Lectures on Quantum Monte Carlo

Riccardo Fantoni*

Dipartimento di Scienze Molecolari e Nanosistemi,
Università Ca' Foscari Venezia,
Calle Larga S. Marta DD2137,
I-30123 Venezia,
Italy

(Dated: March 4, 2021)

We discuss random walk methods in the first lecture. In the second lecture we discuss variational Monte Carlo which is a straightforward application of the Metropolis Monte Carlo method, the only complication being that the wavefunction for fermion systems is a determinant. In the third lecture we discuss the projector Monte Carlo methods. Finally in the fourth lecture we discuss the Path Integral Monte Carlo method.

CONTENTS

Quantum Monte Carlo methods have been developed to calculate properties of quantum many-body systems. The motivation is basically the same as for classical systems. In classical many-body system, direct simulations have proved the only way to get thoroughly reliable information about many-body effects, particularly as the systems get more comptex. 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.

There is hope that QMC will be useful in providing exact results, or at least exact constraints, on properties of many-body quantum systems. Except in a few cases, this hope is not fully realized today. Fermion statistics remain a challenge to the practitioner of simulation techniques. Nonetheless the results are competitive with those from the other methods used for quantum system and provide insight into the origin of properties of quantum systems.

---

* rfantoni@ts.infn.it

There are two basic types of methods that I will discuss. In zero temperature methods (Variational Monte Carlo and Projector Monte Cairlo) one calculates the properties of a single wavefunction. This is for when we need to calculate matrix elements like $\langle\phi|\mathcal{O}|\phi\rangle$. In finite temperature methods (Path Integral Monte Carlo) one takes a trace over the thermal density matrix: $\mathcal{O}\exp(-\beta\mathcal{H})$.

The equivalent to Molecular Dynamics (Quantum Molecular Dynamics) 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 serious approximations. Figure 1 shows a sort of family tree that connects various simulation methods.

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. I cannot even describe everything in these notes. They are provided as a kind of review for the student. More extensive discussion of these topics is to be found in Refs. (Anderson, 1995a,b; Ceperley, 1995; Ceperley and Kalos, 1979; Hammond *et al.*, 1994; Schmidt and Ceperley, 1991; Schmidt and Kalos, 1984).

First a few words on notation. I will always assume that the system is a non-relativistic collection of N particles described by the Hamiltonian:

$$\mathcal{H} = -\lambda\sum_{i=1}^{N}\nabla_i^2 + V(R), \tag{0.1}$$

where $\lambda = \hbar^2/2m$, $V(R) = \sum_{i<j}v(r_{ij})$, and $v(r)$ is a two-body pair potential. Later we will see why relativistic or spin Hamiltonians are more complicated to treat. The symbol R refers to the $3N$ vector of particle coordinates $(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$, and $(\mathbf{r}_i, \sigma_i)$ to the 3 spatial and 1 spin coordinate of particle i. The symbol $\hat{R}$ refers to the corresponding versor. Sometimes I will refer to the exact eigenfunctions and eigenvalues of the Hamiltonian: $(\phi_\alpha(R), E_\alpha)$. A known (computable) trial wave function will be denoted $\Psi(R)$. The symbol $\int$ will imply an integral over the full $3N$ dimensional configuration space of the particles. I will stick to the first quantized notation in the canonical ensemble. A boson wave function is then totally symmetrical under particle exchange and a fermion function is antisymmetrical. The permutation operator acting on particle labels is denoted PR.

Quantum Monte Carlo methods are exclusively examples of Markov processes or random walks. I discuss random walk methods in the first lecture, in the second lecture I discuss variational Monte Carlo which is a straightforward application of the Metropolis Monte Carlo method, the only complication being that the wavefunction for fermion systems is a determinant. 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 straightforward 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. In the third lecture 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. Finally in the fourth lecture I discuss, Path Integral Monte Carlo, another application of the Metropolis algorithm; its complications stem from the "ring-polymer" nature of the paths.

I. RANDOM WALKS

Let us start by reviewing the random walk or equivalently 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 (Metropolis *et al.*, 1953). 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 simple sampling methods is that their efficiency goes to zero as the dimensionality of the space increases. Suppose we want to sample the probability distribution:

$$\pi(s) = \frac{\exp[-S(s)]}{Z}, \tag{1.1}$$

where $S(s)$ is the action, say $\beta V(s)$ for the canonical classical Boltzmann distribution. The partition function Z normalizes the function π in this space and is usually not known. A direct sampling method, requires sampling a

function with a known normalization. Suppose we can directly sample a function $p_m(s)$. One can show that the Monte Carlo variance will depend on the ratio π/p_m is a sharply peaked function and it is very difficult to sample it directly because that would require knowing the normalization, or equivalently the exact partition function of a nearby related system, The efficiency would be related to the ratio of partition functions of the "model" systems to the real system which goes exponentially to zero as the number of particles increases.

