JOURNAL OF COMPUTATIONAL PHYSICS 51, 404-422 (1983)

The Simulation of Quantum Systems with Random Walks: A New Algorithm for Charged Systems

D. CEPERLEY

University of California, Lawrence Livermore National Laboratory,
Livermore, California 94550

Received September 3, 1982

Random walks with branching have been used to calculate exact properties of the ground state of quantum many-body systems. In this paper, a more general Green's function identity is derived which relates the potential energy, a trial wavefunction, and a trial density matrix to the rules of a branched random walk. It is shown that an efficient algorithm requires a good trial wavefunction, a good trial density matrix, and a good sampling of this density matrix. An accurate density matrix is constructed for Coulomb systems using the path integral formula. The random walks from this new algorithm diffuse through phase space an order of magnitude faster than the previous Green's Function Monte Carlo method. In contrast to the simple diffusion Monte Carlo algorithm, it is an exact method. Representative results are presented for several molecules.

I. INTRODUCTION

Forty years ago Fermi [1] suggested that properties of the ground state of a quantum system could be calculated by a random walk. In a more general context, random walks have been recognized to provide Green's functions for certain partial differential equations [2]. After some preliminary research on quantum Monte Carlo in the early 1950s, very little progress was made until Kalos applied [3] some of the techniques from neutron transport Monte Carlo [4]: importance sampling and iterated Green's functions. By importance sampling it is meant that the probability distribution of random walks is changed from ϕ, the ground state wave function, to $\psi\phi$ where ψ is the importance function. The advantages of this transformation will become evident later. The first major application of the method, which is now called Green's Function Monte Carlo (GFMC), was by Kalos, Levesque, and Verlet [5] who calculated the exact ground state energy of approximately 100 hard spheres, at densities near melting. This was followed by extensive simulations of the ground state of superfluid helium which achieved satisfactory agreement with experiment [6]. The particular algorithm used in those calculations will be referred to in this paper as KLV. The methodology and results have been reviewed in Ceperley and Kalos [7].

In recent years interest has shifted to Coulombic systems, in particular small molecules [8–10], the electron gas [11], and many-body hydrogen [12]. In contrast

0021-9991/83 $3.00

to helium 4, in these systems the interaction is known exactly; it is just the Coulomb potential between electrons and nuclei, at least in the nonrelativistic limit. In principle, the method is exact and in practice has been found to give reliable and precise results. The difficulty in applying random walk methods to electronic problems is that the probability distribution of the random walk cannot be made everywhere positive unless the nodes of the antisymmetric ground state are known. Usually the fixed-node approximation is made [11], where it is assumed that the nodes of a trial function are correct. For good trial functions (e.g., those from a self-consistent field method) this approximation has been found to give quite accurate results [9, 10]. This paper does not concern itself with the difficulties introduced by Fermi statistics, but only with the random walk process. In both the fixed-node algorithm and in an exact fermion algorithm, the boson random walk is used; only the boundary conditions are changed. So an improvement in the random walk algorithm will be equally as useful for fermion systems as it is for boson systems.

In most Monte Carlo studies of electronic systems, the exact KLV algorithm has not been used, but instead a simple stochastic process, reminiscent of Fermi's original suggestion, has been employed. In this approach, which will be referred to as diffusion Monte Carlo (DMC), the Schrödinger equation in imaginary time is viewed as a diffusion plus branching process, and a small time approximation to the evolution of the probability distribution is used. The method is only exact as the time step becomes sufficiently small. Indeed some have suggested [9] that a significant error persists even at very small time steps, because of the attractive Coulomb singularity. With small molecules, where very long random walks can be carried out with resulting small error bars on the energy, an exact method is quite desirable. Otherwise the effect of the time step error must be assessed. In addition, recent calculations have removed the fixed-node approximation [13], and then it was found to be difficult to simulate, sufficiently accurately, the simple diffusion process in the nodal region. In fact the difference between the exact and the fixed-node energy, and the error induced by a finite time step, are of the same order of magnitude for reasonable time steps.

However, it is found that the KLV algorithm applied to electronic systems converges an order of magnitude more slowly than diffusion Monte Carlo as measured by the average step size of the random walk. Hence the random walk with KLV needs to be much longer to achieve the same statistical error bars and convergence to the ground state. The slow convergence results from the use, in KLV, of domains from which the new random walk step is sampled [7]. Because the sampling is from the Green's function of an ideal gas, the potential energy must be bounded above in each domain and consequently the domains must be such that no two electrons can "touch." In practice, each electron is constrained to move in a sphere with a diameter less than the distance to the nearest electron and this constraint leads to rather small steps.

It is shown here that the exact Green's function can be related via an integral equation to any trial density matrix, not just the ideal gas one. If the electron–electron cusp is included properly in the trial density matrix, the domains in the

KLV algorithm can be eliminated with the result that the electrons can move much further in one step. Such a trial density matrix can be constructed from the Feynman–Kac formula for the exact density matrix, and it will be shown below that this leads to time steps as long as in the DMC method. Hence it is possible to have both a long time step and an exact simulation.

In Section II the terminology of the branched random walk is introduced and such a walk is related to an integral equation. In Section III, the exact many-body Green's function is related via an integral equation to an arbitrary density matrix. Section IV address the question of the efficiency of a random walk in calculating properties of the ground state. In an efficient random walk, branching is kept to a minimum and the average step size is as large as possible. In this new algorithm there are three arbitrary functions which can be adjusted to optimize the efficiency: the importance function, the trial density matrix, and the sampling function. The importance function should be chosen as close to the ground state as possible and the trial density matrix as close to the exact density matrix as possible. In addition the new points on the random walk should be sampled from the importance-transformed density matrix. Only if all three functions are optimal is the random walk optimal. In Section V a trial density matrix is derived from the path integral formula and details of its computation are given in the Appendix. Finally Section VI compares the results from this new algorithm with those of the KLV and DMC methods for several molecules.

