

Monte Carlo eigenvalue and variance estimates from several functional optimizations

S. A. Alexander, R. L. Coldwell, Hendrik J. Monkhorst, and J. D. Morgan III

Citation: *The Journal of Chemical Physics* **95**, 6622 (1991); doi: 10.1063/1.461532

View online: <http://dx.doi.org/10.1063/1.461532>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/jcp/95/9?ver=pdfcov>

Published by the AIP Publishing

Articles you may be interested in

[Zero-variance zero-bias quantum Monte Carlo estimators for the electron density at a nucleus](#)

J. Chem. Phys. **135**, 134112 (2011); 10.1063/1.3644964

[A simple technique to estimate partition functions and equilibrium constants from Monte Carlo simulations](#)

J. Chem. Phys. **102**, 6189 (1995); 10.1063/1.469063

[Optimized trial functions for quantum Monte Carlo](#)

J. Chem. Phys. **92**, 597 (1990); 10.1063/1.458409

[Variance and correlation length of energy estimators in Metropolis path integral Monte Carlo](#)

J. Chem. Phys. **89**, 7454 (1988); 10.1063/1.455276

[Molecular Schrödinger Equation. VII. Properties of the Energy Variance Function: The Estimation of Energy Eigenvalues](#)

J. Chem. Phys. **47**, 930 (1967); 10.1063/1.1712058



Monte Carlo eigenvalue and variance estimates from several functional optimizations

S. A. Alexander,^{a)} R. L. Coldwell,^{a)} and Hendrik J. Monkhorst^{b)}

Department of Physics, University of Florida, Gainesville, Florida 32611

J. D. Morgan III

Departments of Chemistry and Physics, Harvard University, Cambridge, Massachusetts 02138

and Institute for Theoretical Atomic and Molecular Physics, Harvard-Smithsonian Center for Astrophysics,

60 Garden Street, Cambridge, Massachusetts 02138

and Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716^{c)}

(Received 23 April 1991; accepted 16 July 1991)

Using several simple systems as examples, we show that the choice of optimization functional can have a significant influence on the accuracy of variational Monte Carlo calculations. In addition, we demonstrate that the Monte Carlo analog of the Rayleigh-Ritz procedure, which explicitly orthogonalizes ground and excited states, can be used to produce accurate eigenvalue and variance estimates of excited states.

I. INTRODUCTION

In a variational Monte Carlo calculation the form of a trial wave function, Ψ_i , is adjusted so as to most accurately evaluate an expectation value (within the statistical error). Because the form of the exact eigenfunction is rarely known, such calculations often contain both a systematic error and a statistical error, i.e., variance. Just as in more traditional calculations where the integrals and matrix elements are computed analytically, a systematic error occurs when the form of the trial wave function cannot exactly reproduce the exact wave function. In contrast, another widely used Monte Carlo method, diffusion Monte Carlo,¹⁻⁹ has, in principle, only a statistical error and can obtain the exact expectation value without parameterized trial wave functions. In practice, however, this method uses the best variational wave functions available in order to obtain the most accurate results.

In addition to systematic and statistical errors, another source of error in variational Monte Carlo calculations can occur when the adjustable parameters in Ψ_i are not chosen to most accurately evaluate an expectation value, but instead to satisfy some other criterion. This sort of tradeoff is often made when calculating the energy

$$\langle H \rangle = \sum_i [\Psi_i H \Psi_i / w_i] / \sum_i [\Psi_i^2 / w_i], \quad (1)$$

where

$$H = -1/2 \sum_a \nabla_a^2 - \sum_{aA} Z_a / r_{aA} + \sum_{a<b} 1/r_{ab} + \sum_{A<B} Z_A Z_B / r_{AB} \quad (2)$$

and $\Psi_i = \Psi_i(\mathbf{x}_i)$. In this notation the weight function,

$w_i = w(\mathbf{x}_i)$, is the relative probability of choosing the i th configuration, \mathbf{x}_i , and exactly compensates for not having chosen it randomly. If we want to obtain the most accurate evaluation of Eq. (1), one might expect that the parameters in the trial wave function should be adjusted to minimize this expectation value. In practice, however, an accurate optimization of this quantity requires a large number of configurations and often has a large variance. For this reason, such calculations are seldom performed. One alternative is to minimize the variance.¹⁰⁻¹⁸ These calculations have a lower bound (a variance of zero), they can be optimized using a small set of configurations, and their energy is usually quite close to (but always above) the energy minimum in a given parameter space.

The idea of optimizing the variance dates back at least to Conroy¹⁰ who adjusted the parameters in his trial wave function with respect to the functional

$$\sum_i [(H\Psi_i - E_{in}\Psi_i)^2 / w_i] / \sum_i [\Psi_i^2 / w_i], \quad (3)$$

where the reference energy, E_{in} , is fixed at a value which is close to the desired state. This causes the parameter optimization to start in the proper region. Although Eq. (3) has been used as the optimization functional in several recent calculations,^{16,18} it does not minimize the variance. This was pointed out by Coldwell in 1977^{11,19} who showed that the variance in the energy is instead given by

$$\sum_i [(H\Psi_i - E_{in}\Psi_i)^2 \Psi_i^2 / w_i^2] / \left\{ \sum_i [\Psi_i^2 / w_i] \right\}^2. \quad (4)$$

In principle, any functional of the wave function and the local energy can be used to adjust the parameters in the trial wave function. The ideal functional would be one which accurately evaluates the expectation value, gives a low variance, and can be optimized using a small number of configurations. Because no comparison of optimization functionals has ever been made, we examine in this paper the effect such functionals can have on the accuracy of the energy. In Sec. II we show that Eq. (4) is the variance of the energy and Eq.

^{a)} Also at Institute for Astrophysics and Planetary Exploration, One Progress Blvd. Box 33, Alachua, Florida 32615.

^{b)} Also at Quantum Theory Project, University of Florida, Gainesville, Florida 32611.

^{c)} Permanent address.

(3) is not. Section III contains a brief outline of our Monte Carlo algorithm and of our comparison of several functionals using the helium 2^3S state as a test system. In Sec. IV we examine the behavior of the three most promising functionals for several simple atoms and molecules. Section V describes a method for modifying any straightforward calculation of excited states to explicitly include orthogonality with the lower state(s).

II. COMPUTING THE VARIANCE

To show that Eq. (4) is the variance of the Hamiltonian and that Eq. (3) is not, we present in this section a simple pedagogical example. A more rigorous derivation is given in Appendix A.

Consider a one-dimensional system $0 < x < 1$ with uniform, unweighted data and where $\Psi_1 = 0.05$ and $H\Psi_1/\Psi_1 = 1.0$ if $0 < x < 0.5$ and $\Psi_2 = 0.95$ and $H\Psi_2/\Psi_2 = 0.5$ if $0.5 < x < 1$. The exact energy for this system is simply the sum of the contributions from the two regions, i.e., $\langle H_0 \rangle = [0.05 \cdot 1.0 \cdot 0.05 + 0.95 \cdot 0.5 \cdot 0.95] / [(0.05)^2 + (0.95)^2] = 0.501\,381$. Table I shows that a Monte Carlo calculation with two configurations has a $1/4$ probability of choosing both configurations in the region $0 < x < 0.5$ giving $\langle H \rangle = 1.0$; a $1/4$ probability of choosing the first configuration in the region $0 < x < 0.5$ and the second configuration in the region $0.5 < x < 1$ giving $\langle H \rangle = 0.501\,381$; a $1/4$ probability of choosing the first configuration in the region $0.5 < x < 1$ and the second configuration in the region $0 < x < 0.5$ giving $\langle H \rangle = 0.501\,381$; and a $1/4$ probability of choosing both configurations in the region $0.5 < x < 1$ giving $\langle H \rangle = 0.5$. The average expectation value, Eq. (1), is the average over all possible ways of picking two configurations and yields $\langle H \rangle = 0.625\,690$. This value is close to the exact result and well within the variances pre-

dicted by both Eqs. (3) and (4), i.e., $\sigma_i^2 = 6.25 \times 10^{-2}$ and 3.11×10^{-2} , respectively.

When this same system is studied using ten configurations, we find that $\langle H \rangle = 0.522\,63$ and that Eqs. (3) and (4) give $\sigma_i^2 = 1.13 \times 10^{-3}$ and 2.55×10^{-5} , respectively. In Fig. 1 we plot the histogram of all possible energies which can be calculated from ten configurations. As expected, these energies are distributed closely around the exact value. The height of the histogram at a particular energy is the relative probability of ten configurations yielding this energy. The standard deviation should be such that 65% of the energy values fall inside $E \pm \sigma$ and 35% fall outside. From statistical theory we know that values in a random distribution are placed about the mean value $\langle x_0 \rangle$ like $\exp\{-0.5[(\langle x \rangle - \langle x_0 \rangle)/\sigma]^2\}$ and that σ = full width at half-maximum/2.345.²⁰ The values in Fig. 1 have a full width at half-maximum of about 0.01 and thus a variance of $\sigma_i^2 = 1.82 \times 10^{-5}$. This value is in good agreement with the result produced by Eq. (4) (the numbers do not agree exactly because a finite number of configurations can never produce a pure Gaussian distribution) and is much smaller than the result produced by Eq. (3).

