SOLVING QUANTUM MANY-BODY PROBLEMS WITH RANDOM WALKS

David M. Ceperley

October 1, 1996

Contents

1 Random Walks **3**
1.1 The Metropolis Monte Carlo Method 4
1.2 Dynamical Monte Carlo 9

2 Variational Monte Carlo **11**
2.1 The Pair Product Trial Function 12
2.2 Computational Details of VMC 12
2.3 Optimization of Trial Functions 13
2.4 Problems with Variational Methods 15

3 Projector Monte Carlo **16**
3.1 Importance Sampling 17
3.2 The Fixed-Node and Fixed-Phase Method 19
3.3 Exact Fermion and Excited State Methods 20
3.4 Excited States: CFMC 23
3.5 General Features of the Projection Method 24

4 Lattice models **25**

5 Treatment of Atomic Cores in QMC **26**

6 Beyond the pair-product trial function **27**

Quantum Monte Carlo (QMC) methods have been developed over the past thirty years to calculate properties of quantum many-body systems. The motivation for developing simulation methods is basically the same as it is for classical systems. For the classical many-body problem, direct simulations have proved the only way to get thoroughly reliable information about many-body effects, particularly as the systems get more complex. Quantum systems reduce to classical systems in certain limits (e.g. at high temperature) hence if one needs simulation to do classical systems, one needs simulation to calculate the properties of quantum systems. Quantum simulations are more challenging than classical simulations because not only do we have the problems inherent in sampling a multi-dimensional space, also we do not have an analytic expression for the function to be sampled. The simulation has to accomplish both tasks.

For a few systems QMC has provided crucial information in understanding quantum systems or simply in providing needed numerical data. The correlation energy of the electron gas is maybe the best-known example in solid state physics. But except for these few cases, the hope of QMC is not yet fully realized mainly because of inadequate development of the methodology or of aggressive application of the existing methods. Fermion statistics and quantum dynamics remain a challenge to the practitioner of simulation techniques. Nonetheless the results are more than competitive with those from the other methods used for quantum systems. In addition, the method used by Monte Carlo to "solve" the quantum problem, (essentially we map the system to an equivalent classical system) provides new insights into the origin of properties of quantum systems.

There are two basic types of methods used to simulate quantum systems. In zero temperature methods (Variational Monte Carlo and Projector Monte Carlo) one calculates the properties of a single wavefunction. These methods are applicable when we need to calculate matrix elements like $\langle\phi|\mathcal{O}|\phi\rangle$. On the other hand, in finite temperature methods (Path Integral Monte Carlo) one takes a trace over the thermal density matrix: $\langle\mathcal{O}\exp(-\beta\mathcal{H})\rangle$. The equivalent to Molecular Dynamics (Quantum Molecular Dynamics, QMD) does not exist in any practical sense. In QMD one would take an arbitrary wave function and propagate it forward in time, then compute some expectation values. The difficulty is that the full wavefunction must be kept until it "collapses" with the final measurement. The amount of data needed grows exponentially with the number of particles. One is forced to either simulate very small systems (*i.e.* less than 5 particles) or to make very severe approximations such as assuming that the wavepacket remains localized.

Quantum Monte Carlo methods are exclusively examples of Markov processes or random walks which are discussed in the next section. The main example, the Metropolis method is appropriate when one wants to sample a known, computable function. If one had an exact analytic expression for the many-body wave function, it would then be straight forward to use this method to determine quantum expectation values for that state. However, such is not the case, and one is forced to resort to either more complicated, or more approximate, methods. Following that, I discuss variational Monte Carlo, a straightforward application of the Metropolis Monte Carlo method, the only complication being that the wavefunction for fermion systems is a determinant and the distribution itself must be optimized. In the following sections, I discuss the projector Monte Carlo methods, where the transition rules are set up so that the asymptotic population is the exact ground state wave function for a given Hamiltonian. They involve using branching random walks.

I have recently reviewed[1, 2] an even more powerful method, Path Integral Monte Carlo, which is yet another application of the Metropolis algorithm; space prohibits me from including that description in these notes. I will primarily discuss continuum models, not lattice models, although most of the techniques can be carried over directly. As examples, I will discuss applications of these methods to helium and electronic systems. More extensive discussion of QMC is to be found in refs. [3, 4, 5, 6, 7].

I will always assume that the system is a non-relativistic collection of N particles described by

the Hamiltonian:

$$\mathcal{H} = -\lambda \sum_{i=1}^{N} \nabla_i^2 + \sum_{i<j} v(r_{ij}), \tag{1}$$

where $\lambda = \hbar^2/2m$ and $v(r)$ is a two-body pair potential. A boson wave function is then totally symmetrical under particle exchange and a fermion function is antisymmetrical. The symbol R refers to the $3N$ set of particle coordinates, and (r_i, σ_i) to the 3 spatial and 1 spin coordinate of particle i. The exact eigenfunctions and eigenvalues of the Hamiltonian are written as: $(\phi_\alpha(R), E_\alpha)$. The trial wave function will be denoted $\Psi(R)$.

1 Random Walks

Let me start by reviewing random walk (Markov chains). The application of these ideas have lead to one of the most important and pervasive numerical algorithm to be used on computers: the *Metropolis* algorithm first used by Metropolis, Rosenbluth and Teller in 1953[8]. It is a general method of sampling arbitrary highly-dimensional probability distributions by taking a random walk through configuration space. Virtually all Quantum Monte Carlo simulations are done using either Markov sampling or a generalization of the Metropolis rejection algorithm.

The problem with direct (or independent) sampling methods is that their efficiency goes to zero as the dimensionality of the space increases. Suppose we want to sample the probability distribution:

$$\pi(s) = \frac{exp[-S(s)]}{Z}, \tag{2}$$

where $S(s)$ is called the *action* and s is a variable in the space to be sampled. For classical systems $S(s)$ would be equal to $\beta V(R)$, the classical Boltzmann distribution. The partition function Z normalizes the function π in its space and is usually not known. A direct sampling method would require sampling a function with a known normalization. Suppose we can directly sample a function $p_m(s) \approx \pi(s)$. One can show that the Monte Carlo variance of any quantity will depend on the ratio π/p_m as:

$$variance \propto \frac{<(\pi/p_m)^2>}{<\pi/p_m>^2} \tag{3}$$

(averages taken with respect to p_m.) As the number of degrees of freedom in the system increases, the variance of the direct sampling approach will grow exponentially.

Let us briefly review the properties of Markov chains. In a Markov chain, one changes the state of the system randomly according to a fixed *transition rule*, $\mathcal{P}(s \to s')$, thus generating a random walk through state space, $\{s_0, s_1, s_2 \ldots\}$. (The definition of a Markov process is that the next step is chosen from a probability distribution that depends only on the "present" position. This makes it very easy to describe mathematically.) The process is often called the drunkard's walk. $\mathcal{P}(s \to s')$ is a probability distribution so it satisfies

$$\sum_{s'} \mathcal{P}(s \to s') = 1 \tag{4}$$

and

$$\mathcal{P}(s \to s') \geq 0. \tag{5}$$

Let $f_n(s)$ be the probability distribution of the walker after s steps. Then it is easy to write the evolution of f_n in terms of $\mathcal{P}$.

$$f_{n+1}(s') = \sum_{s} f_n(s) \mathcal{P}(s \to s') \tag{6}$$

or in vector-matrix notation:

$$f_{n+1} = \mathcal{P} f_n = \mathcal{P}^n f_1. \tag{7}$$

If the transition probability is *ergodic*, the distribution f_n converges to a *unique equilibrium state*. That means there is a unique solution to:

$$\sum_s f(s)\mathcal{P}(s \to s') = f(s'). \tag{8}$$

The transition is ergodic if:

1. One can move from any state to any other state in a finite number of steps with a nonzero probability, *i.e.*, there are no barriers that restrict any walk to a subset of the full configuration space.

2. It is not periodic. An example of a periodic rule is if the hopping on a bipartite lattice always proceeds from the A sites to the B sites and vice-versa so that one never forgets which site one started on. Non-periodic rules hold if $\mathcal{P}(s \to s) > 0$; if there is always some chance of staying put.

3. The average return time to any state is finite. This is always true in a finite system (*e.g.* periodic boundary conditions). It would be violated in a model of the expanding universe where the system gets further and further from equilibrium because because there is no possibility of energy flowing between separated regions after the "big bang".

Under these conditions we can show that if $f_n(s)$ is the probability distribution of random walks after n steps, with $f_0(s)$ the initial condition, then:

$$f_n(s) = \pi + \sum_\lambda \epsilon_\lambda^n c_\lambda \phi_\lambda(s), \tag{9}$$

where the $\epsilon_\lambda < 1$. Hence the probability distribution converges exponentially fast to the stationary distribution π. Furthermore, the convergence is monotonic (it does not oscillate). Specifically, what we mean is that the distance between f_n and π is strictly decreasing: $|f_n - \pi| > |f_{n+1} - \pi|$.

The transition probabilities often satisfy the *detailed balance* property for same function: the transition rate from s to s' equals the reverse rate,

$$\pi(s)\mathcal{P}(s \to s') = \pi(s')\mathcal{P}(s' \to s). \tag{10}$$

If the pair $\{\pi(s), \mathcal{P}(s \to s')\}$ satisfy detailed balance and if $\mathcal{P}(s \to s')$ is ergodic, then the random walk must eventually have π as its equilibrium distribution. To prove this fact, sum the previous equation over s and use Eq.(4) to simplify the right-hand-side. Detailed balance is one way of making sure that we sample π; it is a sufficient condition. Some methods work directly with Eq. (8) as we will see.

1.1 The Metropolis Monte Carlo Method

The Metropolis (rejection) method is a particular way of ensuring that the transition rules satisfy detailed balance. It does this by splitting the transition probability into an "a priori" *sampling distribution* $T(s \to s')$ (a probability distribution that we can directly sample) and an *acceptance probability* $A(s \to s')$ where $0 \le A \le 1$.

$$\mathcal{P}(s \to s') = T(s \to s')A(s \to s'). \tag{11}$$

In the generalized Metropolis procedure, (Kalos and Whitlock, 1986), trial moves are accepted according to:

$$A(s \to s') = min\left[1, q(s' \to s)\right], \tag{12}$$

where

$$q(s \to s') = \frac{\pi(s')T(s' \to s)}{\pi(s)T(s \to s')}. \tag{13}$$

It is easy to verify detailed balance and hence asymptotic convergence with this procedure by looking at the 3 cases:

- $s = s'$ (trivial)
- $q(s \to s') \leq 1$
- $q(s \to s') \geq 1$

Two common errors are: first, if you can move from state s to s' then the reverse move must also be possible ($T(s \to s')$ and $T(s' \to s)$ should be zero or non-zero together) and second, moves that are not accepted are rejected and remain at the same location for at least one more step. Accepted or rejected steps contribute to averages in the same way.

Here is the generalized Metropolis algorithm:

1. Decide what distribution to sample ($\pi(s)$) and how to move from one state to another, $T(s \to s')$
2. Initialize the state, pick s_0.
3. To advance the state from s_n to s_{n+1}:
   - Sample s' from $T(s_n \to s')$
   - Calculate the ratio:
   $$q = \frac{\pi(s')T(s' \to s_n)}{\pi(s_n)T(s_n \to s')} \tag{14}$$
   - Accept or reject:
   If $q > 1$ or if $q > u_n$ where u_n is a uniformly distributed r.n. in $(0, 1)$ set $s_{n+1} = s'$.
   Otherwise set $s_{n+1} = s_n$
4. Throw away the first κ states as being out of equilibrium where κ is the "warm-up" time.
5. Collect averages every so often and block them to get error bars.

