Quantum Monte Carlo for molecules: Green’s function and nodal release

D. M. Ceperley and B. J. Alder
Lawrence Livermore National Laboratory, University of California, Livermore, California 94550

(Received 4 June 1984; accepted 1 August 1984)

A random walk algorithm is presented which exactly calculates the properties of a many-electron system. For that purpose both the Green’s function Monte Carlo method and nodal relaxation have been employed and both are described in detail. The scheme is applied to several small molecules, \{He, LiH, Li2, H2O\} and with modest computational effort and simple importance functions, ground state energies are obtained which agree with experimental energies within statistical error bars. The small energy decrease due to nodal release is accurately evaluated by a difference method.

I. INTRODUCTION

Recently Monte Carlo methods have been applied to calculate electron correlations in molecules. Using these methods on the electron gas it has been demonstrated that it is possible to obtain as exact a numerical estimate of the ground state energy of a many particle system as computer time permits. For Fermion systems, a serious difficulty in achieving this aim is due to the wave function having both positive and negative regions so that the underlying probability distribution is not positive. Previously this difficulty when applied to molecular systems has been circumvented by the fixed-node approximation in which some approximate nodal surface separating the positive and negative regions is assumed. Within this nodal surface, considered a boundary, the problem then reduces to a boson ground state problem with a well defined probability distribution. The energy so obtained can be shown to be an upper bound to the true energy. Furthermore, this fixed-node energy was found to be quite accurate, exceeding the best previous variational calculations. In this paper we remove this fixed-node approximation for molecules in a similar way as was done for the electron gas. Furthermore, we also remove the previous numerical approximation of using a finite time step in the diffusion process by using Green’s function Monte Carlo so that the ultimate accuracy of the method can be demonstrated. Removal of the finite time step error is particularly important since this error is frequently comparable to the one caused by the fixed-node approximation.

For the investigation of chemical systems, a high absolute accuracy must be attained. An accuracy of at least $10^{-3}$ hartrees is necessary to determine chemical bonding energies (which are typically $10^{-4}$ hartrees) to within 1%. Even more stringent requirements are frequently needed to decide between various bonding alternatives in a molecule. Further improvements in the statistical accuracies of absolute energies depend on finding better trial functions to be used as importance functions, better Fermion methods, and longer computer runs.

The outline of the paper is as follows: In the next section the Green’s function method is derived and described in detail. We then briefly describe its application to the fixed-node calculations and, in more detail, to the nodal relaxation calculation. Subsequently, description of the choices for the trial and guiding functions leads to a discussion of the results obtained for several small molecules, namely LiH, Li2, H2O, and H3.

Our results demonstrate that a random walk method can determine ab initio the energies of small molecules to an accuracy comparable to the best existing configuration interaction calculations. However, in contrast to those results, the Monte Carlo energies come with a mathematically well defined error estimate. Because the two methods are so different it is to be expected that each will have its own area of applicability but the overlap between the two provides a much needed and reassuring check on the numerical precision of each method.

II. GREEN’S FUNCTION MONTE CARLO (GFMC)

A method that samples the exact ground state wave function for charged systems will be described in this section. The method eliminates the finite time step truncation error of diffusion Monte Carlo (DMC). For simplicity, a discussion of the complications introduced by Fermi statistics is postponed to the next section, where we assume the particles are distinguishable.

In the original GFMC method, called here domain GFMC, the move of a single electron is restricted within a sphere whose radius is smaller than half the distance to the nearest neighboring electron. For molecular systems this restriction of the step size leads to an inefficient, though exact, algorithm. In a previous paper, one of the authors has already discussed how to remove this limitation and this algorithm is rederived here employing a different presentation and notation.

In quantum Monte Carlo calculations a guiding function is sometimes called a trial or an importance function \( \psi_c \) is introduced to increase the efficiency of the process. The exact ground state wave function \( \psi(\mathbf{R}) \) is obtained as a steady state solution for the probability density in configuration space of \( \rho(\mathbf{R}) = \psi_c(\mathbf{R}) \psi(\mathbf{R}) \), where \( \mathbf{R} \) is the \( 3N \) vector of electron coordinates. In DMC, the configurations advance from \( t \) to \( t + \tau \) with three processes: diffusion, drift, and branching. The evolution of the configuration is [Eq. (16) and (17) of Ref. 1]

\[
\rho(\mathbf{R}, t + \tau) = \int d\mathbf{R}' \rho_d(\mathbf{R}, \mathbf{R}', t) \rho(\mathbf{R}', t) \times \exp \left[ - \left( E_L(\mathbf{R}') - E_{\tau}\right) \tau \right]
\]
The exponential in the integrand describes the branching process, which depends on the difference between the local energy $E_L(R) = \psi_0^{-1}(R) \hat{H}\psi_0(R)$ and the trial energy $E_T$, where $H$ is the nonrelativistic Hamiltonian. The diffusion and drift from $R'$ to $R$ is described by the normalized Gaussian distribution

$$\rho_D(R,R';\tau) = (4\pi D\tau)^{-N/2} \times \exp \left\{ -\frac{1}{2} (R - R')^2/4D\tau \right\}$$

in which the drift is

$$F(R') = 2D\psi_0^{-1}(R) \nabla \psi_0(R')$$

where

$$D = \hbar^2/2m.$$  

The steady state distribution [i.e., when $f(R,t+\tau) = f(R,t)$] in DMC is the exact ground state wave function only in the limit that $\tau$ vanishes. Any finite time step $\tau$ will lead to a systematic error in the estimate of the ground state energy.

In order to make this diffusion method free of time step errors, the exact density matrix $\rho(R,R';\tau)$ is used in place of $\rho_D$. The density matrix is defined in operator notation as $\rho = \exp \{ -H\tau \}$ and obeys the Bloch equation in coordinate space

$$\frac{\partial \rho}{\partial t}(R,R';t) = -H\rho(R,R';t)$$

with

$$\rho(R,R';0) = \delta(R - R').$$

As usual, the ground state wave function would be the steady state solution of the equation

$$\phi(R,t+\tau) = \int dR' \rho(R,R';\tau) e^{i\delta t/\tau} \phi(R',t).$$

Applying importance sampling, that is multiplying by $\psi_0$, leads to the evolution of the probability distribution $f$:

$$f(R,t+\tau) = \int dR' \rho(R,R';\tau) e^{i\delta t/\tau} \frac{\psi_0(R')}{\psi_0(R)} f(R',t).$$

Even though the exact density matrix is not known analytically, it is not difficult to sample it exactly. To recover diffusion Monte Carlo, one makes the semiclassical approximation to $\rho$, expands $\psi_0$ about $\psi_0$ and arrives at Eq. (1).

