Lectures on Quantum Monte Carlo

David M. Ceperley

May 1, 1996

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 comples. 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.

There are two basic types of methods that I will discuss. In zero temperature methods (Variational Monte Carlo and Projector Monte Carlo) one calculates the properties of a single wave-function. 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. [20, 21, 22, 53, 51, 52].

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:

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

where $\lambda = \hbar^2/2m$ and $v(r)$ is a two-body pair potential. Later we will see why relativistic or spin Hamiltonians are more complicated to treat. 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. The symbol R refers to the 3N vector of particle coordinates, σ to the N spin coordinates, and (r_i, σ_i) to the 3 spatial and 1 spin coordinate of particle i. 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.

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 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. 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.

1 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[1]. 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:

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

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. $\pi(s)$ 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

$$\sum_{s'} \mathcal{P}(s \to s') = 1 \tag{3}$$

and

$$\mathcal{P}(s \to s') \geq 0. \tag{4}$$

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:

$$\sum_{s} \pi(s)\mathcal{P}(s \to s') = \pi(s'). \tag{5}$$

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 $\mathcal{P}(s \rightarrow s) > 0$; 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 (it 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:

$$f_n(s) = \pi + \sum_{\lambda} \epsilon_{\lambda}^{n} c_{\lambda} \phi_{\lambda}(s), \tag{6}$$

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,

$$\pi(s)\mathcal{P}(s \rightarrow s') = \pi(s')\mathcal{P}(s' \rightarrow s). \tag{7}$$

If the pair $\{\pi(s), \mathcal{P}(s \rightarrow s')\}$ satisfy detailed balance and if $\mathcal{P}(s \rightarrow s')$ 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.(3) 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. (5) as we will see.

1.1 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*** $T(s \rightarrow s')$ (which is a probability distribution that we can sample) and an ***acceptance probability*** $A(s \rightarrow s')$ where $0 \leq A \leq 1$.

$$\mathcal{P}(s \rightarrow s') = T(s \rightarrow s')A(s \rightarrow s'). \tag{8}$$

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

$$A(s \rightarrow s') = min\left[1, q(s' \rightarrow s)\right], \tag{9}$$

where

$$q(s \rightarrow s') = \frac{\pi(s')T(s' \rightarrow s)}{\pi(s)T(s \rightarrow s')}. \tag{10}$$

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:
   $$q = \frac{\pi(s')T(s' \to s_n)}{\pi(s_n)T(s_n \to s')} \tag{11}$$
   - 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.
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:
$$\mathbf{r}' = \mathbf{r} + \mathbf{u} - 1/2\Delta \tag{12}$$
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 probability and the reverse probability if one has a non-uniform 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 $T(s \rightarrow s')$ 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

$$\bar{\mathcal{O}} = \langle \mathcal{O}_k \rangle \tag{13}$$

and

$$\sigma_{\mathcal{O}}^2 = \langle (\mathcal{O}_k - \bar{\mathcal{O}})^2 \rangle \tag{14}$$

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

$$error[\bar{\mathcal{O}}] = \sqrt{\frac{\kappa_{\mathcal{O}} \sigma_{\mathcal{O}}^2}{P}}. \tag{15}$$

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

$$\kappa_{\mathcal{O}} = 1 + 2 \sum_{k=1}^{\infty} \frac{\langle (\mathcal{O}_0 - \bar{\mathcal{O}})(\mathcal{O}_k - \bar{\mathcal{O}}) \rangle}{\sigma_{\mathcal{O}}^2}, \tag{16}$$

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. 15 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,

$$\xi_{\mathcal{O}} = \frac{1}{\kappa_{\mathcal{O}} \sigma_{\mathcal{O}}^2 T}, \tag{17}$$

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

$$D_R = \left\langle \frac{[(R_{n+1} - R_n)]^2}{T} \right\rangle, \tag{18}$$

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:

$$D_R \propto \langle A \rangle \Delta^2. \tag{19}$$

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,

$$T_{HB}(s \to s') = \frac{\pi_{s'}}{C_s}, \tag{20}$$

where the normalization constant is

$$C_s = \sum_{s'' \in \mathcal{N}(s)} \pi_{s''}. \tag{21}$$

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

$$A(s \to s') = \min\left[1, \frac{C_s}{C_{s'}}\right]. \tag{22}$$

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. (21) 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 known as 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

$$-d\pi(R,t)/dt = -\nabla D(R)[\nabla\pi - \beta\mathbf{F}(R)\pi], \tag{23}$$

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. (23) 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:

$$\pi(R)G(R \to R'; t) = \pi(R')G(R' \to R; t), \tag{24}$$

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'.) 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,

$$I_n(R_0, t) = \int dR(R - R_0)^n G(R_0 \to R; t). \tag{25}$$

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,

$$dI_0/dt = 0. \tag{26}$$

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:

$$dI_1/dt = < \beta D\mathbf{F} + \nabla D > . \tag{27}$$

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

$$< R_t > = R_0 + t[\beta \mathbf{F}(R_0) + \nabla D(R_0)] + \mathcal{O}[t^2]. \tag{28}$$

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

$$dI_2/dt = 2 < D > + 2 < (R - R_0)(\beta \mathbf{F} + \nabla D) > . \tag{29}$$

Integrating,

$$< (R - R_0)^2 > = 2D(R_0)t + \mathcal{O}[t^2]. \tag{30}$$

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

$$G_g(R, R_0; t) = \exp[-(R - R_t)(2D(R_0)t)^{-1}(R - R_t)][2\pi t det(D(R_0))]^{-1/2}. \tag{31}$$

According to the central limit theorem, Eqs. 28-30 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 $D(R) = D_0\mathcal{I}$ 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.

