JOURNAL OF COMPUTATIONAL PHYSICS 96, 258–276 (1991)

Monte Carlo Algorithms for Expectation Values of Coordinate Operators*

R. N. BARNETT, P. J. REYNOLDS,† AND W. A. LESTER, JR.

Materials and Chemical Sciences Division,
Lawrence Berkeley Laboratory and Department of Chemistry,
University of California, Berkeley, California 94720

Received March 23, 1990; revised October 25, 1990

Two Monte Carlo algorithms for computing quantum mechanical expectation values of coordinate operators, i.e., multiplicative operators that do not commute with the Hamiltonian, are presented and compared. The first employs a single quantum Monte Carlo (QMC) random walk, while the second involves a variational Monte Carlo (VMC) random walk with auxiliary QMC "side walks." The tagging algorithm used for efficiently tracking descendants of a walker is described in detail for each approach. For the single-walk algorithm it is found that carrying weights together with branching significantly improves efficiency. Exploitation of the correlation between VMC and QMC expectation values is also considered. Large increases in efficiency in the second approach are found when such correlations are incorporated. It is found that both approaches readily yield accuracies and precisions of better than 0.5% for the model systems treated here, namely, H and H_2. The second method, involving a VMC walk with auxiliary QMC walks, is the more efficient for these systems.

I. INTRODUCTION

Over the last 10 to 15 years, Monte Carlo techniques have been increasingly applied to quantum mechanical problems [1]. Of these "quantum Monte Carlo" (QMC) methods, a subclass called Green's function Monte Carlo (GFMC) [2] has been employed in obtaining stochastic solutions of the Schrödinger equation for atomic and molecular systems. The focus of most of these approaches has been the accurate computation of the total electronic energy of small atoms and molecules. Other energy-related quantities have also been computed [3]. However, since energy is only one of many important properties, it is desirable to be able to evaluate expectation values of operators other than the Hamiltonian, H. Unfortunately, for properties whose operators do not commute with H, such as functions of coordinates from which static moments of the charge distribution may be

* This work was supported in part by the Office of Naval Research through an interagency agreement with the U.S. Department of Energy under Contract DE-AC03-76SF00098.

† Present address: Physics Division, Office of Naval Research, 800 North Quincy Street, Arlington, VA 22217.

0021-9991/91 $3.00

obtained, the usual evaluation of quantum Monte Carlo averages (the so-called mixed expectation values) is not exact [4].

Approaches for modifying the GFMC approach to compute expectation values with respect to the square of the wave function, rather than the mixed product of this function and a trial wave function, have been proposed and developed by Kalos and others [5–7]. We follow Liu *et al.* [7] in exploring modifications to the diffusion QMC approach [8, 9] to obtain exact ("pure"), rather than mixed, expectation values [4, 7, 10, 11].

Within the class of GFMC techniques [9, 12–15] diffusion QMC is pedagogically the clearest and is summarized for present purposes. For a full discussion see Refs. [8, 9]. An important antecedent is the work of Anderson [16]. The development below can also be generalized to other forms of GFMC.

Starting with the time-dependent Schrödinger equation in imaginary time multiplied by a trial function, Ψ_T (for importance sampling), one obtains (in atomic units)

$$-\frac{\partial f(\mathbf{R}, t)}{\partial t} = -\frac{1}{2}\nabla^2 f + (E_L(\mathbf{R}) - E_R) f + \frac{1}{2}\nabla \cdot (f F_Q(\mathbf{R})). \tag{1}$$

Equation (1) gives the evolution of $f(\mathbf{R}, t) \equiv \Psi_T(\mathbf{R})\, \Phi(\mathbf{R}, t)$ in imaginary time, where $\Phi(\mathbf{R}, t)$ is the state function. The other quantities are the local energy, $E_L(\mathbf{R}) \equiv \Psi_T(\mathbf{R})^{-1} H \Psi_T(\mathbf{R})$, and the quantum "force" or drift velocity, $F_Q(\mathbf{R}) \equiv 2\Psi_T(\mathbf{R})^{-1} \nabla \Psi_T(\mathbf{R})$; $\mathbf{R}$ is the $3N$-dimensional coordinate vector of the quantum particles (electrons in the calculations here) while E_R, the reference energy, is simply a shift in the zero of energy. Interpreted in the light of the Monte Carlo simulations we will describe, the terms on the right-hand side of Eq. (1) correspond to diffusion, branching, and drift, respectively. For this interpretation to hold requires that f be both real and positive. This is a nontrivial requirement, but it can be satisfied [9, 15]. By following the evolution of f, one notes that after a sufficient time t, the excited-state components in the eigenfunction expansion of $\Phi(\mathbf{R}, t)$ become negligible, leaving f proportional to the ground state. This follows since, in imaginary time, the amplitudes of the excited states relative to the ground state depend on time as $\exp[-t(E_i - E_0)]$. Therefore, the dominant term remaining at large t is

$$f(\mathbf{R}, t) = c_0 \exp[-t(E_0 - E_R)]\, \Psi_T(\mathbf{R})\, \varphi_0(\mathbf{R}), \tag{2}$$

where ϕ_0 is the lowest energy eigenfunction of the Hamiltonian not orthogonal to Ψ_T, and E_0 is the corresponding eigenvalue. The coefficient c_0 results from the expansion of $\Phi(\mathbf{R}, 0)$ in the complete set $\{\phi_i\}$. The time dependence of $f(\mathbf{R}, t)$ is largely eliminated by choosing E_R to approximate E_0 as accurately as possible. Since f converges exponentially to its asymptotic distribution, $\Psi_T \phi_0$ may generally be sampled with great accuracy after modest simulation times.

The procedure by which $f(\mathbf{R}, t)$ is obtained from $f(\mathbf{R}, 0)$ involves the Green's function propagator, $G(\mathbf{R} \to \mathbf{R}', t)$ [7–15]. This function is a solution of Eq. (1)

with the boundary condition $G(\mathbf{R} \to \mathbf{R}', 0) = \delta(\mathbf{R}' - \mathbf{R})$ and gives the probability of moving from $\mathbf{R}$ to $\mathbf{R}'$ in time t. Though an analytic form of G is known for only the simplest systems, one may nevertheless use Monte Carlo to sample the exact G [5–7]. A more straightforward approach, however, is to employ an analytic short-time approximation to G [8, 9], namely, $G_a(\mathbf{R} \to \mathbf{R}', \tau)$, which becomes exact as $\tau \to 0$. Thus, convergence to large t is accomplished by many moves with small time steps τ. Because the asymptotic distribution generally will not correspond exactly to $\Psi_T \phi_0$ for non-zero τ, it is necessary to use sufficiently small steps to suppress the bias below the level of statistical uncertainty. Alternatively, calculations at several τ values may be performed, followed by an extrapolation to $\tau = 0$. Though there is additional overhead, its simplicity relative to exact GFMC makes the short-time approximation useful.