II. BRANCHED RANDOM WALKS

This section introduces the terminology that is used in Green's function Monte Carlo to describe the random walk process and derives the relationship between the "rules" of the random walk and a Green's function. Since the branched random walk can be used to simulate many linear integral equations, this section will be general and not refer at all to the Schrödinger equation. Let R denote a point in $3N$-dimensional space, where N is the number of physical particles (electrons). The process begins with a set of points $\{R_i\}$ sampled from an initial probability distribution $f_1(R)$. This set is referred to as the first generation; the number of such points is the population of the first generation. Given any point R' in this set, a new point R is sampled from the distribution $h(R, R')$, and from these points are calculated a promotional probability q $(0 < q < 1)$, and a multiplicity M (a nonnegative integer). With probability q, M copies of R become members of the second generation; otherwise these M copies are intermediate points. If the multiplicity M is zero, that random walk is terminated. This process of making multiple copies is called branching. Except when taking averages, intermediate points are treated identically to first generation points. The random walk process continues until all of the walks, including the intermediate points, reach the second generation. The population of points in the second generation have a distribution $f_2(R)$ and will be the source for the next iteration, etc.

The Green's function $G(R, R')$ is defined as the number density of points R in the second generation coming from a single point R' in the first generation. The distribution of second generation points is then

$$f_2(R) = \int dR'\, G(R, R')\, f_1(R'), \tag{1}$$

or, in operator notation, $f_2 = G * f_1$, and $f_n = G^{n-1} f_1$.

The number density for a random walk arriving in the second generation in one step (no intermediate points) is

$$H \equiv H(R, R') \equiv h(R, R')\, q(R, R')\, M(R, R'), \tag{2}$$

while the number density of creating an intermediate point in one step is

$$K \equiv K(R, R') \equiv h(R, R')(1 - q(R, R'))\, M(R, R'). \tag{3}$$

The total Green's function G is the sum of one step processes, two step processes, etc.

$$G = H + H * K + H * K * K + \cdots. \tag{4}$$

Assuming this series converges, that all of the intermediate points will eventually reach the second generation, the series can be summed to give

$$G = H + G * K. \tag{5}$$

This operator equation specifies the relationship between the "rules" of the random walk (h, q, and M) and the Green's function G. Taking the coordinate representation of G, H, and K, it becomes an integral equation of the second kind relating the known functions H and K to the unknown Green's function G.

It is quite simple to generalize this algorithm. For example, the multiplicity M need only be positive, not necessarily an integer. Let the number of copies of R, actually made, be $\lfloor M(R, R') + \chi \rfloor$ where $\lfloor a \rfloor$ is the largest integer less than a, and χ is a uniformly distributed random number in $[0, 1]$. Then H is unchanged since

$$\begin{aligned} H(R, R') &= \int_0^1 d\chi\, h(R, R')\, q(R, R')\, \lfloor M(R, R') + \chi \rfloor \\ &= h(R, R')\, q(R, R')\, M(R, R'), \end{aligned} \tag{6}$$

with a similar result for K. Hence G is still given by (5).

III. Green's Function Monte Carlo

The ground state of energy of a quantum many-body system may be estimated using the branched random walk. Green's Function Monte Carlo, as developed by Kalos [3], consists of, first, relating the many-body Green's function for the

Schrödinger equation to the ideal gas Green's function via an integral equation and, second, transforming this Green's function, by multiplication with a known importance function [5], to achieve a better behaved random walk. Following Kalos, the Green's function for the Schrödinger equation with importance sampling is

$$G(R, R') = \psi(R)\,\psi^{-1}(R')(E_T - w)\int_0^{\infty} dt\langle R\,|\,e^{-(\mathscr{H}-w)t}\,|R'\rangle, \tag{7a}$$

or

$$G(R, R') = \psi(R)\,\psi^{-1}(R')\langle R\,|\,\frac{(E_T - w)}{(\mathscr{H} - w)}\,|R'\rangle, \tag{7b}$$

where $\mathscr{H}$ is the many-body Hamiltonian, $\psi(R)$ is the importance function, E_T is the trial energy, and w is a constant ($w < E_T$). Let the eigenvalues and eigenfunctions of $\mathscr{H}$ be ε_α and $|\alpha\rangle$. Then the distribution of walks in the nth generation can be written in terms of these eigenfunctions as

$$f_n = G^{n-1}f_1 = \psi\sum_\alpha \left(\frac{E_T - w}{\varepsilon_\alpha - w}\right)^{n-1} |\alpha\rangle\langle\alpha\,|\,\psi^{-1}f_1\rangle. \tag{8}$$

If $\varepsilon_0 > w$ and $\langle 0\,|\,\psi^{-1}f_1\rangle \neq 0$ (ε_0 and $|0\rangle$ are the ground state eigenvalue and function), then for large enough n the ground state will dominate:

$$f_n \to \psi\,|0\rangle\left(\frac{E_T - w}{\varepsilon_0 - w}\right)^{n-1} \langle 0\,|\,\psi^{-1}f_1\rangle. \tag{9}$$

The population of the nth generation will have a mean value of $\int dR\, f_n$. From Eq. (9), to keep the population stable, for large n (neither expanding or declining) the trial energy must equal the ground state energy ($E_T = \varepsilon_0$). The constant energy w controls the rate at which distribution f_n converges to the ground state as can be seen from Eq. (8). As w approaches ε_0 the higher eigenfunctions are attenuated more per generation. In the next section w will be related directly to the effective time step of the random walk.

The ground state energy can thus be estimated by the growth of the population. Another useful way of estimating ε_0 comes from the averge value of the *local trial energy* ($\psi^{-1}\mathscr{H}\psi$). Since $\mathscr{H}$ is Hermitian the average value of the local trial energy approaches the ground state energy.

$$\begin{aligned}\varepsilon_0 &= \langle\psi\,|\,0\rangle^{-1}\langle\psi|\mathscr{H}\,|0\rangle = \langle 0\,|\,\psi\psi^{-1}\mathscr{H}\psi\rangle\langle\psi\,|\,0\rangle^{-1} \\ &= \lim_{n\to\infty}\,\langle\psi^{-1}\mathscr{H}\psi\rangle_n,\end{aligned} \tag{10}$$

where $\langle\ \cdots\ \rangle_n$ is the average value over the population in the nth generation. In order to achieve good estimates of ε_0, the local trial energy must be as smooth as possible. Hence the importance function ψ should be as close as possible to $|0\rangle$, particularly in regions where the potential energy is singular.