The disagreement above between Eqs. (3) and (4) is due in part to the fact that we have sampled the configurations uniformly. If we had instead sampled them with respect to Ψ^2 , as the Markov chain method does, then the variance produced by both equations would have differed only by a factor N . This freedom to sample a system independently of the wave function forms the heart of the biased-selection method which is discussed at length in the next section.

III. CHOOSING A MINIMIZATION FUNCTIONAL

In this section we examine the behavior of ten functionals using the helium 2^3S state as a test system. To perform

TABLE I. Energy and variance estimates of the model problem described in Sec. II. These values are averages over all ways of choosing two configurations from the two regions.

$\sum_i [\Psi_i H \Psi_i / w_i] / \sum_i [\Psi_i^2 / w_i] = 0.625\,690$		
Region 1	Region 1	$[0.05 \cdot 1.0 \cdot 0.05 + 0.05 \cdot 1.0 \cdot 0.05] / [(0.05)^2 + (0.05)^2] = 1.0$
Region 1	Region 2	$[0.05 \cdot 1.0 \cdot 0.05 + 0.95 \cdot 0.5 \cdot 0.95] / [(0.05)^2 + (0.95)^2] = 0.501\,381$
Region 2	Region 1	$[0.95 \cdot 0.5 \cdot 0.95 + 0.05 \cdot 1.0 \cdot 0.05] / [(0.95)^2 + (0.05)^2] = 0.501\,381$
Region 2	Region 2	$[0.95 \cdot 0.5 \cdot 0.95 + 0.95 \cdot 0.5 \cdot 0.95] / [(0.95)^2 + (0.95)^2] = 0.5$
$\sum_i [(H \Psi_i / \Psi_i - E_{in})^2 \Psi_i^2 / w_i] / \sum_i [\Psi_i^2 / w_i] = 6.25 \times 10^{-2}$		
Region 1	Region 1	$[(1.0 - 0.501\,381)^2 (0.05)^2 + (1.0 - 0.501\,381)^2 (0.05)^2] / [(0.05)^2 + (0.05)^2] = 2.49 \times 10^{-1}$
Region 1	Region 2	$[(1.0 - 0.501\,381)^2 (0.05)^2 + (0.5 - 0.501\,381)^2 (0.95)^2] / [(0.05)^2 + (0.95)^2] = 6.89 \times 10^{-4}$
Region 2	Region 1	$[(0.5 - 0.501\,381)^2 (0.95)^2 + (1.0 - 0.501\,381)^2 (0.05)^2] / [(0.95)^2 + (0.05)^2] = 6.89 \times 10^{-4}$
Region 2	Region 2	$[(0.5 - 0.501\,381)^2 (0.95)^2 + (0.5 - 0.501\,381)^2 (0.95)^2] / [(0.95)^2 + (0.05)^2] = 1.91 \times 10^{-6}$
$\sum_i [(H \Psi_i / \Psi_i - E_{in})^2 \Psi_i^4 / w_i^2] / \left\{ \sum_i [\Psi_i^2 / w_i] \right\}^2 = 3.11 \times 10^{-2}$		
Region 1	Region 1	$[(1.0 - 0.501\,381)^2 (0.05)^4 + (1.0 - 0.501\,381)^2 (0.05)^4] / [(0.05)^2 + (0.05)^2]^2 = 1.24 \times 10^{-1}$
Region 1	Region 2	$[(1.0 - 0.501\,381)^2 (0.05)^4 + (0.5 - 0.501\,381)^2 (0.95)^4] / [(0.05)^2 + (0.95)^2]^2 = 3.79 \times 10^{-6}$
Region 2	Region 1	$[(0.5 - 0.501\,381)^2 (0.95)^4 + (1.0 - 0.501\,381)^2 (0.05)^4] / [(0.95)^2 + (0.05)^2]^2 = 3.79 \times 10^{-6}$
Region 2	Region 2	$[(0.5 - 0.501\,381)^2 (0.95)^4 + (0.5 - 0.501\,381)^2 (0.95)^4] / [(0.95)^2 + (0.05)^2]^2 = 9.54 \times 10^{-7}$

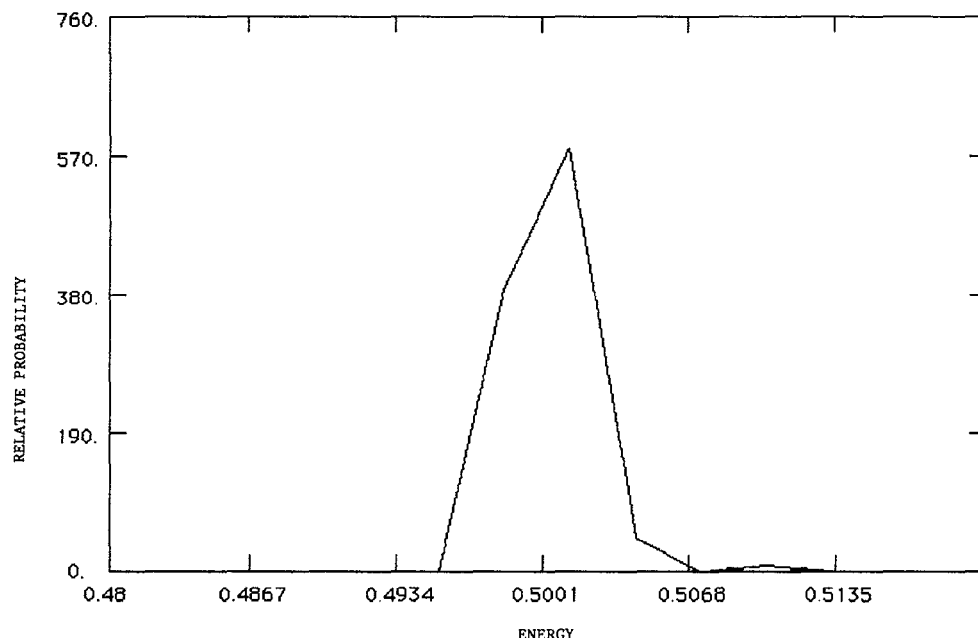


FIG. 1. Histogram of all possible energies which can be computed from ten configurations as described in Sec. II.

these calculations we use biased-selection Monte Carlo—a method which uses a small set of configurations, \mathbf{x}_i , and weights, $w(\mathbf{x}_i)$, to evaluate multidimensional integrals¹¹

$$I = \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_i)/w(\mathbf{x}_i) \rightarrow \int f(\mathbf{x}) d\mathbf{x} \quad N \rightarrow \infty$$

Each configuration is a collection of electronic coordinates, e.g., $\mathbf{x}_i = (\mathbf{r}_1, \mathbf{r}_2)$ for a two-electron system, and contains the usual Cartesian, $\mathbf{r}_j = (x_j, y_j, z_j)$, or spherical, $\mathbf{r}_j = (r_j, \theta_j, \phi_j)$, components. These components are generated using some predetermined guiding function $g(x)$. In its simplest form, the introduction of a guiding function is equivalent to importance sampling.²¹ If $g(x)$ is large where $f(x)$ is large and small where $f(x)$ is small then the configurations will tend to be chosen from those regions which contribute most to the integral. In addition, $f(\mathbf{x}_i)/w(\mathbf{x}_i)$ will be roughly constant which helps to reduce the variance. The only restriction on the choice of $g(x)$ is that it be positive definite everywhere and computable. To study the helium 2^3S state, we would like to have a guiding function which essentially approximates $\Psi(\mathbf{r}_1, \mathbf{r}_2)^2$. For simplicity we choose

$$\begin{aligned} g(r) &= 23.404r^2e^{-3r} \quad \text{when } r = [0,1], \\ &= \text{maximum } (23.404r^2e^{-3r}, 0.042r^4e^{-r}) \\ &\quad \text{when } r = [1,100], \end{aligned} \quad (6)$$

where r is the electron–nuclear distance. The factor e^{-3r} lies between e^{-2r} (a hydrogenic electron around a shielded +1 atom) and e^{-4r} (a hydrogenic electron around a +2 atom) and so assumes that when one electron is close to the nucleus the second electron is even closer 50% of the time. At distances far from the nucleus the factor r^2e^{-r} takes into account the fact that this electron normally sees a nucleus com-

pletely shielded by the other electron and so should be treated as a $2s$ orbital of this hydrogenic-like system. For reasons discussed in the next paragraph, we include an extra factor of r^2 in both parts of Eq. (6). The constants 23.404 and 0.042 are chosen because they equalize the value of the integrals from $[0,1]$ and $[1,100]$ and thus allow both the short-range region and the long-range region to be sampled nonpreferentially.

