Spatial-partitioning-based acceleration for variational Monte Carlo

Dario Bressanini[a]
*Dipartimento di Scienze Chimiche, Fisiche, e Matematiche, Università dell'Insubria, polo di Como,
Via Lucini 3, I-22100 Como, Italy*

Peter J. Reynolds[a],[b]
Physical Sciences Division, Office of Naval Research, Arlington, Virginia 22217

The problem of inherently differing time scales of core and valence electrons in Monte Carlo (MC) simulations is circumvented in a straightforward and intuitive manner. By appropriately subdividing into equivalent subspaces the high-dimensional (many-electron) space in which Monte Carlo integration is done, it is possible to choose completely independent and appropriate sampling times for each ''electron.'' This approach trivially satisfies detailed balance. The partitioning of space is applicable to both variational and Green's function MC. Such a partitioning, however, only provides a significant computational advantage in variational MC. Using this approach we were able to have inner electrons move with reasonably large steps and yet avoid excessive rejection, while outer electrons were moved great distances in few steps. The net result is a large decrease in the sampling autocorrelation time, and a corresponding increase in convergence rate. Results of several standard algorithms are compared with the present acceleration algorithm for the atoms Be and Ne, and the molecule Li_2. © *1999 American Institute of Physics.* [S0021-9606(99)50838-6]

INTRODUCTION

Variational Monte Carlo (VMC) methods allow one to calculate quantum expectation values given a trial wave function.[1] Wave functions of great functional complexity are amenable to this treatment, since analytical integration is not being performed. This greater complexity, including for example, explicit two-body and higher-order correlation terms, in turn allows for a far more compact description of a many-body system, with the benefit of high accuracy. The primary disadvantage of using a Monte Carlo approach is that the calculated quantities contain a statistical uncertainty, which needs to be made small. This can always be done, but at the cost of CPU time, since the statistical uncertainty decreases as $N^{-1/2}$. The term ''variational'' Monte Carlo derives from the use of this type of Monte Carlo sampling to optimize the trial wave function via the variational principle. Despite the inherent statistical uncertainty, a number of very good algorithms have been created that allow one to optimize trial functions.[2–7] The best of these approaches go beyond simply minimizing the energy, and exploit the minimization of the energy variance, which vanishes for energy eigenfunctions.

All total energy methods, whether Monte Carlo or not, suffer from scaling problems.[8–10] That is, as the system being treated increases in size, the computational cost rises as an (often large) power of the system size. Although such behavior is far preferable to that of the exponentially difficult problems in the classes NP and beyond, large-power polynomial scaling is nevertheless a severe roadblock to the treatment of many physically interesting systems. Even significantly faster computers will leave large classes of interesting

problems untouchable. This is the motivation behind the so-called order-N methods now becoming well-known in density functional theory, where N is the number of electrons in the system. While density functional theory provides an approach to electron correlation and is useful in many contexts, often an exact treatment of such correlations, or at least a systematically improvable one, is necessary. Quantum chemical approaches are of the latter variety. Unfortunately, they are among the class of methods that scale with large powers of system size. On the other hand, Monte Carlo methods exist which are either systematically improvable or exact, and these methods scale reasonably well with system size. Generally these methods scale roughly between N^2 and N^3; moreover algorithms with lower powers are in principle possible to implement (e.g., using fast multipole methods to evaluate the Coulomb potential, and the use of localized orbitals together with sparse matrix techniques for the wave function computation).

There is still a problem. This remaining problem is well-known from other contexts. It is often referred to as the multiple time scales problem.[11,12] Possibly the most extreme instance of it occurs in condensed matter physics near a phase transition, where the problem is known as *critical slowing down*. In the VMC [and more generally quantum Monte Carlo (QMC)] context it has come to be known as the large-Z problem. This class of problem occurs in both Monte Carlo and molecular dynamics simulations, and more generally whenever dynamical calculations are performed.

Although, as mentioned above, the various quantum Monte Carlo algorithms scale well with N, they have been shown to scale much more poorly with atomic number Z. A common estimate is that computational time T rises between $Z^{5.5}$ and $Z^{6.5}$. Upon reflection it is clear that the problem is the differing time (as well as distance and energy) scales for

[a]Also at: Department of Physics, Georgetown University, Washington, D.C.
[b]Author to whom correspondence should be addressed.

core and valence electrons. As Z increases, the range of time scales increases as well. In fact, $Z \rightarrow \infty$ is in many ways analogous to a critical point.[10,13,14] As in critical slowing down, an unending hierarchy of time scales ensues as the critical point is approached. This is the problem that must be addressed.

In critical phenomena, the problem has been effectively addressed through a class of acceleration methods, particularly so-called cluster acceleration methods.[15,16] These take advantage of the self-similarity that occurs in the vicinity of a critical point. In the electronic structure problem, there do exist analogous critical points.[17,18] However, ground states of typical systems are not near the regime of these critical points. Thus, a common way to address the large-Z problem has been through the use of effective-core potentials which eliminate the large Z at the outset. This is the standard approach in quantum chemistry and solid-state physics. It is also becoming widely (and effectively) used in quantum Monte Carlo simulations.[9,19–22] However, in Monte Carlo there are many other possible ways to address the problem. The method we discuss here can be used in VMC to avoid entirely the pseudopotential approximation, or can be used in conjunction with it to provide additional computational advantage.

OVERVIEW OF VMC

Since very detailed descriptions of the VMC method are available,[1,23] we only give here a short resume. The essence of VMC is the sampling of a distribution proportional to $|\Psi_T(\mathbf{R})|^2$, where Ψ_T is a given ("trial") wave function—a function of the $3N$-dimensional coordinate $\mathbf{R}$. Once such a distribution is established, expectation values of nondifferential operators may simply be sampled, since

$$\langle \hat{O} \rangle = \int \hat{O}(\mathbf{R}) |\Psi_T(\mathbf{R})|^2 d^{3N}\mathbf{R} \Big/ \int |\Psi_T(\mathbf{R})|^2 d^{3N}\mathbf{R}$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} \hat{O}(\mathbf{R}_i). \tag{1}$$

Differential operators are only slightly more difficult, since we can write

$$\langle \hat{O} \rangle = \frac{\int [\hat{O}\Psi_T(\mathbf{R})/\Psi_T(\mathbf{R})]|\Psi_T(\mathbf{R})|^2 d^{3N}\mathbf{R}}{\int |\Psi_T(\mathbf{R})|^2 d^{3N}\mathbf{R}}$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} [\hat{O}\Psi_T(\mathbf{R}_i)/\Psi_T(\mathbf{R}_i)]. \tag{2}$$

The remaining problem is how to sample the distribution $|\Psi_T(\mathbf{R})|^2$. This is readily done in a number of ways. The most straightforward is simple Metropolis sampling.[24] Specifically, this involves generating a Markov chain of steps by "box sampling" $\mathbf{R}' = \mathbf{R} + \zeta\Delta$, with Δ the box size, and ζ a $3N$-dimensional vector of uniformly distributed random numbers $\zeta \in [-1, +1]$. This is followed by the classic Metropolis accept/reject step, in which $|\Psi_T(\mathbf{R}')/\Psi_T(\mathbf{R})|^2$ is compared to a uniformly distributed random number between zero and unity. The new coordinate $\mathbf{R}'$ is accepted only if the ratio of trial functions squared exceeds the ran-

dom number. Otherwise the new step remains at $\mathbf{R}$. This completes one step of the Markov chain (or random walk). Under very general conditions,[25] such a Markov chain results in an asymptotic equilibrium distribution proportional to $|\Psi_T(\mathbf{R})|^2$.

