INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/854-20

College on Computational Physics

15 May - 9 June 1995

Metropolis Monte Carlo

C. Umrigar

Cornell University
Ithaca, USA

International Center for Theoretical Physics, Trieste
College on Computational Physics, 15th May - 9th June 1995, Lecture Notes
Lecture 1, 5th June 1995

Introduction to quantum Monte Carlo and (generalized) Metropolis Monte Carlo

Cyrus J. Umrigar

Cornell Theory Center and Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853.

I. INTRODUCTION

Why should one be interested in quantum Monte Carlo(QMC) methods? The number of systems for which exact solutions are possible is very limited. Even approximate solutions, such as from perturbation theory are only possible for a limited set of problems. Quantum Monte Carlo methods on the other hand are widely applicable to wide variety of lattice and continuum systems. They have been used in condensed matter physics to study various lattice models for strongly correlated systems and spin systems, various continuum systems such as liquid and solid helium, droplets, and atomic clusters, and to perform electronic structure calculations of atoms, molecules and solids. They are used in nuclear physics and in lattice gauge theory. Very often the bottleneck in QMC methods is computer time. Since QMC methods can usually be easily parallelized they will become increasingly popular as the advent of massively parallel computers leads to increased computational power.

However, QMC is certainly not a panacea. First of all, non-trivial applications of Monte Carlo methods have a statistical error. For some problems these errors can be made to be quite small, for others not. There are many in-between cases where a naive application of existing QMC methods may lead to an unacceptably large error, but where some thought invested in improved methods leads to sufficiently small statistical errors. Quite often knowledge of an approximate solution can be used to reduce the statistical errors through a technique known as *importance sampling*. (Although path-integral Monte Carlo is usually performed without using a trial wavefunction, there too approximate trial wavefunctions can be used to reduce the statistical error in ground state calculations.)

Nor are statistical errors the only source of error. There are systematic errors as well, though in many cases these can be made negligibly small. Some systematic errors, can be extrapolated away or alternatively overcome by using a more sophisticated (and computationally expensive) method. For example, the time-step error in diffusion Monte Carlo can be removed by either extrapolating to zero time-step or alternatively by using the domain Green function Mote Carlo method.

Quite often in using QMC methods one is faced with a trade-off between the statistical error and systematic error. Sometimes a happy compromise can be found, for example in dealing with the population control error. On the other hand, in dealing with the infamous Fermion sign problem the increase in the statistical error that results from attempts to design algorithms that have negligible systematic errors is sufficiently large that in practice one often decides to live with the systematic "fixed node" errors. This is in spite of the fact that much effort has been expended on trying to solve this problem and many ingenious algorithms have been proposed that work for relatively simple problems.

In these lectures we will discuss four quantum Monte Carlo methods, a) Metropolis Monte Carlo, b) path integral Monte Carlo, c) diffusion Monte Carlo and d) domain Green function Monte Carlo. Only the essence of the last three will be given. The emphasis in these lectures will be on understanding the basics of the methods and especially on the common threads of all the methods, rather than on specific applications. In addition to discussing the Monte Carlo methods we will discuss the functional forms of many-body wavefunctions and the method used for optimizing

the free parameters in the wavefunctions. This is important because it is essential to have good trial wavefunctions in order to keep the statistical error (and sometimes also the systematic error) at an acceptable level.

All the QMC methods, that we will discuss, can be viewed as being stochastic implementations of the power method for calculating the dominant eigenvalue and eigenvector of a matrix, *i.e.*, if we repeatedly multiply an arbitrary vector with a matrix we will eventually project out the dominant eigenvector.

The Metropolis method is the simplest of the Monte Carlo methods and accordingly we will discuss it first. In spite of its simplicity, we will see that some thought invested in an intelligent use of the method can result in very large gain in the efficiency of the method.

II. (GENERALIZED) METROPOLIS MONTE CARLO

The Metropolis method [1] was originally used to sample the Boltzmann distribution and this is still in fact the purpose for which it is most frequently used. However, it is in fact a very general method for sampling any *known* distribution. (Later on we will discuss how to sample a distribution which is the solution of a differential equation but for which one does not have an analytic expression.) Many simple distributions can be sampled directly (see *e.g.* Ref. [4]). Direct sampling methods, where feasible, are preferable to the Metropolis method, since the points generated by the latter are serially correlated. However, most distributions of interest cannot be sampled directly and the Metropolis method comes in very handy. It is one of the most frequently used methods in both classical and quantum Monte Carlo and is a component of some of the more sophisticated forms of quantum Monte Carlo, e.g. path integral Monte Carlo.

Why would one want to sample a distribution? Because by so doing, one can calculate expectation values. These may be thermal expectation values or quantum mechanical expectation values. Suppose we want to calculate the energy expectation value for some trial function $\Psi_T(\mathbf{R})$, where $\mathbf{R}$ denotes the the $3n$ particle coordinates. We can write it in the following form

$$\begin{aligned}
\langle H \rangle &= \frac{\int d\mathbf{R}\, \Psi_T^*(\mathbf{R}) H \Psi_T(\mathbf{R})}{\int d\mathbf{R}\, |\Psi_T(\mathbf{R})|^2} \\
&= \frac{\int d\mathbf{R}\, |\Psi_T(\mathbf{R})|^2 \frac{H\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}}{\int d\mathbf{R}\, |\Psi_T(\mathbf{R})|^2} \\
&= \left\langle \frac{H\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})} \right\rangle_{|\Psi_T(\mathbf{R})|^2} \\
&\approx \frac{1}{N} \sum_{i=1}^{N} \frac{H\Psi_T(\mathbf{R}_i)}{\Psi_T(\mathbf{R}_i)} \quad \text{with the } \mathbf{R}_i \text{ sampled from } |\Psi_T(\mathbf{R})|^2 \qquad (1)
\end{aligned}$$

Note that as Ψ_T approaches an eigenstate, the statistical error of the energy estimate tends to zero. In the limit that it equals an eigenstate, just one Monte Carlo point is enough to give us the exact energy! So, if we already know the answer, we can recover it with negligible work.