Once a guiding function has been chosen, the calculation of the configurations and weights is straightforward. For a small atom it is sufficient to take \mathbf{r}_1 and \mathbf{r}_2 “biased as random.” This means finding a point (r_j, θ_j, ϕ_j) such that

$$R = \int_0^{r_j} g(s) ds, \quad \Theta = \theta_j, \quad \Phi = \phi_j, \quad (7)$$

where R , Θ , and Φ are random numbers chosen from the range

$$R = \left[0, \int_0^\infty g(s) ds\right], \quad \Theta = [0, 2\pi], \quad \text{and } \Phi = [-\pi, \pi]. \quad (8)$$

Since the coordinates of each electron are chosen independently, the weight of each configuration is given by

$$w(\mathbf{x}_i) = w(\mathbf{x}_1)w(\mathbf{x}_2), \quad (9)$$

where

$$w(\mathbf{x}_j) = g(r_j) \left[4\pi r_j^2 \int_0^\infty g(s) ds\right]. \quad (10)$$

The factor $4\pi r_j^2$ is needed here to compensate for choosing the points in spherical coordinates in Eq. (7). When $g(r_j)$ is of the form $r_j^2 f(r_j)$, the overall factor of r_j^2 in $g(r_j)$ cancels the r_j^2 in the denominator of Eq. (10). If we make $g(0)$ finite instead of zero, this cancellation will eliminate the electron–nuclear cusp.

One way to tell whether a reasonable guiding function has been chosen is to calculate the effective number of configurations which contribute to the normalization integral

$$N_{\text{eff}} = \left\{ \sum_i [\Psi_i^2/w_i] \right\}^2 / \sum_i [\Psi_i^4/w_i^2]. \quad (11)$$

This quantity will be of order 1 when only a few of the points dominate the integral, and it will be of order N when all the points make roughly equal contributions. In practice, we have found that N_{eff} must be on the order of ten times the number of adjustable parameters and between $0.2N$ and $0.4N$ to perform an accurate optimization. If N_{eff} is too small then too many points are being wasted sampling unimportant regions instead of lowering σ . If N_{eff} is too large then not enough points are sampling the other regions to see if any of them are important. In either case the guiding function(s) should be changed. Fortunately, our experience suggests that the exact form of the guiding function is not critical to the success of the biased-selection method.

For the helium 2^3S state we take a trial wave function of the form

$$\Psi_i = (1 - P_{12}) \sum_k a_k \{ e^{-\alpha r_1 - \beta r_2} r_{12}^{b_k} \times (r_1 - r_2)^{c_k} (r_1 + r_2)^{d_k} \}, \quad (12)$$

where P_{12} is the operator which interchanges the coordinates of the two electrons and where α , β , and a_k are the parameters determined by optimizing the minimization functional. As in a typical Hylleraas calculation, the constants b_k , c_k , and d_k have integral values $0, 1, 2, \dots$ and are chosen beforehand.²² All values up to $b_k + c_k + d_k = 6$ are included except for those where c_k is even. These are excluded because we require that the spatial part of our wave function be symmetric with respect to an interchange of the two electrons. As a result, Eq. (12) contains a total of 36 variational parameters.

During some of these calculations we generated a small number of configurations which have two electrons placed close to one another. Such configurations produce a sudden jump in the variance since the electron-electron potential is nearly singular (r_{ij} very small) and thus the energy becomes unusually large. In Appendix B we describe a simple addi-

tion to our Monte Carlo procedure which eliminates this problem.

In Table II we list the results produced by each functional. All the variances have been computed using Eq. (4) (with $E_{\text{in}} = -2.175$ hartrees) but the linear and nonlinear parameters have been determined so as to minimize each functional. In the optimization step 4000 configurations were used. Once the variational parameters were determined, the random number seed was changed to insure statistical independence and then 1 024 000 configurations were used to calculate the energy. It should be noted, however that our 34-parameter wave function is incapable of giving the "exact" result. In Appendix C we use a Gaussian integration version of the biased-selection method to show that the energy minimum in this parameter space is $-2.175\,225\,3$ hartree, which is 4×10^{-5} hartree higher than the best value in the literature.²³ Of the ten functionals examined in Table II, #1, #5, and #10 have both a low variance and a reasonably good energy. As expected, #1 (the variance functional) produces the lowest variance. This value is 33% lower than #5 (the Conroy functional). In contrast, the energy produced by the variance functional is almost 5 standard deviations above the parameter space limit whereas the Conroy functional value agrees with it within the statistical error.

These results can be explained by a few general principles. The first is that a fixed number of constants can only decrease the integrand of a functional in one region by increasing it in another region. Since we always use a fixed, finite number of configurations in our optimization step, this increase can even occur in a region where the wave function is not sampled. Compared to a direct optimization of the local energy, i.e., $\Psi_i^2(H\Psi_i/\Psi_i - E)$, the $\Psi_i^4(H\Psi_i/\Psi_i - E)^2$ sampling of the variance functional (#1) overweights those regions where Ψ_i is large and underweights those where Ψ_i is small. Similarly, the $(H\Psi_i/\Psi_i - E)^2$ weighting of the variance (#1) and Conroy (#5) functionals will either overemphasize or underemphasize the importance of the local energy. Because our basis functions do not satisfy the electron-electron cusp condition and the region around the electron-electron singularity is not heavily sampled, the largest deviations in our local energies occur when $r_{ij} \rightarrow 0$. When this effect dominates the local energy, both the vari-

TABLE II. The energy of the helium 2^3S state (in hartrees) produced by several minimization functionals.

Functional	Energy
1 $\sum_i [(H\Psi_i - E_{\text{in}}\Psi_i)^2 \Psi_i^2/w_i^2] / [\sum_i \Psi_i^2/w_i]$	$-2.175\,207\,6 \pm 0.000\,003\,4$
2 $\sum_i [(H\Psi_i - E_{\text{in}}\Psi_i)^2 \Psi_i^2/w_i] / \sum_i [\Psi_i^4/w_i]$	$-2.175\,160\,1 \pm 0.000\,004\,2$
3 $\sum_i [(H\Psi_i - E_{\text{in}}\Psi_i)^2 \Psi_i^2] / \sum_i [\Psi_i^4]$	$-2.173\,758\,5 \pm 0.000\,026\,7$
4 $\sum_i [(H\Psi_i - E_{\text{in}}\Psi_i)^2/w_i^2] / \sum_i [\Psi_i^2/w_i^2]$	$-2.175\,096\,0 \pm 0.000\,021\,5$
5 $\sum_i [(H\Psi_i - E_{\text{in}}\Psi_i)^2/w_i] / \sum_i [\Psi_i^2/w_i]$	$-2.175\,224\,0 \pm 0.000\,004\,4$
6 $\sum_i [(H\Psi_i - E_{\text{in}}\Psi_i)^2] / \sum_i [\Psi_i^2]$	$-2.174\,980\,6 \pm 0.000\,009\,8$
7 $\sum_i [(H\Psi_i - E_{\text{in}}\Psi_i)^2 \Psi_i^2/w_i^2] / \sum_i [1/w_i^2]$	Unstable
8 $\sum_i [(H\Psi_i - E_{\text{in}}\Psi_i)^2 \Psi_i^2/w_i] / \sum_i [1/w_i]$	Unstable
9 $\sum_i [(H\Psi_i - E_{\text{in}}\Psi_i)^2 \Psi_i^2] / \sum_i [1]$	$-2.170\,869\,4 \pm 0.000\,179\,0$
10 $\sum_i [H\Psi_i/\Psi_i - E_{\text{in}}] \Psi_i^2/w_i / \sum_i [\Psi_i^2/w_i]$	$-2.175\,216\,8 \pm 0.000\,004\,9$
Parameter space limit	$-2.175\,225\,3$
Drake (Ref. 23)	$-2.175\,229\,378\,236\,790\,7$

ance (#1) and Conroy (#5) functionals will yield roughly the same values. In this system the energies produced by both the Conroy functional (#5) and functional #10 (hereafter referred to as the local-energy functional) are quite close to one another which indicates that deviations in the wave function are the main source of error. This is not surprising since the antisymmetric nature of the triplet state prevents two electrons from approaching one another and thus minimizes the effect of the cusp condition. As the quality of a wave function increases, however, all functionals should eventually produce values which are in close agreement. When a small number of configurations are used to optimize a functional, several problems may appear. Functionals #7 and #8 optimized easily but the resulting wave functions were found to be unstable when evaluated using a larger number of configurations. The reason is that these functionals can be made small by setting $H\Psi_i = E\Psi_i$ mostly in those regions where the wave function is small. This violates a fundamental assumption of the biased-selection method which requires that the wave function be small only where the weight (i.e., the guiding function) is small. This problem does not occur with functional #9 because all points are equally weighted. This functional is also of some interest because it samples only the deviations in the local energy, i.e., $(H\Psi_i/\Psi_i - E)^2$. Such a weighting prevents a minimization routine from setting $\Psi_i^2 = 0$ except in the region where the desired solution is located. In Ref. 17 we used this functional (referred to there as NOPSI) to produce a good, reliable estimate of all nonlinear parameters. From such a starting point we were able to minimize other functionals with far fewer problems than when we attempted to optimize these functionals directly. This ability may be of some use when a less than perfect wave function is available.

