schmidt, M. W.; Ivanic, J.; Ruedenberg, K. Electronic Structure Analysis of the Ground-State Potential Energy Curve of Be(2). *J. Phys. Chem. A* **2010**, *114*, 8687–8696.

(210) Bondybey, V. E.; English, J. H. Laser Vaporization of Beryllium: Gas Phase Spectrum and Molecular Potential of Be2. *J. Chem. Phys.* **1984**, *80*, 568.

(211) Kaledin, L. .; Kaledin, a. .; Heaven, M. .; Bondybey, V. . Electronic Structure of Be2: Theoretical and Experimental Results. *J. Mol. Struct. THEOCHEM* **1999**, *461-462*, 177–186.

(212) Hajgató, B.; Szieberth, D.; Geerlings, P.; De Proft, F.; Deleuze, M. S. A Benchmark Theoretical Study of the Electronic Ground State and of the Singlet-Triplet Split of Benzene and Linear Acenes. *J. Chem. Phys.* **2009**, *131*, 224321.

(213) Koput, J. The Ground-State Potential Energy Function of a Beryllium Dimer Determined Using the Single-Reference Coupled-Cluster Approach. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20311.

(214) Schmidt, M. W.; Ivanic, J.; Ruedenberg, K. Electronic Structure Analysis of the Ground-State Potential Energy Curve of Be(2). *J. Phys. Chem. A* **2010**, *114*, 8687–8696.

(215) Needs, R.; Towler, M.; Drummond, N.; Pablo, L. User's Guide Version 2.5.0. **2009**.

(216) Koput, J.; Peterson, K. A. Ab Initio Prediction of the Potential Energy Surface and Vibration-Rotation Energy Levels of BeH2. *J. Chem. Phys.* **2006**, *125*, 44306.

(217) Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning, T. H.; Wilson, A. K. Gaussian Basis Sets for Use in Correlated Molecular Calculations. VII. Valence, Core-Valence, and Scalar Relativistic Basis Sets for Li, Be, Na, and Mg. *Theor. Chem. Acc.* **2010**, *128*, 69–82.

(218) Ceperley, D. Fermion Nodes. *J. Stat. Phys.* **1991**.

(219) Bressanini, D.; Morosi, G.; Tarasco, S. An Investigation of Nodal Structures and the Construction of Trial Wave Functions. *J. Chem. Phys.* **2005**, *123*, 204109.

(220) Casula, M.; Attaccalite, C.; Sorella, S. Correlated Geminal Wave Function for Molecules: An Efficient Resonating Valence Bond Approach. *J. Chem. Phys.* **2004**, *121*, 7110–7126.

(221) El Khatib, M.; Bendazzoli, G. L.; Evangelisti, S.; Helal, W.; Leininger, T.; Tenti, L.; Angeli, C. Beryllium Dimer: A Bond Based on Non-Dynamical Correlation. *J. Phys. Chem. A* **2014**, *118*, 6664–6673.

(222) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(223) Dykstra, C. An Examination of the Brueckner Condition for the Selection of Molecular Orbitals in Correlated Wavefunctions. *Chem. Phys. Lett.* **1977**, *45*, 466–469.

(224) Handy, N.; Pople, J. Size-Consistent Brueckner Theory Limited to Double Substitutions. *Chem. Phys. …* **1989**, *164*, 185–192.

(225) Ma, a.; Towler, M. D.; Drummond, N. D.; Needs, R. J. Scheme for Adding Electron-Nucleus Cusps to Gaussian Orbitals. *J. Chem. Phys.* **2005**, *122*, 224322.

(226) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Shyjun, S. U.; et al. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

(227) Kim, J.; Esler, K. P.; McMinis, J.; Morales, M. a; Clark, B. K.; Shulenburger, L.; Ceperley, D. M. Hybrid Algorithms in Quantum Monte Carlo. *J. Phys. Conf. Ser.* **2012**, *402*, 012008.

(228) Pachucki, K.; Komasa, J. Relativistic and QED Corrections for the Beryllium Atom. *Phys. Rev. Lett.* **2004**, *92*, 213001.

(229) Kysilka, J.; Rubeš, M.; Grajciar, L.; Nachtigall, P.; Bludský, O. Accurate Description of Argon and Water Adsorption on Surfaces of Graphene-Based Carbon Allotropes. *J. Phys. Chem. A* **2011**, *115*, 11387–11393.

(230) Voloshina, E.; Usvyat, D.; Schütz, M.; Dedkov, Y.; Paulus, B. On the Physisorption of Water on Graphene: A CCSD(T) Study. *Phys. Chem. Chem. Phys.* **2011**, *13*, 12041–12047.

(231) Jenness, G. R.; Karalti, O.; Jordan, K. D. Benchmark Calculations of Water–acene Interaction Energies: Extrapolation to the Water–graphene Limit and Assessment of Dispersion–corrected DFT Methods. *Phys. Chem. Chem. Phys.* **2010**, *12*, 6375–6381.

