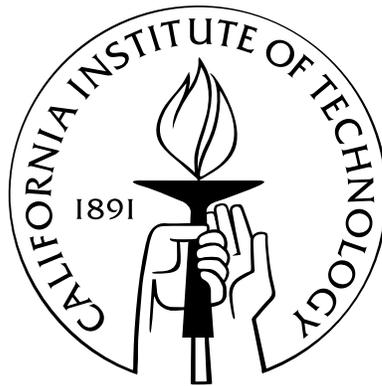


# Development and Applications of Quantum Monte Carlo

Thesis by  
Daniel Ross Fisher

In Partial Fulfillment of the Requirements  
for the Degree of  
Doctor of Philosophy



California Institute of Technology  
Pasadena, California

2010  
(Defended Dec 7, 2009)

© 2010

Daniel Ross Fisher

All Rights Reserved

# Acknowledgements

I would like to thank my friends and family for their support during my graduate studies.

# Abstract

Quantum Monte Carlo is a relatively new class of electronic structure methods that has the potential to calculate expectation values for atomic, molecular, and materials systems to within chemical accuracy. QMC scales as  $O(N^3)$  or better with the size of the system, which is much more favorable than traditional electronic structure methods capable of comparable accuracy. In addition, the stochastic nature of QMC makes it relatively easy to parallelize over multiple processors.

QMC calculations use the Metropolis algorithm to sample the electron density of the system. This method has an inherent equilibration phase, during which the configurations do not represent the desired density and must be discarded. Because the time spent on equilibration increases linearly with the number of processors, this phase limits the efficiency of parallel calculations, making it impossible to use large numbers of processors to speed convergence.

This thesis presents an algorithm that generates statistically independent walker configurations in regions of high probability density, shortening the length of the equilibration phase and ensuring the accuracy of calculations. Shortening the length of the equilibration phase greatly improves the efficiency of large parallel calculations, which will allow QMC calculations to use the next generation of homogeneous, heterogeneous, and distributed computing resources to conduct highly accurate simulations on large systems.

The most common formulation of diffusion Monte Carlo has two sources of error: the time step used to propagate the walkers and the nodes of the trial function. In order to explore these sources of error, DMC calculations were carried out on three pericyclic hydrocarbon reactions using Hartree-Fock, generalized valence bond, and

multiconfiguration self-consistent field trial functions and time steps ranging from  $10^{-4}$  to  $10^{-2}$  au. The results are compared to values from experiment and high quality *ab initio* calculations, as well as the recently developed X3LYP, M06, and XYG3 density functionals. The appropriate time step and trial functions for the reactants, transition states, and products are identified to begin to develop guidelines for researchers carrying out calculations on larger systems.

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>1 Electronic Structure Theory</b>	<b>1</b>
1.1 Quantum Mechanics . . . . .	2
1.1.1 Cusp Conditions . . . . .	3
1.1.2 The Variational Theorem . . . . .	4
1.2 Approximate Methods . . . . .	5
1.2.1 Hartree-Fock . . . . .	6
1.2.2 Post Hartree-Fock Methods . . . . .	8
1.2.2.1 Configuration Interaction . . . . .	8
1.2.2.2 Coupled Cluster . . . . .	9
1.2.2.3 Multiconfiguration SCF . . . . .	10
1.2.3 Perturbation Theory . . . . .	11
1.2.4 Extrapolated Methods . . . . .	12
1.2.5 Density Functional Theory . . . . .	13
<b>2 Parallel Computing</b>	<b>16</b>
2.1 Designing Parallel Algorithms . . . . .	17
2.2 Analyzing Parallel Algorithms . . . . .	18
2.3 Supercomputers . . . . .	19
2.4 Beowulf Clusters and Grid Computing . . . . .	20

<b>3</b>	<b>Random Number Generation</b>	<b>23</b>
3.1	Random Number Generation . . . . .	23
3.1.1	Uniform Random Numbers . . . . .	24
3.1.2	The Transformation Method . . . . .	25
3.1.3	The Von Neumann Method . . . . .	25
3.1.4	The Metropolis Algorithm . . . . .	26
<b>4</b>	<b>Quantum Monte Carlo</b>	<b>28</b>
4.1	Variational Monte Carlo . . . . .	29
4.2	VMC Trial Functions . . . . .	31
4.3	Diffusion Monte Carlo . . . . .	33
4.4	Generating Configurations in DMC . . . . .	36
<b>5</b>	<b>An Optimized Initialization Algorithm to Ensure Accuracy in Quantum Monte Carlo Calculations</b>	<b>40</b>
	<b>Abstract</b>	<b>41</b>
5.1	Introduction . . . . .	42
5.2	The Metropolis Algorithm and the Initialization Catastrophe . . . . .	43
5.3	Walker Initialization . . . . .	46
5.3.1	STRAW . . . . .	48
5.3.2	Equilibration Behavior . . . . .	51
5.3.3	Timing and Spatial Correlation . . . . .	57
5.3.4	Parallel Calculation Efficiency . . . . .	59
5.4	Conclusion . . . . .	62
<b>6</b>	<b>A Quantum Monte Carlo Study of Three Pericyclic Hydrocarbon Reactions</b>	<b>64</b>
	<b>Abstract</b>	<b>65</b>
6.1	Introduction . . . . .	66
6.2	Reactions . . . . .	67

6.3	Experimental and Computational Results . . . . .	68
6.4	Computational Methods . . . . .	69
6.4.1	Quantum Monte Carlo . . . . .	72
6.4.2	QMC Procedures . . . . .	76
6.5	Results and Discussion . . . . .	77
6.5.1	Reaction 1 . . . . .	78
6.5.2	Reaction 2 . . . . .	81
6.5.3	Reaction 3 . . . . .	82
6.5.4	Discussion . . . . .	83
6.6	Conclusion . . . . .	85
<b>7</b>	<b>Conclusion</b>	<b>100</b>
	<b>Bibliography</b>	<b>103</b>

# Chapter 1

## Electronic Structure Theory

The goal of electronic structure theory is to understand the geometries, reactions, and other properties of molecules and materials based on simulations of their electronic structure. The behavior of particles on this scale is governed by the laws of quantum mechanics. Although these laws are well understood, applying them to nontrivial systems leads to equations too complicated to solve exactly. Since exact solutions are not possible, researchers use approximations that trade accuracy for computational tractability.

Approximate electronic structure methods are classified as *ab initio* methods, which are based only on the laws of quantum mechanics, or semiempirical methods, which use experimental results to determine functional forms and fit parameters. Methods of both types are used to understand and predict experimental phenomena such as reaction mechanisms, electrical properties, and biological activity for a wide variety of systems. This chapter contains a very basic introduction to the laws of quantum mechanics and the approximate methods used to apply them to molecular and solid state systems. Further information on quantum mechanics can be found in references [1, 2], while applications to chemistry and materials science are covered in references [3, 4, 5, 6, 7].

## 1.1 Quantum Mechanics

The fundamental postulate of quantum mechanics is that the energy and all other observable properties of an atom or molecule are expressed in its wavefunction, which can be obtained by solving the Schrödinger equation. Exact solutions for the Schrödinger equation are possible for only the simplest systems. Larger systems lead very quickly to equations with too many dimensions to be solvable.

Quantum mechanics dictates several necessary features for  $\psi$ , the wavefunction of a particle. The product of the wavefunction with its complex conjugate,  $\psi^*\psi = |\psi|^2$ , is interpreted as the probability density function for the position of the particle. Since the particle must exist somewhere, integrating  $|\psi|^2$  over all space must give unity. Wavefunctions that satisfy this condition are referred to as normalized. Accordingly, we multiply  $\psi$  by a normalization constant  $c$  so that  $\int d\tau |c\psi|^2 = 1$ .

The mathematical framework of a many particle wavefunction,  $\Psi$ , must account for the fact that electrons are indistinguishable from each other. This means that the spatial probability density,  $\rho$ , cannot vary with the interchange of any two electrons:

$$\rho(1, 2) = |\Psi(1, 2)|^2 = |\Psi(2, 1)|^2 = \rho(2, 1). \quad (1.1)$$

Therefore,

$$\Psi(1, 2) = \pm\Psi(2, 1). \quad (1.2)$$

Particles such as electrons with half-integer spin are fermions, for which the wavefunction is antisymmetric:  $\Psi(1, 2) = -\Psi(2, 1)$ .

The time-dependent Schrödinger equation determines the evolution of the wavefunction of a system with time [8]:

$$i\frac{\partial}{\partial t}|\Psi(X, t)\rangle = \hat{H}|\Psi(X, t)\rangle, \quad (1.3)$$

where  $\hat{H}$  is the Hamiltonian or energy operator of the system,  $t$  is time, and  $X$  is a generalized coordinate that includes the spatial and spin coordinates of the particles

of the system.

The solution to the time-dependent Schrodinger equation can be expanded as

$$|\Psi(X, t)\rangle = \sum_j c_j e^{-iE_j t} |\Phi_j(X)\rangle, \quad (1.4)$$

where the coefficient  $c_j = \langle \Phi_j(X) | \Psi(X, 0) \rangle$  and the  $E_j$  and  $|\Phi_j(X)\rangle$  are the eigenvalues and eigenfunctions of the time-independent Schrodinger equation:

$$\hat{H}|\Phi_j(X)\rangle = E_j|\Phi_j(X)\rangle. \quad (1.5)$$

Because they do not change with time, the eigenfunctions  $|\Phi_j(X)\rangle$  are known as the stationary states of the system. Each stationary state has an associated eigenvalue,  $E_j$ , which can be interpreted as its energy. The stationary states are usually ordered so that  $E_0 \leq E_1 \leq E_2 \leq \dots$ , with the lowest energy state,  $|\Phi_0(X)\rangle$ , being called the ground state.

Because  $\hat{H}$  is a Hermitian operator, its eigenvalues are real and its eigenfunctions are orthogonal to each other and span the space of all possible solutions. They can also be chosen to be normalized, so that

$$\langle \Phi_i(X) | \Phi_j(X) \rangle = \delta_{ij}, \quad (1.6)$$

where  $\delta_{ij}$  is the Kronecker delta:  $\delta_{ij}$  equals 1 if  $i = j$  and 0 otherwise.

### 1.1.1 Cusp Conditions

The Hamiltonian operator for a system of  $N$  electrons and  $K$  nuclei with charges  $Z_L$  and masses  $M_L$  is

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \frac{1}{2} \sum_{L=1}^K \frac{1}{M_L} \nabla_L^2 - \sum_{i=1}^N \sum_{L=1}^K \frac{Z_L}{r_{iL}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{I=1}^K \sum_{J>I}^K \frac{Z_I Z_J}{r_{IJ}}, \quad (1.7)$$

where lowercase indices refer to electrons, uppercase indices refer to nuclei, and  $r_{ij}$  is the distance between particles  $i$  and  $j$ .

In equation 1.7 and throughout this work, atomic units are used, in which  $\hbar = 1$ ,  $m_e = 1$ ,  $|e| = 1$ , and  $4\pi\epsilon_0 = 1$ , where  $m_e$  is the mass of an electron,  $|e|$  is the magnitude of its charge, and  $\epsilon_0$  is the permittivity of free space.

The Coulomb terms in the Hamiltonian diverge when two particles approach each other. The Schrödinger equation can be solved analytically for these configurations because the kinetic and potential terms of the two approaching particles dominate the others. In order for the energy of the system to be finite, a divergence in the kinetic energy must exactly cancel the divergence in the potential. Solving the Schrödinger equation for these configurations to achieve this cancellation leads to the following cusp condition for the wavefunction [9]:

$$\lim_{r_{ij} \rightarrow 0} \frac{\partial \Psi}{\partial r_{ij}} = \frac{\mu_{ij} q_i q_j}{l + 1} \lim_{r_{ij} \rightarrow 0} \Psi, \quad (1.8)$$

where  $\mu_{ij}$  is the reduced mass of the particles,  $q_i$  and  $q_j$  are their charges, and  $l$  is 1 for same-spin electrons and 0 otherwise. An accurate wave function must satisfy the cusp condition for each pair of particles in the system.

### 1.1.2 The Variational Theorem

The most powerful tool researchers have in constructing approximate ground state wavefunctions is the variational principle, which provides a way to compare their quality. The exact eigenfunctions,  $|\Phi_i(X)\rangle$ , of the Hamiltonian span the space of all possible wavefunctions for the system. Therefore, any normalizable trial wavefunction,  $|\Psi_T(X)\rangle$ , that satisfies the boundary conditions of the system can be expanded in terms of the  $|\Phi_i(X)\rangle$ :

$$|\Psi_T(X)\rangle = \sum_i b_i |\Phi_i(X)\rangle, \quad (1.9)$$

$$b_i = \langle \Phi_i(X) | \Psi_T(X) \rangle. \quad (1.10)$$

The expansion can be used to calculate the expectation value for the energy of

the trial wavefunction:

$$\langle E_T \rangle = \frac{\langle \Psi_T(X) | \hat{H} | \Psi_T(X) \rangle}{\langle \Psi_T(X) | \Psi_T(X) \rangle} = \frac{\sum_i |b_i|^2 E_i}{\sum_i |b_i|^2} \geq E_0, \quad (1.11)$$

where the equality applies if  $|\Psi_T(X)\rangle = |\Phi_0(X)\rangle$ .

The expectation value of the energy of a trial wavefunction is an upper bound to the ground state energy. The closer the trial wavefunction is to the actual ground state, the lower its energy will be. This provides a way to approximate the ground state. First, a parametrized wave function is constructed with a form that can easily be evaluated. Then the parameters are adjusted to minimize the expectation value of the energy. This is the closest approximation to the ground state in the space of the adjustable parameters. Physical arguments must be used in choosing the form of the trial wavefunction: it determines the restrictions on the interactions that can be described and therefore represents a model.

## 1.2 Approximate Methods

In most cases, the first simplification to the Schrödinger equation is the Born-Oppenheimer approximation, which makes use of the fact that the masses of nuclei are much greater than that of an electron. The electrons see the heavy, slow-moving nuclei as almost stationary charges, and the nuclei see the much faster electrons as essentially a three-dimensional distribution of charge. The Born-Oppenheimer approximation simplifies the molecular problem by treating the electronic and nuclear motions separately [10].

In this method, one assumes a fixed configuration for the nuclei, and for this configuration solves an electronic Schrödinger equation to find the electronic wave function and energy. This process is repeated for different configurations to give the electronic energy as a function of the positions of the nuclei. The nuclear configuration that minimizes the energy is the equilibrium geometry of the molecule. The electronic energy can be used as the potential energy function in a Schrödinger equation for the nuclear motion, which can be solved to give the molecular vibrational and rotational

energy levels for a given electronic state.

### 1.2.1 Hartree-Fock

The basis for almost all methods to solve the electronic part of the Schrödinger equation is the Hartree-Fock (HF) method. In HF, the trial wavefunction is expressed as an antisymmetric product of normalized, orthogonal molecular orbitals,  $\psi_i$ . The simplest way to construct a trial wavefunction from a set of orbitals is to use a Slater determinant, a framework that ensures the antisymmetry of the overall wavefunction [11]:

$$\Psi^{AS}(x_1, x_2, \dots, x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(x_1) & \psi_2(x_1) & \cdots & \psi_N(x_1) \\ \psi_1(x_2) & \psi_2(x_2) & \cdots & \psi_N(x_2) \\ \vdots & \vdots & & \vdots \\ \psi_1(x_N) & \psi_2(x_N) & \cdots & \psi_N(x_N) \end{vmatrix}. \quad (1.12)$$

In equation 1.12,  $x_i$  contains the space and spin coordinates of electron  $i$ . Since the determinant of a matrix changes sign if two rows or columns are interchanged, the overall wavefunction will have the proper antisymmetry with respect to permutation of the electrons.

The molecular orbitals can be factored into spatial and spin components:

$$\psi(x) = \psi(\vec{r}, \gamma) = \phi(\vec{r}) \chi(\gamma), \quad (1.13)$$

where  $\phi$  is a spatial orbital and  $\chi$  is a spin function, either  $\alpha$  or  $\beta$ . The spatial orbitals are written as linear combinations of basis functions:

$$\phi = \sum_{\mu} c_{\mu} \chi_{\mu}, \quad (1.14)$$

where the  $c_{\mu}$  are the molecular orbital expansion coefficients. The basis functions,  $\chi_{\mu}$ , are usually centered on the nuclei and resemble atomic orbitals, but any normalizable functions can be used. Because of the ease with which they are evaluated, Gaussian

type orbitals (GTO) are usually used as basis functions. GTOs have the following radial part:

$$\chi_{\mu}^{GTO}(r) = d_{\mu} r^{l_{\mu}} \exp(-\alpha_{\mu} r^2). \quad (1.15)$$

Since the derivative of a Gaussian is zero at its origin, these functions cannot satisfy the electron-nuclear cusp condition of section 1.1.1. The resulting multicenter integrals can be evaluated analytically, however, so a large number of Gaussians can be used in the basis set with little computational expense.

Slater type orbitals (STO) have the correct form to satisfy both the electron-nuclear cusp condition and the long range behavior of molecular orbitals, but are not typically used in basis sets because they lead to very complicated integrals in calculations:

$$\chi_{\mu}^{STO}(r) = d_{\mu} r^{l_{\mu}} \exp(-\alpha_{\mu} r). \quad (1.16)$$

Since the molecular orbitals are constructed from the basis functions, the basis set restricts them to certain shapes and regions of space. The more functions in a basis set, the more flexibility it has to approximate molecular orbitals. Larger basis sets generally produce better results in computations, but require more computer time. Since an electron has a finite probability of existing anywhere in space, an infinite basis set would be necessary to completely describe its possible position.

In order to solve for the orbital expansion coefficients, the Hartree-Fock method makes use of the variational principle. Minimizing the expectation value of the energy of the wavefunction leads to a series of equations, which can be written in matrix form:

$$FC = SC\epsilon, \quad (1.17)$$

where each element is a matrix. The Fock matrix,  $F$ , represents the average effects of the field of all the electrons on each orbital. The matrix  $C$  contains the orbital coefficients,  $S$  indicates the overlap between the orbitals, and  $\epsilon$  is a diagonal matrix of the orbital energies.

Both the Fock matrix and the orbitals depend on the molecular orbital coefficients.

Thus equation 1.17 is not linear and must be solved with an iterative procedure called the self-consistent field (SCF) method. First, an initial guess for the orbital coefficients is formed, and the corresponding density matrix is constructed. Using it, the Fock matrix is formed. Then, solving the eigenvalue problem yields a new set of orbital coefficients. This procedure is repeated until both the orbital coefficients and the energy have converged. At this point, the orbitals generate a field that produces the same orbitals. This method produces both occupied and virtual (unoccupied) orbitals. The total number of orbitals formed is equal to the number of basis functions.

Solving the eigenvalue problem is the slowest step of the process. It involves diagonalizing a matrix, a process that scales as  $O(N^3)$ , where  $N$  is the linear size of the matrix. In this case,  $N$  is the number of basis functions.

## 1.2.2 Post Hartree-Fock Methods

The errors of Hartree-Fock are due to the fact that it treats the repulsion of the electrons for each other in an average way and neglects the details of their motion. The shape of the orbital an electron occupies is determined by the potential field of the nuclei and the density of the other occupied orbitals. An electron sees only the “mean field” of the other electrons, which allows them to come close together more often than they should and makes it impossible for the wavefunction to satisfy the electron-electron cusp conditions of section 1.1.1. The difference in energy that would result from properly allowing the electrons to avoid each other is called the *correlation energy*. Several methods go beyond Hartree-Fock and attempt to treat this phenomenon properly.

### 1.2.2.1 Configuration Interaction

The configuration interaction (CI) method uses the virtual orbitals generated by Hartree-Fock in addition to the occupied orbitals to construct a wavefunction as a linear combination of Slater determinants. The determinants are formed by exciting electrons from the ground state occupied orbitals into the virtual orbitals, and the

expansion coefficients are found by diagonalizing the resulting Hamiltonian matrix:

$$\Psi_{CI}(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N) = \sum_m a_m \Psi_m^{AS}(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N). \quad (1.18)$$

If a Slater determinant corresponding to every possible occupation of the orbitals is included in the expansion, the calculation is a “full CI.” In most cases, full CI is impossible because the number of possible Slater determinants is too large.

In practice, CI calculations are usually carried out by including a limited number of determinants in the expansion. A CI singles (CIS) calculation excites one electron at a time into a virtual orbital, a CI doubles (CID) excites two at a time, a CISD calculation includes singles and doubles, a CISDT calculation includes singles, doubles, and triples, etc. CI calculations can provide quantitative results (within 2 kcal/mol) for energies of molecules, but are extremely time consuming and require immense amounts of memory, even for small systems and minimal basis sets. In addition, the correlation energy recovered scales poorly with the number of configurations included.

### 1.2.2.2 Coupled Cluster

In coupled cluster (CC) calculations, the trial wavefunction is expressed as a linear combination of Slater determinants, but an exponential form of an excitation operator is used to generate the configurations and calculate the energy [12]:

$$|\Psi_{CC}\rangle = \exp(\hat{T}) |\Psi_{HF}\rangle. \quad (1.19)$$

The excitation operator,  $\hat{T}$ , makes Slater determinants by exciting electrons from the ground state into virtual orbitals. Equation 1.19 can be expanded in a Taylor series:

$$\begin{aligned} |\Psi_{CC}\rangle &= \exp(\hat{T}) |\Psi_{HF}\rangle \\ &= |\Psi_{HF}\rangle + \hat{T}_1 |\Psi_{HF}\rangle + \left(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right) |\Psi_{HF}\rangle \\ &\quad + \left(\hat{T}_3 + \hat{T}_1\hat{T}_2 + \frac{1}{6}\hat{T}_1^3\right) |\Psi_{HF}\rangle + \dots, \end{aligned} \quad (1.20)$$

where  $\hat{T}_1$  creates single excitations,  $\hat{T}_2$  creates double excitations, and so on. In equation 1.20, the terms are grouped into levels of excitation. At each level of excitation, several terms contribute. At the second level, for example,  $\hat{T}_2$  generates *connected* double excitations, while  $\hat{T}_1^2$  generates two *disconnected* single excitations.

Coupled cluster makes it easy to treat molecules of different sizes with the same level of correlation, which is important for chemical reactions, in which bonds may form or a large molecule may dissociate into fragments. Treating the products and reactants of a reaction consistently is necessary to get accurate energy differences.

Like configuration interaction, coupled cluster calculations are named by the levels of excitation included in the expansion. A CCSD calculation includes single and double excitations, while CCSD(T) includes triples as a perturbation. CCSD(T) is a very popular method for conducting accurate calculations with reasonable cost, and is often used as a benchmark to compare the results of other methods. The expense of CCSD(T) scales as  $O(N^7)$  with the number of basis functions, which limits its application to small molecules and basis sets.

### 1.2.2.3 Multiconfiguration SCF

In a multiconfiguration self consistent field (MCSCF) calculation, the user defines an “active space” consisting of a subset of the electrons and orbitals of a molecule. The active electrons are excited into the active virtual orbitals to form a set of determinants, and both the orbitals and CI expansion coefficients are variationally optimized [13]. If a full CI is carried out on the active space, and all possible occupations of the active orbitals are considered, the calculation is called a complete active space SCF (CASSCF) [14] or fully optimized reaction space (FORS) [15] calculation. Because both the orbitals and CI coefficients are optimized, MCSCF offers the most general approach available to computing electronic structure. The large number of variational parameters makes the optimization a challenge, so users must be careful to only include the electrons and orbitals involved in the reaction under investigation in the active space.

The generalized valence bond (GVB-PP) method can be thought of as a limited

form of MCSCF in which electrons are excited pairwise from valence orbitals into virtual orbitals [16]. Although the selection of configurations is constrained, the optimization procedure for GVB calculations is much more systematic and reliable than a general MCSCF calculation. The GVB wavefunction is the simplest form that allows molecules to dissociate into open shell fragments, which allows it to produce accurate dissociation curves for chemical bonds.

### 1.2.3 Perturbation Theory

Møller-Plesset (MP) perturbation theory is a non-iterative method for calculating the correlation energy of a set of orbitals. In perturbation theory, the Hamiltonian is divided into two parts:

$$\hat{H} = \hat{H}_0 + \lambda\hat{V}, \quad (1.21)$$

where  $\hat{H}_0$  is exactly solvable and  $\lambda\hat{V}$  is a perturbation that is assumed to be small compared to it. The perturbed wavefunction can be expanded as a power series in  $\lambda$ :

$$\Psi = \Psi^0 + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \lambda^3\Psi^{(3)} + \dots \quad (1.22)$$

The perturbed wavefunction is substituted into the Schrödinger equation:

$$\left(\hat{H}_0 + \lambda\hat{V}\right) \left(\Psi^0 + \lambda\Psi^{(1)} + \dots\right) = \left(E^0 + E^{(1)} + \dots\right) \left(\Psi^0 + \lambda\Psi^{(1)} + \dots\right). \quad (1.23)$$

Equating terms with the same power of  $\lambda$  gives formulas for corrections to the energy for varying lengths of the expansion.

In electronic structure theory, the unperturbed Hamiltonian and wavefunction are the Fock operator and its ground state Slater determinant. The perturbation,  $\hat{V}$ , is the Coulomb repulsion between the electrons, which is replaced with the mean field approximation in Hartree-Fock. The second-order correction to the energy involves

integrals between determinants:

$$E^{(2)} = - \sum_t \frac{|\langle \Psi^0 | r_{12}^{-1} | \Psi_t \rangle|^2}{E_t - E^0}, \quad (1.24)$$

where the index  $t$  sums over determinants in which two electrons have been excited into virtual orbitals. It is easy to see from the denominator of equation 1.24 that the greatest contributions to the second-order correction will come from low-lying excited states, whose energy is close to the ground state energy.

Møller-Plesset perturbation theory is referred to by the order of the expansion of the perturbation. Second order (MP2) is commonly used, and third (MP3) and fourth (MP4) order are implemented in many quantum chemistry programs. Multireference MP theory, in which an MCSCF or CI wavefunction is used as the unperturbed wavefunction, has also been developed [17].

While MP2 generally gives good results for molecular geometries and changes in energy for chemical reactions, recent studies comparing levels of perturbation for different chemical systems and basis sets have shown that MP perturbation theory energies are not necessarily convergent in the limit of higher orders of perturbation [18].

### 1.2.4 Extrapolated Methods

Several methods have been developed to approximate an extremely expensive calculation by systematically combining less accurate results. Although multiple calculations are run, the overall cost can be significantly less than that of the single highly accurate calculation.

The complete basis set (CBS) methods address the errors due to using a finite basis set in calculations. They extrapolate to an infinite basis using expressions for the correlation energy recovered for electron pairs as functions of higher angular momentum are included in the basis set [19]. A CBS calculation consists of a Hartree-Fock calculation with a large basis set, an MP2 calculation with a moderate basis set, and higher level calculations with progressively smaller basis sets. The results and several empirical corrections are combined to estimate the results that would be

obtained for a high level calculation with an infinite basis set.

The Gaussian-1 (G1) method approximates a quadratic CISD(T) result with a large basis set using four smaller calculations [20]. It corrects for truncation of the basis set by carrying out MP4 calculations with three different basis sets and for the limited level of correlation by carrying out a quadratic CISD(T) calculation with the smallest basis set. The results are entered into a formula that includes some empirical corrections for the remaining systematic errors to give the G1 energy. G2 [21], G3 [22], and G4 [23] methods have subsequently been developed.

The focal point method explicitly examines the convergence of the energy with respect to both the basis set and level of correlation to estimate the *ab initio* limit within the Born-Oppenheimer approximation [24]. In the focal point procedure, HF energies are extrapolated to the CBS limit, and CCSDT and CCSDT(Q) calculations are carried out using a moderate basis set. The results are combined to estimate the CBS limit of the CCSDT(Q) energy. Corrections for non-Born-Oppenheimer [25] and special relativistic effects [26] are added to give the focal point result.

### 1.2.5 Density Functional Theory

Density functional theory (DFT) is another widely used class of methods for taking into account the effects of electron correlation. DFT is based on the theorem of Hohenberg and Kohn, which proves the existence of a functional that determines the exact electron density and energy for a given a nuclear potential field [27]. Unfortunately, the theorem does not provide the form of the exact functional. While the exact functional would take an electron density as input and return the energy, approximate functionals partition the energy into several terms [28]:

$$E = E^T + E^V + E^J + E^{XC}. \quad (1.25)$$

The first three terms correspond to the kinetic energy, the attraction between the nuclei and the electrons, and the repulsion of the electrons for each other. The fourth is called the *exchange-correlation* term and includes the remaining interactions

between the electrons.