The essential step in the GFMC method is to relate the exact density matrix to an approximate trial density matrix via a linear integral equation (a Fredholm equation of the second kind). This equation can be solved by successive substitutions, thereby expanding an expansion of the exact density matrix in terms of the trial density matrix. By stochastically sampling all terms in the series, the exact density matrix itself can be sampled. Then $f(R,t+\tau)$ can also be sampled, even for finite $\tau$. The general procedure is derived from the von Neumann–Ulam technique of the Monte Carlo solution of linear integral equations, but many of the specific techniques for the quantum many-body problem were developed by Kalos.

Let us introduce a known trial density matrix $\rho_T(R,R',\beta)$ which is a differentiable function of $\beta$ and $R$ and for small time approaches $\rho_T(R,R',0) = \delta(R - R')$. It will then be shown that

$$\rho(T)(R,R';B) = \rho_T(R,R';B) + \int_0^\beta d\beta \int dR \rho(R,R';\beta - t)K(R,R';t).$$

The kernel $K(R,R';t)$ can be interpreted as a transition probability to an intermediate state, $R''$, between $R$ and $R'$, and is given in terms of the trial density matrix by

$$K(R,R';t) = -\frac{1}{\frac{\partial}{\partial t}} \rho_T(R'', R'; t),$$

where $H(R'')$ represents the Hamiltonian operator evaluated at $R''$.

The derivation of the integral equation is accomplished via the Bloch equation, which, since the density matrix is symmetric in $R$ and $R''$, can be rewritten by a change of variables as

$$\frac{\partial}{\partial t} \rho(T)(R,R';\beta - t) = H(R'') \rho(T)(R,R';\beta - t).$$

The following operations then lead to Eq. (6): (i) Multiply both sides by $\rho_T(R'', R'; t)$; (ii) integrate over $R''$; (iii) integrate over $t$ from 0 to $\beta$; (iv) the Hermitian operator $H$ on the right-hand side can be made to act to the left, on $\rho_T$, instead of to the right, on $\rho$. This leads to the $H\rho_T$ term of the kernel; (v) integrate the left-hand side by parts with respect to $t$; (vi) utilize the boundary conditions $\rho_T(R,R',0) = \delta(R - R')$ mentioned earlier. After regrouping, the integral equation results.

This equation can be solved iteratively by a stochastic process, provided the kernel and the trial density matrix are nonnegative everywhere and the kernel is sufficiently small. This, as will be shown shortly, can be arranged for Coulombic systems. The closer the trial density matrix is to the exact density matrix the faster the convergence becomes, since the kernel will vanish for the exact density matrix. The iteration process consists of assuming that $\rho_T$ is the first approximation to the density matrix and thus substituting $\rho_T$ for $\rho$ in the integral of Eq. (6). This leads on the left-hand side to a second approximation to $\rho$. Successive substitutions lead to an expansion of $\rho$ in terms of $\rho_T$, and the kernel is the contribution of the $(n+1)$th iteration relative to the nth. The Monte Carlo process consists of sampling all terms in this series with their proper contribution so that the sampling of $\rho$ will be exact.

For particles interacting with Coulombic forces, one can choose a positive $\rho_T$ such that the kernel is always positive. In fact, the procedure takes any exact density matrix $\rho_0$ for a given reference potential $V_0$ (practically, only the free particle and harmonic oscillator potentials are simple enough) and constructs a trial density matrix from this reference density matrix by correcting for the differences between the actual potential and $V_0$. Our trial density matrix is of the form

$$\rho(T)(R,R';\beta) = \rho_0(R,R';\beta) \exp \{ -U(R,R';\beta) \},$$

where

$$U(R,R'\beta) = \int_0^\beta \rho_T(R,R';t)[V(R'') - V_0(R'')] \rho_0(R,R';t) \rho_T(R,R';\beta - t),$$

and

$$\rho_0(R,R';\beta) = \int_0^{\beta-\tau} d\beta \rho(T)(R,R';\beta) \exp \{ -U(R,R';\beta) \}.\]
This formula is an approximation to the Feynman–Kac expression for the density matrix

$$\rho(R, R^{'}, \beta) = \rho_0(R, R^{'}, \beta) \text{exp} \left[ - \int_0^t dt \, V(t, R) \right],$$  \hspace{1cm} (11)

where the average is over all Brownian walks from $R$ to $R^{'},$ and $V(t)$ is the potential energy. The approximation of taking the average into the exponent is strictly valid only in the high temperature (small time step) limit, however, it is a considerable improvement on the previous GFMC procedure of ignoring the potential correction entirely and using $\rho(T) = \rho_0,$ the free particle density matrix. $V(t)$ is given by Eq. 2 with $F = 0.$ For a system of hard sphere particles, for which the domain GFMC method (which uses for $\rho_0$ the free particle density matrix) was developed,

it is necessary to restrict moving within domains such that the potential remains bounded. However, for the softer Coulomb potential a more accurate trial density matrix allows for larger moves.

It can be shown by straightforward, though tedious, manipulations that the kernel for this choice of $\rho(T)$ [i.e., Eq. (9)] is positive and has the simple form:

$$K_1(R, R^{'}, t) = D \sum_i (\mathbf{V}, U) \rho(T) \langle R, R^{'}, t \rangle.$$ \hspace{1cm} (12)

With the ideal gas choice for $\rho_0,$ the integrals in Eq. (10) can be carried out, being Gaussian, leading to a pairwise additive $U$ that is for the Coulomb potential a confluent hypergeometric function

$$U(R, R^{'}, \beta) = \sum_{r_1} \sum_{r_2} \beta \epsilon_2 \epsilon_1 \int_0^1 d\lambda \, W(\lambda, \lambda'),$$ \hspace{1cm} (13a)

where

$$W(\lambda, \lambda') = \frac{1}{X} \text{erf} \left( \frac{X}{4 \sqrt{4D\lambda (1 - \lambda')^2}} \right),$$ \hspace{1cm} (13b)

$$X(\lambda) = \left( 1 - \lambda \right) \tau_{e_2} + \lambda \tau_{e_1},$$ \hspace{1cm} (13c)

and $e_i$ is the charge of particle $i,$ $\mathbf{V}$ is the numerical evaluation of $U.$ The analytical evaluation of $U$ has been discussed in the Appendix of Ref. 7.