IV. RESULTS

In this section we compare the results from the variance (#1), Conroy (#5) and local-energy (#10) functionals for a number of simple systems. Table III gives our energies for the helium 2^3S state and those from a number of other calculations.^{16,23–25} Although the relative merits of each functional have already been discussed in Sec. III, we first wish to note that our variances decrease by a factor of 2 when the number of configurations increase by a factor of 4. This

drop is a trademark of a correct biased-selection calculation and indicates that the error per configuration has become independent of N .²¹ When this is not the case it means that there are inadequately sampled regions such as untreated singularities in the electron–nuclear and electron–electron potentials or other regions of configuration space where $w_i \rightarrow 0$ faster than Ψ_i^2 . Since each of our configurations is statistically independent, this feature also enables us to easily compute how many configurations would be needed to reach a given variance.

We also wish to point out that by adjusting the variational parameters to minimize a combination of both the energy and the variance rather than the variance alone, i.e., $(1-x)*\sigma + x*E$ where x is the percentage of the energy added, the variance functional (#1) can produce energies closer to the parameter space limit. The two drawbacks with this procedure, however, are that it gives a larger variance and the results differ from a variance minimization only when the difference in energy between the variance minimum and the energy minimum is larger than the Monte Carlo variance. Typically this means that as a greater percentage of the energy is mixed into the calculation, more configurations are needed in the optimization step. We found that neither 4 000 nor 16 000 configurations, although sufficient to minimize any of the functionals in Sec. III, would allow even 1% of the energy to be included. This is in marked contrast to our earlier calculations on nuclear bound states where the values from the variance + energy optimizations were significantly better than those from the variance-only optimizations.¹⁷ We believe that this indicates that the nuclear problem, where the probability of two particles coming close to one another is relatively small because the nucleon–nucleon potential is repulsive as $r_{ij} \rightarrow 0$, is intrinsically easier than the Coulombic problem which has both attractive and repulsive potentials and where particles are often close to one another.

For our calculations on the helium singlet ground state, we used the same guiding function as Eq. (6) and a trial wave function of the form

$$\Psi_t = (1 + P_{12}) \sum_k a_k \{e^{-\alpha r_1 - \beta r_2} r_{12}^{b_k} \times (r_1 - r_2)^{c_k} (r_1 + r_2)^{d_k}\}. \quad (13)$$

TABLE III. The energy of the helium 2^3S state (in hartrees). N is the number of configurations.

N	Variance functional (#1)	Conroy functional (#5)	Local-energy functional (#10)
16 000	$-2.175\,213\,9 \pm 0.000\,027\,5$	$-2.175\,247\,6 \pm 0.000\,038\,3$	$-2.175\,262\,5 \pm 0.000\,043\,0$
64 000	$-2.175\,218\,2 \pm 0.000\,014\,1$	$-2.175\,242\,4 \pm 0.000\,018\,9$	$-2.175\,237\,1 \pm 0.000\,021\,0$
256 000	$-2.175\,202\,0 \pm 0.000\,006\,8$	$-2.175\,216\,9 \pm 0.000\,009\,0$	$-2.175\,212\,6 \pm 0.000\,009\,9$
1 024 000	$-2.175\,207\,6 \pm 0.000\,003\,3$	$-2.175\,224\,0 \pm 0.000\,004\,4$	$-2.175\,216\,8 \pm 0.000\,004\,9$
Pekeris (Ref. 24)		$-2.175\,229\,378\,237^a$	
Thakkar and Smith (Ref. 25)		$-2.175\,229\,376^b$	
Drake (Ref. 23)		$-2.175\,229\,378\,236\,790\,7^c$	
Umrigar <i>et al.</i> (Ref. 16)		$-2.175\,226\,(2)^d$	

^a 1078 functions in perimetric coordinates.

^b 55 random-tempered Slater-type geminals.

^c 616 Hylleraas functions with multiple sets of exponents.

^d Variational Monte Carlo.

All values up to $b_k + c_k + d_k = 6$ are included except for those where c_k is odd. These are excluded because we require that the spatial part of our wave function be antisymmetric with respect to an interchange of the two electrons. As a result, Eq. (13) contains a total of 52 variational parameters. In Eq. (4) E_{in} was set to -2.904 hartree. In the optimization step 4 000 configurations were used. The results from the three functionals are presented in Table IV and are not much different from those for the triplet ground state. The variance from the variance functional (#1) is about 19% lower than the values given by both the Conroy (#5) and local-energy (#10) functionals and the energy is about 1.5 standard deviations above the values in the literature.^{16,24,26-29} In contrast, the energy given by both the Conroy (#5) and local-energy (#10) functional agree with the nonrelativistic limit within the statistical error.

In Table V we list our results for the first excited singlet (2^1S) state of helium. Both the guiding function and the form of the trial wavefunction were taken to be the same as for the singlet ground state. Other than setting $E_{in} = -2.146$ hartree this calculation was performed in exactly the same manner as the earlier ones. The results for this system, however, show several important differences. The most obvious is that the results from the local-energy functional (#10) now look much more like the results from the variance functional (#1). In addition, the energy from the variance functional (#1) is closer to the nonrelativistic limit^{23,25,30} than is the value from the Conroy functional (#5) and its variance is almost 27% smaller. Since an accurate description of this system requires many more terms than we have allowed, it is not surprising that the energy produced by the variance functional (#1) is almost 8 standard deviations above the exact value. The reason the variance functional (#1), rather than the Conroy functional (#5), produces the best energy for this state is that its Ψ^4 weighting deemphasizes those regions where Ψ is small. At a dissociation threshold the wave function becomes relatively constant over a large region (and therefore small compared to the wave function of a bound state). Near a dissociation threshold an optimization with the Conroy functional (#5) is thus more likely to jump to the dissociated wave function

than an optimization with the variance functional (#1). The local energy functional (#10) avoids this problem by being more dependent on the accuracy of $|H\Psi_i/\Psi_i - E|$. As the desired state gets closer to a dissociation level, however, all functionals eventually fail. This occurred when we tried to examine the second excited singlet (3^1S) state of helium.

For the H_2 ground state at internuclear distance 1.4 a.u. we took E_{in} to be -1.174 hartree and our guiding function to be

$$g(r) = 12.371r^2e^{-2r} \quad \text{when } r = [0,1], \\ = \text{maximum} (12.371r^2e^{-2r}, 0.042r^4e^{-r}) \\ \text{when } r = [1,100]. \quad (14)$$

Since only one electron will usually be around each atom at small distances, we assume no shielding, i.e., e^{-2r} . At large electron-nuclear distances the outer electron typically sees both atoms and the inner electron so we use the same form as helium, i.e., r^2e^{-r} . The constants 12.371 and 0.042 are chosen to sample these regions in a 1:1 ratio. One way to implement the biased-selection algorithm in molecular systems is to first choose an atom at random, e.g., at \mathbf{R}_k . Each configuration is then taken to be the sum of \mathbf{R}_k and the components in Eq. (7). To eliminate the possibility that the weight of this configuration might go to zero faster than Ψ_i^2 in some region of space (which would make a large part of the integral's value to come from a single configuration) Eq. (10) must be replaced by¹²

$$w(\mathbf{x}_j) = \frac{1}{\text{Atoms}} \sum_{k=1}^{\text{Atoms}} g(|\mathbf{r}_j - \mathbf{R}_k|) / \left[4\pi |\mathbf{r}_j - \mathbf{R}_k|^2 \right. \\ \left. \times \int_0^\infty g(s) ds \right]. \quad (15)$$

For our trial wave function we took

$$\Psi_t = (1 + P_{12}) \sum_k a_k \{ e^{-\alpha r_{1A} - \beta r_{2B}} r_{12}^{b_k} (r_{1A} - r_{1B})^{c_k} \\ \times (r_{1A} + r_{1B})^{d_k} (r_{2A} - r_{2B})^{e_k} (r_{2A} + r_{2B})^{f_k} \}, \quad (16)$$

where A and B are the positions of the two nuclei. This form

TABLE IV. The energy of the helium 1^1S state (in hartrees). N is the number of configurations.

N	Variance functional (#1)	Conroy functional (#5)	Local-energy functional (#10)
16 000	$-2.903\,734\,1 \pm 0.000\,095\,2$	$-2.903\,747\,4 \pm 0.000\,116\,7$	$-2.903\,757\,7 \pm 0.000\,114\,4$
64 000	$-2.903\,750\,1 \pm 0.000\,050\,1$	$-2.903\,776\,6 \pm 0.000\,056\,7$	$-2.903\,784\,7 \pm 0.000\,056\,0$
256 000	$-2.903\,726\,5 \pm 0.000\,023\,4$	$-2.903\,748\,8 \pm 0.000\,028\,2$	$-2.903\,749\,1 \pm 0.000\,027\,8$
1 024 000	$-2.903\,707\,4 \pm 0.000\,011\,2$	$-2.903\,734\,3 \pm 0.000\,013\,3$	$-2.903\,734\,3 \pm 0.000\,013\,2$
Pekeris (Ref. 24)		$-2.903\,724\,375^a$	
Rybak <i>et al.</i> (Ref. 26)		$-2.903\,723\,8^b$	
Thakkar and Smith (Ref. 27)		$-2.903\,724\,363^c$	
Baker <i>et al.</i> (Ref. 28)		$-2.903\,724\,377\,034\,118\,4^d$	
Umrigar <i>et al.</i> (Ref. 16)		$-2.903\,722\,(2)^e$	
Skinner <i>et al.</i> (Ref. 29)		$-2.903\,55(61)^f$	

^a 1078 functions in perimetric coordinates.