(232) Nozieres, P.; Pines, D. Correlation Energy of a Free Electron Gas. *Phys. Rev.* **1958**, *111*, 442–454.

(233) Kresse, G.; Harl, J. Accurate Bulk Properties from Approximate Many-Body Techniques. *Phys. Rev. Lett.* **2009**, *103*, 4–7.

(234) Lebègue, S.; Harl, J.; Gould, T.; Ángyán, J. G.; Kresse, G.; Dobson, J. F. Cohesive Properties and Asymptotics of the Dispersion Interaction in Graphite by the Random Phase Approximation. *Phys. Rev. Lett.* **2010**, *105*, 1–4.

(235) Ma, J.; Michaelides, A.; Alfè, D.; Schimka, L.; Kresse, G.; Wang, E. Adsorption and Diffusion of Water on Graphene from First Principles. *Phys. Rev. B* **2011**, *84*, 033402.

(236) Min, seung kyy; Lee, eun cheol; Lee, han myoung; Kim, dong young; Kim, D.; Kim, kwang S. Complete Basis Set Limit of Ab Initio Binding Energies and Geometrical Parameters for Various Typical Types of Complexes. *J. Comput. Chem.* **2008**, *29*, 1208–1221.

(237) Head-Gordon, M.; Pople, J. a.; Frisch, M. J. MP2 Energy Evaluation by Direct Methods. *Chem. Phys. Lett.* **1988**, *153*, 503–506.

(238) Deegan, M. J. O.; Knowles, P. J. Perturbative Corrections to Account for Triple Excitations in Closed and Open Shell Coupled Cluster Theories. *Chem. Phys. Lett.* **1994**, *227*, 321–326.

(239) 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.

(240) Pople, J. A.; Head-gordon, M. Krishnan RAGHAVACHARI, Gary W. TRUCKS. **1989**, *157*, 479–483.

(241) Feller, D.; Jordan, K. D. Estimating the Strength of the Water/Single-Layer Graphite Interaction. *J. Phys. Chem. A* **2000**, *104*, 9971–9975.

(242) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular Potentials Based on Symmetry-Adapted Perturbation Theory with Dispersion Energies from Time-Dependent Density-Functional Calculations. *J. Chem. Phys.* **2005**, *123*.

(243) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van Der Waals Complexes. *Chem. Rev.* **1994**.

(244) King, H. W. *CRC Handbook of Computational Chemsitry and Physics*; 2007.

(245) Courty, A.; Mons, M.; Dimicoli, I.; Piuzzi, F.; Gaigeot, M. P.; Brenner, V.; de Pujo, P.; Millié, P. Quantum Effects in the Threshold Photoionization and Energetics of the

Benzene-H2O and Benzene-D2O Complexes: Experiment and Simulation. *J. Phys. Chem. A* **1998**, *102*, 6590–6600.

(246) Feller, D. Strength of the Benzene - Water Hydrogen Bond. *J. Phys. Chem. A* **1999**, *103*, 7558–7561.

(247) Hachmann, J.; Cardoen, W.; Chan, G. K.-L. Multireference Correlation in Long Molecules with the Quadratic Scaling Density Matrix Renormalization Group. *J. Chem. Phys.* **2006**, *125*, 144101.

(248) Jenness, G. R.; Jordan, K. D. DF-DFT-SAPT Investigation of the Interaction of a Water Molecule to Coronene and Dodecabenzocoronene: Implications for the Water−Graphite Interaction. *J. Phys. Chem. C* **2009**, *113*, 10242–10248.

(249) Perera, F. P. Environment and Cancer: Who Are Susceptible? *Science* **1997**, *278*, 1068–1073.

(250) Craats, A. M. Van De; Warman, J. M.; Müllen, K.; Geerts, Y.; Brand, J. D. Rapid Charge Transport Along Self-Assembling Graphitic Nanowires. *Adv. Mater.* **1998**, *10*, 36–38.

(251) Zacharia, R.; Ulbricht, H.; Hertel, T. Interlayer Cohesive Energy of Graphite from Thermal Desorption of Polyaromatic Hydrocarbons. *Phys. Rev. B - Condens. Matter Mater. Phys.* **2004**, *69*, 1–7.

(252) Podeszwa, R.; Szalewicz, K. Accurate Interaction Energies for Argon, Krypton, and Benzene Dimers from Perturbation Theory Based on the Kohn–Sham Model. *Chem. Phys. Lett.* **2005**, *412*, 488–493.

(253) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H2S. *J. Phys. Chem. A* **2009**, *113*, 10146–10159.

(254) Park, Y. C.; Lee, J. S. Accurate Ab Initio Binding Energies of the Benzene Dimer. *J. Phys. Chem. A* **2006**, *110*, 5091–5095.