The notion of *importance sampling* is that one can evaluate an integral by writing the integrand as the product of two terms. The first term is a distribution that can be sampled directly. A simple arithmetic average of the second term, using points sampled from the first yields an estimate of the integral.

$$\int d\mathbf{R}\, F(\mathbf{R}) = \frac{\int d\mathbf{R}\, \rho(\mathbf{R}) f(\mathbf{R})}{\int d\mathbf{R}\, \rho(\mathbf{R})} \qquad (2)$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{R}_i) \quad \text{with the } \mathbf{R}_i \text{ sampled from } \rho(\mathbf{R}), \qquad (3)$$

where $f(\mathbf{R}) = \frac{F(\mathbf{R})}{\rho(\mathbf{R})} \int d\mathbf{R}\, \rho(\mathbf{R})$. The idea is to put as much of the variations of the integrand as possible into the first term ρ, so that the second term f fluctuates as little as possible, thereby yielding an estimate with a small statistical error. Note that although we do not need to know the integral of ρ in order to sample it, we do need to know the integral in order to define f and this in practice limits the available choices for ρ. The example in the previous paragraph is an instance of importance sampling, but it is more than that. As the quality of the wavefunction is improved, not only is the statistical error reduced, but the expectation value itself is improved. Note that for an operator which does not commute with the Hamiltonian, one does not achieve a zero-variance estimator in the limit that Ψ_T is an eigenstate.

We consider a system, the states of which are described by a set of continuous and/or discrete variables. We will use discrete notation, but most of what follows will be equally valid for continuum systems. We wish to sample the states from a known distribution $\rho(i)$, where i stands for the collection of all the variables that characterize the system. Starting from an initial state i the final state f will be sampled from the elements of the i^{th} column of a stochastic matrix M. This is a Markov process since the proposed move depends only on the current state i and not on the prior history.

Definitions

• A *stochastic* matrix is one whose elements are non-negative and whose columns sum to 1, *i.e.* $\sum_f M(f|i) = 1$. So, starting from any initial state, the system must evolve to some state (including the initial state) with unit probability.

• A *reducible* matrix is one that can be brought into block triangular form by permutations of rows and columns.

The matrix M must be such that if we start with the desired distribution ρ, the states obtained after one application of M to ρ are also sampled from ρ. The necessary and sufficient condition for this is

$$\sum_f M(f|i)\ \rho(i) = \sum_f M(i|f)\ \rho(f) \quad \text{for all states } i \tag{4}$$

This says that the total flow from state i into all states, must equal the total flow from all states into state i. In practice, a more stringent condition is imposed. *Detailed balance* is a sufficient (but not necessary) condition:

$$M(f|i)\ \rho(i) = M(i|f)\ \rho(f) \tag{5}$$

This expresses the condition that for any pair of states i and f the probability of being in a state i and making a transition to a state f is equal to the probability of the reverse process. It is clear, that if one starts with the correct distribution, detailed balance implies that we will continue to sample the correct distribution. What we now show is that the stochastic property of M and detailed balance imply that any starting distribution will evolve into the desired distribution ρ. The proof applies only to finite matrices.

Since M is stochastic, the vector with all elements equal to 1 is a left eigenvector of M with eigenvalue 1. It follows from the Perron-Frobenium theorem that an irreducible matrix with non-negative elements which satisfies detailed balance [5] must have a dominant eigenvalue that is positive and non-degenerate and the corresponding eigenvector must have only non-negative components. For any matrix the left and right eigenvectors corresponding to different eigenvalues are orthogonal. So, the components of all other eigenvectors cannot all have the same sign. Hence, the dominant eigenvalue of M is 1. Now, using detailed balance and stochasticity of M we obtain

$$\sum_i M(f|i)\ \rho(i) = \sum_i M(i|f)\ \rho(f) = \rho(f). \tag{6}$$

So, ρ is a right eigenvector of M with eigenvalue 1. Since it is the dominant right eigenvector, the distribution must evolve into ρ.

We have shown that an arbitrary distribution evolves into ρ when it is repeatedly acted upon by M. The rate at which the initial distribution evolves to the desired distribution ρ and the autocorrelation time of estimates of various observables is governed by the other eigenvalues. The ideal situation (never realized in practical applications) is that

all the other eigenvalues are zero. In that case the distribution evolves to ρ in a single Monte Carlo step and every measurement is independent.

Note that both the Metropolis method and the *projector* Monte Carlo methods we will discuss later are stochastic implementations of the power method for projecting out the dominant eigenvector of a matrix (or integral kernel). In the Metropolis method one designs a matrix which projects out a known distribution, whereas in the projector Monte Carlo methods one projects out the (unknown) solution to a differential equation. In the former case one has a great deal of freedom in designing an efficient algorithm, since there are no restrictions on M other than that it should be ergodic and satisfy detailed balance. In the latter case the available choices are much more limited. Unfortunately, reseachers often opt for the simplest form of the Metropolis method and only rarely exploit fully the freedom in the choice of M. Accordingly, we will discuss this in detail.