From the above description of the standard Metropolis VMC simulation algorithm, it is clear that the attempted move of an electron covers a volume which is independent of its position. This means that the optimal move size is a trade-off between the best move size for electrons far from the nucleus (i.e., *valence* or *outer* electrons), which need to be large since the accessible region of configuration space is very large, and the best move size for the electrons close to the nucleus (i.e., *core* or *inner* electrons). These latter moves must be small, since the relevant region of configuration space is quite limited, and also because the wave function changes rapidly near the nucleus, meaning that large moves would cause a high rejection rate.

This situation is only mildly improved when one switches to the commonly used, and otherwise more efficient, Langevin simulation scheme.[26] This scheme is a generalization of the standard Metropolis algorithm in which a Langevin equation containing drift and diffusion (i.e., a "quantum" force term and white noise) is employed for the transition matrix from $\mathbf{R}$ to $\mathbf{R}'$. Although the quantum force depends on position, the size of an attempted move is still determined by the step size (now the *time*-step size). Using a single time step for all the electrons still implies a certain degree of negotiation between inner and outer electrons in choosing the best global time step. The inner electrons still end up dominating the dynamics, and slowing down the outer electrons.

There are many ways one can think of improving the simple algorithm. Several methods have been explored with differing degrees of success. For example, one can render the attempted moves position-dependent, which subsequently entails the need for a modified coordinate system to maintain detailed balance.[27] Another approach, borrowed from high-energy theory, has been to modify the VMC dynamics while keeping the steady-state unchanged.[10,12,28] One can also radically change the algorithm by mixing a molecular dynamics approach with VMC[29] or using a feedback method.[30] Here we explore an intuitive and straightforward new approach.

THEORETICAL APPROACH

Partitioning the space

As a result of the antisymmetry of an electronic wave function, there are multiple regions of ($3N$-dimensional) space which are equivalent. Specifically, up to a sign, the value of the wave function is the same when any two coordinates representing like-spin electrons are interchanged. This results in $N_{up}! N_{down}!$ equivalent volumes or domains. Since it is hard visualizing in high-dimensional spaces, it is worth pointing out here that we are *not* talking about *nodal volumes*, that is, regions of the $3N$-dimensional space surrounded by a hypersurface where the wave function is zero. Though these volumes may be equivalent (e.g., in 1D), generally the nodal volumes are connected sets of the volumes
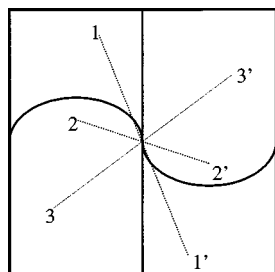
FIG. 1. Different partitioning schemes, illustrated with the inversion operator in a 2D box.

we are discussing here. In fact, there are suggestions that the number of distinct nodal volumes for all ground state atomic or molecular systems is just two.[31,32] The volumes we are distinguishing are the following: given a point $\mathbf{R}$ in configuration space, there are another $N_{up}!N_{down}! - 1$ points generated by permutations of the indices. We can think of these points as belonging to different regions, or subspaces, of the full space. If we can explicitly construct such subspaces, then the integration over the entire $3N$-dimensional space is redundant, since for any operator $\hat{O}$ which is totally symmetric with respect to the exchange of two identical particles

$$\langle \hat{O} \rangle = \frac{\int_{\text{all space}} \Psi_T^*(\mathbf{R}) \hat{O} \Psi_T(\mathbf{R}) d^{3N}\mathbf{R}}{\int_{\text{all space}} \Psi_T^*(\mathbf{R}) \Psi_T(\mathbf{R}) d^{3N}\mathbf{R}}$$

$$= \frac{\int_{\text{any subspace}} \Psi_T^*(\mathbf{R}) \hat{O} \Psi_T(\mathbf{R}) d^{3N}\mathbf{R}}{\int_{\text{any subspace}} \Psi_T^*(\mathbf{R}) \Psi_T(\mathbf{R}) d^{3N}\mathbf{R}}, \qquad (3)$$

meaning we only need to integrate over a single subspace.

Such subspaces are not uniquely defined. The following example can help in visualizing this fact. Consider a two-dimensional box centered at the origin as the analogue of our $3N$-dimensional space, and the inversion operator $\hat{i}$ (the operator which takes $\mathbf{R}$ to $-\mathbf{R}$) as analogous to the permutation operator. Assume that in this space $\Psi(\hat{i}\mathbf{R}) = -\Psi(\mathbf{R})$ and $V(\hat{i}\mathbf{R}) = V(\mathbf{R})$. This system can be thought of as a particle in a two-dimensional box with an additional symmetry constraint. For each point $\mathbf{R}$ there is a ''corresponding point'' $\mathbf{R}' = \hat{i}\mathbf{R}$. Two (of an infinite number) of equally valid subspace constructions are shown in Fig. 1. Each divides the space into two regions such that the points $\mathbf{R}$ and $\mathbf{R}' = \hat{i}\mathbf{R}$ are in opposite regions. In the example illustrated, the points 1, 2, and 3 all belong to the same subspace if we divide the volume with the vertical line. The other curve shown still divides all the points from their counterparts, however now point 1' rather than 1 sits in the same volume as points 2 and 3. Because of the inversion symmetry, integration over any one of these subspaces is equivalent to integration over the whole configuration space! Yet none of these curves needs to be a nodal line for the system. Only an even lower-dimensional surface (here the origin) is guaranteed to be on

the nodal surface. Now we will see how this simple fact can be used to our advantage to help avoid the time-scale problem. Let us concentrate on an atomic system. Later in this paper we will discuss the modifications needed to treat molecules.

Separating time scales

In none of the earlier attempts to accelerate VMC (as described in the overview section above) was an attempt made to make the electrons ''distinguishable'' within the simulation. Based on the idea of dividing space into equivalent subspaces, this is now possible. Here we explore such an approach.