Then the acceptance probability a constant diffusion is given by:

$$A = min\left[1, exp(-\beta(V(\mathbf{r}') - V(\mathbf{r})) - \beta(F(\mathbf{r}) + F(\mathbf{r}'))(2(\mathbf{r}' - \mathbf{r}) - \beta D(F' - F))/4)\right]. \tag{32}$$

The acceptance ratio goes to unity at small t. 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,

$$T_S(R) = \sqrt{(2\pi)^{3m} det(\mathbf{A})} e^{-(R-\bar{R})(2\mathbf{A})^{-1}(R-\bar{R})}, \tag{33}$$

where the 3×3 positive-definite covariance matrix $\mathbf{A}$ and the mean position vector $\bar{R}$ can be arbitrary. Suppose we solve equation (29) to one higher order:

$$\mathbf{A} = 2\mathcal{I}t - t^2 \nabla\nabla V(R_0). \tag{34}$$

We can sample the multivariate Gaussian distribution. 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 $T(R \to R')$ 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.

2 Variational Monte Carlo

We now turn to the first and simplest of the Quantum Monte Carlo methods, Variational Monte Carlo (VMC). The VMC method was first used by McMillan[8] to calculate the ground state properties of liquid ^{4}He and then generalized to fermion systems by Ceperley *et al.*[9]. 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:

$$E_V = \frac{\int \Psi^*(R)\mathcal{H}\Psi(R)}{\int \Psi^*(R)\Psi(R)} \geq E_0. \tag{35}$$

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 $\int |\Psi \mathcal{H}|^2$ should exist. The last integral is only required to exist for a Monte Carlo evaluation of the integrals. 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. 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:

$$\pi(R) = \frac{|\Psi(R)|^2}{\int |\Psi(R)|^2}. \tag{36}$$

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

$$E_V = \int \pi(R) E_L(R) = \langle E_L(R) \rangle_\Psi , \tag{37}$$

where the local energy of Ψ is defined as:

$$E_L(R) = \Psi^{-1} \mathcal{H} \Psi(R). \tag{38}$$

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 in 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.

2.1 The Pair Product Trial Function

Now consider a system interacting with a one-body (*e. g.* an external potential) 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:

$$\Psi(R, \sigma) = \exp[-\sum_{i<j} u(r_{ij})] det[\theta_k(r_i, \sigma_i)], \tag{39}$$

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[10] wave function, although Bijl[11] 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 $\theta_k(r, \sigma)$ are in principle determined by minimizing the variational energy.

2.2 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:

$$q(R \to R') = \mid \Psi(R')/\Psi(R) \mid^2 = \exp[-2\sum_{j \neq i}(u(r_i' - r_j) - u(r_i - r_j))] \mid \sum_k \theta_k(r_i')C_{ki} \mid^2, \tag{40}$$

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

$$\sum_k \theta_k(r_i)C_{kj} = \delta_{jk} \tag{41}$$

By the elementary properties of determinants 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[9]

$$C'_{jk} = C_{jk} + [\delta_{ji} - b_j]C_{ik}/b_i \tag{42}$$

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

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

$$E_L(R) = V(R) + \lambda \sum_i [\nabla_i^2 U - \sum_k \nabla_i^2 \theta_k(r_i)C_{ki} - G_i^2], \tag{43}$$

where $G_i = -\nabla_i U + \sum_k \nabla_i \theta_k(r_i)C_{ki}$, and $U = \sum u(r_{ij})$. 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 $\nabla_i^2 \theta_k(r_i)$, will simplify. Finally note that using Green's identity allows several alternative ways[9] 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. (43) 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. (Exercise)

2.3 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[23].

- The dispersion of the local energy: $\sigma^2 \int [(\mathcal{H}-E_V)\Psi]^2$. 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.

$$-\lambda\nabla^2\theta_k(\mathbf{r}) + v_m(r) - e_k)\theta_k(\mathbf{r}) = 0 \tag{44}$$

Let us examine the dominant terms in the local energy eq. (43) as 2 particles are brought together. In a good trial function the singularities in the kinetic energy must cancel the the singularities of the potential energy. In this limit the local energy will have the form:

$$E_L(R) = v(r) + 2\lambda\nabla^2 u(r) - 2\lambda(\nabla u(r))^2 + \ldots, \tag{45}$$

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: $u(r) = ((2\epsilon\sigma^2)/(25\lambda))^{1/2}(\sigma/r)^5$. For the Coulomb potential this equation can be used to derive the cusp condition.