^b 100 fully optimized Gaussian-type geminals.

^c 66 random-tempered Slater-type geminals.

^d 476 Hylleraas functions with logarithmic terms.

^e Variational Monte Carlo.

^f Diffusion Monte Carlo.

TABLE V. The energy of the helium 2^1S state (in hartrees). N is the number of configurations. Orthogonality with the ground state is not imposed.

N	Variance functional (#1)	Conroy functional (#5)	Local-energy functional (#10)
16 000	$-2.145\,821\,7 \pm 0.000\,072\,5$	$-2.145\,836\,0 \pm 0.000\,089\,2$	$-2.145\,830\,8 \pm 0.000\,077\,6$
64 000	$-2.145\,871\,2 \pm 0.000\,033\,8$	$-2.145\,751\,6 \pm 0.000\,043\,3$	$-2.145\,854\,9 \pm 0.000\,036\,0$
256 000	$-2.145\,884\,8 \pm 0.000\,017\,1$	$-2.145\,802\,3 \pm 0.000\,021\,7$	$-2.145\,862\,5 \pm 0.000\,018\,1$
1 024 000	$-2.145\,880\,6 \pm 0.000\,008\,6$	$-2.145\,782\,3 \pm 0.000\,010\,8$	$-2.145\,868\,0 \pm 0.000\,009\,1$
Pekeris (Ref. 30)		$-2.145\,974\,037^a$	
Thakkar and Smith (Ref. 25)		$-2.145\,973\,824^b$	
Drake (Ref. 23)		$-2.145\,974\,046\,054\,143^c$	

^a615 functions in perimetric coordinates.^b55 random-tempered Slater-type geminals.^c616 Hylleraas functions with multiple sets of exponents.

is similar to the one used by James and Coolidge.³¹ All values up to $b_k + c_k + d_k + e_k + f_k = 4$ were selected. This resulted in a total of 128 variational parameters which were optimized with 4000 configurations. The values presented in Table VI show that for this system the variance functional (#1) and the Conroy functional (#5) produce very similar results. Both energies also agree within their error with the best results from the literature.^{18,32–34} The close agreement between these two functionals is to be expected since the rather primitive wave function considered in Eq. (16) approximates the electron–electron cusp condition by one or two terms of the form $(1 - ar_{12} - br_{12}^2)$ for $r_{12} < 1$ a.u. For systems where Ψ_i is relatively small for $r_{12} > 1$ a.u. (most atoms) this approximation will be reasonably accurate. For more diffuse systems where Ψ_i is relatively large at large values of r_{12} (most molecules—in the H_2 singlet ground state the electrons are correlated over the entire length of the molecule, i.e., 1.4 a.u.), this form cannot adjust fast enough to be constant for $r_{12} > 1$ a.u. and still be correct as $r_{12} \rightarrow 0$. In such cases the cusp condition will not be satisfied and its error will dominate the calculation. As a result, the $(H\Psi_i/\Psi_i - E)^2$ weighting of the variance functional (#1) and the Conroy functional (#5) should produce similar results and the local-energy functional (#10), because of its $|H\Psi_i/\Psi_i - E|$ weighting, quite different results.

The next system we consider is the first singlet excited ($B^1\Sigma_u^+$) state of H_2 . The guiding function and the form of

the trial wave function were taken to be the same as the ground state. Since this form of the guiding function is not ideal, we had to optimize the nonlinear parameters with 16 000 configurations in order to get a sufficiently high N_{eff} . E_{in} was set to -0.7058 hartree. Because the first dissociation threshold is relatively far away, we had no convergence problems with any of the functionals. The values in Table VII show that in contrast to the H_2 ground state, the Conroy (#5) and local-energy (#10) functionals now yield energies which are very similar. The reason for this is that the wave function of this system increases as a function of r_{12} over a range of several a.u. As a result, the wave function in Eq. (16) is able to satisfy the cusp condition and so deviations in the wave function are the main source of error. These deviations are also the reason why the variance functional (#1) produces a variance for this state which is almost 55% lower than the value given by the Conroy functional (#5) and an expectation value which is significantly below that produced by other calculations.^{35,36} This apparent breakdown in the variational Monte Carlo method has occurred because our final wave function is not a pure eigenfunction but a mixture of both ground and excited states. To some extent this problem is present in all Monte Carlo calculations which do not explicitly orthogonalize ground and excited states in the minimization step. We consider this subject in more detail in Sec. V and propose one solution.

For the ground state of the H_3^+ ion we examined the

TABLE VI. The energy of the H_2 ground state (in hartrees). N is the number of configurations. The internuclear distance is 1.4 a.u.

N	Variance functional (#1)	Conroy functional (#5)	Local-energy functional (#10)
16 000	$-1.174\,540\,6 \pm 0.000\,100\,1$	$-1.174\,577\,9 \pm 0.000\,100\,2$	$-1.174\,345\,9 \pm 0.000\,174\,4$
64 000	$-1.174\,399\,8 \pm 0.000\,050\,1$	$-1.174\,438\,4 \pm 0.000\,051\,1$	$-1.174\,196\,6 \pm 0.000\,088\,3$
256 000	$-1.174\,445\,2 \pm 0.000\,025\,4$	$-1.174\,453\,0 \pm 0.000\,025\,8$	$-1.174\,174\,4 \pm 0.000\,044\,7$
Kolos <i>et al.</i> (Ref. 32)		$-1.174\,475\,668^a$	
Frye <i>et al.</i> (Ref. 33)		$-1.174\,474\,67^b$	
Alexander <i>et al.</i> (Ref. 34)		$-1.174\,474\,85^c$	
Sun <i>et al.</i> (Ref. 18)		$-1.172\,2(1)^d$	
Sun <i>et al.</i> (Ref. 18)		$-1.174\,51(10)^e$	

^a249 explicitly correlated functions in elliptical coordinates.^bLinear Gaussian-type geminal calculation.^c900 random-tempered Gaussian-type geminals.^dVariational Monte Carlo.^eDiffusion Monte Carlo.

TABLE VII. The energy of the $\text{H}_2 B^1\Sigma_u^+$ state (in hartrees). N is the number of configurations. The internuclear distance is 1.4 a.u. Orthogonality with the ground state is not imposed.

N	Variance functional (#1)	Conroy functional (#5)	Local-energy functional (#10)
64 000	$-0.717\,500\,3 \pm 0.000\,322\,5$	$-0.686\,894\,9 \pm 0.000\,587\,9$	$-0.687\,962\,9 \pm 0.000\,680\,6$
256 000	$-0.717\,828\,5 \pm 0.000\,173\,2$	$-0.686\,971\,9 \pm 0.000\,311\,0$	$-0.688\,304\,6 \pm 0.000\,335\,7$
Kolos and Wolniewicz (Ref. 35)		$-0.705\,772\,684^a$	

^a249 explicitly correlated functions in elliptical coordinates.

equilibrium configuration, which has D_{3h} symmetry and a bond distance of 1.65 a.u. Our guiding function for this system is

$$g(r) = 12.371r^2e^{-2r} \quad \text{when } r = [0,1],$$

$$= \text{maximum } (12.371r^2e^{-2r}, 91.410r^2e^{-4r})$$

$$\text{when } r = [1,100]. \quad (17)$$

In this case the behavior at small distances is the same as in H_2 . At larger distances, however, we take the function e^{-4r} which is the form for a system with a charge of $+2$, i.e., one electron shielding the three hydrogen atoms. We do not include the r^2 in this case so as to cover more of the interesting intermediate space and not the region far away. The constants 12.371 and 91.410 were chosen so as to sample these two regions in a 1:1 ratio. This form is the result of some experimentation and should not be considered ideal. For our trial wave function we took

$$\Psi_t = (1 + P_{12}) \sum_k a_k \{ e^{-\alpha r_{1A} - \beta r_{2B}} r_{12}^{b_k} (r_{1A} - r_{1B})^{c_k} \\ \times (r_{1A} + r_{1B})^{d_k} (r_{2A} - r_{2B})^{e_k} (r_{2A} + r_{2B})^{f_k} \\ \times (r_{1C} - r_{2C})^{g_k} (r_{1C} + r_{2C})^{h_k} + [A \leftrightarrow C] + [B \leftrightarrow C] \}, \quad (18)$$

where A , B , and C are the positions of the three nuclei. This form is a three-atom generalization of Eq. (16) and again should not be considered ideal. All values up to $b_k + c_k + d_k + e_k + f_k + g_k + h_k = 3$ were selected. This resulted in a total of 122 variational parameters which were optimized with 4 000 configurations. E_{in} was taken to be -1.344 hartree. In Table VIII we show that for this system, as for the H_2 ground state, both the variance functional (#1) and the Conroy functional (#5) produce very simi-

lar results and within their error these energies agree with the best results from the literature.^{34,37,38} Since this state is more diffuse than the H_2 ground state, it is not surprising that the same cusp-dominated behavior is also found here.