The actual numerical procedure involves some modifications of the integral equation, [Eq. (6)] to get it into a form more convenient for the calculation of the exact ground state. For that purpose, the density matrix weighted by $e^{E[R] - E[R']} = \psi(\beta, \epsilon_2 - 1/\beta) \psi(\beta, \epsilon_2) \rho(R, R^{'}, \beta)$ is required [see Eq. (5)]. Furthermore, we are not interested in evaluating $\int_0^t$ for a particular value of $t$ nor are we required to choose a constant value of the time step as previously in DMC. Instead it is convenient to obtain a time independent form of the equations by integrating over the time variable. We also multiply by an arbitrary time weighting function, chosen for convenience to be an exponential $e^{-\beta/\Delta}$ so that the average step is $\Delta.$ Thus, the function, properly normalized, which is used to propagate the configurations is the time independent Green's function

$$\hat{\rho}(R, R^{'}) = \int_0^\infty d\beta \, \rho(T) \psi(\beta, \epsilon_2 - 1/\beta) \psi(\beta, \epsilon_2) \rho(R, R^{'}, \beta) \psi(\beta, \epsilon_2) \psi(\beta, \epsilon_2 - 1/\beta).$$ \hspace{1cm} (14)

Note that $\hat{\rho}$ is the Laplace transform of $\rho$ in time and is important sampled in space, but depends on $\psi(\beta)$ in a simple way. This is important in the release-node algorithm. The integral equation [Eq. (6)] for the density matrix can then be transformed to the following integral equation for the Green's function:

$$\hat{\rho}(R, R^{'}) = \hat{\rho}(T) \rho(R, R^{'}) + \Delta \int dR'' \rho(R, R'') \hat{K}(R'', R^{'}) \rho(R^{'}, R^{'}) \psi(\beta, \epsilon_2 - 1/\beta).$$ \hspace{1cm} (15)

where $\hat{\rho}(T)$ and $\hat{K}$ are defined as $\hat{\rho}$ in Eq. (14) with $\rho(T)$ and $K,$ respectively, substituted for $\rho.$ This equation is obtained by changing the order of time integration and taking advantage of the "convenient" property of the exponential: $e^{-\beta/\Delta}$ $= e^{-\beta/\Delta} e^{-1/\beta} - 1/\beta.$ With this Green's function the evolution of the probability distribution is

$$f_{n+1}(R) = \int dR'' \hat{\rho}(R, R') f_n(R').$$ \hspace{1cm} (16)

The distribution $f_n$ is referred to as the $n$th generation.

The random walk method for the combined process of solving the integral equations [Eqs. (15)] and (16)] for the steady state $f,$ consists of first choosing $E, \psi, \rho(T)$ and selecting from a variational calculation (i.e., with respect to $\psi(\beta)$) an appropriate number of initial configurations, denoted $R'$, usually several hundred. The calculation then proceeds as follows: (i) Sample a time step from the distribution $\Delta + 1 e^{-\beta/\Delta}.$ (ii) Diffuse and drift all the coordinates of one of the initial configurations for that selected time step, as in DMC (i.e., with probability $\rho(D)$) arriving at a new point in phase space $R.$ (iii) Branch with multiplicity calculated from

$$m_D(R, R', t) = \psi(D) \rho(T) \rho(T, R'') \psi(D) \rho(T, R'').$$ \hspace{1cm} (17)

These configurations referred to as the direct ones, correspond to the first term in the integral equation. The reason for the appearance of $\rho(D)$ in the denominator is to correct the approximate $\rho(D)$ used in step (ii) to the approximate $\hat{\rho}$ by multiplying and dividing the integrand in Eq. (14) for $\hat{\rho}$ by $\rho(D).$ One can not sample $\rho(D)$ because its normalization is unknown. Normalization requires being able to integrate $\rho(T)$ over all space, which is equivalent to evaluation of a classical partition function [see Eq. (9)], generally not an easy task. This step differs from DMC only in changing the branching formula. Since branching is an integer process, and $m_D$ is not an integer, it must be converted to an integer in such a way that the mean value is not changed. This is done by adding a random number in $[0, 1]$ and then rounding off. The configurations so generated constitute part of the next generation. (At this stage it would be part of the first.) (iv) Generate intermediate configurations by branching with multiplicity

$$m_I(R, R', t) = \frac{K(R, R', t) m_D(R, R', t) \Delta}{\rho(T, R, t)}.$$ \hspace{1cm} (18)

This corresponds to the first iteration of Eq. (15) since the first approximation to $\hat{\rho}$, namely $\hat{\rho}(T),$ has been substituted in the integral. An intermediate point, is considered the same as an initial point in step (i). Note that these replicas are not counted in any averaging process, but, are only sources for new random walks. (v) Each of the intermediate configurations generated in step (iv), if any, undergo again all the previous steps, generating both further direct and intermediate configurations. The former would be added to the first generation as in step (iii), while the latter are again sources for
more direct and intermediate points. Of course the average value of \( m_I \) must be less than one, otherwise this process would never finish. When all intermediate points have been processed, \( \tilde{\rho}(R, R') \) has been correctly sampled. The additional direct configurations generated through the intermediate steps correct the Green's function \( \tilde{\rho}_T \) to \( \tilde{\rho} \). (v) When all of the original configurations, their intermediate configurations, the intermediate configurations generated by intermediate configurations, etc. have been processed, the first generation \( f_1(R) \) has been sampled and those resulting configurations become the source for the second generation. As in DMC, eventually the populations must be renormalized through random deletions and copying to restore it to its original size, in the process of which \( E_T \) can also be adjusted. This can lead to a bias if it is done too often \(^7\) and if the population is too small.

There is one final change which improves the efficiency. It can be shown that if \( \psi_0 \) is close to the ground state, the average of the sum of the direct and intermediate multiplicities will be close to unity. \(^5\) The fluctuations in branching can then be reduced by letting the creation of direct and intermediate points be exclusive events. The number of total copies made at each step is \( m_I + m_D \), and with probability \( m_D/(m_I + m_D) \) all of these copies are direct points; otherwise they are intermediate points.

Although from this lengthy description it might appear that the Green's function method is rather different from DMC, from the programming view the differences are not major. These differences are: (i) A variable time step sampled from an exponential distribution is used. This is clearly an advantage if the average time step is larger, since this allows quicker sampling of configuration space. For the small chemical molecules investigated here, it was indeed found that the average time step could be an order of magnitude larger than in DMC (ii) The branching formula is different from DMC. It can be shown that in the vanishing time step limit they become identical, but only if second order terms are included in the width of the Gaussian \( \rho_\rho \) (ii) The price to be paid for eliminating the finite time step is that \( \rho_T \) and \( K \) have to be calculated at each point and that intermediate points are created which do not contribute to averages. In practice this roughly doubles the computer time per step. (iv) In the DMC method \(^1\) electrons were moved one by one. Here the electrons must be moved simultaneously, which is, however, more efficient for small systems.