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 for a probability distribution fixed in "time" and depending only on the "present". 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 that it satisfies

$$\sum_{s'} \mathcal{P}(s \to s') = 1, \tag{1.2}$$

and

$$\mathcal{P}(s \to s') \geq 0, \tag{1.3}$$

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

$$\sum_s \pi(s)\mathcal{P}(s \to s') = \pi(s'). \tag{1.4}$$

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 so that one never forgets which site one started on. Non-periodic rule holds 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 bomrdary conditions). It would be violated in a model of the expanding universe where the system gets further and further from equilibrium 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:

$$f_n(s) = \pi + \sum_\lambda \epsilon_\lambda^n c_\lambda \phi_\lambda(s), \tag{1.5}$$

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 stricttly 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,

$$\pi(s)\mathcal{P}(s \to s') = \pi(s')\mathcal{P}(s' \to s), \tag{1.6}$$

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. (1.2) 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 the equilibrium Eq. (1.4) as we will see.

II. METROPOLIS

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

$$\mathcal{P}(s \to s') = T(s \to s')A(s \to s'),\tag{2.1}$$

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

$$A(s \to s') = \min[1, q(s' \to s)],\tag{2.2}$$

where

$$q(s \to s') = \frac{\pi(s')T(s' \to s)}{\pi(s)T(s \to s')}.\tag{2.3}$$

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 (both $T(s \to s')$ and $T(s' \to s)$ should be zero or non-zero together) and secondly 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:

$$q = \frac{\pi(s')T(s' \to s_n)}{\pi(s_n)T(s_n \to s')}\tag{2.4}$$

   - Accept or reject:
     If $q > 1$ or if $q > u_n$ where u_n is a uniformly distributed random number (RN) 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

5. Collect averages every so often, and block them to get error bars

Consider the sampling of a classical Boltzman 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, 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:

$$\mathbf{r}' = \mathbf{r} + \Delta(\mathbf{u} - 1/2)\tag{2.5}$$

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

Some things to note about Metropolis:

- 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 in the move of one hard sphere. Then the probability of getting an acceptance with N hard spheres is $(1-p)^N = \exp(N \ln(1-p))$. In order to get any acceptances 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 proobability and the reverse probability if one has a nonuniform transition probability. Also note that we cannot calculate the normalization of π-it is never needed. Only ratios enter in.

- 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 try 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 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

$$\bar{\mathcal{O}} = \langle \mathcal{O} \rangle \tag{2.6}$$

and

$$\sigma_{\mathcal{O}}^2 = \langle (\mathcal{O}_k - \bar{\mathcal{O}})^2 \rangle \tag{2.7}$$

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

$$\mathrm{error}[\bar{\mathcal{O}}] = \sqrt{\frac{\kappa_{\mathcal{O}} \sigma_{\mathcal{O}}^2}{P}}. \tag{2.8}$$

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

$$\kappa_{\mathcal{O}} = 1 + 2 \sum_{k=1}^{P} \frac{\langle (\mathcal{O}_0 - \bar{\mathcal{O}})(\mathcal{O}_k - \bar{\mathcal{O}}) \rangle}{\sigma_{\mathcal{O}}^2}, \tag{2.9}$$

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. (2.8) 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.

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,

$$\xi_{\mathcal{O}} = \frac{1}{\kappa_{\mathcal{O}} \sigma_{\mathcal{O}}^2 T}, \tag{2.10}$$

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 an intuitive measure of convergence. Let us define the *diffusion constant D_R* of an algorithm by

$$D_R = \left\langle \frac{(R_{n+1} - R_n)^2}{T} \right\rangle, \tag{2.11}$$

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:

$$D_R \propto \langle A \rangle \Delta^2. \tag{2.12}$$

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 deline 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,

$$T_{HB}(s \to s') = \frac{\pi(s')}{C(s)}, \tag{2.13}$$

where the normalization constant is

$$C(s) = \sum_{s'' \in \mathcal{N}(s)} \pi(s''). \tag{2.14}$$

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

$$A(s \to s') = \min\left[1, \frac{C(s)}{C(s')}\right]. \tag{2.15}$$

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. (2.14) 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, butt with a transition rule which is able to be exacuted quickly

III. 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 known as Brownian dynamics and smart Monte Carlo are more naturally regarded as local dynamical random walks. The Smoluchowski equation

$$\frac{d\pi(R,t)}{dt} = \nabla D(R)[\nabla \pi - \beta \mathbf{F}(R)\pi], \tag{3.1}$$

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 Markovian, this must be the only solution.

Let us define the Green's function: $G(R, R_0; t)$ is the solution to Eq. (3.1) 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:

$$\pi(R)G(R \to R'; t) = \pi(R')G(R' \to R; t), \tag{3.2}$$

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 that 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'.) Then G can be used for a transition probability and it will always give an acceptance probability of unity. Also it gives interesting dynamics (not MD but dynamics of viscous particles always in contact with a heat bath).

The Smoluchoski equation leads to an interesting process but we can only calculate G in the short-time limit. In the following I explain a general procedure for devising an algorithm of sampling G. Let us calculate the moments of G,

$$I_n(R_0, t) = \int dR(R - R_0)^n G(R_0 \to R; t). \tag{3.3}$$

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 $\langle \ldots \rangle$). We assume there are no absorbing surfaces of the random walks. Then,

$$dI_0/dt = 0. \tag{3.4}$$

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:

$$(dI_1/dt)_k = \langle \nabla_j D_{kj} + D_{kj} F_j \beta \rangle, \tag{3.5}$$

where we use the Einstein summation convention over the repeated indexes. Let us assume that $\mathbf{F}$ and ∇D are slowly varying. Then we can replace them by the values at the intial points and integrate:

$$(R_t)_k = \langle R_k \rangle = (R_0)_k + t \langle \nabla_j D_{kj}(R_0) + D_{kj}(R_0) F_j(R_0)\beta \rangle + \mathcal{O}[t^2]. \tag{3.6}$$