$$e_i e_j + 2(\lambda_i + \lambda_j)\frac{u_{ij}(0)}{dr_{ij}} = 0 \tag{46}$$

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:

$$E_V = E_F + \sum_k (v_k - \lambda k^2 u_k)(S_k - 1) \tag{47}$$

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[13] the optimal pseudopotential at long wavelengths:

$$2\rho u_k = -\frac{1}{S_{0k}} + \left[\frac{1}{S_{0k}} + \frac{2\rho v_k}{\lambda k^2}\right]^{1/2}. \tag{48}$$

For a short-ranged potential, (e.g. liquid helium), v_k can be replaced by a constant and and we find the Reatto-Chester[14] 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 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[2, 13] 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[20, 9] 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: $|\Psi_t(R; a_0)|^2$. Then the variational energy (or variance) for another trial function can be calculated by using the same set of points:

$$E_v(a) = \frac{\sum_j w(R_j, a) E_L(R_j, a)}{\sum_j w(R_j, a)}, \tag{49}$$

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

2.4 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 three- body (or polarization) term with the functional form of a squared force:

$$U_3(R) = -\sum_i \left[\sum_j \xi(r_{ij}) \vec{r_{ij}} \right]^2 . \tag{50}$$

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.

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

Table 1: The energies of liquid ^{4}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. 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 0.01K. The numbers are from Moroni et al. [19]

$terms$	E_V	$[E_V - E_0]/[2T]$	E_{FN}	$ref.$
u	$-1.15(4)$	5.7%	$-2.20(3)$	[16]
$u - optimized$	$-1.30(3)$	5.7%	$-2.20(3)$	[19]
u, ξ	$-1.780(17)$	4.6%	$-2.20(3)$	[16]
u, η	$-1.730(4)$	3.7%	$-2.37(1)$	[16]
u, ξ, η	$-2.163(6)$	1.3%	$-2.37(5)$	[17]
$exp.$	-2.47	0.0%	-2.47	

Table 2: 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. BCS refers to the spin-paired trial functions in eqs. (14-15). 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 with the Fixed-Node Green's Function Monte Carlo to be described in the next section. The numbers in parenthesis are the statistical error in units of 0.01K.

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:

$$det[\theta_k(\vec{s}_i, \sigma_i)], \tag{51}$$

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

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.

2.5 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 helium 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 1 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 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! This allows an element of human bias; the optimization is stopped when the expected result is obtained.

- 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.

3 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 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 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 $\psi_0(R)$:

$$\psi_{n+1}(R) = e^{-\tau(\mathcal{H}-E_T)}\psi_n(R) = \int dR' G(R, R')\psi_n(R'). \tag{52}$$

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:

$$\psi_n(R) = \sum_\alpha \phi_\alpha(R) < \phi_\alpha|\psi_0 > e^{-n\tau(E_\alpha - E_T)}. \tag{53}$$

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:

$$\lim_{n\to\infty} \psi_n(R) = \phi_0(R) < \phi_0|\psi_0 > e^{-n\tau(E_0 - E_T)}. \tag{54}$$

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 one N gets larger than a few, one must do the integral with Monte Carlo. The interpretation of Eq. (52) 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 $G(R_1, R_0)$. In the limit that the time step approaches zero, a coordinate space representation of the Green's function is:

$$< R|e^{-\tau(\mathcal{H}-E_T)}|R' >= (4\pi\lambda\tau)^{-3N/2} e^{-\frac{(R-R')^2}{4\lambda\tau}} e^{-\tau(V(R)-E_T)} \; + \; O(\tau^2), \tag{55}$$

The iteration equation, eq. (52) , 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. (55) and then branching. The number of copies of R' in the next generation is the integer part of

$$m = u + \exp[-\tau(V(R) - E_T)] \tag{56}$$

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

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*	$[1-\tau(\mathcal{H}-E_T)]$	[28]

Table 3: 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 at the origin and to have the same derivative.

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.

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:

$$E_T = E_0 + \kappa \ln(P^*/P), \tag{57}$$

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 it is too large, one can bias the distribution.

3.1 Importance Sampling

The above scheme, first suggested by Fermi, was actually tried out in the first days of computing some forty years ago [29]. 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 [25] 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. (52) by Ψ, the trial function, and define $f_n(R) = \Psi(R)\psi_n(R)$. Then:

$$f_{n+1} = \Psi e^{-\tau(\mathcal{H}-E_T)}\psi_n = \int dR' \tilde{G}(R,R') f_n(R') \tag{58}$$

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 $f_0(R) = \Psi^2(R)\Psi(R)\psi_0(R)$. It is easily shown by differentiating $\tilde{G}$ with respect to τ that it satisfies the evolution equation:

$$-\frac{\partial \tilde{G}(R,R_0;\tau)}{\partial \tau} = -\sum_i \lambda_i \nabla_i [\nabla_i \tilde{G} + 2\tilde{G}\nabla_i \ln(\Psi(R))] + [E_L(R) - E_T]\tilde{G}, \tag{59}$$

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 sufficiently 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:

$$R_{n+1} = R_n + \chi + \lambda\tau\nabla\ln(\Psi(R_n)^2), \tag{60}$$

where χ is a normally distributed random vector with variance $2\lambda\tau$ and zero mean.