We conclude this section with a discussion of the calculation of the ground state energy. The energy may be calculated both from the average population change and the average local energy, although the variance of the two estimators is different. If \( f_n \) is the distribution in the \( n \)th generation, let us define the energy in the \( n \)th generation as

\[
E_n = \frac{\int dR \rho_\psi_0(R) \frac{\partial}{\partial R} f_n(R)}{\int dR f_n(R)}.
\]  

(19)

Since \( f_n \) converges to \( \phi_\psi_0 \), \( E_n \) converges to the ground state energy. The population of \( f_n \) is

\[
P_n = \int dR f_n(R).
\]  

(20)

Then it is easy to show using the properties of the Green's function \( \tilde{\rho} \) that

\[
E_n[P] = E_T + \frac{1}{\Delta} \left( P_{n-1} - P_n \right).
\]  

(21)

This is the growth estimator for the energy, since it relates the ground state energy to the trial energy and the population change. Thus the trial energy is adjusted correctly if the population does not change.

The local energy estimate follows from the definition (19), by letting \( H \) act to the left:

\[
E_n[\psi_0] = \langle E_L(R) \rangle_{f_n},
\]  

(22)

where the brackets imply an average over \( f_n \). Usually the local energy estimate has a smaller variance than the growth estimate, because sampling errors \( (\rho_\rho \neq \rho_T) \) cause additional fluctuations in the population. However, it is extremely useful to compute both estimates, as their equality is a global check on the program.

In the next section, we will use different functions to guide the random walk and to estimate the energy. The function used to supply the antisymmetry and calculate the energy will be called the trial function \( \psi_T \). The local energy estimator using \( \psi_T \), for a random walk guided by \( \psi_0 \), is defined as

\[
E_n[\psi_T] = \frac{\int dR \psi_T(R) H f_n(R)}{\int dR \psi_T(R) f_n(R)} \frac{\langle W(R) \rangle}{\langle W(R) \rangle_{f_n}}.
\]  

(23)

By the second equality, the ground state energy is the asymptotic average \( (\text{large} n) \) of the weighted local energy of \( \psi_T \) (i.e., \( H\psi_T/\psi_T \)), where the weight is \( W(R) = \psi_T(R)/\psi_0(R) \).

III. ANTISYMMETRIC BOUNDARY CONDITIONS

A. The fixed-node method

A simple though approximate way of introducing Fermi statistics is the fixed-node method. It consists of finding the solution to the Schrödinger equation with the constraint that the wave function vanish whenever \( \psi_T(R) \), an antisymmetric trial function, is zero. In the DMC method \(^1\) this means that the nodes of \( \psi_T \) (where \( \psi_T = 0 \)) act as an absorbing barrier to all random walks. The random walks are terminated whenever they cross these nodes. The energy calculated in this approximation \( E_{\text{PN}} \) is an upper bound to the exact Fermi energy \(^2\) and, as we shall see later, if \( \psi_T \) is well chosen, a very tight upper bound. Of course, if the nodes of \( \psi_T \) happen to coincide with those of the exact Fermion ground state (or an excited state) as they do, for example, in one dimensional systems, the fixed-node energy will exactly equal the corresponding eigenvalue.

Even in the GFMC method, the calculated \( E_{\text{PN}} \) is not the exact fixed-node energy because the boundary conditions are not exactly obeyed. A move from \( R \) to \( R' \) should be absorbed by a nearby node, even though \( R \) and \( R' \) are on the same side, because the random walk from \( R \) to \( R' \) could have crossed and recrossed within one time step. Hence the exact fixed-node energy will only be obtained in the limit that the time step goes to zero. An exact method, with a nonvanish-
ing time step, requires that $\rho_{f}(R,R')\beta$ vanish whenever $\psi_{f}(R)$ or $\psi_{f}(R')$ do. Formally this can be accomplished by choosing $\rho_{f}$ to be the free particle density matrix with the fixed-node boundary conditions. However, then $\rho_{f}$ cannot be analytically calculated as done in Eq. (13) and some approximation must be made for the value of the integral in Eq. (10). Such an approximation has been implemented and found to improve the accuracy of the fixed-node energy for a given time step. However, such calculations will not be discussed since the exact fixed-node energy is not of primary interest here. In the release-node method the fixed-node distribution is considered merely an approximate starting distribution and there are no time step errors.

B. The nodal relaxation method

In the nodal relaxation method the random walks are allowed to cross the nodes of $\psi_{f}$, but instead of deleting the walks instantly, as in the fixed-node method, they are allowed to continue for some time. However, their contribution to any average will change sign each time they cross a node. Thus, walks that have crossed an odd number of times after they were in the fixed-node source contribute negatively and those that have crossed an even number of times contribute positively. In contrast to the fixed-node method, asymmetry is exactly taken into account. However, the process is numerically unstable since after many generations there will be approximately equal numbers of even and odd walks. As a result, any expectation value will become increasingly noisy. For this reason, this way of incorporating asymmetry has been called a transient estimate. If the trial energy equals the correct Fermi ground state energy, the difference between the numbers of even and odd walks will be constant upon each application of the Green’s function, however, the total number of walks will grow geometrically since the trial energy appropriate to unconstrained walks is the boson ground state energy. The relative growth in population after $M$ generations is from Eq. (21) equal to $[1 - (E_{f} - E_{g})\Delta]^{-M}$. Thus, the method is practically restricted to systems where either the trial nodes are relatively accurately located so that the relaxation to the true nodes occurs quickly or the Pauli principle is relatively unimportant so that $(E_{f} - E_{g})$ is small. Then the exact Fermi energy can be projected out accurately as the difference between the even and odd walks before the branching process overwhelms the computational resources. In this process, the original nodes are changed because of the unequal fluxes of the even and odd walks across a node. No attempt has been made to locate the new nodal surface because of the difficulty in visualizing these highly dimensional surfaces.

The function which guides the random walks must always be positive, since otherwise the Green’s function will not be a probability distribution. Let us assume the guiding function $\psi_{G}$ is a symmetric function of the electron coordinates and let the trial function $\psi_{f}$ be antisymmetric under electron interchange. It will be used to introduce the antisymmetric boundary conditions. In the release-node process only the guiding function controls the walks, which are free to cross the nodes. However, weights are attached to the walks so that $\psi_{f}$ can be used to project out the Fermion ground state. Let $\tilde{\rho}_{f}(R,R')$ be the Green’s function obtained from $\rho$ using $\psi_{f}$ as an importance function [in Eq. (14)]. Then it is easily related to the $\psi_{G}$ Green’s function ($\tilde{\rho}$):

$$\tilde{\rho}_{f}(R,R') = \frac{W(R)}{W(R')} \tilde{\rho}(R,R'),$$

where $W(R)$ is again given by $\psi_{f}(R)\psi_{G}(R)$. Suppose the point $R$ conditional on $R'$ has been sampled from $\tilde{\rho}$. To generate averages with respect to $\rho_{f}$ we merely include the weight factor $W(R) / W(R')$. If $\tilde{\rho}_{f}$ is applied many times to a symmetric initial distribution then the resulting distribution will converge to $\psi_{f}\phi_{f}$, where $\phi_{f}$ is the lowest energy Fermion wave function (the initial distribution $\phi_{FN}$ must overlap with $\phi_{f}$).