Consider the sampling of a classical Boltzmann distribution, $exp(-\beta V(s))$. In the original Metropolis procedure, $T(s \to s')$ was chosen to be a constant distribution inside a cube and zero outside. This is the *classic* rule: a single atom at a single time slice is displaced uniformly and the cube side Δ is adjusted to achieve 50% acceptance. Since T is a constant, it drops out of the acceptance formula. So the update rule is:

$$\mathbf{r}' = \mathbf{r} + (\mathbf{u} - \frac{1}{2})\Delta \tag{15}$$

with the acceptances based on $q = exp(-\beta(V(s') - V(s))$. Moves that lower the potential energy are always accepted. Moves that raise the potential energy are often accepted if the energy cost (relative to $k_B T = 1/\beta$) is small. Hence the random walk does not simply roll downhill. Thermal fluctuations can drive it uphill.

Some things to note about Metropolis:

- The acceptance ratio (number of successful moves/total number of trials) is a key quantity to keep track of and to quote. Clearly if the acceptance ratio is very small, one is doing a lot of work without moving through phase space. On the other hand, if the acceptance ratio is close to 1, you could probably use

  larger steps and get faster convergence. There is a rule-of-thumb that it should be 1/2, but in reality we have to look at the overall efficiency.

- One nice feature is that particles can be moved one at a time. Note that N steps of Metropolis takes the same amount of time as 1 step of Molecular Dynamics. Consider what would happen if we moved N hard spheres all together. Let p be the probability of getting an overlap (and hence rejection) in the move of one hard sphere. Then the probability of getting an acceptance with N hard spheres is $(1-p)^N = exp(N \ln(1-p))$ assuming no correlation. In order to get a reasonable acceptance ratio one would have to decrease δ so that $p \approx 1/N$ which would require extremely small steps.

- Note that we need both the forward probability and the reverse probability if one has a non-uniform transition probability. Also note that we cannot calculate the normalization of π nor is it ever needed. Only ratios enter in.

- One can show that the Metropolis acceptance formula is optimal among formulas of this kind which satisfy detailed balance.

- In some systems, it is necessary to have several different kinds of moves, for example, moves that change path variables and other moves that change the permutation. So it is necessary to generalize the Metropolis procedure to the case in which one has a *menu* of possible moves. There are two ways of implementing such a menu. The simplest is to choose the type of move randomly, according to some fixed probability. For example, one can choose the particle to be updated from some distribution. One must include in the definition of $T(s \to s')$ the probability of selecting that move from the menu (unless you can argue that it cancels out.) A more common procedure is to go through all possible atoms systematically. After one *pass*, moves of all coordinates have been attempted once. In this case, individual moves do not satisfy detailed balance but it is easy to show that composition of moves is valid as long as each type of move individually satisfies detailed balance. Having many types of moves makes the algorithm much more robust, since before doing a calculation one does not necessarily know which moves will lead to rapid movement through phase space.

Since asymptotic convergence is easy to guarantee, the main issue is whether configuration space is explored thoroughly in a reasonable amount of computer time. Let us define a measure of the convergence rate and of the efficiency of a given Markov process. This is needed to compare the efficiency of different transition rules, to estimate how long the runs should be, and to calculate statistical errors. The rate of convergence is a function of the property being calculated. Generally one expects that there are local properties which converge quickly and other properties (such as order parameters near a phase boundary) which converge very slowly.

Let $\mathcal{O}(s)$ be a given property and let its value at step k of the random walk be $\mathcal{O}_k$. Let the mean and intrinsic variance of $\mathcal{O}$ be denoted by

$$\bar{\mathcal{O}} = \langle \mathcal{O}_k \rangle \tag{16}$$

and

$$\sigma^2_{\mathcal{O}} = \langle (\mathcal{O}_k - \bar{\mathcal{O}})^2 \rangle \tag{17}$$

where the averages $\langle \ldots \rangle$ are over π. These quantities depend only on the distribution π, not on the Monte Carlo procedure. We can show that the standard error of the estimate of the average, $\bar{\mathcal{O}}$, over a Markov chain with P steps, is

$$error[\bar{\mathcal{O}}] = \sqrt{\frac{\kappa_{\mathcal{O}} \sigma_{\mathcal{O}}^2}{P}}. \tag{18}$$

The *correlation time*, $\kappa_{\mathcal{O}}$, defined as

$$\kappa_{\mathcal{O}} = 1 + 2 \sum_{k=1}^{\infty} \frac{\langle (\mathcal{O}_0 - \bar{\mathcal{O}})(\mathcal{O}_k - \bar{\mathcal{O}}) \rangle}{\sigma_{\mathcal{O}}^2}, \tag{19}$$

gives the average number of steps to decorrelate the property $\mathcal{O}$. The correlation time will depend crucially on the transition rule and has a minimum value of 1 if one can move so far in configuration space that successive values are uncorrelated. In general, the number of independent steps which contribute to reducing the error bar from Eq. (18) is not P but P/κ.

Hence to determine the true statistical error in a random walk, one needs to estimate the correlation time. To do this it is very important that the total length of the random walk be much greater than $\kappa_{\mathcal{O}}$. Otherwise the result and the error will be unreliable. Runs in which the number of steps is $P \gg \kappa_{\mathcal{O}}$ are called *well converged*. In general, there is no mathematically rigorous procedure to determine κ. Usually one must determine it from the random walk. It is a good practice occasionally to run very long runs to test that the results are well converged. Error bars can be conveniently determined by blocking the result over enough steps so that successive blocks are independently distributed.

The correlation time defined above is an equilibrium average. There is another correlation time relevant to Markov chains, namely, how many steps it takes to reach equilibrium from some starting state. Normally this will be at least as long as the equilibrium correlation time, but in some cases it can be much longer. The simplest way of testing convergence is to start the random walk from several, radically different, starting places and see if a variety of well-chosen properties converge to the same values. A starting place appropriate for a dense liquid or solid is with all the atoms sitting on lattice sites. However, it may take a very large number of steps for the initial solid to melt. Metastability and hysteresis are characteristic near a (first-order) phase boundary. A random starting place is with placing each variable randomly in the total space. It may be very difficult for the system to go to the equilibrium distribution from this starting place. More physical starting places are well-converged states at neighboring densities and temperatures.

The *efficiency* of a random-walk procedure (for the property $\mathcal{O}$) is defined as how quickly the errors bars decrease as a function of computer time,

$$\xi_{\mathcal{O}} = \frac{1}{\kappa_{\mathcal{O}} \sigma_{\mathcal{O}}^2 T}, \tag{20}$$

where T is the computer time per step. Hence the efficiency is independent of the length of the calculation and is the figure-of-merit for a given algorithm. The efficiency depends not only on the algorithm but also on the computer and the implementation. Methods that generate more steps per hour are, other things being equal, more efficient. We are fortunate to live in a time when the efficiency is increasing because of rapid advances in computers. Improvements in algorithms can also give rise to dramatic increases in efficiency. If we ignore how much computer time a move takes, an optimal transition rule is one which minimizes $\kappa_{\mathcal{O}}$, since $\sigma_{\mathcal{O}}^2$ is independent of the sampling algorithm.

There are advantages in defining an *intrinsic efficiency* of an algorithm since one does not necessarily want to determine the efficiency for each property separately. It is best to optimize an algorithm to compute a whole spectrum of properties. Diffusion of paths through phase space provides at least a intuitive measure of convergence. Let us define the *diffusion constant* D_R of an algorithm by

$$D_R = \left\langle \frac{[(R_{n+1} - R_n)]^2}{T} \right\rangle, \tag{21}$$

where $R_{n+1} - R_n$ is the total change in one Monte Carlo step and T is the CPU time per step. Note that this change is zero if a move is rejected. For the "classic" Metropolis procedure we see that the diffusion constant is roughly:

$$D_R \propto \langle A \rangle \Delta^2. \tag{22}$$

Hence one wants to increase Δ until the acceptance ratio starts decreasing too rapidly. This leads to an optimal choice for Δ. The values of these diffusion constants depend not only on the computer and the algorithm, but also on the physics. Diffusion of the atoms in a solid is much less than in a liquid, irrespective of the algorithm.

Usually transition rules are local; at a given step only a few coordinates are moved. If we try to move too many variables simultaneously, the move will almost certainly be rejected, leading to long correlation times. Given a transition rule, we define the *neighborhood*, $\mathcal{N}(s)$, for each point in state space as the set of states s' that can be reached in a single move from s. (It is essential for detailed balance that the neighborhoods be reflexive. If s' is in the neighborhood of s, then s is in the neighborhood of s'.) With the *heat-bath* transition rule, one samples elements from the neighborhood with a transition probability proportional to their equilibrium distribution,

$$T_{HB}(s \to s') = \frac{\pi_{s'}}{C_s}, \tag{23}$$

where the normalization constant is

$$C_s = \sum_{s'' \in \mathcal{N}(s)} \pi_{s''}. \tag{24}$$

Then one sees, by substitution into the acceptance probability formula, that the acceptance probability will be

$$A(s \to s') = \min\left[1, \frac{C_s}{C_{s'}}\right]. \tag{25}$$

If the neighborhood of s equals the neighborhood of s' then all moves will be accepted. For all transition rules with the same neighborhoods, the heat-bath rule will converge to the equilibrium distribution fastest and have the smallest correlation time. Within the neighborhood, with heat bath one comes into equilibrium within a single step.

This heat-bath rule is frequently used in lattice spin models where one can easily compute the normalization constant, C_s needed in the acceptance ratio formula and to perform the sampling. The heat-bath approach is not often used in continuum systems because the normalizations are difficult to compute; note that the integral in Eq. (24) extends over all space. In Monte Carlo on a classical system, the new atom could be anywhere in the box. One has to compute a one-particle partition function at each step. A repulsive potential will cut holes in the uniform distribution where another atom is present. Although it would be possible to develop sophisticated ways of sampling T_{HB}, it has been found more efficient to further approximate T_{HB} by some function that can be sampled quickly and let the Metropolis algorithm correct the sampling, since all that

matters in the end is the efficiency. For continuum systems the idea is to find a method close to the heat-bath rule, so that the correlation time is small, but with a transition rule which is able to be executed quickly.

1.2 Dynamical Monte Carlo

Let me introduce a different way of generating random walks, based on an evolution equation. In nature, equilibrium distributions are generated by an evolution process. The diffusion Monte Carlo algorithm and the classical simulation methods of Brownian dynamics and smart Monte Carlo are more naturally regarded as local dynamical random walks.

Suppose we want to sample the distribution $\exp(-\beta V(R))$. The Smoluchowski equation

$$-d\pi(R,t)/dt = -\nabla D(R)[\nabla\pi - \beta\mathbf{F}(R)\pi], \tag{26}$$

is the unique "master" equation which is:

- local in space
- goes to the Boltzmann distribution
- is Markovian.

Here $D(R)$ is, in general, a many-body tensor (usually taken to be a constant diagonal tensor) and $\mathbf{F} = -\nabla V$ is the force.

The asymptotic solution of $\pi(R,t)$ will be $\pi(R) \propto \exp(-\beta V(R))$. It is easy to see that this distribution satisfies $d\pi/dt = 0$. If we assume the process is ergodic, since it is also Markovian, this must be the only solution.