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

$$\exp(-\tau(E_L(R_n) - E_T)) + u, \tag{61}$$

where u is a uniformly distributed random number in $(0, 1)$.

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 Eq. (57).
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,' $V_{mix} =< \phi_0|V|\Psi >$, and the variational estimator [20]. For example the potential energy is calculated as:

$$< \phi_0|V|\phi_0 >= 2 < \phi_0|V|\Psi > - < \Psi|V|\Psi > +\mathcal{O}([\phi_0 - \Psi]^2). \tag{62}$$

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.

3.2 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 $\phi_0(R)$ 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, $\tilde{G}(R, R')$ 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[30]; 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. See Table 1.

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 $\phi(R) = 0$. 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[31]. 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. [31].

3.3 Exact Fermion 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.

The transient estimate (TE) method calculates the ratio:

$$E_{TE}(t) = \frac{\int \Psi \mathcal{H} e^{-t(\mathcal{H}-E_T)} \Psi}{\int \Psi e^{-t(\mathcal{H}-E_T)} \Psi} \tag{63}$$

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 $E_{TE}(t) \geq E_0$. Also the energy converges exponentially fast in t:

$$\lim_{t\to\infty} E_{TE}(t) = E_0 + O(e^{-tE_g}) \tag{64}$$

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. (59) 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. (59) and will be assumed to be a reasonable boson ground state trial function, while the trial function appears in eq. (63). The Ψ_G importance-sampled Green's function is:

$$\tilde{G}(R,R';t) = \Psi_G(R) < R|e^{-t(\mathcal{H}-E_T)}|R' > \Psi_G^{-1}(R'), \tag{65}$$

and we can rewrite eq. (63) as:

$$E_{TE}(t) = \frac{\int \sigma(R) E_{LT}(R)\tilde{G}(R,R';t)\sigma(R')\Psi_G^2(R')}{\int \sigma(R)\tilde{G}(R,R';t)\sigma(R')\Psi_G^2(R')} \tag{66}$$

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.

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. (66).

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

3. Propagate the walk forward an amount of time, t with the Green's function, $\tilde{G}(R,R';t)$. 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 t is computed as:

$$E_{TE}(t) = \frac{< [E_{LT}(R) + E_{LT}(R')]\sigma(R)\sigma(R') >}{2 < \sigma(R)\sigma(R') >}, \tag{67}$$

   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[30, 32] 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. (63) 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 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 $\exp(-t(E_B - E_T))$ 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:

$$\epsilon_{stat} = e^{-t(E_F - E_B)}. \tag{68}$$

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. (64,68) we find the total error will decrease as:

$$\epsilon \quad \propto \quad P^{-\eta} \qquad \eta = \frac{E_g}{2(E_F - E_B + E_g)}. \tag{69}$$

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[33] 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[34] 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' 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 terms 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 $\mathcal{O}(N)$. 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 $\mathcal{O}(N)$. Variational MC calculations of fermion systems are $calo(N^3)$ in general, but the exponent would be smaller if localized spin-orbits are used. The Hubbard model at half filling on a bipartite lattice[35] 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 Lattice models

Let me briefly discuss the application of these methods to lattice problems. Most of the methods described here work also for lattice models, for example variational Monte Carlo 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:

$$\tau \leq \frac{2}{E_{max} - E_T} \tag{70}$$

where E_{max} is the maximum energy to avoid negative matrix elements. 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. [28].

The conditions on a lattice model not to have a sign problem are easy to state. The 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.

$$< s|\mathcal{H}|s' > \leq \ 0 \quad \forall \ \ s \neq s'. \tag{71}$$

(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. As far as I know, there has not been a systematic search through local basis transformations to see if some of the other interesting lattice models might be solvable.

The fixed-node 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 2 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, van Bemmel and co-workers [37] 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:

$$V_{eff}(s) = V(s) + \sum_{s' \in SF} < s|\mathcal{H}|s' > \Psi(s')/\Psi(s) \tag{72}$$

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 common methods for performing simulations of lattice models. These are based on applying the Stratonovitch-Hubbard transformation[35] 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.

3.5 Complex Wavefunctions

In this section I briefly mention the generalization of these methods to situations where the wavefunction is necessarily complex. Variational methods are straightforward in principle: one simply samples the square of the modulus of the wavefunction. See, for example [38]. 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 fixed by a variational wavefunction 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:

$$V_{eff}(R) = V(R) + \lambda = \sum_i [\nabla_i \phi(R) + \mathbf{A}(\mathbf{r}_i)]^2 \tag{73}$$

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. [39]. An application to a vortex in superfluid helium is discussed in ref. [40]

3.6 Treatment of Atomic Cores in QMC

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 $\approx Z^6$ 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 two 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 [42] 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.

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.* [43], 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 virtues of the DMC method immediately apply. For example, the fixed-node approximation 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 [44].