Because the nodal release process must occur quickly, it is essential to have the initial distribution as close as possible to this final distribution. We take our initial population from the output of the fixed-node process, $f_{FN}$ times a weight to bias the distribution by $\psi_{f}$ instead of $\psi_{G}$.

$$\tilde{\rho}_{f}(R) = |W(R)|f_{FN}(R).$$

Successive release-node generations are gotten by applying $\tilde{\rho}_{f}$ to this initial distribution.

$$\tilde{\rho}_{M+1}(R) = \int dR' \tilde{\rho}_{f}(R,R')g_{M}(R').$$

Therefore the release node weight for a random walk, starting in the fixed-node population at the point $R$, being guided by $\psi_{G}$ for $M$ generations and arriving at the point $R$ is $W(R)\sigma_{0}$, where $\sigma_{0} = \text{sign} [\psi_{f}(R)] = \pm 1$. Thus the weight of the point $R$ can be positive or negative depending on whether it has crossed an even or odd number of times. The release-node energy is defined as in Eq. (23), with $g$ replacing $f$, and its local energy estimator is

$$E_{M}^{\text{RN}} = \sum_{R,M} \frac{W(R)\sigma_{0}E_{f}(R)}{M}. \tag{27}$$

The sum is over all walks $M$ generations after release, and $E_{f}(R) = \psi_{f}^{-1}H\psi_{f}$ is the local energy of the trial function. Because of the growth of the number of configurations, $E_{M}^{\text{RN}}$ can be computed only up to some maximum value of $M$, say $S_{\text{max}}$. After that the walks are terminated. Also note the initial value of $E_{M}^{\text{RN}}$ is the fixed node energy $E_{0}^{\text{RN}} = E_{FN}$.

In a fixed-node calculation the guiding function $\psi_{G}$ can be set equal to $|\psi_{f}|$. In the release-node calculation such a choice would lead to an infinite variance in the energy estimate and in the branching ratio (i.e., there is a nonzero probability that the process of bringing intermediate points to the next generation will never finish). This divergence can be readily seen in the expression for the variance of the energy estimator [Eq. (27)]. The divergent term in the variance of $E_{M}^{\text{RN}}$ is

$$\int \phi_{M}(W\sigma_{0}W_{f}E_{f})^{2} = \int \phi_{M}^{2}[H\psi_{f}]^{2} / \psi_{G}, \tag{28}$$

where $\phi_{M}\psi_{G}$ is the total density of walks after nodal release (i.e., without the weights). Because the boundaries are not absorbing, $\phi_{M}$ will be positive everywhere. If $\phi_{G}$ were to vanish on the nodal surface, the integral would diverge.
logarithmically. There is no problem in a fixed-node calculation since $\phi_\alpha$ is forced to have the same nodes as $\psi_T$. Since the divergence of the variance is weak, $\psi_\alpha$ can still be chosen to be close to $|\psi_T|$ everywhere except very near the nodes. An equivalent problem occurs with the fluctuations in branching. The release node distribution $g_M(R)$ is independent of $\psi_\alpha$ for all $M$. Thus a guiding function which impedes node crossing will be accompanied by occasional large branching since the total contribution of the walks that cross is independent of $\psi_\alpha$. The choice of $\psi_\alpha$ will be discussed in the next section.

Although the fixed-node and release-node calculations have been presented as two separate processes they are in fact carried out concurrently, which not only simplifies the programming but also has the enormous advantage of reducing the statistical error in the difference between the fixed-node and release-node energies. In both the fixed-node and release-node processes, the random walk is guided by $\psi_\alpha$. Therefore in the release-node process, if a random walk by chance does not cross the nodes of $\psi_T$, it can still be considered a fixed-node source until it actually does cross a node. Thus, in determining the release-energy one can average not only over the initial ensemble of points, but also over the generation number which is considered as the initial condition for the release-node process. Each direct point $R$ can contribute to many values of $M$ in the release-node energy estimator Eq. (27) and this lowers the statistical error. To carry out this process, let us define $S$ as the last generation for which the ancestor of $R$ was in the fixed-node population. That is, previous to the generation $S$ there were no node crossings, while a node crossing and possible recrossing occurred while sampling generation $S + 1$. Then we can write for the release-node energy

$$E_{RN}^{M} = \frac{\sum_{R:S+S+M} W(R) | \sigma_0 \rangle E_{\tau}^{RN}(R) \}}{\sum_{R:S+S+M} W(R) | \sigma_0 \rangle}$$

(29)

where the point $R$ has generation number $i$ and the summations are over all points $R$ such that $i < S + M$. In other words, $M$ generations previously, the ancestor of $R$ must have been in the fixed-node population for $R$ to contribute to $E_{RN}^{M}$. Also $\sigma_0$ is the sign of $\psi_T$ then. If branching occurs, $S$ and $\sigma_0$ are copied into all branches. The walks are terminated after $i$ equals $S + S_{\max}$. Hence $M$ is less than or equal to $S_{\max}$.

This rearrangement reduces the error of each $E_{RN}^{M}$, because more configurations contribute to each sum, but more significantly the error of the difference $E_{RN}^{M} - E_{RN}^{M-1}$ is much reduced because large numbers of walks cancel out. It is easily shown from the previous equation that

$$E_{RN}^{M} - E_{RN}^{M-1} = \frac{\sum_{R:S+S+M} W(R) | \sigma_0 \rangle [E_{\tau}^{RN}(R) - E_{RN}^{M-1}] \}}{\sum_{R:S+S+M} W(R) | \sigma_0 \rangle}$$

(30)

The summation in the numerator is over all walks exactly $M$ generations after first crossing a node. Thus the number of terms in the numerator is, for small $M$, much less than those in the denominator and the error in the difference will be proportional to the difference, and not to the error in the total energy. This helps tremendously in deciding whether $E_{RN}^{M}$ has converged to its asymptotic value.