In principle, a pure density functional method would deal directly with the electron density, a function of the three spatial variables. No orbitals would be involved, and calculations would scale linearly with the size of the system. In practice, however, a method similar to Hartree-Fock is used. The wavefunction is written as a Slater determinant of orbitals, and the Fock operator is replaced with one that takes the effects of electron correlation into account.

The exchange-correlation energy of equation 1.25 is separated into exchange and correlation terms. The exchange energy arises from the interactions between same spin electrons, which are kept apart by the antisymmetry of the spatial part of the wave function. The correlation energy is due to the interactions between opposite spin electrons.

The exchange and correlation energy terms are calculated by functionals of the density. The basis for most functionals is the *local density approximation*, in which electrons uniformly occupy a volume with a positive background charge to keep the overall charge neutral. For this system, the exchange energy has a simple form:

$$E_{LDA}^X = -\frac{3}{2} \left( \frac{3}{4\pi} \right)^{\frac{1}{3}} \int d\tau \rho^{\frac{4}{3}}. \quad (1.26)$$

Local correlation functionals are more complicated, but are also in use [29].

The electron density of atoms and molecules, however, is not uniform, so researchers have developed exchange and correlation functionals that use the gradient of the density as well as its value [30, 31].

Some of the most accurate density functional methods in use are *hybrid functionals*, in which the Hartree-Fock definition of the exchange energy, which is based on molecular orbitals, is included as a component of the exchange-correlation energy [32, 33]. The exchange-correlation energy term for B3LYP, one of the most popular density functional methods, includes local, gradient corrected, and Hartree-

Fock terms:

$$E_{B3LYP}^{XC} = E_{LDA}^X + c_0 (E_{HF}^X - E_{LDA}^X) + c_X \Delta E_{B88}^X + E_{VWN3}^C + c_C (E_{LYP}^C - E_{VWN3}^C), \quad (1.27)$$

where  $\Delta E_{B88}^X$  is a gradient-corrected exchange term,  $E_{VWN3}^C$  is a local correlation term, and  $E_{LYP}^C$  is a gradient-corrected correlation term. The coefficients  $c_0$ ,  $c_X$ , and  $c_C$  were fit to experimental data.

Density functional theory is a very popular way for researchers to include electron correlation in calculations and obtain results that are accurate enough for many applications with moderate computational expense. These methods have been applied successfully to a large variety of systems and have been a benefit to many areas of research. While post-Hartree-Fock methods can always be improved by including more configurations, using a larger basis set, or calculating higher orders of perturbation, DFT suffers from the fact that there is no systematic way to improve its results. New density functionals are continuously being developed [34, 35, 36], but none give results with errors consistently less than 4 to 5 kcal/mol for molecular systems. If more accurate results are necessary, different methods must be used.

# Chapter 2

## Parallel Computing

The extraordinary increase in computing power available to researchers over the last fifty years has revolutionized engineering, astronomy, biology, chemistry, physics, economics, and many other fields. The long term trend in the number of circuits that can be placed on an integrated circuit inexpensively was described in 1965 by Gordon Moore:

The complexity for minimum component costs has increased at a rate of roughly a factor of two per year .... Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years. That means by 1975, the number of components per integrated circuit for minimum cost will be 65,000. I believe that such a large circuit can be built on a single wafer [37].

In 1975, Moore's prediction for the time to double the number of transistors on a circuit was revised to 18 months. The trends in almost every measure of electronic devices, such as processing speed, memory capacity, and computing performance per unit cost, are closely related to Moore's law.

It has often been predicted that chip designers would not be able to keep up with Moore's law. Gordon Moore himself has stated that the rate of increase in computing power cannot be sustained indefinitely, but it has been sustained through 2009, with

chip makers predicting new processors consistent with Moore's law for another ten years.

Not satisfied with the computing power available in a single processor, researchers have developed techniques for parallel computing, in which multiple processors connected with a network are used together to solve a problem. This chapter describes some of the ways parallel programs are designed and analyzed. Additional information can be found in reference [38].

## 2.1 Designing Parallel Algorithms

In order for an algorithm to be executed in parallel, the programmer must decompose the work into tasks and identify which tasks can be executed concurrently. The concurrent tasks can be assigned to different processors to be executed. In order for a processor to be able to complete its task, the appropriate instructions, input, and output must be communicated.

The *granularity* of a problem refers to the number and size of the tasks into which it can be decomposed. The *degree of concurrency* is the number of tasks that can be executed simultaneously. This number is usually less than the total number of tasks due to dependencies among them.

There are many techniques for decomposing a problem into tasks. The nature of the problem determines how it can be divided and how the different tasks interact with each other. A problem for which a fine-grained decomposition into independent tasks is possible is well suited for parallel computing, and will benefit greatly from being carried out on multiple processors. A less ideal application may benefit less from the parallel environment. In especially unfavorable circumstances, such as if many processors are idle while they wait for another task to supply them with input, or if interprocessor communication saturates the bandwidth of the network, an application may take longer to execute in parallel than on a single processor.

## 2.2 Analyzing Parallel Algorithms

Using twice as many processors to execute a program rarely results in it completing in half the time or generating twice as much useful output. Overhead expenses, unavoidable in the parallel environment, subtract from the performance. Computer scientists have developed several metrics to measure the expense and performance of parallel algorithms.

The most basic measure of performance is the *speedup*, the ratio of the serial execution time,  $T_S$ , to the parallel execution time,  $T_P$ :

$$\text{Speedup} = \frac{T_S}{T_P}. \quad (2.1)$$

The largest source of overhead is usually communication of data between processors. In addition, some processors may become idle if they finish their task and must wait for a new one. The parallel algorithm may also have to carry out excess computation compared to the serial algorithm. For example, if the result of a certain calculation must be available to each task, it may have to be carried out separately on each processor in a parallel calculation, while the serial algorithm only has to carry out the calculation once.

The overhead for a parallel algorithm is the difference between the parallel and serial costs:

$$T_o = pT_P - T_S, \quad (2.2)$$

where  $p$  is the number of processors.

An important measure for the effectiveness of a parallel algorithm is its *efficiency*, which is the ratio of the serial cost to the parallel cost:

$$E = \frac{T_S}{pT_P} \quad (2.3)$$

$$= \frac{T_S}{T_S + T_o}. \quad (2.4)$$

Because every algorithm has at least some serial component, the parallel overhead

increases with the number of processors. As can be seen in equation 2.4, an increase in overhead causes a decrease in efficiency. The loss of efficiency leads to decreasing returns as more processors are used to execute an algorithm.

The decreasing gain in performance as the number of processors increases is expressed in a slightly different way in *Amdahl's Law* [39]. If  $P$  is the fraction of an algorithm that can be parallelized and  $S$  is the fraction that must be computed serially, the speedup can be written as a function of the number of processors:

$$\text{Speedup}(p) = \frac{S + P}{S + \frac{P}{p}} \quad (2.5)$$

$$= \frac{1}{S + \frac{P}{p}}. \quad (2.6)$$

As the number of processors increases, the speedup asymptotically approaches  $\frac{1}{S}$ . According to this formula, if the serial portion of an algorithm is 10%, the greatest possible speedup is ten times, no matter how many processors are used. As a result, much of the effort in designing parallel algorithms goes into parallelizing as much of the work as possible.

After examining Eqs 2.4 and 2.6, it would be easy to become skeptical as to the viability of massively parallel computers, since the benefit of using more processors is bounded. In practice, however, the size of the problem usually increases with the number of processors. When given more processors, researchers will usually increase the size or complexity of the problem to keep the run time approximately constant. As the problem size increases, the fraction of the run time spent on overhead decreases, which improves the efficiency for large numbers of processors.

## 2.3 Supercomputers

Since 1993, the Top500 list has kept track of the most powerful supercomputers in the world [40]. The United States Department of Energy has constructed several of the highest ranking machines to conduct simulations on nuclear weapons through

the Advanced Simulation and Computing program [41]. These are homogeneous machines, meaning they are constructed from one type of processor, with huge memory and fast interconnection hardware.

Several calculations presented later in this work were carried out using Blue Gene/L at Lawrence Livermore National Laboratory. The unclassified portion of this machine consists of 81,920 IBM PowerPC processors running at 700 MHz [42]. Although the processor speed is slow, the great number of processors gives uBGL almost 230 TFlops of computing power.

The most powerful computer in the world at the moment is Roadrunner at Los Alamos National Laboratory. Roadrunner has 13,824 1.8GHz AMD Opteron processors to handle operations, communications, and some computation and 116,640 IBM PowerXCell 8i processors to handle floating point-operations. Roadrunner is the first machine to have over 1 petaflop sustained performance [43].

Not to be outdone, LLNL has announced they will be constructing Sequoia, a Blue Gene/Q machine that will exceed 20 petaflops, to go online in 2011. Sequoia will have more computing power than the current Top500 list combined [44].

The rate of escalation in computing power is easy to see. Access to the DOE machines is difficult to obtain, however, and most researchers do not have the resources to construct and maintain this sort of supercomputer. The advances in processors, interconnection hardware, and management software brought about by the DOE project have improved the performance of the machines an individual research group can afford.

## 2.4 Beowulf Clusters and Grid Computing

In contrast to the massively expensive homogeneous computers of the previous section, researchers can assemble a low cost cluster using off-the-shelf processors and connection hardware in a Beowulf framework [45]. Such a cluster can be homogeneous if one type of processor is used, or heterogeneous if the processors are not equivalent. This sort of cluster is also scalable, as researchers can add processors as

applications demand and resources allow, or retire processors as they become obsolete.

Another development in parallel computing is the use of loosely coupled, widely distributed grids of processors to carry out computations. Some examples include SETI@home [46] and Folding@home [47], which use idle internet connected computers to search for extraterrestrial intelligence and simulate protein folding. BOINC is a project at UC Berkeley that has tools to help researchers develop software for and connect to distributed volunteer computing resources [48].

Parallel algorithms must have certain characteristics in order to perform well in a heterogeneous, loosely coupled environment. An application that must communicate large amounts of data among the tasks or is unable to balance the work between processors running at different speeds will encounter large overhead costs and perform very poorly. A parallel algorithm with low communications requirements and the ability to use processors running at different speeds will be able to efficiently use inexpensive computing resources to carry out large computing jobs.

The bulk of the computing effort in traditional electronic structure methods such as those discussed in chapter 1 is spent diagonalizing matrices. This operation is very difficult to parallelize, since each step involves all of the rows. As a result, calculations such as DFT and coupled cluster are unable to efficiently use more than a few tens of processors. Since coupled cluster scales as  $O(N^7)$  with the size of the system, the inability to use large numbers of processors prevents researchers from carrying out highly accurate calculations on large systems, such as nanodevices or biological systems.

Quantum Monte Carlo (QMC) is an alternate approach to electronic structure simulations that calculates expectation values stochastically rather than analytically. The stochastic nature of QMC makes it well suited for parallel implementation. QMC can, in principle, calculate exact expectation values and scales as  $O(N^3)$  with the size of the system. QMC can be formulated to have very small memory and communications requirements and automatically balance the work between processors running at different speeds [49, 50]. The favorable scaling of QMC and the ability to efficiently use large numbers of processors will allow it to provide highly accurate expectation

values for systems too large for other methods.

## Chapter 3

# Random Number Generation

Random numbers have applications in areas such as cryptography, electronic gaming, and statistical sampling and analysis. In addition, stochastic, or non-deterministic, simulations can be used to model many types of physical and mathematical systems. In these simulations, the behavior of some part of the system is randomly generated. Because of the essential role played by random numbers, they are grouped into a class called “Monte Carlo” methods. Electronic structure applications include Variational Monte Carlo (VMC), in which the parameters of a trial wavefunction are optimized, and Diffusion Monte Carlo (DMC), which has the potential to calculate exact expectation values for many-body quantum mechanical systems.

### 3.1 Random Number Generation

Truly random numbers can be generated based on unpredictable physical phenomena, such as the noise of an analog circuit, the decay of radioactive nuclei [51], or background atmospheric radio noise [52]. Computers, on the other hand, only operate based on programmed instructions. They can generate sequences of “pseudorandom” numbers that lack patterns, but are determined by a formula. Statistical tests have been developed to detect correlations in sequences of numbers. The quality of a pseudorandom number generator is judged by which tests for randomness its sequences pass.

### 3.1.1 Uniform Random Numbers

Uniform random numbers lie within a specified range, usually 0 to 1, with all numbers in the range having the same probability of being generated. Virtually every scheme to generate random numbers with respect to a desired probability density relies on converting uniform random numbers.

The most common way to generate uniform random numbers is with a linear congruential, or modulo, generator, which generates a series of integers,  $\{I_0, I_1, I_2, \dots\}$ , by the recurrence relation

$$I_{j+1} = aI_j + c \pmod{m}, \quad (3.1)$$

where  $m$ ,  $a$ , and  $c$  are positive integers called the modulus, multiplier, and increment. They define the linear congruential generator. The first integer,  $I_0$ , is called the *seed*. Using the same seed with a certain generator will always give the same sequence of numbers.

Clearly,  $I_j < m$  for all  $j$ . Therefore, the algorithm can generate at most  $m$  distinct integers. The sequence of integers is transformed into uniform random numbers between 0 and 1 by letting  $u_j = \frac{I_j}{m}$ .

The sequence  $\{I_j\}$  generated by equation 3.1 will eventually repeat itself with a period  $p$  that is less than or equal to  $m$ . If  $m$ ,  $a$ , and  $c$  are properly chosen, the period will be of maximal length. Several rules have been developed and implemented to maximize  $p$  and give the best results in statistical tests for randomness [53].

Poor choices of  $a$ ,  $c$ , and  $m$ , can result in random number sequences with very short periods. Many linear congruential generators implemented as library routines in compilers have been shown to be deeply flawed and give poor results in statistical tests.

### 3.1.2 The Transformation Method

Monte Carlo simulations often require random numbers distributed with respect to a given probability density function,  $\rho(x)$ . The most efficient way to generate such a sequence is with the transformation method, which directly converts uniform random numbers to the desired density.

The cumulative distribution function represents the probability that a point in the given density is less than or equal to  $y$ :

$$P(y) = \int_{-\infty}^y dx \rho(x). \quad (3.2)$$

If  $\rho(x)$  is normalized,  $P(y)$  will increase monotonically from 0 to 1.

To generate a random number,  $w$ , distributed with respect to  $\rho(x)$ , a uniform random number,  $u$ , is generated. Then  $w = P^{-1}(u)$ , where  $P^{-1}$  is the inverse of  $P$ . This method requires that the function  $P$  be known and invertible, which is the case for some very simple distributions, such as exponential or Gaussian distributions. For more complicated functions, different algorithms must be used.

### 3.1.3 The Von Neumann Method

The Von Neumann, or rejection, method is a less efficient but more generally applicable way to generate points with respect to a probability density function that is known and can be calculated. The cumulative distribution function does not have to be known or invertible.

In order to use this method, one first finds a function,  $h(x)$ , that is everywhere greater than and preferably close to the desired probability density function,  $\rho(x)$ , and for which the transformation method can be used. A random number,  $z$ , is generated with respect to  $h(x)$  and the ratio  $A(z) = \frac{\rho(z)}{h(z)}$  is calculated. Because  $h(x)$  is always greater than  $\rho(x)$ , this ratio will be between 0 and 1.

The number  $z$  is accepted as a member of the probability density  $\rho$  with probability  $A(z)$ . This last step involves generating a uniform random number,  $u$ , and accepting

$z$  if  $u < A(z)$  and rejecting  $z$  if  $u > A(z)$ . The effect of the rejection step is to weight the density  $h(x)$  by  $\frac{\rho(x)}{h(x)}$  so that  $\rho(x)$  emerges. This method is very simple, but will lead to excessive rejection and be very inefficient if  $h(x)$  is not close to  $\rho(x)$ , in which case the acceptance probability  $A(z)$  will often be small. This loss of efficiency is particularly important for high-dimensional spaces.

### 3.1.4 The Metropolis Algorithm

Quantum Monte Carlo calculations require random electronic configurations distributed with respect to the quantum mechanical probability density, the square of the magnitude of the electronic wavefunction. This is an extremely complicated and tightly coupled  $3N$ -dimensional function, where  $N$  is the number of electrons. Furthermore, it has appreciable magnitude only in a very small fraction of the total available configuration volume. The transformation and rejection methods are unable to efficiently generate random points with respect to this sort of probability density.

In order to distribute electronic configurations with respect to their quantum mechanical probability density, the idea of generating statistically independent configurations must be abandoned. Instead, a Markov chain is used, in which each new configuration is generated with respect to a probability distribution depending on the previous configuration. The sequence of configurations forms a “random walk” that is proportional to the desired density. Because each configuration depends on the one before it, they will have some degree of *serial correlation*, which must be considered when the variance of quantities derived from these configurations is calculated.

A Markov chain is defined in terms of the transition probability  $T(x \rightarrow x')$  for having the point  $x'$  after the point  $x$  in the chain. The transition probabilities depend only on the current state of the system and are independent of time and the history of the walk. The Metropolis algorithm is a series of rules for generating a Markov chain of points distributed with respect to a desired probability density function,  $\rho(x)$  [54]. A Markov chain will converge to the desired density if its transition probabilities

satisfy the following relationship:

$$T(x \rightarrow x') \rho(x) = T(x' \rightarrow x) \rho(x'). \quad (3.3)$$

Eq 3.3 is known as the detailed balance condition. Using it, the probability for accepting a proposed move from  $x$  to  $x'$  is

$$A(x \rightarrow x') = \min \left( 1, \frac{T(x' \rightarrow x) \rho(x')}{T(x \rightarrow x') \rho(x)} \right). \quad (3.4)$$

It should be noted that in equation 3.4, only the ratio  $\frac{\rho(x')}{\rho(x)}$  is calculated, rather than the values  $\rho(x)$  and  $\rho(x')$  separately. As a result, the probability density function  $\rho$  does not have to be normalized.

In the simplest version of the Metropolis algorithm, the transition probabilities are chosen so that  $T(x \rightarrow x') = T(x' \rightarrow x)$ . The acceptance probability can be increased by using *importance sampling* algorithms, which manipulate the transition probabilities to direct the proposed moves into regions of high density [55].

The Metropolis algorithm guarantees the Markov chain will equilibrate to a stationary distribution, which will represent the desired probability density function. This method allows virtually any probability density to be sampled, which makes it an invaluable tool for high dimensional simulations. The Metropolis algorithm is commonly used in simulations of liquids and disordered materials, as well as in molecular dynamics and quantum Monte Carlo.

# Chapter 4

## Quantum Monte Carlo

Quantum Monte Carlo (QMC) is a relatively new class of methods for conducting highly accurate quantum mechanical simulations on atomic and molecular systems. Variational and diffusion Monte Carlo, the most commonly used electronic structure QMC variants, use stochastic methods to optimize wavefunctions and calculate expectation values [56] and can provide energies to within chemical accuracy [57, 58, 59]. Because QMC trial functions do not have to be analytically integrable, there is considerable freedom as to their form. The inclusion of explicit interparticle coordinates, which is impossible with traditional electronic structure methods, allows QMC trial functions to have a very compact form compared to SCF wavefunctions of comparable accuracy [60].

The computational expense of QMC calculations scales with the size of the system as  $O(N^3)$  or better [61, 62, 63, 64]. Mean field methods capable of comparable accuracy, such as coupled cluster, scale much less favorably, as  $O(N^6)$  to  $O(N!)$ . Since QMC is a stochastic method, it lends itself naturally to parallelization across multiple processors. Although QMC is not “perfectly parallel,” as has been claimed [65], the parallel overhead function can be very small, and large numbers of processors can be used with high efficiency. The use of large numbers of processors allows QMC calculations to finish in a reasonable amount of time, despite the slow convergence of Monte Carlo. The combination of favorable scaling and parallelizability of QMC make it possible to conduct highly accurate simulations on systems that are too large for other methods.

Many parallel scientific applications require large memory and fast interprocessor communication to run. Supercomputers that satisfy these needs are very expensive to construct and maintain. The most powerful supercomputers in the world today are owned by the United States Department of Energy, which has committed vast resources to constructing them in order to conduct simulations on nuclear weapons [42, 43, 44].

Quantum Monte Carlo, however, can be formulated to run with very small memory and interprocessor communication requirements [49, 50]. It is reasonable to envision an inexpensive QMC specific supercomputer made of nodes with no hard drives and inexpensive connection hardware. Such a machine would give researchers who do not have access to national lab computers the ability to conduct QMC calculations.

## 4.1 Variational Monte Carlo

Variational quantum Monte Carlo (VMC) uses the Metropolis algorithm to minimize the expectation value of the energy of a trial wavefunction with respect to its adjustable parameters. Because the high dimensional integrals are done using Monte Carlo methods, some of the restrictions on the form of the wavefunction that are necessary when the integrals are evaluated analytically can be relaxed.

The expectation value for the energy of a trial electronic wavefunction,  $|\Psi_T\rangle$ , is

$$\langle E \rangle = \frac{\langle \Psi_T | \hat{H} | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} = \frac{\int d\vec{R} \Psi_T^* (\vec{R}) \hat{H} \Psi_T (\vec{R})}{\int d\vec{R} \Psi_T^* (\vec{R}) \Psi_T (\vec{R})}, \quad (4.1)$$

where  $\hat{H}$  is the Hamiltonian operator for the system and  $\vec{R}$  is a vector containing the  $3N$  spatial coordinates of the  $N$  electrons of the molecule. The *local energy* of an electronic configuration is defined as

$$E_L (\vec{R}) = \frac{\hat{H} \Psi_T (\vec{R})}{\Psi_T (\vec{R})}. \quad (4.2)$$

If  $\Psi_T(\vec{R})$  is an eigenfunction of  $\hat{H}$ , the local energy will be constant with respect to  $\vec{R}$ . Using the local energy, the expectation value can be rewritten:

$$\langle E \rangle = \frac{\int d\vec{R} \Psi_T^*(\vec{R}) \Psi_T(\vec{R}) E_L(\vec{R})}{\int d\vec{R} \Psi_T^*(\vec{R}) \Psi_T(\vec{R})} \quad (4.3)$$

$$= \frac{\int d\vec{R} |\Psi_T(\vec{R})|^2 E_L(\vec{R})}{\int d\vec{R} |\Psi_T(\vec{R})|^2} \quad (4.4)$$

$$= \int d\vec{R} \rho_{VMC}(\vec{R}) E_L(\vec{R}), \quad (4.5)$$

where  $\rho_{VMC}(\vec{R})$  is the probability density for the configuration  $\vec{R}$ :

$$\rho_{VMC}(\vec{R}) = \frac{|\Psi_T(\vec{R})|^2}{\int d\vec{R} |\Psi_T(\vec{R})|^2}. \quad (4.6)$$

VMC employs the Metropolis algorithm to generate a series of electronic configurations,  $\{\vec{R}_i\}$ , distributed with respect to  $\rho_{VMC}(\vec{R})$ . The expectation value of the energy can then be evaluated as

$$\langle E \rangle_{VMC} = \frac{1}{M} \sum_{i=0}^M E_L(\vec{R}_i) \pm O\left(\frac{1}{\sqrt{M}}\right). \quad (4.7)$$

As the wavefunction is sampled, the expectation value of the energy will fluctuate within its statistical uncertainty, which makes comparing different sets of variational parameters difficult. This effect can be mitigated by using correlated sampling, in which expectation values for several sets of parameters are calculated using one set of configurations [66]. Correlated sampling allows the difference between the energies of two sets of parameters to be calculated with much less variance than if the two expectation values are compared after being calculated separately.

Because the local energy for an eigenfunction of  $\hat{H}$  is constant, the variance of its expectation value will be zero. As a trial wavefunction is optimized and it approaches the exact ground state, its local energy will vary less strongly with  $\vec{R}$  and its variance will decrease. As a result, the variance of the energy can be used as a criterion to

optimize the parameters of the wavefunction rather than its expectation value. This method works well for Monte Carlo optimization because the exact minimum value of the variance of the energy is known, while the minimum of the expectation value for the energy is unknown. Some optimization methods minimize a combination of the energy and its variance [67]. Algorithms that sample the derivatives of the energy with respect to the adjustable parameters in the wavefunction can be used to speed convergence and ensure the true minimum is found [68, 69].

## 4.2 VMC Trial Functions

As discussed in section 1.2.1, quantum mechanical wavefunctions have space and spin components. Since the QMC Hamiltonian does not have spin terms, we integrate over spin and consider only the spatial component. The spatial trial functions used in VMC typically have the form

$$\Psi_T = \left( \sum_i c_i \Psi_i^{AS} \right) J, \quad (4.8)$$

where the  $c_i$  are CI expansion coefficients, the  $\Psi_i^{AS}$  are Slater determinant wavefunctions, and  $J$  is a symmetric function of the distances between particles called the Jastrow function. This results in an overall antisymmetric function with explicit interparticle terms. The  $c_i$  and  $\Psi_i^{AS}$  can be obtained through standard electronic structure methods such as Hartree-Fock, GVB, MCSCF, CI, or DFT.

There are many adjustable parameters in  $\Psi_T$ . The  $c_i$ , the orbital coefficients and basis functions in the  $\Psi_i^{AS}$ , and the parameters in the Jastrow function can all be optimized through VMC. Even when correlated sampling is used, optimizing a wavefunction with a large number of adjustable parameters is a challenging task. The form of the trial wavefunction must be chosen with care, so that time is not wasted optimizing parameters that have little effect on the expectation value of the energy.

The two body Jastrow function is written as

$$J = \exp \left[ \sum_{i,j} u_{ij}(r_{ij}) \right], \quad (4.9)$$

where the sum is over all pairs of particles,  $r_{ij}$  is the distance between particles  $i$  and  $j$ , and  $u_{ij}(r_{ij})$  is a function that describes the interaction between particles  $i$  and  $j$ .

The two body Jastrow function makes it straightforward to construct trial wavefunctions that satisfy the quantum mechanical cusp conditions for pairs of particles [9]. Satisfying these cusp conditions removes the singularities in the local energy that occur when particles collide, which lowers the variance of the energy.

If Gaussian orbitals are used in the SCF part of the wavefunction, the cusp condition for particles  $i$  and  $j$  approaching each other leads to the following condition for the function  $u_{ij}$ :

$$\lim_{r_{ij} \rightarrow 0} \frac{\partial u_{ij}(r_{ij})}{\partial r_{ij}} = -\frac{\mu_{ij} q_i q_j}{l + 1}, \quad (4.10)$$

where  $\mu_{ij}$  is the reduced mass of the particles,  $q_i$  and  $q_j$  are their charges, and  $l$  is 1 for same spin electrons and 0 otherwise.

A form for the two-body correlation function commonly used for molecular systems is the Padé-Jastrow function:

$$u_{ij}(r_{ij}) = \frac{c_{ij} r_{ij} + \sum_{k=2}^N a_{ij,k} r_{ij}^k}{1 + \sum_{l=1}^M b_{ij,l} r_{ij}^l}, \quad (4.11)$$

where the  $a_{ij,k}$  and  $b_{ij,l}$  are adjustable parameters. The constant  $c_{ij}$  is set to the value of the cusp condition for the particles  $i$  and  $j$ . To ensure that the limit as  $r \rightarrow \infty$  remains finite,  $M$  and  $N$  are usually chosen so that  $M \geq N$ . Many other forms for correlation functions are in use, including some with scaled variables and some with three- and higher-body terms [70, 71].

In addition to allowing  $\Psi_T$  to have explicit interelectronic coordinates, QMC allows freedom in the form of the orbitals that make up the Slater determinant part of the trial wavefunction. Because they are convenient to evaluate, Gaussian basis functions are used by most SCF programs to construct orbitals.

Gaussian functions have zero derivative at their origin, so molecular orbitals constructed from Gaussian basis functions are unable to satisfy the electron-nucleus cusp condition. Although these cusps can be satisfied by two-body correlation functions, replacing the Gaussian orbitals with exponential functions that satisfy the cusp near the nuclei gives much better results in QMC calculations [72]. When the orbitals are modified in this way, the electron-nucleus cusp values in the two body correlation functions are set to zero.