V. EXCITED STATE CALCULATIONS

In any straightforward Monte Carlo calculation of excited states there is always the possibility that the final result will not be a pure excited state but rather some linear combination of ground and excited states. This problem can be eliminated by enforcing explicit orthogonality between states during the optimization process or by performing a Rayleigh–Ritz calculation afterwards just as in a traditional basis set calculation. The Monte Carlo analog of the Rayleigh–Ritz procedure constructs an improved trial wave function from a sum of trial wave functions calculated in the usual manner

$$\Psi_R = \sum_k c_k \Psi_k. \quad (19)$$

The Rayleigh–Ritz energy is found by finding the values of c_k which diagonalize

$$E_R = \langle \Psi_R H \Psi_R \rangle / \langle \Psi_R \Psi_R \rangle. \quad (20)$$

The matrix elements $\langle \Psi_i H \Psi_j \rangle$ and $\langle \Psi_i \Psi_j \rangle$ are evaluated by Monte Carlo integration and so their cost is only slightly more expensive than one energy evaluation.³⁹ The diagonalization of Eq. (20) produces a true variational energy for the excited states together with an improved wave function. Following the usual method of calculating errors in functions which contain terms with statistically independent errors,⁴⁰ the statistical error associated with E_R is given by the expression

TABLE VIII. The energy of the H_3^+ ground state (in hartrees). N is the number of configurations. The system is in an equilateral triangle configuration with $R = 1.65$ a.u.

N	Variance functional (#1)	Conroy functional (#5)	Local-energy functional (#10)
16 000	$-1.343\,586\,5 \pm 0.000\,240\,0$	$-1.343\,651\,0 \pm 0.000\,258\,6$	$-1.343\,193\,4 \pm 0.000\,326\,0$
64 000	$-1.343\,660\,7 \pm 0.000\,122\,7$	$-1.343\,722\,2 \pm 0.000\,130\,3$	$-1.342\,887\,3 \pm 0.000\,153\,9$
256 000	$-1.343\,762\,7 \pm 0.000\,061\,8$	$-1.343\,819\,8 \pm 0.000\,065\,9$	$-1.343\,085\,7 \pm 0.000\,078\,2$
Frye <i>et al.</i> (Ref. 37)		$-1.343\,827\,9^a$	
Alexander <i>et al.</i> (Ref. 34)		$-1.343\,822\,0^b$	
Traynor and Anderson (Ref. 38)		$-1.343\,85(5)^c$	

^aLinear Gaussian-type geminal calculation.

^b700 random-tempered Gaussian-type geminals.

^cDiffusion Monte Carlo.

$$\sigma_R^2 = \sum_i (E_R - E_i)^2, \quad (21)$$

where E_i is the energy obtained from the Rayleigh–Ritz procedure when one of the matrix elements in either the Hamiltonian or the overlap is changed by its error. The error of each of these matrix elements is most easily calculated using the same procedure used in Appendix A, i.e., removing the i th configuration. Because the trial wave functions are normalized, the $\langle \Psi_i^2 \rangle$ matrix elements have no error. The square of the error in the $\langle \Psi_i \Psi_j \rangle$ matrix elements, however, is given by

$$\delta \mathbf{S}_{ij}^2 + (1/4) \mathbf{S}_{ij}^2 [\delta \mathbf{S}_{ii}^2 + \delta \mathbf{S}_{jj}^2], \quad (22)$$

where

$$\delta \mathbf{S}_{ii}^2 = [\langle \Psi_i^2 \rangle - \langle \Psi_i \rangle^2] / \langle \Psi_i^2 \rangle, \quad (23)$$

$$\delta \mathbf{S}_{ij}^2 = [\langle (\Psi_i \Psi_j)^2 \rangle - \langle \Psi_i \Psi_j \rangle^2 / N] / [\langle \Psi_i^2 \rangle \langle \Psi_j^2 \rangle], \quad (24)$$

and

$$\mathbf{S}_{ij}^2 = \langle \Psi_i \Psi_j \rangle^2 / [\langle \Psi_i^2 \rangle \langle \Psi_j^2 \rangle]. \quad (25)$$

Similarly, the square of the error in the $\langle \Psi_i H \Psi_j \rangle$ matrix elements (for all values of i and j) is given by

$$\begin{aligned} & [(1/4) E_{ij}^2 \{ \mathbf{AA}_{iiii} / \langle \Psi_i^2 \rangle^2 + 2 \mathbf{AA}_{ijij} / (\langle \Psi_i^2 \rangle \langle \Psi_j^2 \rangle) \\ & + \mathbf{AA}_{jjjj} / \langle \Psi_j^2 \rangle^2 \} - E_{ij} \{ (\mathbf{AH}_{ijij} + \mathbf{AH}_{jiij}) / \langle \Psi_i^2 \rangle \\ & + (\mathbf{AH}_{jjji} + \mathbf{AH}_{ijjj}) / \langle \Psi_j^2 \rangle \} + (\mathbf{HH}_{ijij} + \mathbf{HH}_{jjji} \\ & + 2 \mathbf{HH}_{ijji})] / (4 \langle \Psi_i^2 \rangle \langle \Psi_j^2 \rangle), \end{aligned} \quad (26)$$

where

$$\mathbf{HH}_{ijkl} = \langle \Psi_i \Psi_j H \Psi_k H \Psi_l \rangle, \quad (27)$$

$$\mathbf{AH}_{ijkl} = \langle \Psi_i \Psi_j \Psi_k H \Psi_l \rangle, \quad (28)$$

$$\mathbf{AA}_{ijkl} = \langle \Psi_i \Psi_j \Psi_k \Psi_l \rangle, \quad (29)$$

$$E_{ij} = [E_{in}(\text{state } i) + E_{in}(\text{state } j)] \langle \Psi_i \Psi_j \rangle. \quad (30)$$

In Table IX we present our Rayleigh–Ritz results. These values were calculated using only the ground and excited state wave functions discussed in the previous section. For helium the matrix elements were computed using 1 024 000 configurations. In this system the Rayleigh–Ritz procedure raised the excited-state energy from the variance (#1) and local-energy (#10) functionals slightly and from the Conroy functional (#5) by a negligible amount. This indicates that all our trial wave functions contained only a

small ground state contamination. Within their errors all of the ground state energies are the same as before, although the variance functional (#1) appears to have improved slightly. For H_2 the appropriate matrix elements were computed using 256 000 configurations. In this case the Rayleigh–Ritz procedure corrects the clearly unphysical excited-state result produced by the variance functional (#1) and leaves all other calculations unchanged within their errors.

VI. CONCLUSIONS

We have described the biased-selection method and shown how it can be applied to both atoms and molecules. Our discussion includes a derivation of the variance of a Monte Carlo calculation when weighted data is used. Calculations which use some other definition do not correctly estimate their statistical error. We have shown that the biased-selection method differs from the Markov-chain approach in two important respects. The first is that the guiding function differs from the wave function. As shown in Appendix C, this feature can be used to conceptually connect the biased-selection method to more traditional numerical integration methods. In addition, it allows us to more intensely sample important regions of space (e.g., the cusps) by adjusting only the sampling function without changing the form of the wave function. The second difference is that each configuration in the biased-selection method is statistically independent. This makes our error estimates more valid than those of the Markov-chain approach where each configuration differs only slightly from the preceding one. It also prevents the parameter optimization from yielding very different results simply by using a different number of configurations.

Using a number of simple systems as examples, we have shown that the choice of optimization functional can have a significant influence on the accuracy of variational Monte Carlo calculations. We also noticed that the variance functional (#1) is sensitive to the weighting but that both the Conroy functional (#5) and the local-energy functional (#10) are not. This feature may be useful in some applications. Had we chosen to sample the cusps in the weighting, for instance, we believe that the energies produced by the variance functional (#1) and the Conroy functional (#5) would become closer to one another. Because we have used relatively simple guiding functions and trial wave functions which do not satisfy the cusp conditions, the Conroy func-

TABLE IX. Rayleigh–Ritz calculations. Energies are in hartrees.