Let us define the Green's function: $G(R, R_0; t)$ as the solution to Eq. (26) with the boundary condition at zero time: $G(R, R_0; 0) = \delta(R - R_0)$. We can prove that the Green's function satisfies detailed balance:

$$\pi(R)G(R \to R'; t) = \pi(R')G(R' \to R; t), \tag{27}$$

for any value of t. (To do that one writes the evolution equation for the symmetrized Green's function: $(\pi(R)/\pi(R'))^{1/2}G(R \to R'; t)$, and sees the right hand side of the master equation is an Hermitian operator which implies that the symmetrized Green's function is symmetric in R and R'.) If G is used for a transition probability it will always give acceptances. Also it gives interesting dynamics (not MD but dynamics of viscous particles in contact with a heat and momentum bath).

The Smoluchowski equation leads to an interesting process. To use it we need to calculate G in the short-time limit. In the following I explain a general procedure for devising an algorithm for sampling G by calculating the moments of G,

$$I_n(R_0, t) = \int dR(R - R_0)^n G(R_0 \to R; t). \tag{28}$$

First take the time derivative of this equation, use the master equation on the r.h.s., and Green's theorem to get a simple integral over G on the r.h.s (we interpret this as an average $< \ldots >$). We assume there are no absorbing surfaces of the random walks. Then,

$$dI_0/dt = 0. \tag{29}$$

This implies the normalization of G is always one so the evolution describes a process which neither creates nor destroys walks. The next moment is:

$$dI_1/dt = < \beta D\mathbf{F} + \nabla D > . \tag{30}$$

Let us assume that $\mathbf{F}$ and ∇D are slowy varying. Then we can replace them by the values at the initial points and integrate:

$$< R_t >= R_0 + t[\beta \mathbf{F}(R_0) + \nabla D(R_0)] + \mathcal{O}[t^2]. \tag{31}$$

The equation for the second moment (in general a second rank tensor) is:

$$dI_2/dt = 2 < D > +2 < (R - R_0)(\beta \mathbf{F} + \nabla D) > . \tag{32}$$

Integrating,

$$< (R - R_0)^2 >= 2D(R_0)t + \mathcal{O}[t^2]. \tag{33}$$

According to the central limit theorem, Eqs. 31-33 are all that is needed to simulate the random walk if the time step t is sufficiently small. Hence we can construct a Gaussian distribution with the correct mean and covariance. The solution at small time is:

$$G_g(R, R_0; t) = \exp[-(R - R_t)(2D(R_0)t)^{-1}(R - R_t)][2\pi t det(D(R_0))]^{-1/2}. \tag{34}$$

We have not yet discussed the diffusion tensor. For simplicity, one normally assumes that $D(R) = D_0\mathcal{I}$ is a constant, unit tensor. In this case D_0 can be absorbed into the units of time. Physically more complicated tensors are related to "hydrodynamic" interactions and will lead to different dynamics but the same static properties.

Although the exact Green's function will obey detailed balance automatically, our approximate one for non-zero t will not. Its acceptance acceptance probability (for constant diffusion) is given by:

$$A = min\left[1, exp(-\beta(V(\mathbf{r}') - V(\mathbf{r})) - \beta(F(\mathbf{r}) + F(\mathbf{r}'))(2(\mathbf{r}' - \mathbf{r}) - \beta D(F' - F))/4)\right]. \tag{35}$$

The acceptance ratio goes to unity at small t.

One can find more accurate approximations to the exact Green's function by including off-diagonal components in the second moment. These correspond to coupling between the variables. We can choose for a transition probability the most general correlated Gaussian in $3n$ variables,

$$T_S(R) = \sqrt{(2\pi)^{3m} det(\mathbf{A})} e^{-(R-\bar{R})(2\mathbf{A})^{-1}(R-\bar{R})}, \tag{36}$$

where the 3×3 positive-definite covariance matrix $\mathbf{A}$ and the mean position vector $\bar{R}$ can be arbitrary. Suppose we solve equation (32) to one higher order in the time step. One obtains:

$$\mathbf{A} = 2\mathcal{I}t - t^2 \nabla\nabla V(R_0). \tag{37}$$

We can sample the multivariate Gaussian distribution as follows. One Cholesky-factorizes the covariance matrix as $\mathbf{A} = \mathbf{S}\mathbf{S}^T$, where $\mathbf{S}$ is an upper triangular matrix. Then if χ is a vector of Gaussian random numbers with zero mean and unit variance, $\mathbf{S}\chi + \bar{R}$ has the desired mean and variance. The diagonal divisors in the Cholesky decomposition of $\mathbf{A}$ are needed to find the actual value of $T(R \rightarrow R')$ and the acceptance probability for a move. The effect of interactions is to push the mean position of an atom away from its current position if other particles are nearby. Similarly, the covariance is changed by interactions with neighboring particles. In directions where the curvature of the potential is positive, the cage of surrounding atoms results in a narrower Gaussian's being sampled. It is likely that this transition probability will find the easy directions for the particles to move.

2 Variational Monte Carlo

We now turn to the simplest Quantum Monte Carlo method, Variational Monte Carlo (VMC). The VMC method was first used by McMillan[12] to calculate the ground state properties of liquid ^{4}He and then generalized to fermion systems by Ceperley *et al.*[13]. It is a relatively simple generalization from a classical Monte Carlo simulation to VMC.

The variational theorem which gives its name to VMC states that for Ψ a proper trial function, the expectation of the Hamiltonian with respect to the trial function is an upper bound to the exact ground state energy:

$$E_V = \frac{\int \Psi^*(R)\mathcal{H}\Psi(R)}{\int \Psi^*(R)\Psi(R)} \geq E_0. \tag{38}$$

In VMC we use the Metropolis algorithm to sample the distribution:

$$\pi(R) = \frac{\mid \Psi(R) \mid^2}{\int \mid \Psi(R) \mid^2}. \tag{39}$$

One can easily see that the variational energy is simply the average value of the *local residual energy* over this distribution,

$$E_V = \int \pi(R) E_L(R) = \langle E_L(R) \rangle_\pi \,, \tag{40}$$

where the local residual energy of Ψ is defined as:

$$E_L(R) = \Psi^{-1}\mathcal{H}\Psi(R). \tag{41}$$

Note that as the trial function approaches an exact eigenfunction, $\Psi \rightarrow \phi_\alpha$, the local residual energy approaches a constant, the energy eigenvalue, E_i. This leads to a very important and general property of VMC, the *zero variance property*: as we improve the trial function, the Monte Carlo estimate of the variational energy converges more rapidly with the number of steps in the random walk. Of course, in this limit the upper bound is also becoming closer to the true energy. (In fact the variance of the local energy is usually proportional to $E_V - E_0$.) It is because of the zero variance property that Quantum Monte Carlo calculations of energies can be much more precise than Monte Carlo calculations of classical systems. Fluctuations are only due to inaccuracies in the trial function.

Let me summarize the conditions that Ψ must satisfy to be useful for QMC.

1. $\mathcal{H}\Psi$ must be well defined everywhere. Hence both Ψ and $\nabla\Psi$ must be continuous wherever the potential is finite otherwise differentiating will give singular terms. One must be particularly careful at the edges of the periodic box and when two particles approach each other. Otherwise E_V could lie above or below the true energy.

2. The integrals $\int |\Psi|^2$, $\int \Psi^*\mathcal{H}\Psi$, and $\int |\Psi\mathcal{H}|^2$ should exist. The existence of these integrals should be demonstrated analytically. If the last integral is infinite the central limit theorem may not hold and hence error bars will not be meaningful. Again, examine the limit as two particles approach each other and at infinity.

3. Ψ must have the proper symmetry: $\Psi(R) = (-1)^P\Psi(PR)$ for fermions and the right behavior at the periodic boundaries. Since the boson ground state is the lowest of all symmetries, one will still satisfy the variational principle even with an unsymmetrical trial function. As an example, one can tie atoms to crystal lattice sites. Although this wavefunction does not have bose symmetry, it still gives an upper bound since the boson energy is lowest of all.

For a lattice spin model, only item 3 is applicable.

2.1 The Pair Product Trial Function

What should we use for a trial wave function? First note that the ground state of a real Hamiltonian (*i.e.* no magnetic fields) can always be made real and non-negative. This implies that the ground state has Bose symmetry. Consider a system interacting with a one-body (*e. g.* an external potential) and two-body potentials and suppose the potential is composed of repulsive interactions at short range, like between helium atoms. Then the wavefunction should vanish when any pair of atoms approaches each other. This is the motivation which led Bijl[15] to propose using a pair product wavefunction. Each factor should be similar to the solution of the two-body wavefunction.

The pair product trial function is the simplest generalization of the Slater determinant and the ubiquitous form for the trial function in QMC:

$$\Psi(R,\sigma) = \exp[-\sum_{i<j} u(r_{ij})] det[\theta_k(r_i,\sigma_i)], \tag{42}$$

where $\theta_k(r,\sigma)$ is the kth spin-orbital and $u(r)$ is the *pseudopotential* or pair-correlation factor. This function also goes by the name of a Jastrow[14] wave function. Closely related forms are the Gutzwiller function for a lattice, or the Laughlin function in the fractional quantum hall effect. Both $u(r)$ and $\theta_k(r,\sigma)$ are to be chosen by minimizing the variational energy or other quantity.

2.2 Computational Details of VMC

First, how do the particles move in VMC? On a lattice one can make a random hop of a particle or a spin flip. In the *classic* Metropolis procedure for a continuum system, one moves the particles one at a time by adding a random vector to a particle's coordinate, where the vector is either uniform inside of a cube or is a normally distributed random vector centered around the old position. The move for the ith particle is accepted with probability:

$$q(R \rightarrow R') = \mid \Psi(R')/\Psi(R) \mid^2 = \exp[-2\sum_{j\neq i}(u(\mathbf{r}'_i - \mathbf{r}_j) - u(\mathbf{r}_i - \mathbf{r}_j))] \mid \sum_k \theta_k(\mathbf{r}'_i) C_{ki} \mid^2, \tag{43}$$

C is equal to the cofactor matrix divided by the determinant. Remembering our linear algebra, the matrix, C, is also the transposed inverse to the Slater matrix defined as:

$$\sum_k \theta_k(r_i) C_{kj} = \delta_{jk}. \tag{44}$$

Now the evaluation of a general determinant takes $O(N^3)$ operations. The evaluation of the fermion part of the acceptance ratio will take only $O(N)$ operations if C is already calculated. So it pays to keep C current as particles are being moved. If a move is accepted, C needs to be updated[13] using the formula:

$$C'_{jk} = C_{jk} + [\delta_{ji} - b_j] C_{ik}/b_i \tag{45}$$

where $b_j = \sum_k \theta_k(\mathbf{r}') C_{ki}$. (Remember it is particle i which is being moved.) This takes $O(N^2)$ operations. Hence to attempt moves for all N particles (a pass) takes $O(N^3)$ operations.