### 4.3 Diffusion Monte Carlo

Diffusion Monte Carlo (DMC) does not rely on the variational principle to calculate expectation values, but its convergence depends on accurate trial functions.

DMC starts with the time dependent Schrödinger equation:

$$i \frac{\partial}{\partial t} |\Psi(\vec{R}, t)\rangle = \hat{H} |\Psi(\vec{R}, t)\rangle. \quad (4.12)$$

With a change of variables to imaginary time,  $\tau = it$ , equation 4.12 takes the form of a diffusion equation:

$$-\frac{\partial}{\partial \tau} |\Psi(\vec{R}, \tau)\rangle = \hat{H} |\Psi(\vec{R}, \tau)\rangle. \quad (4.13)$$

The formal solution to equation 4.13 can be written:

$$|\Psi(\vec{R}, \tau)\rangle = e^{-\tau \hat{H}} |\Psi(\vec{R}, 0)\rangle. \quad (4.14)$$

At some time  $\tau_1$ , the state  $|\Psi(\vec{R}, \tau_1)\rangle$  is expanded in eigenstates of the Hamiltonian:

$$|\Psi(\vec{R}, \tau_1)\rangle = \sum_i c_i |\Phi_i\rangle, \quad (4.15)$$

where

$$\hat{H} |\Phi_i\rangle = E_i |\Phi_i\rangle, \quad (4.16)$$

and

$$c_i = \langle \Phi_i | \Psi(\vec{R}, \tau_1) \rangle. \quad (4.17)$$

The expansion in equation 4.15 is substituted into equation 4.14:

$$|\Psi(\vec{R}, \tau_1 + d\tau)\rangle = \sum_i c_i e^{-E_i d\tau} |\Phi_i\rangle. \quad (4.18)$$

As equation 4.18 is propagated with  $\tau$ , contributions to  $\Psi(\vec{R}, \tau)$  with  $i > 0$  will die out exponentially, leaving  $\Phi_0$ , the ground state.

Propagating equation 4.18 with Monte Carlo methods is inefficient because the potential part of the Hamiltonian varies widely throughout configuration space and diverges when charged particles approach each other. Efficient DMC calculations use importance sampling, in which  $\Psi_T(\vec{R})$ , a trial function that approximates the ground state, is used as a guide function. A mixed distribution is defined:

$$\rho_{DMC}(\vec{R}) = \frac{\Phi_0(\vec{R}) \Psi_T(\vec{R})}{\int d\vec{R} \Phi_0(\vec{R}) \Psi_T(\vec{R})}. \quad (4.19)$$

The mixed expectation value for an operator,  $\hat{A}$ , has the form

$$\langle A \rangle_{DMC} = \frac{\langle \Phi_0 | \hat{A} | \Psi_T \rangle}{\langle \Phi_0 | \Psi_T \rangle}. \quad (4.20)$$

For operators that commute with the Hamiltonian, the DMC expectation value equals the expectation value of the true ground state:

$$\langle A \rangle_{DMC} = \frac{\langle \Phi_0 | \hat{A} | \Psi_T \rangle}{\langle \Phi_0 | \Psi_T \rangle} = \frac{\langle \Phi_0 | \hat{A} | \Phi_0 \rangle}{\langle \Phi_0 | \Phi_0 \rangle}. \quad (4.21)$$

The DMC expectation value for the energy can be rewritten in a manner similar to the VMC expectation value:

$$\langle E \rangle_{DMC} = \frac{\langle \Phi_0 | \hat{H} | \Psi_T \rangle}{\langle \Phi_0 | \Psi_T \rangle} \quad (4.22)$$

$$= \frac{\int d\vec{R} \Phi_0(\vec{R}) \hat{H} \Psi_T(\vec{R})}{\int d\vec{R} \Phi_0(\vec{R}) \Psi_T(\vec{R})} \quad (4.23)$$

$$= \frac{\int d\vec{R} \Phi_0(\vec{R}) \Psi_T(\vec{R}) \frac{\hat{H} \Psi_T(\vec{R})}{\Psi_T(\vec{R})}}{\int d\vec{R} \Phi_0(\vec{R}) \Psi_T(\vec{R})} \quad (4.24)$$

$$= \int d\vec{R} \rho_{DMC}(\vec{R}) E_L(\vec{R}). \quad (4.25)$$

A series of electronic configurations is generated with respect to  $\rho_{DMC}(\vec{R})$ , which allows the expectation value to be evaluated. Generating electronic configurations with respect to  $\rho_{DMC}$  will be discussed in the next section.

Interpreting  $\rho_{DMC}(\vec{R})$  as a probability density is only possible if it is nonnegative for all  $\vec{R}$ . For bosons, this property is easily satisfied because the ground state wavefunction has one sign everywhere in configuration space. If the trial wavefunction has the same sign,  $\rho_{DMC}(\vec{R})$  will be nonnegative for all  $\vec{R}$ . Ground state wavefunctions for fermions, however, have positive and negative regions separated by nodes. If the nodes of  $\Psi_T(\vec{R})$  and  $\Phi_0(\vec{R})$  are identical, the two functions will have the same sign in every nodal region, and  $\rho_{DMC}(\vec{R})$  will be nonnegative for all  $\vec{R}$ .

If the nodal structures of  $\Psi_T(\vec{R})$  and  $\Phi_0(\vec{R})$  are different,  $\rho_{DMC}(\vec{R})$  will have positive and negative regions. This is known as the *fermion problem* in DMC. In VMC, the magnitude squared of the trial function is sampled, so there is no analogous nodal problem.

The nodal surface of an electronic wavefunction is a  $(3N - 1)$ -dimensional hypersurface where the wavefunction vanishes. The spatial antisymmetry of the wavefunction defines a set of  $(3N - 3)$ -dimensional hyperpoints embedded in the nodal surface. Although these points are known, no general techniques exist for constructing a trial wavefunction with the same nodal structure as the true ground state.

The simplest and most widely used solution to this problem is the *fixed-node* approximation, in which the nodes of the true ground state are assumed to be the same as the nodes of the trial wavefunction. When this approximation is used,  $\Phi_0(\vec{R})$  becomes the ground state wavefunction consistent with the boundary condition that

it vanish at the nodes of  $\Psi_T(\vec{R})$ . The fixed-node approximation is enforced in a DMC calculation by rejecting any proposed move that crosses a node and causes  $\Psi_T(\vec{R})$  to change sign. The resulting energy lies above the exact energy and is variational in the nodal structure of the trial function [73, 74].

Other solutions to the nodal problem that do not rely on the fixed node approximation have been developed. For example, the *transient estimator* method propagates two bosonlike walker ensembles, representing the positive and negative parts of  $\Phi_0(\vec{R})$ . This method is not stable with respect to  $\tau$ , the imaginary time variable in which the ensembles are propagated, because both parts of the simulation converge to the nodeless boson ground state. Expectation values can only be calculated during the intermediate regime before this occurs, which limits the statistical accuracy that can be attained.

For most small molecules, the nodes of trial wavefunctions obtained by standard SCF methods are of good enough quality for fixed node DMC calculations to yield results within chemical accuracy [57, 59]. In some cases, such as the beryllium atom, multi-configuration wavefunctions are needed to obtain nodes of sufficient quality.

## 4.4 Generating Configurations in DMC

Electronic configurations are generated with respect to  $\rho_{DMC}$  using the distribution

$$f(\vec{R}, \tau) = \Phi(\vec{R}, \tau) \Psi_T(\vec{R}), \quad (4.26)$$

where  $|\Phi(\vec{R}, \tau)\rangle$  is a solution to the time-dependent Schrödinger equation, equation 4.12.

The distribution  $f(\vec{R}, \tau)$  is a solution to a Fokker-Planck equation:

$$-\frac{\partial}{\partial \tau} f(\vec{R}, \tau) = (\hat{L} - E_T) f(\vec{R}, \tau), \quad (4.27)$$

where

$$\hat{L} = -\frac{1}{2}\nabla^2 + \nabla \cdot V(\vec{R}) + E_L(\vec{R}). \quad (4.28)$$

$V(\vec{R})$  is the *local velocity* of the trial function at  $\vec{R}$ :

$$V(\vec{R}) = \frac{\nabla \Psi_T(\vec{R})}{\Psi_T(\vec{R})}, \quad (4.29)$$

and  $E_L(\vec{R})$  is its local energy:

$$E_L(\vec{R}) = \frac{\hat{H}\Psi_T(\vec{R})}{\Psi_T(\vec{R})}. \quad (4.30)$$

In the case  $\Psi_T(\vec{R}) = 1$ , equations 4.27 and 4.28 reduce to equation 4.13, the Schrödinger equation in imaginary time.

The operator  $\hat{L}$  defines an eigenvalue equation:

$$\hat{L}|K_i(\vec{R})\rangle = \kappa_i|K_i(\vec{R})\rangle. \quad (4.31)$$

The  $K_i(\vec{R}) = \Phi_i(\vec{R})\Psi_T(\vec{R})$  and the  $\kappa_i = E_i$ , where the  $\Phi_i(\vec{R})$  and  $E_i$  are the eigenvectors and eigenvalues of the Hamiltonian.

Equations 4.27 and 4.28 describe a diffusion process in a potential. As with equation 4.13, the formal solution to equation 4.27 can be written:

$$f(\vec{R}, \tau) = e^{-\tau(\hat{L}-E_T)}f(\vec{R}, 0). \quad (4.32)$$

This solution can be expanded in the eigenvectors of  $\hat{L}$ :

$$f(\vec{R}, \tau) = \sum_i c_i e^{-\tau(\kappa_i - E_T)} K_i(\vec{R}) \quad (4.33)$$

$$= \sum_i c_i e^{-\tau(E_i - E_T)} \Phi_i(\vec{R}) \Psi_T(\vec{R}), \quad (4.34)$$

where

$$c_i = \langle K_i(\vec{R}) | f(\vec{R}, 0) \rangle = \langle \Phi_i(\vec{R}) | \frac{f(\vec{R}, 0)}{\Psi_T(\vec{R})} \rangle. \quad (4.35)$$

It is easy to see that if  $E_T = E_0$ , contributions to  $f(\vec{R}, \tau)$  from the  $\Phi_i(\vec{R})$  with  $i > 0$  will die off exponentially as  $\tau$  increases, leaving the desired density,  $\rho_{DMC}(\vec{R}) = \Phi_0(\vec{R}) \Psi_T(\vec{R})$ .

In order to propagate Eq. 4.32 with  $\tau$  and obtain  $\rho_{DMC}(\vec{R})$ , it is rewritten in integral form:

$$f(\vec{Y}, \tau + d\tau) = e^{d\tau \cdot E_T(\tau + d\tau)} \int d\vec{R} G(\vec{Y}, \vec{R}, d\tau) f(\vec{R}, \tau), \quad (4.36)$$

where  $G(\vec{Y}, \vec{R}, d\tau)$  is the Green's function corresponding to the operator  $\hat{L}$ . Unfortunately, this Green's function, like the Green's functions for most complicated physical processes, cannot be written for arbitrary  $d\tau$ .

The three terms of equation 4.28 describe diffusion, drift, and branching processes. Green's functions can be written for each process individually, and an approximate Green's function can be written as their product:

$$\begin{aligned} G(\vec{Y}, \vec{R}, \tau) &\approx \frac{1}{(2\pi\tau)^{3N/2}} \delta[\vec{Z} - \vec{R} - V(\vec{R}) d\tau] \\ &\times \int d\vec{Z} \exp\left[\frac{(\vec{Y} - \vec{Z})^2}{2d\tau}\right] \\ &\times e^{-\frac{1}{2}[E_L(\vec{Y}) + E_L(\vec{R})]d\tau} + O(d\tau^2). \end{aligned} \quad (4.37)$$

The factorization of the Green's function neglects the fact that the terms of  $\hat{L}$  do not commute, so equation 4.37 is exact only in the limit  $d\tau \rightarrow 0$ . Equation 4.34, however, is only exact in the limit  $\tau \rightarrow \infty$ . Any choice of time step is a tradeoff between these two considerations. In practice, runs with several values of  $d\tau$  must be done, and the results are extrapolated to  $d\tau = 0$ .

During a DMC calculation,  $f(\vec{R}, \tau)$  is represented by an ensemble of walkers,

each consisting of an electronic configuration and a statistical weight:

$$f(\vec{R}, \tau) = \sum_n w_{n,\tau} \delta(\vec{R} - \vec{R}_{n,\tau}). \quad (4.38)$$

Each iteration in a DMC calculation consists of four stages: drift, diffusion, weighting, and branching. In the drift step, the electrons are moved according to the time step and the local velocity. In the diffusion step, the electrons are moved to new positions with transition probabilities given by the kinetic part of the Green's function. The weighting and branching step takes into account the potential part of the Green's function.

After a walker is moved from configuration  $\vec{R}$  to  $\vec{Y}$ , its weight is calculated based on the local energy at  $\vec{R}$  and  $\vec{Y}$ . In the branching step, walkers with high weight give birth to new walkers, while low weight walkers are deleted.

The trial energy,  $E_T$ , serves as a normalization factor and is adjusted after each step based on the sum of the weights of the walkers in order to keep the population stable. The average value of  $E_T$  after many steps will converge to the ground-state energy.

Several DMC algorithms, each with slightly different schemes for factoring the Green's function, proposing configurations, calculating the weights, and branching the walkers have been published [66, 75]. The DMC calculations presented later in this work use a combination of Umrigar's DMC algorithm [55] and the reweighting method of Assaraf et al. [76].

## Chapter 5

# An Optimized Initialization Algorithm to Ensure Accuracy in Quantum Monte Carlo Calculations

# Abstract

Quantum Monte Carlo (QMC) calculations require the generation of random electronic configurations with respect to a desired probability density, usually the square of the magnitude of the wavefunction. In most cases, the Metropolis algorithm is used to generate a sequence of configurations in a Markov chain. This method has an inherent equilibration phase, during which the configurations are not representative of the desired density and must be discarded. If statistics are gathered before the walkers have equilibrated, contamination by nonequilibrated configurations can greatly reduce the accuracy of the results. Because separate Markov chains must be equilibrated for the walkers on each processor, the use of a long equilibration phase has a profoundly detrimental effect on the efficiency of large parallel calculations.

The stratified atomic walker initialization (STRAW) shortens the equilibration phase of QMC calculations by generating statistically independent electronic configurations in regions of high probability density. This ensures the accuracy of calculations by avoiding contamination by nonequilibrated configurations. Shortening the length of the equilibration phase also results in significant improvements in the efficiency of parallel calculations, which reduces the total computational run time. For example, using STRAW rather than a standard initialization method in 512 processor calculations reduces the amount of time needed to calculate the energy expectation value of a trial function for a molecule of the energetic material RDX to within 0.01 au by 33%.

## 5.1 Introduction

Quantum Monte Carlo methods for simulating the electronic structure of molecules [77, 78] can in principle provide energies to within chemical accuracy ( $\sim 2$  kcal/mol) [57, 58, 59]. The computational expense of QMC scales with system size as  $O(N^3)$  or better [61, 62, 63, 64], albeit with a large prefactor. This is much more favorable than other electronic structure methods capable of comparable accuracy, such as coupled cluster, which tend to scale very poorly with the size of the system, generally  $O(N^6$  to  $N!)$  [79]. Moreover, the stochastic nature of QMC makes it relatively easy to parallelize over a large number of processors, which can allow calculations to finish in a reasonable amount of time despite the slow convergence of Monte Carlo.

As supercomputing resources improve and become more accessible to researchers [42, 80], QMC will become a powerful tool for conducting accurate simulations on chemically interesting systems. Recent efforts have focused making these calculations more straightforward and efficient on heterogeneous and homogeneous computers. To this end, a finite all-electron QMC program, QMcBeaver, has been written and used to develop and demonstrate several new algorithms [49, 50, 81].

Before statistics gathering begins in a QMC calculation, the walkers must be allowed to equilibrate so that their configurations are proportional to the desired density. It is impossible to calculate accurate expectation values if nonequilibrated configurations contaminate the statistics. In order to ensure their statistical independence, the walkers must equilibrate separately. This makes the equilibration phase a serial step of the calculation and a major limiting factor in the efficiency of parallel calculations. These considerations make it imperative that the equilibration process be fast and reliable. For example, we show that for the energetic material RDX, approximately 30,000 iterations are necessary for equilibration when the initial configurations are generated by a standard method.

We present here a simple method for choosing initial electronic configurations designed to reduce the length of the equilibration phase of calculations. The Stratified Atomic Walker initialization (STRAW) for quantum Monte Carlo calculations uses a

shell model to distribute the electrons. When STRAW is used in RDX calculations, 100 iterations are sufficient for equilibration.

Avoiding contamination by nonequilibrated configurations in quantum Monte Carlo calculations ensures their accuracy, and reducing the cost of equilibration makes calculations with large numbers of processors much more efficient. Improving the parallel efficiency of these calculations makes better use of computer resources and will broaden the range of systems for which quantum Monte Carlo calculations are practical.

## 5.2 The Metropolis Algorithm and the Initialization Catastrophe

Quantum Monte Carlo calculations center around the random generation of electronic configurations with respect to quantum mechanical probability densities. In this work, we focus on variational Monte Carlo (VMC), in which the trial wavefunction is sampled in order to optimize its adjustable parameters [67, 68, 69].

VMC trial functions usually have the form  $\Psi_{VMC} = \Psi_{SCF}J$ , where  $\Psi_{SCF}$  is one or a sum of Slater determinant wavefunctions obtained by a standard electronic structure method such as Hartree-Fock (HF), density functional theory (DFT), or multiconfiguration self-consistent field (MCSF). The Jastrow factor,  $J$  [70, 71, 82], is a symmetric function of the interparticle coordinates meant to account for quantum mechanical cusp conditions [9] and short range correlations.

The expectation value for the energy of this trial function is

$$\langle E \rangle = \frac{\langle \Psi_{VMC} | \hat{H} | \Psi_{VMC} \rangle}{\langle \Psi_{VMC} | \Psi_{VMC} \rangle} = \frac{\int_{-\infty}^{\infty} d\vec{x} \Psi_{VMC}^* (\vec{x}) \hat{H} \Psi_{VMC} (\vec{x})}{\int_{-\infty}^{\infty} d\vec{x} \Psi_{VMC}^* (\vec{x}) \Psi_{VMC} (\vec{x})}, \quad (5.1)$$

where  $\vec{x}$  is a  $3N$ -dimensional vector of the positions of the  $N$  electrons in the molecule. Because the Jastrow factor includes explicit interparticle coordinates, equation 5.1 cannot be separated into independent electron problems and solved using the standard SCF procedure. Instead, the expectation value is evaluated stochastically [75].

The *local energy* of a configuration,  $\vec{x}$ , is defined as  $E_L(\vec{x}) = \frac{\hat{H}\Psi_{VMC}(\vec{x})}{\Psi_{VMC}(\vec{x})}$ . Using this quantity, the expectation value of the energy can be rewritten:

$$\langle E \rangle = \frac{\int_{-\infty}^{\infty} d\vec{x} |\Psi_{VMC}(\vec{x})|^2 E_L(\vec{x})}{\int_{-\infty}^{\infty} d\vec{x} |\Psi_{VMC}(\vec{x})|^2} = \int_{-\infty}^{\infty} d\vec{x} \rho_{VMC}(\vec{x}) E_L(\vec{x}), \quad (5.2)$$

with

$$\rho_{VMC}(\vec{x}) = \frac{|\Psi_{VMC}(\vec{x})|^2}{\int_{-\infty}^{\infty} d\vec{x} |\Psi_{VMC}(\vec{x})|^2}. \quad (5.3)$$

The expectation value now has the form of a weighted average. A series of  $M$  electronic configurations,  $\{\vec{x}_i\}$ , is generated with respect to  $\rho_{VMC}$  and used to evaluate the expectation value of the energy:

$$\langle E \rangle = \frac{1}{M} \sum_{i=1}^M E_L(\vec{x}_i) \pm O\left(\frac{1}{\sqrt{M}}\right). \quad (5.4)$$

The VMC probability density,  $\rho_{VMC}$ , is an extremely complicated,  $3N$ -dimensional function. An effective way to generate electronic configurations with respect to this type of function is to use a Markov chain, which is defined in terms of the transition probability  $T(\vec{x} \rightarrow \vec{y})$  of having the configuration  $\vec{y}$  after  $\vec{x}$  in the chain. The Metropolis algorithm [54] is a method for generating a Markov chain of points distributed with respect to a desired probability density. It states that a Markov chain will converge to a desired density,  $f(\vec{x})$ , if its transition probabilities satisfy the following relationship:

$$T(\vec{x} \rightarrow \vec{y}) f(\vec{x}) = T(\vec{y} \rightarrow \vec{x}) f(\vec{y}). \quad (5.5)$$

Equation 5.5 is known as the detailed balance condition. The most commonly used formula for calculating the probability of accepting a proposed move from  $\vec{x}$  to  $\vec{y}$  that satisfies detailed balance is

$$A(\vec{x} \rightarrow \vec{y}) = \min \left[ 1, \frac{w(\vec{y} \rightarrow \vec{x}) f(\vec{y})}{w(\vec{x} \rightarrow \vec{y}) f(\vec{x})} \right], \quad (5.6)$$

where  $w(\vec{x} \rightarrow \vec{y})$  is the probability for proposing a move from  $\vec{x}$  to  $\vec{y}$ .

In this work, we use the accelerated Metropolis algorithm developed by Umrigar

and coworkers [55, 83] to propose configurations and calculate  $w(\vec{x} \rightarrow \vec{y})$ . This algorithm allows different length scales for the motions of core and valence electrons, which increases the size of the time step that can be used in a calculation while maintaining a high acceptance rate.

The Metropolis algorithm guarantees that the Markov chain will equilibrate to the desired distribution, but does not provide any criteria to predict the number of iterations necessary for equilibration or to determine when it has occurred. It is vital to avoid contamination by nonequilibrated points in calculations, because it is impossible to calculate accurate expectation values using configurations that do not represent the desired density.

The equilibration time will depend strongly on the choice of the initial configuration,  $\vec{x}_0$ . If  $\vec{x}_0$  is in a region of low probability density, repeated iterations using equation 5.6 will guide the chain into regions of higher probability density. The chain is equilibrated when it reaches a region whose probability density is high enough that sampling it is consistent with the desired probability density and the total number of iterations. Clearly, the number of iterations required for equilibration can be minimized by making an intelligent choice for  $\vec{x}_0$ .

In Monte Carlo simulations, a *walker* is an entity that defines the state of the system at a particular instant. In QMC, a walker consists of a  $3N$ -dimensional electronic configuration. An ensemble of walkers is used to carry out the integration, with each one tracing out an independent Markov chain in configuration space. In a parallel calculation, an ensemble of walkers is equilibrated and propagated on each processor, and the results are gathered to obtain the global results.

In the QMcBeaver program, each processor must have at least one walker, and the number of walkers per processor is a user defined constant. Since the number of walkers increases linearly with the number of processors, the computational effort devoted to equilibration increases as well. The impact of the equilibration phase on the efficiency of a parallel calculation was predicted and demonstrated by Feldmann and Kent [50], and we follow their derivation.

Since separate Markov chains must be equilibrated on each processor, the total

equilibration time scales as  $O(N_{Processors})$ . The time devoted to generating statistics,  $T^{Propagate}$ , scales as  $O(1)$  because the number of independent samples needed to achieve a certain level of convergence does not change with the number of processors. From this, the efficiency, or fraction of the total calculation time devoted to useful work,  $\epsilon$ , is

$$\epsilon = \frac{T^{Propagate}}{T^{Initialize} + T^{Equilibrate} + T^{Synchronize} + T^{Communicate} + T^{Propagate}} \quad (5.7)$$

$$\approx \frac{O(1)}{O(N_{Processors}) + O(1)}. \quad (5.8)$$

Since the synchronization and communication costs for QMC calculations are extremely small, the main threat to efficiency in parallel calculations will be the equilibration time. In order to use a large number of processors efficiently, an algorithm for quickly generating equilibrated, statistically independent electronic configurations for the walkers is necessary. The next section examines how initial walkers are generated in several QMC programs and considers possibilities for improvement.

### 5.3 Walker Initialization

The walker initialization algorithm originally implemented in the QMcBeaver program works as follows: the electrons of the molecule are assigned to the nuclei according to the density implied by the SCF wavefunction. Each nucleus and its electrons are treated as an atom, and the electrons are distributed with respect to a three-dimensional Gaussian centered on the nucleus whose variance is related to the covalent radius of that atom. The configuration is discarded and a new one is generated if substituting the locations of the electrons into the Slater determinant part of the wavefunction results in a singularity [81]. This happens if there is any linear dependence among the columns of the determinant, which can happen if two parallel spin electrons are too close to each other. We will refer to this method as the Gaussian atomic walker initialization (GAWI).

The initialization algorithm of Casino, a QMC program developed at Cambridge, assigns the electrons to atoms and then places the electrons randomly within spheres centered on the atoms [84].

QMAGIC, a QMC program developed at UC Berkeley and the Lawrence Berkeley National Laboratory, uses an initialization method similar to GAWI [85]. The electrons are distributed with respect to three-dimensional Gaussians centered on the nuclei, and then the configuration is checked to ensure no two particles are closer than a tolerance distance to each other. Zori, a new QMC program developed in the same research group, distributes electrons randomly in spheres of the atomic covalent radius and checks to make sure no electron-electron distance is smaller than a threshold. A configuration is discarded if its local energy is not within a given range of an estimate of the energy of the system [86]. This test is probably effective in eliminating some unfavorable initial configurations, but requires additional user specified parameters and could cause the walker initialization to scale badly if a large fraction of the configurations generated were discarded.

These initialization methods give satisfactory performance in calculations on small molecules using moderately large computers. In these calculations, the equilibration phase is a small part of the total computational expense and does not have a severely detrimental effect on the efficiency. As the size of the molecules and the number of processors increase, however, the fraction of the total time spent equilibrating can become significant. By improving the way initial configurations are chosen, the length of the equilibration phase can be reduced, which will improve the parallel scaling and efficiency of calculations using large numbers of processors.

The walker initialization algorithms described above suffer from several deficiencies. Most importantly, because all the electrons of an atom are distributed with respect to the same probability distribution, the electrons tend not to avoid each other in the initial configurations. For opposite spin electrons, this is unfavorable because of their coulomb repulsion. For parallel spin electrons, however, it is even worse. The antisymmetry of the wavefunction dictated by the Pauli principle forces the wavefunction to go to zero as two parallel spin electrons approach each other. In

addition, these methods ignore the structure of the energy levels, in which there will be certain numbers of electrons mostly within annular shells.

Because the initialization methods of this section share these deficiencies, we assume that their performance will be similar, and will use GAWI to represent them in comparisons.

### 5.3.1 STRAW

The Stratified Atomic Walker initialization (STRAW) is a method for generating initial electronic configurations that addresses the problems described above. In STRAW, the electrons are assigned to the nuclei as in the other methods. Care is taken to ensure that, for an overall neutral molecule, each atom is neutral. The atoms are treated separately, and the electrons are partitioned into energy levels, with one alpha spin and one beta spin electron in the first energy level, up to four alpha spins and four beta spins in the second energy level, and so on. The electrons in each energy level are distributed using the transformation method with respect to probability densities in spherical coordinates:  $r, \theta, \phi$ . The transformation method directly converts uniform random numbers on the interval  $(0, 1)$  to random numbers distributed with respect to a desired probability density using the inverse of its cumulative distribution function [53].

To obtain the radial densities for the energy levels, Hartree-Fock/6-311G\*\* calculations were carried out for each atom in the first three rows of the periodic table using Jaguar [87]. The occupied atomic orbitals were localized by the Boys procedure, which creates orbitals with maximum insensitivity to changes in distant nuclear charges [88]. For molecules, the resulting orbitals are localized around the chemical bonds and in the atomic lone pair regions. In our case, the Boys procedure hybridizes the valence orbitals of the atom.