Naïvely trying to assign a different time step (and so a different time scale) to different electrons does not work, of course. Given a symmetric or antisymmetric wave function, two identical particles (here like-spin electrons) can exchange positions without changing the probability of the configuration. Thus, assigning larger time steps to electrons starting out in the valence region at the beginning of the simulation would not accomplish our goal, since ultimately such electrons exchange positions with inner electrons, with no energy penalty. Once this happens the electrons are taking inappropriate step sizes. In terms of our previous discussion of subspaces, we can restate this fact saying that in the standard algorithm, for any subspace division, electrons can generally cross the subspace boundaries. We can, however, enforce the boundaries and constrain particles to stay in certain subspaces.

Thus far, however, nothing suggests that constraining moves to subspaces would be better for the efficiency of the algorithm than simply integrating over the whole configuration space. Nevertheless, since the subspace division is to a large extent arbitrary, there is hope that a good choice can in fact help. Specifically, we seek to construct a subspace such that the electrons in the outer regions of 3-space, away from the nucleus, and likewise the electrons close-in, near the nucleus, each stay in their relative places through the action of the constraint. This would enable us to assign different time steps to the different electrons, and let them explore their respective regions of configuration space with the most appropriate step sizes. This is actually quite readily done, as the example below illustrates. Moreover, this generalizes immediately to a practical scheme for constructing subspaces for any atom, and with little modification, for molecules.

An example: The Be atom

Let us take as a practical example the Be atom in its ground state. After having (arbitrarily) assigned spin up to electrons 1 and 2, and spin down to electrons 3 and 4, we are left with a configuration space in which it is possible to define four equivalent subspaces. Given a point in one subspace, we can generate the symmetry-related points by permuting electrons 1 and 2 and/or 3 and 4.

Our chemical intuition tells us that, on the average, electrons 1 and 2 are not likely to be both close to the nucleus simultaneously. This is because they have the same spin, and we expect the beryllium ''core'' to be composed of electrons

of opposite spin. Of course, the same reasoning applies to the other electron pair. We can now assign, say, electrons 1 and 3 to the core, and electrons 2 and 4 to the valence. Having done so, we can define our subspace as the set of points $\mathbf{R}$ in configuration space for which the first electron is always closer to the nucleus than the second, and the third closer than the fourth. Specifically, we can define the space

$$\Omega_1(\mathbf{R}) = \{\mathbf{R}: \ r_1 < r_2 \ \text{AND} \ r_3 < r_4\}$$
$$r_i = \sqrt{x_i^2 + y_i^2 + z_i^2}.$$
$$(4)$$

In a similar way, one can define the other three equivalent subspaces,

$$\Omega_2 = \{\mathbf{R}: \ r_1 < r_2 \ \text{AND} \ r_3 > r_4\}$$
$$\Omega_3 = \{\mathbf{R}: \ r_1 > r_2 \ \text{AND} \ r_3 < r_4\} \qquad (5)$$
$$\Omega_4 = \{\mathbf{R}: \ r_1 > r_2 \ \text{AND} \ r_3 > r_4\}.$$

An equivalent integration can be performed over any of these subspaces. The simulation in e.g., the domain Ω_1 can be done rejecting any configuration where $(r_1 > r_2)$ or $(r_3 > r_4)$. Since electrons 2 and 4 are outer electrons (and in the above sense will always remain so during the simulation), we can assign to them a larger time step than the one we give to the inner electrons. Separately optimizing these time steps increases the efficiency of the overall simulation.

Any starting configuration, i.e., a walker in VMC, is a single point in the $3N$-dimensional space, and thus resides in a single subspace. Subsequent moves need only enforce the boundaries by rejecting any attempts to cross them. It is easy to see that detailed balance is trivially satisfied. One way to see this is to regard the present algorithm as the standard VMC algorithm, without any constraint (thereby satisfying detailed balance), applied to a wave function that vanishes outside the boundaries. Since all the subspaces are equivalent, Eq. (3) tells us that all expectation values for this restricted wave function are the same as those computed over the full space.

One point is worth noting here: we have designed a partition of configuration space through a set of constraints defined in three-dimensional space rather than $3N$-dimensional space. This is a desirable feature of any partitioning scheme, since this greatly simplifies the practical implementation of the algorithm, and also leads to a nice physical interpretation. However, since other division schemes of the $3N$-dimensional space are possible, we must take particular care when devising the constraints in three-dimensional space, in particular to ensure that no configurations are left out. Consider for example the following division scheme for electron 1 and 2, which superficially looks similar to the previous one. We can constrain the two electrons, e.g., to be on different sides of a given fixed plane passing through the nucleus, say the plane xy. Apart from the fact that such a partition would be useless for our purposes, it also is wrong, since it leaves out all the configurations where the two electrons are on the same side. A correct, though still useless, scheme is to constrain the second electron to be always to the "left" of the first, with respect to a given plane.

Heavier atoms

As the number of electrons increases, so does the number of equivalent subspaces, and with that our freedom in choosing them. In particular, we can combine subspaces to increase efficiency. Again, we prefer to give a concrete example for pedagogical purposes: let us consider the neon atom. With five spin-up electrons and five spin-down electrons the number of equivalent subspaces is $5!^2 = 14\,400$. We can choose any one as our integration space, imposing, e.g., the constraint $r_1 < r_2 < r_3 < r_4 < r_5$ AND $r_6 < r_7 < r_8 < r_9 < r_{10}$, in analogy to what was done for the Be atom. However, using again our chemical intuition, we expect the neon core to be composed of two electrons, and the "valence" space of the other eight. Since the outer electrons share the same three-dimensional region of space, we expect very little gain in imposing the above overly restrictive boundary conditions. Instead, a more physically-sound partition would be always to keep electron 1 closer to the nucleus than all the other spin-up electrons, while electron 6 is kept closer to the nucleus than all the other spin-down electrons. The net effect is that we have merged some of the smaller equivalent subspaces to build a bigger subspace.

Loosely speaking, we might expect that a good partition is one in which we prevent electrons from changing "shells," while we leave free the electrons within a shell to explore all the "shell" space. This should be more efficient as well, because we avoid unnecessary rejections which would be caused by crossings among electrons with the same time scale. Thus, going to still larger Z, for the argon atom we would divide the spin-up electrons (and similarly the spin-down ones) into three groups, and would impose constraints such that electrons in any group never exchange their "role."