Having obtained the asymptotic distribution f, one may sample "mixed" expectation values of a time-independent operator A by averaging $\Psi_T^{-1} A \Psi_T$ over $f = \Psi_T \phi_0$, i.e.,

$$A_m \equiv \lim_{N \to \infty} \frac{1}{N} \sum_i \Psi_T(\mathbf{R}_i)^{-1} A \Psi_T(\mathbf{R}_i) \tag{3a}$$

$$= \frac{\int \Psi_T(\mathbf{R}) \phi_0(\mathbf{R}) [\Psi_T(\mathbf{R})^{-1} A \Psi_T(\mathbf{R})] \, d\mathbf{R}}{\int \Psi_T(\mathbf{R}) \phi_0(\mathbf{R}) \, d\mathbf{R}} = \frac{\langle \phi_0 | A | \Psi_T \rangle}{\langle \phi_0 | \Psi_T \rangle}. \tag{3b}$$

This is to be distinguished from the exact or "pure" expectation value $A_p = \langle \phi_0 | A | \phi_0 \rangle$. (Here and throughout this paper all wave functions are assumed to be normalized.) Since ϕ_0 is an eigenfunction of H, $A_m = A_p$ if A is an Hermitian operator that commutes with H. This follows since, for $A\phi_0 = a_0 \phi_0$, one has

$$\langle \phi_0 | A | \Psi_T \rangle / \langle \phi_0 | \Psi_T \rangle = a_0 = \langle \phi_0 | A | \phi_0 \rangle. \tag{4}$$

For operators that do not commute with H, e.g., the dipole and higher moment operators, the mixed expectation value is only an approximation to the pure one [7]. The mixed average is accurate to first order in quantities which depend on the difference function, $\delta \equiv \phi_0 - \Psi_T$, i.e., the quantities $\langle \phi_0 | \delta \rangle$ and $\langle \phi_0 | A | \delta \rangle$. Also accurate to first order in δ is the "trial" expectation value, $A_T \equiv \langle \Psi_T | A | \Psi_T \rangle$, computed by sampling from $|\Psi_T|^2$ in a procedure often referred to as variational Monte Carlo (VMC). VMC may be cast in a form algorithmically identical to diffusion QMC, except that the branching, cf. Eq. (1), is suppressed. The usefulness of A_T in this context is that it can be combined with A_m to obtain an estimate of A_p accurate to *second* order in δ, namely [4, 10],

$$A_s \equiv 2A_m - A_T = A_p + O(\delta^2). \tag{5}$$

Though this approximation is generally better than either the trial or mixed estimators, it may still be poorer than necessary for reliable numerical predictions. For this reason it is desirable to compute pure expectation values exactly

Nevertheless, early work [4] essentially abandoned pure expectation values because of their large statistical noise relative to second-order estimators (5). Thus, to obtain the desired accuracy of the pure estimators efficient algorithms are required. These are explored in this paper. Other approaches have also been recently considered [11].

The remainder of this paper consists of four sections. In Section II, we present the theory and an efficient algorithm for sampling the ratio ϕ_0/Ψ_T *during* the QMC walk. In this approach, the computation of pure expectation values requires minimal change from the QMC approach used for mixed averages. No auxiliary walks are required, only a tagging algorithm for keeping track of offspring. In Section III, we describe a second algorithm that enables the computation of trial, mixed, and pure expectation values in a single calculation. Section IV discusses an approach for reducing statistical error in averages of odd functions (of which the dipole moment is an example). Finally, Section V presents our results for two model systems. We calculate the moments $\langle r\rangle$, $\langle z^2\rangle$, and $\langle r^2\rangle$ of H and $\langle z^2\rangle$ and $\langle r^2\rangle$ of H_2. The efficiency and advantages of each method are discussed. The Appendix treats the advantages associated with carrying weights for several generations before branching is performed.

II. Pure Expectation Values with a Single QMC Walk

Since the QMC approach described in Section I yields only mixed expectation values, the ratio ϕ_0/Ψ_T must be sampled in order to obtain pure expectation values. One has

$$\frac{\langle A(\phi_0/\Psi_T)\rangle_f}{\langle \phi_0/\Psi_T\rangle_f} = \frac{\langle \Psi_T|\, A(\phi_0/\Psi_T)\,|\phi_0\rangle}{\langle \Psi_T|\phi_0\rangle} \Big/ \frac{\langle \Psi_T|\,\phi_0/\Psi_T\,|\phi_0\rangle}{\langle \Psi_T|\phi_0\rangle}$$

$$= \langle \phi_0|\, A\,|\phi_0\rangle = A_p, \tag{6}$$

where the subscript f denotes the average over $\Psi_T\phi_0$. As shown by Liu *et al.* [7], ϕ_0/Ψ_T may be obtained from the asymptotic number of descendants resulting from a QMC walk which starts at $\mathbf{R}$. For completeness, we modify their proof for diffusion QMC.

An initial distribution given by a single point at $\mathbf{R}$, is

$$f_{\mathbf{R}}(\mathbf{R}', 0) = \Psi_T(\mathbf{R}')\,\Phi(\mathbf{R}', 0) = \delta(\mathbf{R}' - \mathbf{R}). \tag{7}$$

Expanding the initial state function, Φ, in the complete set of (normalized) eigenfunctions of H yields

$$\delta(\mathbf{R}' - \mathbf{R}) = \Psi_T(\mathbf{R}') \sum c_j \phi_j(\mathbf{R}'). \tag{8}$$

The expansion coefficients may be obtained by multiplying Eq. (8) by $\phi_i(\mathbf{R}')/\Psi_T(\mathbf{R}')$ and integrating over $\mathbf{R}'$, giving

$$c_i = \phi_i(\mathbf{R})/\Psi_T(\mathbf{R}). \tag{9}$$

From the asymptotic form of f, given by Eq. (2), we note that $f_{\mathbf{R}}(\mathbf{R}', 0)$ evolves to

$$f_{\mathbf{R}}(\mathbf{R}', t) = c_0 \exp[-t(E_0 - E_R)]\ \Psi_T(\mathbf{R}')\,\phi_0(\mathbf{R}'), \tag{10}$$

with c_0 given by Eq. (9). Integrating Eq. (10) over all space "counts" the current number of descendants (at time t), or for large t, the asymptotic population $P(\mathbf{R})$ of a walker starting at $\mathbf{R}$, i.e.,

$$P(\mathbf{R}) \equiv \int f_{\mathbf{R}}(\mathbf{R}')\,d\mathbf{R}' = \phi_0(\mathbf{R})/\Psi_T(\mathbf{R}) \exp[-t(E_0 - E_R)]\langle \Psi_T | \phi_0 \rangle. \tag{11}$$

Returning to Eq. (6) for the evaluation of A_p, we see that ϕ_0/Ψ_T may be replaced by P since the overlap integral and time dependence present in P will cancel.