The equation for the second moment is:

$$dI_2/dt = 2\langle D_{ii} + (R - R_0)_i [\nabla_j D_{ij} + D_{ij} F_j \beta] \rangle, \tag{3.7}$$

Integrating,

$$\langle (R - R_0)^2 \rangle = 2\langle D_{ii}(R_0) \rangle t + \mathcal{O}[t^2]. \tag{3.8}$$

The solution at small time is a Gaussian distribution with the above mean and covariance.

$$G_g(R, R_0; t) = \exp[-(R - R_t)(4\langle D_{ii}(R_0)\rangle t)^{-1}(R - R_t)][4\pi \langle D_{ii}(R_0)\rangle t]^{-3N/2}. \tag{3.9}$$

According to the central limit theorem, Eqs. (3.6)-(3.8) are all that is needed to simulate the random walk if the time step t is sufficiently small.

We have not yet discussed the diffusion tensor. For simplicity, one normally assumes that we have $D(R) = D_0 \mathcal{I}$ 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.

Suppose we want to sample the distribution $\exp(-\beta V(R))$. Then the acceptance probability is given by:

$$A(n) = \min\{1, \exp[-\beta(V(R') - V(R))]G(R, R_{n\tau}; \tau)/G(R', R_{n\tau}; \tau)\}, \tag{3.10}$$

so that the acceptance at large n ratio goes to unity at small τ. One can possibly make more accurate schemes by including off-diagonal components in the second moment. We can choose for a transition probability the most general correlated Gaussian in $3N$ variables,

$$T_S(R) = [(6N\pi)^{3N} \det \mathbf{A}]^{-1/2} e^{-(R-\bar{R})(6N\mathbf{A})^{-1}(R-\bar{R})}, \tag{3.11}$$

where the $3N \times 3N$ positive definite covariance matrix A and the mean position vector $\bar{R}$ can be arbitrary. Suppose we solve equation (3.8) to one higher order, we find at constant diffusion,

$$\mathbf{A} = 2(D_0 \mathcal{I} t + \beta^2 \mathbf{F}(R_0)\mathbf{F}(R_0)t^2). \tag{3.12}$$

We can sample the multivariate Gaussian distribution. One Cholesky factorizes the covariance matrix as $\mathbf{A} = \mathbf{SS}^T$, 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 $T(R \to R')$ 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.

IV. VARIATIONAL MONTE CARLO

We now turn to the first and simplest of the Qumtum Monte Carlo methods, Variational Monte Carlo (VMC). The VMC method was first used by McMillan (McMillan, 1965) to calculate the ground state properties of iquid ^{4}He and then generalized to fermion systems by Ceperley et al. (Ceperley *et al.*, 1977). It is a relatively simple generalization from a classical Monte Carlo simulation to VMC. The variational theorem says that for Ψ a proper trial function, the variational energy of the trial function is an upper bound to the exact ground state energy:

$$E_V = \frac{\int \Psi^*(R)\mathcal{H}\Psi(R)}{\int \Psi^*(R)\Psi(R)} \geq E_0. \tag{4.1}$$

One occasionally sees mistakes in the literature so let me remind you of the conditions that the trial function must satisfy:

1. $\mathcal{H}\Psi$ is well defined everywhere which means that both Ψ and $\nabla\Psi$ must be continuous wherever the potential is finite.

2. The integrals $\int |\Psi|^2$, $\int \Psi^*\mathcal{H}\Psi$, and $\int |\mathcal{H}\Psi|^2$ should exist. The last integral is only required to exist for a Monte Carlo evaluation of the integrais. If it does not exist the statistical error of the energy will be infinite.

3. Ψ has the proper symmetry: $\Psi(R) = (-1)^P \Psi(PR)$ for fermions and the right behavior at the periodic boundaries.

In the continuum, it is important to show analytically that properties 1 and 2 hold everywhere, particularly at the edges of the periodic box and when two particles approach each other. Otherwise either the upper bound property may not be guaranteed or the Monte Carlo error estimates are not valid. For a lattice spin model, only item 3 is applicable.

The variational method is then quite simple. Use the Metropolis algorithm to sample the square of the wave function:

$$\pi(R) = \frac{|\Psi(R)|^2}{\int |\Psi(R)|^2}. \tag{4.2}$$

Then the variational energy is simply the average value of the local residual energy over this distribution,

$$E_V = \int \pi(R) E_L(R) = \langle E_L(R) \rangle_\Psi, \tag{4.3}$$

where the local energy of Ψ is defined as:

$$E_L(R) = \Psi^{-1}\mathcal{H}\Psi(R). \tag{4.4}$$

Variational Monte Carlo(VMC) has a very important *zero variance property*: as the trial function approaches an exact eigenfunction, $\Psi \to \phi_\alpha$, the local energy approaches the eigenvalue everywhere, $E_L(R) \to E_\alpha$, and the Monte Carlo estimate of the variational energy converges more rapidly with the number of steps n the random walk. Of course, in this limit the upper bound is also becoming closer to the true energy. 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.