	Variance functional (#1)	Conroy functional (#5)	Local-energy functional (#10)
He 1 ¹ S	−2.903 720 1 ± 0.000 015 8	−2.903 734 1 ± 0.000 013 5	−2.903 734 4 ± 0.000 013 3
He 2 ¹ S	−2.145 851 1 ± 0.000 018 3	−2.145 782 1 ± 0.000 010 8	−2.145 850 9 ± 0.000 015 0
He literature	−2.903 724 377 034 118 4 (Ref. 28) −2.145 974 046 054 143 (Ref. 23)		
H ₂ X ¹ Σ _g ⁺	−1.174 445 2 ± 0.000 026 0	−1.174 453 1 ± 0.000 026 0	−1.174 174 4 ± 0.000 045 0
H ₂ B ¹ Σ _u ⁺	−0.673 581 5 ± 0.000 791 3	−0.686 964 0 ± 0.000 311 2	−0.688 261 3 ± 0.000 337 0
H ₂ literature	−1.174 475 668 (Ref. 32) −0.705 772 684 (Ref. 35)		

tional (#5) produced the most accurate eigenvalue in almost every system we examined. Under these conditions this functional should be much more accurate than the others when larger, more complex systems are examined.

Although the major emphasis of this work has been to examine the behavior of various optimization functionals, our results are in reasonable agreement with previous conventional and Monte Carlo calculations. In addition, our results show that any straightforward calculation of an excited state produces a wave function which includes a contribution from the ground state and other lower states unless some form of explicit orthogonality has been included. In most cases this mixing may be small and the results may appear reliable but as our calculation of the first excited singlet ($B^1\Sigma_u^+$) state of H_2 illustrates, it is also possible to obtain nonphysical results. To eliminate this problem we have investigated the Monte Carlo analog of the Rayleigh–Ritz procedure. This method produces both accurate eigenvalue and variance estimates of excited states.

In this paper we have made no attempt to obtain the most accurate energy for any of the systems examined. Since the major part of the error in our calculations comes from the electron–electron cusp, one possible improvement would be to change the guiding function so as to more heavily sample this region. The other possibility would be to construct trial wave functions which satisfy the cusp condition. Our experience suggests, however, that any restriction on the form of the trial wave function tends to raise both the energy and the error per configuration unless this form is sufficiently flexible. This disadvantage will more than offset the advantages of allowing the configurations to sample other regions of space. Based on the error per configuration of our current trial wave functions, it should be much easier to improve both our energies and variances by using better trial wave functions rather than simply adding many more Hylleraas (or Hylleraas-like) basis functions, optimizing over more configurations and then beating down the error with still more configurations. We believe that the problem of choosing good trial wave function forms remains, at present, the largest hurdle to doing very accurate variational Monte Carlo calculations.

ACKNOWLEDGMENTS

We would like to thank Dr. Cyrus Umrigar and Dr. P. J. Reynolds for a number of helpful comments and the staff of the Northeast Regional Data Center for their support in running our program on the University of Florida IBM 3090. This work has been supported by the National Science Foundation (NSF) Grant No. PHY-8608155 and by a grant from the NSF for the Institute for Theoretical Atomic and Molecular Physics at the Harvard–Smithsonian Astrophysical Observatory and at Harvard University. One of us (J. D. M.) would like to thank Dr. Dudley Herschbach for the hospitality shown him at the Harvard Department of Chemistry.

APPENDIX A: CALCULATING THE STANDARD DEVIATION FORMULA

The difference between the expectation value of the operator A

$$\langle A\{\mathbf{x}_i\} \rangle = \sum_i [\Psi_i A \Psi_i / w_i] / \sum_i [\Psi_i^2 / w_i] \quad (A1)$$

and its exact value can be expanded as

$$\langle A\{\mathbf{x}_i\} \rangle - A_{\text{exact}} = \sum_i \frac{\partial \langle A \rangle}{\partial \mathbf{x}_i} e_i = \sum_i \delta_i, \quad (A2)$$

where e_i is the variance of the i th point. Assuming that δ_i and δ_j are uncorrelated for $i \neq j$, the standard deviation can thus be written as

$$\sigma^2 = \sum_i \delta_i^2. \quad (A3)$$

When $\partial \langle A \rangle / \partial \mathbf{x}_i$ is calculated in terms of the matrix being inverted, Eq. (A3) can also be used to derive the standard least square fit formula.²⁰ For Monte Carlo calculations though, it is more convenient to estimate δ_i directly by equating it to the deviation in $\langle A \rangle$ caused by removing the i th configuration, i.e.,

$$\begin{aligned} \delta_i &= \sum_{j \neq i} [\Psi_j A \Psi_j / w_j] / \sum_{j \neq i} [\Psi_j^2 / w_j] \\ &\quad - \sum_k [\Psi_k A \Psi_k / w_k] / \sum_k [\Psi_k^2 / w_k], \\ &= \frac{\{\sum_k [\Psi_k A \Psi_k / w_k] \Psi_i^2 / w_i - \sum_k [\Psi_k^2 / w_k] \Psi_i A \Psi_i / w_i\}}{\{(\sum_k [\Psi_k^2 / w_k])^2 + \sum_k [\Psi_k^2 / w_k] \Psi_i^2 / w_i\}}, \end{aligned} \quad (A4)$$

and thus

$$\sigma^2 = \sum_i [\Psi_i A \Psi_i / w_i - \langle A \rangle \Psi_i^2 / w_i]^2 / \left\{ \sum_i [\Psi_i^2 / w_i] \right\}^2. \quad (A5)$$

This expression agrees exactly with Conroy's definition of the variance when the integrand is taken to be $\Psi_i H \Psi_i / w_i$ [Eq. (14) in Ref. 10] but differs from his optimization function [Eq. (2) in Ref. 10]. Equation (A5) can also be rewritten in terms of some input value A_{in}

$$\begin{aligned} \sigma^2 &= \sum_i [\Psi_i A \Psi_i / w_i - (\langle A \rangle + A_{in} - A_{in}) \Psi_i^2 / w_i]^2 / \left\{ \sum_i [\Psi_i^2 / w_i] \right\}^2 \\ &= \sum_i [\Psi_i (A - A_{in}) \Psi_i / w_i]^2 / \left\{ \sum_i [\Psi_i^2 / w_i] \right\}^2 \\ &\quad - \sum_i [(A_{in} - \langle A \rangle) \Psi_i^2 / w_i]^2 / \left\{ \sum_i [\Psi_i^2 / w_i] \right\}^2. \end{aligned} \quad (A6)$$

Unless A_{in} is quite far from $\langle A \rangle$, the second term in this expression will be very small and so can be ignored. This is in apparent contradiction to Ref. 18 where Sun *et al.* claim that A_{in} should be set slightly above the exact result to obtain the best results and Ref. 16 where Umrigar *et al.* claim that A_{in} should be slightly below the exact result. Since most of our error is caused by the electron–electron cusp, setting our reference energy slightly low would cause our optimization to ignore many of the effects of this cusp. This would give us a better energy but a higher variance. Similarly if we were to set our reference energy slightly high we would get a worse energy but a lower variance. It should be stressed, however,

that our results depend only weakly on the value of A_{in} and that this behavior will be true only when the potential is repulsive. If we were to take this cusp into account when choosing our configurations (much like we do for the electron–nuclear potential) or if the trial wave function were to satisfy the cusp condition then this dependence would be even weaker.

APPENDIX B: REMOVING SINGULARITIES FROM THE HAMILTONIAN

In all Monte Carlo methods the straightforward integration of functions with singularities presents a number of problems because as the number of configurations increases, the probability of getting one which is close to the singularity also increases. In the biased-selection method the appearance of a configuration close to a singularity causes the variance to suddenly increase. For the Hamiltonian in Eq. (2) this behavior can be caused only by the electron–electron potential since the electron–nuclear potential has already been “neutralized” by our method of selecting the configurations. The problems with all other singularities can be easily corrected, however, by replacing each singularity with the constant which yields the unchanged value for the integral, assuming Ψ to be constant, when r is less than some small value. For the Coulomb potential this means replacing r with $2\epsilon/3$ when $r < \epsilon$ (we typically set $\epsilon = 0.01$ a.u.), i.e.,

$$\int V(r)\Psi^2 d^3r = 4\pi\Psi^2 \int_0^\epsilon \frac{r^2}{r} dr = 4\pi\Psi^2 \int_0^\epsilon \frac{3}{2\epsilon} r^2 dr. \quad (B1)$$

In order to preserve the delicate cancellation which makes $H\Psi/\Psi$ nearly constant, this replacement must be done in both the potential and in the Laplacian. A similar procedure has been implemented in other Monte Carlo calculations.^{41,42}

APPENDIX C: THE “GAUSSIAN BIASED-SELECTION METHOD”

In this paper we have described how integrals can be estimated using Monte Carlo integration points chosen by the biased-selection method. One advantage of our formalism is that it also be used with more traditional quadrature procedures. For example, the norm of a two electron atom can be written using the Hylleraas coordinates $s = r_1 + r_2$, $t = r_1 - r_2$, and $u = r_{12}$ as⁴³

$$\text{Norm} = \pi^2 \int_0^\infty ds \int_0^s du \int_{-u}^u dt u(s^2 - t^2)^2 \Psi(s, t, u)^2. \quad (C1)$$