Tagging Algorithm for Counting Descendants

In order to count efficiently the descendants of a walker *during* the QMC walk, we have developed the following algorithm, which we have previously described briefly in Ref. [10]. A similar approach has been advocated recently by Runge and Runge [17]. For an arbitrary walker at time 0, its descendants at time t later (or, equivalently, N steps later where $N\tau = t$) are counted to obtain $P \propto \phi_0/\Psi_T$. This process may be repeated for different points sampled from $\Psi_T\phi_0$. However, this would yield an additional start-up cost by propagating each initial point a distance of N time steps to reach the asymptotic domain of P. Instead, our algorithm uses each step of a single (branching) walk; that is, by propagating for additional time steps to a time $t + \tau$, the "N-distant" offspring of the first generation (at time τ) may be computed. Likewise, at time $t + 2\tau$, N-distant offspring populations may be determined for walkers which may be thought of as beginning at time 2τ. Therefore, after an initial investment of N steps, additional N-distant values of ϕ_0/Ψ_T may be sampled for each subsequent step in the walk. Furthermore, since N itself is arbitrary, convergence of the asymptotic population as a function of N may also be determined.

The branching of the QMC walk, however, requires tagging walkers. When the branching yields several walkers, say at a time $m\tau$, one must know which of those at $(m+N)\tau$ are descended from which of those at $m\tau$. To store this information a "family tree" is created as the walk progresses, and each walker is labelled so that its location in the tree is specified. This labelling is accomplished with two tags [10]. The first, θ $(0 \leqslant \theta < \Theta)$, gives (together with t) a walker's location in the tree, while the second, δ, specifies that area (range) within the tree in which descendants of this walker will be placed. The tagging begins by setting $\theta(k = 1, t = 0) = 0$, where the index k labels the walker, and $\delta(k = 1) = \Theta$. If a walker "dies," i.e., has no

offspring, then no subsequent values of θ and δ are derived from it. When the kth walker has n_k (>0) immediate offspring, the values of θ and δ assigned to these daughters are obtained from $\theta(k, t)$ and $\delta(k)$ according to

$$\theta(M+l, t+\tau) = \theta(k, t) + \frac{l-1}{n_k}\,\delta(k) \tag{12a}$$

and

$$\delta(M+l) = \delta(k)/n_k, \tag{12b}$$

where l ranges from 1 to n_k. The tags are properly sequenced if $M \equiv \sum_{i=1}^{k-1} n_i$, where M is the partial sum (up to walker $(k-1)$) of the number of walkers in the new generation. The example in Fig. 1 follows a few generations out from a single walker at the origin. The shadings indicate the range in which each walker's offspring are placed. Using the algorithm described, the descendants of a walker with $\theta = \theta(k, t)$ and $\delta = \delta(k)$ lie in the range, S, given by

$$S = [\theta(k, t),\ \theta(k, t) + \delta(k)). \tag{13a}$$

From Eqs. (12), $\theta(k, t) + \delta(k) \leqslant \theta(k+1, t)$. (The equality holds when no walkers die.) Since the descendants of the next walker, $(k+1)$, possess values of $\theta \geqslant \theta(k+1, t)$, the θ values of the walkers descended from walker k also lie within

$$S' = [\theta(k, t),\ \theta(k+1, t)). \tag{13b}$$

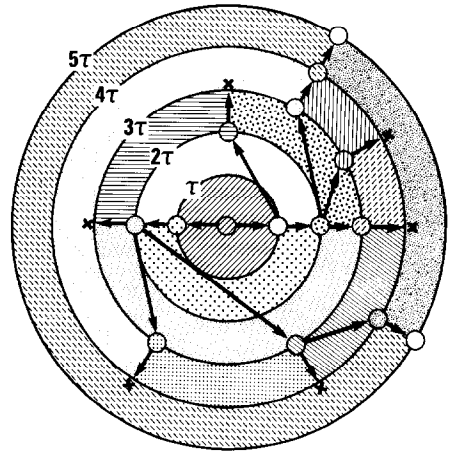

FIG. 1. Tagging algorithm for single QMC walk method. A "family tree" for a single walker starting at the origin is shown. Each ring outward corresponds to one generation (or an increase in time by τ). Location in the diagram identifies a walker with its $\theta(k, t)$ label. All descendants of a walker will be in the range of angles from θ to $\theta + \delta$, making identification of progeny possible for all future generations. This enables one to determine convergence to asymptotic populations in a single calculation. Shadings show the angular descendant space for each walker. An ensemble of trees may be readily treated as described in the text.

With this choice, only θ values need to be stored to compute the range in which the descendants of a given walker are located. This choice is illustrated in Fig. 1.

To verify that the range in Eq. (13a) follows from Eqs. (12), we consider the values of θ possessed by the descendants, at time $t+N\tau$, of a walker with $\theta=\theta(k,t)$ and $\delta=\delta(k)$. Repeated use of Eqs. (12) readily shows that a particular descendant, m, will be located at

$$\theta(m,t+N\tau)=\theta(k,t)+\frac{l_1-1}{n_1}\delta(k)+\frac{l_2-1}{n_2}\cdot\frac{\delta(k)}{n_1}+\cdots+\frac{l_L-1}{n_L}\cdot\frac{\delta(k)}{n_1n_2\cdots n_{L-1}}. \tag{14}$$

The labels l_i, $\{1\leqslant l_i\leqslant n_i\}$, correspond to the ancestral lineage of m, while n_i (>0) give the number of daughters in each family leading to m. That is, there are n_1 daughters of walker k; daughter l_1, who has n_2 daughters, is the direct ancestor of m in this generation and so on. The minimum value of $\theta(m,t+N\tau)$, corresponding to the initial walker never dying, is obtained by setting $\{l_i\}=1$, which yields $\theta_{\min}(m,t+N\tau)=\theta(k,t)$. The maximum θ, corresponding to the walker with the largest value of θ at each time step never dying, results from setting each l_i to its maximum value (n_i) yielding

$$\begin{aligned}\theta_{\max}(m,t+N\tau)&=\theta(k,t)+\frac{n_1-1}{n_1}\delta(k)+\frac{n_2-1}{n_2}\cdot\frac{\delta(k)}{n_1}+\cdots+\frac{n_L-1}{n_L}\cdot\frac{\delta_k}{n_1n_2\cdots n_{L-1}}\\&=\theta(k,t)+\delta(k)\left(1-\frac{1}{n_1n_2\cdots n_L}\right).\end{aligned} \tag{15}$$

Therefore, the range of walker k is indeed given by Eq. (13a). Since this range is contained within $[\theta(k,t),\ \theta(k+1,t))$, the "descendant spaces" of walkers never overlap.