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:

$$H_{val} = H_{loc} + W \tag{74}$$

with the local part given by:

$$H_{loc} = \sum_i \left[-\frac{1}{2}\nabla_i^2 + \sum_I v_{loc}(r_{iI}) + \frac{1}{2}\sum_{j\neq i}\frac{1}{r_{ij}} \right]. \tag{75}$$

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

$$\langle R|W|R'\rangle = \sum_{I,i}\sum_{\ell} \frac{2\ell+1}{4\pi} v_\ell(r_{iI}) \frac{\delta(r_{iI} - r'_{iI})}{r_{iI} r'_{iI}} P_\ell(\hat{\mathbf{r}}_{iI}\cdot\hat{\mathbf{r}}'_{iI}) \tag{76}$$

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, and Fahy *et al* [52, 32] 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[45] and by Hammond *et al.* [46] to define a new transformed effective core potential by a projection onto a trial function

$$V_{eff}(R) = \Psi_T^{-1}(R) \int dR' \langle R|W|R'\rangle \Psi_T(R) \tag{77}$$

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 [47], silicon and carbon clusters [49, 50], nitrogen solids [48]. 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. [41, 51]. A recent book on the subject is ref. [52].

3.7 Problems with Projection 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. In 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 time 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, with 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 state. 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 t is the release node projection time, then we can move the nodes a distance of at most $\sqrt{6N\lambda t}$.

- 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 Path Integral Monte Carlo

This section is an abridged version of the longer review article "Path integrals in the theory of condensed helium" [53].

In the rest of this review I will discuss what I consider to be the the most powerful quantum simulation method: path integral Monte Carlo. First consider what we must do to simulate a quantum system at finite temperature. We must sample an energy eigenstate with probability proportional to the Maxwell-Boltzmann distribution: $exp(-\beta E_i)$ and then sample the spatial distribution from $|\psi(R)|^2$. A simpler procedure is to recall the properties of the thermal density matrix. Recall that all static properties (in principle, dynamical properties also) of a quantum system in thermal equilibrium are obtainable from the thermal density matrix. If this sounds unfamiliar, the reader might wish to review the material in Feynman[54]. In this section, we detail the basic mathematical properties of the density matrix, give the relationship between the density matrix, path integrals, and the statistical mechanics of classical "polymers," explain how Bose symmetry is expressed with path integrals and fix the notation and terminology of our description. In thermal equilibrium, the probability of a given state i being occupied is $e^{-E_i/k_B T}$, with T the temperature. Hence the equilibrium value of an operator O is

$$\langle O \rangle = Z^{-1} \sum_i < \phi_i | O | \phi_i > e^{-\beta E_i} \tag{78}$$

where the partition function is

$$Z = \sum_i e^{-\beta E_i} \tag{79}$$

and $\beta = 1/k_B T$. The position-space density matrix is

$$\begin{aligned} \rho(R, R'; \beta) &= < R | e^{-\beta \mathcal{H}} | R' > \\ &= \sum_i \phi_i^*(R) \phi_i(R') e^{-\beta E_i}, \end{aligned} \tag{80}$$

In the position representation, the expectation of O becomes

$$\langle O \rangle = Z^{-1} \int dR dR' \rho(R, R'; \beta) < R|O|R' > \tag{81}$$

The following simple, exact property of density matrices is the basis of the path-integral method. The product of two density matrices is a density matrix:

$$e^{-(\beta_1+\beta_2)\mathcal{H}} = e^{-\beta_1 \mathcal{H}} e^{-\beta_2 \mathcal{H}}. \tag{82}$$

Written for positions, one has a convolution,

$$\rho(R_1, R_3; \beta_1 + \beta_2) = \int dR_2 \rho(R_1, R_2; \beta_1) \rho(R_2, R_3; \beta_2). \tag{83}$$

The path-integral formula for the many-body density matrix is arrived at by using the product property M times, giving an expression for the density matrix at a temperature T, in terms of density matrices at a temperature MT. In operators,

$$e^{-\beta \mathcal{H}} = \left(e^{-\tau \mathcal{H}}\right)^M, \tag{84}$$

where the ***time step*** is $\tau = \beta/M$. Written in the position representation,

$$\rho(R_0, R_M; \beta) = \int \ldots \int dR_1 dR_2 \ldots dR_{M-1} \rho(R_0, R_1; \tau) \rho(R_1, R_2; \tau) \ldots \rho(R_{M-1}, R_M; \tau). \tag{85}$$

If M is finite we have a discrete-time path. If the limit $M \to \infty$ is taken, one has a continuous path $\{R_t\}$ where $0 \leq t \leq \beta$. But note that Eq. (85) is ***exact for any*** $M \geq 1$. The second property that is needed by path integrals is that, for MT large enough, we can write down a sufficiently accurate approximation to the density matrix. Thus we shall be able to write down an explicit form for the low-temperature density matrix which, however, involves many additional integrals. Suppose the Hamiltonian is split into two pieces, $\mathcal{H} = \mathcal{T} + \mathcal{V}$, where $\mathcal{T}$ and $\mathcal{V}$ are the kinetic and potential operators.