The local energy, needed to evaluate the variational energy is calculated by applying the Hamiltonian to the trial function. We get:

$$E_L(R) = V(R) + \lambda \sum_i [\nabla_i^2 U - \sum_k \nabla_i^2 \theta_k(\mathbf{r}_i) C_{ki} - G_i^2], \tag{46}$$

where $G_i = -\nabla_i U + \sum_k \nabla_i \theta_k(\mathbf{r}_i) C_{ki}$, and $U = \sum u(\mathbf{r}_{ij})$. Thus the inverse matrix is also needed to determine the local energy. Very often the orbitals are taken to be exact solutions to an external one-body potential:

$$-\lambda \nabla^2 \theta_k(\mathbf{r}) = (e_k - v(\mathbf{r}))\theta_k(\mathbf{r}) \tag{47}$$

Then the term $\sum_i \nabla_i^2 \theta_k(\mathbf{r}_i) C_{ki} = (e_k - v(\mathbf{r}_i))\theta_k(\mathbf{r}_i)$ simplifies. Finally note that using Green's identity allows several alternative ways[13] of calculating the variational energy. While some of them are simpler and do not involve so many terms, for a sufficiently good trial function, the local energy estimator of Eq. (46) has the lowest variance. The transformed energy expressions give useful tests of the computer program and the convergence of the random walk. It is very important to test that the first and second derivatives of the trial function are computed correctly. This is a very common source of error. It is best to do this test automatically whenever the trial function form is changed. Calculation of the derivatives by numerical differentiation is not recommended because of the loss of numerical precision and the slowness of this approach.

2.3 Optimization of Trial Functions

Optimization of the parameters in a trial function is crucial for the success of VMC. Bad upper bounds do not give much physical information. Good trial functions will be needed in the Projector Monte Carlo method. First we must decide on what to optimize and then how to perform the optimization. There are several possibilities of the quantity to optimize and depending on the physical system, one or other of the criteria may be best.

- The variational energy: E_V. If the object of the calculation is to find the least upper bound one should minimize E_V. There is a general argument suggesting that the trial function with the lowest variational energy will maximize the efficiency of Projector Monte Carlo[23].
- The variance of the local energy: $\sigma^2 = \int |\mathcal{H}\Psi|^2 - E_V^2$. If we assume that every step on a QMC calculation is statistically uncorrelated with the others, then the variance of the average energy will equal σ^2/p where p is the number of steps. The minimization of σ^2 is statistically more robust than the variational energy because it is a positive definite quantity with zero as a minimum value. One can also minimize a linear combination of the variance and the variational energy.
- The overlap with the exact wave function: $\int \Psi\phi$. If we maximize the overlap we find the trial function closest to the exact wave function in the least squares sense. This is the preferred quantity to optimize if you want to calculate correlation functions, not just ground state energies since then the VMC correlation functions will be closest to the true correlation functions. Optimization of the overlap will involve a Projector Monte Carlo calculation to determine the change of the overlap with respect to the trial function so that it is not often used.

Let us now consider what properties the optimal pseudopotential, u^*, has. If we suppose we assume that the spin-orbits come from an exact solution of a one-body potential, then the local energy expression simplifies. In particular, examine the dominant terms in the local energy Eq. (46) as 2 particles are brought together. We get

$$E_L(R) = v(r) + 2\lambda\nabla^2 u(r) - 2\lambda(\nabla u(r))^2 + \ldots, \tag{48}$$

where r is the distance separating the particles. To keep the local residual energy finite the singularities in the kinetic energy must cancel the the singularities of the potential energy. We see that

$e^{-u(r)}$ will equal the solution to the 2-body Schroedinger equation. If two particles are sufficiently close together, other particles are irrelevant.

For He atoms interacting with a short range Lennard-Jones potential $4\epsilon(\sigma/r)^{12}$, the small distance behavior will be:

$$u(r) = ((2\epsilon\sigma^2)/(25\lambda))^{1/2}(\sigma/r)^5. \tag{49}$$

Charged particles have a different rule, they obey the *cusp condition*:

$$e_i e_j + 2(\lambda_i + \lambda_j)\left.\frac{u_{ij}}{dr_{ij}}\right|_0 = 0 \tag{50}$$

The effect of fluctuations caused by the kinetic energy is always to make the wavefunction smoother and less singular than the potential energy. Thus an r^{-12} potential becomes in the wave function r^{-5}. The r^{-1} coulomb potential becomes a constant.

To describe the long wavelength behavior of the optimal $u(r)$ one uses a description in terms of collective coordinates such as phonons, or plasmons. We can write the variational energy in Fourier space as:

$$E_V = E_F + \sum_k (S_k - 1)(v_k - \lambda k^2 u_k) \tag{51}$$

where E_F is the fermion energy in the absence of correlation, v_k and u_k are the fourier transforms of $v(r)$ and $u(r)$, and S_k is the static structure factor for a given $u(r)$. Minimizing E_V with respect to u_k and making the RPA assumption of how S_k depends on u_k: $S_k^{-1} = S_{0k}^{-1} + 2\rho u_k$ where ρ is the particle density and S_{0k} is the structure factor for uncorrelated fermions, we obtain[17] the optimal wavefunction at long wavelengths:

$$2\rho u_k = -\frac{1}{S_{0k}} + \left[\frac{1}{S_{0k}} + \frac{2\rho v_k}{\lambda k^2}\right]^{1/2}. \tag{52}$$

For a short-ranged potential, (e.g. liquid helium), v_k can be replaced by a constant and and we find the Reatto-Chester[18] form: $u(r) \propto r^{-2}$. But for a charged system, where $v_k \propto k^{-2}$, then $u(r) \propto r^{-1}$. Careful studies[17] have shown that this zero-parameter wavefunction is excellent for both the 2 and 3 dimensional electron gas where it satisfies both the optimal large r-behavior and the short-distance cusp condition.

The optimal wavefunction is long-ranged so that correlation extends beyond the edge of the simulation box. The ground state energy is little affected by this tail in the wave function because it is screened. On the other hand, response functions, such as the dielectric function or the static structure factor are crucially dependent on using the correct long-range properties. In order to maintain the upper bound property, the correlation function must be properly periodic in the simulation cell. For high accuracy results and physically correct properties in the long wavelength limit, the Ewald image method[9, 17] is needed to represent the correct long-range behavior of the optimal trial function.

For more complex systems, a purely Monte Carlo optimization method is needed to find a good trial function. The most direct method consists of running independent VMC runs using different variational parameters. One can fit the energies to a polynomial, performing more calculations near the predicted minimum and iterating until convergence in parameter space is attained. The difficulty with this direct approach is that close to the minimum the independent statistical errors will mask the variation with respect to the trial function parameters. This is because the derivative of the energy with respect to trial function parameters is very poorly calculated. Also, it is difficult to optimize in this way functions involving more than 3 variational parameters because so many independent runs are needed to cover the parameter space.

A correlated sampling method, known as reweighting[3, 13] is much more efficient. One samples a set of configurations $\{R_j\}$ (usually several thousand points at least) according to some distribution function, usually taken to be the square of the wavefunction for some initial trial function: $|\Psi_t(R;a_0)|^2$. Then the variational energy (or variance) for trial function nearby in parameter space can be calculated by using the same set of points:

$$E_v(a) = \frac{\sum_j w(R_j,a) E_L(R_j,a)}{\sum_j w(R_j,a)}, \tag{53}$$

where the weight factor $w(R) = |\Psi_T(R;a)/\Psi_T(R;a_0)|^2$ and the local energy is $E_L(R,a)$. The weight factors take into account that the distribution function changes as the variational parameters change. One then can use a minimizer to find the lowest variational energy or variance as a function of a keeping the points fixed. However there is an instability: if the parameters move too far away, the weights span too large of a range and the error bars of the energy become large. The number of effective points of a weighted sum is:

$$N_{eff} = (\sum w_j)^2 / \sum w_j^2. \tag{54}$$

If this becomes much smaller than the number of points, one must resample and generate some new points. When minimizing the variance, one can also simply neglect the weight factors. Using the reweighting method one can find the optimal value of wavefunction containing tens of parameters.

2.4 Problems with Variational Methods

The variational method is very powerful, and intuitively pleasing. One posits a form of the trial function and then obtains an upper bound. In contrast to other theoretical methods, no further approximations are made. The only restriction on the trial function is that the trial function be computable in a reasonable amount of time. There is no sign problem associated with fermi statistics in VMC. To be sure, the numerical work has to be done very carefully which means that convergence of the random walk has to be tested and dependence on system size needs to be understood.

One of the problems with VMC is that it favors simple states over more complicated states. As an example, consider the liquid-solid transition in helium at zero temperature. The solid wave function is simpler than the liquid wave function because in the solid the particles are localized so that the phase space that the atoms explore is much reduced. This biases the difference between the liquid and solid variational energies for the same type of trial function, (*e.g.* a pair product form) since the solid energy will be closer to the exact result than the liquid. Hence the transition density will be systematically lower than the experimental value. Another illustration is the calculation of the polarization energy of liquid ^{3}He. The wave function for fully polarized helium is simpler than for unpolarized helium because antisymmetry requirements are higher in the polarized phase so that the spin susceptibility computed at the pair product level has the wrong sign!

The optimization of trial functions for many-body systems is time consuming, particularly for complex trial functions. In a one component system (say the electron gas) one only has to optimize a single $u(r)$ function since the orbitals are determined by symmetry. By contrast in the H_2O molecule, one has 5 different 3-dimensional orbitals (some related to each other by symmetry) and a 6-dimensional correlation function ($u(\mathbf{r}_i,\mathbf{r}_j)$). Clearly it is quite painful to fully optimize all these functions! Here I am not speaking of the computer time, but of the human time to decide which terms to add, to program them and their derivatives in the VMC code. This allows an element of human bias into VMC; the VMC optimization is more likely to be stopped when the expected

Method		$G(R,R')$	*ref.*
Diffusion	*DMC*	$\exp[-\tau(\mathcal{H}-E_T)]$	[26, 27]
Green'sFunction	*GFMC*	$[1+\tau(\mathcal{H}-E_T)]^{-1}$	[24, 25]
Power	*PMC*	$[1-\tau(\mathcal{H}-E_T)]$	[29]

Table 1: The Green's functions for various projection methods. τ is the timestep, and E_T is the trial energy. They have all been normalized to be unity and to have the same derivative for $\tau = 0$.

result is obtained. The basis set problem is still plaguing quantum chemistry even at the SCF level where one only has 1-body orbitals. VMC shares this difficulty with basis sets as the problems get more complex.

Finally, the variational energy is insensitive to long range order. The energy is dominated by the local order (nearest neighbor correlation functions). If one is trying to compare the variational energy of a trial function with and without long range order, it is extremely important that both functions have the same short-range flexibility and both trial functions are equally optimized locally. Only if this is done, can one have any hope of saying anything about the long range order. The error in the variational energy is second order in the trial function, while any other property will be first order. Thus variational energies can be quite accurate while correlation functions are not very accurate.

As a consequence, the results typically reflect what was put into the trial function. Consider calculating the momentum distribution. Suppose the spin-orbitals have a Fermi surface. Then the momentum distribution of the pair product trial function will also have a Fermi surface although it will be renormalized. This does not imply that the true wave function has a sharp Fermi surface. Only for localized spin-orbitals will a gap appear.

3 Projector Monte Carlo

We now turn to a potentially more powerful method where a function of the Hamiltonian projects out the the ground state, hence the name, projector Monte Carlo. The nomenclature of the various quantum Monte Carlo methods is not at all standardized. Table I shows the operators that have been used as projectors, or Green's functions. For simplicity I will only discuss Diffusion Monte Carlo although most of what I say carries over immediately to the other projectors.