In order to sample the Green's function G (Eq. (7)) with a branched random walk it must be related via the integral equation (5) to known functions H and K. In KLV, H was chosen to be the ideal gas Green's function. Here the Green's function is related to any trial density matrix. To obtain this relationship between H and K consider T_β, a known operator, continuous in β, with $T_0 = I$ and $T_\infty = 0$. The following operator identity comes from integration by parts

$$(\mathscr{H} - w)^{-1} = \int_0^\infty d\beta\, T_\beta - \int_0^\infty d\beta (\mathscr{H} - w)^{-1} \left(\mathscr{H} - w + \frac{d}{d\beta}\right) T_\beta. \tag{11}$$

It is easy to see that Eq. (11) will relate the many-body Green's function G (Eq. (7)) to H with the integral equation (5) if H is chosen to be

$$H(R, R') = \psi(R)\, \psi^{-1}(R')(E_{\mathrm{T}} - w) \int_0^\infty d\beta \langle R |\, T_\beta \,| R' \rangle. \tag{12}$$

The kernel must then be

$$K(R, R') = -\psi(R)\, \psi^{-1}(R') \int_0^\infty d\beta \langle R | \left(\mathscr{H} - w + \frac{d}{d\beta}\right) T_\beta \,| R' \rangle. \tag{13}$$

Now H must be as close to G as possible to minimize the number of intermediate points. In fact convergence of the branched random walk is only ensured if the kernel K is sufficiently small.[1] Also T_β must be chosen so that K is positive everywhere, since promotional probabilities q greater than one or multiplicities M less than zero cannot be allowed. Clearly the best choice for T_β is $\exp(-(\mathscr{H} - w)\beta)$, proportional to the density matrix at a temperature β^{-1}. But since the density matrix for a many-body system is also unknown (remember H must be a computable function), assume

$$\langle R |\, T_\beta \,| R' \rangle = \exp(-N_\beta)\, \rho_{\mathrm{T}}(R, R'; \beta), \tag{14}$$

where ρ_{T} is a known (computable) trial density matrix and N_β a normalization function[2] needed to keep K positive. Define the trial-density-matrix-local-energy as

$$E_\beta(R, R') \equiv \rho_{\mathrm{T}}^{-1} \left(\mathscr{H} + \frac{\partial}{\partial\beta}\right) \rho_{\mathrm{T}}(R, R'; \beta). \tag{15}$$

If ρ_{T} were the exact density matrix, then E_β would be zero everywhere. Let us assume that E_β can be bounded everywhere by B_β. That is,

$$B_\beta \geqslant E_\beta(R, R') \qquad \text{for all } R,\, R'. \tag{16}$$

---

[1] The largest eigenvalue of K must be less than unity.

[2] In principle N_β and B_β could be functions of R'. For simplicity they will be assumed to depend only on β.

The kernel from Eq. (13) is then

$$K(R,R') = \psi(R')^{-1}\,\psi(R)\int_0^{\infty} d\beta\langle R\,|\,T_\beta\,|R'\rangle(w + N_\beta - E_\beta), \tag{17}$$

and will be positive everywhere if ψ and ρ_{T} are, and if

$$N_\beta = \int_0^{\beta} dt(B_t - w). \tag{18}$$

The parameter β up to this point has only been an integration variable in the definition of H and K (Eqs. (12), (13)). Those integrations will be performed by sampling β from the joint distribution $h(R,R';\beta)$. In fact it will be shown below that β is the diffusion time of the random walk and corresponds to imaginary time in the time-dependent Schrödinger equation.

Finally, the promotial probability and the multiplicity can be related to ρ_{T}, ψ, and h using Eqs. (2), (3), (12), (17), (18).

$$q = (1 + (B_\beta - E_\beta)/(E_{\mathrm{T}} - w))^{-1}, \tag{19}$$

$$M = \frac{\psi(R)\,\psi^{-1}(R')\,e^{-N_\beta}\rho_{\mathrm{T}}(R,R';\beta)(B_\beta - E_\beta + E_{\mathrm{T}} - w)}{h(R,R';\beta)}. \tag{20}$$

This constitutes a more general algorithm than KLV Green's Function Monte Carlo, since the Green's function has been related to any trial density-matrix ρ_{T}. In fact the KLV algorithm can be recovered by choosing ρ_{T} to be the free particle density matrix inside a domain $D(R')$

$$\begin{aligned} \left(-\nabla^2 + \frac{\partial}{\partial\beta}\right)\rho_{\mathrm{T}}(R,R';\beta) &= 0, \qquad R \in D(R'), \\ \rho_{\mathrm{T}}(R,R';\beta) &= 0, \qquad R \notin D(R'). \end{aligned} \tag{21}$$

The domain D in KLV [5] is a Cartesian product of N spheres centered around each particle with radii chosen so that the spheres do not overlap. Since ρ_{T} has a discontinuous gradient at the surface of D, E_β will contain surface terms which give rise to a new type of intermediate point, where the random walk moves to the surface of D. It is these surface moves which limit the step size ($|R - R'|$). The advantage of the KLV algorithm is that E_β is the potential energy and need only be bounded inside the domain D. In the present approach one must bound E_β everywhere. However, for electronic systems it is possible to construct an accurate trial density matrix with a bound which allows step sizes an order of magnitude larger than in KLV, thus achieving quicker convergence of the random walk. From a programming point of view the new GFMC algorithm differs from KLV only in the greater flexibility of sampling new points and in the explicit formulas for q and M, Eqs. (19), (20).

IV. OPTIMIZATION OF THE RANDOM WALK

The branched random walk will calculate the exact ground state energy only if the initial population is infinite, and if the process is advanced an infinite number of generations. In any finite simulation there are statistical errors, convergence error, and bias. A statistical error is inherent in any Monte Carlo calculation since an average over a finite number of samples has a variance inversely proportional to the number of samples. Convergence error results if f_n for practical values of n has not converged to $\psi|0\rangle$. By bias, in this paper, it is meant that the probability density in the nth generation is not proportional to f_n because of the finite number of walks. The as yet undetermined functions in the random walk algorithm, namely ψ, ρ_{T}, and h, are to be chosen to minimize these errors.

The variance of the ground state energy estimate is

$$\mathrm{var}(\bar{\varepsilon}_0) < \langle(\psi^{-1}\mathscr{H}\psi - \varepsilon_0)^2\rangle/(\kappa N^*), \tag{22}$$

where N^* is the total number of samples which contribute to the average and κ $(0 < \kappa < 1)$ is a factor to account for the correlation between members of the sample. If every point on the random walk were independent of every other point then κ would be unity. However, because of branching and finite step sizes κ usually is less than unity. It is clear that to reduce these correlations branching is to be avoided, since coverage of phase space is lost when walks are terminated, and clustering results from duplication of walks. Also, to reduce this variance, the average number of intermediate points should be reduced since they do not contribute to the average. But it is most crucial to choose ψ so that fluctuations in the local energy are small. Finally, the computer time needed to evaluate ρ_{T}, E_β, and h will limit the total number of steps in the simulation.