With the assignment of labels to walkers as described above, asymptotic populations are readily sampled. At an arbitrary time, t_0, the number of descendants of point k at a later time t_0+t is given by

$$P(\mathbf{R}_k,t)=\sum_i I_i,\qquad I_i=1,\quad \text{if}\quad \theta(k,t)\leqslant\theta(i,t_0+t)<\theta(k+1,t),$$
$$I_i=0,\qquad \text{otherwise.} \tag{16}$$

As mentioned earlier, only one tree is necessary. However, to take advantage of the vector capabilities of current machines, a group of family trees may be created and asymptotic populations of several independent points obtained. This is easily accomplished for M_p initial points by setting

$$\theta_i(1,0)=(i-1)\Theta,\qquad 1\leqslant i\leqslant M_p. \tag{17}$$

For each initial point $\delta = \Theta$, new values of θ and δ are computed according to Eqs. (12). Again, the descendant spaces are given by either (13a) or (13b), though with integer multiples of Θ added to each bound.

Finally, it should be noted that with this algorithm there is the need for some additional memory overhead in the storage of $\theta(k, t)$. This can become excessive when, in performing the QMC walk, the time step τ must be very small to obtain low bias. However, because successive values of the function to be sampled, $A(\mathbf{R})$, are highly correlated, A and P need not be sampled at every point along the walk. By sampling only every n steps, these memory requirements are reduced by a factor of n.

Branching Algorithms with Weighting

To reduce the statistical error in sampled values of ϕ_0/Ψ_T, we explore a variant of the usual branching generally used in QMC walks. Upon completing a move, $\mathbf{R} \to \mathbf{R}'$, the most common implementation of branching is to obtain an integer, I_b, which specifies the number of walkers at $\mathbf{R}'$. The number of descendants, I_b, is $\text{int}[b(\mathbf{R}, \mathbf{R}') + \xi]$, where $b(\mathbf{R}, \mathbf{R}')$ is the weight of $\mathbf{R}'$ relative to $\mathbf{R}$, and ξ is is a uniform random variate between 0 and 1. In the diffusion QMC approach, $b(\mathbf{R}, \mathbf{R}')$ is the branching factor from G_a [8, 9]. While this rounding is correct on average, i.e., $\bar{I}_b = b$, a "microscopically" exact procedure is to weight each walker by the product of its current weight and the branching factor b. The drawback to this purely weighting procedure is that, since the product of these weights tends to either 0 or ∞, efficiency is lost with computations on walkers that contribute very little information due to their low (absolute or relative) weights.

A combination of branching and weighting, however, *is* useful. In this case we omit integer rounding until a weight becomes exceedingly small or large. When the weight w becomes large, an integer I_w is determined from it, as described above; however the daughters are assigned weights of I_w/w, rather than unity, so that no loss of information occurs. When the weight becomes smaller than a threshold value, integer rounding *is* applied. The benefit of this modification is that the variance of the asymptotic populations, P, now given as the sum of the weights of descendants, is noticeably reduced. This leads to improved precision in pure expectation values. See the Appendix for an analysis of the comparative variances.

III. Pure Expectation Values by VMC with QMC "Side Walks"

In this section we explore an approach that begins with VMC, rather than QMC, sampling of points from the distribution $|\Psi_T|^2$. These walks are very efficient, can employ large time steps, and have no bias. The points obtained are then initial points for QMC "side walks," performed to obtain P and thereby ϕ_0/Ψ_T. Before implementing the QMC walk, the starting coordinates of the initial points are stored so that the VMC walk may be continued after values of ϕ_0/Ψ_T are sampled. Since values of P are computed only for points at the beginning of each QMC walk,

the labelling process of the previous section is greatly simplified. For example, for M walkers drawn from $|\Psi_T|^2$, $\theta(i)=i$ and $\delta(i)=0$. Under these conditions the descendants of the ith walker are simply those at the end of the QMC walk with $\theta=i$, cf. Fig. 2.

Since initial points are selected from $|\Psi_T|^2$, the factor $|\phi_0/\Psi_T|^2$ is necessary to obtain pure expectation values. To obtain it, at least two *independent* samplings of ϕ_0/Ψ_T must be performed [11b]. This is because the square of the asymptotic population P is not an unbiased estimate of $|\phi_0/\Psi_T|^2$ even though, on average, P is given by ϕ_0/Ψ_T. One way to show this is to write

$$P(\xi)=\phi_0/\Psi_T+\eta(\xi), \tag{18}$$

where ξ corresponds to a specific QMC walk and $\eta(\xi)$ is the fluctuation in this walk. For convenience the time dependence is not displayed. Denoting the probability distribution of walks by $h(\xi)$, there results

$$\int h(\xi)\,P(\xi)\,d\xi=\phi_0/\Psi_T+\int h(\xi)\,\eta(\xi)\,d\xi=\phi_0/\Psi_T. \tag{19}$$

The term, $\int h(\xi)\,\eta(\xi)\,d\xi$ vanishes because the average $\bar{\eta}=0$ (i.e., $\bar{P}=\phi_0/\Psi_T$). Note, however, that squaring the population yields

$$\int h(\xi)\,P^2(\xi)\,d\xi=|\phi_0/\Psi_T|^2+\int h(\xi)\,\eta^2(\xi)\,d\xi, \tag{20}$$

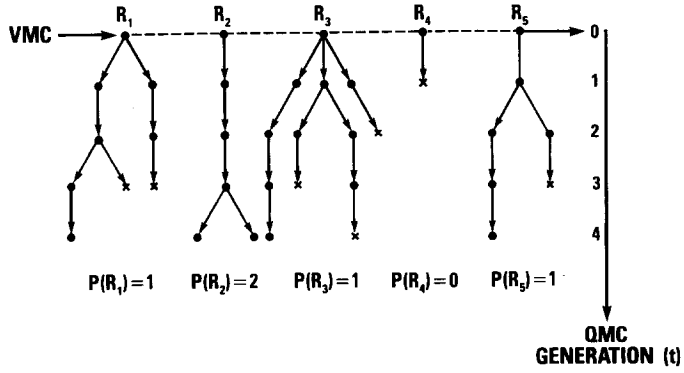

FIG. 2. Tagging algorithm for VMC with QMC side walks method. The family trees generated are shown for five points sampled from $|\Psi_T|^2$. Since only the descendants of the points beginning each QMC walk are tracked, the tagging algorithm is very simple, as indicated here and discussed in the text. Weights are carried with the branching walkers so that the asymptotic population of a point is the sum of weights of its descendants at a sufficiently large t. For each initial point, *two* QMC walks are employed in order to obtain two statistically independent samplings of ϕ_0/Ψ_T and, therefore, an unbiased estimate of $|\phi_0/\Psi_T|^2$. As in the single QMC walk approach, results may readily be computed for several convergence times (t).

in which the second term on the right-hand side of Eq. (20) does *not* vanish. On the other hand, the product of two asymptotic populations obtained independently is

$$\int h(\xi_1)\, P(\xi_1)\, d\xi_1 \cdot \int h(\xi_2)\, P(\xi_2)\, d\xi_2 = |\phi_0/\Psi_T|^2. \tag{21}$$