In order to proceed further, it is useful to write the elements of the Markov matrix M as the product of the elements of a proposal (apriori transition) matrix T and an acceptance Matrix A, $M(f|i) = A(f|i)\ T(f|i)$. $M(f|i)$ and $T(f|i)$ are stochastic matrices, but $A(f|i)$ is not. The detailed balance equation is now

$$A(f|i)\ T(f|i)\ \rho(i) = A(i|f)\ T(i|f)\ \rho(f) \tag{7}$$

or

$$\frac{A(f|i)}{A(i|f)} = \frac{T(i|f)\ \rho(f)}{T(f|i)\ \rho(i)}\ . \tag{8}$$

There are an infinity of choices for the acceptance matrix that satisfy this equation. The optimal choice, which maximizes the acceptance, is the one prescribed by Metropolis *et al.* [1]

$$A(f|i) = \min\left\{1, \frac{T(i|f)\ \rho(f)}{T(f|i)\ \rho(i)}\right\}. \tag{9}$$

In later work, other choices for $A(f|i)$ have sometimes been made, *e.g.*

$$A(f|i) = \frac{T(i|f)\ \rho(f)}{T(i|f)\ \rho(f) + T(f|i)\ \rho(i)}. \tag{10}$$

However, it is clear that (for a given choice of the proposal matrix T) all choices other than that in Eq. 9 lead to a slower evolution of the system and consequently a less efficient algorithm. Despite this, I have seen at least one paper in the literature that attempts to study empirically whether the choice given in Eq. 9 or Eq. 10 is the more efficient one!

Now we make several observations as regards the Metropolis method.

1. The distribution ρ to be sampled need not be normalized. The Metropolis method automatically samples $\rho/\int\rho$ and can only be used to calculate expectation values of the form $\int\rho f/\int\rho$.
2. The rate of convergence to the desired distribution and the autocorrelation time of estimates of observables is governed by the sub-dominant eigenvalues of M.
3. The autocorrelation time must be taken into account in estimating the statistical errors.
4. There is folklore that an optimal acceptance ratio is close to 1/2. However, I have found instances where the optimal is as small as 0.2 or as large as 0.9. A much better criterion is to maximize the rate at which the system diffuses through configuration space $\langle(\vec{R}_f - \vec{R}_i)^2\rangle \approx A Delta^2$, where A is the average acceptance and Δ is the average size of the proposed moves. The real measure of goodness is of course to minimize the autocorrelation time for the observables of interest.

5. Using any admixture of different kinds of moves, such that each of them satisfies detailed balance and they are collectively ergodic is legal.

6. The heat bath algorithm is a special case of (generalized) Metropolis. $T(f|i) \propto \rho(f)$ for only a small set of accessible states in a domain $D(i)$ in the neighborhood of i:

$$T(f|i) = \begin{cases} \rho(f)/\sum\rho(f) & \text{if } f \epsilon D(i) \\ 0 & \text{otherwise} \end{cases}$$

Then

$$A(f|i) = \frac{\sum_{D(i)}\rho(f)}{\sum_{D(f)}\rho(i)} \tag{11}$$

If the sum over the accessible states from i and f is the same, the acceptance is unity. The heat bath algorithm is usually used for lattice systems, where the sum over a restricted set of final states can be easily performed.

Since the optimal choice of A for a given choice of T is obvious, we will discuss in the rest of the lecture what is an efficient choice for T. The original Metropolis method [1] employed a transition matrix $T(i|f)$ which is symmetric in i and f, in which case the factors of T in Eq. 9 cancel. Hence, we have called the method presented here, the *generalized* Metropolis method. Unfortunately many present day applications also make this inefficient choice. The first observation (to my knowledge) in the literature that $T(i|f)$ need not be symmetric did not occur until 1970 [2], although if one formulates the problem in the manner given above, it seems totally obvious. The ideas developed will be illustrated by applying them to electronic structure problems. There one has two additional complications to deal with. a) One has multiple length scales – the core electrons move on a length scale of $1/Z$ where Z is the nuclear charge, whereas the valence electrons move on a length scale of unity. b) The distribution to be sampled $|\Psi_T(\mathbf{R})|^2$ has non-analyticities which invalidate our analysis based on a Taylor expansion of the distribution. The method used to construct an efficient proposal matrix that solves these problems is contained in Ref. [6] and therefore will not be repeated in these notes.

---

[1] N. Metropolis *et al.*, *J. Chem. Phys.* **21**, 1087 (1953).

[2] W.K. Hastings, *Biometrika* **57**, 97 (1970).

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

[4] M.H. Kalos and P.A. Whitlock, *Monte Carlo Methods*, Vol. 1, (Wiley, 1986).

[5] We include the detailed balance condition here to rule out cyclic matrices which have eigenvalues that are degenerate in absolute magnitude. For example, the eigenvalues of an n^{th} order cyclic stochastic matrix (*i.e.*a matrix that can be transformed into a matrix with one's on the superdiagonal and the bottom left corner by a permutation of rows and columns and zero's elsewhere) are the n^{th} roots of unity. Since a cyclic matrix cannot satisfy detailed balance for any ρ they are ruled out.

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

Accelerated Metropolis Method

C. J. Umrigar
Theory Center and Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853
(Received 19 April 1993)

It is shown that the freedom in the choice of the proposal matrix in the generalized Metropolis method can be used to greatly enhance the efficiency of the method. For example, the difficulties associated with the existence of multiple length scales in electronic structure calculations can be avoided by making an intelligent choice. Results are presented for Ne, Ar, and Li_2.