From the eigenvalue expansion of the Green's function (Eq. (8), it can be seen that convergence to the ground state is controlled by $(\varepsilon_0 - w)/(\varepsilon_\alpha - w)$ for those excited states $|\alpha\rangle$ with a component in the initial distribution f_1. Typically, f_1 is taken to be proportional to $|\psi|^2$. For quick convergence, w should be as close as possible to ε_0 and ψ close to $|0\rangle$. It is important that ψ have the correct long range properties since otherwise convergence can be expected to be slow.

A bias results from too frequent renormalization of the population of walks. The generational population will fluctuate even if the trial energy is exact because of fluctuations in branching. Since the population is not constrained, it will drift up and down, and if the process is run for enough generations, it will eventually either be too large to continue, or zero. Hence the population needs to be renormalized every so often, or stabilized by some other means. However this is done, a bias will be introduced, because the probability of removing or adding a walk in order to stabilize the population will be correlated with the position of the walk through its past history. This bias will depend on the ratio of the renormalization time to the correlation time of the branching. This renormalization time[3] is proportional to the average population divided by the mean squared fluctuations in the branching ($P/\langle (M-1)^2\rangle$). Hence to reduce bias branching must be kept to a minimum.

Relationship to a Diffusion Process

As was noted by Fermi [1] and recently used by Anderson [8] and others [9–11] the ground state can also be simulated by a diffusion and branching process. The intimate relationship between this diffusion process and Green's Function Monte Carlo allows one to determine the sampling function h in GFMC. The time-dependent importance-sampled Green's function used in diffusion Monte Carlo [11] (DMC) is

$$g_\beta(R,R') = \psi(R)\,\psi^{-1}(R')\langle R|\,e^{-\beta(\mathscr{H}-\varepsilon_0)}\,|R'\rangle. \tag{23}$$

Then G is simply the Laplace transform of g_β

$$G(R,R') = \tau_G^{-1}\int_0^\infty d\beta\, g_\beta(R,R')\,e^{-\beta/\tau_G}, \tag{24}$$

where the Laplace transform variable $\tau_G = (\varepsilon_0 - w)^{-1}$ is defined as the generational time step. Since β is being sampled from some distribution, GFMC can be regarded as a variable time step version of DMC with the average time step being τ_G.

It is easy to verify that g_β will satisfy the diffusion and branching equation

[redacted]

The total population resulting from a single point R' a time β later is

$$P_\beta(R') = \int dR\, g_\beta(R,R'), \tag{26}$$

then from Eq. (25) and Green's theorem

$$\frac{\partial}{\partial\beta}P_\beta(R') = -\langle \psi^{-1}\mathscr{H}\psi - \varepsilon_0\rangle_\beta\, P_\beta(R'), \tag{27}$$

where $\langle\ \cdots\ \rangle_\beta$ means the average over the population at time β, i.e. the average over $\hat{g}_\beta = g_\beta/P_\beta$. Thus, the average change in the population in DMC, and through

---

[3] Assume each point in a population of size P_0 branches on the average once with a variance ε. The variance in the total population will be $P_0\varepsilon$ and the distribution of the new population will be proportional to $\exp(-(P^{1/2}-P_0^{1/2})^2/2\varepsilon)$. Hence the variable $P^{1/2}$ executes a one-dimensional random walk with step size $\varepsilon^{1/2}$ and the average number of generations needed until $P=0$ the first time is P/ε.

Eq. (24) in GFMC, is determined by the local energy. The crucial role of importance sampling is to control the branching process. If importance sampling were absent ($\psi = 1$) the change in population would be determined by the bare potential. This means that in many-body systems, an exponentially small number of walks would branch very many times, and the rest would branch to zero.

Suppose the trial density matrix were equal to the exact density matrix, $\rho_T = \rho_\beta$. Then the mean value of the multiplicity in Eq. (20) would be $P_\beta(R')$ and the optimal sampling function for R and β would be $h_1(\beta)\,\hat{g}_\beta(R, R')$, where

$$h_1(\beta) = -\frac{d}{d\beta} \exp\left[-\int_0^\beta d\tau(B_\tau + 1/\tau_G)\right]. \tag{28}$$

To sample this distribution[4] generate χ, a uniformly distributed random number in $[0, 1]$, and solve for β from

$$-\ln \chi = \int_0^\beta d\tau(B_\tau + 1/\tau_G). \tag{29}$$

From Eq. (28), the effect of a poor trial density matrix, as measured by the bounding function B_β, is to cause small diffusion times β and hence slow convergence and high statistical correlations. To avoid many intermediate points τ_G must be chosen such that

$$\tau_G \langle B_{\tau_G} - E_{\tau_G} \rangle < 1. \tag{30}$$

The optimal spatial sampling function $\hat{g}_\beta(R, R')$ is the solution to Eq. (25) with no average branching term. Assuming the local energy is constant in the sampling region, then

$$\frac{\partial \hat{g}_\beta}{\partial \beta} = \sum_i \frac{\hbar^2}{2m} \nabla_i(\nabla_i \hat{g}_\beta - 2\hat{g}_\beta \psi^{-1} \nabla_i \psi), \tag{31}$$

with the boundary conditions $\hat{g}_0(R) = \delta(R, R')$. This is the Smoluchowski equation [14] which describes Brownian motion in an external potential proportional to $\ln |\psi|^2$.