A representative orbital for each energy level was chosen and expressed as a sum

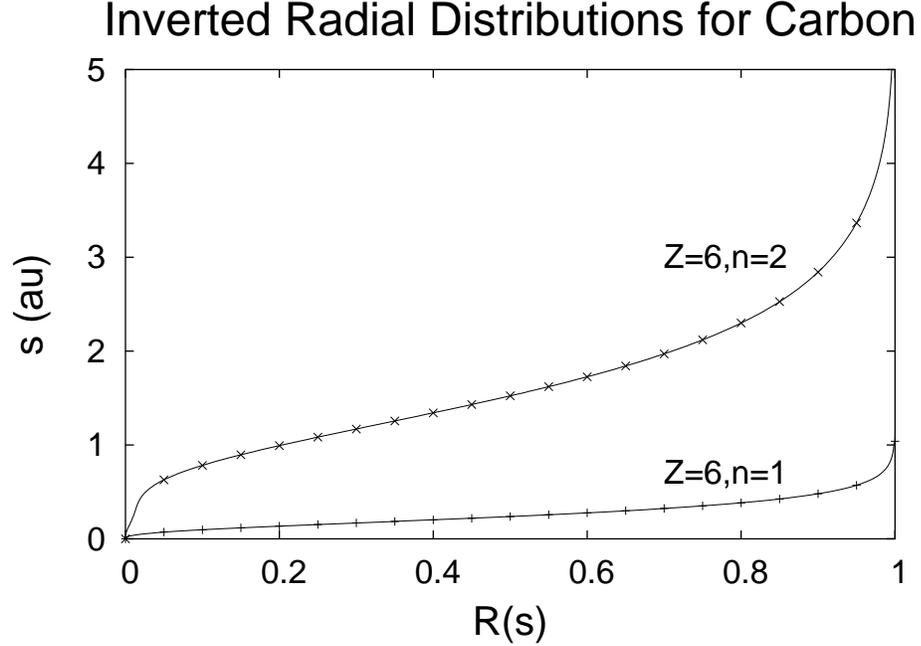


Figure 5.1: The inverted radial distributions for the first and second energy levels of carbon. To generate the radial coordinate for an electron in one of these energy levels, we generate a uniform random number in the range (0,1) and then evaluate the appropriate inverted distribution.

of primitive Gaussians:

$$\psi = \sum_i d_i x^{a_i} y^{b_i} z^{c_i} \exp(-\alpha_i r^2). \quad (5.9)$$

The  $d_i$  are the expansion coefficients and the exponents  $a_i$ ,  $b_i$ , and  $c_i$  determine the symmetry of the primitive Gaussians. The square of the orbital is its probability density:

$$\rho = |\psi|^2 = \sum_{i,j} d_i d_j x^{a_i+a_j} y^{b_i+b_j} z^{c_i+c_j} \exp[-(\alpha_i + \alpha_j) r^2]. \quad (5.10)$$

Converting the probability density into spherical coordinates and integrating over the angles yields the radial marginal probability distribution of the orbital:

$$R(s) = \sum_{i,j} d_i d_j \int_0^{2\pi} d\phi \cos \phi^{a_i+a_j} \sin \phi^{b_i+b_j} \int_0^\pi d\theta \sin \theta^{a_i+a_j+b_i+b_j+1} \cos \theta^{c_i+c_j} \int_0^s dr r^{2+a_i+a_j+b_i+b_j+c_i+c_j} \exp[-(\alpha_i + \alpha_j) r^2]. \quad (5.11)$$

The integrals over the angles were done analytically, and the radial integrals were evaluated numerically by a change of variables from the incomplete gamma function. Because the orbitals are normalized and their probability densities are always positive,  $R(s)$  increases monotonically with  $r$  from zero to one.

Radial probability distributions for each energy level of each atom were tabulated and inverted by interchanging the coordinates. For example, the inverted radial marginal distributions for the first and second energy levels of carbon are shown in figure 5.1. In order to generate the radial distances for the electrons in an energy level, QMcBeaver fits a cubic spline to the appropriate tabulated inverted distribution. A uniform number in the range  $(0, 1)$  is generated for each electron and converted to a radial coordinate by evaluating the spline.

The transformation method is also used to generate the angular coordinates for the electrons. Probability densities in  $\theta$  and  $\phi$  for  $s$ ,  $sp$ ,  $sp^2$ , and  $sp^3$  hybrid orbitals were found in terms of the real spherical harmonics [89] and integrated analytically. The results were tabulated and inverted. As with the radial distributions, splines are fit to the tabulated inverted distributions and used to generate the angular coordinates of the electrons in the energy level.

The probability densities in  $\theta$  and  $\phi$  are chosen for each electron so that they avoid each other, with parallel spin electrons having higher priority. For example, if there are three alpha and two beta electrons in an energy level, the three alpha electrons are distributed with respect to the angular probability distributions of the three  $sp^2$  orbitals in the  $xz$  plane, while the two beta electrons are distributed with respect to those of the  $sp$  orbitals along the  $y$  axis.

Once the radial and angular coordinates for the electrons of an energy level have been assigned, they are converted to Cartesian coordinates. The entire energy level is then given a random rotation about a random axis. This rotation is easily computed using quaternions and prevents the distribution from becoming skewed along any axis or plane.

STRAW has been implemented in QMcBeaver, an open source program [81]. Researchers interested in further details of the algorithm are encouraged to download

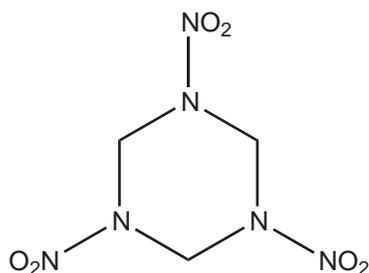


Figure 5.2: The RDX molecule.

and examine the source code.

### 5.3.2 Equilibration Behavior

The computational experiments described in this section comparing the performance of the initialization methods were conducted using QMcBeaver [81].

The VMC trial functions used in this section have the form  $\Psi_{VMC} = \Psi_{SCF}J$ , where  $\Psi_{SCF}$  is a HF/6-311G\*\*++ wavefunction calculated using Jaguar [87] and  $J$  is a Pade-Jastrow correlation function with terms for each pair of particles in the molecule:

$$J = \exp\left(\sum_i \sum_{j<i} u_{ij}\right), \quad (5.12)$$

$$u_{ij} = \frac{c_{ij}r_{ij}}{1 + b_{ij}r_{ij}}. \quad (5.13)$$

In order to satisfy the cusp condition [9] for an electron approaching a nucleus, we set  $c = -Z$  for the electron-nuclear  $u$  functions, where  $Z$  is the charge of the nucleus. Similarly, we set  $c = \frac{1}{2}$  for opposite spin electron pairs and  $c = \frac{1}{4}$  for same spin electron pairs.

For opposite spin electron pairs, we use  $b = 3.0$ , and for same spin electron pairs and all nuclear-electron terms, we use  $b = 100.0$ . Our experience is that these values work reasonably well for ground states of molecules composed of atoms from the first three rows of the periodic table.

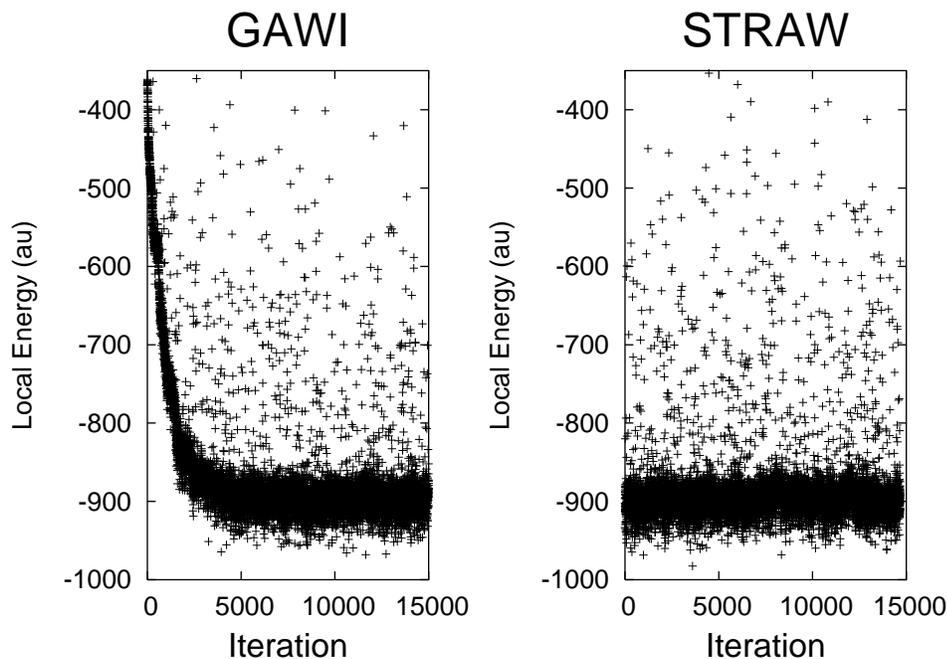


Figure 5.3: Local energies of two RDX walkers. The walker initialized with GAWI starts off with very high local energy and approaches equilibration after several thousand steps, while the walker initialized with STRAW reaches a constant distribution very quickly.

This is a very simple trial wavefunction, and its parameters are not optimized. In order to calculate accurate electronic properties for these molecules, the trial function could be improved by modifying the orbitals to satisfy the electron-nucleus cusp condition [72], using a better Jastrow form [71, 82], and optimizing its parameters [67, 68, 69]. In this work, however, we are focusing on equilibration and our ability to sample a wavefunction, so the simple trial function is sufficient.

In the calculations of this section, we use a time step of 0.001 au for both the equilibration and propagation phases, which results in propagation phase acceptance probabilities of 85% for  $\text{SiCl}_4$  and 93% for RDX. Methods such as using a larger time step during the equilibration phase can be used to accelerate equilibration. In order to simplify comparisons between initialization methods, however, we use a constant time step in all of our calculations.

The effort that has gone into the more complicated initialization scheme pays off

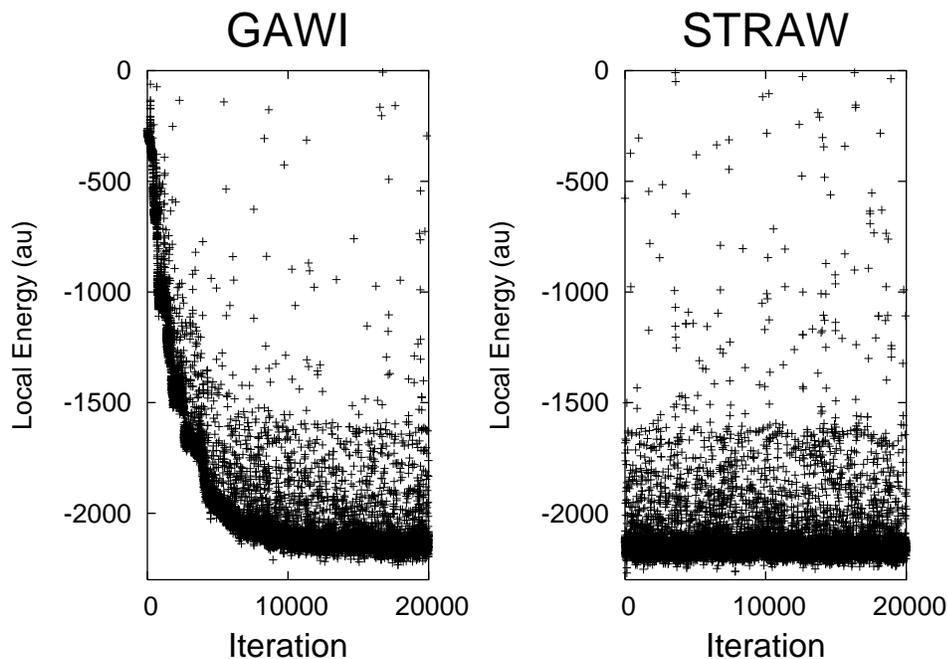


Figure 5.4: Local energies of SiCl<sub>4</sub> walkers. The local energy of the walker initialized with GAWI starts in a high energy region and approaches a steady state after several thousand steps, while the walker initialized with STRAW is equilibrated very quickly.

handsomely. Figures 5.3 and 5.4 show the behavior of the local energy of VMC walkers initialized with GAWI and STRAW. In each case, we find that walkers initialized with GAWI require several thousand steps to reach an equilibrium distribution, while walkers initialized with STRAW require very few.

Figure 5.3 shows the behavior of the local energy of two walkers during calculations on hexhydro-1,3,5-trinitro-1,3,5-triazine, or RDX (figure 5.2) [90], an energetic material. The local energy of the walker initialized with GAWI approaches a steady state after several thousand steps. This figure clearly shows the importance of avoiding contamination by the high energy nonequilibrated configurations in the beginning of the calculation. In contrast, the distribution of local energies for the walker initialized with STRAW is constant throughout the run. The initial configuration is in a region of high probability density and low local energy, and the long equilibration phase we see in the case of the GAWI walker is eliminated.

In order to test the effectiveness of STRAW on a molecule with atoms from the

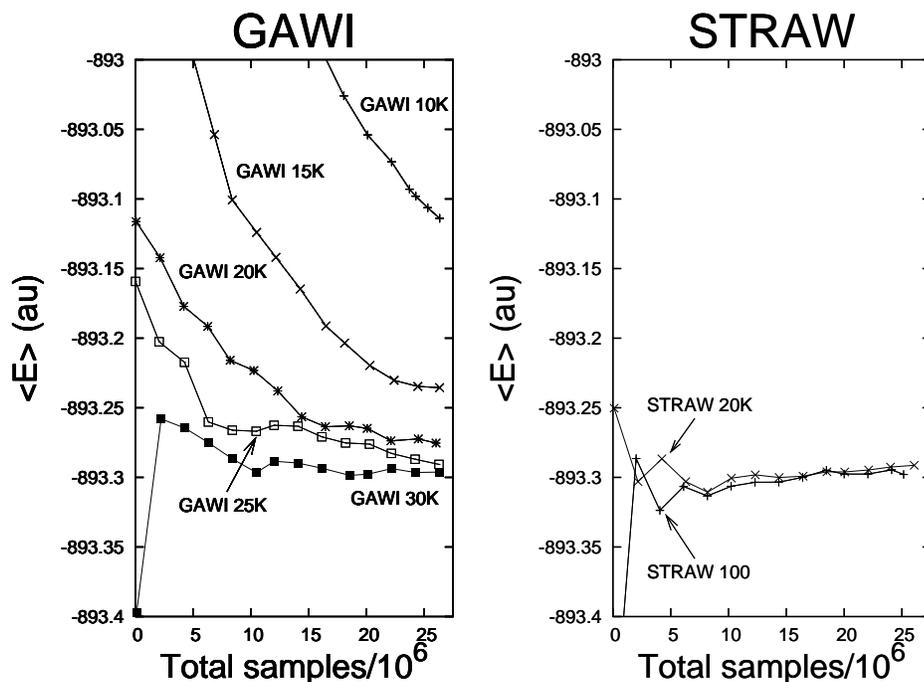


Figure 5.5: VMC calculations on RDX were carried out using 512 processors and 5 walkers per processor. Ensembles initialized with GAWI require 30,000 equilibration steps before contamination by high energy samples is eliminated. The ensemble initialized with STRAW is equilibrated after 100 steps.

third row of the periodic table, a series of calculations was carried out with the  $\text{SiCl}_4$  molecule. Figure 5.4 shows the behavior of the local energy of walkers initialized with GAWI and STRAW. Once again, we see that the local energy of the walker initialized with GAWI approaches equilibration after several thousand steps, while the local energy of the walker initialized with STRAW reaches an equilibrium distribution very quickly.

Figures 5.3 and 5.4 examine only one walker for each initialization method. The results are encouraging, but a visual examination of the local energy is hardly a quantitative measure of equilibration. In addition, realistic QMC calculations on the molecules of this section will use ensembles of thousands to hundreds of thousands of walkers. In order to compare the behavior of ensembles of walkers generated by GAWI and STRAW, VMC calculations on the RDX molecule were carried out using the ASCI-QSC supercomputer at the Los Alamos National Laboratory. This machine

Initialization	Eq steps	$\langle E \rangle$ (au)	Total samples
GAWI	10K	$-893.114 \pm 0.0122$	26,324,421
GAWI	15K	$-893.235 \pm 0.0103$	26,334,855
GAWI	20K	$-893.275 \pm 0.0102$	26,024,948
GAWI	25K	$-893.291 \pm 0.0184$	26,272,857
GAWI	30K	$-893.296 \pm 0.0101$	26,291,124
STRAW	100	$-893.298 \pm 0.0099$	25,145,777
STRAW	20K	$-893.291 \pm 0.0117$	26,071,024

Table 5.1: VMC calculations on RDX were carried out using 512 processors with 5 walkers per processor to compare different initialization methods and equilibration lengths. Calculations with too few equilibration steps are contaminated by nonequilibrated samples and do not agree with calculations that are allowed to equilibrate. RDX calculations initialized with GAWI require 30,000 steps to equilibrate, while 100 steps are sufficient when STRAW is used.

is composed of 256 4 CPU HP/Compaq Alphaserer ES45s running at 1250 MHz. Calculations using 512 processors, 5 walkers per processor, and varying equilibration lengths were run until about 26 million samples were collected. The results are summarized in table 5.1.

Expectation values calculated using equilibrated walkers should be approximately independent of time, with random fluctuations. A long term, low frequency drift in an expectation value as samples are collected is a sign of contamination by nonequilibrated configurations. Figure 5.5 shows the energy expectation value vs the number of samples collected for these calculations. The left side of the figure shows the calculations initialized with GAWI. In the calculations with less than 30,000 equilibration steps, we see a monotonic decrease in the expectation value of the energy as points are collected. These calculations are contaminated with high energy, nonequilibrated configurations from the beginning of the calculation, and the expectation value decreases as equilibrated samples are added. The energy expectation value in the calculation with 30,000 equilibration steps has the desired behavior, fluctuating about the limit with no long term drift.

The calculations initialized with STRAW used 100 and 20,000 equilibration steps. We use a minimum of 100 equilibration steps in our calculations as a safety margin to

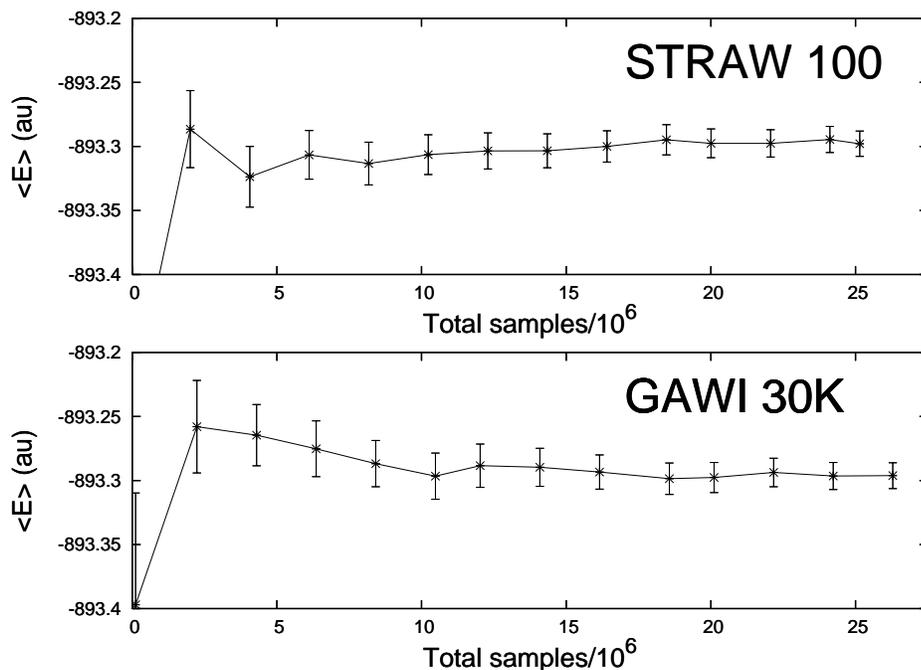


Figure 5.6: The energy expectation values and standard deviations for two uncontaminated RDX VMC calculations are shown. The calculations used 512 processors with 5 walkers per processor. The eventual answer is within one standard deviation of the expectation value at almost every point during the calculations. This shows that for uncontaminated calculations, the standard deviation of the energy expectation value is a good measure of its convergence.

be sure that each walker has at least one accepted move during the equilibration phase. The behavior of the energy expectation value in these calculations is very similar to that of the calculation initialized with GAWI using 30,000 equilibration steps. These three calculations show no signs of contamination. In Table 5.1, we see that their expectation values all agree to within one standard deviation of each other. These results demonstrate that 100 steps is sufficient for equilibration for RDX ensembles initialized with STRAW, while 30,000 equilibration steps are necessary when GAWI is used.

In QMcBeaver, standard deviations for expectation values are calculated using DDDA [49], which averages samples into blocks in order to account for their serial correlation. If we examine the results for the calculations initialized with GAWI using 10,000, 15,000, and 20,000 equilibration steps in table 5.1, we see that their

energy expectation values do not agree with each other or those of the equilibrated calculations to within one standard deviation. This is important because it shows that the standard deviation calculated during a contaminated calculation does not necessarily reflect the inaccuracy of its expectation value. If a researcher specifies an equilibration phase that is too short and nonequilibrated configurations contaminate the statistics, the expectation values will be inaccurate, and their standard deviations will not be a reliable measure of their inaccuracy.

The energy expectation value in the calculation that used GAWI and 25,000 equilibration steps agrees with the equilibrated results to within one standard deviation, but its behavior in figure 5.5 still shows signs of contamination by high energy samples in the beginning of the run.

In contrast, figure 5.6 replots the energy expectation value for two of the uncontaminated calculations. The error bars show the standard deviation of the expectation value. For these uncontaminated calculations, the eventual answer is within the range  $\langle E \rangle \pm \sigma \langle E \rangle$  at almost every point. In an uncontaminated calculation, we see that the standard deviation calculated by DDDA as the calculation progresses is a good measure of the level of convergence of the expectation value.

### 5.3.3 Timing and Spatial Correlation

Because it is more complicated than GAWI, STRAW takes more time to generate an initial electronic configuration for a walker. The new initialization method would be of little use if the time it took to generate an initial configuration was greater than the time saved in equilibration steps. Although coordinates are generated for each electron individually, the use of splines makes the process very inexpensive. Generating an initial configuration using STRAW requires less time than two VMC iterations for each of the molecules examined in this work.

The equilibration phase of a QMC calculation allows the walkers to become independent of their initial configurations and, by extension, each other. Since our objective is to shorten the equilibration phase of the calculation, an important objec-

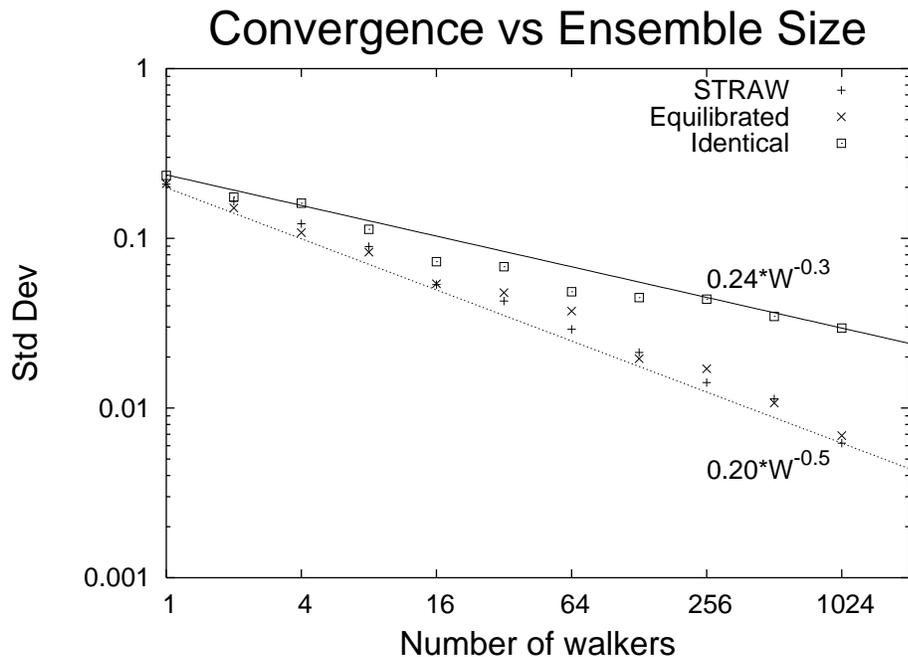


Figure 5.7: Standard deviation of energy expectation values for single processor ethanol calculations using equilibrated, identical, and STRAW ensembles after 200,000 propagation steps. The points for the equilibrated and STRAW ensembles are very close to the function  $0.20 \cdot W^{-0.5}$ , which shows that walkers generated by STRAW are statistically independent of each other.

tion to using STRAW could be raised if it led to *spatial correlation*, or any kind of statistical dependence within the ensemble of walkers.

Testing for spatial correlation in an ensemble of walkers is difficult. Vectors can be tested for spatial correlation by taking dot products, but a comparison of electronic configurations must take into account the indistinguishability of identical particles and the symmetry of the molecule. We avoid these difficulties by instead examining the statistical consequences of spatial correlation. If the walkers are indeed independent of each other, we expect that for a fixed number of iterations, the standard deviation of the energy expectation value will be proportional to  $\frac{1}{\sqrt{W}}$ , where  $W$  is the number of walkers. Any spatial correlation among the walkers will result in a different trend.

Single processor VMC calculations with 200,000 propagation steps and different ensemble sizes were carried out using the ethanol molecule. The trial function has the form  $\Psi_{VMC} = \Psi_{SCF}J$ , where  $\Psi_{SCF}$  is a HF/6-311G\*\*++ wavefunction calculated

using Jaguar [87] and  $J$  is the Jastrow function described in equations 5.12 and 5.13. One series of calculations was initialized with GAWI and allowed to equilibrate for 200,000 steps. We assume the walkers in these ensembles are independent of each other and their initial configurations. A second series was started with ensembles of identical walkers. These ensembles start with perfect spatial correlation. A third series of calculations was initialized with STRAW and used 100 equilibration steps.

Figure 5.7 shows the results for the different ensembles. The points for the equilibrated and STRAW ensembles are close to each other and the function  $0.20 \cdot W^{-0.5}$ , which is what we expect for independent walkers. The points for the identical ensembles, on the other hand, are very close to the function  $0.24 \cdot W^{-0.3}$ . Because they do not sample as much configuration space as independent walkers, ensembles with a high degree of spatial correlation generate less information than ensembles that are independent. Although the equilibration phase is very short, the statistical behavior of the STRAW ensembles is very similar to that of the equilibrated ensembles and to the behavior expected of independent walkers.

The initial electronic configurations generated by STRAW are statistically independent of each other and in regions of high enough probability density that a long equilibration phase is not necessary. The initialization algorithm is based on general principles of electronic structure, such as energy levels and the Pauli principle. It does not, however, generate configurations directly with respect to  $\rho_{VMC}$ , and is not meant to substitute for Metropolis sampling.

### 5.3.4 Parallel Calculation Efficiency

The equilibration phase of a QMC calculation is an inherently serial step: the walkers on each processor must be equilibrated individually, so adding more processors increases the time spent on this phase of the calculation. Knowing the appropriate number of iterations to exclude is vital, because leaving out too many wastes computer time, while leaving out too few will result in nonequilibrated values contaminating the statistics.