PACS numbers: 71.10.+x, 31.15.+q

Introduction.—In many branches of physics, simulations of systems with multiple length scales are very time consuming. Accelerated methods, that involve making collective moves of the degrees of freedom, have been developed for lattice problems [1–3]. Here we present an accelerated method generally applicable to continuum problems and demonstrate its efficiency by applying it to calculations of the electronic energy of atoms and molecules. In this case, much of the interesting physics or chemistry is related to the valence electrons but the size of the Monte Carlo (MC) moves is restricted by the much shorter length scale of the $1s$ core electrons [4]. It is shown that an accelerated Metropolis algorithm, wherein each electron attempts moves that are proportional to its distance from the nearest nucleus, enhances greatly the rate at which the system evolves.

The generalized Metropolis algorithm.—The importance of the Metropolis method [5] in computational science is due to the fact that it is a simple and powerful method for sampling *any* known distribution $f(\mathbf{R})$, where $\mathbf{R}$ labels the degrees of freedom of the system, which may be continuous or discrete. We review here a generalization [6,7] of the Metropolis algorithm which yields an infinite family of algorithms depending on the choice of the proposal matrix $T(\mathbf{R}_f|\mathbf{R}_i)$. We show that the efficiency of the method can be enhanced greatly by a suitable choice of $T(\mathbf{R}_f|\mathbf{R}_i)$.

Let $T(\mathbf{R}_f|\mathbf{R}_i)=S(\mathbf{R}_f|\mathbf{R}_i)/I(\mathbf{R}_i)$, where $I(\mathbf{R}_i)=\int d\mathbf{R}_f\, S(\mathbf{R}_f|\mathbf{R}_i)$, be the probability for an attempted move from $\mathbf{R}_i$ to $\mathbf{R}_f$ and let $A(\mathbf{R}_f|\mathbf{R}_i)$ be the probability for the move to be accepted. Assuming ergodicity, the equilibrium distribution is $f(\mathbf{R})$, provided that the ratio of acceptance probabilities is chosen to satisfy detailed balance

$$\frac{A(\mathbf{R}_f|\mathbf{R}_i)}{A(\mathbf{R}_i|\mathbf{R}_f)}=\frac{f(\mathbf{R}_f)}{f(\mathbf{R}_i)}\frac{T(\mathbf{R}_i|\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i)}=\frac{f(\mathbf{R}_f)}{f(\mathbf{R}_i)}\frac{S(\mathbf{R}_i|\mathbf{R}_f)}{S(\mathbf{R}_f|\mathbf{R}_i)}\frac{I(\mathbf{R}_i)}{I(\mathbf{R}_f)}. \tag{1}$$

Note here that we have complete freedom [7] in the choice of the attempt probabilities T. *For a given choice of* T, the optimal choice for the acceptance probabilities, is given by

$$A(\mathbf{R}_f|\mathbf{R}_i)=\min\left\{\frac{f(\mathbf{R}_f)}{f(\mathbf{R}_i)}\frac{T(\mathbf{R}_i|\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i)},1\right\}. \tag{2}$$

For this choice, $A(\mathbf{R}_f|\mathbf{R}_i)$ is largest and therefore the system evolves the fastest. Many applications of the Metropolis method (and this is the form in which it was originally formulated [5]), make the simple but inefficient choice that T is symmetric, in which case the factors of T in Eq. (2) cancel.

The drawback of the Metropolis method is that the points sampled are sequentially correlated, resulting in a loss of computational efficiency. If it takes, on the average, T_{corr} MC moves for an observable to be decorrelated, then the effective number of independent observations in a MC run of length N is only N/T_{corr}. It is clearly advantageous to reduce the autocorrelation time T_{corr}. This can be done by either increasing the average size of the proposed moves or by increasing the acceptance of the moves. In order to prevent the acceptance from getting too small, it is common practice to restrict the moves to be in the neighborhood of $\mathbf{R}_i$ by choosing $S(\mathbf{R}_f|\mathbf{R}_i)$ to be nonzero only within a domain $D(\mathbf{R}_i)$ of volume $\Omega(\mathbf{R}_i)$ around $\mathbf{R}_i$. For a given functional form of $S(\mathbf{R}_f|\mathbf{R}_i)$ the acceptance decreases as $\Omega(\mathbf{R}_i)$ (and therefore the average size of the proposed moves) increases, so, there exists an optimal $\Omega(\mathbf{R}_i)$ for which the system evolves the fastest.

In this Letter we propose a $S(\mathbf{R}_f|\mathbf{R}_i)$ with both large moves and large acceptance. $S(\mathbf{R}_f|\mathbf{R}_i)$ should be viewed as being a function of $\mathbf{R}_f$ which depends parametrically on $\mathbf{R}_i$. Our task is to find a functional form for $S(\mathbf{R}_f|\mathbf{R}_i)$ such that $\int d\mathbf{R}_f S(\mathbf{R}_f|\mathbf{R}_i)$ is known, $T(\mathbf{R}_f|\mathbf{R}_i)$ can be sampled directly, the resulting proposed moves are large and the acceptance in Eq. (2) is large on average. It was observed in Ref. [7] that if $I(\mathbf{R}_i)$ is independent of $\mathbf{R}_i$ then the choice $S(\mathbf{R}_f|\mathbf{R}_i)\sim f(\mathbf{R}_f)$ results in $A(\mathbf{R}_f|\mathbf{R}_i)/A(\mathbf{R}_i|\mathbf{R}_f)\approx 1$. However, the only way to have $I(\mathbf{R}_i)$ be independent of $\mathbf{R}_i$ and $S(\mathbf{R}_f|\mathbf{R}_i)\sim f(\mathbf{R}_f)$ is to have $S(\mathbf{R}_f|\mathbf{R}_i)$ be independent of $\mathbf{R}_i$. However, for most $f(\mathbf{R})$ of interest, it is not possible to find a function that approximates it sufficiently well over the entire domain of $F(\mathbf{R})$ and which can be sampled directly. Hence, as mentioned before, $S(\mathbf{R}_f|\mathbf{R}_i)$ is chosen to be