A sequence of trial functions is defined by repeatedly applying the projector, $G(R,R')$: to some initial state $\psi_0(R)$:

$$\psi_{n+1}(R) = e^{-\tau(\mathcal{H}-E_T)}\psi_n(R) = \int dR' G(R,R')\psi_n(R'). \tag{55}$$

The effect on the trial function of the Green's function is seen by expanding the trial function in the set of exact eigenfunctions ϕ_α of the Hamiltonian. The *nth* iterate is:

$$\psi_n(R) = \sum_\alpha \phi_\alpha(R)\langle\phi_\alpha|\psi_0\rangle e^{-n\tau(E_\alpha-E_T)}. \tag{56}$$

The Green's functions shown in the table will project out the state of lowest energy having a non-zero overlap with the initial trial function:

$$\lim_{n\to\infty}\psi_n(R) = \phi_0(R)\langle\phi_0|\psi_0\rangle e^{-n\tau(E_0-E_T)}. \tag{57}$$

The role of the *trial energy*, E_T is to keep the overall normalization of ψ_n fixed, which implies $E_T \approx E_0$. The *timestep*, τ, controls the rate of convergence to the ground state. It must be taken small to allow us to make accurate approximations to the Green's function.

Since the evaluation of the Green's function involves a $3N$ dimensional integral, if N is larger than two or three one must do the integral with Monte Carlo. The interpretation of Eq. (55) is very similar to the Markov chain we discussed earlier except than the normalization is not fixed. The probability of starting a random walk at R_1 is $\psi_0(R_1)$. For the moment let us discuss bosons where ψ_0 is non-negative. To sample $\psi_1(R)$, we choose moves from R_0 to R_1 from the Green's function $G(R_1, R_0)$. Projector MC is different than a Markov process in that the Green's function is non-normalized as we will see. It does describe a Markov process in ensemble space.

Trotter's theorem[22] says that in the limit of small time step we are allowed to consider the kinetic energy and potential energy terms independently. In the limit that the time step approaches zero, a coordinate space representation of the Green's function is:

$$< R|e^{-\tau(\mathcal{H}-E_T)}|R' >= (4\pi\lambda\tau)^{-3N/2} e^{-\frac{(R-R')^2}{4\lambda\tau}} e^{-\tau(V(R)-E_T)} \; + \; \mathcal{O}(\tau^2), \tag{58}$$

The iteration equation, Eq. (55) , has a simple interpretation in terms of branching random walks since the first factor is the Green's function for diffusion and the second is multiplication of the distribution by a positive scalar. Luckily both are non-negative so a probabilistic interpretation is possible. Such is not the case for arbitrary Hamiltonians. The branching process makes projector Monte Carlo differ from a Markov process: walks are allowed to split and to die.

The computer algorithm is quite simple: an ensemble of configurations is constructed with a Metropolis sampling procedure for $\psi_0(R)$. This is the *zeroth generation*, *i.e.* $n = 0$. The number of configurations is the *population* of the zeroth generation, P_0. Points in the next generation are constructed by sampling the Gaussian distribution in Eq. (58) and then branching. The number of copies of R' in the next generation is the integer part of

$$m = u + \exp[-\tau(V(R) - E_T)] \tag{59}$$

where u is a uniform random number in $(0, 1)$. If we average over u, we see that the average density of points so sampled is precisely equal to the Green's function. If the potential energy is less than the ground state energy, duplicate copies of the configuration may be generated. In succeeding generations, these walks propagate independently of each other. In places of high potential energy, random walks are terminated.

This procedure is a Markov process where the state of the walk in the nth generation is given by $\{P_n; R_1, R_2, \ldots R_{P_n}\}$. Hence it has a unique stationary distribution, constructed to be the ground state wave function. As a result the population (the number of walkers) fluctuates from step to step. It executes a undirected random walk and if uncontrolled, will either reach zero or go off to infinity. The trial energy, E_T, must be adjusted to keep the population within computationally acceptable limits. This is done by adjusting the trial energy. The smoothest way of doing this feedback is:

$$E_T = E_0 + \kappa \ln(P^*/P), \tag{60}$$

where P is the current population, P^* is the desired population, E_0 is the best guess of the ground state energy, and κ is a feedback parameter adjusted to be small as possible while achieving the goal of stabilizing the population around the target, P^*. If κ is too large, one can bias the distribution.

3.1 Importance Sampling

The algorithm as described above was first suggested by Fermi, and actually tried out in the first days of computing some forty years ago [30]. But it fails on many-body systems because the

potential is unbounded. For example, a coulomb potential can go to both positive and negative infinity as two charges approach each other. In the case that like charges come close together branching will kill the entire configuration. In the case of unlike charges, the configuration can branch into an unlimited number of copies and those copies will continue branching. Even with a bounded potential the method becomes very inefficient as the number of particles increases since the branching factor grows with the number of particles. Mathematically it is correct, but computationally unstable and inefficient.

There is a simple cure discovered by Kalos [25] for GFMC, and extended by Ceperley and Alder to Diffusion Monte Carlo and fermion systems[27]. *Importance sampling* multiplies the underlying probability distribution by a known, approximate solution called the *trial* or *guiding* function, $\Psi(R)$. Multiply Eq. (55) by Ψ, the trial function, and define $f_n(R) = \Psi(R)\psi_n(R)$. Then:

$$f_{n+1} = \Psi e^{-\tau(\mathcal{H}-E_T)}\psi_n = \int dR' \tilde{G}(R,R') f_n(R') \tag{61}$$

where $\tilde{G}(R,R') = \Psi^{-1} e^{-\tau(\mathcal{H}-E_T)}\Psi$ is the importance-sampled Green's function and the initial conditions are $f_0(R) = \Psi(R)\psi_0(R)$. It is easily shown by differentiating $\tilde{G}$ with respect to τ that it satisfies the evolution equation:

$$-\frac{\partial \tilde{G}(R,R_0;\tau)}{\partial \tau} = -\sum_i \lambda_i \nabla_i [\nabla_i \tilde{G} + 2\tilde{G}\nabla_i \ln(\Psi(R))] + [E_L(R) - E_T]\tilde{G}, \tag{62}$$

where $E_L(R)$ is the local-energy defined in the VMC section. As we discused earlier, Trotter's theorem[22] says that for short enough time steps each term on the right-hand side can be considered as an independent process in the random walk. The three terms on the right-hand side correspond to diffusion, drifting and branching. We have already discussed diffusion and branching. We just have to add drift to the previous algorithm.

The Diffusion Monte Carlo (DMC) algorithm is:

1. The ensemble is initialized by sampling from $\Psi^2(R)$ using VMC.

2. The points in the configuration are advanced in time as:

$$R_{n+1} = R_n + \chi + \lambda\tau\nabla \ln(\Psi(R_n)^2), \tag{63}$$

   where χ is a normally distributed $3N$ dimensional random vector with variance $2\lambda\tau$ and zero mean. The last term is the drift.

3. The number of copies of each configuration is the integer part of

$$\exp(-\tau(E_L(R_n) - E_T)) + u, \tag{64}$$

   where u is a uniformly distributed random number in $(0,1)$ and E_T is the current trial energy. As the trial function approaches the exact eigenfunction, the branching factor approaches unity; thus a sufficiently good trial function can control the branching.

4. The energy is calculated as the average value of the local energy: $E_0 = \langle E_L(R_n)\rangle$.

5. The trial energy is periodically adjusted to keep the population stable using Eq. (60).

6. To obtain ground state expectations of quantities other than the energy, one must correct the average over the DMC walk using the *mixed estimator*, $V_{mix} = \langle\phi_0|V|\Psi\rangle$, and the variational estimator [3]. For example the potential energy is calculated as:

$$\langle\phi_0|V|\phi_0\rangle \approx 2\langle\phi_0|V|\Psi\rangle - \langle\Psi|V|\Psi\rangle + \mathcal{O}([\phi_0 - \Psi]^2). \tag{65}$$

The first term on the LHS is the mixed estimator produced by the projector Monte Carlo, the second term the variational estimate. If the mixed estimator equals the variational estimator then the trial function has maximum overlap with the ground state. For a strictly positive quantity like the density or pair correlation function we can use another estimator:

$$\langle\phi_0|\rho|\phi_0\rangle \approx \frac{[\langle\phi_0|\rho|\Psi\rangle]^2}{\langle\Psi|\rho|\Psi\rangle} + \mathcal{O}([\phi_0 - \Psi]^2) \tag{66}$$

to keep the extrapolated density positive.

Note that repeated use of step 2 alone would generate a probability density proportional to Ψ^2, if we turn off the branching we recover VMC. One can substantially improve the DMC algorithm given above by using rejections to put in the exact detailed balance property as detailed in ref. [47]. Further recent improvements are described in ref[28].

In the GFMC algorithm introduced by Kalos there is no error resulting from taking a finite timestep which makes it very useful for performing precise energy calculations. Its essence is identical to the above algorithm. The new algorithmic features of GFMC are the introduction of intermediate points and the sampling of the value of the timestep. But the other features are very similar.

3.2 The Fixed-Node and Fixed-Phase Method

We have not discussed at all the problem posed by fermi statistics or complex-valued wavefunctions to the projector Monte Carlo method. First let us consider the difficulty in implementing the non-importance sampled algorithm. The initial condition $\phi_0(R)$ is not a probability distribution since a fermion trial function will have an equal volume of positive and negative regions (assuming it can be made real at all.) We must use the initial sign of the wave function as a weight for the random walk. That leads to an exact but slowly converging algorithm that we will discuss in the next subsection.

Importance sampling cures this defect of the initial condition since the initial distribution $|\Psi(R)|^2$ is positive, but the importanced sampled Green's function, $\tilde{G}(R, R')$ can be negative, if in a step from R to R' we have $\Psi(R)\Psi(R') < 0$. If we sample the absolute value of the Green's function, then sign can be used as a weight. The walk will count negatively towards averages until such a time as it recrosses the node. This leads to a growing statistical variance for all matrix elements. There is a simple way to avoid the sign: forbid moves in which the sign of the trial function changes. This is the fixed-node (FN) approximation[26].

In a diffusion process, forbidding node crossings puts a zero boundary condition on the evolution equation for the probability. This solves the wave equation with the boundary conditions that the solution vanish wherever the trial function vanishes. The calculated energy will be an upper bound to the exact ground state energy[31], in fact the best possible upper bound with the given boundary conditions. One is exactly solving the wave equation inside of all the nodal regions, but there is a mismatch of the derivative of the solution across the boundary. With the FN method, we do not necessarily have the exact fermion energy, but the results are much superior to those of VMC. No longer do we have to optimize two-body correlation factors, three-body terms etc., since the nodes

of the trial function are unchanged by those terms. The fixed-node method gets rid of the basis set problem to a large extent. One generally finds that the systematic error in the FN calculation is three to ten times smaller than it would be for a well-optimized VMC energy.

The nodes play a very important role since, as we have seen, if the nodes were exactly known, the many-fermion system could be treated by Monte Carlo methods without approximation. Let me briefly recap a few basic facts about nodal surfaces. First note that the ground state wave function can be chosen real in the absence of magnetic fields; the nodes are the set of points where $\phi(R) = 0$. Since this is a single equation, the nodes are in general a $3N-1$ dimensional hypersurface. (A common confusion is between these many-body nodes and those of the spin-orbits which are 2D surfaces in a 3D space.) When any two particles with the same spin are at the same location the wave function vanishes. These coincident planes, with $\mathbf{r}_i = \mathbf{r}_j$ are $3N-3$ dimensional hypersurfaces. In 3D space they do not exhaust the nodes, but are a sort of scaffolding. The situation is very different in 1D where the set of nodes is usually equal to the set of coincident hyperplanes. Fermions in 1D are equivalent to 1D bosons with a no-exchange rule.