(255) Hobza, P.; Selzle, H.; Schlag, E. Potential Energy Surface of the Benzene Dimer: Ab Initio Theoretical Study. *J. Am. …* **1994**, 3500–3506.

(256) Lee, E. C.; Kim, D.; Jurecka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. Understanding of Assembly Phenomena by Aromatic-Aromatic Interactions: Benzene Dimer and the Substituted Systems. *J. Phys. Chem. A* **2007**, *111*, 3446–3457.

(257) *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Cundari, T. R., Eds.; Reviews in Computational Chemistry; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008; Vol. 26.

(258) Chakarova, S. D.; Schröder, E. Van Der Waals Interactions of Polycyclic Aromatic Hydrocarbon Dimers. *J. Chem. Phys.* **2005**, *122*, 54102.

(259) Chakarova-Käck, S. D.; Vojvodic, A.; Kleis, J.; Hyldgaard, P.; Schröder, E. Binding of Polycyclic Aromatic Hydrocarbons and Graphene Dimers in Density Functional Theory. *New J. Phys.* **2010**, *12*, 013017.

(260) Gonzalez, C.; Lim, E. C. A Quantum Chemistry Study of the van Der Waals Dimers of Benzene, Naphthalene, and Anthracene: Crossed (D 2 D) and Parallel-Displaced (C 2 H

) Dimers of Very Similar Energies in the Linear Polyacenes. *J. Phys. Chem. A* **2000**, *104*, 2953–2957.

(261) Podeszwa, R.; Szalewicz, K. Physical Origins of Interactions in Dimers of Polycyclic Aromatic Hydrocarbons. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2735–2746.

(262) Gonzalez, C.; Lim, E. C. Evaluation of the Hartree−Fock Dispersion (HFD) Model as a Practical Tool for Probing Intermolecular Potentials of Small Aromatic Clusters: Comparison of the HFD and MP2 Intermolecular Potentials. *J. Phys. Chem. A* **2003**, *107*, 10105–10110.

(263) Valiev, R. R.; Kopbalina, K. B.; Cherepanov, V. N.; Ibraev, N. K.; Mazhenov, N. A. Theoretical Investigation of the Structural and Spectroscopic Properties of Anthracene Dimers. *Russ. Phys. J.* **2014**, *57*, 95–99.

(264) Matsuoka, T.; Kosugi, K.; Hino, K.; Nishiguchi, M.; Ohashi, K.; Nishi, N.; Sekiya, H. Electronic Spectra of Jet-Cooled Anthracene Dimer: Evidence of Two Isomers in the Electronic Ground State. *J. Phys. Chem. A* **1998**, *102*, 7598–7602.

(265) Chakraborty, T.; Lim, E. C. Study of van Der Waals Clusters of Anthracene by Laser-Induced Fluorescence in a Supersonic Jet: Evidence for Two Structurally Different Dimers. *J. Phys. Chem.* **1993**, *97*, 11151–11153.

(266) Podeszwa, R.; Bukowski, R.; Szalewicz, K. Potential Energy Surface for the Benzene Dimer and Perturbational Analysis of Π-Π Interactions. *J. Phys. Chem. A* **2006**, *110*, 10345–10354.

(267) Erlekam, U.; Frankowski, M.; von Helden, G.; Meijer, G. Cold Collisions Catalyse Conformational Conversion. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3786–3789.

(268) Sorella, S.; Casula, M.; Rocca, D. Weak Binding between Two Aromatic Rings: Feeling the van Der Waals Attraction by Quantum Monte Carlo Methods. *J. Chem. Phys.* **2007**, *127*, 014105.

(269) Wagner, L. K.; Mitas, L. Energetics and Dipole Moment of Transition Metal Monoxides by Quantum Monte Carlo. *J. Chem. Phys.* **2007**, *126*.

(270) Booth, G. H.; Thom, A. J. W.; Alavi, A. Fermion Monte Carlo without Fixed Nodes: A Game of Life, Death, and Annihilation in Slater Determinant Space. *J. Chem. Phys.* **2009**, *131*, 054106.

(271) Shepherd, J. J.; Booth, G.; Gr, A.; Alavi, A. A Full Configuration Interaction Perspective on the Homogeneous Electron Gas. **2011**, 1–5.

(272) Booth, G. H.; Grüneis, A.; Kresse, G.; Alavi, A. Towards an Exact Description of Electronic Wavefunctions in Real Solids. *Nature* **2013**, *493*, 365–370.

(273) Daday, C.; Smart, S.; Booth, G. H.; Alavi, A.; Filippi, C. Full Configuration Interaction Excitations of Ethene and Butadiene: Resolution of an Ancient Question. *J. Chem. Theory Comput.* **2012**, *8*, 4441–4451.

(274) Fracchia, F.; Filippi, C.; Amovilli, C. Size-Extensive Wave Functions for Quantum Monte Carlo: A Linear Scaling Generalized Valence Bond Approach. *J. Chem. Theory Comput.* **2012**, *8*, 1943–1951.