nonzero only in some domain $D(\mathbf{R}_i)$ of volume $\Omega(\mathbf{R}_i)$ around $\mathbf{R}_i$, so that the proposed moves are within this domain. In that case $I(\mathbf{R}_i) \approx S(\mathbf{R}_i|\mathbf{R}_i)\Omega(\mathbf{R}_i)$ and Eq. (1) becomes

$$\frac{A(\mathbf{R}_f|\mathbf{R}_i)}{A(\mathbf{R}_i|\mathbf{R}_f)} \approx \frac{f(\mathbf{R}_f)}{f(\mathbf{R}_i)} \frac{S(\mathbf{R}_i|\mathbf{R}_f)}{S(\mathbf{R}_f|\mathbf{R}_i)} \frac{S(\mathbf{R}_i|\mathbf{R}_i)}{S(\mathbf{R}_f|\mathbf{R}_f)} \frac{\Omega(\mathbf{R}_i)}{\Omega(\mathbf{R}_f)}, \tag{3}$$

from which it is apparent that the choice

$$S(\mathbf{R}_f|\mathbf{R}_i) = g_i(\mathbf{R}_f)/\sqrt{\Omega(\mathbf{R}_f)} \tag{4}$$

yields $A(\mathbf{R}_f|\mathbf{R}_i)/A(\mathbf{R}_i|\mathbf{R}_f) \approx 1$, if $g_i(\mathbf{R}_f) \sim \sqrt{f(\mathbf{R}_f)}$ within $D(\mathbf{R}_i)$ [8]. To be more precise, it can be shown [9] by Taylor expansion that if the logarithmic derivatives of $g_i(\mathbf{R}_f)$ at $\mathbf{R}_i$ equal those of $\sqrt{f(\mathbf{R}_f)}$ then the average acceptance goes as $1+\mathcal{O}(\Delta^m)$, where Δ is the linear dimension of $D(\mathbf{R}_i)$. In general, $m=2$, but if $D(\mathbf{R}_i)$ is inversion symmetric with $\mathbf{R}_i$ at center, then $m=3$. This is a considerable improvement compared to using a symmetric $S(\mathbf{R}_f|\mathbf{R}_i)$ or choosing $S(\mathbf{R}_f|\mathbf{R}_i) = f(\mathbf{R}_f)$ for either of which the average acceptance goes as $1+\mathcal{O}(\Delta)$.

In the case of electronic structure calculations the probability distribution $f(\mathbf{R})$ is $|\Psi(\mathbf{R})|^2$, where $\Psi(\mathbf{R})$ is a trial wave function and $\mathbf{R}$ specifies the $3n$ electron coordinates. Hence Eq. (4) becomes

$$S(\mathbf{R}_f|\mathbf{R}_i) = |\Phi_i(\mathbf{R}_f)|/\sqrt{\Omega(\mathbf{R}_f)}, \tag{5}$$

where $\Phi_i(\mathbf{R}_f) \sim \Psi(\mathbf{R}_f)$ in $D(\mathbf{R}_i)$ and has logarithmic derivatives at $\mathbf{R}_i$ that match those of $\Psi(\mathbf{R}_f)$. We now discuss the explicit forms of $S(\mathbf{R}_f|\mathbf{R}_i)$ for which we present results here. For the first two choices $\Omega(\mathbf{R}_i)$ is constant, independent of $\mathbf{R}_i$.

Simple metropolis.—The simplest form is $S(\mathbf{R}_f|\mathbf{R}_i)$ equal to a constant when $\mathbf{R}_f$ is within a $3n$-dimensional hypercube (box) of linear dimension 2Δ centered at $\mathbf{R}_i$ and zero elsewhere. Then $T(\mathbf{R}_f|\mathbf{R}_i) = (2\Delta)^{-3n}$ within the box and zero elsewhere. Aside from the question of whether one or all particles are moved in a single MC update, this is the original Metropolis method [5] and it is the form in which it has most frequency been used.

Cartesian coordinate directed Metropolis.—According to Eq. (5), an improved form of S is $S(\mathbf{R}_f|\mathbf{R}_i) = |\Psi_i(\mathbf{R}_f)|$ within the box. The simplest choice of $\Phi_i(\mathbf{R}_f)$ is a linear approximation (in each of the $3n$ Cartesian coordinate directions) to $\Psi(\mathbf{R}_f)$ at $\mathbf{R}_i$.

Two features of the wave function restrict the size of the attempted moves. First, probable electron configurations have two core electrons at a distance of $\mathcal{O}(1/Z)$ Bohr radii from each nucleus, whereas valence electrons are typically a distance $\mathcal{O}(1)$ from the nearest nucleus [10]. Hence, if the same value of Δ is used for all the electrons, as is usually the case, then the core electrons set the length scale for all the electrons, else a large fraction of the proposed moves are rejected. Second, $\Psi(\mathbf{R})$ has a derivative discontinuity when an electron is at a nucleus, which renders any Taylor series approximation, in Cartesian coordinates, of $\Psi(\mathbf{R})$, very inaccurate. (There is also a derivative discontinuity when two electrons overlap, but this is less problematic since electrons repel each other.) A natural solution to both these problems is provided by the use of spherical polar rather than Cartesian coordinates in proposing the Monte Carlo moves.