Molecules

All the theoretical considerations regarding the subdivision of configuration space into equivalent subspaces, which we gave for atoms, are still valid for the molecular case. This is so because these considerations were based only on the Pauli principle and not on any particular potential. What needs to be modified, of course, is the prescription on how to divide the configuration space in an efficient way. In particular, if we want to keep the useful picture of electrons in shells we need to choose, e.g., an origin from which to measure the distance to the electrons. We could measure all the distances from the heaviest nucleus of the molecule, and implement the constraints described above. This should be sufficient for a molecule with only one heavy atom and other very light atoms. In general, however, we have different nuclei with various atomic numbers and different cores belonging to the various nuclei. A more physically motivated approach is thus to assign the various core electrons to the corresponding nuclei, and to treat the remaining electrons as belonging to the "valence" space for the entire molecule. The core electrons can, as before, be assigned to multiple shells.

Let us once again construct a practical example: we will divide the configuration space of the Li_2 molecule. We can

imagine this molecule as having two electrons always close to the first nucleus, two electrons likewise close to the second nucleus, and lastly, two electrons free to explore the region outside the two cores. Considering the spin-up electrons, our subspace is then defined as having electron 1 closer to nucleus A than either of the other two; electron 2 is then the remaining one which is closest to nucleus B; and electron 3 is the one farthest away from the two nuclei. This partition is valid as long as the two centers are equivalent (as in the case of Li_2). However, as in the case of the atomic shells, we are unnecessarily restricting the core electrons (1 and 2), as these have the same time step behavior anyway. Thus, it is sufficient on physical grounds to merely prevent the penetration of the third (outer) electron into either of the cores.

RESULTS AND DISCUSSION

We have tested the proposed algorithm on the Be and Ne atoms and on the Li_2 molecule. Since the main purpose of this algorithm is to separate the different time scales of the electrons, it is natural to compare the average movement of the electrons in different regions with respect to those in a standard algorithm, one which has the same time scale for all the electrons. Such a microdiagnostic approach has been suggested recently.[33]

Usually, when making efficiency comparisons, one compares against the ''box Metropolis'' algorithm, in which a uniform move within a box is accepted or rejected only after *all* the electrons have been moved to a new location. The microdiagnostic analysis can be very useful in monitoring such a simulation, to check that all the electrons move reasonable distances, and to ensure that the run time of the simulation is sufficient to allow a meaningful sampling of configuration space. However, since our algorithm's moves are diffusion Monte Carlo based, the standard algorithm we wish to compare against should involve a time step rather than a box (length scale). To do so, we use as our standard algorithm an all-electron Metropolis with moves chosen from a Gaussian distribution whose mean is the box size. The value of a time step τ controls the box size. In fact, this is just the all-electron version of standard diffusion Monte Carlo with branching set to unity.

In addition to any microdiagnostic measures, it is also useful to have a global measure of the efficiency of a sampling algorithm. A useful quantity for this is the autocorrelation time[27] of the local energy. This quantity depends not only on the algorithm, but also on the trial wave function employed. Thus it is necessary to compare different sampling methods while employing the same trial wave function. In all the work described here we have employed simple self-consistent-field plus electron-Jastrow forms for the trial wave functions. Choosing approximately optimum time step sizes for each algorithm required only very short simulations to fix the average acceptance ratio for each move at close to 50%.

Of course, measuring the correlation length for an operator other than the Hamiltonian provides a different global efficiency measure; nevertheless the energy is usually the
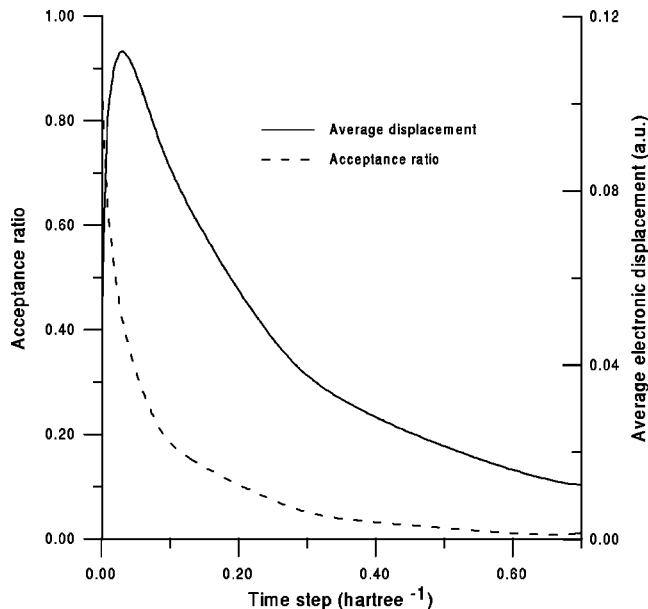

FIG. 2. Plots of mean electronic displacement and acceptance ratio versus time step size for the Be atom. Here the algorithm used is a simple Metropolis moving all electrons at once.

most important quantity in which one is interested. The focus of these investigations has been on the energy correlation (or more properly, decorrelation) time.

We have implemented our partitioning algorithm within the framework of both the Metropolis and the Langevin algorithms. In each case we compare the results obtained to those of algorithms which move all electrons at once and with those which move one electron at a time. Although it takes roughly twice as much computer time to move one electron at a time (*versus* an all-electron move), the former algorithm is the more efficient. This is well-known, but provides a framework in which to observe decorrelation times.

Diagnoses and cure: A detailed analysis

In order to better appreciate why this method can help in alleviating the problem of multiple time scales, we present a detailed analysis of several different simulations of the beryllium atom, showing the causes of the problem and how our proposed algorithm eliminates it.

Beginning with a Metropolis algorithm that moves all the electrons at once, with a fixed time step, Fig. 2 shows the acceptance ratio and the mean displacement obtained. Mean displacement is defined for a single pass of a single electron. As is apparent, the acceptance ratio quickly drops with increasing τ. Moreover, the average displacement is quite small. Note, however, that the old rule of thumb that an acceptance ratio of about 50% is optimal in a Metropolis simulation, is satisfied here. Figure 3 shows the acceptance ratio as a function of *distance from the nucleus*, for different values of τ. One can see that the acceptance ratio is high in the core ($r<1$) only if τ is sufficiently small. In the valence region the acceptance ratio remains fairly constant.