An approximate solution to Eq. (31) is obtained by expanding $\ln \psi(R)$ to second order in $(R - R')$ and neglecting higher order terms. Assuming the matrix of second derivatives $(\partial_\alpha \partial_\beta \ln \psi(R))$ is positive definite, the solution is the product of $3N$ harmonic-oscillator density matrices in the directions given by the eigenfunctions of this matrix [14]. In practice diagonalizing this matrix is a significant computation, and it has been found that one can neglect all but the diagonal terms without signifi-

---

[4] In many situations it is desirable to impose an upper limit on the time step β since ρ_T or h may be very inaccurate for $\beta > \beta_c$. One can set $B_\beta = \infty$ for $\beta > \beta_c$ in Eqs. (18), (19), (29). Thus $q = 0$ for $\beta = \beta_c$, and an intermediate point is generated. The same result is obtained by changing the limits of integration in Eq. (11) to be $[0, \beta_c]$.

cantly changing the fluctuations in the branching. For each particle i its new coordinate is then given by

$$\mathbf{r}_i = \mathbf{r}'_i + (\beta\hbar^2/m_i)\,\psi^{-1}\,\nabla\psi + \boldsymbol{\chi}_i, \tag{32}$$

where $\boldsymbol{\chi}_i$ is a normally distributed random vector with zero mean and variance

$$\langle\chi_i^2\rangle = (\beta\hbar^2/m_i)(1 + \beta\hbar^2\nabla^2 \ln \psi/3m_i). \tag{33}$$

Special treatment must be given in regions where $\nabla \ln|\psi|$ or $\nabla^2 \ln|\psi|$ are very large, such as near encounters of two particles or near a node. In addition, it has been found useful to enhance the probability of large moves in order to ensure the multiplicity M is bounded as $|R - R'|$ becomes large.

V. THE TRIAL DENSITY MATRIX

In Section IV it was shown that fluctuations in branching come from inadequacies in ρ_{T}, ψ, and h. However, the time step, and hence the convergence of the walk, is determined by ρ_{T} alone through the upper bound to the trial-density-matrix-local-energy. The trial density matrix must approximate well the exact density matrix for values of β on the order of the average time step τ_{G}. For small β, the density matrix has approximately the semiclassical form $\rho_0 \exp(-\beta V)$, where V is the potential energy and ρ_0 is the ideal gas density matrix. For charged systems this form does not have a bound because it has the wrong behavior at small interparticle distances. If the potential energy is a sum of pair interactions ($V = \sum_{i<j} v_{ij}$), a pair product form will, however, give a satisfactory trial density matrix, since the exact density matrix has this form at small enough β. Assume

$$\rho_{\mathrm{T}}(R, R'; \beta) = \rho_0(R, R'; \beta) \exp\left[-\sum_{i<j} u(\mathbf{r}_{ij}, \mathbf{r}'_{ij}; \beta)\right], \tag{34}$$

where $\mathbf{r}_{ij}$ is the relative distance between particles i and j. The trial-density-matrix-local-energy can then be written in the form

$$E_\beta = \sum_{i<j} \theta_\beta(\mathbf{r}_{ij}, \mathbf{r}'_{ij}) - \sum_i \frac{\hbar^2}{2m_i}\mathbf{F}_i^2, \tag{35}$$

where

$$\theta_\beta(\mathbf{r}, \mathbf{r}') = v(\mathbf{r}) - \frac{\partial u}{\partial\beta} + \frac{\hbar^2}{2\mu}\nabla^2 u - \frac{(\mathbf{r} - \mathbf{r}')}{\beta}\nabla u, \tag{36}$$

and

$$\mathbf{F}_i = \sum_{j \neq i} \nabla_i u(\mathbf{r}_{ij}, \mathbf{r}'_{ij}; \beta), \tag{37}$$

and μ is the reduced mass of particles i and j. Then E_β will be bounded by zero if the two body problem, $\theta_\beta = 0$, is solved since the squared gradient term is negative.[5]

Notice that if Eq. (36) is solved, u will be linear in the interaction v. Hence if the exact density matrix is expanded in a power series in the interaction, u will be the linear term. The path integral formula for the density matrix (Feynman–Kac) [15] gives such an expansion

$$\rho(R, R'; \beta) = \rho_0(R, R'; \beta) \left\langle \exp\left[-\int_0^\beta dt\, V(t) \right] \right\rangle_{R' \to R}, \tag{38}$$

where the average ($\langle \cdots \rangle$) is over all Weiner paths from R' to R in time β and $V(t)$ is the total potential energy at time t on this path. Assume $V(R)$ is pairwise additive and has a Fourier transform. Expanding the exponential in a power series gives

$$\rho = \rho_0 \sum_{n=0}^{\infty} \frac{(-)^n}{n!} \int_0^\beta dt_1 \cdots \int_0^\beta dt_n \langle V(t_1) \cdots V(t_n) \rangle_{R' \to R}, \tag{39}$$

then the solution to the equation, $\theta_\beta = 0$, will be

$$u(\mathbf{r}, \mathbf{r}'; \beta) = \int_0^\beta dt \int d^3\mathbf{r}''\, v(\mathbf{r}'') \frac{\rho_0(\mathbf{r}'', \mathbf{r}'; t)\, \rho_0(\mathbf{r}, \mathbf{r}''; \beta - t)}{\rho_0(\mathbf{r}, \mathbf{r}'; \beta)}, \tag{40}$$

where ρ_0 is the two-particle ideal-gas density matrix

$$\rho_0(\mathbf{r}, \mathbf{r}'; \beta) = (2\pi\hbar^2\beta/\mu)^{-3/2} \exp[-\mu(\mathbf{r} - \mathbf{r}')^2/(2\hbar^2\beta)]. \tag{41}$$

The density matrix $\rho_0 e^{-u}$ is a much better approximation to the exact density matrix than $\rho_0(1 - u)$ since all of the higher order averages in Eq. (39) are approximated by their uncorrelated versions (i.e., $\langle V(t_1)\, V(t_2) \rangle \simeq \langle V(t) \rangle^2$). This is accurate as long as distances are larger than $\beta^{1/2}$.

Equation (40) for u can be simplified by the use of Fourier transforms to give

$$u(\mathbf{r}, \mathbf{r}'; \beta) = \beta \int_0^1 d\tau\, W((1 - \tau)\, \mathbf{r}' + \tau\mathbf{r}; \beta\tau), \tag{42}$$

where $W(\mathbf{r}; t)$ is a time-dependent smoothed potential with Fourier transform

$$W_k = v_k \exp\left[-\frac{\hbar^2 k^2 t(\beta - t)}{2\mu\beta} \right]. \tag{43}$$

Thus the pair density matrix is the integral along a straight line from $\mathbf{r}'$ to $\mathbf{r}$ of the

---

[5] This bound is actually attained since the squared gradient term vanishes as all of the electrons are removed from each other and from the nuclei. Since $E_\beta < 0$ one can easily show that $\rho_T \leqslant \rho$. For Coulomb systems, one can easily show that this trial density matrix satisfies the conditions assumed in Section III, namely, $T_0 = I$ and $T_\infty = 0$.

smoothed potential $W(\mathbf{r}; t)$. In the case of the Coulomb potential ($v(r) = e_i e_j/r$) one has

$$W(\mathbf{r}, t) = \frac{e_i e_j}{r} \operatorname{erf}\left[r \left(\frac{\mu\beta}{2\hbar^2 t(\beta - t)} \right)^{1/2} \right]. \tag{44}$$

The properties of u for the Coulomb case, and a procedure for calculating its value, are discussed in the Appendix.