Spherical polar coordinate directed Metropolis.—We now describe a choice for $S(\mathbf{R}_f|\mathbf{R}_i)$ which allows each electron to make a move appropriate to its length scale and which avoids the derivative discontinuity in $\Psi(\mathbf{R})$. Each electron moves in a volume which is the intersection of a cone which subtends an angle $2\theta_M$ at the closest nucleus and a concentric spherical annulus. Let $r_{k,i}$ be the initial distance of the kth electron from the nucleus closest to it. The proposed move will use spherical polar coordinates centered at that nucleus. The radial moves are made in the interval $(r_{k,i}/\Delta_r, r_{k,i}\Delta_r)$. Hence the size of the radial moves is proportional to $r_{k,i}$. For atoms, this ensures that if a move is possible, then the reverse move is also possible. We will discuss the necessary changes for molecules and solids later. The advantages of moving in this volume are that the size of the valence electron moves is not restricted by the core electrons and that Taylor expansions of the wave function are valid in the region of interest since $\Psi(\mathbf{R})$ does not have derivative discontinuities in spherical polar coordinates at nuclei.

Since

$$\Omega(\mathbf{R}_f) = \prod_{k=1}^{n} (2\pi/3)(1-\cos\theta_M)(\Delta_r^3 - 1/\Delta_r^3) r_{k,f}^3,$$

according to Eq. (5), $S(\mathbf{R}_f|\mathbf{R}_i) = |\Phi_i(\mathbf{R}_f)|/\prod_{k=1}^{n} r_{k,f}^{3/2}$. In order to simplify the sampling, $\Phi_i(\mathbf{R}_f)$ is chosen to have the form

$$\Phi_i(\mathbf{R}_f) = \prod_{k=1}^{n} W_i(\phi_{k,f}|\theta_{k,f}, r_{k,f}) V_i(\theta_{k,f}|r_{k,f}) U_i(r_{k,f}).$$

For each electron k, the radial coordinate $r_{k,f}$ is sampled from $|U_i(r_{k,f})|\sqrt{r_{k,f}}$, then $\theta_{k,f}$ is sampled from $|V_i(\theta_{k,f}|r_{k,f})|\sin\theta_{k,f}$ conditional upon $r_{k,f}$ and finally $\phi_{k,f}$ is sampled from $|W_i(\phi_{k,f}|\theta_{k,f}, r_{k,f})|$ conditional on $r_{k,f}$ and $\theta_{k,f}$. The precise choice of the functions W_i, V_i, U_i is described elsewhere [9].

It is possible to make further improvements by making θ_M a function $r_{k,f}$ and $r_{k,i}$. The reason it is advantageous to do so is that most trial functions have a finite discontinuity of magnitude Z in the local energy when two electrons approach a nucleus [11], Z being the nuclear charge. In this limit the local energy is Z hartrees higher when the two electrons and the nucleus lie along a straight line with the electrons on opposite sides of the nucleus than with the electrons on the same side. It is desirable to average over this discontinuity as rapidly as possible by making large angular moves when $r_{av} \ll 1/Z$, where $r_{av} = (r_{k,i} + r_{k,f})/2$. Hence $\cos\theta_M$ is chosen to be

$$\cos\theta_M = \cos\theta_m - \frac{1+\cos\theta_m}{1+(Zr_{av})^2}, \quad (6)$$

where θ_m is fixed, which has the limiting behaviors $\theta_M = \theta_m$ when $r_{av} \gg 1/Z$ and $\theta_M = \pi$ when $r_{av} \ll 1/Z$. In spite of the fact that when an electron is close to a nucleus, the angular move is made over the entire surface of a sphere, the acceptance of these moves is close to one [9]. On the other hand when moves are made in Cartesian coordinates, the acceptance for the electrons close to the nucleus is very low [9].

Finally we note that although each of the above algorithms assumes that all the electrons are moved during each Monte Carlo update, it is trivial to modify the algorithms to move only one electron at each MC step or in fact any number in between.

Generalization to molecules and solids.—The additional complication for molecules and solids is that the closest nuclei to each of the n electrons at $\mathbf{R}_i$ need not be the closest nuclei to the corresponding electrons at $\mathbf{R}_f$. For some fraction of these the reverse move from $\mathbf{R}_f$ to $\mathbf{R}_i$ is not possible, i.e., $S(\mathbf{R}_i|\mathbf{R}_f) = 0$, because whereas $\mathbf{R}_f$ lies in $D(\mathbf{R}_i)$, $\mathbf{R}_i$ may not lie in $D(\mathbf{R}_f)$. In that case detailed balance demands that the move from $\mathbf{R}_i$ to $\mathbf{R}_f$ be rejected, i.e., $A(\mathbf{R}_f|\mathbf{R}_i) = 0$. Since these rejections may be performed on purely geometrical grounds (they do not require calculating the wave function or its derivatives at $\mathbf{R}_f$), this does not result in an appreciable loss of efficiency.

Results.—The efficiency of the algorithms is inversely proportional to the autocorrelation time of observables of interest. Table I shows the autocorrelation time of the energy for four algorithms and four wave functions. The four algorithms are (1) the simple Metropolis algorithm moving all electrons at each MC step, (2) the Cartesian coordinate directed Metropolis algorithm moving all electrons at each MC step, (3) the Cartesian coordinate directed Metropolis algorithm moving only one electron at each MC step, and (4) the spherical-polar coordinate directed Metropolis algorithm moving only one electron at each MC step. The four wave functions used are (1) a simple Ne wave function, (2) a good Ne wave function, (3) a simple Ar wave function, and (4) a simple Li_2 wave function. The simple wave functions consist of a determinant multiplied by a simple Jastrow function which is a function of the interelectron distances only, whereas the good wave function consists of a determinant multiplied by a more complicated Jastrow function which is a function of both the interelectron distances and the electron-nuclear distances [12].