We can approximate the exact density matrix by the product of the density matrices for $\mathcal{T}$ and $\mathcal{V}$ alone. One might worry that this will lead to an error in the limit as $M \to \infty$ with small errors building up to a finite error. According to the Feynman-Kacs or Trotter formula[55], one does not have to worry:

$$e^{-\beta(\mathcal{T}+\mathcal{V})} = \lim_{M\to\infty} \left[e^{-\tau\mathcal{T}} e^{-\tau\mathcal{V}}\right]^M. \tag{86}$$

Let us now write the primitive approximation in position space,

$$\rho(R_0, R_2; \tau) \approx \int dR_1 \left\langle R_0 | e^{-\tau\mathcal{T}} | R_1 \right\rangle \left\langle R_1 | e^{-\tau\mathcal{V}} | R_2 \right\rangle \tag{87}$$

and evaluate the kinetic and potential density matrices. Since the potential operator is diagonal in the position representation, its matrix elements are trivial:

$$\left\langle R_1 | e^{-\tau\mathcal{V}} | R_2 \right\rangle = e^{-\tau V(R_1)} \delta(R_2 - R_1). \tag{88}$$

The kinetic matrix can be evaluated using the eigenfunction expansion of $\mathcal{T}$. For the moment, consider the case of distinguishable particles in a cube of side L with periodic boundary conditions.

Then the exact eigenfunctions and eigenvalues of $\mathcal{T}$ are $L^{-3N/2}e^{iK_{\mathbf{n}}R}$ and $\lambda K_{\mathbf{n}}^2$, with $K_{\mathbf{n}} = 2\pi\mathbf{n}/L$ and $\mathbf{n}$ a $3N$-dimensional integer vector. Then

$$\begin{aligned} \left\langle R_0|e^{-\tau\mathcal{T}}|R_1\right\rangle &= \sum_{\mathbf{n}} L^{-3N} e^{-\tau\lambda K_n^2 - iK_n(R_0-R_1)} && (89) \\ &= (4\pi\lambda\tau)^{-3N/2} \exp\left[-\frac{(R_0-R_1)^2}{4\lambda\tau}\right]. && (90) \end{aligned}$$

Equation (90) is obtained by approximating the sum by an integral. This is appropriate only if the thermal wavelength of one step is much less than the size of the box,

$$\lambda\tau \ll L^2. \tag{91}$$

Using Eqs. (85), (87), (88),and (90) we arrive at the discrete path-integral expression for the density matrix in the primitive approximation:

$$\rho(R_0, R_M; \beta) = \int dR_1 \ldots dR_{M-1} (4\pi\lambda\tau)^{-3NM/2} \exp\left(-\sum_{m=1}^{M}\left[\frac{(R_{m-1}-R_m)^2}{4\lambda\tau} + \tau V(R_m)\right]\right). \tag{92}$$

This expression relates the quantum density matrix at any temperature to integrals over the path $R_1 \ldots R_{M-1}$ of something that is like a classical Maxwell-Boltzmann distribution function. This is the famous mapping from a quantum system to a classical system. The Feynman-Kacs formula, to be used later, is obtained by taking the limit $M \to \infty$, making a continuous path.

Of particular importance for the Monte Carlo evaluation is the following corollary of the convolution property: if the density matrix is non-negative for any time step τ, by which we mean $\rho(R_1, R_2; \tau) \geq 0\ \forall\ (R_1, R_2)$, then the density matrix is non-negative for all positive multiples of τ. But we see that the density matrix in the primitive approximation is non-negative, so that the density matrix at all temperatures must be non-negative.

All the approximations are controllable. The price we have to pay for having an explicit expression for the density matrix is additional integrations; altogether $3N(M-1)$. Without techniques for multidimensional integration, nothing would have been gained by expanding the density matrix in a path. Fortunately, simulation methods can accurately treat such integrands. Since we have a non-negative integrand [see Eq. (92)] the time to do a Monte Carlo calculation (with a predefined error) will scale roughly linearly with the number of integrals. It is feasible to make M rather large, say in the hundreds or thousands, and thereby systematically reduce the time step error.

The ***time step*** is defined as:

$$\tau \equiv \beta/M \tag{93}$$

and a single R_k is referred to as the k^{th} ***time slice***. Again R_k represents the $3N$ positions of the N particles: $R_k = \{\mathbf{r}_{1,k}, \ldots, \mathbf{r}_{N,k}\}$ and $\mathbf{r}_{i,k}$, ***a bead***, is the position of the i^{th} particle in the k^{th} time slice. The ***path*** is the sequence of points $\{R_0, R_1, \ldots, R_{M-1}, R_M\}$. The ***time*** associated with the point R_k is defined as $t_k = k\tau$.