Figure 4 illustrates the problem by showing the mean displacement as a function of r. What it shows is well worth emphasizing, even if it is well-known. Specifically, in order
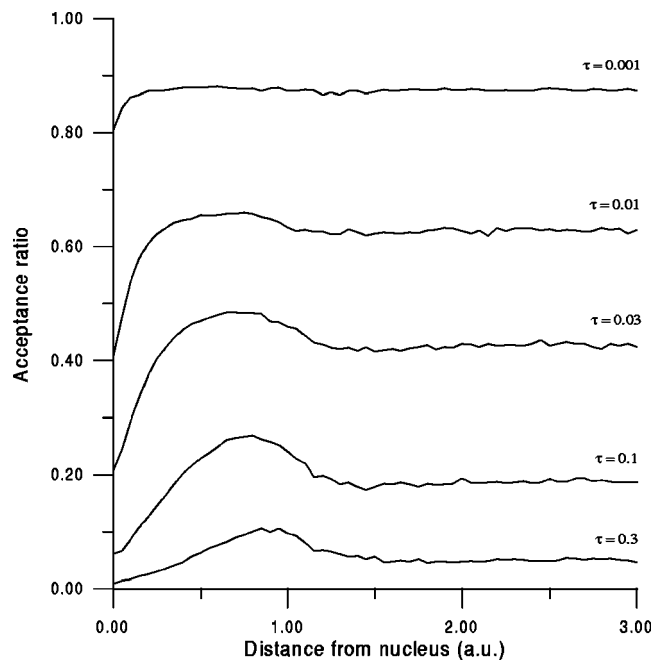
FIG. 3. Acceptance ratio as a function of distance from nucleus for various time steps in a simulation of the Be atom. The algorithm is as in Fig. 2, namely all electrons move at once. The short length scale fluctuations seen are statistical, and are on the order-of-magnitude of the error bars.
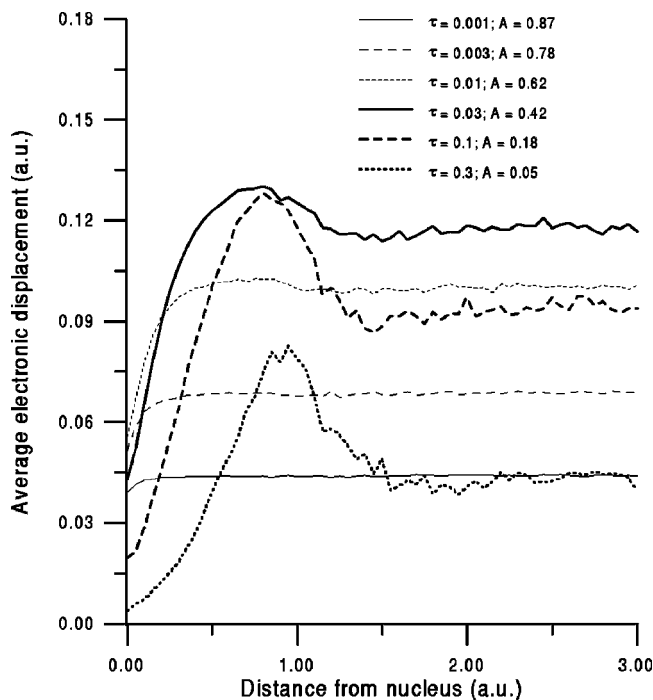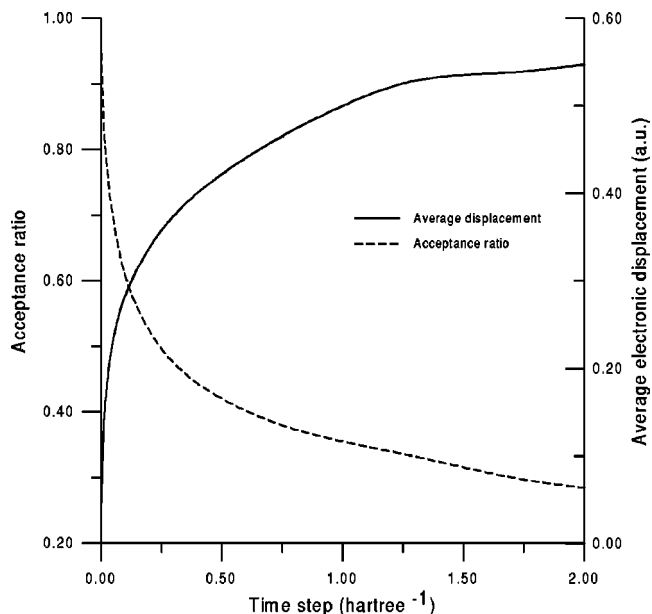

FIG. 5. Plots of mean electronic displacement and acceptance ratio versus time step size for the Be atom, now obtained using a Metropolis algorithm moving one electron at a time.

FIG. 4. The mean electronic displacement for Be as a function of distance from the nucleus. It is apparent that to have any significant movement, particularly in the core, τ must be small. Thus, we see that all-electrons-at-once Metropolis is ''core''-dominated. Such simulation is inefficient because one requires very small times steps in order to sample well everywhere. Note that such very small time steps greatly limit the distance traveled in a single move, regardless of location relative to the nucleus. At the other extreme, large time steps (darker curves) increase the probability that a move will be rejected, as seen in the previous figure. In this limit, the functional form reflects in part the underlying shape of the electron distribution.

to have any significant movement, particularly *in the core*, τ must be small (here <0.1). In other words, this algorithm is core-dominated: if we try to raise the time step, the acceptance ratio drops due to bad core moves, and the walker as a whole cannot move. (Recall, we are moving all electrons at once.) On the other hand, if the acceptance ratio is large (note values of A in Fig. 4) the simulation is inefficient because of the very small moves. Thus, either the simulation is inefficient, or the $3N$-dimensional space is badly sampled. As is typical, somewhere around $A=0.5$ is an optimal (but not necessarily good) tradeoff.

We turn now to what is well-known as a more efficient algorithm. We again use Metropolis, but moving (and accepting) one electron at a time. Figure 5 is the analog of Fig. 2, but the difference between them is dramatic. Of course, the acceptance ratio is now that of a single electron move. It decreases with τ, as expected. However the mean displacement increases with τ for a long way, until it finally reaches a plateau (not shown in the plot) and dies off for $\tau>3.0$. Note also that the displacement here is larger than in the all-electrons-at-once case. Figure 6 shows the details. As before, the core electrons move only for small τ (here $\tau <0.1$) causing the sampling of the valence space to be inefficient. On the other hand, selecting the best τ for the valence space ($\tau>0.7$), or even the best τ overall, results in the core electrons essentially not moving at all. What we are looking for is an algorithm that moves the electrons with the best time step both in the core *and* the valence.