The measure of efficiency of the algorithm is the autocorrelation time which is determined as follows. The entire MC run (after discarding the equilibration updates) consists of N MC updates that are divided into N_b blocks each consisting of N_s MC steps for each of the n electrons. The local energy is measured after each MC update. The autocorrelation time is given by $T_{corr} = N_s(\sigma_b/\sigma)^2$, where σ and σ_b are the rms fluctuations of the individual energies and the block average energies, respectively. N_s must be chosen such that $N_s \gg T_{corr}$. The autocorrelation times presented in Table I were obtained using values of N_s that were at least 100 times greater than T_{corr}. It was found that using $N_s \approx 10\,T_{corr}$ resulted in estimates of T_{corr} that were too low by as much as 20%. When MC moves consist of moving one electron at a time it takes twice the computer time to move all the electrons as compared to when they are all moved at once [7]. Hence Table I has values of $T^*_{corr} = 2T_{corr}$ for algo-

TABLE I. Autocorrelation times for the four wave functions and the four algorithms discussed in the text. In order to have a fair comparison, $T^*_{corr} = T_{corr}$ for algorithms 1 and 2 and $T^*_{corr} = 2T_{corr}$ for algorithms 3 and 4. In algorithm 4, $\theta_m = \pi/2$ for Ne and Ar and $\theta_m = \pi$ for Li_2. $\bar{A}$ is the average acceptance. The uncertainty in T^*_{corr} is typically 10% of its value.

Wave function	Algorithm	Δ	Δ_r	$\bar{A}$	T^*_{corr}
Ne simple	1	0.25	···	0.288	84
$\bar{E} = -128.716$ hartree	2	0.3	···	0.661	28
$E_{corr} = 43\%$	3	0.8	···	0.769	13
$\sigma = 1.8$ hartree	4	···	5	0.708	2.0
Ne good	1	0.25	···	0.290	37
$\bar{E} = -128.901$ hartree	2	0.3	···	0.663	11
$E_{corr} = 91\%$	3	0.8	···	0.771	7.2
$\sigma = 0.91$ hartree	4	···	5	0.708	1.7
Ar simple	1	0.12	···	0.307	190
$\bar{E} = -527.1$ hartree	2	0.2	···	0.474	44
$E_{corr} = 37\%$	3	0.5	···	0.813	13
$\sigma = 4.0$ hartree	4	···	5	0.620	2.2
Li_2 simple	1	0.75	···	0.268	210
$\bar{E} = -14.9476$ hartree	2	1	···	0.614	56
$E_{corr} = 61\%$	3	2	···	0.616	32
$\sigma = 0.41$ hartree	4	···	8	0.775	5.8

rithms 3 and 4. We note that σ and the variational energy $\bar{E}$ depend on the trial wave function but not on the algorithm used. T_{corr} and the acceptance $\bar{A}$ depend on both the algorithm and the trial wave function.

For each of the algorithms the values of Δ or Δ_r and θ_m were optimized to yield the smallest possible T_{corr}. Table I shows the values of T^*_{corr} for approximately optimal values of the parameters. Moderate variations of the parameters about their optimal values affects the efficiency of the algorithm only slightly. For example, changing the value of Δ_r from 5 to 4 altered the value of T_{corr} by less than 10% for each of the wave functions. For the four wave functions, the autocorrelation times get smaller by factors of 42, 22, 86, and 36, respectively, as we progress from algorithm 1 to 4 and by factors of 6.5, 4.2, 5.9, and 5.5 in going from algorithm 3 to 4.

Note that for each of the four algorithms T_{corr} is smaller for the good Ne wave function than for the simple Ne wave function. The reason for this is that the good Ne wave function has a more rapidly varying local energy (but with a smaller amplitude of course) and consequently it takes fewer Monte Carlo steps to wander from a region where the local energy is too high to one where it is too low and vice versa. Hence the gain in efficiency from improving the wave function is greater than would be supposed by merely comparing their respective values of σ.

The variational energy for the good Ne wave function is -128.9005 ± 0.0005 hartree, corresponding to 91% of the correlation energy. This is the lowest energy calculated to date by variational Monte Carlo for Ne. The fact that the energy is good is due to the high quality of the wave function, but the fact that the energy could be determined with a small statistical error, in just a few hours on a workstation, is testimony to the efficiency of the new algorithm.

Comparison of the results for algorithms 2 and 3 shows, as has been noted before [7], that for systems with many electrons, and for this class of algorithms, it is more efficient to move one electron at a time rather than all at once.

For algorithm 1 there is a considerable increase in T_{corr} in going from Ne to Ar. On the other hand for algorithm 4, T_{corr} increases very little. This does not mean that calculations of heavy atoms can be performed as rapidly as those of light atoms. The time for evaluating determinants in the wave functions scales as Z^3 and the fluctuation in the local energy σ is empirically found to scale roughly as Z or $Z^{1.5}$. Consequently the computer time required to obtain results with a fixed statistical uncertainty increases as Z^5 or Z^6 if T_{corr} is independent of Z and yet more rapidly otherwise.