A. The Pair Product Trial function

Now consider a system interacting with a one-body (e.g. an external potentiai) and two-body potentials. Froebinius proved many years ago that the ground state of a real Hamiltonian (i.e. no magnetic fields) can always be made non-negative. This implies that the ground state has Bose symmetry.

The pair product trial function is the simplest generalization of the slater determinant and the ubiquitous form for the trial function in variational Monte Carlo:

$$\Psi(R, \sigma) = \exp\left[-\sum_{i<j} u(r_{ij})\right] \det[\theta_k(\mathbf{r}_i, \sigma_i)], \tag{4.5}$$

where $\theta_k(\mathbf{r}_i, \sigma_i)$ is the kth spin-orbital and $u(r)$ is the "pseudopotential" or pair-correlation factor. This function also goes by the name of Jastrow (Jastrow, 1955) wave function, although Bijl (Bijl, 1940) much earlier described the motivation for its use in liquid ^{4}He. Closely related forms are the Gutzwiller function for a lattice, or the Laughlin function in the fractional quantum Hall effect. Both $u(r)$ and $\theta_k(\mathbf{r}_i, \sigma_i)$ are in principle determined by minimizing the variational energy.

B. Details

I will only mention a few details concerning VMC. First, how do the particles move? On a lattice one can make a random hop of a particle or a spin flip. In the continuum, in the classic Metropolis procedure, one moves the particles one at a time by adding a random vector to the particle's coordinate, where the vector is either uniform inside of a

cube centered about the old coordinate, or is a normally distributed random vector centered around the old position. Assuming the first kind of move for the ith particle, the trial move is accepted with probability:

$$q(R \to R') = |\Psi(R')/\Psi(R)|^2$$

$$= \exp[-2\sum_{i<j}(u(r'_{ij}) - u(r_{ij}))]\left|\frac{\det[\theta_k(\mathbf{r}'_i, \sigma'_i)]}{\det[\theta_k(\mathbf{r}_i, \sigma_i)]}\right|^2$$

$$= \exp[-2\sum_{i<j}(u(r'_{ij}) - u(r_{ij}))]|\det[\sum_k \theta_k(\mathbf{r}'_j, \sigma'_j)C_{ki}]|^2 \qquad (4.6)$$

where the matrix, C, is the transposed inverse to the Slater matrix

$$\sum_k \theta_k(\mathbf{r}_i, \sigma_i)C_{kj} = \delta_{ij}. \qquad (4.7)$$

By the elementary properties of deteminants C is also proportional to the cofactor matrix, so we have used it to calculate the acceptance probability. Let me remind the reader that the evaluation of a general determinant takes $O(N^3)$ operations. The evaluation of the fermion part of the acceptance ratio will take $O(N)$ operations if C is kept current. If a move is accepted, C needs to be updated (Ceperley *et al.*, 1977)

$$C'_{jk} = C_{jk} + \sum_i(\delta_{ij} - b_j)C_{ik}/b_i, \qquad (4.8)$$

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

The local energy, needed to evaluate the variational energy is calculated by applying the Hamitonian to the trial function:

$$E_L(R) = V(R) + \lambda\sum_i[\nabla_i^2 U - (\nabla_i U)^2 + 2(\nabla_i U)(\nabla_i B)/B - (\nabla_i^2 B)/B], \qquad (4.9)$$

where $B = \det[\theta_k(\mathbf{r}_j, \sigma_j)]$, and $U = \sum_{i<j} u(r_{ij})$. Thus the inverse matrix is also needed to determine the local energy. Very often the orbitals are taken to be exact solutions to some model problem, in which case the term $\nabla_i^2 B$, will simplify. Finally note that using Green's identity allows several alternative ways (Ceperley *et al.*, 1977) 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. (4.9) will always have the lowest variance. The other forms of the energy give useful tests of the computer program and the convergence of the random walk.

C. Optimization of Trial Functions

Optimization of the parameters in a trial function is crucial for the success of the variational method and important for the Projector Monte Carlo method. 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. Clearly one minimizes E_V if the object of the calculation is to find the least upper bound. There are also some general arguments suggesting that the trial function with the lowest variational energy will maximize the efficiency of Projector Monte Carlo (Ceperley, 1986).

- The dispersion of the local energy: $\int |(\mathcal{H} - E_V)\Psi|^2$. If we assume that every step on a QMC calculation is statistically uncorrelated with the others, the dispersion is proportional to the variance of the calculation. There are some indications that minimization of the dispersion is statistically more robust than the variational energy because it is a positive definite quantity with zero as a minimum value.

- The overlap with the exact wave function: $\int \Psi\phi$. This is equivalent to finding the trial function which is 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. Optimization of the overlap will involve a Projector Monte Carlo calculation to determine it, which is a more computer intensive step.