Let us now consider our proposed algorithm. We partition the space, adding our constraints as discussed earlier, namely for Be that $r_1<r_2$ and $r_3<r_4$. Note that for now we still use the *same* τ for all the electrons. However, we can now distinguish outer from inner electrons—they no longer exchange. Figure 7 should be compared to Fig. 5. If nothing else, this method allows for a useful diagnostic. We can now
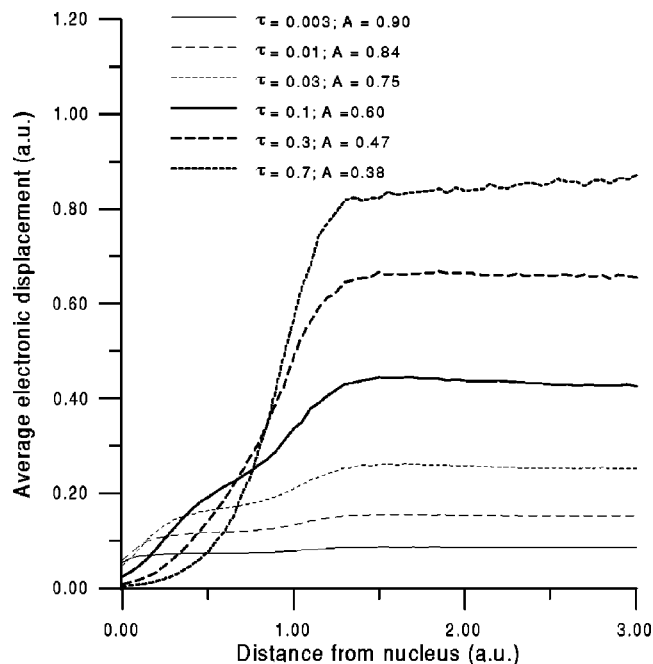
FIG. 6. As in Fig. 4, the mean electronic displacement for Be atom as a function of distance from the nucleus. The algorithm, however, is now Metropolis moving one electron at a time. As can be seen, time steps which are too short still limit the distance traveled. However, large time steps (see darker set of curves), which increase the distance valence electrons can travel, increase the probability that a move will be rejected near the nucleus. The entire set of electrons, however, is no longer constrained by the core electrons.

find the best τ for the core and the best τ for the valence regions separately. This indicates that the one-electron-at-a-time algorithm is valence dominated: in the mean total displacement only the valence moves give a sizable contribution.

With this partitioning we are now in a position to use different values of τ for the different regimes. The plots of Fig. 8 show what happens when one separately optimizes these time steps: specifically we chose $\tau=0.045$ for the core and $\tau=2.5$ for the valence. These values were chosen to bring the respective acceptance ratios to 50%. It can be seen in Fig. 8 that as a function of radial position, the total acceptance ratio is quite graceful, always staying around 50%, and actually passes through 50% twice, once at the center of each orbital. (The plot is superposed with a graph of the electronic density so that one can better see the limits of the core and valence regions.) It can also be seen that the diffusion length follows the overall maximum of the individual mean displacement curves. This is ''the best of both worlds'': the acceptance ratio and displacement follow the small τ behavior in the core and the large τ behavior in the valence region.

All three of the above algorithms were repeated using Langevin Monte Carlo[34] instead of Metropolis. Actually, the Langevin approach as implemented is a hybrid of traditional Langevin and Metropolis. This hybrid maintains the desirable Metropolis property of having no time-step bias. But the Langevin character results in better behavior overall with respect to decorrelating moves. All the same modifications can be made here as in the above, pure Metropolis case. One can move all electrons at once, one electron at a time, and
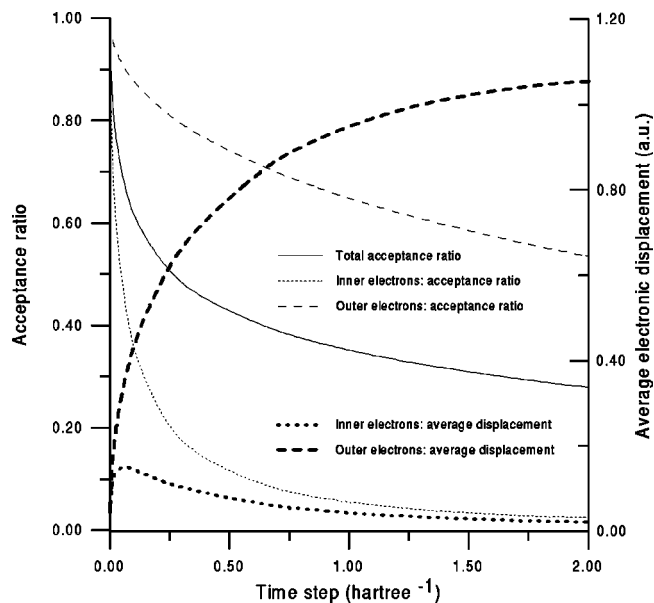

FIG. 7. As in Fig. 5, plots of mean electronic displacement and acceptance ratio versus time step size for the Be atom. The algorithm is Metropolis moving one electron at a time. Having partitioned the space we can now separately plot displacement and acceptance ratio for inner and outer electrons.

give each electron a separate time-step size in a suitably chosen partitioning of the space.

Results

The results for the beryllium atom are summarized in Table I. By exploiting the inherently different time scales of
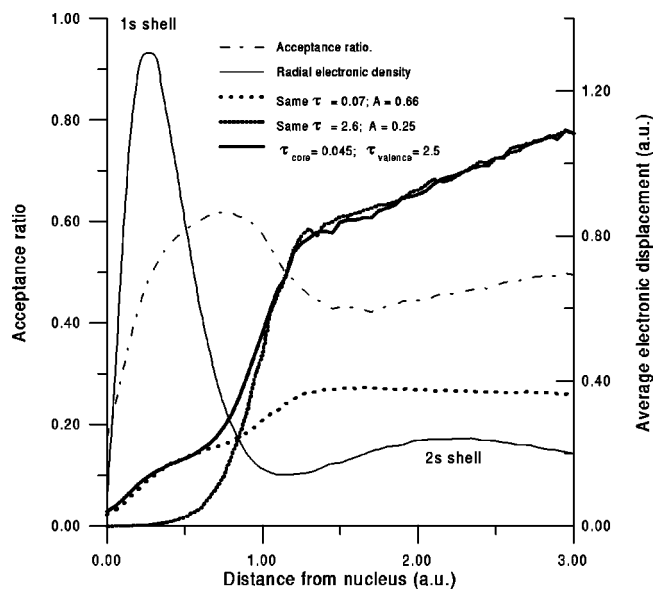

FIG. 8. As in Fig. 6, Be atom mean displacement for different choices of time steps. Because the partitioning algorithm described in the text allows different size time steps for core and valence electrons, a ''best of both worlds'' curve for the average displacement versus distance from the nucleus is possible (bold, solid line). Also indicated is the acceptance ratio versus distance from nucleus for this choice (dash-dot line). The light solid line indicates the radial electronic density to provide perspective on the regions of enhanced sampling.

J. Chem. Phys., Vol. 111, No. 14, 8 October 1999

Variational Monte Carlo   6187

TABLE I. Time to decorrelate moves for Be with various algorithms. Values of τ are in units of hartree^{-1}. Correlation time is dimensionless, and measures the number of steps required to effectively decorrelate two energy measurements.