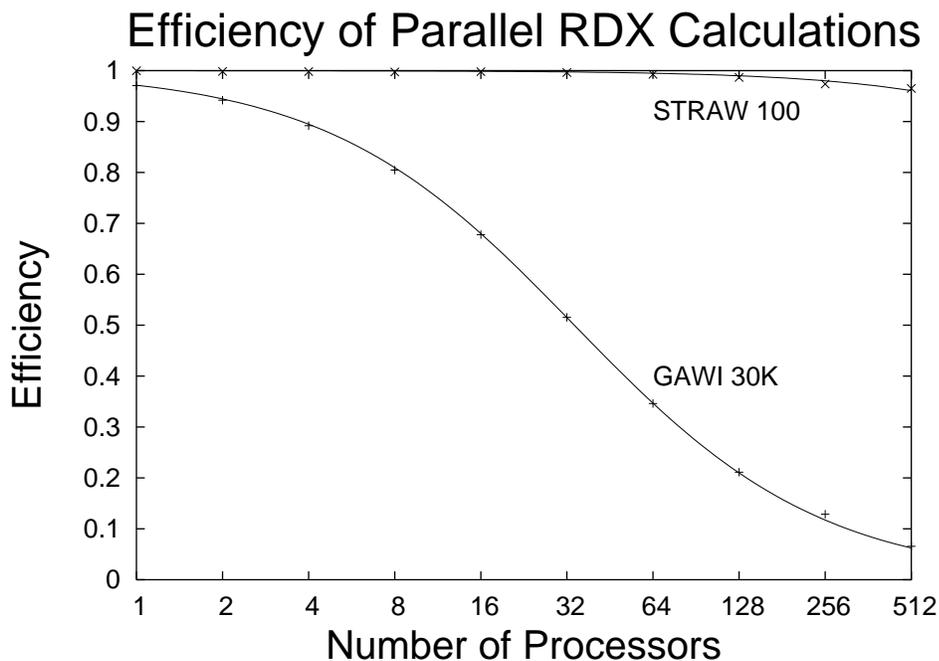


Figure 5.8: RDX calculations with one walker per processor were run until 1,000,000 total samples were collected. The calculations initialized with GAWI used 30,000 equilibration steps, while the calculations initialized with STRAW used 100 equilibration steps. Decreasing the number of equilibration steps greatly improves the efficiency of calculations with large numbers of processors. The data are fit to  $\epsilon(N_{Processors}) = \frac{a}{a + N_{Processors}}$  with  $a = 34.0$  for GAWI and  $a = 12,514.0$  for STRAW.

The calculations of sections 5.3.2 and 5.3.3 give us confidence that STRAW can generate independent initial configurations for RDX and  $\text{SiCl}_4$  in regions of high enough density that one hundred equilibration steps is sufficient before calculating expectation values. We expect that the electronic structure of other molecules composed of atoms from the first three rows of the periodic table will be similar enough to these examples to allow STRAW to be successful for them as well.

To demonstrate the effect of shortening the equilibration phase of a calculation, a scaling experiment was performed on ASCI-QSC. VMC calculations on RDX were conducted using 1,000,000 total propagation steps and 1 walker per processor. Following the results of Section 5.3.2, 30,000 equilibration steps were used in the calculations initialized with GAWI, while 100 equilibration steps were used with STRAW. The efficiency of each calculation was found using equation 5.7.

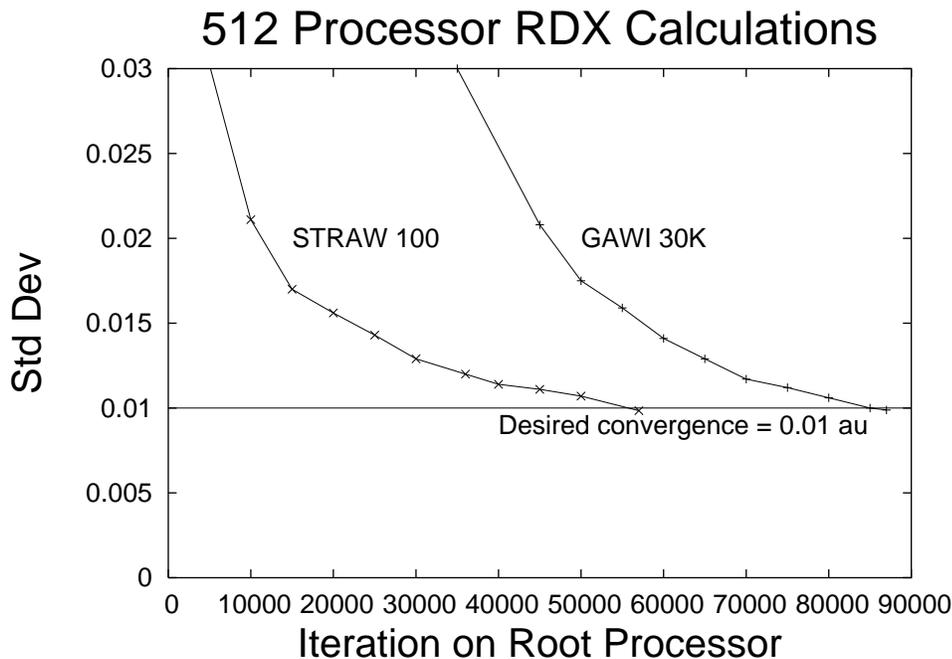


Figure 5.9: VMC calculations on RDX were carried out using 512 processors and 5 walkers per processor. The total iterations on the  $x$  axis include the equilibration phase of the calculations. Initializing the walkers with STRAW decreases the wall clock time needed to calculate the RDX energy expectation value to within 0.01 au from 9.4 hours to 6.3 hours, an improvement of 33%.

The points in figure 5.8 were fit to the function  $a/(a + N_{Processors})$ . The value for  $a$  for GAWI is 34.0, while for STRAW it is 12,514.0. This result clearly shows the effect of reducing the number of equilibration steps on the efficiency of parallel calculations. The experiment has a short statistics gathering phase, which makes it scale particularly badly as the number of processors increases. In a realistic calculation on RDX, many more steps will have to be used before the expectation values converge to within chemical accuracy. A calculation with a longer statistics gathering phase will scale more favorably as the number of processors increases, which can be seen by examining equation 5.7. As computers with large numbers of processors come into general use [80, 42], however, the equilibration phase will limit the efficiency of any calculation.

The most important consequence of reducing the length of the equilibration phase with STRAW is that the improvement in parallel efficiency will speed the calcula-

Initialization	Eq steps	$\langle E \rangle$ (au)	Steps on root proc	Eff	Wall clock time
GAWI	30K	$-893.291 \pm 0.0099$	85500	65.0%	9.40h
STRAW	100	$-893.287 \pm 0.0098$	57048	99.8%	6.30h

Table 5.2: RDX calculations using 512 processors and 5 walkers per processor were run until 0.01 au convergence in the energy expectation value was achieved. The calculation initialized with GAWI used 30,000 equilibration steps, while the calculation initialized with STRAW used 100 equilibration steps. The calculation initialized with STRAW took 6.3 hours to converge, while the calculation initialized with GAWI took 9.4 hours.

tion of converged expectation values. Using an automatic method to terminate the calculation based on the convergence of the energy expectation value [49, 50], RDX calculations using 512 processors with five walkers per processor were run until the expectation value of the energy converged to within 0.01 au or 6.27 kcal/mol. The calculations initialized with GAWI used 30,000 equilibration steps, while the calculations initialized with STRAW used 100 equilibration steps.

Table 5.2 summarizes the results from these calculations, and figure 5.9 shows the standard deviation of the energy expectation value vs total iterations on the root processor. The total iterations include the equilibration phase, and we see that the two calculations have very similar convergence behavior, with the calculation initialized with GAWI offset by about 30,000 iterations compared to the one initialized with STRAW. The calculation initialized with STRAW converged to the desired level in 6.3 hours with 99.8% efficiency, while the calculation initialized with GAWI took 9.4 hours with 65.0% efficiency.

## 5.4 Conclusion

We have presented and tested STRAW, a simple and automatic method for generating initial electronic configurations for QMC calculations. STRAW is based on the structure of the energy levels of atoms and distributes the electrons in annular shells. The electrons in each energy level are distributed with respect to probability distri-

butions in the angular coordinates so that they avoid each other. The configurations generated by STRAW are statistically independent of each other and are in regions of high probability density, which reduces the length of the equilibration phase of the calculation, during which the statistics must be discarded. STRAW has been implemented in QMcBeaver, an open source QMC program [81].

Using an appropriate equilibration length is vital, because when the statistics are contaminated by nonequilibrated configurations, both the expectation values and their standard deviations can be inaccurate. STRAW simplifies the job of the user to specify the equilibration length by generating initial configurations that show no signs of contamination or spatial correlation after an equilibration phase of one hundred iterations.

Shortening the equilibration phase increases the efficiency of parallel QMC calculations and decreases the amount of computer time needed to calculate converged expectation values. For example, using STRAW instead of a standard initialization method in 512 processor calculations decreases the time needed to calculate the energy expectation value of a trial function for an RDX molecule to within 0.01 au from 9.4 hours to 6.3 hours, an improvement of 33%.

Using STRAW improves the parallel scaling of QMC and will increase the efficiency of calculations using tens to hundreds of thousands of processors. This will, in turn, allow highly accurate simulations on a broader range of chemically interesting systems than is possible today. QMC results will be useful as benchmarks for training force fields for molecular dynamics simulations and developing new density functional (DFT) methods. There are several classes of systems that have proven elusive for current DFT methods [91, 92]. Reproducing QMC results for these systems will be an important goal for the next generation of DFT methods.

Clearly, many other schemes for generating initial electronic configurations for QMC calculations are possible. The tests for equilibration of the Markov chain, initialization time, and spatial correlation described in the sections 5.3.2 and 5.3.3 will provide a basis for comparison of future initialization schemes.

## Chapter 6

# A Quantum Monte Carlo Study of Three Pericyclic Hydrocarbon Reactions

# Abstract

Diffusion quantum Monte Carlo calculations using Hartree-Fock, generalized valence bond, and multiconfiguration self-consistent Field trial functions were carried out for three pericyclic hydrocarbon reactions. The enthalpies of activation and reaction are compared to experimental, CCSD(T), and CBS-QB3 results, as well as those of B3LYP and the recently introduced X3LYP, XYG3, and M06 family of density functional methods.

For all three reactions, B3LYP geometries and zero point energies combined with DMC electronic energies calculated with the appropriate trial function result in accuracy comparable to CCSD(T) and CBS-QB3. HF trial functions are sufficient for C-C  $\sigma$  bonds, while GVB trial functions are necessary for  $\pi$  bonds. For molecules with multiple  $\pi$  bonds and transition states with several bonds being formed or broken, MCSCF trial functions must be used.

## 6.1 Introduction

The advances in computing power and electronic structure theory in recent years have increased the role of simulations in understanding systems in chemistry and materials science. Electronic structure calculations can provide information on the geometry, molecular orbitals, and vibrations of not only stable molecules, but also transition states and reaction intermediates that are impossible to observe experimentally. Understanding the properties of transition states is crucial in areas such as the development of new catalysts and energetic materials. Transition states are difficult cases for electronic structure methods, with the errors for most methods larger by a factor of three or four for transition states than for stable molecules.

Quantum Monte Carlo (QMC) is a class of stochastic electronic structure methods that can, in principle, calculate expectation values to within chemical accuracy [57, 59]. Although the expense of QMC will keep it from replacing traditional methods such as Density Functional Theory (DFT) for routine calculations, its favorable scaling [62, 63, 64] and parallelizability [65] will allow QMC calculations on systems too large for other comparably accurate methods. QMC has the potential to resolve disagreements and provide benchmark results when other electronic structure methods are too expensive or not reliable enough for a certain application.

The QMC variants used in this work are variational Monte Carlo (VMC), in which the adjustable parameters of a trial wavefunction are optimized, and diffusion Monte Carlo (DMC), which simulates a diffusion process to sample the exact ground state wavefunction of a system. The most common formulation of DMC uses two main approximations. First, the factorization of the Green's function that propagates the walkers is exact only for a time step of zero. In order to propagate the walkers and sample configuration space, however, a finite time step must be used. As the size of the time step increases, configuration space is sampled more quickly, but the time step error increases. The second source of error is the fixed node approximation, in which the nodal structure of the exact ground state is assumed to be the same as that of the SCF part of the trial wavefunction. In this work, we explore both sources

of error for stable molecules and transition states by conducting DMC calculations on three pericyclic hydrocarbon reactions. Trial functions were constructed for each reaction using Hartree-Fock (HF), generalized valence bond (GVB) and multiconfiguration self-consistent field (MCSCF) wavefunctions and used in DMC calculations with time steps from  $10^{-2}$  to  $10^{-4}$  au. The time step and fixed-node errors in the activation barriers and overall reaction energies are analyzed to develop guidelines for calculations on larger systems.

There are many interesting systems in chemistry and materials science that are too large for highly accurate SCF methods, and for which DFT methods are unreliable. DMC has the potential to provide accurate expectation values in these cases. Because the DMC results will become the benchmarks against which other calculations are judged, it is essential that they be accurate. In order to carry out accurate and efficient DMC calculations, researchers will have to use the appropriate trial function and time step for the system being studied. Studying reactions involving small molecules, for which several time steps and trial functions can be compared, and for which the results of experiments and other high quality calculations are available, will provide a base of knowledge for researchers addressing larger problems.

## 6.2 Reactions

Figure 6.1 shows the three reactions studied in this work. Reaction 1 is the 2+2 cycloaddition of ethylene to form cyclobutane. Several mechanisms have been investigated for this reaction. The supra-supra pathway with  $D_{2h}$  symmetry is a classic example of a reaction forbidden by orbital symmetry [93]. In the supra-supra pathway with  $C_{2v}$  symmetry, a biradical tetramethylene chain is formed, which closes to form cyclobutane. In the supra-antara pathway, which we examine in this work, the C-C bond of one ethylene twists during the course of the reaction. This mechanism is allowed by orbital symmetry, but has a very high activation barrier because bonding cannot be maintained as the reaction proceeds. In the transition state, the four carbons have a dihedral angle of about 40 degrees.

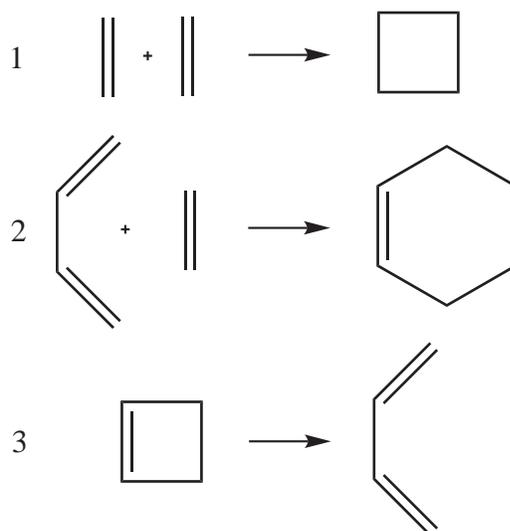


Figure 6.1: Reactions 1, 2, and 3.

Reaction 2 is the 4+2 cycloaddition of ethylene and butadiene to form cyclohexene. This thermally allowed reaction is the prototype Diels-Alder reaction and represents an important class of reactions in organic synthesis. Reaction 3 is the ring opening of cyclobutene to form butadiene. For reaction 3, we examine the symmetry-allowed conrotatory pathway, in which the two terminal  $\text{CH}_2$  groups rotate in the same direction. The disrotatory pathway is forbidden by orbital symmetry.

Although these reactions involve small molecules, with at most six carbon atoms, the breaking and forming of  $\sigma$  and  $\pi$  bonds make them difficult cases for density functional methods, with most methods predicting reaction enthalpies with errors of four to five kcal/mol. We compare the DMC results with those of B3LYP and the recently developed X3LYP, XYG3, and the M06 family of density functional methods.

### 6.3 Experimental and Computational Results

In all three mechanisms for reaction 1, bonds must be broken for the reaction to happen. As a result, the activation enthalpies are very high and the reaction is difficult to carry out under thermal conditions. Reliable experimental results for the enthalpies of activation and reaction are not available. Using the experimental enthalpies of

formation of ethylene [94] and cyclobutane [95], we calculate  $\Delta H_{0K} = -16.5$  kcal/mol for reaction 1. In 2002, Sakai calculated an MP2/CAS/6-311+G(d,p) 0K activation enthalpy for the supra-antara pathway of 77.6 kcal/mol [96]. In 2006, Sirjean et al. calculated a CBS-QB3 [97] 0K enthalpy for reaction 1 of -16.66 kcal/mol [98].

Detailed experimental and computational results for reactions 2 and 3 were assembled by Guner et al. in 2003 [99]. They compared HF, MP2, CASSCF, CASPT2, B3LYP, BPW91, MPW1K, KMLYP, and CBS-QB3 results with experiment for 11 pericyclic hydrocarbon activation and reaction enthalpies. Reactions 2 and 3 of this work are reactions 1 and 7, respectively, of theirs. We use their experimental values of  $\Delta H_{0K} = -39.6$  kcal/mol and  $\Delta H_{0K}^\ddagger = 23.3 \pm 2$  kcal/mol (later revised to 25.0 kcal/mol [100]) for reaction 2 and  $\Delta H_{0K} = -10.6 \pm 1$  kcal/mol and  $\Delta H_{0K}^\ddagger = 31.9 \pm 0.2$  kcal/mol for reaction 3.

## 6.4 Computational Methods

The structures of the reactants, products, and transition states of reactions 1, 2, and 3 were optimized with Jaguar [87] using B3LYP DFT [33] and the 6-311G\*\* basis set [101]. Frequency calculations were carried out to verify the optimized geometries and calculate zero point energies. All stable molecules had no negative frequencies, and the transition states for reactions 2 and 3 each had one negative frequency corresponding to the desired reaction. The transition state for reaction 1 had two negative frequencies, one corresponding to the reaction and the another corresponding to a rocking motion. Zero point energies were calculated using unscaled frequencies.

*Ab initio* SCF calculations were conducted with GAMESS [102] using the B3LYP/6-311G\*\* geometries to provide trial functions for the DMC calculations. The aug-cc-pVTZ basis set was obtained from the EMSL website [103, 104] and used for all of the wavefunctions. HF wavefunctions were calculated for all molecules to provide a “zero correlation” starting point. GVB and MCSCF wavefunctions were calculated to see the effect of correlated trial functions on the QMC results.

In a GVB-PP calculation, geminal pairs are defined, each of which consists of two

orbitals and two singlet paired electrons. The first orbital is usually a bonding orbital occupied in the HF configuration. The second orbital is usually the corresponding antibonding orbital, orthogonal to the first and unoccupied in the HF configuration. The two orbitals are combined to form two singlet paired GVB orbitals that allow the electrons to avoid each other [16].

For reaction 1, two sets of GVB wavefunctions were calculated. The first defined two geminal pairs for the product, transition state, and reactants, containing the electrons and orbitals involved in the reaction. The  $\pi$  and  $\pi^*$  orbitals of ethylene were used as a geminal pair. Two nonneighboring  $\sigma$  bonds of cyclobutane were correlated with their  $\sigma^*$  orbitals, and the corresponding intermediate bonds of the transition state were correlated with their antibonding orbitals. Because these wavefunctions involved two electron pairs and four orbitals, they are labeled GVB 2,4. Because the C-C bonds of cyclobutane are equivalent, a second set of GVB wavefunctions were calculated with four geminal pairs for each molecule. All four  $\sigma$  bonds of cyclobutane, the  $\sigma$  and  $\pi$  orbitals of ethylene, and the corresponding four orbitals of the transition state were correlated. These wavefunctions are labeled GVB 4,8.

GVB wavefunctions were calculated for the molecules of reactions 2 and 3 with geminal pairs for the electrons and orbitals involved in the reaction. Three pairs in six orbitals were used for reaction 2, and two pairs in four orbitals were used for reaction 3.

In an MCSCF calculation, an active space consisting of a subset of the orbitals and electrons of a molecule is defined. A CI calculation is carried out in which the active orbitals are occupied. The orbitals and CI expansion coefficients are optimized, giving a very general description of the electronic structure of the molecule [13]. If a full CI calculation is used, and all possible occupations of the active orbitals are considered, the calculation is called a Complete Active Space SCF (CASSCF) [14] or Fully Optimized Reactive Space (FORS) [15] calculation.

CASSCF wavefunctions were calculated for the molecules using the same active orbitals as the GVB calculations. CASSCF wavefunctions with an active space of four electrons in four orbitals and eight electrons in eight orbitals were calculated for

reaction 1. An active space of six electrons in six orbitals was used for reaction 2, and four electrons in four orbitals for reaction 3. For simplicity, the MCSCF wavefunctions were labeled in the same manner as the GVB wavefunctions.

Density functional theory (DFT) is currently the most popular way for researchers to include the effects of electron correlation in calculations. The small expense, accuracy, and favorable scaling of DFT calculations have enabled theorists to make significant contributions to chemistry, physics, and materials science. DFT methods are based on the theorem of Hohenberg and Kohn, which proves the existence of a functional of the electron density that will give the exact energy [27]. Practical implementations do not deal with the electron density directly, but use the orbital formulation of Kohn and Sham to express the wavefunction [28]. New density functionals are constantly being introduced with parameters optimized for certain classes of reactions, but the goal of consistent results within chemical accuracy for a broad range of systems has not yet been achieved. The most widely used DFT method is B3LYP, introduced by Becke in 1993 [33].

In this work, we carried out calculations on reactions 1, 2, and 3 using three recently introduced density functional methods.

The X3LYP functional is based on the exact form of the exchange energy density for an electron density decaying with Gaussian-like behavior at long range [34]. An exchange functional with the correct behavior is described as a linear combination of the Becke [30] and Perdew-Wang [105] exchange functionals. X3LYP was designed to improve the accuracy for noncovalent interactions, such as hydrogen bonds and electrostatic and van der Waals interactions, for use in simulating the binding of ligand molecules with proteins. X3LYP was demonstrated to have excellent results for nonbonded systems such as noble gas dimers and water clusters, as well as for heats of formation, ionization potentials, and electron affinities.

The M06 suite is a family of four density functional methods, each parameterized for different systems. M06 is parameterized for both metals and nonmetals, while M06-2X has twice the nonlocal exchange and is intended for nonmetals only [35]. M06-L is a local functional, which reduces the computational expense for large sys-

tems [106]. M06-HF includes the full HF exchange energy for the Kohn-Sham orbitals, which makes it suitable for one electron systems and long range charge transfer excited states [107].

The XYG3 functional includes an exact exchange term as well as information about the unoccupied Kohn-Sham orbitals in a second-order perturbation theory term [36]. The PT2 term causes the method to scale less favorably than other density functionals, as  $O(N^5)$  instead of  $O(N^4)$ , but gives XYG3 extremely high accuracy for enthalpies of formation and reaction barriers.

### 6.4.1 Quantum Monte Carlo

The distinguishing feature of QMC calculations is the use of electronic configurations generated randomly with respect to quantum mechanical probability densities to calculate expectation values [56, 66]. In variational Monte Carlo (VMC), the configurations are used to optimize the adjustable parameters of the wavefunction.

First, the expression for the expectation value of the energy of a trial wavefunction is rewritten as a weighted average:

$$\langle E \rangle = \frac{\int d\vec{R} \Psi_T(\vec{R}) \hat{H} \Psi_T(\vec{R})}{\int d\vec{R} |\Psi_T(\vec{R})|^2} \quad (6.1)$$

$$= \frac{\int d\vec{R} |\Psi_T(\vec{R})|^2 E_L(\vec{R})}{\int d\vec{R} |\Psi_T(\vec{R})|^2} \quad (6.2)$$

$$= \int d\vec{R} \rho_{VMC}(\vec{R}) E_L(\vec{R}), \quad (6.3)$$

where  $\hat{H}$  is the Hamiltonian operator for the system and  $\vec{R}$  is a  $3N$ -dimensional vector containing the coordinates of the  $N$  electrons of the molecule. In equation 6.3,  $\rho_{VMC}(\vec{R})$  is the probability density for the electronic configuration  $\vec{R}$ :

$$\rho_{VMC}(\vec{R}) = \frac{|\Psi_T(\vec{R})|^2}{\int d\vec{R} |\Psi_T(\vec{R})|^2}. \quad (6.4)$$

$E_L(\vec{R})$  is its local energy:

$$E_L(\vec{R}) = \frac{\hat{H}\Psi_T(\vec{R})}{\Psi_T(\vec{R})}. \quad (6.5)$$

A series of  $M$  independent electronic configurations,  $\{\vec{R}_i\}$ , is generated with respect to  $\rho_{VMC}$  using the Metropolis algorithm [54]. The configurations are used to calculate the expectation value of the energy:

$$\langle E \rangle = \frac{1}{M} \sum_i^M E_L(\vec{R}_i) \pm O\left(\sqrt{\frac{1}{M}}\right). \quad (6.6)$$

Because the trial functions are evaluated stochastically, they do not have to be analytically integrable, which gives researchers considerable freedom in choosing their form. In most cases, the trial functions are written in the following form:

$$\Psi_T = \left( \sum_i c_i \Psi_i^{SCF} \right) J, \quad (6.7)$$

where the  $\Psi_i^{SCF}$  are one or a small number of Slater determinant wavefunctions calculated by traditional electronic structure methods such as HF, GVB, or MCSCF. The Jastrow function,  $J$ , is a symmetric function of the interparticle coordinates that accounts for short range correlations and allows the trial function to satisfy the quantum mechanical cusp conditions for collisions between particles [9].

Most VMC methods employ correlated sampling, in which expectation values for several sets of parameters are calculated with one set of configurations. This technique allows the differences between sets of parameters to be determined with much higher precision than if the results from separate runs are compared [66]. Algorithms that minimize a combination of the expectation value of the energy and its variance to optimize the adjustable parameters of the trial functions Jastrow have been shown to be effective and efficient [67].

In Diffusion Monte Carlo (DMC), a mixed distribution is defined:

$$f_{DMC}(\vec{R}, \tau) = \Phi(\vec{R}, \tau) \Psi_T(\vec{R}), \quad (6.8)$$

where  $\Psi_T(\vec{R})$  is a trial function that approximates the ground state of the system and  $\Phi(\vec{R}, \tau)$  satisfies the time-dependent Schrödinger equation for the system.

The mixed distribution satisfies a Fokker-Planck equation:

$$-\frac{\partial}{\partial \tau} f(\vec{R}, \tau) = \left[ -\frac{1}{2} \nabla^2 + \nabla \cdot V(\vec{R}) - S(\vec{R}) \right] f(\vec{R}, \tau), \quad (6.9)$$

where  $V(\vec{R})$  is the local velocity of the trial function:

$$V(\vec{R}) = \frac{\nabla \Psi_T(\vec{R})}{\Psi_T(\vec{R})}, \quad (6.10)$$

and  $S(\vec{R})$  is defined in terms of its local energy:

$$S(\vec{R}) = E_T - E_L(\vec{R}). \quad (6.11)$$

$E_T$  is a shift in energy that approximates the true ground state energy.

In order to propagate equation 6.9 with  $\tau$ , an equivalent integral equation is written:

$$f(\vec{Y}, \tau + d\tau) = e^{d\tau E_T(\tau + d\tau)} \int d\vec{R} G(\vec{Y}, \vec{R}, d\tau) f(\vec{R}, \tau), \quad (6.12)$$

where  $G(\vec{Y}, \vec{R}, d\tau)$  is the Green's function for the case  $E_T = 0$ .

The three terms on the right side of equation 6.9 describe diffusion, drift, and branching processes. The exact Green's function cannot be written, but it can be approximated by a product of Green's functions for the three individual processes:

$$\begin{aligned} G(\vec{Y}, \vec{R}, d\tau) &\approx \frac{1}{(2\pi\tau)^{3N/2}} \delta[\vec{Z} - \vec{R} - V(\vec{R})d\tau] \times \\ &\int d\vec{Z} \exp\left[\frac{(\vec{Y} - \vec{Z})^2}{2d\tau}\right] \times \\ &e^{-\frac{1}{2}[E_L(\vec{Y}) + E_L(\vec{R})]d\tau} + O(d\tau^2). \end{aligned} \quad (6.13)$$

Because the diffusion, drift, and branching terms do not commute, equation 6.13 is

exact only for  $d\tau = 0$ . In DMC calculations, expectation values are calculated for several values of  $d\tau$  and extrapolated to  $d\tau = 0$ . Because configuration space is sampled more slowly, the number of iterations needed to achieve a certain level of convergence increases when the time step size is decreased.

As  $\tau \rightarrow \infty$ ,  $f_{DMC}$  approaches  $f_0 = \Phi_0\Psi_T$ , where  $\Phi_0$  is the true ground state wavefunction for the system. For operators that commute with the Hamiltonian, expectation values calculated using this distribution equal those of the exact ground state. In order for  $f_0$  to be sampled, it must be interpreted as a density. Since a density cannot be negative,  $\Psi_T$  and  $\Phi_0$  must have the same sign throughout configuration space.

Many electron wavefunctions have positive and negative regions separated by nodes, on which they have zero value. In order for  $\Psi_T$  and  $\Phi_0$  to have the same sign throughout configuration space, they must have the same nodal structure. Unfortunately, it is impossible to construct a trial function with the nodal structure of the exact ground state.