I will now review the analytic properties of the optimal pseudopotential, u^*. Suppose we assume that the spin-orbits come from an exact solution of a one-body potential

$$-\lambda\nabla^2\theta_k(\mathbf{r}) + [v(r) - e_k]\theta_k(\mathbf{r}) = 0. \tag{4.10}$$

Let us examine the dominant terms in the local energy Eq. (4.9) as 2 particles are brought together. In a good trial function the singularities in the kinetic energy must cancel the singularities of the potential energy. In this limit the local energy will have the form:

$$E_L(R) = v(r) + 2\lambda\nabla^2 u(r) - 2\lambda[\nabla u(r)]^2 + \ldots, \tag{4.11}$$

where r is the distance separating the particles. An intuitive result emerges: $e^{-u(r)}$ will equal the solution to the 2-body Schroedinger equation. For He atoms interacting with a Lennard-Jones potential $4\epsilon(\sigma/r)^{12}$, at small distances this gives: $u(r) = \sqrt{2\epsilon\sigma^2/25\lambda}(\sigma/r)^5$. For the Coulomb potential $e_i e_j/r$ this equation can be used to derive the cusp condition

$$e_i e_j + 2(\lambda_i + \lambda_j) \left.\frac{du(r_{ij})}{dr_{ij}}\right|_{r_{ij}=0} = 0. \tag{4.12}$$

Now let us turn to the large r behavior of the optimal $u(r)$ where a description in terms of collective coordinates (phonons, or plasmons) is appropriate. The variational energy can be written as:

$$E_V = E_F + \sum_k (v_k - \lambda k^2 u_k)(S_k - 1) \tag{4.13}$$

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 (Ceperley, 1978) the optimal pseudopotential at long wavelengths:

$$2\rho u_k = -\frac{1}{S_{0k}} + \sqrt{\frac{1}{S_{0k}} + \frac{2\rho v_k}{\lambda k^2}}. \tag{4.14}$$

For a short-ranged potential, (e.g. liquid helium), v_k can be replaced by a constant and we find the Reatto-Chester (Reatto and Chester, 1966) form: $u(r) \propto r^{-2}$. But for a charged system, where $v_k \propto k^{-2}$, then $u(r) \propto r^{-1}$.

This raises a very important point which we will not have space to go into. Optimal pseudopotentials are always long-ranged in the sense that correlation will extend beyond the simulation box. The ground state energy is little affected by this tail in the wave function, but response functions, such as the dielectric function or the static structure factor are crucially dependent on using the corret 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 (Allen and Tildesley, 1987; Ceperley, 1978) is needed to represent the correct long-range behavior of the optimal trial function.

It is possible to carry out further analysis of the optimal correlation factor using the Fermi-Hypernetted-Chain equation. However, at intermediate distances or for highly correlated or complex systems, a purely Monte Carlo optimization method is needed. The simplest such method consists of running independent VMC runs with a variety of different variational parameters, fitting the resulting energies with a quadratic form, doing more calculations at the predicted minimum, until convergence in parameter space is attained. The difficulty is that close to the minimum the independent statistical errors will mask the variation with respect to the trial function parameters. The derivative of the variational energy with respect to trial function parameters is very poorly calculated. Also, it is difficult to optimize by hand, functions involving more than 3 variational parameters.

A correlated sampling method, known as reweighting (Ceperley *et al.*, 1977; Ceperley and Kalos, 1979) solves this problem. 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: $|\Psi(R; a_0)|^2$. Then the variational energy (or variance) for another trial function can be calculated by using the same set of points:

$$E_V(a) = \frac{\sum_j w(R_j, a)E_L(R_j, a)}{\sum_j w(R_j, a)}, \tag{4.15}$$

where the weight factor $w(R, a) = |\Psi(R; a)/\Psi(R; a_0)|^2$ 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. There is a dangerous 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 $N_{\text{eff}} = (\sum w_j)^2 / \sum w_j^2$. If this becomes much smaller than the number of points, one must resample and generate some new points. Using the reweighting method one can find the optimal value of wavefunction containing tens of parameters.

D. Beyond the pair-product trial function

Relatively little has been done to take the variational results beyond the two-body level. I will describe several of the recent directions. The Possibilities for improving the pair-product trial function in a homogenous one-component system are relatively limited. The dominant term missing in the trial function for a bosonic system is a thre-body (or polarization) term with the functional form of a squared force:

$$U_3(R) = -\sum_i \left[\sum_j \xi(r_{ij}) \mathbf{r}_{ij} \right]^2 . \tag{4.16}$$

The new variational function $\xi(r)$ can be shown to be roughly given by $\xi(r) = du(r)/dr$. The overall functional form of the polarization (the form of a squared force) makes it rapid to compute: the computational time being of the same order of magnitude as the 2 body pair function.

For a fermion system, the interaction can affect the nodes of the interacting wave function. The simplest correction in a homogenous system is known as "backflow". The particle coordinates in the slater determinants become "quasi-particle" coordinates:

$$\det[\theta_k(\mathbf{s}_i, \sigma_i)], \tag{4.17}$$

where the "quasi-particle" coordinates are defined by: $\mathbf{s}_i = \mathbf{r}_i + \sum_j \eta(r_{ij}) \mathbf{r}_{ij}$. 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.

E. 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 essential approximations need to be made and there are no restrictions on the trial function except that it 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. To motivate the methods to be described in the next section, let me list some of the intrinsic problems with the variational method.

- The variational method favors simple states over more complicated states. One of the main uses of simulations is to determine when and if a zero-temperature phase transition will occur. 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 means that the difference between the liquid and solid variational energies for the same type of trial function, (e.g. a pair product form) the solid energy will be closer to the exact result than the liquid and 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 jellium is simpler than for unpolarized helium 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 very time consuming, particularly for complex trial functions. In the one component system (say the electron gas) one only has to optimize a single $u(r)$ function, the orbitals are determined by symmetry. In the H_2O molecule, one has 5 difierent 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! This allows an element of human bias; the optimization is stopped when the expected result is obtained.