Nodal volumes of ground state wave functions possess a tiling property[32]. To define this property first pick a point, R_0, which does not lie on the nodes. Consider the set of points which can be reached from R_0 by a continuous path with $\phi(R) \neq 0$. This is the volume in phase space accessible to a fixed-node random walk starting at R_0. Now consider mapping this volume with the permutation operator (only permute like spins), *i. e.* relabel the particles. The tiling theorem says that this procedure completely fills phase space, except, of course, for the nodes. Thus one does not have to worry about where the random walk started; all starting places are equivalent. This theorem applies for any fermion wave function which is the ground state for some local Hamiltonian. Excited states, ground states of non-local Hamiltonians, or arbitrary antisymmetric functions need not have the tiling property. More extensive discussion of fermion nodes and some pictures of cross-sections of free particle nodes are given in ref. [32].

Let us now consider systems for which the wavefunction is necessarily complex-valued. Two examples are when there is a strong magnetic field and when one wants to work in a state of fixed non-zero linear or angular momentum such as a vortex. The generalization of the VMC method is straightforward in principle: one simply samples the square of the modulus of the wavefunction. See, for example ref. [43]. One complication is in finding good wavefunctions, particularly in periodic boundary conditions since now the phase of the wavefunction can be periodic or more generally quasi-periodic. For the projector MC methods, the fixed-node method can be generalized to the ***fixed-phase*** method. Here the phase is specified by a variational wavefunction such as HF and the modulus is exactly solved for using the Diffusion Monte Carlo method. All that needs to be changed (over the zero field situation) is to add an additional term to the potential energy equal to:

$$V_{eff}(R) = V(R) + \lambda \sum_{i=1}^{n} [\nabla_i \phi(R) + \mathbf{A}(\mathbf{r}_i)]^2 \tag{67}$$

where $\phi(R)$ is the phase and $\mathbf{A}$ the vector potential. If the phase is exact, the exact energy is obtained even if the trial modulus was not exact. Otherwise, the best upper bound over all functions with that phase is found. Applications to quantum Hall systems are discussed in ref. [44]. An application to a vortex in superfluid helium is discussed in ref. [45]

3.3 Exact Fermion and Excited State Methods

As accurate as the FN method might be, it is still unsatisfactory since one does not know how the assumed nodal structure will affect the final result. One might guess that long-range properties, such as the existence or non-existence of a fermi surface will be determined by the assumed nodes.

The FN algorithm only improves the bosonic correlations of the trial function, and may not change the genuine fermion features. There are some fairly simple ways of improving on the FN method, but their use is limited to small systems, though by small it may be possible to do rather accurate "exact calculations" of fifty or more particles. We explain this limitation below.

The transient estimate (TE) method calculates the ratio:

$$E_{TE}(t) = \frac{\int \Psi \mathcal{H} e^{-t(\mathcal{H}-E_T)}\Psi}{\int \Psi e^{-t(\mathcal{H}-E_T)}\Psi} \tag{68}$$

where $\mathcal{H}$ is the exact Hamiltonian (not the fixed-node Hamiltonian) and Ψ is an antisymmetric trial function. Clearly the variational theorem applies so that $E_{TE}(t) \geq E_0$. In fact you can show that the energy converges exponentially fast and monotonically to the exact energy as a function of projection time:

$$\lim_{t\to\infty} E_{TE}(t) = E_0 + \mathcal{O}(e^{-tE_g}). \tag{69}$$

Here E_g is the gap to the next excited state with the same quantum numbers as the fermion ground state. In a Fermi liquid, this is the gap to the state with the same momentum, parity and spin obtained by making 2 particle-hole excitations.

Since we do not know precisely where the nodes are, the walks must be able to go everywhere in configuration space and so the drift term in Eq. (62) must not diverge at the nodes. Hence we must distinguish between the antisymmetric trial function that is used to calculate the energy, $\Psi(R)$, (this is always assumed to be our best variational function) and a strictly positive guide function, $\Psi_G(R)$, used to guide the walks. The guide function appears in the drift and branching terms of Eq. (62) and will be assumed to be a reasonable boson ground state trial function, while the trial function appears in Eq. (68). The Ψ_G importance-sampled Green's function is:

$$\tilde{G}(R,R';t) = \Psi_G(R) < R|e^{-t(\mathcal{H}-E_T)}|R' > \Psi_G^{-1}(R'), \tag{70}$$

and we can rewrite Eq. (68) as:

$$E_{TE}(t) = \frac{\int \sigma(R) E_{LT}(R) \tilde{G}(R,R';t)\sigma(R')\Psi_G^2(R')}{\int \sigma(R)\tilde{G}(R,R';t)\sigma(R')\Psi_G^2(R')} \tag{71}$$

where $\sigma(R) = \Psi(R)/\Psi_G(R)$ and $E_{LT}(R)$ is the local energy of Ψ. In the limit, $\Psi_G \to |\Psi|$, $\sigma(R)$ equals the sign of the trial function at the point R.

The transient estimate algorithm is:

1. Sample configuration R' from the square of the guide function with VMC. That corresponds to the rightmost factor in Eq. (71).

2. Record the initial weight of the walk, $\sigma(R')$.

3. Propagate the walk forward an amount of time, t with the Green's function, $\tilde{G}(R,R';t)$ by taking many sufficiently many time steps. If a branch occurs, each branch will count separately.

4. The total contribution of the walk arriving at R is $\sigma(R)\sigma(R')$. The energy at projection time t is:

$$E_{TE}(t) = \frac{< [E_{LT}(R) + E_{LT}(R')]\sigma(R)\sigma(R') >}{2 < \sigma(R)\sigma(R') >}, \tag{72}$$

where the averages are over all random walks generated by this process.

We see that the contribution of the walk is positive if the walk crosses an even number of nodes (or does not cross at all) and is negative if it crosses once or an odd number of times. This is how the nodes of the true wave function can differ from those of the trial function because of an unequal diffusion of walks from the negative to positive regions at particular points on the nodes.

The release node (RN) algorithm[31, 33] is an improvement on this TE method for ground state fermion calculations. Instead of starting the projection from the trial function, one begins the projection from the fixed-node solution. There are several advantages. First, boson correlation within the fixed-nodes is already optimized, thus the projection time is only determined by the time to adjust the position of the nodes. Second, one can directly calculate the difference between the exact result and the fixed-node solution. It turns out that this is given by the local energy of walks as they cross the nodes. Thus the difference is obtained with more statistical accuracy than either energy alone which allows the convergence to be carefully monitored. Finally, the release node method can be conveniently integrated into a fixed-node program. The only modifications are to introduce a guide function, and to keep track of the energy as a function of time since nodal crossing.

However, there are serious problems with both the TE and RN method. Let us examine how the statistical error of Eq. (68) depends on the projection time. It is not hard to see that the value of both the numerator and denominator are asymptotically proportional to $\exp(-t(E_F - E_T))$. Thus to keep the normalization fixed the trial energy must be equal to E_F. But, because the guide function allows the walks to cross the nodes, the population will increase as $\exp(-t(E_B - E_T))$ where E_B is the boson energy. From this, one can demonstrate that the signal-to-noise ratio vanishes exponentially fast. This is a general result. In any fermion scheme, as soon as negative weights are introduced the statistical error will grow as:

$$\epsilon_{stat} \propto e^{-t(E_F - E_B)}. \tag{73}$$

The behavior is physically easy to understand. Our estimator depends on finding differences between random walks crossing an even or an odd number of times. As soon as there is substantial mixing, the difference becomes harder and harder to see. Note that the exponential growth rate depends on a total energy difference. This implies that the transient estimate algorithm is guaranteed to fail if N is sufficiently large; the statistical errors will be too large. Nonetheless reliable results have been obtained for systems of 54 fermions.

The convergence problem is actually a bit more subtle since the projection time, t, can be optimized. The projection time should be chosen to give approximately equal statistical errors and systematic errors coming from non-convergence of the projection. Taking these errors from eqs. (69,73) we find the total error will decrease as:

$$\epsilon \quad \propto \quad P^{-\eta} \qquad \eta = \frac{E_g}{2(E_F - E_B + E_g)}. \tag{74}$$

where P is the total number of steps in the random walk. Only for bosons will $\eta = 1/2$. Any excited state will converge at a slower rate. Note that $\eta \propto 1/N$ for a fermion system. Inverting this relation, we find that the computer time needed to achieve a given error will increase exponentially with N.

One possibility for improving this convergence is to use all of the information given in the function, $E_{TE}(t)$, rather then just the value of the energy at the largest time. Crudely speaking, we can fit this function with a sum of exponentials and thereby try to extract the asymptotic limit. This "inverse Laplace transform" problem is well-known to be numerically unstable. It has been suggested[34] in the context of Quantum Monte Carlo for lattice models that the proper way to

perform such a function fit is with the maximum entropy statistical method, wherein a model of the expected density of states is used to bias the result, thereby regularizing the fitting problem. We[35] have applied these ideas to the TE and RN methods on simple problems and shown that they do indeed reduce the statistical and systematic errors.

There have been many attempts to "solve" the fermion sign problem. For example, one can try to pair positive and negative random walks in the TE method. This is difficult in many dimensions simply because the volume of phase space is so large than random walks rarely approach each other and no such schemes have yet succeeded for more than a few particles.

There is some confusion about the nature of the "fermion" or "sign" problem. Both the TE and RN methods do converge to the exact fermion energy. A proper statement of the fermion sign problem is in terms of complexity: how long does it take to achieve a given error estimate and how does the computer time scale with the number of fermions? In the TE method, the computer time to reach a given precision grows exponentially with the number of fermions. I would say that a solution of the fermion problem would be an approximation free algorithm which scales as some low power of the number of fermions.

Properties of classical systems can be simulated in time $\mathcal{O}(N)$. Simulations of equilibrium properties of quantum bosons at zero or non-zero temperature are also $\mathcal{O}(N)$. A Heisenberg model on a bipartite lattice, or any 1D fermion system is $\mathcal{O}(N)$. VMC calculations of fermion systems are $\mathcal{O}(N^3)$ in general, but the exponent would be smaller if localized spin-orbits are used. The Hubbard model at half filling on a bipartite lattice[39] is $\mathcal{O}(N^3)$ using the projection Monte Carlo method and auxiliary field techniques. This is the only non-trivial fermion problem solved. Known algorithms for general fermion systems are $\mathcal{O}(e^{\kappa N})$. Barring a breakthrough, one can still reduce the rate of exponential growth, κ, or use the TE or RN methods to gain confidence in FN and VMC calculations of much larger systems.

3.4 Excited States: CFMC

We have discussed to this point only calculations of ground state properties, or more correctly ground states with a given symmetry. Those are the states where one can use VMC and the FNA. If we try to apply either VMC or FN-DMC to excited states, one must always keep the wavefunction orthogonal to lower states. In the case of MC calculations, the lower states may not be explicitly known. However the MacDonald theorem provides a way both to minimize the energy and to keep all states orthogonal. It is a generalization of the TE method that is capable of calculating a spectrum of excited state properties from a single random walk.

The Correlation Function Quantum Monte Carlo method (CFMC) introduced by Ceperley and Bernu [36] starts with a basis of m trial functions, hopefully having a strong overlap with the first m exact eigenfunctions. MacDonald's theorem says that in any finite basis, if both the Hamiltonian and the overlap matrices are simultaneously diagonalized the resulting nth eigenvalue is an upper bound to the nth exact state. This is commonly used in SCF method in chemistry, for example. In QMC we use the Green's function $exp(-t\mathcal{H})$ to make the original basis much better.

$$\tilde{f}_i(t) = exp(-t\mathcal{H}) f_i \tag{75}$$

One then uses a single DMC trajectory to calculate the Hamiltonian and overlap matrices of the projected basis as a function of projection time. The name arises because all of the excited state properties are determined from the correlation function of the basis sets along the imaginary time trajectory. At zero projection time what enters are equal time correlation; the method is a generalization of the variational Monte Carlo method.