Algorithm	τ_{core}	$\tau_{valence}$	Correlation time
Metropolis	0.030	0.03	50
Metropolis: individual electron moves	0.100	0.10	12
As above, with separated time scales	0.045	2.50	6
Langevin	0.07	0.07	17
Langevin: individual electron moves	0.10	0.10	7
As above, with separated time scales	0.13	3.50	3.5

TABLE III. Time to decorrelate moves for Li$_2$ with various algorithms. Units as in Table I.

Algorithm	τ_{core}	$\tau_{valence}$	Correlation time
Metropolis	0.05	0.05	35
Metropolis: individual electron moves	0.10	0.10	11.5
As above, with separated time scales	0.08	2.00	8.5
Langevin	0.07	0.07	10
Langevin: individual electron moves	0.30	0.30	6
As above, with separated time scales	0.20	2.00	3.5

core and valence electrons we see that we were able to greatly reduce the time needed to decorrelate moves. The remaining correlation time is not due to the different time scales in the system but instead to aspects of electronic motion within the shells. To further reduce the correlation time, one could couple this algorithm with another specifically designed to alleviate this problem. The Appendix describes one such attempt.

Table II illustrates the results for the neon atom. As discussed earlier we replace the naïve constraint $r_1 < r_2 < r_3 < r_4 < r_5$ AND $r_6 < r_7 < r_8 < r_9 < r_{10}$, which would be the analogy to what was done for the Be atom with the more physically-sound partition $r_1 < r_2, r_3, r_4, r_5$ AND $r_6 < r_7, r_8, r_9, r_{10}$. This derives from chemical intuition leading to the expectation that the neon core is composed of two electrons, with a valence space of the other eight. Thus we end up with a less restrictive constraint which follows the shell structure of the atom. Again we see the effect of separate time scales in decorrelating the Monte Carlo moves. The larger Z for Ne results in a greater effect from the acceleration algorithm.

Finally we present the results for the Li$_2$ molecule. Table III illustrates the results for the autocorrelation time for this simple diatomic. We can imagine this molecule as having two electrons which are always close to the first nucleus (i.e., its core electrons), two electrons likewise close to the second nucleus, and lastly, two electrons free to explore the region

TABLE II. Time to decorrelate moves for Ne with various algorithms. Units as in Table I.

Algorithm	τ_{core}	$\tau_{valence}$	Correlation time
Metropolis	0.005	0.005	100
Metropolis: individual electron moves	0.100	0.100	9.5
As above, with separated time scales	0.006	0.010	5.5
Langevin	0.01	0.01	29
Langevin: individual electron moves	0.03	0.03	6.5
As above, with separated time scales	0.01	0.10	2.5

outside the two cores. Considering just the spin-up electrons, we can define our subspaces by having electron 1 closer to nucleus A than either of the other two; electron 2 is that one of the remaining two which is closest to nucleus B; and electron 3 is then the one, in some sense, farthest away from both nuclei. However, we are still unnecessarily restricting the core electrons, as the two cores are equivalent, and so have the same time-step behavior. Thus, it is sufficient on physical grounds to merely prevent the penetration of the third (outer) electron into either of the cores.

CONCLUSIONS

We have shown that a great improvement in the efficiency of standard VMC algorithms can be achieved by combining a very simple partitioning of the 3-space of the electrons with appropriate time steps for electrons within each partition. This improvement is comparable to what has been obtained using other acceleration approaches which are considerably more difficult to implement. In general, a decrease by a factor of 10 in the autocorrelation time is found over naïve algorithms. This amounts to an effective speedup of simulations by an order-of-magnitude.

Using the present algorithm, core, and the valence electrons—and more generally electrons in different shells and different regions of space—can be made to move at their own optimum rates, independent of the time steps of other electrons. Moreover, the configuration space can be divided on either a physical basis or using other practical criteria. The algorithm is extremely easy to apply, and at almost no additional computational cost. Detailed balance remains trivially satisfied. Application to molecules is not any more difficult. Finally, the algorithm can be used in conjunction with virtually any sampling approach, such as Metropolis and Langevin as demonstrated here, or with others yet to be invented.

In essence, this approach overcomes the autocorrelation resulting from the mixture of time scales due to core/valence exchange. However, within any given shell the algorithm does nothing, so there is room for improvement. In fact, it can be noted that the residual autocorrelation appears to derive mainly from the core.

Future work needs to focus on study of the effect of constrained movement on other sampling methods, on inves-

tigations with larger atoms and more complex molecules, and with other division schemes. One division criterion worth investigation might be the assignment of electrons to specific nuclei, e.g., on a chemical basis, and enforcement of constraints for each nucleus.

Finally, it is significant to note that this acceleration approach, apparently unlike all previous schemes, *can* be extended to full Green's function or diffusion quantum Monte Carlo. The basis for doing so is updating all electrons to the same absolute time (thus, requiring varying numbers of steps for electrons in the different partitions) before calculating a branching factor for the net move. However, the need to synchronize the steps, as indicated, would appear to reduce the efficiency of the acceleration over that described here for VMC. This requires further exploration.

ACKNOWLEDGMENTS

APPENDIX

In this Appendix we describe an alternative approach to accelerate VMC simulations. Our intent was to combine this approach with that described in the text to achieve an even greater decoupling of the steps. However, little additional improvement was discovered. Investigating further we found that, used by itself, this approach is almost as good as that described in the body of this paper. It simply provides little advantage in combination. Nevertheless, this method is simple enough, and a reasonable starting point for further enhancement, that we find it worthwhile to describe it in some detail.

The goal is to modify the Langevin algorithm in such a way that the angular and radial moves of the electrons are decoupled as much as possible. A similar approach has been pursued by Umrigar[27] through a geometric construction.

In the standard Langevin algorithm, each random walker undergoes a displacement from point $\mathbf{R}$ to $\mathbf{R}'$ following the equation

$$\mathbf{R}' = \mathbf{R} + D\tau\mathbf{F}(\mathbf{R}) + \sqrt{2D\tau}\,\chi, \tag{A1}$$

where $\mathbf{F}$ is the usual quantum force, D is the diffusion constant, τ is the step size, and χ is a vector of Gaussian random variables with zero mean and unit width. To decouple the angular movement from the radial movement, while at the same time retaining the simplicity of the Langevin algorithm, we use the quantum force in Eq. (A1) to determine only the *direction* of the movement (more specifically, the angular displacement). We independently choose the radial displacement from a separate probability distribution $P(r \rightarrow r')$.

The steps of the algorithm are the following:

(i) Move from the initial point $\mathbf{R} \equiv (r, \theta, \varphi)$ to an intermediate point $\mathbf{R}_1 \equiv (r_1, \theta_1, \varphi_1)$ using

$$\mathbf{R}_0 = \mathbf{R} + D\tau\mathbf{F}(\mathbf{R}), \tag{A2}$$

$$\mathbf{R}_1 = \mathbf{R}_0 + \sqrt{2D\tau}\,\chi. \tag{A3}$$

This is the usual Langevin movement, where for later convenience we have designated $\mathbf{R}_0$ as the position after the quantum force drift movement alone. Keep the angles (θ_1, φ_1) and discard r_1.