Thus, by sampling A and two values of P at points selected from $|\Psi_T|^2$, trial, mixed, and pure expectation values may be computed. These averages are, respectively,

$$A_T = \langle \Psi_T |\, A \,| \Psi_T \rangle,$$

$$A_m = \langle \Psi_T |\, [P(\xi_1) + P(\xi_2)] A \,| \Psi_T \rangle / \langle \Psi_T |\, [P(\xi_1) + P(\xi_2)] \,| \Psi_T \rangle,$$

and

$$A_p = \langle \Psi_T |\, P(\xi_1)\, P(\xi_2) A \,| \Psi_T \rangle / \langle \Psi_T |\, P(\xi_1)\, P(\xi_2) \,| \Psi_T \rangle. \tag{22}$$

Although two walks and two samplings of P are required for A_p, an increase in efficiency is possible, since A and P are sampled from a VMC walk which possesses no time-step bias and allows large time steps to be used [9], thereby sampling configuration space more efficiently. Another benefit is that in this approach (which we hereafter denote V + QMC) the statistical error in $A_p - A_T$ is generally much smaller than that of A_p alone, since A_p and A_T are correlated. This is useful because trial expectation values can be quickly computed in a separate VMC calculation to high precision. Therefore, adding $A_p - A_T$ from V + QMC to a value of A_T from VMC alone (referred to herafter as "correlated V + QMC") can yield a significant reduction in the statistical error of A_p.

The degree of correlation observed between A_p and A_T, and thus the efficiency of correlated V + QMC, is reduced as the statistical error of the asymptotic populations (and therefore of A_p) increases. The population, P, after N steps is given by the (N-fold) product of the branching factors of G_a. The statistical error of the branching factor, namely σ_b, arises from fluctuations of the local energy, and is related to the statistical error in the population, σ_P, according to

$$\sigma_P / P = \sqrt{N}\, \sigma_b / b. \tag{23}$$

Since $t = N\tau$, the statistical error in P increases as $\sqrt{t}$. Therefore, the usefulness of correlated V + QMC depends on the time required to converge to ϕ_0/Ψ_T and the fluctuations in the local energy, as reflected in σ_b/b.

A final point of interest for the correlated approach concerns the most efficient amount of computation on $(A_p - A_T)$ versus A_T. The efficiency, E_f, for a computation time of T and a resulting variance of V $(= \sigma^2)$, is given by $E_f = (VT)^{-1}$. Therefore, when computing a pure expectation value as the sum of $(A_p - A_T)$ and A_T, one obtains

$$E_f = (T_{p-T} + T_T)^{-1}\, (V_{p-T} + V_T)^{-1}. \tag{24}$$

For $T_{p-T}+T_T=T$, and since $V_{p-T}\propto T_{p-T}^{-1}$ and $V_T\propto T_T^{-1}$, the efficiency may be written as a function of only T_{p-T} or T_T allowing the values of these quantities which maximize E_f to be obtained. The optimum computation times also yield

$$T_{p-T}V_T=T_TV_{p-T} \tag{25}$$

as a useful expression for when the efficiency is greatest. Unfortunately, performing computations which satisfy Eq. (25) requires a priori knowledge of the statistical errors. Nevertheless, initial estimates of the statistical errors in (A_p-A_T) and A_T can serve as guidelines for the amount of computer time best spent on each quantity.

IV. Moments with Odd Powers

In this section we explore more efficient approaches for computing expectation values of odd powered coordinate operators. For such expectation values, contributions from different regions of space may partially or entirely cancel. Therefore, sampling techniques which exploit this cancellation are preferred.

The one-dimensional computation of $\langle x\rangle=\int x\rho(x)\,dx$, serves as an example. As the symmetry of $\rho(x)$ about $x=0$ increases (for simplicity, considering only symmetry about the origin), the degree of cancellation of the integrand also increases. Therefore, lower variance estimates may be obtained by sampling a new distribution, ρ' (e.g., by sampling ρ'/ρ from the original distribution ρ) which exploits this cancellation. The function ρ' need not be a probability density; it is only required that $\int x\rho'(x)\,dx=\langle x\rangle$.

Probably the most straightforward approach is to choose ρ' as the antisymmetric component of ρ, namely $\rho'(x)=\rho_A(x)=\frac{1}{2}[\rho(x)-\rho(-x)]$. Since for an odd function such as x, the symmetric component of ρ vanishes upon integration, one has

$$\langle x\rangle=\int x\rho_A(x)\,dx. \tag{26}$$

In sampling from ρ, $\langle x\rangle$ may now be computed by averaging $x\rho_A(x)/\rho(x)$. Note that if ρ is an even function, all sampled values vanish identically so that the correct result is obtained with no statistical error. More generally, of course, ρ is not fully symmetric, and only a finite statistical error reduction ensues. For example, consider a one-dimensional distribution that, like a heteronuclear molecule, possesses unequal exponentially decaying tails at large $\pm x$. The antisymmetric part of ρ_A then also possesses such tails, though of differing sign. Thus little error reduction is expected in this case. In addition, sampled values of $x\rho_A(x)/\rho(x)$ will be unbounded in regions where $\rho(x)$ is small but $\rho(-x)$ is not, thus increasing statistical error. Therefore, depending on $\rho(x)$, the reduction in statistical error may or may not be significant.

The problems with the naive choice of $\rho' = \rho_A$ (that the spread in ρ' is not optimally reduced and that ρ'/ρ may be poorly behaved) are addressed by a new choice of ρ', given by

$$\tilde{\rho}(x) = \begin{cases} \rho(x) - \rho(-x) & \text{if } \rho(x) \geqslant \rho(-x) \\ 0 & \text{if } \rho(x) < \rho(-x). \end{cases} \tag{27}$$

It is easily shown that $\langle x \rangle = \int x\tilde{\rho}(x)\,dx$. However, the spread in $\tilde{\rho}$ should be less than that of ρ or ρ_A. If tails are present in the original distribution, one of them is immediately eliminated. Also, $\tilde{\rho}$ possesses the zero-variance property (that ρ_A also does) when ρ is an even function. Finally, by construction the quantity to be averaged in sampling from ρ, namely $x\tilde{\rho}/\rho$, will remain bounded for small ρ, since $0 \leqslant \tilde{\rho}/\rho \leqslant 1$.

To take a simple example, consider $\rho(x) = \frac{1}{2}(1+x)$, $|x| \leqslant 1$. From the definitions above, $\rho_A(x) = x/2$ on $|x| \leqslant 1$ and $\tilde{\rho}(x) = x$ on $0 \leqslant x \leqslant 1$. The variance obtained from $\tilde{\rho}$ is

$$V_{\tilde{\rho}} \equiv \int_{-1}^{1} \rho(x)[x\tilde{\rho}(x)/\rho(x)]^2\,dx \; - \; \langle x \rangle^2 \approx 0.11. \tag{28}$$

Efficiency is thus doubled over sampling from ρ directly, for which $V_\rho \approx 0.22$. In contrast, V_{ρ_A} is infinite because of the singularity in ρ_A/ρ at $x = -1$.