The simplest solution to the nodal problem is known as the fixed node approximation, in which the nodes of the exact wavefunction are assumed to be the same as those of the trial function. The nodes are enforced by rejecting any proposed move that crosses a node and changes the sign of  $\Psi_T$ .

Fixed node DMC imposes a boundary condition on the ground state wavefunction that it vanish at the nodes of the trial function. The simulation will converge to the best possible solution to the Schrödinger equation within the nodal structure of the trial function. The resulting energy will be an upper bound to the true energy and will be variational with respect to the nodal structure [73]. It has been shown that the error in the fixed node energy is second order in the error of the nodes [74].

Because the Jastrow function is symmetric with respect to particle interchange, the SCF part of the trial function determines its nodal structure. For a DMC calculation to give accurate expectation values, it is essential that the trial function use the appropriate SCF wavefunction.

### 6.4.2 QMC Procedures

A Jastrow function similar to that of Drummond et al. [71] was used for all of the trial functions. Their two body Jastrow is a polynomial that goes to zero at a cutoff distance:

$$\chi_{ij}(r_{ij}) = (r_{ij} - L_{ij})^C \Theta(L_{ij} - r_{ij}) \left( \alpha_{0ij} + \left[ \frac{\Gamma_{ij}}{(-L_u)^C} + \frac{\alpha_{0ij}C}{L_{ij}} \right] r_{ij} + \sum_{l=2}^N \alpha_{lij} r_{ij}^l \right), \quad (6.14)$$

where  $r_{ij}$  is the distance between particles  $i$  and  $j$ .  $\Theta$  is the Heaviside function, and  $L_{ij}$  is the cutoff distance.  $\Gamma_{ij}$  enforces the cusp condition for the particles: it is set to  $\frac{1}{2}$  for opposite spin electrons,  $\frac{1}{4}$  for same spin electrons, and the opposite of the nuclear charge if particles  $i$  and  $j$  are an electron and a nucleus. If  $C = 2$ , the gradient of  $\chi_{ij}$  is continuous at the cutoff but the second derivative is discontinuous. If  $C = 3$ , both the gradient and second derivative are continuous at the cutoff, making the local energy also continuous. The cutoff distance and coefficients  $\alpha_{lij}$  are adjustable.

We used  $C = 3$  and  $N = 8$  for every Jastrow function and scaled the interparticle distance by letting  $s_{ij} = \frac{r_{ij}}{L_{ij}}$ . We found the cutoff distance and other parameters much easier to optimize when the distances were scaled.

The molecular orbitals were modified near the nuclei using the procedure of Ma et al. [72]. The part of an orbital arising from  $s$ -type basis functions centered on a particular nucleus was replaced within a radius of correction of that nucleus by a function of the following form:

$$\tilde{\phi} = C + \text{sgn} [\tilde{\phi}(0)] \exp[p(r)], \quad (6.15)$$

where  $\text{sgn} [\tilde{\phi}(0)]$  is  $\pm 1$ , reflecting the sign of the replacement orbital at the nucleus.  $C$  is a shift chosen so that the replacement orbital does not change sign within the radius of correction. The coefficients in the polynomial  $p(r)$  are calculated by optimizing the behavior of the local energy while requiring the replacement orbital to satisfy the electron-nucleus cusp condition and the value and first and second derivatives to be continuous at the radius of correction. With these modifications, the orbitals satisfy

the electron-nucleus cusp condition, so the electron-nucleus cusp parameters in the Jastrow are set to zero.

The quantum Monte Carlo calculations were carried out with QMcBeaver, a QMC program developed in the Goddard group at Caltech [81]. The interprocessor communication and statistics gathering were done using QMC-MW, a manager-worker model that automatically balances the workload between the processors [50]. The statistics were analyzed by the dynamic distributed decorrelation algorithm (DDDA) [49], a reformulation of the Flyvberg-Peterson blocking algorithm [108] that greatly reduces the amount of data that has to be communicated when the results are gathered. The initial electronic configurations for the walkers were generated with STRAW, an algorithm that generates electronic configurations in regions of high probability density [109].

The Jastrow parameters and CI expansion coefficients were optimized with the linear method of Umrigar et al. [69]. The DMC algorithm of Umrigar et al. was used to propose and accept electronic configurations [55], and the algorithm of Assaraf et al. was used to reweight and branch the walkers [76].

## 6.5 Results and Discussion

VMC calculations were carried out for every trial function to optimize the CI coefficients and adjustable Jastrow parameters. The VMC optimizations were run using four processors with an ensemble of 100 walkers per processor. Using the optimized trial functions, DMC calculations were carried out on supercomputers at Los Alamos and Lawrence Livermore National Laboratories. DMC calculations run on Coyote at LANL and hera at LLNL used 512 processors with 100 walkers per processor, while DMC calculations run on uBGL at LLNL used 16,384 processors with 100 walkers per processor. The DMC calculations were typically run for 12 hours at a time and restarted from checkpoint files until the energy expectation value converged to within  $1.5 \times 10^{-4}$  au or about 0.09 kcal/mol. DMC calculations were run with time steps of 0.01, 0.003, and 0.001 au for every molecule. In addition, DMC calculations with

time steps of 0.0003 and 0.0001 au were run for ethylene, cyclobutane, cyclobutene, and butadiene. Limitations in available computer time prevented these calculations from being run for cyclohexene and the transition states.

The DMC expectation values for each trial function and time step are shown in figures 6.2 through 6.11. The expectation values for each trial function were fit to the function  $a + b(d\tau) + c(d\tau)^2$ . The time step error for every trial function is positive and increases with the time step. The fit function should increase with  $d\tau$ , but in some cases, it has a maximum between  $d\tau = 0.003$  and  $d\tau = 0.01$ . Adding terms to the polynomial did not alleviate this behavior or improve the fit of the function. Since the function is being used to extrapolate the expectation values to the  $d\tau = 0$  limit and the fit is uniformly excellent in the small  $d\tau$  region, the incorrect behavior of the function for large  $d\tau$  was ignored.

### 6.5.1 Reaction 1

Figure 6.2 shows the DMC results for ethylene. It is easy to see the nodal and time step errors in the calculations. The HF trial function has a nodal error, fairly constant with respect to time step, of approximately 0.004 au or 2.5 kcal/mol compared to the correlated trial functions. The DMC/GVB 1,2 and DMC/MCSCF 1,2 results coincide for every time step. These wavefunctions differ only in their localization, which should not affect their DMC energy. It is interesting to note the GVB 2,4 trial function, which has lower SCF and VMC energies than the GVB 1,2 trial function, has a DMC energy approximately 0.35 kcal/mol *higher* than that of the one pair GVB trial function. The two pair MCSCF trial function has a DMC energy approximately 0.2 kcal/mol *lower* than the one pair trial functions. Trial functions with lower SCF energy do not necessarily have higher quality nodes for DMC calculations.

The DMC results for TS1 are shown in figure 6.3. The results for the MCSCF 2,4 and MCSCF 4,8 trial functions are almost identical for every time step. The DMC/GVB 2,4 results are approximately 6 kcal/mol higher than the DMC/MCSCF energies. The GVB 4,8 trial function, which has lower SCF and VMC energies than

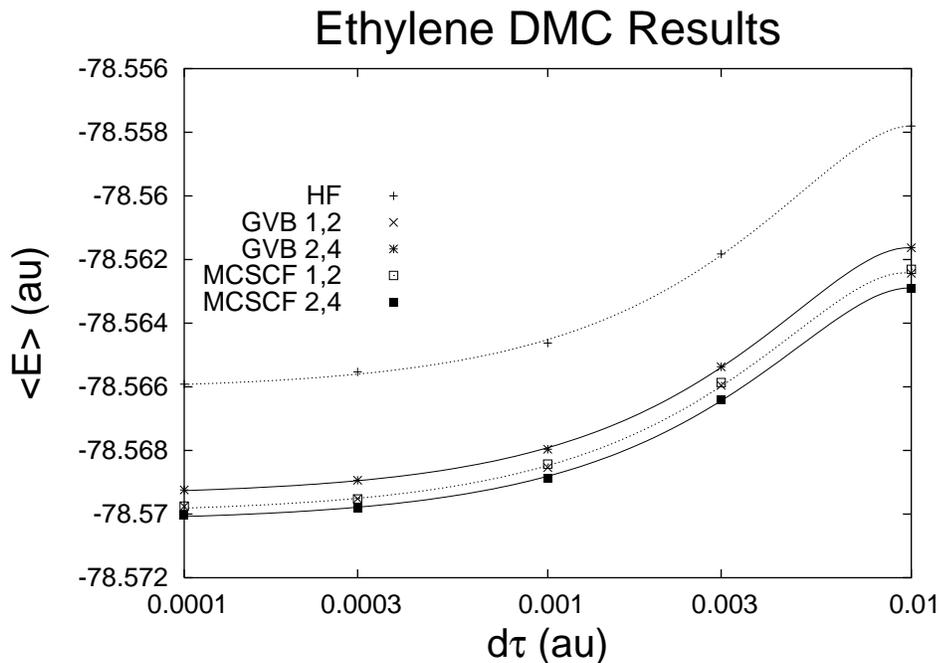


Figure 6.2: DMC expectation values for the energy of ethylene. The expectation values are fit to the function  $a + b(d\tau) + c(d\tau)^2$ . The expectation values for the GVB 1,2 and MCSCF 1,2 trial functions coincide at every time step. The energy for the GVB 2,4 trial function is slightly higher than that of the GVB 1,2 trial function at every time step, and the energy for the HF trial function is considerably higher than those of the correlated trial functions at every time step.

the GVB 2,4 trial function, has a DMC energy approximately 1 kcal/mol higher than the GVB 1,2 trial function. Once again, we see adding GVB pairs decreases the quality of the nodes of the trial function and raises the DMC energy. The energies for the HF trial function are about 1 kcal/mol above the GVB 4,8 DMC energies.

Figure 6.4 shows the DMC results for cyclobutane. The DMC results for all five trial functions are within 1 kcal/mol of each other for every time step. Cyclobutane is the only saturated hydrocarbon studied in this work, and these results show correlating C-C  $\sigma$  bonds does not significantly change the nodes for DMC calculations.

Figure 6.5 shows the enthalpies at 0K of activation and reaction for reaction 1, calculated using DMC electronic energies and B3LYP/6-311G\*\* geometries and zero point energies. The results for each time step and trial function are plotted with points, and the results for each trial function extrapolated to  $d\tau = 0$  are plotted

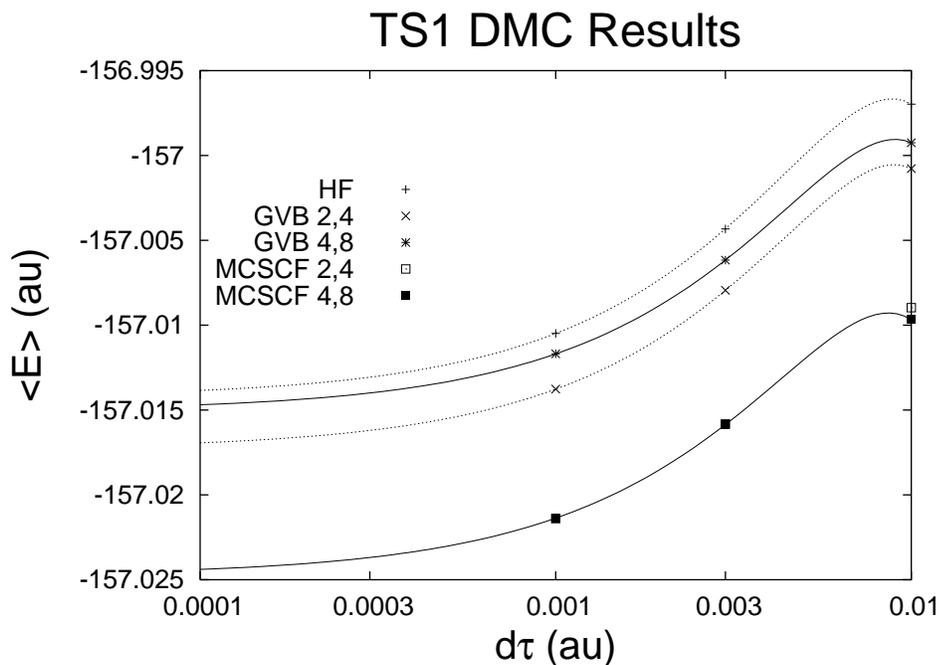


Figure 6.3: DMC expectation values for the energy of TS1. The expectation values are fit to the function  $a + b(d\tau) + c(d\tau)^2$ . The expectation values for the MCSCF 2,4 and MCSCF 4,8 trial functions coincide for the 0.003 and 0.001 au time steps. The energy for the GVB 4,8 trial function is higher than that of the GVB 2,4 trial function for every time step.

with dotted lines. The experimental value for the overall reaction enthalpy is plotted as a solid line. The DMC results for all four correlated trial functions are within 1 kcal/mol of the experimental value, but the HF/DMC result is about 4 kcal/mol too low. The large nodal error in the HF trial function for ethylene skews the result for the reaction far outside the desired accuracy.

We do not have an experimental or high quality *ab initio* activation enthalpy for reaction 1. The DMC/GVB activation enthalpies are about 79 kcal/mol, while the DMC/MCSCF results are about 74 kcal/mol. The DMC/HF value of 76 kcal/mol is closest to the MP2//CAS/6-311G(d,p) result of 77.6 kcal/mol [96], but we do not consider that value to be definitive.

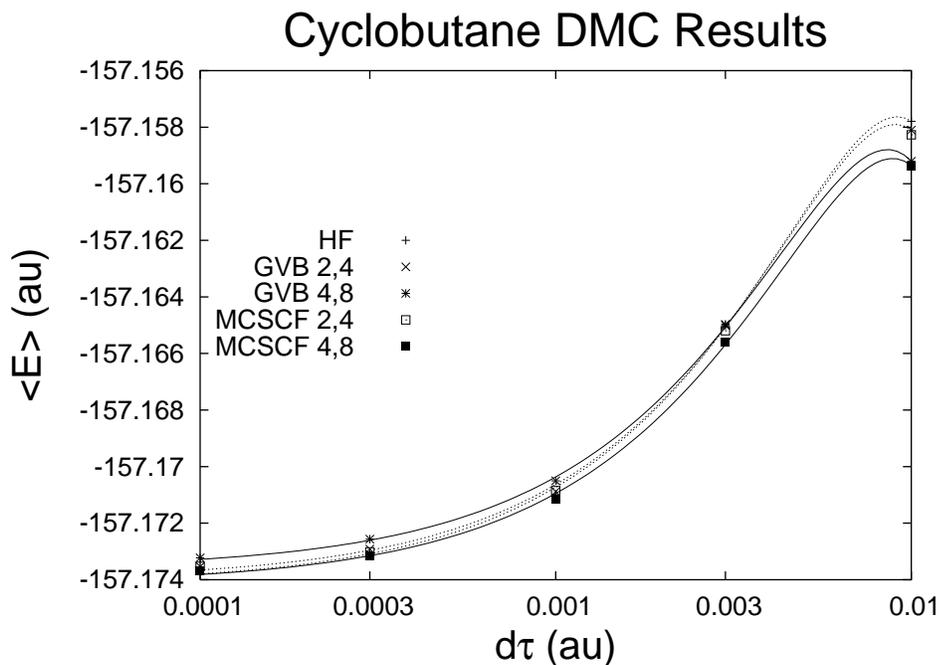


Figure 6.4: DMC expectation values for the energy of cyclobutane. The expectation values are fit to the function  $a + b(d\tau) + c(d\tau)^2$ . The energies for all five trial functions are very similar at every time step. Correlated trial functions do not significantly change the DMC energy for cyclobutane.

### 6.5.2 Reaction 2

The DMC results for butadiene are shown in figure 6.6. The DMC/GVB 2,4 energies are about 4 kcal/mol below the HF values and about 1 kcal/mol above the DMC/MCSCF 2,4 energies. Figure 6.7 shows the DMC results for TS2. The DMC/GVB 3,6 energies are about 1.25 kcal/mol below the DMC/HF energies and about 6.25 kcal/mol above the DMC/MCSCF 3,6 energies. In figure 6.8, the DMC/GVB 3,6 and DMC/MCSCF 3,6 energies for cyclohexene are almost identical, and the DMC/HF energies are about 2.5 kcal/mol higher. In these three figures, we see that for cyclohexene, a molecule with one C-C  $\pi$  bond, the GVB and MCSCF trial functions have almost identical results, while for the transition state, in which three  $\pi$  bonds are being broken and two  $\sigma$  bonds and one  $\pi$  bond are being formed, the DMC/MCSCF 3,6 energy is significantly lower than the DMC/GVB 3,6 energy. The nodal error for the GVB 2,4 trial function for butadiene is about 1 kcal/mol compared to the MCSCF 2,4 trial

function.

The enthalpies of activation and reaction for reaction 2 are shown in figure 6.9. The results for each time step and trial function are plotted with points, and the results for each trial function extrapolated to  $d\tau = 0$  are plotted with dotted lines. The experimental and CBS-QB3 values are plotted with solid lines. For the overall enthalpy, the experimental, CBS-QB3, and correlated DMC results are all within about 2 kcal/mol of each other. The DMC/HF results are about 4 kcal/mol too low. The DMC/MCSCF 3,6 results are about 2 kcal/mol higher than the experimental value, but only about 1 kcal/mol higher than the CBS-QB3 result. For the activation enthalpy, the DMC/MCSCF 3,6 and DMC/HF results are within about 0.5 kcal/mol of each other, and within about 2 kcal/mol of experiment. The DMC/GVB 3,6 result is about 5 kcal/mol higher than the experimental value. The CBS-QB3 result is about 2 kcal/mol lower than the experimental value.

### 6.5.3 Reaction 3

Figure 6.10 shows the DMC results for cyclobutene. For this molecule, the DMC/GVB 2,4 and DMC/MCSCF 2,4 energies are almost identical for every time step, while the DMC/HF energies are about 2.5 kcal/mol higher. The results for TS3 are shown in figure 6.11. For this transition state, the DMC/GVB 2,4 energy is about 2.5 kcal/mol below the DMC/HF result and about 2.2 kcal/mol above the DMC/MCSCF 2,4 energy.

The results for the activation and overall enthalpies for reaction 3 are shown in figure 6.12. The DMC results for each trial function and time step are plotted with points, and the results for each time step extrapolated to  $d\tau = 0$  are plotted with dotted lines. The experimental and CBS-QB3 results are plotted with solid lines. In this case, the DMC/HF results for the overall change in enthalpy are within 0.5 kcal/mol of experiment, while the DMC/GVB 2,4 and DMC/MCSCF 2,4 results are about 2 and 3 kcal/mol too low, respectively. The correlated DMC results, however, are both within 1 kcal/mol of the CBS-QB3 value. For the activation en-

thalpy, the DMC/MCSCF 2,4 results are about 2 kcal/mol above the experimental and CBS-QB3 values. The DMC/GVB 2,4 and DMC/HF values are about 4.5 and 6.5 kcal/mol, respectively, above the experimental and CBS-QB3 values.

#### 6.5.4 Discussion

Conducting DMC calculations for several time steps and extrapolating the results to  $d\tau = 0$  greatly increases their computational expense. Although the DMC expectation values for the individual molecules have considerable time step errors, they cancel out when energy differences for reactions are calculated. In the enthalpies of activation and reaction for reactions 1, 2, and 3, the DMC results for every time step are within 1 kcal/mol of the extrapolated value. There is no consistent trend in the extrapolated expectation values compared to their values at finite time steps. Since extrapolating the energies to  $d\tau = 0$  does not significantly change the results for reactions, it is unnecessary when energy differences are being considered, and a single time step can be used. For the rest of this work, only the DMC results with  $d\tau = 0.01$  au will be considered.

Table 6.1 contains the SCF, VMC, and DMC expectation values for the energy of each trial function. The VMC and DMC calculations used a time step of 0.01 au. The percentage of the correlation energy recovered by the Jastrow function is remarkably constant for every trial function, at about 77%. This result indicates the role of the Jastrow function is very similar in all of the trial functions considered. The trial functions represent the electronic structure of chain and cyclical, saturated and unsaturated, stable hydrocarbons and transition states. The consistency of the ability of the Jastrow to recover correlation energy in these diverse molecules suggests it acts primarily within the atoms, and does not have much effect in the bonding regions. If the Jastrow only influences the electronic structure within the atoms, it may be possible to optimize a set of electron-nuclear correlation functions using atomic or simple molecular calculations, and use them without optimization in larger systems. If the number of parameters to be optimized were reduced to only those in

the electron-electron correlation functions and the CI expansion coefficients without sacrificing accuracy, optimization would be simplified and calculations could proceed more quickly to the DMC phase.

Table 6.2 contains the *ab initio*, QMC, and DFT results for the activation and overall enthalpy changes for reaction 1. The DMC/GVB and DMC/MCSCF results are all within about 1 kcal/mol of experiment. The B3LYP result is about 4 kcal/mol too high, while the M06 results are between 5.5 and 2.5 kcal/mol too low. XYG3 is the best of the DFT methods, with an error of -0.7 kcal/mol. The CBS-QB3 result is within 0.2 kcal/mol of experiment, while the error for CCSD(T)/6-31G\* is -1.7 kcal/mol.

The results for reaction 2 are shown in table 6.3. For the overall enthalpy change, the DMC/GVB 3,6 result is within about 0.1 kcal/mol of the experimental value. The DMC/MCSCF 3,6 result is about 1.3 kcal/mol higher than experiment, but is within 0.03 kcal/mol of the CBS-QB3 value. The DMC/HF enthalpy is about 4.3 kcal/mol below experiment, and the CCSD(T)/6-31G\* result is about 0.8 kcal/mol below experiment. The B3LYP result is about 9 kcal/mol higher than the experimental value, while M06, M06-2X, and M06-HF are 2 to 2.5 kcal/mol lower than experiment. The M06-L and XYG3 values are about 1 kcal/mol higher than than experiment.

While the DMC/GVB 3,6 result for the overall enthalpy change of reaction 2 agrees with experiment, the DMC/GVB 3,6 activation enthalpy is about 6.5 kcal/mol too high. The DMC/HF activation enthalpy is about 1 kcal/mol too high, and the DMC/MCSCF 3,6 result is about 1.5 kcal/mol too high. The CBS-QB3 activation enthalpy is about 2.1 kcal/mol below experiment, and the CCSD(T)/6-31G\* result is about 2.5 kcal/mol above experiment. The B3LYP activation enthalpy is about 2 kcal/mol higher than experiment, while the M06 results are 1.5 to 7.5 kcal/mol lower. The XYG3 activation enthalpy is about 0.6 kcal/mol above the experimental value.

Table 6.4 contains the results for reaction 3. The DMC/HF overall enthalpy change is about 0.1 kcal/mol lower than experiment. The DMC/GVB 2,4 and DMC/MCSCF 2,4 values are about 1.8 and 2.8 kcal/mol, respectively, lower than

experiment, but are both within 1 kcal/mol of the CBS-QB3 value. The CCSD(T)/6-31G\* enthalpy is about 1.2 kcal/mol below experiment. The B3LYP enthalpy change is about 4.5 kcal/mol lower than experiment, while the M06 results are all within about 1 kcal/mol of experiment. The XYG3 result is about 2.4 kcal/mol below the experimental value.

The DMC/MCSCF 2,4 activation enthalpy is about 2.5 kcal/mol higher than experiment, while the GVB 2,4 and HF values are about 4.5 and 6 kcal/mol higher than experiment, respectively. The B3LYP activation enthalpy is the closest of all methods to experiment, agreeing within 0.16 kcal/mol. The M06-HF activation enthalpy is about 0.5 kcal/mol higher than experiment, while the other M06 methods are 3 to 4 kcal/mol higher. The XYG3 activation enthalpy is about 1.4 kcal/mol above the experimental value.

Table 6.5 shows the differences from the experimental values for the DMC/MCSCF and DFT activation and overall enthalpies for reactions 1, 2, and 3. The result with the lowest error for each quantity is in bold font.

## 6.6 Conclusion

In summary, we were able to calculate  $\Delta H_{0K}$  and  $\Delta H_{0K}^\ddagger$  values to within experimental accuracy for three difficult pericyclic hydrocarbon reactions using B3LYP/6-311G\*\* geometries and zero point energies and DMC electronic energies. The DMC trial functions consisted of a two body Jastrow and an antisymmetric wavefunction constructed with HF, GVB, or MCSCF and the aug-cc-pVTZ basis set. A time step of 0.01 au was found to be acceptable, making extrapolation to  $d\tau = 0$  unnecessary.

Because of the formation and breaking of C-C  $\sigma$  and  $\pi$  bonds, Reactions 1, 2, and 3 are difficult cases for DFT. Compared to the experimental reaction enthalpies, B3LYP has errors of about 4, 9, and -4.5 kcal/mol, respectively. X3LYP performs somewhat better, with errors of about 3, 7, and -4 kcal/mol. The M06 family of functionals is usually more accurate than B3LYP, but no one of the four methods is consistently better than B3LYP or the most accurate among the M06 family. The

results for the new XYG3 functional are excellent, with errors of about -1, 1, and -2.5 kcal/mol.

DMC using MCSCF trial functions has errors of about 1, 1, and -3 kcal/mol for the three reactions. If the overall enthalpy for reaction 3 is compared to CBS-QB3 instead of experiment, the DMC/MCSCF error is about -1 kcal/mol. DMC/GVB 3,6 has the most accurate result for the overall enthalpy of reaction 2, and DMC/HF is the closest to experiment for reaction 3, but DMC/MCSCF is the only method that performs consistently for all three reactions. When the appropriate trial function is used, the errors of DMC for reactions 1, 2, and 3 are comparable to CBS-QB3 and CCSD(T)/6-31G\*. The favorable scaling and parallelizability of DMC, however, will allow it to be applied to much larger systems than the ones in this work. CBS-QB3 and CCSD(T) are limited to systems with less than about twelve heavy atoms.

Based on the DMC results for individual molecules, some principles emerge as to the trial functions necessary to give accurate results. First, HF is sufficient for C-C  $\sigma$  bonds. Defining GVB pairs for C-C  $\sigma$  bonds raises the DMC energy, and correlating them with MCSCF does not lower the DMC energy significantly. Second, C-C  $\pi$  bonds must be correlated to give accurate trial functions. If there is one  $\pi$  bond, a GVB trial function is sufficient. If the molecule contains more than one  $\pi$  bond, an MCSCF trial function must be used. For transition states, all partially formed bonds must be correlated. If more than one bond is being formed or broken, an MCSCF trial function is once again necessary.

The DMC time step errors for individual molecules are significant. For example, using a time step of 0.01 au raises the expectation value for the energy of ethylene by about 5 kcal/mol compared to the  $d\tau = 0$  limit. These errors tend to cancel out when energy differences for reactions are considered. For reactions 1, 2, and 3, extrapolating the results to  $d\tau = 0$  did not change them significantly from the  $d\tau = 0.01$  au values. Being able to use a single time step greatly decreases the amount of computer time needed to calculate enthalpy differences.

To further simplify DMC calculations and make these excellent results available for larger and more interesting systems, it will be helpful to investigate the effect of

the basis set on the expectation values. The most time consuming parts of the QMC algorithm scale as  $O(N^3)$  with the number of basis functions, so decreasing the size of the basis set will make the trial functions easier to construct and speed up the QMC calculations. In addition, the VMC and DMC results in table 6.1 suggest it may be possible to develop a single set of electron-nuclear correlation functions that could be applied to all molecules. Eliminating these parameters from the optimization phase of QMC calculations would make them less expensive and simpler for nonexperts to carry out.

Finally, the GVB and particularly the MCSCF trial functions in this work included large numbers of configurations. Increasing the length of the CI expansion greatly increases the time needed to evaluate and optimize a trial function. While it was shown that HF trial functions are not acceptable for transition states or molecules with C-C  $\pi$  bonds, it is likely that the CI expansions could be truncated after a fairly small number of terms without sacrificing accuracy. Comparing DMC expectation values for trial functions with different CI expansion lengths will give researchers guidelines to identify which configurations must be included, and which can safely be ignored.