The release-node calculation thus involves the additional determination of the antisymmetric trial function $\psi_T$ and its local energy at every point of the random walk. Furthermore, the initial sign of $\psi_T$ must be recorded as well as

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Atom</th>
<th>$X$</th>
<th>$Y$</th>
<th>$\beta_0$</th>
<th>Spin</th>
<th>$\omega_K$</th>
<th>$\nu_K$</th>
<th>$c_x$</th>
<th>$c_y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_2$</td>
<td>H</td>
<td>-1.757</td>
<td>0</td>
<td>10.12</td>
<td>$\alpha$</td>
<td>2.19</td>
<td>0</td>
<td>2.03</td>
<td>0</td>
</tr>
<tr>
<td>$H_2$</td>
<td>H</td>
<td>0</td>
<td>0</td>
<td>$\alpha$</td>
<td>2.19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$H_2$</td>
<td>H</td>
<td>1.757</td>
<td>0</td>
<td>$\beta$</td>
<td>3.18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$H_2$</td>
<td>H</td>
<td>2.84</td>
<td>0</td>
<td>6.707</td>
<td>$\alpha$</td>
<td>1.67</td>
<td>1.028</td>
<td>-2.84</td>
<td>0</td>
</tr>
<tr>
<td>$H_2$</td>
<td>H</td>
<td>0</td>
<td>0</td>
<td>$\alpha$</td>
<td>2.18</td>
<td>2.24</td>
<td>0.729</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$H_2$</td>
<td>H</td>
<td>1.03</td>
<td>0</td>
<td>$\beta$</td>
<td>1.957</td>
<td>0.503</td>
<td>0.646</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$H_2$</td>
<td>H</td>
<td>3.156</td>
<td>1.36</td>
<td>6.23</td>
<td>$\alpha$</td>
<td>2.66</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$H_2$</td>
<td>H</td>
<td>-0.81</td>
<td>0</td>
<td>$\alpha$</td>
<td>1.55</td>
<td>1.27</td>
<td>3.156</td>
<td>1.36</td>
<td></td>
</tr>
<tr>
<td>$H_2$</td>
<td>H</td>
<td>+0.81</td>
<td>0</td>
<td>$\beta$</td>
<td>2.66</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LiH</td>
<td>Li</td>
<td>0</td>
<td>0</td>
<td>1.0358</td>
<td>$\alpha\beta$</td>
<td>0.8937</td>
<td>1.463</td>
<td>3.013</td>
<td>0</td>
</tr>
<tr>
<td>LiH</td>
<td>Li</td>
<td>3.0</td>
<td>0</td>
<td>$\alpha\beta$</td>
<td>0.568</td>
<td>0.414</td>
<td>0.01</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>LiH</td>
<td>Li</td>
<td>2.526</td>
<td>0</td>
<td>0.5766</td>
<td>$\alpha\beta$</td>
<td>0.509</td>
<td>0.414</td>
<td>-2.525</td>
<td>0</td>
</tr>
<tr>
<td>LiH</td>
<td>Li</td>
<td>-2.526</td>
<td>0</td>
<td>$\alpha\beta$</td>
<td>0.509</td>
<td>0.414</td>
<td>-2.525</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>LiH</td>
<td>Li</td>
<td>3.33</td>
<td>1.49</td>
<td>$\alpha\beta$</td>
<td>3.33</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>H$_2$O</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0.217</td>
<td>SCF/STO determinant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H$_2$O</td>
<td>H</td>
<td>1.4315</td>
<td>1.1094</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H$_2$O</td>
<td>H</td>
<td>-1.4315</td>
<td>1.1094</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE II. Parameters of the random walk. $\Delta$ is the generational time step in inverse hertz, $S_{\text{max}}$ is the number of generations from the first node crossing, random walks are deleted. $N$ is the length of the computer run in thousands of generations. $P$ is the average ensemble size.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$\Delta$</th>
<th>$S_{\text{max}}$</th>
<th>$N$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{H}_2(\text{I})$</td>
<td>0.2</td>
<td>20</td>
<td>100</td>
<td>400</td>
</tr>
<tr>
<td>$\text{H}_2(\text{II})$</td>
<td>0.2</td>
<td>20</td>
<td>23</td>
<td>400</td>
</tr>
<tr>
<td>$\text{H}_2(\text{III})$</td>
<td>0.2</td>
<td>32</td>
<td>50</td>
<td>400</td>
</tr>
<tr>
<td>LiH</td>
<td>0.0192</td>
<td>50</td>
<td>8</td>
<td>220</td>
</tr>
<tr>
<td>Li$_2$</td>
<td>0.0074</td>
<td>80</td>
<td>11</td>
<td>200</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>0.01</td>
<td>5</td>
<td>2</td>
<td>400</td>
</tr>
</tbody>
</table>

The generation number $S$ when the walk first crosses a node. The trial energy for the walks which have crossed a node is never changed and the population of those walks is not renormalized. All walks are followed $S_{\text{max}}$ generations after their first crossing at which time they are deleted. Thus there is no renormalization bias in the release node energy. The standard errors on $E_{\text{RN}}$, $E_{\text{MN}}$, $E_{\text{RN}} - E_{\text{MN}}$, and $E_{\text{RN}} - E_{\text{PN}}$ should be computed individually (since the errors are not directly computable from each other) by dividing up the entire calculations into a number of statistically independent blocks.

IV. THE TRIAL AND GUIDING FUNCTIONS

A proper trial wave function must be antisymmetric under the interchange of two electrons and possess a second derivative. The optimal trial function minimizes the variance of the total energy for a given amount of computer time. Both the variance of the local energy and the relaxation time from the fixed-node energy to the exact energy must be minimized. There is a tradeoff between complexity of the trial function, as measured by the computer time needed to evaluate it, and its accuracy (i.e., the variance of the local energy). Hence there is a premium for accurate compact trial functions.

The trial functions used in our calculations are of the pair product (Slater–Jastrow) form, also used in the earlier fixed-node paper (designated there as $\psi_{\text{III}}$). That is

$$\psi_{\text{R}}(R) = \exp \sum_{i<j} \left( w_{ij} (r_i - r_j) \right) \det \left[ \phi_k(r_i) \right], \quad (31)$$

where the sum runs over all pairs of particles, both electrons and nuclei. The correlation factor $u(r)$ is of the form

$$u_{ij}(r) = a_{ij} r/(1 + b_{ij} r), \quad (32)$$

where the parameters are determined by the cusp conditions at $r = 0$. That is, $a_{ij} = e^2/8D$ if $i$ and $j$ are electrons of the same spin, $a_{ij} = e^2/4D$ if $i$ and $j$ are electrons of different spin, $a_{ij} = -Ze^2/2D$ if $i$ is a nucleus of charge $Z_i$ and $j$ is an electron, and $b_{ij} = a_{ij}^{1/2} / \beta_{ij}$ where $\beta_{ij}$ is a variational parameter. The orbitals in the determinant are

$$\phi_k(r) = \exp \left[ - \left( r - x_k \right)^2 / (w_k^2 + v_k |r - x_k|) \right], \quad (33)$$

where $x_k$ is the center of the orbital and $w_k$ and $v_k$ are parameters. For the water molecule orbitals used were $2\zeta$ functions from an SCF–STO calculation.