Having obtained these matrices, one solves the generalized eigenvalue problem for these matrices. It can be shown that the resulting eigenvalues approach the exact eigenvalues from above, monotonically and exponentially fast. However the statistical noise is also growing exponentially fast, just as it would for the TE method (of which CFMC is a generalization.) The method has been applied to small molecules [37] and the excitations of the two dimensional electron gas[38]. It is possible to use the excited state CFMC method to get exact results for systems in strong magnetic fields or for when the wavefunction is complex-valued.

3.5 General Features of the Projection Method

Let me briefly summarize some of the strengths and weakness of the projection method. The fixed-node result is guaranteed to be closer to the exact answer than the starting variational trial function. Since the FN algorithm automatically includes bosonic correlation, the results are much less likely to have a systematic bias than does VMC results. There is also the possibility of new physics coming out of the simulation. For example, one may observe a particular type of correlation completely absent from the trial function. It is always good to pay close attention to correlation functions computed by DMC since they are a good way of monitoring the quality of a trial function. But DMC is slower than VMC because the timestep needs to be smaller to have a good approximation to the Green's function. The cost in computer time is typically a factor of 2 to 10.

Although the projected probability distribution converges to the exact answer mathematically, in practice, this does not always occur in a finite length simulation of a many-body systems. The situation is similar to that of a classical simulation near a phase boundary. Metastable states exist and can have a very long lifetime. In addition the importance sampling biases the result. If the trial function describes a localized solid, even after complete convergence, the correlation functions will show solid-like behavior. Careful observation will reveal liquid-like fluctuations indicating the presence of the other state. The ability to perform simulations in a metastable state is useful but the results must be interpreted with caution. Although the fixed-node approximation dramatically improves energies, other properties, such as the momentum distribution may not be improved. To explore the metal-insulator phase transition with FN-DMC, one must come up with a sequence of nodes spanning the transition and use the upper bound property of the fixed-node approximation. In both VMC and DMC there is a premium for good trial functions; that is the most straightforward way of making progress in solving the many-fermion problem.

Importance sampling is only a partial cure to the unbounded fluctuations of the branching method. As N increases, sooner or later the branching becomes uncontrollable. Most projector Monte Carlo calculations have fewer than several hundred fermions. The finite temperature Path Integral Monte Carlo based on the Metropolis method does not suffer from the problem of uncontrolled branching. Release node calculations only improve the nodes locally. If t is the release node projection time, then we can move the nodes a distance of at most $\sqrt{6N\lambda t}$. One expects that global properties of the nodes will take a much longer time to be fixed by the projection operator.

The projector methods can only calculate energies exactly. For all other properties one must extrapolate out the effect of the importance sampling. This is a real problem if one is interested in obtaining asymptotic behavior of correlation functions. There are ways of getting around some of these problems but none are totally satisfactory. The Path Integral finite temperature methods are much superior to Projector Monte Carlo for calculating correlation functions.

4 Lattice models

Most of the methods that we have described here work also for lattice models, for example VMC and importanced-sample projector Monte Carlo. One important difference is that it is convenient to use the power Green's function to project out the ground state because the energy spectrum is bounded from both above and below. In this method, the Hamiltonian is directly used to hop the spins without time step error. The time step must be chosen to obey:

$$\tau \leq \frac{2}{E_{max} - E_T} \tag{76}$$

where E_{max} is the maximum energy. (We need to specify these to avoid negative elements in the Green's function.) Since the maximum energy is proportional to the number of sites, $\tau \propto 1/N$. This is normal since after N time steps, all spins on the average will be updated, just like in a classical Monte Carlo of a lattice model. Importance sampling enters in the same way. Details can be found in ref. [29].

The conditions on a lattice model not to have a sign problem are easy to state. The importanced sampled Green's function must be non-negative so it can be interpreted as a probability. This implies that the off-diagonal elements of the Hamiltonian should be non-positive.

$$< s|\mathcal{H}|s' > \leq \ 0 \quad \forall \ \ s \neq s'. \tag{77}$$

(We will discuss a more general relation in a moment). Of course, we can choose to do the random walk in any convenient basis, so the question becomes: is there any local basis that can be shown to satisfy the above inequalities? The exact eigenfunction basis satisfies these conditions but we do not know how to transform into that basis unless the eigenfunctions are known. Anyhow the eigenfunction basis is non-local and would not scale very well with the number of lattice sites.

The FN approximation is different for a lattice model because random walks can directly pass from one nodal region to the other without crossing a place where the trial function vanishes. A walker could pick up an unwanted minus sign if there are two many-body configurations (s, s'), with $\langle s|\mathcal{H}|s'\rangle\Psi(s)\Psi(s') > 0$. These are called sign-flip hops. Recently ten Haff *et al.* [41] have shown that for a lattice model, it is possible to modify the Hamiltonian is such a way that the fixed-node energy is an upper bound for a lattice model. The sign-flip matrix elements are set to zero and an extra potential is added to the diagonal:

$$V_{eff}(s) = V(s) + \sum_{s' \in SF} < s|\mathcal{H}|s' > \Psi(s')/\Psi(s) \tag{78}$$

This is the generalization of the continuum fixed-node method to an arbitrary lattice model. It has the properties that it gives a lower energy than the variational method but still an upper bound to the ground state energy. Hence, it gives the exact answer if Ψ is exact. However in contrast to the continuum, the magnitude of the wavefunction, not just its sign, enters. Going to a lattice does not at all change the TE and RN methods. They are useful ways of estimating the fixed-node approximation for a lattice model.

What I have not discussed are the most common methods for performing simulations of lattice models. These are based on applying the Stratonovitch-Hubbard transformation[39] to $e^{-t\mathcal{H}}$. An auxiliary field is introduced in place of the electron-electron interaction. Except for the case of the half-filled Hubbard model on a bi-partite lattice, sign problems remain. Recently there have been some developments of fixed-node methods for lattice models with in the auxiliary field treatment using a method known as Constrained Path Monte Carlo[42]. These methods are useful because the lattice Hamiltonian is bounded. Auxiliary field methods have not yet been successful for off-lattice problems.

5 Treatment of Atomic Cores in QMC

For QMC methods to have broad application in condensed matter physics, methods for heavy atoms are needed. The core electrons pose a problem for QMC methods because the core energy is much larger than chemical energies and the relevant distance scale of core states is much smaller. The scaling of computer time grows roughly as Z^6 with the atomic number, Z. Obviously, all-electron calculations quickly become intractable (at least to reach a fixed accuracy on the energy) as Z increases.

The core electrons create several basic problems. The first one is that the very small size of the core region requires a different strategy for sampling the core region otherwise the time step that controls the movement of electrons will scale as Z^{-2}. Although this might be technically difficult it is not the main obstacle. One can modify the propagator[28] so that it reflects the strong localization of the core charge and thus to a large extent avoid substantial slowing down of the simulations.

Far more severe are the local energy fluctuations caused by the strong potentials and large kinetic energies in the core. Because of a rapidly changing density it is very difficult (although, perhaps, not impossible) to design a trial function which can decrease these fluctuations. Even though correlation is relatively less important in the core, on the absolute scale it is still very large. The core, because of the high density, large potentials and large kinetic energy, is always the strongest fluctuating term of the local energy. Fortunately, for most valence properties the core remains practically inert and has a negligible impact on the valence properties. This fact can be used to eliminate the core electrons from the calculations and replace them with effective core Hamiltonians. Finally there is the additional expense of carrying along extra electrons not to speak of their relativistic effects.

In LDA calculations, pseudopotentials (or effective core potentials) are almost always used to increase the efficiency of calculations, even for calculations involving hydrogen! This allows smoother wave functions which in turn reduces the number of basis functions. It has been found that transferability (the ability of a pseudo-atom to mimic a full-core atom) is governed by norm conservation, and pseudopotentials are constructed so that the pseudo-orbitals match the full-core orbitals outside the core.

Bachelet *et al.* [48], in the *pseudo-Hamiltonian* approach, proposed to replace the action of the core on the valence states by an effective single electron Hamiltonian. The most general one-electron Hamiltonian which is local, spherically symmetric and Hermitian, has a local effective ionic potential and a spatially varying radial and tangential mass. Outside the atomic cores the potential becomes Coulombic and the mass becomes the usual scalar constant mass. The freedom in the effective ionic potential, the tangential and the radial mass can be used to tune the pseudo-Hamiltonian to mimic the action of the core electrons on the valence electrons. The approach has a great advantage in that the resulting valence Hamiltonian is local and all the technology of the DMC method immediately apply. For example, the FNA gives an upper bound and release-node calculations can then converge to the exact answer. Calculations on silicon have demonstrated the practicality and accuracy of this approach [49]. The disadvantage of the pseudo-Hamiltonian is that one does not have very much flexibility in matching the core response to valence electrons with different angular momentum because the restrictions on the mass tensor are too severe, especially for first row and transition metal atoms, *i.e.* for the atoms with strong nonlocalities.

The usual form of a valence-only Hamiltonian is:

$$H_{val} = H_{loc} + W \tag{79}$$

with the local part given by:

$$H_{loc} = \sum_i \left[-\frac{1}{2}\nabla_i^2 + \sum_I v_{loc}(r_{iI}) + \frac{1}{2}\sum_{j\neq i} \frac{1}{r_{ij}} \right] . \tag{80}$$

The nonlocal pseudopotential operator W includes pseudopotentials $v_\ell(r)$ for a small number of the lowest symmetry channels labeled by ℓ (usually spd)

$$\langle R|W|R'\rangle = \sum_{I,i}\sum_\ell \frac{2\ell+1}{4\pi} v_\ell(r_{iI}) \frac{\delta(r_{iI}-r'_{iI})}{r_{iI}r'_{iI}} P_\ell(\hat{\mathbf{r}}_{iI}\cdot\hat{\mathbf{r}}'_{iI}) \tag{81}$$

where P_ℓ is the Legendre polynomial. Therefore the valence states of different symmetry experience different potentials in the core region. The variational Monte Carlo can accommodate such Hamiltonians without major problems. Fahy *et al* [32, 52] used nonlocal pseudopotentials for the first VMC simulations of solids.

The nonlocality, however, is a problem for the DMC simulations because the matrix element for the evolution of the imaginary-time diffusion is not necessarily positive. For realistic pseudopotentials the matrix elements are indeed negative and thus create a sign problem (even for one electron) with consequences similar to those of the fermion sign problem.

In order to circumvent this problem it was proposed by Hurley and Christiansen[50] and by Hammond*et al.* [51] to define a new transformed effective core potential by a projection onto a trial function

$$V_{eff}(R) = \Psi_T^{-1}(R) \int dR' \langle R|W|R'\rangle \Psi_T(R) \tag{82}$$

The new effective potential is explicitly many-body but local and depends on the trial function. However, the DMC energy with V_{eff} will not necessarily be above the true eigenvalue of the original H_{val} and will depend on the quality of $\Psi_T(R)$.