(ii) Generate the new coordinate r' from the old r using a (selectable, and still to be determined) probability distribution $P(r \rightarrow r')$.

(iii) Move to the new trial point $\mathbf{R}' \equiv (r', \theta_1, \varphi_1) \equiv (r', \theta', \varphi')$.

(iv) Accept or reject the move using the Metropolis algorithm, with acceptance probability given by

$$A(\mathbf{R} \rightarrow \mathbf{R}')$$

$$= \min\left(1, \frac{\Psi^2(\mathbf{R}')P(r' \rightarrow r)P_\Omega((\theta', \varphi') \rightarrow (\theta, \varphi))}{\Psi^2(\mathbf{R})P(r \rightarrow r')P_\Omega((\theta, \varphi) \rightarrow (\theta', \varphi'))}\right), \tag{A4}$$

where P_Ω is the probability for the angular movement.

What we need now is the analytical form for P_Ω. Since in step (i) we move from $\mathbf{R} = (r, \theta, \varphi)$ to $\mathbf{R}_1 \equiv (r_1, \theta_1, \varphi_1) = (r_1, \theta', \varphi')$ with probability

$$T(\mathbf{R} \rightarrow \mathbf{R}_1) = (4\pi D\tau)^{-3N/2}e^{-(\mathbf{R}_1 - \mathbf{R} - D\tau\mathbf{F}(\mathbf{R}))^2/4D\tau}, \tag{A5}$$

we need to integrate over all possible radial positions r_1 in order to get the probability of getting a certain pair (θ, φ). The integration is trivially done passing into spherical coordinates. Calling α the angle between $\mathbf{R}_1$ and $\mathbf{R}_0$ we get

$$P_\Omega((\theta, \varphi) \rightarrow (\theta', \varphi'))$$

$$= e^{-r_0^2\sin^2(\alpha)/4D\tau}(1 + \mathrm{Erf}(r_0\cos(\alpha)/\sqrt{4D\tau}))$$

$$\times \sin(\alpha)(\tau + r_0^2\cos^2(\alpha))\sqrt{D\pi\tau}$$

$$+ D\tau r_0\sin(2\alpha)e^{-r_0^2/4D\tau}. \tag{A6}$$

All that remains is the selection of the distribution $P(r \rightarrow r')$. Once we choose this, we can independently optimize the angular step size and the radial step distribution.

We tested this algorithm using different radial transition distributions, $P(r \rightarrow r')$, including both a simple Gaussian and a box $[r'/\Delta, r'\Delta]$, where Δ is here an effective step size. This approach (using either of the radial distributions) gives quite good results. When tested on the beryllium and neon atoms, we obtained a correlation time of about 3.5 for both systems, similar to what we obtained with the partitioning algorithm. Nevertheless, we were not able to further decrease the correlation time by combining the two approaches, despite the effectiveness of spatial partitioning when used alone.

[1] D. Bressanini and P. J. Reynolds, in *Advances in Chemical Physics* (Wiley, New York, 1999), Vol. 105, p. 37.

[2] S. Y. Huang, Z. W. Sun, and W. A. Lester, Jr., J. Chem. Phys. **92**, 597 (1990).

[3] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988).

[4] A. Harju, B. Barbiellini, S. Siljamaki, R. M. Nieminen, and G. Ortiz, Phys. Rev. Lett. **79**, 1173 (1997).

[5] H. Bueckert, S. M. Rothstein, and J. Vrbik, Can. J. Chem. **70**, 366 (1992).

[6] R. N. Barnett, Z. W. Sun, and W. A. Lester, Jr., Chem. Phys. Lett. **273**, 321 (1997).

[7] R. Bianchi, D. Bressanini, P. Cremaschi, M. Mella, and G. Morosi, Int. J. Quantum Chem. **57**, 321 (1996).

[8] D. M. Ceperley, J. Stat. Phys. **43**, 815 (1986).

[9] B. L. Hammond, P. J. Reynolds, and W. A. Lester, Jr., J. Chem. Phys. **87**, 1130 (1987).

[10] P. J. Reynolds, Int. J. Quantum Chem. **24**, 679 (1990).

[11] M. E. Tuckerman, G. J. Martyna, and B. J. Berne, J. Chem. Phys. **97**, 1990 (1992).

[12] G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Phys. Rev. D **32**, 2736 (1985).

[13] P. J. Reynolds, in *Computational Physics and Cellular Automata*, edited by A. Pires, D. P. Landau, and H. J. Herrmann (World Scientific, Singapore, 1990), p. 144.

[14] J. P. Neirotti, P. Serra, and S. Kais, J. Chem. Phys. **108**, 2765 (1998).

[15] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).

[16] R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

[17] G. Murthy and S. Kais, Chem. Phys. Lett. **290**, 199 (1998).

[18] P. Serra and S. Kais, Phys. Rev. A **55**, 238 (1997).

[19] P. A. Christiansen, J. Chem. Phys. **88**, 4867 (1988).

[20] L. Mitas, E. L. Shirley, and D. M. Ceperley, J. Chem. Phys. **95**, 3467 (1991).

[21] T. Yoshida and K. Iguchi, J. Chem. Phys. **88**, 1032 (1988).

[22] C. W. Greeff and W. A. Lester, Jr., J. Chem. Phys. **109**, 1607 (1998).

[23] B. L. Hammond, W. A. Lester, Jr., and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, Singapore, 1994).

[24] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. M. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[25] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods* (J. Wiley & Sons, New York, 1986).

[26] D. M. Ceperley, G. V. Chester, and M. H. Kalos, Phys. Rev. B **16**, 3081 (1977).

[27] C. J. Umrigar, Phys. Rev. Lett. **71**, 408 (1993).

[28] M. Mella, A. Luchow, and J. B. Anderson, Chem. Phys. Lett. **265**, 467 (1997).

[29] D. Bressanini, M. Mella, and M. Morosi (unpublished).

[30] M. L. Stedman, W. M. C. Foulkes, and M. Nekovee, J. Chem. Phys. **109**, 2630 (1998).

[31] D. M. Ceperley, J. Stat. Phys. **63**, 1237 (1991).

[32] W. A. Glauser, W. R. Brown, W. A. Lester, Jr., D. Bressanini, B. L. Hammond, and M. L. Koszykowski, J. Chem. Phys. **97**, 9200 (1992).

[33] Z. W. Sun, M. M. Soto, and W. A. Lester, Jr., J. Chem. Phys. **100**, 1278 (1994).

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