As mentioned earlier, the guiding function must be a symmetric function which does not vanish anywhere and which should be close to $|\psi_{\text{R}}|$ except near the nodes. However, when $\psi_{\text{R}}$ is small because the electron density is small, as, for example, when an electron is far removed from the molecule, then $\psi_{\text{G}}$ should also be small since that represents an improbable configuration. A form for $\psi_{\text{G}}$ which has the desired characteristics is

$$\psi_{\text{G}}(R) = |\psi_{\text{R}}(R)| \left( 1 + \exp \left[ - \sum_i \rho(r_i) / \det [\phi_k(r_i)]^2 \right] \right)^{1/2}. \quad (34)$$

where $\rho(r) = \sum_k |\phi_k(r)|^2$ is an estimate of the electron density and $\Theta$ is a switching parameter. If the value of $\Theta$ is chosen to be too small, excessive branching will occur near the nodes. On the other hand, if $\Theta$ is too large $\psi_{\text{G}}$ could be quite different from $|\psi_{\text{R}}|$ and there could be too many walks in the nodal region, leading to a loss of efficiency. We chose $\Theta$ so that for about 10% of the configurations $\Theta \Pi \rho(r_i) > \det [\phi_k(r_i)]^2$.

TABLE III. Results. $E_{\text{RN}}$, $E_{\text{PN}}$, and $E_{\text{MN}}$ are the variational energy with the trial functions of Table I, the fixed-node energy and the release-node energy, respectively, in hertz with the standard error in the last digit given in parenthesis. The energy drop in going from the fixed-node to the release-node calculation, with its own error bars is given by $E_{\text{PN}} - E_{\text{MN}}$. $E_{\text{MN}}$ is the best estimate of the ground state energy derived either from experiment corrected for nuclear motion and relativistic effects or extrapolated CI calculations (H$_2$). $E_{\text{CI}}$ is the best available configuration interaction upper bound.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$E_{\text{RN}}$</th>
<th>$E_{\text{PN}}$</th>
<th>$E_{\text{MN}}$</th>
<th>$E_{\text{PN}} - E_{\text{MN}}$</th>
<th>$E_{\text{CI}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{H}_2(\text{I})$</td>
<td>-1.633(4)</td>
<td>-1.658(3)</td>
<td>-1.659(1)</td>
<td>0.0009(2)</td>
<td>-1.659 19$^a$</td>
</tr>
<tr>
<td>$\text{H}_2(\text{II})$</td>
<td>-1.605(7)</td>
<td>-1.623(3)</td>
<td>-1.624(3)</td>
<td>0.0005(2)</td>
<td>-1.624 31$^a$</td>
</tr>
<tr>
<td>$\text{H}_2(\text{III})$</td>
<td>-1.625(6)</td>
<td>-1.660(2)</td>
<td>-1.661(2)</td>
<td>0.0011(2)</td>
<td>-1.661 94$^a$</td>
</tr>
<tr>
<td>LiH</td>
<td>-8.04(18)</td>
<td>-8.067(1)</td>
<td>-8.071(1)</td>
<td>0.004(1)</td>
<td>-8.070 5$^a$</td>
</tr>
<tr>
<td>Li$_2$</td>
<td>-14.95(23)</td>
<td>-14.990(2)</td>
<td>-14.994(2)</td>
<td>0.004(1)</td>
<td>-14.996 7$^a$</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>-76.20(3)</td>
<td>-76.39(1)</td>
<td>-76.43(2)</td>
<td>0.04(1)</td>
<td>-76.437 6$^a$</td>
</tr>
</tbody>
</table>

V. RESULTS

Release-node calculations using GFMC have been carried out for the molecules H$_3$, LiH, Li$_2$, and H$_2$O. In all calculations the Born–Oppenheimer approximation is invoked, the nuclei being at fixed positions and of infinite mass. The geometries of the molecules and the trial function parameters are given in Table I. LiH, Li$_2$, and H$_2$O are at their equilibrium positions. H$_3$ (I) is at the saddle point for the reaction H + H$_2$, while H$_3$ (II) and H$_3$ (III) are at two other arrangements where CI calculations have been performed. Table II contains details about the random walks. Table III contains the results for the variational, fixed-node, and release-node energies, and comparisons with the best previous estimates of the ground state energies, as well as the

FIG. 1. The energy (in hartrees) vs the number of generations since node release for the molecule H$_3$(I) at the saddle point. Shown are the results for three different trial functions with (O) referring to that function whose parameters are tabulated in Table I. The experimental energy is rather uncertain so the results of a configuration interaction calculation upper bound and extrapolation are shown.

FIG. 2. The energy (in hartrees) vs the number of generations since node release for the molecule LiH. The results for three different trial functions are shown with (O) referring to that function whose parameters are tabulated in Table I. The parameters in the other two functions were deliberately deoptimized to raise the fixed-node energy.
Experimental energies corrected for relativistic and nuclear motion effects.

Figures 1–3 show the convergence of the energy from the fixed-node value to the exact result. For all molecules a variety of trial functions were used to ascertain whether any of these gave significantly lower fixed-node energies or smaller error bars and to check that the release-node energies were invariant with respect to the trial and guiding functions. Many more runs are included in the averages of Table III than shown in the figures. For the molecules LiH and Li₂, the trial functions were deliberately deoptimized, hence the errors bars and the fixed-node energies are larger for those.

**FIG. 3.** The energy (in hartrees) vs the number of generations since node release for the molecule Li₂. The results for two different trial functions are shown with (O) referring to that function whose parameters are tabulated in Table I.