A number of VMC and DMC calculations of atomic, molecular and solid systems have been carried out by this approach. This includes sp and transition element atoms [52], silicon and carbon clusters [54, 55], nitrogen solids [53]. Our experience indicates that with sufficient number of valence electrons one can achieve a high *final* accuracy. This, however, requires using $3s$ and $3p$ in the valence space for the $3d$ elements and, possibly, $2s$ and $2p$ states for elements such as Na. Once the core is sufficiently small, the systematic error of the fixed node approximation is larger than the systematic error from pseudopotentials and their subsequent projection in the DMC algorithm. Recent reviews of applications of QMC to chemistry are in refs. [46, 6]. A recent book on the subject is ref. [7].

6 Beyond the pair-product trial function

Relatively little has been done to take the variational results beyond the two-body level. The possibilities for improving the pair-product trial function in a homogeneous one-component system are relatively limited. I will describe several of the recent directions.

The dominant term missing in the trial function for a bosonic system is a three- body (or polarization) term with the functional form of a squared force:

$$U_3(R) = -\sum_i \left[\sum_j \xi(r_{ij}) \vec{r_{ij}} \right]^2 . \tag{83}$$

terms	E_V	$[E_V - E_0]/[2T]$
$u(McMillan)$	$-5.702(5)$	5.1%
$u(optimized)$	$-6.001(16)$	4.1%
u, ξ	$-6.901(4)$	0.86%
DMC	$-7.143(4)$	0.0%

Table 2: The energies of liquid ^{4}He in Kelvin/atom at zero pressure and zero temperature with various forms of trial functions[21]. In the first column u refers to pair correlations, ξ implies that three body terms were included. The second column shows the variational energies and the third column the percentage of the energy missed by the trial function. The numbers in parenthesis are the statistical error in units of 0.001K.

terms	E_V	$[E_V - E_0]/[2T]$	E_{FN}	*ref.*
u	$-1.15(4)$	5.7%	$-2.20(3)$	[19]
$u(optimized)$	$-1.30(3)$	5.7%	$-2.20(3)$	[21]
u, ξ	$-1.780(17)$	4.6%	$-2.20(3)$	[19]
u, η	$-1.730(4)$	3.7%	$-2.37(1)$	[19]
u, ξ, η	$-2.163(6)$	1.3%	$-2.37(5)$	[20]
$exp.$	-2.47	0.0%	-2.47	

Table 3: The energies of liquid ^{3}He in Kelvin/atom at zero pressure and zero temperature with various forms of trial functions. In the first column u refers to pair correlations, ξ implies that three body terms were included and η means backflow terms were included. The second column shows the variational energies and the third column the percentage of the energy missed by the trial function. The fourth column shows the results obtained with FN-DMC. The numbers in parenthesis are the statistical error in units of 0.01K.

The new function $\xi(r)$ can be shown to be roughly given by $\xi(r) \approx du(r)/dr$. Because the polarization has the form of a squared force it is rapid to compute. The computational time is roughly the same as the product trial wave function.

For a fermion system, the interaction can shift the antisymmetric part away from the non-interacting Slater determinant. The simplest correction in a homogeneous system is known as "backflow". The particle coordinates in the Slater determinants become "quasi-particle" coordinates:

$$det[\theta_k(\vec{s}_i, \sigma_i)], \tag{84}$$

where the 'quasi-particle' coordinates are defined by:

$$\vec{s}_i = \vec{r}_i + \sum_j \eta(r_{ij})\vec{r}_{ij}. \tag{85}$$

Backflow is needed to satisfy local current conservation. However the computation of the determinant and energy become much more complex, because each element of the Slater matrix now depends on all the electron coordinates.

Table 1 gives VMC energies for ^{4}He and Table 2 for ^{3}He, for a variety of trial functions. It is important to realize that the kinetic and potential energies are almost completely cancelling out,

liquid helium is very weakly bound. The third column $(E_V - E_0)/(2T)$ is a measure of the accuracy of the trial function, where $T = 12.3K$ is the kinetic energy and $E_0 = -2.47K$ is the ground state energy. This ratio is independent of how the zero of potential energy is defined and is equal to the percentage error in the upper bound for a harmonic potential. The chief motivation for the simulation of ^{3}He is that the results can rather directly be compared with experiment, assuming of course that the assumed inter-atomic potential is known accurately enough. There is a gratifying convergence toward experiment as more terms are added to the trial function. The most important terms beyond the pair-product level are the backflow terms. Similar results have been obtained using backflow and polarization terms on the two dimensional electron gas[56].

I acknowledge financial support from NSF-DMR94-224-96 and ONR-N00014-92J-1320. Publications from the University of Illinois Quantum Monte Carlo group are available at http://www.ncsa.uiuc.edu/Apps/CM homepage.html

References

[1] Ceperley, D. M. , Rev. Mod. Phys. **67**, 279 (1995).

[2] Ceperley, D. M., in "Simulation in Condensed Matter Physics and Chemistry", Ed. K. Binder and G. Ciccotti, 1996.

[3] D. M. Ceperley and M. H. Kalos, *Monte Carlo Methods in Statistical Physics*, ed. K. Binder, Springer-Verlag (1979).

[4] K. E. Schmidt and M. H. Kalos, *Monte Carlo Methods in Statistical Physics II, Topics in Current Physics*, ed. K. Binder, Springer-Verlag (1984).

[5] K. E. Schmidt and D. M. Ceperley, *Monte Carlo Methods III*, ed. K. Binder, Springer-Verlag,(1991).

[6] J.B. Anderson, in *Understanding Chemical Reactivity*, ed. by S.R. Langhoff, Kluwer, Dordrecht, 1995; Int. Rev. Phys. Chem., **14**, 85 (1995)

[7] B.L. Hammond, W.A. Lester, Jr., and P.J. Reynolds, *Monte Carlo Methods in ab initio quantum chemistry*, World Scientific, Singapore, 1994

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

[9] Allen, M. P., and D. J. Tildesley, 1987, *Computer Simulation of Liquids* (Oxford University, New York). pgs. 114-123.

[10] Kalos, M. H., and P. A. Whitlock, 1986, *Monte Carlo Methods Volume I: Basics* (Wiley, New York). pgs. 73-86.

[11] Hammersley, J. M., and D. C. Handscomb, 1964, *Monte Carlo Methods* (Chapman and Hall, London). pgs. 113-122.

[12] W. L. McMillan, Phys. Rev. A **138**, 442 (1965).

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

[14] R. Jastrow, Phys. Rev. **98**, 1479 (1955).

[15] A. Bijl, Physica **7**, 869 (1940).

[16] D. Ceperley, J. of Stat. Phys. **43**, 815 (1986).

[17] D. M. Ceperley, Phys. Rev. **B18**, 3126 (1978).

[18] L. Reatto , G. V. Chester, Phys, Lett. **22**,276 (1966).

[19] K. E. Schmidt,M. A. Lee, M. H. Kalos, and G.V. Chester, Phys. Rev. Lett. **47** , 807 (1981).

[20] R.M. Panoff and J. Carlson, Phys. Rev. Letts. **62** , 1130 (1989).

[21] S. Moroni, S. Fantoni and G. Senatore, Europhys. Lett. **30**, 93 (1995); Phys. Rev. B **52**, 13547 (1995).

[22] Trotter, H. F., 1959, Proc. Am. Math. Soc. **10**, 545.

[23] D. Ceperley, J. of Stat. Phys. **43**, 815 (1986).

[24] M. H. Kalos, Phys. Rev. **128**, 1791 (1962); J. Comp. Phys. **2**, 257 (1967).

[25] M. H. Kalos, D. Levesque and L. Verlet, Phys. Rev. A **9**, 2178 (1974).

[26] J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975); J. Chem. Phys. **65**, 4122 (1976).

[27] D. M. Ceperley, B. J. Alder, Phys. Rev. Letts. **45**, (1980).

[28] C.J. Umrigar, M.P. Nightingale, and K.J. Runge, J. Chem. Phys. **99**, 2865 (1993)

[29] N. Trivedi and D. M. Ceperley, Phys. Rev. **B41**, 4552 (1990).

[30] M. D. Donsker and M. Kac, J. Res. Natl. Bur. Stan. **44**, 551(1950).

[31] D. M. Ceperley, *Recent Progress in Many-Body Theories*, ed. J. Zabolitzky, Springer-Verlag (1981).

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

[33] D. M. Ceperley and B. J. Alder, J. Chem. Phys. **81**,5833 (1984).

[34] R. N. Silver, D. S. Sivia, and J. E. Gubernatis, Phys. Rev. **B41**, 2380 (1990).

[35] M. Caffarel and D. Ceperley, J. Chem. Phys. **97**, 8415 (1992).

[36] Ceperley, D. M. and Bernu, B., J. Chem. Phys. **89**, 6316 (1988).

[37] B. Bernu, D. M. Ceperley and W. A. Lester Jr., J. Chem. Phys. **93**, 552 (1990).

[38] Kwon, Y., Ceperley, D. M. and Martin, R. M., Phys. Rev. **B53**, 7376 (1996).

[39] J. Hirsch, Phys. Rev. **B28**, 4059 (1983); Phys. Rev. Letts. **51**, 1900 (1983); Phys. Rev. **B31**, 4403 (1985).

[40] G. An and J.M.J. van Leeuwen, Phys. Rev. **B44**,9140 (1991).

[41] ten Haaf, D. F. B., H. J. M. van Bemmel, J. M. J. van Leeuwen and W. van Saarloos, and D. M. Ceperley, Phys. Rev. **B 51**, 13039, (1995).

[42] Shang, S. J. Carlson and J. Gubernatis, Phys. Rev. Letts. **74**, 3074 (1991).

[43] Zhu, X. and S. G. Louie, Phys. Rev. Lett. **70**, 335 (1993).

[44] Ortiz, G., Ceperley, D. M. and Martin R. M., "2-D Fermions in a Magnetic Field:The Fixed-Phase Method" Phys. Rev. Letts. **71**, 2777 (1993).

[45] Ortiz, G. and D. M. Ceperley, Phys. Rev. Letts. **75**, 4642 (1995).

[46] Ceperley, D. M. and L. Mitas,*New Methods in Computational Quantum Mechanics*, Advances in Chemical Physics, ed. by S. Rice and I. Prigogine, 1995.

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

[48] G.B. Bachelet, D.M. Ceperley, and M.G.B. Chiocchetti, Phys. Rev. Lett. **62**, 2088 (1989)

[49] X.-P. Li, D.M. Ceperley, and R.M. Martin, Phys. Rev. B **44**, 10929 (1991)

[52] S. Fahy, X.W. Wang, and S.G. Louie, Phys. Rev. Lett. **61**, 1631 (1988)

[32] S. Fahy, X.W. Wang, and S.G. Louie, Phys. Rev. B **42**, 3503 (1990)

[50] M.M. Hurley and P.A. Christiansen, J. Chem. Phys. **86**, 1069 (1987)

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

[52] L. Mitas, in *Computer Simulation Studies in Condensed-Matter Physics V*, ed. by D.P. Landau, K.K. Mon and H.B. Schuttler (Springer, Berlin, 1993)

[53] L. Mitas and R.M. Martin, Phys. Rev. Lett., **72**, 2438 (1994)

[54] J.C. Grossman, L. Mitas, Phys. Rev. Lett., **74**, 1323 (1995)

[55] J.C. Grossman, L. Mitas, and K. Raghavachari, Phys. Rev. Letts. **75**, 3870 (1995).

[56] Kwon, Y., D. M. Ceperley, and R. M. Martin, Phys. Rev. **B48**, 12037 (1993).