A more realistic example is offered in Fig. 3, where ρ is the square of a trial function for LiH at a fixed distance from the internuclear axis (x). Comparing the solid line, ρ, to the dotted, $\tilde{\rho}$, shows the narrowing of the distribution achieved, even

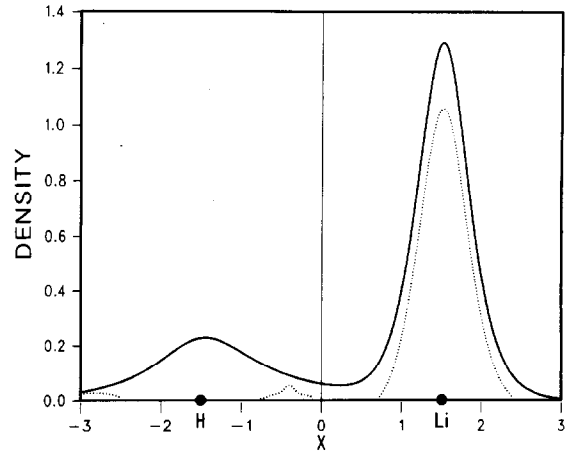

FIG. 3. Equivalent distributions for the computation of $\langle x \rangle$ for LiH. The x coordinate is the internuclear axis. The solid line represents the original distribution, ρ, while the dotted one represents the difference distribution $\tilde{\rho}$, cf. Eq. (27). The average value of x is identical over both distributions, but the variances are not.

though ρ is far from being an even fuction. A numerical computation for the variance yields $V_{\tilde{\rho}}(x) \approx \frac{1}{2} V_{\rho}(x)$. Applying these ideas is most readily accomplished within the V + QMC approach. Specifically, values of ϕ_0/Ψ_T are required for each point **R** sampled from $|\Psi_T|^2$ *and* at symmetry points related to **R** by reflection or inversion. If the symmetry is sufficiently great, the extra computation resulting from sampling at additional points will be compensated by reductions in the statistical error in the averages. The most pronounced reductions will be found for charge distributions which are nearly symmetric. For example, for a molecule such as CO, a substantial reduction in statistical error versus a straightforward sampling of $|\phi_0|^2$ should be observable employing the technique described here.

V. Results and Discussion

For purposes of evaluating the different approaches, we have computed a number of simple moments of the charge distribution of H and H_2. The H atom trial function is chosen as a 1*s* Slater orbital with an exponent detuned to 0.95. For H_2 the trial function is constructed as follows (cf. Table I). A 1*s* Slater orbital is centered on each atom and at the midpoint of the internuclear axis. The linear coefficients are obtained from an SCF calculation using the HONDO program [18]. In addition, a simple Jastrow function of electron–electron and electron–nuclear coordinates is also used, namely,

$$J(r_{12}, r_{i\alpha}) = \exp\left(\frac{ar_{12}}{1+br_{12}} - \sum_{i\alpha} \frac{\lambda r_{i\alpha}}{1+vr_{i\alpha}}\right). \tag{29}$$

In Eq. (29), Roman indices denote electrons, while Greek indices denote nuclei and the bond function. The final form of the trial function is

$$\Psi_T(1, 2) = \psi(1)\,\psi(2)\,J(1, 2), \tag{30}$$

where ψ is a molecular orbital.

As we indicated in the Introduction, there are a variety of GFMC implementations for which our algorithms described here apply. For the computations we

TABLE I

Parameters of the H_2 Trial Wave Function

c(H)	0.48610	*a*	0.50
c(BF[a])	0.11089	*b*	0.50
Exponent	1.19000	λ	0.15
		v	1.00

[a] The bond function is located at the midpoint of the internuclear axis.

report below we have chosen the diffusion QMC approach described in detail in Ref. [9]. The only difference from that description is that the trial energy in the present algorithm is not updated, but rather it is kept fixed at the initial value.

Trial, mixed, and pure expectation values obtained for H and H_2 are presented in Table II. Where analytic values of A_T or A_m are not known, VMC and QMC results have been computed, respectively. The values in Table II are thus indicative of the accuracy of the various starting points for our pure approaches. As can be seen, the second-order estimator, while far more accurate than trial or mixed averages, is still noticably biased.

Pure expectation values of $\langle r \rangle$, $\langle z^2 \rangle$, and $\langle r^2 \rangle$ for H and of $\langle z^2 \rangle$ and $\langle r^2 \rangle$ for H_2, computed by the methods discussed in Sections II and III, are presented in Tables III and IV. Since the efficiency of the techniques is our primary concern, detailed studies of time step bias and the effects of finite converge time, t, are not undertaken here. Despite this, accuracies are generally better than 0.5%. However, it is precisely these biases that account for the small difference between the pure QMC and exact results. In Table III we address these biases by extrapolation of both the convergence time and the time step size for one of the four approaches (correlated V + QMC). The final extrapolated results show no visible bias. In addition, error bars are smaller than the bias of the second-order estimators given in Table II, making that bias visible. This is important, for otherwise the pure estimator would have no real advantage over the second-order one. The ability to

TABLE II

Comparison of Trial, Mixed, and Pure Expectation Values for H and H_2[a]

	H		
	$\langle r \rangle$	$\langle z^2 \rangle$	$\langle r^2 \rangle$
Trial	1.5789	1.1080	3.3241
Mixed	1.5385	1.0519	3.1558
Second-order[b]	1.4981	0.9958	2.9875
Pure	1.5000	1.0000	3.0000
	H_2		
	$\langle z^2 \rangle$		$\langle r^2 \rangle$
Trial (VMC)	1.0787(6)		2.6228(11)
Mixed (QMC)[c]	1.0491(8)		2.5809(14)
Second-order[b]	1.0195(10)		2.5390(18)
Pure (exact)[d]	1.0230		2.5464

[a] Results for H are the exact, analytic values.
[b] See Eq. (5).
[c] Time step for QMC results is $0.01h^{-1}$.
[d] Ref. [19].

reduce these error bars sufficiently comes in part from the use of good quality trial wavefunctions and, as discussed more below, by correlated sampling and the partial carrying of weights. Thus pure expectation values need no longer be abandoned in favor of the less noisy but biased second-order estimators.