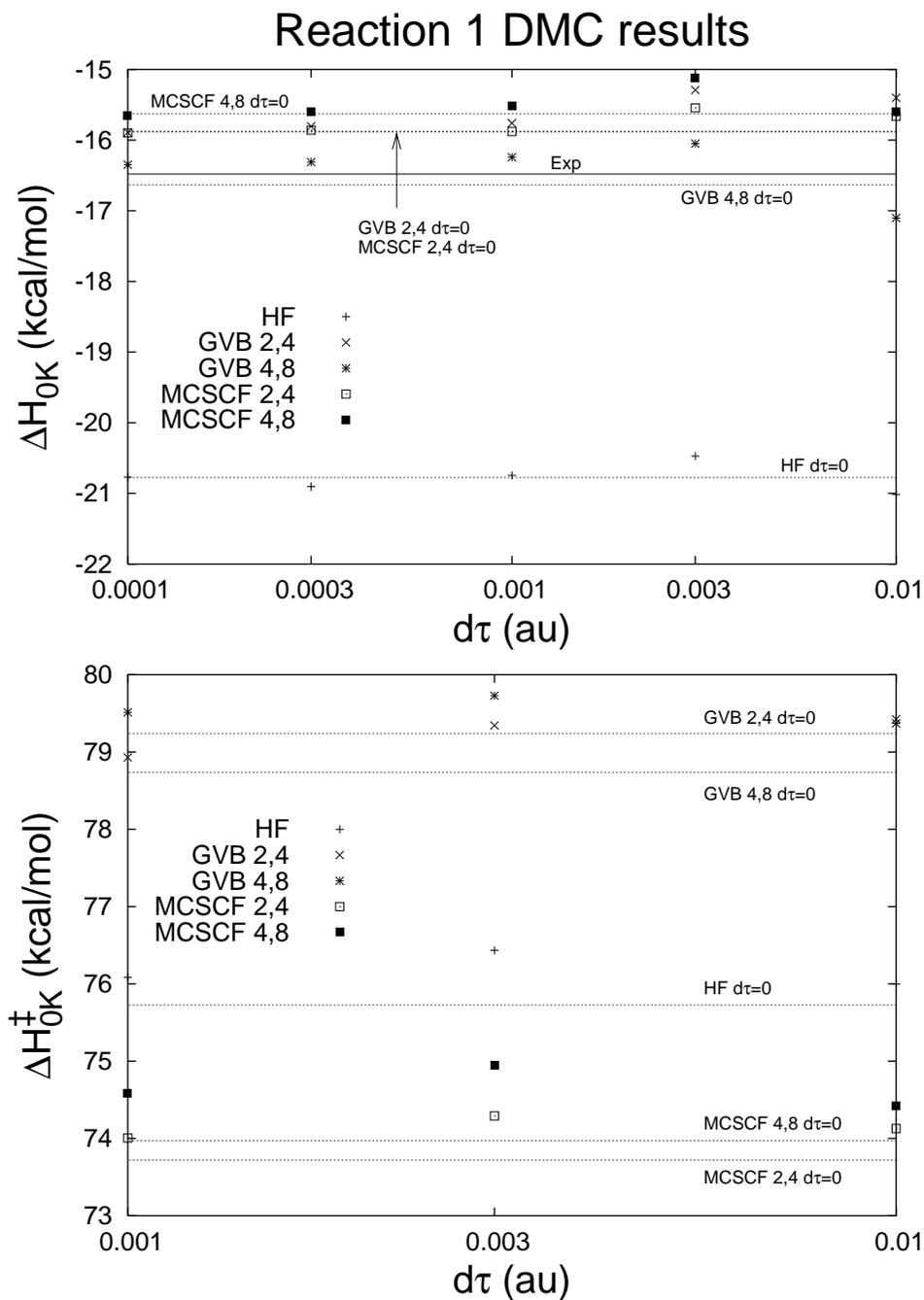


Figure 6.5: DMC results for the activation and overall enthalpies of reaction 1. The dotted lines are the values for each level of correlation extrapolated to  $d\tau = 0$ . All four correlated trial functions give overall DMC enthalpies within 1 kcal/mol of experiment, but the value for the HF trial function is about 4 kcal/mol too low.

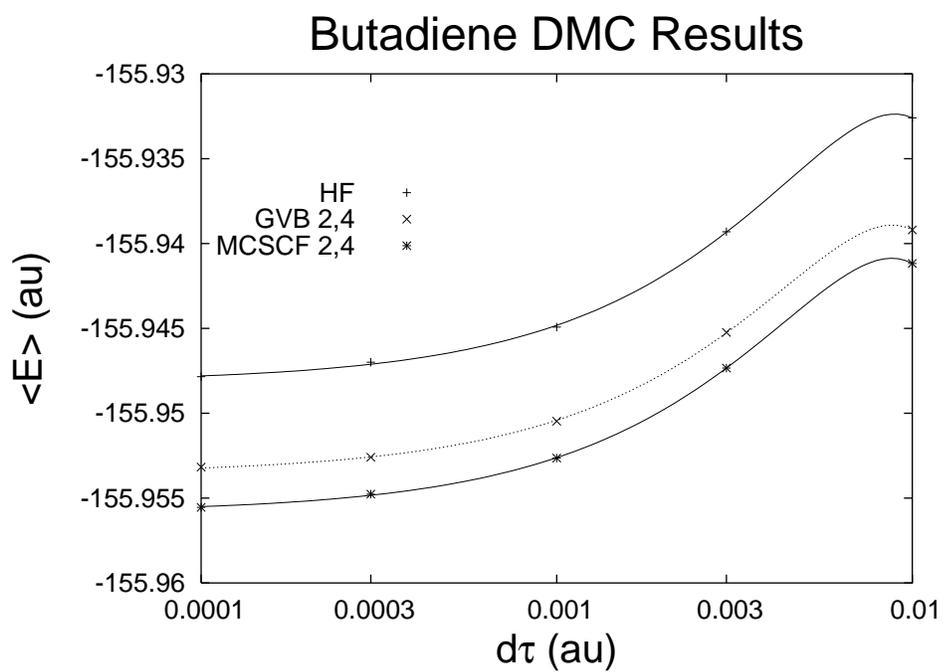


Figure 6.6: DMC expectation values for the energy of butadiene. The expectation values are fit to the function  $a + b(d\tau) + c(d\tau)^2$ . In this case, increasing the correlation of the trial function lowers the DMC energy for every time step.

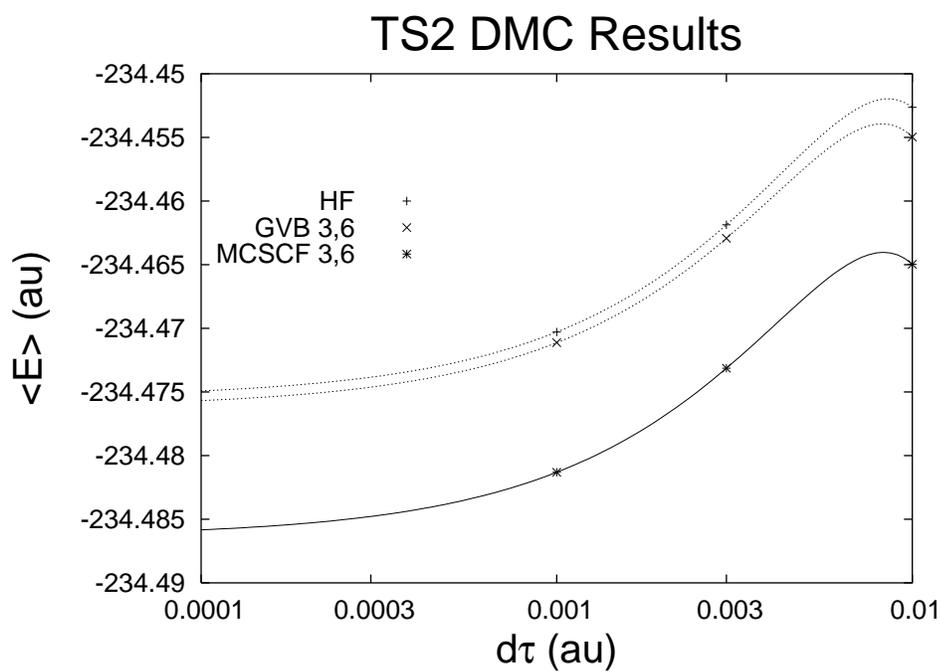


Figure 6.7: DMC expectation values for the energy of TS2. The expectation values are fit to the function  $a + b(d\tau) + c(d\tau)^2$ . The DMC/HF and DMC/GVB 3,6 DMC energies are very close for every time step, but the DMC/MCSCF 3,6 expectation values are significantly lower.

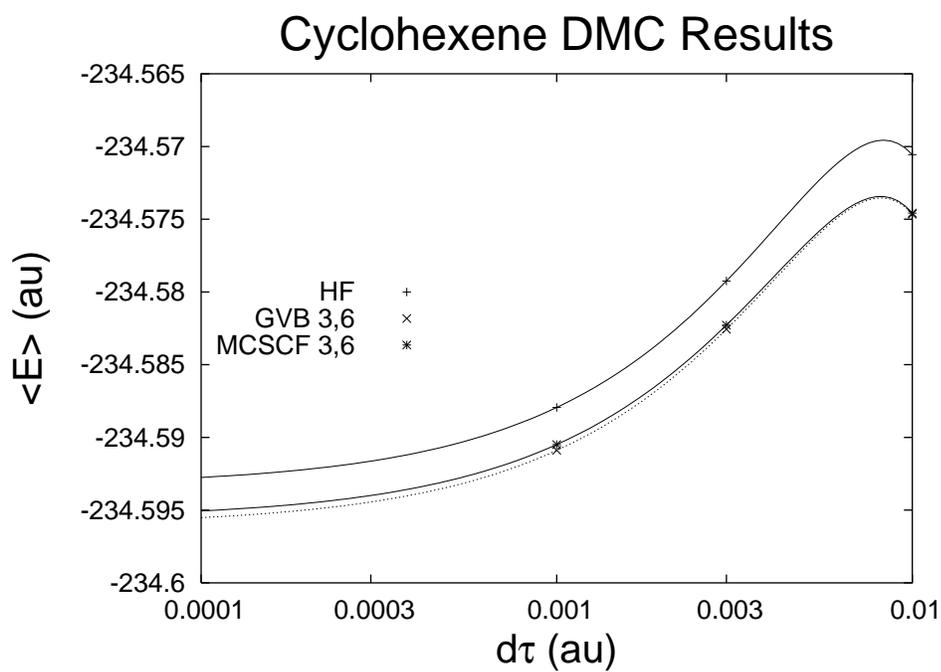


Figure 6.8: DMC expectation values for the energy of cyclohexene. The expectation values are fit to the function  $a+b(d\tau)+c(d\tau)^2$ . In contrast to TS2, the DMC/GVB 3,6 and DMC/MCSCF 3,6 expectation values are similar, while the HF expectation values are higher.

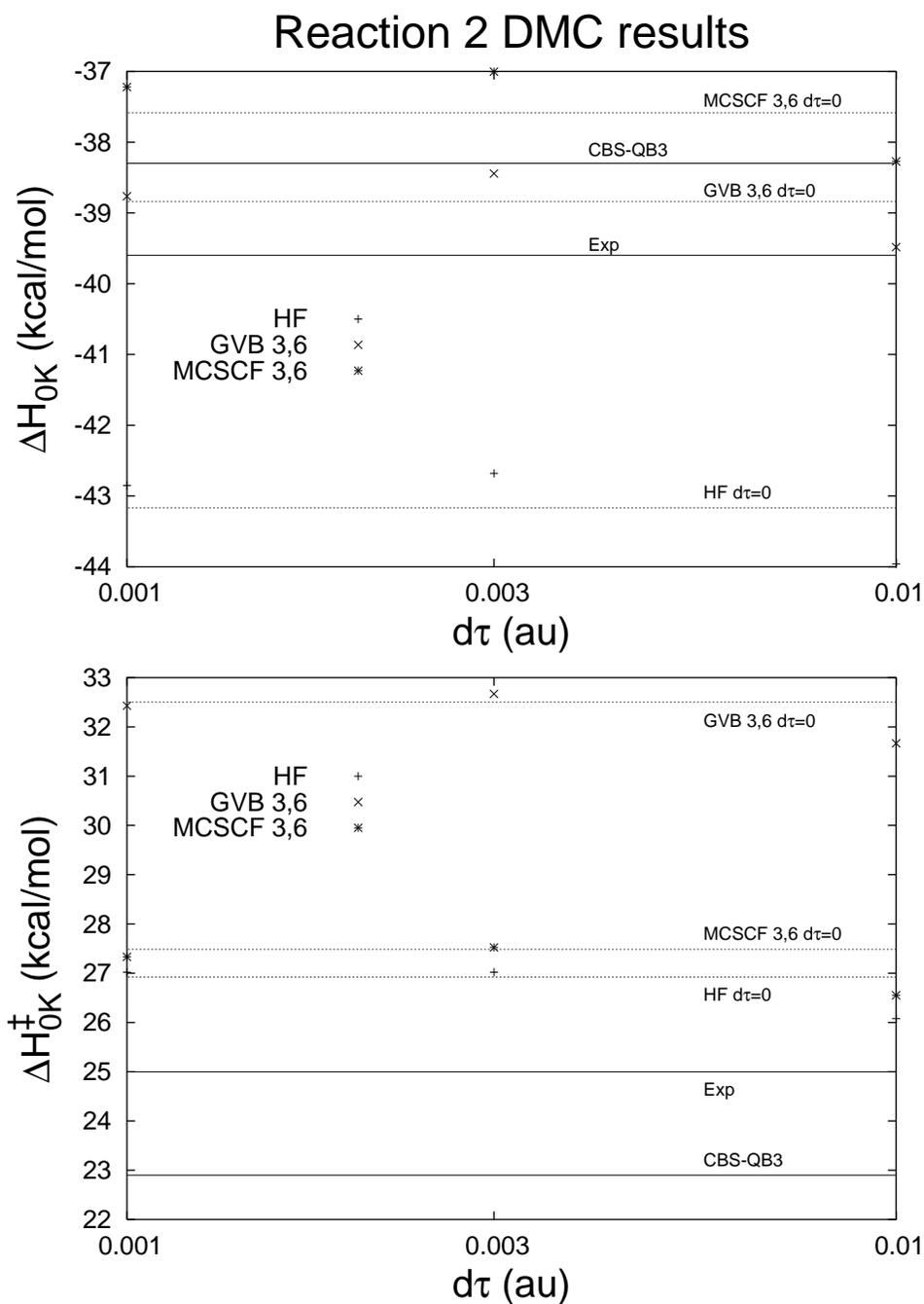


Figure 6.9: DMC results for the activation and overall enthalpies of reaction 2. The dotted lines are the values for each level of correlation extrapolated to  $d\tau = 0$ , and the solid lines are the experimental and CBS-QB3 results. The DMC/GVB 3,6 result is within 1 kcal/mol of experiment for the overall enthalpy change, but about 7.5 kcal/mol too high for the activation barrier.

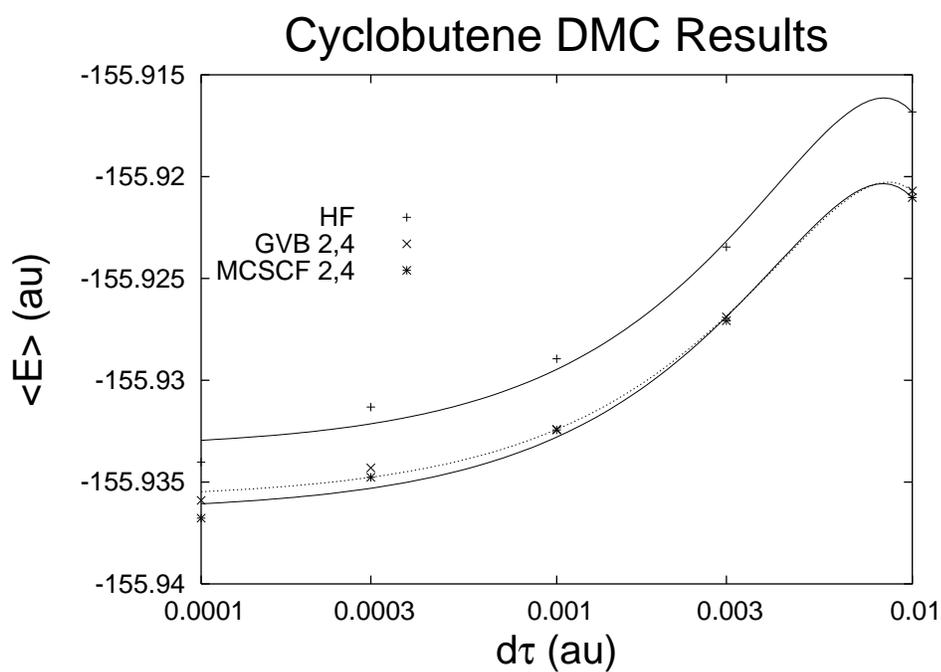


Figure 6.10: DMC expectation values for the energy of cyclobutene. The expectation values are fit to the function  $a + b(d\tau) + c(d\tau)^2$ . The expectation values for the GVB 2,4 and MCSCF 2,4 trial functions are almost identical for every time step, while the DMC/HF expectation values are higher.

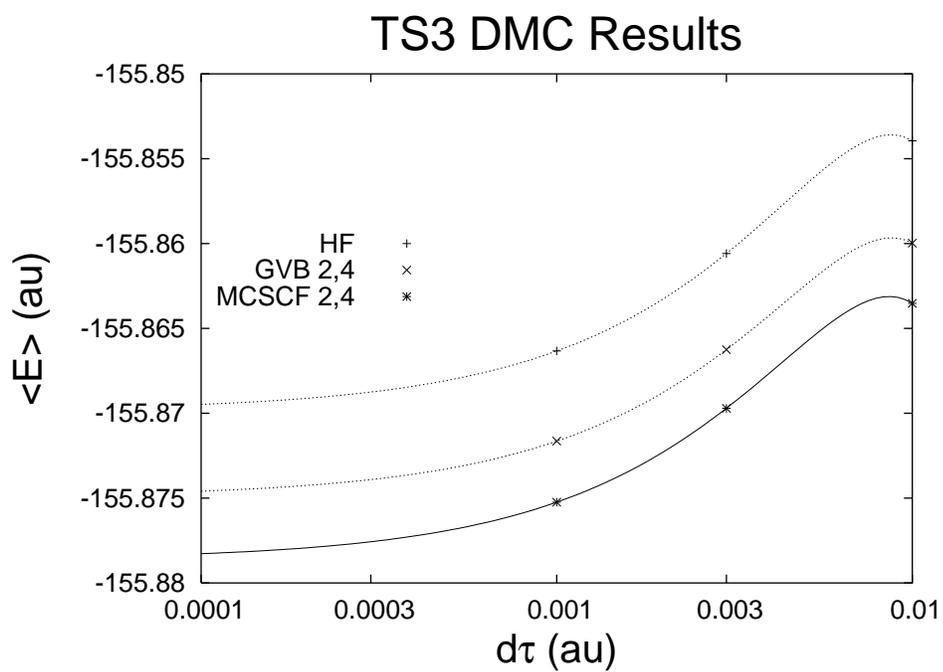


Figure 6.11: DMC expectation values for the energy of TS3. The expectation values are fit to the function  $a + b(d\tau) + c(d\tau)^2$ . Increasing the correlation of the trial function lowers the DMC expectation value for every time step.

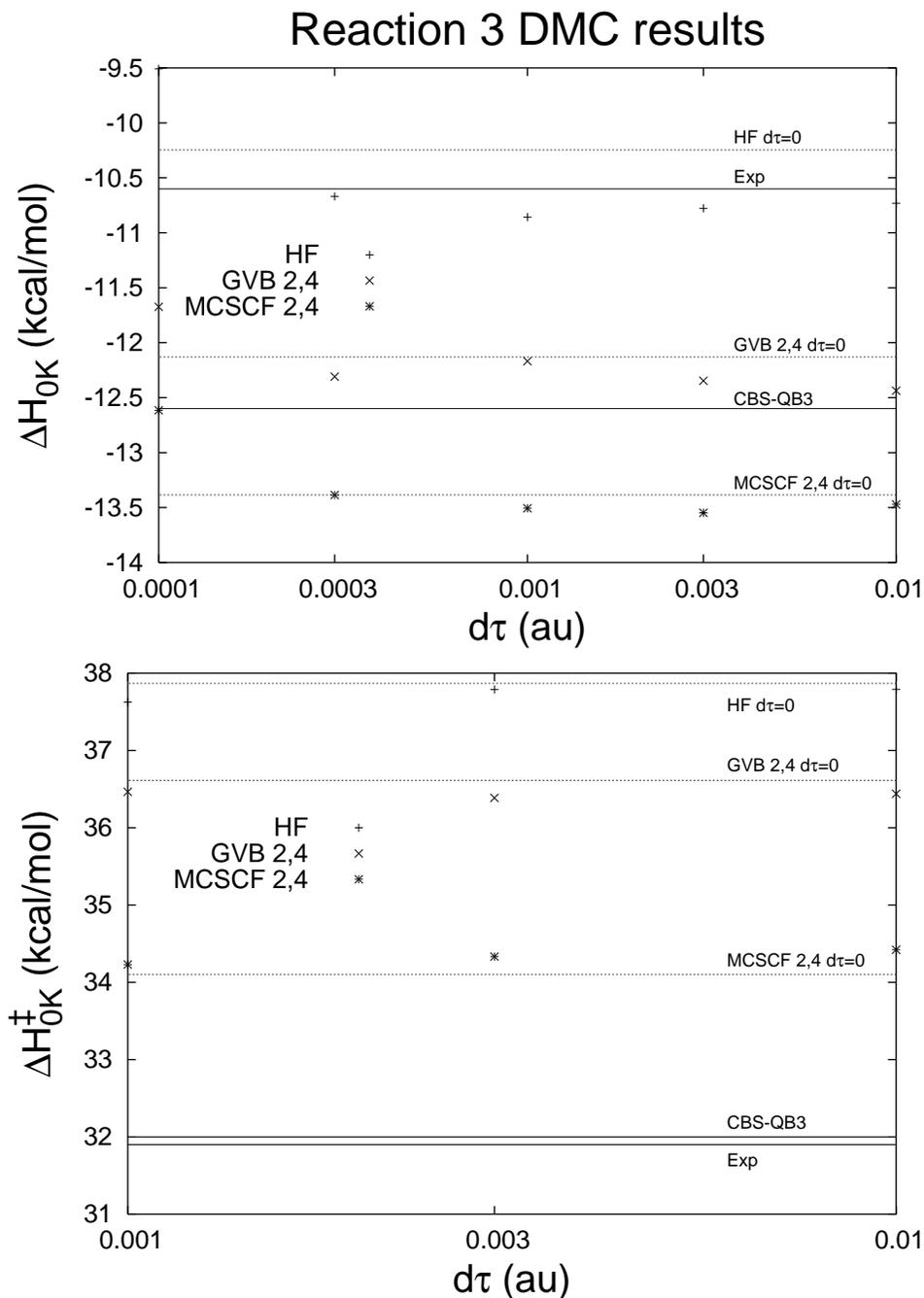


Figure 6.12: DMC results for the activation and overall enthalpies of reaction 3. The dotted lines are the values for each level of correlation extrapolated to  $d\tau = 0$ , and the solid lines are the experimental and CBS-QB3 results. The DMC/HF result is within 0.5 kcal/mol of experiment for the overall enthalpy, but off by about 6 kcal/mol for the activation barrier. The correlated trial functions' results are closer to the CBS-QB3 result than the experimental value for the overall enthalpy.

Molecule	SCF wavefunction	$N_{det}$	$\langle E \rangle_{SCF}$	Params	$\langle E \rangle_{VMC}$	$\langle E \rangle_{DMC}$	% $E_{corr}$
Ethylene	HF	1	-78.0652	36	-78.4461	-78.5578	77.3
	GVB 1,2	2	-78.0921	37	-78.4574	-78.5624	77.7
	GVB 2,4	4	-78.1027	39	-78.4589	-78.5616	77.6
	MCSCF 1,2	2	-78.0921	37	-78.4574	-78.5623	77.7
	MCSCF 2,4	12	-78.1178	45	-78.4629	-78.5629	77.5
TS1	HF	1	-155.9731	36	-156.7596	-156.9970	76.8
	GVB 2,4	4	-156.0219	39	-156.7774	-157.0008	77.2
	GVB 4,8	16	-156.0467	51	-156.7809	-156.9993	77.1
	MCSCF 2,4	16	-156.0413	47	-156.7902	-157.0090	77.4
	MCSCF 4,8	61	-156.0885	72	-156.7994	-157.0097	77.2
Cyclobutane	HF	1	-156.1515	36	-156.9323	-157.1578	77.6
	GVB 2,4	4	-156.1849	39	-156.9411	-157.1581	77.7
	GVB 4,8	16	-156.2171	41	-156.9475	-157.1592	77.5
	MCSCF 2,4	12	-156.1855	45	-156.9400	-157.1583	77.6
	MCSCF 4,8	75	-156.2325	77	-156.9498	-157.1594	77.4
Butadiene	HF	1	-154.9799	36	-155.7125	-155.9326	76.9
	GVB 2,4	4	-155.0288	39	-155.7342	-155.9392	77.5
	MCSCF 2,4	16	-155.0332	47	-155.7367	-155.9412	77.5
TS2	HF	1	-232.9672	36	-234.1120	-234.4526	77.1
	GVB 3,6	8	-233.0269	43	-234.1251	-234.4550	76.9
	MCSCF 3,6	51	-233.0510	68	-234.1425	-234.4650	77.2
Cyclohexene	HF	1	-233.1005	36	-234.2376	-234.5706	77.3
	GVB 3,6	8	-233.1578	42	-234.2522	-234.5747	77.2
	MCSCF 3,6	50	-233.1586	67	-234.2530	-234.5746	77.3
Cyclobutene	HF	1	-154.9551	36	-155.6979	-155.9168	77.2
	GVB 2,4	4	-154.9989	39	-155.7123	-155.9207	77.4
	MCSCF 2,4	8	-154.9997	40	-155.7132	-155.9210	77.4
TS3	HF	1	-154.8840	36	-155.6325	-155.8539	77.2
	GVB 2,4	4	-154.9340	39	-155.6518	-155.8600	77.5
	MCSCF 2,4	20	-154.9423	49	-155.6553	-155.8635	77.5

Table 6.1: QMC results using HF, GVB, and MCSCF trial functions. Energies reported in atomic units. All calculations used the aug-cc-pVTZ basis set and B3LYP/6-311G\*\* geometries. QMC calculations used a 0.01 au time step and the 2 body Jastrow described in Eq 6.14, and were run until the energy converged to within  $1.5 \times 10^{-4}$  au. The number of parameters optimized in the Jastrow function and CI expansion for each trial function is shown in the Params column. The percent of the correlation energy recovered by the Jastrow was calculated using the formula  $\% E_{corr} = 100 \frac{\langle E \rangle_{VMC} - \langle E \rangle_{DMC}}{\langle E \rangle_{SCF} - \langle E \rangle_{DMC}}$ . The percent of correlation energy recovered by the Jastrow is remarkably constant across all trial functions, which suggests the Jastrow acts mostly within the atoms.

Geom	Energy	ZPE	$\Delta H_{0K}^\ddagger$	$\Delta H_{0K}$
Exp [94, 95]				-16.48
MP2//CAS/6-311G(d,p) [96]			77.6	
CBS-QB3 [98]				-16.66
CCSD(T)/6-31G* [110]				-18.18
B3LYP	B3LYP	B3LYP	74.37	-12.55
	HF		100.18	-7.85
	GVB 2,4		103.41	5.01
	GVB 4,8		101.14	-1.84
	MCSCF 2,4		91.21	4.66
	MCSCF 4,8		93.81	7.37
	DMC/HF		76.00	-21.02
	DMC/GVB 2,4		79.42	-15.40
	DMC/GVB 4,8		79.37	-17.10
	DMC/MCSCF 2,4		74.13	-15.66
	DMC/MCSCF 4,8		74.42	-15.61
	M06		70.63	-22.02
	M06-2X		71.78	-21.11
	M06-HF		69.37	-18.94
	M06-L		70.58	-20.94
X3LYP	X3LYP	X3LYP	73.89	-13.79
B3LYP	XYG3	B3LYP	73.21	-17.21

Table 6.2: SCF and DMC results for the enthalpies of activation and reaction at 0K for reaction 1. All enthalpy differences reported in kcal/mol. The geometry, energy, and ZPE columns contain the methods used to optimize geometries, calculate electronic energies, and calculate zero point energies, respectively. B3LYP, M06, and X3LYP calculations used the 6-311G\*\* basis set. The XYG3 calculations used B3LYP/6-311+G(d,p) geometries and frequencies scaled by 0.9877 with XYG3/6-311+G(3df,2p) electronic energies [111]. HF, GVB, and MCSCF calculations used the aug-cc-pVTZ basis set. DMC calculations used a 0.01 au time step.