In conclusion it was shown that an efficient choice of the proposal matrix in the generalized Metropolis method is $S(\mathbf{R}_f|\mathbf{R}_i) \sim \sqrt{f(\mathbf{R}_f)/\Omega(\mathbf{R}_f)}$. Large gains in efficiency were demonstrated in calculations of the electronic energy of atoms and molecules.

I thank Malvin Kalos for his encouragement and Malvin Kalos, Peter Nightingale, Karl Runge, and Kevin Schmidt for useful discussions. This work was supported by the Office of Naval Research. The computations were performed on IBM RS-6000 workstations provided by the Cornell National Supercomputer Facility and Cornell-IBM Computing for Scientific Research Joint Study.

[1] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. 58, 86 (1987).
[2] U. Wolff, Phys. Rev. Lett. 62, 361 (1989).
[3] See review by A. D. Sokal, in *Computer Simulation Studies in Condensed Matter Physics, Recent Developments*, edited by D. P. Landau, K. K. Mon, and H. B. Schüttler, Springer Proceedings in Physics Vol. 33 (Springer, Berlin, 1988).
[4] P. J. Reynolds, Int. J. Quant. Chem. Symp. 24, 679 (1990); G. G. Batrouni and P. J. Reynolds (to be published) have attacked this problem by a different method to that used here.
[5] N. Metropolis *et al.*, J. Chem. Phys. 21, 1087 (1953).
[6] W. K. Hatings, Biometrika 57, 97 (1970).
[7] D. Ceperley, G. V. Chester, and M. H. Kalos, Phys. Rev. B 16, 3081 (1977).
[8] M. Rao and B. J. Berne, J. Chem. Phys. 71, 129 (1979), introduced an adjustable parameter λ in their force-bias method. Although they do not describe their work in these terms, we note that they obtained better results for $\lambda = 1/2$ corresponding to $S(\mathbf{R}_f|\mathbf{R}_i) \sim \sqrt{f(\mathbf{R}_f)}$ than for $\lambda = 1$ corresponding to $S(\mathbf{R}_f|\mathbf{R}_i) \sim f(\mathbf{R}_f)$.
[9] C. J. Umrigar (unpublished).
[10] The terms *valence electron* or *core electron* are only used to indicate the instantaneous distance of an electron from the nearest nucleus. During the course of a Monte Carlo run a given electron will be in both the valence and core regions part of the time.
[11] C. J. Myers, C. J. Umrigar, J. P. Sethna, and J. D. Morgan, Phys. Rev. A 44, 5537 (1991).
[12] C. J. Umrigar, Int. J. Quant. Chem. Symp. 23, 217 (1989), and references therein.

PATH INTEGRAL MONTE CARLO: Problem

Charusita Chakravarty, IIT Delhi

1. Simple Harmonic Oscillator

For a single particle in a harmonic potential, implement the Fourier path integral Monte Carlo algorithm. A sample program to do this is also available. For a 1-dimensional harmonic oscillator with mass m and potential energy $V(x) = 0.5kx^2$, the analytical result is

$$\langle K \rangle = \langle V \rangle = \langle E \rangle / 2 = 0.25\hbar\omega \coth(0.5\beta\hbar\omega)$$

2. Convergence Strategies: Partial Averaging

The partial averaging algorithm can be implemented for complex potentials by replacing the potential along a path $V(x(u))$ by

$$V_{eff}(x(u) = V(x(u)) + 0.5\sigma^2(u)V''(x(u))$$

where $V''(x(u))$ is the second derivative of the potential and

$$\sigma^2(u) = \frac{u}{\beta m}(\beta\hbar - u) - \sum_{k=1}^{k_{max}} \sigma_k^2 \sin^2(\frac{k\pi u}{\beta\hbar})$$

For a multi-dimensional generalisation of this algorithm, see J. Chem. Phys., **85**, 4567 (1986).

Implement this procedure for the harmonic oscillator system to see the improvement in efficiency i.e. the reduction in the number of Fourier coefficients required for convergence. Remember that when deriving the expressions for the thermal average of the total and kinetic energy, the partial averaging correction term in the effective potential must be considered.

3. Incorporating exchange of identical particles in Path Integral Monte Carlo

Let the density matrix for N distinguishable particles be labelled by D and for N indistinguishable particles be labelled by I. Then

$$\rho_I(\mathbf{x}, \mathbf{x}'; \beta) = \frac{1}{N!} \sum_P \xi^P \rho_D(\mathbf{x}, P\mathbf{x}'; \beta) \tag{5}$$

where P is the parity of the permutation and ξ is +1 for bosons and -1 for fermions.The effect of permutation P on the elements of the vector $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ is denoted by $P\mathbf{x}$. ξ^P is always +1 for bosons but for fermions, ξ^P equals -1 and +1 for odd and even permutations respectively. This result can be used to show how the Fourier path integral approach for distinguishable particles can be adapted to incorporate exchange effects for bosons and fermions (J. Chem. Phys., **85**, 4567 (1986).

The difference between fermionic and bosonic systems can be illustrated by considering non-interacting quantum particles in a harmonic potential. For a 1-dimensional system with $\hbar = m = k = 1$ and a potential given by $V(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{N} \frac{1}{2} x_i^2$, the average total, $\langle E \rangle$, kinetic, $\langle K \rangle$, and potential, $\langle V \rangle$, energies are

$$\langle E \rangle / 2 = \langle K \rangle = \langle V \rangle = \sum_{l=1}^{N} \frac{l}{4} \coth \frac{l}{2T} - \xi \frac{N(N-1)}{8}$$

Use either the algorithm given in the above reference, or any other discretised path version of it, to construct a path integral Monte Carlo program for this model system.