Now consider the situation, common in simulations, of periodic boundary conditions. There are two effects of the boundaries. First, the ideal gas density matrix (Eq. (41)) must be replaced by its periodic counterpart ρ_0^P:

$$\rho_0^P(\mathbf{r}, \mathbf{r}'; \beta) = \sum_{n=-\infty}^{\infty} \rho_0(\mathbf{r}, \mathbf{r}' + Ln; \beta), \tag{45}$$

where n is an integer vector and L is the set of three primitive vectors of the Bravais lattice of the periodic system. Second, the potential energy becomes periodic. Usualy in Coulomb systems the Ewald potential is used

$$v^P(\mathbf{r}) = \sum_n{}' \, v(\mathbf{r} + Ln), \tag{46}$$

where a neutralizing background is added and the sum in Eq. (46) is performed in such a manner as to assure convergence [16]. The periodic pair density matrix can be written in terms of the infinite one using Eqs. (39), (40), (44), and (45).

$$u^P(\mathbf{r}, \mathbf{r}'; \beta) = \frac{\sum_{n_1 n_2} \rho_0(\mathbf{r} + Ln_1, \mathbf{r}' + Ln_2; \beta)\, u(\mathbf{r} + Ln_1, \mathbf{r}' + Ln_2; \beta)}{\rho_0^P(\mathbf{r}, \mathbf{r}'; \beta)}. \tag{47}$$

Again the procedure for calculating the value of this function for a Coulomb system is given in the Appendix.

VI. RESULTS

The branched random walk with the path integral density matrix has been programmed and tested on several small molecules and compared with the other methods (KLV and DMC). As explained in the Introduction, for those molecules with more than two electrons, the fixed-node approximation was made in order to converge to an antisymmetric distribution. For all four molecules the importance function used in these calculations was ψ_{III} of [10].

The diffusion Monte Carlo algorithm employed in this calculation was exactly that of [10]. There is some ambiguity in choosing the time step in this method, since there is a trade-off between accuracy of the algorithm and statistical error. A systematic study [10] of time step dependence has been made for the molecule H_2. As a rule of thumb, the time step is chosen so that the rejection ratio is less than 1%. (Rejection is an ad hoc method of enforcing detailed balance of the diffusion process.) With this in mind a time step $\beta = 0.01$/Hartree has been used with all four molecules.

In the KLV algorithm, for any given configuration R, a sphere radius a_i around electron i is computed from the distances to nearby electrons and a bound to the potential energy is then found with those radii. Let r_i^* be the minimum distance between electron i and any other electron and let E_i be the magnitude of the electric field that electron i feels. Then a satisfactory bound on the potential is (with units $e = 1$)

$$B = V(R) + \sum_i \left[a_i E_i + 2(a_i/r_i^*)^2 (r_i^* - 2a_i)^{-1}\right]. \tag{48}$$

The sphere radii a_i were chosen to approximately maximize the average diffusion time; this leads to the equation for a_i

$$a_i^3 \frac{dB}{da_i} = 2\pi^2, \tag{49}$$

which has the power series solution

$$a_i = \tfrac{1}{2} r_i^* \left[1 - \lambda_i + \tfrac{3}{2}\lambda_i^2 - \tfrac{17}{8}\lambda_i^3 \cdots\right], \tag{50}$$

where $\lambda_i = (r_i^*/8)^{1/2}/\pi$. The bound B is not rigorous for a many-electron system, but [illegible] molecules. The algorithm for choosing the radii is nearly optimal, since the average diffusion time changes very little if $a_i = r^*/2$, the maximum possible.

The ground state energies for the three methods were estimated using Eq. (10). Five generations at the beginning of the run were discarded in computing the mean value. In all cases the energy estimates from the three methods were the same within the statistical errors and are given in Table I. Table I also shows the average time step, the average promotional probability, and the average error of the three methods on the systems H_2, He, LiH, and Li_2. The average error is defined as the product of the error of the estimate of the energy times the square root of the total CPU time (in CDC 7600 sec). The variance was calculated by dividing the total computer run into 50 blocks and calculating the dispersion of the block energies. Because the error of the variance is large, these error estimates are rather uncertain. The average population size was in the range of 250 to 1000 systems and the average diffusion time per block was about $\frac{1}{2}$ hartree^{-1}.

The comparison of averge error shows that the improved GFMC algorithm does lessen the CPU time needed to achieve a given error bar in the energy, typically by a factor of 4 to over KLV and from 1.5 to 10 over DMC. It should be stressed that these error estimates probably have an error of about 25%. The increase in the diffusion time from KLV to the present GFMC algorithm is quite dramatic, a factor of ten being typical.

TABLE I

System	H_2	He	LiH	Li_2
$\bar{\varepsilon}_0$	−1.174(1)	−2.903(2)	−8.066(4)	−14.980(7)
β_{DMC}	0.01	0.01	0.01	0.01
ζ_{DMC}	0.079	0.025	0.15	0.021
$\langle\beta\rangle_{\mathrm{KLV}}$	0.053	0.022	0.0068	0.0037
$\langle q\rangle_{\mathrm{KLV}}$	0.20	0.16	0.29	0.32
ζ_{KLV}	0.030	0.041	0.22	0.45
$\langle\beta\rangle$	0.55	0.43	0.087	0.043
$\langle q\rangle$	0.70	0.74	0.80	0.82
ζ	0.017	0.015	0.05	0.17
$\langle(M-1)^2\rangle^{1/2}$	0.49	1.46	1.23	0.41
$\langle(M-1)^2\rangle^{1/2}_{\mathrm{DMC}}$	0.10	0.05	0.023	0.020

Note. The results of runs of the three different methods on the systems H_2, He, LiH, and Li_2. $\bar{\varepsilon}_0$ is the estimated ground state energy (fixed-node approximation) with the number in parentheses representing the statistical error in the last digit. β_{DMC}, $\langle\beta\rangle_{\mathrm{KLV}}$, and $\langle\beta\rangle$ are the average time steps in the three methods. $\langle q\rangle_{\mathrm{KLV}}$ and $\langle q\rangle$ are the average promotional probabilities. ζ_{DMC}, ζ_{KLV}, and ζ are the average errors of the three methods, where ζ is the estimated error bar of the energy times the square root of the CPU times (CDC 7600) in seconds. $\langle(M-1)^2\rangle$ is the rms multiplicity for GFMC and $\langle(M-1)^2\rangle^{1/2}_{\mathrm{DMC}}$ is the rms multiplicity in a DMC calculation at the same timestep. All units are atomic units: energies in hartrees, time in inverse hartrees.