Geometry	Energy	ZPE	$\Delta H_{0K}^\ddagger$	$\Delta H_{0K}$
Exp [99]			25.0	-39.6
CBS-QB3 [99]			22.9	-38.3
CCSD(T)/6-31G* [112, 113]		B3LYP	27.48	-40.36
B3LYP	B3LYP	B3LYP	27.10	-30.80
	HF		51.24	-28.48
	GVB 3,6		61.41	-16.77
	MCSCF 3,6		49.06	-14.50
	DMC/HF		26.08	-43.96
	DMC/GVB 3,6		31.67	-39.48
	DMC/MCSCF 3,6		26.55	-38.27
	M06		23.38	-41.48
	M06-2X		21.60	-42.19
	M06-HF		17.26	-42.21
	M06-L		22.36	-38.56
X3LYP	X3LYP	X3LYP	26.34	-32.37
B3LYP	XYG3	B3LYP	24.44	-38.55

Table 6.3: SCF and DMC results for the enthalpies of activation and reaction at 0K for reaction 2. All enthalpy differences reported in kcal/mol. The geometry, energy, and ZPE columns contain the methods used to optimize geometries, calculate electronic energies, and calculate zero point energies, respectively. B3LYP, M06, and X3LYP calculations used the 6-311G\*\* basis set. The XYG3 calculations used B3LYP/6-311+G(d,p) geometries and frequencies scaled by 0.9877 with XYG3/6-311+G(3df,2p) electronic energies [111]. HF, GVB, and MCSCF calculations used the aug-cc-pVTZ basis set. DMC calculations used a 0.01 au time step.

Geometry	Energy	ZPE	$\Delta H_{0K}^\ddagger$	$\Delta H_{0K}$
Exp [99]			31.9	-10.6
CBS-QB3 [99]			32.0	-12.6
CCSD(T)/6-31G* [110]				-11.83
B3LYP	B3LYP	B3LYP	32.15	-15.11
	HF		42.94	-16.39
	GVB 2,4		39.07	-19.61
	MCSCF 2,4		34.37	-21.87
	DMC/HF		37.79	-10.73
	DMC/GVB 2,4		36.44	-12.44
	DMC/MCSCF 2,4		34.42	-13.47
	M06		35.24	-10.27
	M06-2X		35.37	-10.91
	M06-HF		32.54	-11.68
	M06-L		36.07	-10.14
X3LYP	X3LYP	X3LYP	32.47	-14.87
B3LYP	XYG3	B3LYP	33.73	-12.98

Table 6.4: SCF and DMC results for the enthalpies of activation and reaction at 0K for reaction 3. All enthalpy differences reported in kcal/mol. The geometry, energy, and ZPE columns contain the methods used to optimize geometries, calculate electronic energies, and calculate zero point energies, respectively. B3LYP, M06, and X3LYP calculations used the 6-311G\*\* basis set. The XYG3 calculations used B3LYP/6-311+G(d,p) geometries and frequencies scaled by 0.9877 with XYG3/6-311+G(3df,2p) electronic energies [111]. HF, GVB, and MCSCF calculations used the aug-cc-pVTZ basis set. DMC calculations used a 0.01 au time step.

Method	Reaction 1		Reaction 2		Reaction 3	
	$\Delta H_{0K}^\ddagger$	$\Delta H_{0K}$	$\Delta H_{0K}^\ddagger$	$\Delta H_{0K}$	$\Delta H_{0K}^\ddagger$	$\Delta H_{0K}$
B3LYP		3.93	2.10	8.80	<b>0.25</b>	-4.51
X3LYP		2.69	1.34	7.23	0.56	-4.27
XYG3		<b>-0.73</b>	<b>0.56</b>	1.05	1.83	-2.38
M06		-5.54	-1.62	-1.88	3.34	0.33
M06-2X		-4.63	-3.40	-2.59	3.47	<b>-0.31</b>
M06-HF		-2.46	-7.74	-2.61	0.64	-1.08
M06-L		-4.46	-2.64	<b>1.04</b>	4.17	0.46
DMC/MCSCF		0.82	1.55	1.32	2.52	-2.87

Table 6.5: Differences from experiment for the DMC/MCSCF and DFT activation and overall enthalpies for reactions 1, 2, and 3. The result with the lowest error for each quantity is in bold font.

# Chapter 7

## Conclusion

The QMC calculations presented in this thesis were carried out using QMcBeaver, a program written in the Goddard group to develop, demonstrate, and apply new QMC algorithms [81]. Quantum Monte Carlo has the potential to calculate expectation values to within experimental accuracy, and its favorable scaling and parallelizability will allow it to be applied to much larger systems than comparably accurate traditional electronic structure methods.

QMC is a relatively new class of methods, with new algorithms being developed by a small number of experts, most of whom have written their own QMC programs. Most applications of QMC to chemical and materials systems have been done by these developers to demonstrate their algorithms, not by researchers interested in the systems themselves. The computational expense and theoretical complexity of QMC have kept it from becoming a “black box” method for nonexperts to use.

Algorithms developed using QMcBeaver have made progress in bringing QMC to nonexpert users by providing simple, automatic tools for setting up and carrying out calculations. The Dynamic Distributed Decorrelation Algorithm (DDDA) automatically calculates the standard deviation of expectation values during a QMC calculation, taking the serial correlation of the samples into account, while greatly reducing the amount of data that has to be communicated among the processors to gather results [49]. The manager-worker parallelization (QMC-MW) automatically balances the work between processors running at different speeds, allowing the efficient use of heterogeneous computers. QMC-MW also makes it possible to terminate

a calculation based on the convergence of the expectation values rather than the completion of a certain number of iterations [50].

In chapter 5 of this work, we demonstrated the importance of initial walker configurations to the efficiency and accuracy of QMC calculations. STRAW is a simple, automatic method that requires no user input to generate statistically independent initial walker configurations in regions of high density and low local energy [109]. Avoiding contamination by configurations that do not represent the desired density ensures the accuracy of the results and allows the efficient use of large numbers of processors. DDDA, QMC-MW, and STRAW combine to make it straightforward to set up QMC calculations that will efficiently use the next generation of homogeneous supercomputers, inexpensive heterogeneous beowulf clusters, and distributed computing resources.

The two sources of error in a DMC calculation are the time step and the nodal structure of the SCF part of the trial function. In order to have confidence in the results of their calculations, researchers need guidelines as to the appropriate time steps and trial functions to use for the functional groups in their systems. In chapter 6, we explored the time step and nodal errors for three pericyclic hydrocarbon reactions. DMC results calculated with HF, GVB, and MCSCF trial functions were compared to experiment, high quality *ab initio* calculations, and the recently introduced X3LYP, M06, and XYG3 DFT functionals. From the results, it was determined that the time step error cancels out when energy differences are considered, making extrapolation to zero time step unnecessary. HF trial functions were shown to be acceptable for C-C  $\sigma$  bonds, but to have a large nodal error for  $\pi$  bonds. GVB trial functions are sufficient for molecules with one  $\pi$  bond, while MCSCF wavefunctions must be used for molecules with multiple  $\pi$  bonds and transition states with several bonds being broken and formed. When the appropriate trial function is used, DMC results are consistently as accurate as CCSD(T) and CBS-QB3 for the three hydrocarbon reactions.

In order to allow researchers to construct trial functions for larger and more complicated molecules, the nodal errors for more functional groups must be investigated.

Carrying out DMC calculations with a variety of trial functions for small model systems, for which experimental and high quality *ab initio* results are available for comparison, can provide this information.

All of the trial functions in chapter 6 used the aug-cc-pVTZ basis set, which is probably larger than necessary to give accurate DMC results. In addition, the MCSCF trial functions used long CI expansions. A systematic comparison of the results and rate of convergence of DMC calculations carried out with different basis sets and CI expansion lengths will allow researchers to determine which basis functions can be eliminated and where CI expansions can be truncated without sacrificing accuracy. Since each determinant must be inverted to evaluate the trial function value and matrix inversion scales as  $O(N^3)$  with the number of basis functions, using smaller basis sets and shorter CI expansions will greatly reduce the computational expense of DMC calculations.

Finally, the results of chapter 6 suggest the possibility of developing a set of electron-nuclear Jastrow parameters to be used for all molecules without reoptimization. If this “generic Jastrow” could be used, only the electron-electron and CI expansion parameters would have to be optimized for each system, greatly reducing the complexity and expense of the parameter optimization phase of QMC calculations.

Quantum Monte Carlo has the potential to become a very important tool for computational scientists. The high accuracy of QMC combined with its ability to efficiently use the next generation of computational resources will allow it to provide accurate expectation values to understand reaction mechanisms and train density functional and force field methods. The continuing development of algorithms to make QMC more accurate, straightforward, and efficient will bring it into common use among researchers in chemistry and materials science.

# Bibliography

- [1] D. J. Griffiths. *Introduction to Quantum Mechanics*. Prentice Hall, Upper Saddle River, NJ, 1995.
- [2] C. Cohen-Tannoudji, B. Diu, and F. Laloe. *Quantum Mechanics*. Wiley-Interscience, Paris, France, 1977.
- [3] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry*. Dover, Mineola, NY, 1996.
- [4] J. P. Lowe. *Quantum Chemistry*. Academic Press, London, UK, 1993.
- [5] J. M. Thijssen. *Computational Physics*. Cambridge University Press, Cambridge, UK, 2001.
- [6] D. A. McQuarrie. *Quantum Chemistry*. University Science Books, Sausalito, CA, 1983.
- [7] J. B. Foresman and A. Frisch. *Exploring Chemistry with Electronic Structure Methods*. Gaussian, Wallingford, CT, 1996.
- [8] E. Schrödinger. Quantisierung als eigenwertproblem. *Annalen der Physik*, 384:361, 1926.
- [9] R. T. Pack and W. B. Brown. Cusp conditions for molecular wavefunctions. *The Journal of Chemical Physics*, 45:556, 1966.
- [10] M. Born and J. R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389:457, 1927.

- [11] J. C. Slater. The theory of complex spectra. *Physical Review*, 34(10):1293, 1929.
- [12] H. Kümmel. Origins of the coupled cluster method. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 80:81, 1991.
- [13] M. W. Schmidt and M. S. Gordon. The construction and interpretation of MCSCF wavefunctions. *Annual Reviews of Physical Chemistry*, 49:233, 1998.
- [14] B. O. Roos and P. R. Taylor. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chemical Physics*, 48(2):157, 1980.
- [15] L. M. Cheung, K. R. Sundberg, and K. Ruedenberg. Electronic rearrangements during chemical reactions. II. Planar dissociation of ethylene. *International Journal of Quantum Chemistry*, 16:1103, 1979.
- [16] W. A. Goddard III and R. C. Ladner. A generalized orbital description of the reactions of small molecules. *Journal of the American Chemical Society*, 93:6750, 1971.
- [17] P. M. Kozłowski and E. R. Davidson. Consideration in constructing a multireference second-order perturbation theory. *Journal of Chemical Physics*, 100(5):3672, 1994.
- [18] M. L. Leininger, W. D. Allen, H. F. Schaefer, and C. D. Sherrill. Is Møller-Plesset perturbation theory a convergent ab initio method? *Journal of Chemical Physics*, 112(21):9213, 2000.
- [19] J. W. Ochterski, G. A. Petersson, and J. A. Montgomery Jr. A complete basis set model chemistry. V. Extensions to six or more heavy atoms. *Journal of Chemical Physics*, 104(7):2598, 1996.

- [20] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss. Gaussian-1 theory: A general procedure for prediction of molecular energies. *Journal of Chemical Physics*, 90(10):5622, 1989.
- [21] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople. Gaussian-2 theory for molecular energies of first- and second-row compounds. *Journal of Chemical Physics*, 94(11):7221, 1991.
- [22] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassellov, and J. A. Pople. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *Journal of Chemical Physics*, 109(18):7764, 1998.
- [23] L. A. Curtiss, P. C. Redfern, and K. Raghavachari. Gaussian-4 theory. *Journal of Chemical Physics*, 126(8):084108, 2007.
- [24] A. G. Császár, W. D. Allen, and H. F. Schaefer III. In pursuit of the ab initio limit for conformational energy prototypes. *Journal of Chemical Physics*, 108(23):9751, 1998.
- [25] W. Kutzelnigg. The adiabatic approximation. I. the physical background of the Born-Handy ansatz. *Molecular Physics*, 90(6):909, 1997.
- [26] G. Tarczay, A. G. Császár, W. Klopper, and H. M. Quincy. Anatomy of relativistic energy corrections in light molecular systems. *Molecular Physics*, 99(21):1769, 2001.
- [27] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136:B864, 1964.
- [28] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140:A1133, 1965.
- [29] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: A critical analysis. *Canadian Journal of Physics*, 58:1200, 1980.

- [30] A. D. Becke. Density-functional exchange approximation with correct asymptotic behavior. *Physical Review A*, 38:3098, 1988.
- [31] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37:785, 1988.
- [32] A. D. Becke. A new mixing of Hartree-Fock and local density-functional theories. *Journal of Chemical Physics*, 98:1372, 1993.
- [33] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *Journal of Chemical Physics*, 98:5648, 1993.
- [34] X. Xu and W. A. Goddard III. The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proceedings of the National Academy of Sciences*, 101(9):2673, 2004.
- [35] Y. Zhao and D. G. Truhlar. 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 functionals. *Theoretical Chemistry Accounts*, 120:215–241, 2007.
- [36] Y. Zhang, X. Xu, and W. A. Goddard III. Doubly hybrid density functional for accurate descriptions of nonbond interactions, thermochemistry, and thermochemical kinetics. *Proceedings of the National Academy of Sciences*, 106(13):4963, 2009.
- [37] G. E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8):4, 1965.
- [38] A. Grama, A. Gupta, G. Karypis, and V. Kumar. *Introduction to Parallel Computing*. Pearson Education, Essex, UK, 2003.

- [39] G. Amdahl. Validity of the single processor approach to achieving large-scale computing capabilities. *AFIPS Conference Proceedings*, 30:483–485, 1967.
- [40] Top 500 supercomputer sites. <http://www.top500.org>, June 1993. Dec 15, 2009.
- [41] Advanced simulation and computing program. <http://www.lanl.gov/ASC>. Dec 15, 2009.
- [42] Blue gene. <http://www.research.ibm.com/bluegene>. Dec 15, 2009.
- [43] Roadrunner. <http://www.lanl.gov/roadrunner>. Dec 15, 2009.
- [44] Sequoia. [http://asc.llnl.gov/computing\\_resources/sequoia](http://asc.llnl.gov/computing_resources/sequoia). Dec 15, 2009.
- [45] Beowulf.org. <http://www.beowulf.org>. Dec 15, 2009.
- [46] Seti@home. <http://setiathome.berkeley.edu>. Dec 15, 2009.
- [47] Folding@home distributed computing. <http://folding.stanford.edu>. Dec 15, 2009.
- [48] Open-source software for volunteer computing and grid computing. <http://www.boinc.berkeley.edu>. Dec 15, 2009.
- [49] D. R. Kent IV, R. P. Muller, A. G. Anderson, W. A. Goddard III, and M. T. Feldmann. Efficient algorithm for ‘on-the-fly’ error analysis of local or distributed serially-correlated data. *Journal of Computational Chemistry*, 28(14):2309, 2007.
- [50] M. T. Feldmann, J. C. Cummings, D. R. Kent IV, R. P. Muller, and W. A. Goddard III. Manager-worker-based model for the parallelization of Quantum Monte Carlo on heterogeneous and homogeneous networks. *Journal of Computational Chemistry*, 29(1):8, 2008.
- [51] John Walker. Hotbits: Genuine random numbers, generated by radioactive decay. <http://www.fourmilab.ch/hotbits>, May 1996. Dec 15, 2009.

- [52] Mads Haahr. True random number service. <http://www.random.org>, 1998. Dec 15, 2009.
- [53] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C: The art of scientific computing*. Cambridge University Press, Cambridge, UK, 2002. [www.nr.com](http://www.nr.com).
- [54] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087, 1953.
- [55] C. J. Umrigar, M. P. Nightengale, and K. J. Runge. A diffusion Monte Carlo algorithm with very small time-step errors. *The Journal of Chemical Physics*, 99(4):2865, 1993.
- [56] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum Monte Carlo simulations of solids. *Reviews of Modern Physics*, 73(1):33, 2001.
- [57] J. C. Grossman. Benchmark Quantum Monte Carlo calculations. *Journal of Chemical Physics*, 117(4):1434, 2002.
- [58] S. I. Lu. Accurate atomization energies and dipole moments from Ornstein-Uhlenbeck Diffusion Monte Carlo calculations for small first-row polyatomic molecules. *The Journal of Chemical Physics*, 118(21):9528, 2002.
- [59] S. Manten and A. Lüchow. On the accuracy of the fixed-node Diffusion Quantum Monte Carlo method. *The Journal of Chemical Physics*, 115(12):5362, 2001.
- [60] E. A. Hylleraas. Neue berechnung der energie des heliums im grundzustande, sowie des tiefsten terms von ortho-helium. *Zeitschrift für Physik*, 54(5-6):347, 1929.
- [61] S. Manten and A. Lüchow. Linear scaling for the local energy in Quantum Monte Carlo. *The Journal of Chemical Physics*, 119(3):1307, 2003.

- [62] A. J. Williamson, R. Q. Hood, and J. C. Grossman. Linear-scaling Quantum Monte Carlo calculations. *Physical Review Letters*, 87(24):246406, 2001.
- [63] F. A. Reboredo and A. J. Williamson. Optimized nonorthogonal localized orbitals for linear scaling quantum Monte Carlo calculations. *Physical Review B*, 71:121105, 2005.
- [64] A. Aspuru-Guzik, R. Salomon-Ferrer, B. Austin, and W. A. Lester Jr. A sparse algorithm for the evaluation of the local energy in quantum Monte Carlo. *Journal of Computational Chemistry*, 26(7):708, 2005.
- [65] L. Smith and P. R. C. Kent. Development and performance of mixed OpenMP/MPI Quantum Monte Carlo code. *Concurrency: Practice and Experience*, 12:1121, 2000.
- [66] B. L. Hammond, W. A. Lester Jr., and P. J. Reynolds. *Monte Carlo Methods in Ab Initio Quantum Chemistry*. World Scientific, Hackensack, NJ, 1994.
- [67] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins. Optimized trial wave-functions for quantum Monte Carlo calculations. *Physical Review Letters*, 60(17):1719, 1988.
- [68] C. J. Umrigar and C. Filippi. Energy and variance optimization of many-body wave functions. *Physical Review Letters*, 94:150201, 2004.
- [69] J. Toulouse and C. J. Umrigar. Optimization of quantum Monte Carlo wave functions by energy minimization. *The Journal of Chemical Physics*, 126:084102, 2007.
- [70] C. J. Huang, C. J. Umrigar, and M. P. Nightingale. Accuracy of electronic wave functions in quantum Monte Carlo: The effect of high-order correlations. *The Journal of Chemical Physics*, 107(8):3007, 1997.
- [71] N. D. Drummond, M. D. Towler, and R. J. Needs. Jastrow correlation factor for atoms, molecules, and solids. *Physical Review B*, 70:235119, 2004.

- [72] A. Ma, M. D. Towler, N. D. Drummond, and R. J. Needs. Scheme for adding electron-nucleus cusps to Gaussian orbitals. *The Journal of Chemical Physics*, 122:224322, 2005.
- [73] M. Suzuki. Transfer-matrix method and Monte Carlo simulation in quantum spin systems. *Physical Review B*, 31:2957, 1985.
- [74] J. W. Moskowitz, K. E. Schmidt, M. A. Lee, and M. H. Kalos. A new look at correlation energy in atomic and molecular systems. II. The application of the Green's function Monte Carlo method. *Journal of Chemical Physics*, 77:343, 1982.
- [75] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester Jr. Fixed-node quantum Monte Carlo for molecules. *Journal of Chemical Physics*, 77(11):5593, 1982.
- [76] R. Assaraf, M. Caffarel, and A. Khelif. Diffusion Monte Carlo methods with a fixed number of walkers. *Physical Review E*, 61(4):4566, 2000.
- [77] J. B. Anderson. Quantum chemistry by random walk.  $\text{H } ^2\text{P}$ ,  $\text{H}_3^+ \text{D}_{3h} \text{ } ^1\text{A}'_1$ ,  $\text{H}_2 \text{ } ^3\Sigma_u^+$ ,  $\text{H}_4 \text{ } ^1\Sigma_g^+$ ,  $\text{Be } ^1\text{S}$ . *The Journal of Chemical Physics*, 65(10):4121, 1976.
- [78] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Physical Review Letters*, 45(7):566, 1980.
- [79] K. Morokuma, D. G. Musaev, T. Vreven, H. Basch, M. Torrent, and D. V. Khoroshun. Model studies of the structures, reactivities, and reaction mechanisms of metalloenzymes. *IBM Journal of Research and Development*, 45(3/4):367, 2001.
- [80] Teragrid. <http://www.teragrid.org>. Dec 15, 2009.
- [81] M. T. Feldmann, D. R. Kent IV, D. R. Fisher, and A. G. Anderson. QM<sup>c</sup>Beaver v20051107 ©, 2001-2008. <http://sourceforge.net/projects/qmcbeaver>.

- [82] C. R. Myers, C. J. Umrigar, J. P. Sethna, and J. D. Morgan III. Fock's expansion, Kato's cusp conditions, and the exponential ansatz. *Physical Review A*, 44(9):5537, 1991.
- [83] C. J. Umrigar. Accelerated metropolis method. *Physical Review Letters*, 71(3):408, 1993.
- [84] R. Needs, G. Rajagopal, M. D. Towler, P. R. C. Kent, and A. Williamson. CASINO, the Cambridge Quantum Monte Carlo code, version 1.1.0, 2000.
- [85] R. N. Barnett, B. L. Hammond, P. J. Reynold, L. Terray, and W. A. Lester Jr. Qmagic, the Quantum Monte Carlo code, version 1.1.0, 2000.
- [86] A. Aspuru-Guzik, R. Salomon-Ferrer, B. Austin, R. Perusquia-Flores, M. A. Griffin, R. A. Oliva, D. Skinner, D. Domin, and W. A. Lester Jr. Zori 1.0: A parallel quantum Monte Carlo electronic structure package. *Journal of Computational Chemistry*, 26(8):856, 2005.
- [87] Jaguar, version 7.5, 2008. Schrödinger LLC, New York, NY.
- [88] S. F. Boys. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Reviews of Modern Physics*, 32(2):296, 1960.
- [89] L. Pauling. The nature of the chemical bond. Application of results obtained from the quantum mechanics and from a theory of paramagnetic susceptibility to the structure of molecules. *Journal of the American Chemical Society*, 53(4):1367, 1931.
- [90] A. Strachan, E. M. Kober, A. C. T. van Duin, J. Oxgaard, and W. A. Goddard III. Thermal decomposition of RDX from reactive molecular dynamics. *The Journal of Chemical Physics*, 122:054502, 2005.
- [91] R. A. King, T. D. Crawford, J. F. Stanton, and H. F. Schaefer III. Conformations of [10]annulene: More bad news for density functional theory and

- second-order perturbation theory. *Journal of the American Chemical Society*, 121:10788, 1999.
- [92] J. Černý and P. Hobza. The X3LYP extended density functional accurately describes H-bonding but fails completely for stacking. *Physical Chemistry Chemical Physics*, 7:1624, 2005.
- [93] R. B. Woodward and R. Hoffman. The conservation of orbital symmetry. *Angewandte Chemie International Edition in English*, 8(11):781, 1969.
- [94] L. V. Gurvich, I. V. Veyts, and C. B. Alcock. *Thermodynamic Properties of Individual Substances, Fourth Edition*. Hemisphere Publishing Co, New York, NY, 1989.
- [95] M. Frenkel, K. N. Marsh, R. C. Wilhoit, G. J. Kabo, and G. N. Roganov. *Thermodynamics of Organic Compounds in the Gas State*. Thermodynamics Research Center, College Station, TX, 1994.
- [96] S. Sakai. Theoretical analysis of cyclic reaction mechanisms of two ethylenes. *International Journal of Quantum Chemistry*, 90:549, 2002.
- [97] J. A. Montgomery Jr., M. J. Frisch, J. W. Ochterski, and G. A. Petersson. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *Journal of Chemical Physics*, 110(6):2822, 1999.
- [98] B. Sirjean, P. A. Glaude, M. F. Ruiz-Lopez, and R. Fournet. Detailed kinetic study of the ring opening of cycloalkanes by CBS-QB3 calculations. *Journal of Physical Chemistry A*, 110:12693, 2006.
- [99] V. Guner, K. S. Khuong, A. G. Leach, P. S. Lee, M. D. Bartberger, and K. N. Houk. A standard set of pericyclic reactions of hydrocarbons for the benchmarking of computational methods: The performance of ab initio, density functional, CASSCF, CASPT2, and CBS-QB3 methods for the prediction of activation barriers, reaction energetics, and transition state geometries. *Journal of Physical Chemistry A*, 107:11445, 2003.

- [100] D. H. Ess and K. N. Houk. Activation energies of pericyclic reactions: Performance of DFT, MP2, and CBS-QB3 methods for the prediction of activation barriers and reaction energetics of 1,3-dipolar cycloadditions, and revised activation enthalpies for a standard set of hydrocarbon pericyclic reactions. *Journal of Physical Chemistry A*, 109:9542, 2005.
- [101] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *Journal of Chemical Physics*, 72(1):650, 1980.
- [102] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery. General atomic and molecular electronic structure system. *Journal of Computational Chemistry*, 14:1347, 1993. <http://www.msg.chem.iastate.edu/GAMESS/GAMESS.html>.
- [103] D. Feller. The role of databases in support of computational chemistry calculations. *Journal of Computational Chemistry*, 17(13):1571, 1996.
- [104] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. L. Windus. Basis set exchange: A community database for computational sciences. *Journal of Chemical Information and Modeling*, 47(3):1045, 2007.
- [105] J. P. Perdew. *Electronic Structure of Solids '91*, page 11. Akademie, Berlin, 1991.
- [106] Y. Zhao and D. G. Truhlar. A new local density functional for main group thermochemistry, transition metal bonding, thermochemical kinetics, and non-covalent interactions. *Journal of Chemical Physics*, 125:194101, 2006.
- [107] Y. Zhao and D. G. Truhlar. Density functional for spectroscopy: No long-range self-interaction error, good performance for Rydberg and charge-transfer states,

- and better performance on average than B3LYP for ground states. *Journal of Physical Chemistry A*, 110:13126, 2006.
- [108] H. Flyvberg and H. Peterson. Error estimates on averages of correlated data. *The Journal of Chemical Physics*, 91:461, 1989.
- [109] D. R. Fisher, D. R. Kent IV, M. T. Feldmann, and W. A. Goddard III. An optimized initialization algorithm to ensure accuracy in quantum Monte Carlo calculations. *Journal of Computational Chemistry*, 29(14):2335, 2008.
- [110] R. D. Johnson III. NIST computational chemistry comparison and benchmark database, NIST standard reference database number 101, 2006. <http://srdata.nist.gov/cccbdb>.
- [111] Xin Xu. Private communication.
- [112] R. D. J. Froese, S. Humbel, M. Svensson, and K. Morokuma. IMOMO(G2MS): A new high-level G2-like method for large molecules and its applications to Diels-Alder reactions. *Journal of Physical Chemistry A*, 101:227, 1997.
- [113] R. D. J. Froese, J. M. Coxon, S. C. West, and K. Morokuma. Theoretical studies of Diels-Alder reactions of acetylenic compounds. *Journal of Organic Chemistry*, 62:6991, 1997.