In the single-walk calculations of H (cf. Table III) even the worst (t, τ) combination yields second moments accurate to 1% or better. These small errors at convergence time $t = 5h^{-1}$ (relative to the 5% error in the mixed averages, i.e., $t = 0$) imply that values of P are close to their asymptotic limit, ϕ/Ψ_T. Doubling the convergence time to $10h^{-1}$ further increases the accuracy of the second moments to within $< 0.5\%$. Though this improvement indicates that further convergence is possible, the smallness of the gain (relative to the total change from the mixed average) verifies that values of P are indeed well converged. However, in the V + QMC approaches at $t = 5h^{-1}$, the errors are noticeably larger than those resulting from the single QMC walk algorithm. This is most likely due to the

TABLE III

Pure Expectation Values for H by Various Monte Carlo Methods[a]

Method	(t, τ)[b]	$\langle r \rangle$	$\langle z^2 \rangle$	$\langle r^2 \rangle$
Single QMC walk:	(5, 0.050)	1.5058(14)	1.0075(32)	3.025(6)
branching without weights	(10, 0.050)	1.5038(15)	1.0066(31)	3.015(7)
	(10, 0.025)	1.5025(21)	1.0029(41)	3.010(9)
Single QMC walk:	(5, 0.050)	1.5052(11)	1.0084(25)	3.024(5)
branching with weights	(10, 0.050)	1.5040(12)	1.0024(25)	3.018(6)
	(10, 0.025)	1.5023(15)	1.0040(33)	3.009(7)
V + QMC	(5, 0.050)	1.5074(15)	1.0150(40)	3.033(6)
	(10, 0.050)	1.5032(22)	1.0018(51)	3.012(9)
	(10, 0.025)	1.5026(23)	1.0008(53)	3.012(11)
Correlated	(5, 0.050)	1.5094(5)	1.0144(11)	3.045(2)
V + QMC[c]	(10, 0.050)	1.5024(7)	1.0029(21)	3.010(4)
	(∞, 0.050)	1.5017(7)	1.0015(21)	3.006(4)
	(10, 0.025)	1.5020(9)	1.0025(24)	3.009(5)
	(∞, 0.025)	1.5013(9)	1.0014(24)	3.005(5)
	(∞, 0)	1.5009(9)	1.0013(24)	3.004(5)
Exact[d]		1.5000	1.0000	3.000

[a] Units of length and time are Bohr and h^{-1}, respectively. Statistical errors, in paretheses, represent one standard deviation in the mean of a Gaussian distribution and are normalized to correspond to 10 min of computation time on a Cray X-MP.

[b] t denotes propagation time from Ψ_T; τ is the time step.

[c] For extrapolation to $t \to \infty$ five values of $t(= 2, 4, 6, 8, 10)$ were fit to an exponential $a + be^{-\alpha t}$ representing the effect of the decay of the higher energy eigenstates. Time step extrapolation is simply linear. Error bars on extrapolated values reflect those of the last point in the sequence. Combining all errors can lead to uncertainties slightly different from those quoted.

[d] See Table II.

requirement of sampling (ϕ_0/Ψ_T) twice, compounding the error of incompletely converged values of P. This effect appears to be no longer significant by $t = 10h^{-1}$, where the accuracies of all methods are statistically equivalent. Moreover, as shown in the table, one may also extrapolate to $t \to \infty$ with a few more data points. Time step bias must be independently considered. That little time step bias is present may be seen when reducing τ from 0.050 to $0.025h^{-1}$: only a small improvement is obtained (generally less than the statistical error) in the already accurate pure expectation values.

All H_2 calculations were performed with $\tau = 0.01h^{-1}$. This time step introduces very little error as demonstrated by the accuracy of the results of Table IV. Since the statistical errors are often larger than the differences between means corresponding to different convergence times, trends are difficult to discern. Generally, it is found that errors in the means are $\leqslant 1\%$ at $t = 2h^{-1}$ and less than statistical error ($< 0.5\%$) at $t = 4$ and $6h^{-1}$.

The results of Tables III and IV show that each algorithm readily produces accurate expectation values. For a comparison of the efficiency of the various algorithms, all statistical errors in Tables III and IV correspond to the same amount of computation (10 minutes on a single processor of a Cray X-MP) and therefore provide a direct measure of relative efficiency.

It is immediately apparent that small and consistent improvement in precision results for both H and H_2 when weights are carried in the branching QMC walk.

TABLE IV

Pure Expectation Values for H_2[a]

Method	t	$\langle z^2 \rangle$	$\langle r^2 \rangle$
Single QMC walk: branching without weights	2	1.028(9)	2.554(14)
	4	1.024(8)	2.549(14)
	6	1.021(9)	2.549(15)
Single QMC walk: branching with weights	2	1.034(6)	2.554(10)
	4	1.020(5)	2.535(9)
	6	1.028(6)	2.545(12)
V + QMC	2	1.034(5)	2.563(9)
	4	1.022(8)	2.548(12)
	6	1.027(8)	2.550(21)
Correlated V + QMC	2	1.033(2)	2.562(4)
	4	1.026(3)	2.549(6)
	6	1.026(6)	2.550(10)
Exact[b]		1.023	2.546

[a] Units are given in Table III. The time step is $\tau = 0.01h^{-1}$ for all results presented; the statistical errors correspond to 10 min of computation on a Cray X-MP.

[b] Ref. [19]. Also see Table II.

The average increase in efficiency (the squared ratios of statistical errors averaged across τ and t) is roughly 60% for H and 130% for H_2. This improvement does not appear to be strongly dependent on the choices for the upper and lower bounds of the weights. In the single-walk calculations these weights were not allowed to exceed 2. The lower bound was chosen as either 0.1 or 0.4, and no noticeable change in the efficiency was found depending on this choice.

Comparisons of efficiency between the V + QMC and correlated V + QMC approaches are also given by Tables III and IV. For H, the average increase in efficiency with correlated V + QMC over V + QMC is a factor of 10 at $t = 5h^{-1}$ and a factor of 6 at $t = 10h^{-1}$. For H_2, the increase in efficiency is a factor of 6 at $t = 2h^{-1}$, but decreases to 3 at $t = 6h^{-1}$. As expected, reductions in statistical error obtained by exploiting the correlation between trial and pure expectation values decrease with increasing convergence time. Nonetheless, gains in efficiency remain substantial even at the larger convergence times.

While the most efficient techniques within the single- and double-walk algorithms are discernible, the question of which *class* is better is not so immediately answerable. For H, the smallest statistical errors are obtained by the correlated V + QMC approach. For H_2, however, the superiority of the correlated V + QMC technique is lost by $t = 6h^{-1}$. Since convergence to accurate results is obtained for H_2 by $t = 4h^{-1}$, the correlated V + QMC technique remains the best approach for H_2 as well as H. However, as seen above, the efficiency of the correlated V + QMC method versus other QMC approaches is strongly dependent on the length of the QMC walk required for convergence of the populations to ϕ_0/Ψ_T and on how quickly the statistical error in $(A_p - A_T)$ increases with t. Hydrogenic systems in general have optimal convergence properties and allow for the longest chains of correlated sampling. Larger Z atoms and systems with more complex wavefunctions may likely take too long to converge to be competitive. Therefore, the single-walk method (with weighting) may ultimately be the more competitive for these other systems. This should, however, be investigated further in future studies.