TABLE I The Green's function for various projection methods. τ is the timestep, and E_T is the trial energy. They have all been normalized to be unity at the origin and to have the same derivative.

Method	$G(R, R'; \tau)$	Ref.
Diffusion DMC	$\exp[-\tau(\mathcal{H} - E_T)]$	(Anderson, 1975, 1976; Ceperley and Alder, 1980)
Green's function GFMC	$[1 + \tau(\mathcal{H} - E_T)]^{-1}$	(Kalos, 1962, 1967; Kalos *et al.*, 1974)
Power PMC	$[1 - \tau(\mathcal{H} - E_T)]$	(Trivedi and Ceperley, 1990)

- 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.

- You almost always get out what is put in. 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.

V. PROJECTOR MONTE CARLO

In the last lecture, I discussed the variational Monte Carlo method. Now I will turn to a potentially more powerful method where a function of the Hamiltonian projects out the ground state, hence the name, projector Monte Carlo. In fact, the nomenclature of the various quantum Monte Carlo methods is not at all standardized. The 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, $G(R, R'; \tau)$: to some initial state $\psi_0(R)$:

$$\psi_{n+1}(R) = e^{-\tau(\mathcal{H} - E_T)} \psi_n(R) = \int dR' \, G(R, R'; \tau) \psi_n(R'). \tag{5.1}$$

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

$$\psi_n(R) = \sum_\alpha \phi_\alpha(R) \langle \phi_\alpha | \psi_0 \rangle e^{-n\tau(E_\alpha - E_T)}. \tag{5.2}$$

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

$$\lim_{n \to \infty} \psi_n(R) = \phi_0(R) \langle \phi_0 | \psi_0 \rangle e^{-n\tau(E_0 - E_T)}. \tag{5.3}$$

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.

Now the application of the Green's function involves a $3N$ dimensional integral. Hence once N gets larger than a few, one must do the integral with Monte Carlo. The interpretation of Eq. (5.1) is very similar to the Markov chain we discussed earlier. The probability of starting a random walk at R_1 is $\psi_0(R_1)$ (for the moment let us discuss the case where ψ_0 is non-negative, the boson case.) To sample $\psi_1(R)$, we choose moves from R_0 to R_1 from the Green's function $G(R_1, R_0; \tau)$. In the limit that the time step approaches zero, a coordinate space representation of the Green's function is:

$$\langle R | e^{-(\mathcal{H} - E_T)} | R' \rangle = (4\pi\lambda\tau)^{-3N/2} e^{-\frac{(R - R')^2}{4\lambda\tau}} e^{-\tau[V(R) - E_T]} + O(\tau^2), \tag{5.4}$$

The iteration equation, Eq. (5.1), 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. (5.4) and then branching. The number of copies of R' in the next generation is the integer part of

$$m = u + \exp[-\tau(V(R) - E_T)], \tag{5.5}$$

where u is a uniform random number in $(0, 1)$. If the potential energy is less than the ground state energy, duplicate copies of the configuration are 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 tne 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.

The population (number of walkers) fluctuates from step to step. The trial energy E_T, must be adjusted to keep the population within computationally acceptable limits. This is done by adjusting the trial energy as

$$E_T = E_0 + \kappa \ln(P^*/P), \tag{5.6}$$

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 as small as possible while achieving the goal of stabilizing the population around the target, P^*. If it is too large, one can bias the distribution.

A. Importance Sampling

The above scheme, first suggested by Fermi, was actually tried out in the first days of computing some forty years ago (Donsker and Kac, 1950). 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. Even with a bounded potential the method becomes very inefficient as the number of particles increases since the branching factor grows. But there is a very simple cure discovered by Kalos (Kalos *et al.*, 1974) for GFMC, but equally applicable to any projector method. *Importance sampling* multiplies the underlying probability distribution by a known, approximate solution which we call the *trial* or *guiding* function, $\Psi(R)$. Multiply Eq. (5.1) by Ψ, the trial function, and define $f_n(R) = \Psi(R)\psi_n(R)$. Then:

$$f_{n+1} = \Psi e^{-\tau(\mathcal{H}-E_T)}\psi_n = \int dR' \, \tilde{G}(R, R'; \tau) f_n(R), \tag{5.7}$$

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

$$-\frac{\partial \tilde{G}(R, R'; \tau)}{\partial \tau} = -\lambda \nabla[\nabla \tilde{G} - 2\tilde{G}\nabla \ln(\Psi(R))] + [E_L(R) - E_T]\tilde{G}, \tag{5.8}$$

where $E_L(R)$ is the local-energy defined in the previous lecture. As we discused earlier, we can consider each term on the right-hand side as a 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. As the trial function approaches the exact eigenfunction, the branching factor approaches unity; thus a sufiiciently good trial function can control the branching.

The importance sampled DMC algorithm is

1. The ensemble is initialized with a VMC sample from $\Psi^2(R)$.

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

$$R_{n+1} = R_n + \chi + \lambda\tau \nabla \ln(\Psi^2(R_n)), \tag{5.9}$$

where χ is a normally distributed random vector with variance $2\lambda\tau$ and zero mean. In order to take into account of the diffusion and the drifting.