There is a much greater fluctuation in the branching multiplicity in the present GFMC method than in DMC. Table I shows the rms value of the multiplicities for GFMC. In DMC [8] the multiplicity is

$$M_{\mathrm{DMC}} = \exp[-\beta(\tfrac{1}{2}(E_{\mathrm{L}}(R) + E_{\mathrm{L}}(R')) - \varepsilon_0)], \tag{51}$$

where $E_{\mathrm{L}}(R)$ is the local energy at R. The rms value of M_{DMC} for β equal to the mean diffusion time in GFMC is also shown in Table I. Hence the exact method has fluctuations in branching from 5 to 50 times that of DMC. In fact, for these molecules, the fluctuations in the branching ratios, rather than the local trial density matrix energy, determine the maximum diffusion time, since it is unwise to pick a diffusion time (τ_{G}) such that the rms value of M is much greater than unity. The Neumann series (Eq. (3)) may not converge in such a case.

The multiplicities in DMC of course are only approximate. The correct formula is

$$M_{\mathrm{DMC}} = \langle e^{-\int_0^\beta dt(E_{\mathrm{L}}(t) - \varepsilon_0)}\rangle_{R\to R'}, \tag{52}$$

where $\langle \cdots \rangle_{R\to R'}$ means the average over all drifting random walks from R to R'. It may be that the two point formula of Eq. (51) simply underestimates the fluctuations of the local energy in the neighborhood of R and R'. Hence the difference in the branching ratios may be simply due to the difference between an approximate and an exact method. It seems likely that a major reduction of the fluctuations in branching can be achieved by a better trial density matrix ρ_T and by a much better sampling of $\hat{g}_\beta$. In any case the average error seems not to have been affected too seriously by the fluctuations in branching and no difficulties have been encountered in stabilizing the population.

In summary, this generalized Green's Function Monte Carlo method has been shown to be more efficient than the KLV method while still being exact. The efficiency is somewhat better than the diffusion Monte Carlo.

APPENDIX

This appendix derives some of the properties of the pair Coulomb trial density matrix and indicates how it can be computed. It is well known that the Coulomb potential possesses a unique constant of motion, the Runge–Lenz vector [17]. This implies that the Coulomb density matrix $\rho(\mathbf{r}_1, \mathbf{r}_2; \beta)$ depends only on $r_1 + r_2$ and r_{12} and not on $r_1 - r_2$ as it would for any other potential. This symmetry must also hold for $u(\mathbf{r}_1, \mathbf{r}_2; \beta)$ since it is the linear term in a power series expansion of the density matrix. It is convenient to work with the variables

$$\begin{aligned} \gamma &= e_i e_j (2\beta\mu)^{1/2}/\hbar, \\ y &= (r_1 + r_2 - r_{12})(2\hbar^2\beta/\mu)^{-1/2}, \qquad\qquad (\mathrm{A1}) \\ s &= r_{12}(2\hbar^2\beta/\mu)^{-1/2}, \end{aligned}$$

since for these variables the physical region is the quarter plane ($y > 0$, $s > 0$). Large values of s are not required since the sampling density is proportional to $\exp(-s^2)$. From Eqs. (42) and (44)

$$u(r_1, r_2; \beta) = \gamma \int_0^1 d\lambda\, \mathrm{erf}(\tfrac{1}{2}[y + 2s\lambda]\,[\lambda(1-\lambda)]^{-1/2})/[y + 2s\lambda]. \qquad (\mathrm{A2})$$

Using the definition of the error function and interchanging the integrals this can be rewritten in the form

$$u = \gamma \left\{ \ln(1 + 2s/y)/2s - \pi^{-1/2} s^{-1} \int_{x_0}^{\infty} dx\, e^{-x^2} \ln\left(\frac{1+\varepsilon}{1-\varepsilon}\right) \right\}, \qquad (\mathrm{A3})$$

where

$$x_0 = (y^2 + 2sy)^{1/2}, \qquad \varepsilon = \frac{(x^2 - x_0^2)^{1/2}}{x(1 + y/s)}. \qquad (\mathrm{A4})$$

As long as $x_0 \gg 1$ the second term can be neglected since it will be of order $\exp(-x_0^2)$. An expansion of the logarithm term in the integrand in powers of ε leads to an asymptotic expansion of u in terms of confluent hypergeometric functions [18].

$$u = \gamma \left\{ \ln(1 + 2s/y)/2s - x_0 \pi^{-1/2} s^{-1} e^{-x_0^2} \sum_{k=1,3,5}^{\infty} \frac{\Gamma(1 + k/2)}{k(1 + y/s)^k} U\left(1 + \frac{k}{2}, \frac{3}{2}, x_0^2\right) \right\}. \tag{A5}$$

Along the diagonal ($s = 0$) only the $k = 1$ term survives, hence u is a confluent hypergeometric function there. Note that $\exp(-x_0^2)$ is just proportional to the probability density that a Weiner path will hit the origin ($\mathbf{r} = 0$) in going from $\mathbf{r}_1$ to $\mathbf{r}_2$. If this is sufficiently small, all but the first term in Eq. (A3) can be neglected.

To compute u in the course of the random walk, u, $\partial_y u$, $\partial_s u$, and $\partial_y \partial_s u$ have been tabulated in a 20×20 grid for $0 < y < 3.5$ and $0 < s < 3.5$. In constructing this table the spacial values $\partial_y u(0, s) = -\gamma$, $\partial_y u(y, 0) = \partial_s u(y, 0)$, and $u(0, 0) = \gamma\pi^{1/2}$ are useful. Inside the grid, a bicubic Hermite interpolation formula [19] determines u from the values on the grid. Outside the grid the asymptotic logarithmic term is used. The error of this procedure is about 1 part in 10^{-6}.