APPENDIX: Variance of Rounded versus Unrounded Weights

Here we compare the variance obtained for the weight when it is integer rounded (keeping the correct expectation value) versus when its full real value is kept. The difference is significant when it is necessary to sample asymptotic populations.

Consider a weight w and its associated probability density function, $f(w)$. The mean and variance of w are given by

$$\bar{w} = \int w f(w)\, dw$$

and

$$V_w = \int w^2 f(w)\, dw - \bar{w}^2. \tag{A1}$$

Let $I_w(\xi) \equiv \text{int}(w+\xi)$ be the integer-rounded weight. Here ξ is a uniform random variate between 0 and 1. The mean value of $I_w(\xi)$ over the uniform distribution of ξ's is given as

$$\bar{I}_w = \int_0^1 I_w(\xi)\, d\xi = \int_0^1 \text{int}(w+\xi)\, d\xi = w. \tag{A2}$$

In contrast to Eq. (A1), the variance of the rounded weight is

$$V_{I_w} = \int \left[\int_0^1 I_w^2(\xi)\, d\xi \right] f(w)\, dw - \left[\int \bar{I}_w f(w)\, dw \right]^2. \tag{A3}$$

Defining the remainder $r(w)$ by $w = \text{int}(w) + r(w)$, one obtains for the integral over ξ,

$$\begin{aligned} \int_0^1 I_w^2(\xi)\, d\xi &= \int_0^{1-r} [\text{int}(w+\xi)]^2\, d\xi + \int_{1-r}^1 [\text{int}(w+\xi)]^2\, d\xi, \\ &= (1-r)[\text{int}(w)]^2 + r[\text{int}(w)+1]^2 \\ &= w^2 + r - r^2. \end{aligned} \tag{A4}$$

Eq. (A3) now becomes

$$V_{I_w} = \int [w^2 + r(w) - r^2(w)]\, f(w)\, dw - \bar{w}^2 \tag{A5}$$

or

$$V_{I_w} = V_w + \int [r(w) - r^2(w)]\, f(w)\, dw. \tag{A6}$$

Since $r(w) - r(w)^2 \geqslant 0$, and $f(w)$ is by definition positive definite, $V_{I_w} \geqslant V_w$. This quantifies the lower variance in the weights when integer rounding is avoided. Though the additional variance due to a single integer-rounded weight is small, the reduction in the variance of the asymptotic populations (which are essentially the products of many weights) can be significant.

A somewhat different but related argument for carrying the weights for a time is given in Ref. [17]. We thank one of the referees for bringing this paper to our attention. The very nice analysis given there of the population dynamics of independent "families" of random walkers will likely be useful for determining the optimal time the weights should be accumulated before branching is done.

REFERENCES

1. An excellent overview of developments in quantum Monte Carlo is provided by the thematic issue, *J. Statist. Phys.* **43** (1986); see also, W. A. LESTER, JR. AND B. L. HAMMOND, *Annu. Rev. Phys. Chem.* **41**, 283 (1990).
2. D. M. CEPERLEY AND M. H. KALOS, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer-Verlag, New York, 1979); J. W. MOSKOWITZ AND K. E. SCHMIDT, in *Monte*

Carlo Methods in Quantum Problems, edited by M. H. Kalos (Reidel, New York, 1984); K. E. SCHMIDT AND M. H. KALOS, in *Applications of the Monte Carlo Method in Statistical Physics*, edited by K. Binder (Springer-Verlag, New York, 1985); D. M. CEPERLEY AND B. J. ALDER, *Science* **231** 555 (1986).

3. P. J. REYNOLDS, M. DUPUIS, AND W. A. LESTER, JR., *J. Chem. Phys.* **82**, 1983 (1985); R. N. BARNETT, P. J. REYNOLDS, AND W. A. LESTER, JR., *J. Chem. Phys.* **82**, 2700 (1985); **84**, 4992 (1986); D. R. GARMER AND J. B. ANDERSON, *J. Chem. Phys.* **86**, 7237 (1987).
4. P. A. WHITLOCK, D. M. CEPERLEY, G. V. CHESTER, AND M. H. KALOS, *Phys. Rev. B* **19**, 5598 (1979).
5. M. H. KALOS, *J. Comput. Phys.* **1**, 257 (1966).
6. M. H. KALOS, *Phys. Rev. A* **2**, 250 (1970).
7. K. S. LIU, M. H. KALOS, AND G. V. CHESTER, *Phys. Rev. A* **10**, 303 (1974).
8. D. M. CEPERLEY, "The Stochastic Solution of the Many-Body Schrödinger Equation for Fermions," in *Recent Progress in Many-Body Theories*, edited by J. G. Zabolitzky, M. de Llano, M. Forbes, and J. W. Clark (Springer, Berlin, 1981), p. 262.
9. P. J. REYNOLDS, D. M. CEPERLEY, B. J. ALDER, AND W. A. LESTER, JR., *J. Chem. Phys.* **77**, 5593 (1982).
10. P. J. REYNOLDS, R. N. BARNETT, B. L. HAMMOND, AND W. A. LESTER, JR., *J. Statist. Phys.* **43**, 1017 (1986) (The tagging algorithm is described briefly in this paper).
11. (a) A. L. L. EAST, S. M. ROTHSTEIN, AND J. VRBIK, *J. Chem. Phys.* **89**, 4880 (1988); (b) P. J. REYNOLDS, *J. Chem. Phys.* **92**, 2118 (1990) (This comment points out a deficiency in Ref. 11(a)).
12. D. M. CEPERLEY AND B. J. ALDER, *Phys. Rev. Lett.* **45**, 566 (1980).
13. J. W. MOSKOWITZ, K. E., SCHMIDT, M. A. LEE, AND M. H. KALOS, *J. Chem. Phys.* **77**, 349 (1982).
14. D. M. CEPERLEY, *J. Comput. Phys.* **51**, 404 (1983).
15. D. M. CEPERLEY AND B. J. ALDER, *J. Chem. Phys.* **81**, 5833 (1984).
16. J. B. ANDERSON, *J. Chem. Phys.* **63**, 1499 (1975); **65**, 4121 (1976).
17. K. J. RUNGE AND R. J. RUNGE, "Forward Walking GFMC Calculation of the Staggered Order in the Heisenberg Antiferromagnet," in *Quantum Simulations of Condensed Matter Phenomena*, edited by J. D. Doll and J. E. Gubernatis (World Scientific, Singapore, 1990), p. 300.
18. *NRCC Software Catalog* **1**, 60 (1980); LBL-10811, UC-4 (available from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, Virginia 22161; price code: A05).
19. W. KOLOS AND L. WOLNIEWICZ, *J. Chem. Phys.* **43**, 2429 (1965).