Since $\Psi^2 \rightarrow e^{-(\alpha+\beta)s}$ as $s \rightarrow \infty$, it is appropriate to evaluate the integral over s using a Gauss–Laguerre quadrature and those over u and t using a Gauss–Legendre quadrature. This yields

$$\pi^2 \sum_{i=1}^{N_s} \sum_{j=1}^{N_u} \sum_{k=1}^{N_t} \Psi(s_i, u_j, t_k)^2 e^{-s_i} s_i u_j^2 \times (s_i^2 - t_k^2) \omega_i \omega_j \omega_k / (2\alpha + 2\beta), \quad (C2)$$

where $s_i = \xi_i / (\alpha + \beta)$, $u_j = s_i (1 + x_j) / 2$ and $t_k = u_j x_k$

and where ξ_i and ω_i are the Gauss–Laguerre quadrature points and weights and x_j and ω_j (and x_k and ω_k) are the Gauss–Legendre points and weights. For convenience we take the ξ ’s to range from 0 to ∞ and the x ’s from -1 to 1 . Eq. (C2) can be easily converted into the biased-selection expression

$$\sum_{m=1}^{N_s N_u N_t} \Psi(\mathbf{x}_m)^2 / w(\mathbf{x}_m). \quad (C3)$$

In such a calculation the configurations are generated by taking all possible combinations of s_i , u_j and t_k and from these computing the appropriate components of \mathbf{r}_1 and \mathbf{r}_2 . The weight is simply taken to be

$$w(\mathbf{x}_m) = (2\alpha + 2\beta) [\pi^2 e^{-s_i} s_i u_j^2 (s_i^2 - t_k^2) \omega_i \omega_j \omega_k]. \quad (C4)$$

With this collection of configurations and weights we can use the same code which produced our Monte Carlo results but avoid the slow convergence of such calculations.

To illustrate the capabilities of the Gaussian biased-selection method we now calculate the energy minimum for the helium 2^3S state. For the wave function described in Sec. III the integral over t involves terms up to $e^{(\alpha-\beta)t^{12}}$, the integral over u involves terms up to u^{14} and the integral over s involves terms up to s^{16} . This implies that N_t must be at least 6, N_u at least 7 and N_s at least 8. For the ground state where $\alpha = \beta$ these values would probably be sufficient but the triplet has $\alpha - \beta = 0.7$ and so a slightly larger set is needed. We found that $N_s = 9$, $N_u = 10$, and $N_t = 10$ to be just a little too small for the eight digit accuracy required here but that $N_s = 11$, $N_u = 10$, and $N_t = 10$ seems to be enough. To avoid any possibility of numerical error, however, we optimized all 36 coefficients in the helium triplet wavefunction with $N_s = 15$, $N_u = 16$, and $N_t = 16$ and then checked them with $N_s = 18$, $N_u = 18$, and $N_t = 18$. The final value for the energy in this parameter space is $-2.175\,225\,3$ hartree. To obtain a value closer to the exact result a larger basis set is needed. In addition to the energy we also calculated N_{eff} . This value was less than $0.1N$ which further supports our observation that it is not desirable to set $N_{\text{eff}} = N$ as does the Markov chain.

¹ J. W. Moskowitz, K. E. Schmidt, M. A. Lee, and M. H. Kalos, *J. Chem. Phys.* **77**, 349 (1982); **85**, 2868 (1985).

² K. E. Schmidt and J. W. Moskowitz, *J. Stat. Phys.* **43**, 1027 (1986).

³ P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, Jr., *J. Chem. Phys.* **77**, 5593 (1982).

⁴ P. J. Reynolds, M. Dupuis, and W. A. Lester, Jr., *J. Chem. Phys.* **82**, 1983 (1985).

⁵ B. H. Wells, in *Methods in Computational Chemistry*, edited by Stephen Wilson (Plenum, New York, 1987), Vol. 1.

⁶ R. N. Barnett, P. J. Reynolds, and W. A. Lester, Jr., *J. Chem. Phys.* **84**, 4992 (1986).

⁷ R. N. Barnett, P. J. Reynolds, and W. A. Lester, Jr., *J. Chem. Phys.* **82**, 2700 (1985); **91**, 2004 (1987).

⁸ R. J. Harrison and N. C. Handy, *Chem. Phys. Lett.* **113**, 257 (1985).

⁹ D. R. Garmer and J. B. Anderson, *J. Chem. Phys.* **86**, 4025, 7237 (1987).

¹⁰ H. Conroy, *J. Chem. Phys.* **41**, 1327, 1331, 1336, 1341 (1964).

¹¹ R. L. Coldwell, *Int. J. Quantum Chem. S* **11**, 215 (1977).

¹² R. L. Coldwell and R. E. Lowther, *Int. J. Quantum Chem. S* **12**, 329 (1978).

¹³ R. L. Coldwell, *Int. J. Quantum Chem. S* **13**, 705 (1979).

¹⁴ K. A. Wilson and R. L. Coldwell, *J. Comp. Phys.* **61**, 483 (1985).

- ¹⁵ M. C. Huang, R. L. Coldwell, and M. Katoot, *Nucl. Phys. B* **309**, 733 (1988).
- ¹⁶ C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, *Phys. Rev. Lett.* **60**, 1719 (1988).
- ¹⁷ S. A. Alexander and R. L. Coldwell, *Int. J. Quantum Chem. S* **23**, 169 (1989).
- ¹⁸ Z. Sun, S. Y. Huang, R. N. Barnett, and W. A. Lester, Jr., *J. Chem. Phys.* **93**, 3326 (1990).
- ¹⁹ When the configurations are chosen from the wave function, i.e., $w(\mathbf{x}_i) = \Psi_i(\mathbf{x}_i)^2$ as in the Markov chain method, Eqs. (3) and (4) differ only by a factor of N .
- ²⁰ P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 1969), p. 44.
- ²¹ J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (Methuen, London, 1964).
- ²² See, for example, E. A. Hylleraas, *Z. Phys.* **54**, 347 (1929).
- ²³ G. F. W. Drake, *Nucl. Instrum Methods B* **31**, 7 (1988).
- ²⁴ C. L. Pekeris, *Phys. Rev.* **115**, 1216 (1959).
- ²⁵ A. J. Thakkar and V. H. Smith, Jr., *Phys. Rev. A* **15**, 16 (1977).
- ²⁶ S. Rybak, K. Szalewicz, and B. Jeziorski, *J. Chem. Phys.* **91**, 4779 (1989).
- ²⁷ A. J. Thakkar and V. H. Smith, Jr., *Phys. Rev. A* **15**, 1 (1977).
- ²⁸ J. D. Baker, D. E. Freund, R. N. Hill, and J. D. Morgan III, *Phys. Rev. A* **41**, 1247 (1990).
- ²⁹ D. W. Skinner, J. W. Moskowitz, M. A. Lee, P. A. Whitlock, and K. E. Schmidt, *J. Chem. Phys.* **83**, 4668 (1985).
- ³⁰ C. L. Pekeris, *Phys. Rev.* **127**, 509 (1962).
- ³¹ H. M. James and A. S. Coolidge, *J. Chem. Phys.* **1**, 825 (1933); **3**, 129 (1935); see also, L. Pauling and E. B. Wilson, *Introduction to Quantum Mechanics* (McGraw-Hill, New York, 1935), p. 349.
- ³² W. Kolos, K. Szalewicz, and H. J. Monkhorst, *J. Chem. Phys.* **84**, 3278 (1986).
- ³³ D. Frye, G. C. Lie, and E. Clementi, *J. Chem. Phys.* **91**, 2369 (1989).
- ³⁴ S. A. Alexander, H. J. Monkhorst, R. Roeland, and K. Szalewicz, *J. Chem. Phys.* **93**, 4230 (1990).
- ³⁵ W. Kolos and L. Wolniewicz, *Can. J. Phys.* **53**, 2189 (1975).
- ³⁶ This state has been previously examined using diffusion Monte Carlo (but at a different geometry) by R. M. Grimes, B. L. Hammond, P. J. Reynolds, and W. A. Lester, Jr., *J. Chem. Phys.* **85**, 4749 (1986) and S. Y. Huang, Z. Sun, and W. A. Lester, Jr., *ibid.* **92**, 597 (1990).
- ³⁷ D. Frye, A. Preiskorn, G. C. Lie, and E. Clementi, *J. Chem. Phys.* **92**, 4948 (1990).
- ³⁸ C. A. Traynor and J. B. Anderson, *Chem. Phys. Lett.* **147**, 389 (1988).
- ³⁹ This procedure is similar to the one recently used to compute vibrational excited states by B. Bernu, D. M. Ceperley, and W. A. Lester, Jr., *J. Chem. Phys.* **93**, 552 (1990).
- ⁴⁰ R. L. Coldwell and G. J. Bamford *The Theory and Operation of Spectral Analysis Using ROBFIT* (AIP, New York, 1991).
- ⁴¹ J. Vrbik, D. A. Legare, and S. M. Rothstein, *J. Chem. Phys.* **92**, 1221 (1990).
- ⁴² C. A. Traynor, J. B. Anderson, and B. M. Boghosian, *J. Chem. Phys.* **94**, 3657 (1991).
- ⁴³ B. L. Moisewitsch, *Variational Principles* (Wiley, New York, 1966), p. 179.