**FIG. 4.** The change in the release-node energy (in hartrees) every four generations vs the number of generations since release for the molecule H₂/I. This is the first finite difference of Fig. 1.
trial functions. All calculations within the respective error bars converged to the same release-node energy, which is consistent with the best estimate of the ground state (nonrelativistic, Born–Oppenheimer) energy. Of course, assuming the estimate of the energy is distributed normally, there is a 32% probability that the exact answer will lie outside of one standard error and a 5% probability that it will lie outside two standard errors. Note that one of our molecules \( \text{Li}_2 \) does show such a deviation. The error bars in the release-node energy are, however, different for different trial functions. In these calculations \( S_{\text{max}} \) has been chosen so that about half of the steps are fixed-node and half are release-node. Hence the geometrical population increase has only begun when the crossing walks are terminated. For this reason the error of the release-node energy is only about 30% greater than that of the fixed-node energy. For these molecules, doubling \( S_{\text{max}} \) would approximately take twice as much computer time. The error bars for a run of the order of \( 10^7 \) moves was found to be 0.1 kcal/mol for \( \text{H}_3 \) and 1 kcal/mol for \( \text{LiH} \) and \( \text{Li}_2 \). This indicates that the percentage uncertainty relative to the total energy is comparable for all those molecules for equal length runs. For \( \text{H}_3 \), the Monte Carlo energy at the saddle point relative to the separated system is \( 9.65 \pm 0.08 \) kcal/mol which agrees well with the \( 9.59 \pm 0.06 \) kcal/mol estimate from an extrapolation of configuration interaction calculations.\(^{15}\) For the other two points on the potential surface the agreement is comparable to that at the saddle point. The nodes in one of our trial functions for \( \text{H}_3 \) are identical to those used in a fixed-node diffusion Monte Carlo calculation\(^ {16}\) and, as expected, the fixed-node energy is in agreement with these calculations. Our lowest fixed-node energy for \( \text{H}_3 (\text{I}) \) with the localized orbitals of Eq. (33) are in error by 0.7 kcal/mol. Reynolds et al.\(^ {17}\) has found a fixed-node energy much closer to the exact energy using a double zeta (SCF-STO) determinant. We have verified with a GFMC release-node calculation that the fixed-node energy with this trial function is only \( 0.15 \pm 0.05 \) kcal/mol above the release-node result. Also comparison

FIG. 5. The change in the release node energy (in hartrees) every five generations vs the number of generations since node release for the molecule LiH. This is the first finite difference of Fig. 2.

FIG. 6. The change in the release node energy (in hartrees) every eight generations vs the number of generations since node release for the molecule Li_2. This is the first finite difference of Fig. 3.
can be made with previous fixed-node diffusion Monte Carlo calculations on LiH and Li₂, where the energy gain in releasing the nodes is 2.5 kcal/mol. In general fixed-node calculations are of comparable accuracy to the best (nonextrapolated) configuration interaction calculations.

As discussed in the previous section the difference of the release-node energy from one generation to the next can be calculated more accurately than the energy by itself. These differences with their error bars are shown in Figs. 4–6. From these figures it is possible to accurately ascertain when the release-node process has converged to a steady state value by noting when the slope reaches zero. Note that generations are grouped before averages and error bars are computed. The decay of the energy is exponential, with a time constant inversely proportional to the excitation energy to the next state with the same symmetry as the ground state. The initial value of the difference is a clear reflection of the quality of the nodes of the trial function. It is possible to use this detailed information to improve the trial function by, for example, adding more determinants.

The release-node scheme encounters greater difficulties with atoms and molecules containing higher Z elements because, first of all, the error bars are proportional to the total energy and hence become larger in absolute value for a comparable run at lower Z. Secondly, because of the Pauli principle, the difference between the Fermi and Boson energy increases for larger Z, leading to a larger exponential factor in the population growth on nodal release compared to lower Z atoms. This means that nodal relaxation has to occur in a much shorter time in order to be able to still project out accurately the difference between the positive and negative populations. This difficulty is demonstrated in a calculation for H₂O with an STO type trial wave function in Figs. 7 and 8. Although the results are in reasonable agreement with experiments, there is no internal indication from the slope of the energy release (Fig. 8) that the nodal release process has converged because the error bars increase too fast with generation number even in the difference calculation. Although with a fourfold increase in computer time $S_{\text{max}}$ could be doubled, since the error bars are also growing rapidly for longer $S_{\text{max}}$, even much larger computer runs are required to establish an accurate ground state energy for longer release times. However, the determination of an accurate ground state energy for water is within reach by taking advantage of the parallel processing feature of the present generation of supercomputers, through which an increase of the order of 100 in calculational power is possible.

There are several other directions that can be pursued to increase the efficiency of these calculations. One of those is to replace the inner core electrons by a pseudopotential at the price of making these calculations no longer exact in the limit of infinitely long computer runs. The correlation of the valence electrons would still be treated exactly, and this may be for many chemical purposes by sufficiently accurate. In any case, such an approximation has been used in many previous theoretical calculations and its accuracy could be determined itself by Monte Carlo calculations. If the approximation of eliminating the inner electrons proves to be reliable, the statistical fluctuations in the Monte Carlo calculation would be considerably reduced as would the restrictions imposed by the Pauli principle on nodal release. Another scheme to be explored is to reduce the population growth on nodal release by the elimination of all random walks which frequently cross nodes. It has already been numerically demonstrated that the release-node energy does not perceptibly change when such walks are deleted. This means that the lowering of the energy from the fixed-node value is caused by random walks which cross the nodes only a few times. Once
this intuitive argument, based on the node correction being due to the net flux of node crossings, can be made rigorous, the elimination of walks which frequently cross (and hence do not contribute to the net flux) would reduce the population growth sufficiently to allow the nodal release calculation to be carried out for longer times and with smaller error bars. Another possibility to reduce population growth on nodal release is to annihilate a given positive and negative walk by forcing them, once they are found to be near each other in phase space, to the same point in phase space with the correct probability. Although this scheme works for low dimensional phase space where the probability of being close is not too small, in high dimensional problems formidable difficulties have to be overcome. Further elaboration of this process is given in Ref. 14. Still another way to improve the calculation is to improve the accuracy of the trial functions, including their node location, possibly with an adaptive Monte Carlo scheme. Finally, we are in the process of computing relative energies directly with correlated random walks. This method will be particularly useful in calculating potential energy surfaces, polarizibilities, etc.


13This result can be generalized: $\rho_0$ can be taken as any differentiable function of $R$ and $\beta$ (not necessarily a density matrix), but then $U$ of Eq. (10) is the average, over all random walks with $V \ln \rho_0$ as a drift term of $\rho_0^{-1} [H + \partial / \partial \varphi] \rho_0$.
15Let $\varphi_0$ be an exact eigenfunction with eigenvalue $E_\varphi$ and let $R'$ be sampled from $\rho_0 \{R, R', \varphi \}$ starting from a configuration $R$. Then using Eqs. (7), (17), and (18) it is easily demonstrated that the average over $R'$ of $m_\varphi + m_\varphi$ is unity, independent of $R, \rho_0$, and $\varphi$.