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

$$u + \exp[-\tau(E_L(R_n) - E_T)], \tag{5.10}$$

where u is a uniformly distributed random number in $(0, 1)$. This takes care of 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 as in Bq. (5.6).

6. To obtain ground state expectations of quantities other than the energy, one must correct the average over the DMC walk using the so-called "mixed estimator", $V_{\mathrm{mix}} = \langle \psi_0 | V | \Psi \rangle$, and the variational estimator (Ceperley and Kalos, 1979). For example the potential energy is calculated as:

$$\langle \psi_0 | V | \psi_0 \rangle = 2\langle \psi_0 | V | \Psi \rangle - \langle \Psi | V | \Psi \rangle + O[(\psi_0 - \Psi)^2]. \tag{5.11}$$

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.

Note that repeated use of step 2 alone would generate a probability density proportional to Ψ^2, i.e. if we turn off the branching we recover VMC. 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.

B. The Fixed-Node Method

We have not discussed at all the problem posed by Fermi statistics to the projector Monte Carlo method. First let us consider the difficulty in implementing the non-importance sampled algorithm. The initial condition $\psi_0(R)$ is not a probability distribution since a fermion trial function will have an equal volume of positive and negative regions. Hence 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. The initial distribution $|\Psi(R)|^2$ is positive, but the Green's function, $\tilde{G}(R, R'; \tau)$ can be negative if a step changes the sign of Ψ. Thereafter a minus sign will be attached to the walk which will lead 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.

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 it vanish wherever the trial function vanishes. One can easily demonstrate that the resulting energy will be an upper bound to the exact ground state energy (Ceperley, 1981); the best Possible upper bound with the given boundary conditions. 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. One is exactly solving the wave equation inside the fixed-nodal regions, but there is a mismatch of the derivative of the solution across the boundary. The nodes have an unequal pressure on the two sides unless the nodes are exact. Where comparison has been done between the energy of the trial function, the energy of the fixed-node approximation, and the exact energy, 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 obviously 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 $\psi(R) = 0$. 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 (Ceperley, 1991). 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 $\psi(R) \neq 0$. 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 cross-sections of free particle nodes are given in Ref. (Ceperley, 1991).

C. Exact Fermion Method

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 small it may be possible to do rather accurate "exact calculations" of fifty or more particles.

The transient estimate (TE) method calculates the ratio:

$$E_{\text{TE}}(\tau) = \frac{\int \Psi \mathcal{H} e^{-\tau(\mathcal{H}-E_T)} \Psi}{\int \Psi e^{-\tau(\mathcal{H}-E_T)} \Psi}, \tag{5.12}$$

where $\mathcal{H}$ is the exact Hamiltonian (not the fixed nodes one) and Ψ is an antisymmetric trial function. Clearly the variational theorem applies so that $E_{\text{TE}}(\tau) \geq E_0$. Also the energy converges exponentially fast in τ:

$$\lim_{\tau \to \infty} E_{\text{TE}}(\tau) = E_0 + O(e^{-\tau E_g}), \tag{5.13}$$

where 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. It is obtained by making 2 particle-hole excitations.

For a method to find self consistently its own nodes, the walks must be able to go anywhere, and so the drift term in Eq. (5.8) 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. (5.8) and will be assumed to be a reasonable boson ground state trial function, while the trial function appears in Eq. (5.12). The Ψ_G importance sampled Green's function is:

$$\tilde{G}(R, R'; \tau) = \Psi_G(R) \langle R | e^{-\tau(\mathcal{H}-E_T)} | R' \rangle \Psi_G^{-1}(R'), \tag{5.14}$$

and we can rewrite Eq. (5.12) as:

$$E_{\text{TE}}(\tau) = \frac{\int \sigma(R) E_{\text{LT}}(R) \tilde{G}(R, R'; \tau) \sigma(R') \Psi_G^2(R')}{\int \sigma(R) \tilde{G}(R, R'; \tau) \sigma(R') \Psi_G^2(R')} \tag{5.15}$$

where $\sigma(R) = \Psi(R)/\Psi_G(R)$ and $E_{\text{LT}}(R)$ 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.

Recapping, 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. (5.15).

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

3. Propagate the walk forward an amount of time, τ with the Green's function, $\tilde{G}(R, R'; \tau)$. If a branch occurs, each branch will count separately.

4. The weight of the walk arriving at R is $\sigma(R)\sigma(R')$. The energy at time τ is computed as:

$$E_{\text{TE}}(\tau) = \frac{\langle [E_{\text{LT}}(R) + E_{\text{LT}}(R')]\sigma(R)\sigma(R')\rangle}{2\langle \sigma(R)\sigma(R')\rangle}, \tag{5.16}$$

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

We see that the weight 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. Hence the nodes of the true wave function can differ from those of the trial function if there is an unequal diffusion of walks from the negative and positive regions.