For periodic systems the pair trial density matrix is given in terms of the infinite system one by Eq. (47). The time step β is usually small enough so that only one term in the sum of Eq. (45) is important since there is a very small probability of a particle diffusing the size of a box length in one step. For an accuracy of 10^{-6} in u^P this demands that

$$L > 7.4(\hbar^2\beta/2\mu)^{1/2}. \tag{A6}$$

Then with this assumption

$$u^P(\mathbf{r}_1, \mathbf{r}_2; \beta) = \sum_n u(\mathbf{r}_1 + Ln, \mathbf{r}_2 + Ln; \beta). \tag{A7}$$

Using the asymptotic expansion for u and the definition of the Ewald image potential $v^P(\mathbf{r})$, and again using (A6),

$$\begin{aligned} u^P(\mathbf{r}_1, \mathbf{r}_2; \beta) = \beta \int_0^1 d\lambda\, v^P(\lambda \mathbf{r}_1 + (1 - \lambda)\, \mathbf{r}_2) \\ + u(\mathbf{r}_1, \mathbf{r}_2; \beta) - \frac{\gamma}{2s} \ln(1 + 2s/y). \end{aligned} \tag{A8}$$

The integral in (A8) can be estimated in terms of its value and derivative at the end points by the Euler–MacClaurin formula [18].

$$u^P(\mathbf{r}_1, \mathbf{r}_2; \beta) = \frac{\beta}{2}(v^P(\mathbf{r}_1) + v^P(\mathbf{r}_2)) - \frac{\beta}{12}(\nabla v^P(\mathbf{r}_1) - \nabla v^P(\mathbf{r}_2))\,\mathbf{r}_{12}$$

$$+ u(\mathbf{r}_1, \mathbf{r}_2; \beta) - \beta e_i e_j \left\{ \frac{1}{2}(r_1^{-1} + r_2^{-1}) + \frac{1}{12}\left(\frac{\mathbf{r}_1}{r_1^3} - \frac{\mathbf{r}_2}{r_2^3}\right)\mathbf{r}_{12} \right\}, \tag{A9}$$

where the same approximation has been made for the logarithm term so that singularities at the origin do not occur. The error in the Euler–MacClaurin formula is less than

$$\left[\frac{\hbar^2}{2\mu}\beta\right]^2 \frac{e_i e_j \beta}{720 r^5}. \tag{A10}$$

The density matrix has been broken into a periodic term (Ewald potential and electric field) and a short ranged correction which is independent of the boundaries. With periodic boundary conditions the trial-density-matrix-local-energy contains an additional constant term

$$E_\beta = \frac{\pi\beta\hbar^2}{3\Omega} \sum_{i \neq j} \frac{e_i e_j}{\mu_{ij}} - \sum_i \frac{\hbar^2}{2m_i} \mathbf{F}_i^2, \tag{A11}$$

where $\mathbf{F}$ is defined in Eq. (37) and Ω is the volume of the periodic cell. By charge neutrality, this term is only appreciable for a one-component plasma and comes from the neglect, in the trial density matrix, of the plasmons.

ACKNOWLEDGMENTS

The author thanks Dr. E. L. Pollock for many fruitful discussions on the Coulomb density matrix and for suggestions concerning this manuscript. This work was performed under the auspices of the U. S. Department of Energy by the Lawrence Livermore National Laboratory under contract No. W-7405-ENG-48.

REFERENCES

1. N. METROPOLIS AND S. ULAM, *J. Amer. Stat. Assoc.* **44** (1949), 335.
2. R. COURANT, K. FRIEDRICHS, AND H. LEWY, *Math. Ann.* **100** (1928), 32.
3. M. H. KALOS, *Phys. Rev.* **128** (1962), 1791; *J. Comput. Phys.* **1** (1966), 257; *Phys. Rev. A* 2 (1970), 250.
4. G. GOERTZEL AND M. H. KALOS, *in* "Progress in Nuclear Energy" (D. J. Hughes, Ed.), 368, Series I, Vol. 2, Pergamon, New York, 1958.
5. M. H. KALOS, D. LEVESQUE, AND L. VERLET, *Phys. Rev. A* **9** (1974), 2178.
6. P. A. WHITLOCK, M. H. KALOS, G. V. CHESTER, AND D. M. CEPERLEY, *Phys. Rev. B* **19** (1979), 5598.
7. D. M. CEPERLEY AND M. H. KALOS, *in* "Monte Carlo Methods in Statistical Physics" (K. Binder, Ed.), 145, Springer-Verlag, New York/Berlin, 1979.
8. J. B. ANDERSON, *J. Chem. Phys.* **63** (1975), 1499; *J. Chem. Phys.* **65** (1976), 4121; *J. Chem. Phys.* **73** (1980), 3897; *J. Chem. Phys.* **74** (1981), 6307.
9. J. W. MOSKOWITZ, K. E. SCHMIDT, M. A. LEE, AND M. H. KALOS, *J. Chem. Phys.* **77** (1982), 349.
10. P. J. REYNOLDS, D. M. CEPERLEY, B. J. ALDER, AND W. A. LESTER, *J. Chem. Phys.* **77** (1982), 5593.
11. D. M. CEPERLEY AND B. J. ALDER, *Phys. Rev. Lett.* **45** (1980), 566.
12. D. M. CEPERLEY AND B. J. ALDER, *Physica B & C* **107** (1981), 875.
13. D. M. CEPERLEY, to be published.
14. S. CHANDRASEKHAR, *Rev. Mod. Phys.* **15** (1943), 1.
15. R. P. FEYNMAN, "Statistical Mechanics," p. 72, Benjamin, New York, 1972; M. KAC, "Probability and Related Topics in Physical Science," p. 165, Interscience, New York, 1959.
16. S. G. BRUSH, E. TELLER, AND H. L. SAHLIN, *J. Chem. Phys.* **45** (1966), 2101.
17. L. HOSTLER, *J. Math. Phys.* **8** (1967), 642.
18. M. ABRAMOWITZ AND I. A. STEGUN, "Mathematical Functions," Dover, New York, 1965.
19. C. DE BOOR, "A Practical Guide to Splines," p. 49, Springer-Verlag, New York, 1978.