A ***link m*** is a pair of time slices (R_{m-1}, R_m) separated by time τ. The ***action*** of a link is defined as ***minus*** the logarithm of the ***exact*** density matrix:

$$S^m \equiv S(R_{m-1}, R_m; \tau) \equiv -\ln[\rho(R_{m-1}, R_m; \tau)]. \tag{94}$$

Then the (exact) path-integral expression becomes

$$\rho(R_0, R_M; \beta) = \int dR_1 \ldots dR_{M-1} \exp\left[-\sum_{m=1}^{M} S^m\right]. \tag{95}$$

There will be contributions to S^m coming from each term of the Hamiltonian. It is convenient to separate out the kinetic action from the rest of the action. The exact ***kinetic action*** for link m will be denoted K^m,

$$K^m = \frac{3N}{2}\ln(4\pi\lambda\tau) + \frac{(R_{m-1} - R_m)^2}{4\lambda\tau}. \tag{96}$$

The ***inter-action*** is then defined as what is left:

$$U^m = U(R_{m-1}, R_m; \tau) = S^m - K^m. \tag{97}$$

The approximation [Eq. (87)] of allowing the kinetic and potential energies to commute will be called the ***primitive approximation.*** In the primitive approximation, the inter-action is

$$U_1^m = \frac{\tau}{2}[V(R_{m-1}) + V(R_m)]. \tag{98}$$

We have symmetrized U_1^m with respect to R_m and R_{m-1}.

We can interpret the path-integral expression, Eq. (92), as a classical configuration integral; the action is analogous to a classical potential-energy function divided by k_BT. In the classical analog, the kinetic link action corresponds to a spring potential connecting beads representing the same atom in successive time slices. The classical system is a chain of beads connected with springs. We call such a chain a ***polymer.*** In fact, the bead-spring model of real-life polymers has had a long and useful history. The potential action represents forces between beads of different atoms, keeping the polymers out of each other's way (for a repulsive potential). The potential is represented by an inter-polymeric, potential which is peculiar from the classical point of view in that it interacts only at the same "time" and only between beads on different chains.

Thermodynamical properties, or static properties diagonal in configuration space, are determined by the trace of the density matrix, *i.e.*, the integral of Eq. (92) over R_0 with $R_0 = R_M$. The formula for diagonal elements of the density matrix then involves a path that returns to its starting place after M steps: a ***ring polymer.***

Because the partition function of the quantum system is equal to the partition function of the classical system, and because of the central importance of the partition function in statistical mechanics, there is an exact, systematic procedure for understanding many properties of quantum systems purely in terms of classical statistical mechanics.

The same word applied to the quantum system and the classical system can mean quite different things. To further avoid confusion we do not refer to the "energy" of the polymer model, but to its action. Another confusing term is entropy. The entropy of a quantum system decreases with temperature. But at low temperature, the corresponding polymer system is becoming more disordered. The confusion arises because the "temperature" of the polymer model is not equal to the quantum temperature.

To translate what we mean by temperature into the polymer model we must find how β appears in the action. It is best not to see how the time step appears in the action because the time step is fixed by requiring that the action be accurate. Hence the spring constant and the interbead potential should be fixed as temperature varies. This means that β will be proportional to the number of time slices. The lower the temperature, the more beads on the polymer. Zero temperature corresponds to infinitely long chains. One might worry that sooner or later space will be completely filled by beads. This is not a problem because only beads at the same "time" interact, and hence any given bead always sees N other beads. Time is a word that can have at least three different meanings: real time in the quantum system, the "imaginary time" of the path integrals, and the time related to how the path is moved in the computer program. We shall call this last time, steps, moves, or sweeps. If we confuse the first two meanings of time, a word can have exactly the opposite meaning in the quantum and polymer systems. For example, the "velocity" of a bead is usefully defined as its displacement from one time slice to the next, divided by τ. But with this definition atoms that are "fast" correspond to low-energy atoms, because they are spread out and their kinetic energy is small. On the other hand, particles that are trapped in a small region have a small "velocity" and a high energy. The inversion of meaning comes because path integrals are in imaginary time. The kinetic energy in the primitive approximation is

$$\langle \mathcal{T} \rangle = \frac{3N}{2\tau} - \frac{m}{2} \left\langle \left(\frac{R_i - R_{i-1}}{\hbar\tau} \right)^2 \right\rangle . \tag{99}$$

Kinetic energy is a constant ***minus*** the square of the "velocity." The constant needs to be there so that the total kinetic energy will always be positive. It is possible for a single realization of a path to have a negative kinetic energy, by being spread out more than usual, but the average over all paths must be positive.

4.1 Path Integrals and superfluidity

The density matrices up to this point have been appropriate to distinguishable (Boltzmann) particle statistics, since the indistinguishability of particles was not taken into account. For Bose systems only totally symmetric eigenfunctions $\phi_i(R)$ contribute to the density matrix; those such that $\phi_i(PR) = \phi_i(R)$ where P is a permutation of particle labels, *i.e.*, $PR = (\mathbf{r}_{P_1}, \mathbf{r}_{P_2} \ldots \mathbf{r}_{P_N})$. Define the particle symmetrization operator

$$\mathcal{P}\phi(R) = \frac{1}{N!} \sum_P \phi(PR). \tag{100}$$

If the Hamiltonian is symmetric under particle exchange, all states are either even or odd with respect to a given permutation. Then $\mathcal{P}$ will project out Bose states. If we apply $\mathcal{P}$ to the density matrix, we will obtain the bosonic density matrix. Written in position space this is

$$\rho_B(R_0, R_1; \beta) = \frac{1}{N!} \sum_P \rho(R_0, PR_1; \beta) \tag{101}$$

where ρ_B is the boson density matrix and ρ is the boltzmannon density matrix.