The release node (RN) algorithm (Ceperley, 1981, 1991) is an improvement on this TE method. Instead of starting the projection from the trial function, one begins the projection from the fixed-node solution. There are several advantages. First of all 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 the Eq. (5.12) depends on the projection time. It is not hard to see that the value of both the numerator and denominator are asymptotically proportional to $\exp[-\tau(E_F - E_T)]$. Thus to keep the normalization fixed our 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 $\exp[-\tau(E_B - E_T)]$ 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:

$$\epsilon_{\text{stat}} = e^{-\tau(E_F - E_B)}. \tag{5.17}$$

The behavior is physically easy to understand. Our estimator depends on finding difrerences 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, τ, 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. (5.13,5.17) we find the total error will decrease as:

$$\epsilon \propto P^{-\eta} \quad \eta = \frac{E_g}{2(E_F - E_B + E_g)}. \tag{5.18}$$

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, $E_{\text{TE}}(\tau)$, 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 (Silver *et al.*, 1990) 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 (Caffarel and Ceperley, 1992) 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 that 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 problem" in the literature. Note that the TE and RN methods do converge to the exact fermion energy. A proper statement of the fermion sign problem is in tenns of

complexity theory. Namely how long does it take to achieve a given error estimate, and, more precisely, how does this scale with the number of fermions. Clearly one of the important tasks of simulations is to calculate properties of systems near phase transitions so the ability to do large systems is crucial. In the TE method, the computer time to reach a given precision grows exponentially with the number of fermions. I would say that a complete solution of the fermion problem would be an approximation free algorithm which scales as some low power of the number of fermions.

Let me just briefly mention the computational complexity of simulations of a few physical systems. Properties of classical systems can be simulated in time $O(N)$. Simulations of equilibrium properties of quantum bosons at zero or non-zero temperature are also $O(N)$. A Heisenberg model on a bipartite lattice, or any 1D fermion system is $O(N)$. Variational MC calculations of fermion systems are $O(N^3)$ in general, but the exponent would be smaller if localized spin-orbits are used. The Hubbard model at half filling on a bipartite lattice (Hirsch, 1983a,b, 1985) is $O(N^3)$ 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 $O(e^{\kappa N})$. 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. Solid ^{4}He can be simulated with a permanent of gaussians centered on the crystal sites with variable widths, which requires $O(N2^N)$ operations, even worst than liquid fermions.

D. Problems with projections methods

The projection method shares many of the same problems with the variational method. In fact it is useful to think of the projection method as a "super-variational" method. n both VMC and DMC there is a premium for good trial functions; that is the most straightforward way of making progress to solving the many-fermion problem.

- 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 the human bias than with VMC. There is also the possibility of new things coming out of the simulation. For example, one may observe a particular type of correlation completely absent from the trial function. Hence it is always good to pay close attention to correlation functions computed by DMC since that it is a good way of learning what is missing in the trial function. But it is slower than VMC because the timestep needs to be smaller. The cost in computer Lime is typically a factor of 2 to 10.

- Although the probability distribution does converge to the exact answer, in practice, this does not always occur in any given calculation 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. However, wiLh DMC the importance sampling always 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 states. The ability to perform simulations in a metastable state is useful but the results must be interpreted with caution.

- 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.

- 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.

- Release node calculations only improve the nodes locally. If τ is the release node projection time, then we can move the nodes a distance of at most $\sqrt{6N\lambda\tau}$.

- 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.

E. Path Integral Monte Carlo

See Ref. (Ceperley, 1995) and attached notes on the simulation of jellium at finite temperature.

ACKNOWLEDGMENTS

REFERENCES

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

Anderson, J B (1975), J. Chem. Phys. **63**, 1499.

Anderson, J B (1976), J. Chem. Phys. **65**, 4122.

Anderson, J B (1995a), Int. Rev. Phys. Chem. **14**, 85.

Anderson, J B (1995b), *Understanding Chemical Reactivity*, edited by S. R. Langhoff (Kluwer, Dordrecht).

Bijl, A (1940), Physica **7**, 869.

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

Ceperley, D M (1978), Phys. Rev. B **18**, 3126.

Ceperley, D M (1981), *Recent Progress in Many-Body Theories*, edited by J. Zabolitzky (Springer-Verlag).

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

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

Ceperley, D M (1995), "Path integrals in the theory of condensed helium," Rev. Mod. Phys. **67**, 279.

Ceperley, D M, and B. J. Alder (1980), Phys. Rev. Lett. **45**, 566.

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

Ceperley, D M, and M. H. Kalos (1979), *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer-Verlag).

Donsker, M D, and M. Kac (1950), J. Re. Natl. Bur. Stan. **44**, 551.

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

Hirsch, J (1983a), Phys. Rev. B **28**, 4059.

Hirsch, J (1983b), Phys. Rev. Lett. **51**, 1900.

Hirsch, J (1985), Phys. Rev. B **31**, 4403.

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

Kalos, M H (1962), Phys. Rev. **128**, 1791.

Kalos, M H (1967), J. Comp. Phys. **2**, 257.

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

Kalos, M H, and P. A. Whitlock (1986), *Monte Carlo Methods*, Vol. I (Jon Wiley & Sons, Inc.).

McMillan, W L (1965), Phys. Rev. A **138**, 1479.

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

Reatto, L, and G. V. Chester (1966), Phys. Lett. **22**, 276.

Schmidt, K E, and D. M. Ceperley (1991), *Monte Carlo Methods III*, edited by K. Binder (Springer-Verlag).

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

Silver, R N, D. S. Silva, and J. E. Gubernatis (1990), Phys. Rev. B **41**, 2380.

Trivedi, N, and D. M. Ceperley (1990), Phys. Rev. B **41**, 4552.