A straightforward evaluation of the permutation sum is out of the question once N gets large, since there will be $N!$ terms. Fortunately, each term in the sum is positive, so we can ***sample*** the permutations in the sum. A bosonic simulation consists of a random walk through the path space ***and*** the permutation space. For Fermions the cancellation between the contributions of even and odd permutations generally rules out a Monte Carlo evaluation of the integrand without some major modification, to be discussed in the last section.

The partition function for a Bose system has the form

$$Z_B = \frac{1}{N!} \sum_{\mathcal{P}} \int dR_0 \ldots dR_{M-1} \exp\left(- \sum_{m=1}^{M} S^m \right), \tag{102}$$

with new boundary conditions on path closure: $PR_m = R_0$. Paths are allowed to close on any permutation of their starting positions. The partition function includes contributions from all $N!$ closures. At high temperature the identity permutation dominates, while at zero temperature all permutations have equal contributions. In the classical isomorphic system, ring polymers can "cross-link." (We only mean to be suggestive: cross-linking of real polymers is quite different.) A two-atom system of M links can be in two possible permutation states: either two separate ring polymers, each with M links, or one larger polymer with $2M$ links.

Any permutation can be broken into a product of cyclic permutations. Each cycle corresponds to several polymers "cross-linking" and forming a larger ring polymer. Quantum mechanically the liquid does this to lower its kinetic energy. In the classical language, cross-linking takes place to maximize the "entropy"; there are many more cross-linked configurations than non-cross-linked ones. According to Feynman's 1953 theory[56] the superfluid transition is represented in the classical system by the formation of macroscopic polymers, *i.e.* those stretching across an entire system and involving on the order of N atoms. What we shall see in the following sections is the explicit dependence of superfluid properties on these macroscopic exchanges. ***Monomers*** are atoms not involved in an exchange-atoms i such that $P_i = i$. We shall find that the average monomer density is directly related to the free energy of an isotopic impurity.

In the absence of interaction, the size of a path (or polymer) is its thermal wavelength,

$$\Lambda_\beta = (2\beta\lambda)^{1/2}. \tag{103}$$

When the size of the polymer equals the interpolymer spacing, roughly $\rho^{-1/d}$, it is at least possible for the polymers to link up by exchanging end points. This relationship, $\Lambda_\beta = \rho^{-1/d}$, defines the degeneracy temperature

$$T_D = \frac{\rho^{2/d}\hbar^2}{mk_B}. \tag{104}$$

For temperatures higher than T_D, quantum statistics (either bosonic or Fermionic) are not very important.

In a liquid state, T_D gives a surprisingly good estimate of the superfluid transition temperature. For ideal Bose condensation in three dimensions, $T_c/T_D = 3.31$. For liquid ^{4}He at saturated-vapor-pressure (SVP) conditions (essentially zero pressure), $T_c/T_D = 2.32$.

Qualitatively, one can understand why there will be a phase transition when the temperature is low enough. From Feynman[56]: "A single large polygon of r sides contributes a very small amount y^r with $y < 1$. But a large polygon can be drawn in more ways than a small one. Increasing the length r by one increases the number of polygons available by a factor say s (perhaps 3 or 4) although the contribution of each is multiplied by y. Thus if $sy < 1$ (high T) large polygons are unimportant. As T falls, suddenly when $sy = 1$ the contributions from very large polygons (limited by the size of the container) begin to be important. This produces a transition."

4.2 The momentum distribution

London supposed the superfluid transition to be the analog of the transition that occurs in an ideal Bose gas, where below the transition, a finite fraction of particles occupy the zero-momentum state. It is hard to understand how particles with strong repulsive interactions could behave like free particles. Penrose and Onsager[57] defined Bose condensation in an interacting system as the macroscopic occupation of a single-particle state, namely the state of zero momentum. Using Feynman's partition function and arguments concerning cycle length distribution, they showed that there would be Bose condensation below T_c but not above. They estimated that at zero temperature 8% of the atoms have precisely zero momentum.

The condensate fraction has a simple meaning in terms of path-integrals. The probability density of observing a single atom with momentum k is defined as

$$n_{\mathbf{k}_1} = (2\pi)^{-Nd} \int d\mathbf{k}_2 \ldots d\mathbf{k}_N \left| \int dR\phi(R)e^{-iKR} \right|^2, \tag{105}$$

where $\phi(R)$ is the many-body wave function. If we perform the integrals $d\mathbf{k}_2 \ldots d\mathbf{k}_N$ and thermally occupy the many-body states we find

$$n_{\mathbf{k}} = \frac{1}{\Omega(2\pi)^d} \int d\mathbf{r}_1 d\mathbf{r}'_1 e^{-i\mathbf{k}(\mathbf{r}_1 - \mathbf{r}'_1)} n(\mathbf{r}_1, \mathbf{r}'_1), \tag{106}$$

where the single-particle density matrix is

$$n(\mathbf{r}_1, \mathbf{r}'_1) = \frac{\Omega}{Z} \int d\mathbf{r}_2..d\mathbf{r}_N \rho(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, \mathbf{r}'_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; \beta). \tag{107}$$

According to Eq.(106), the momentum distribution is the Fourier transform of an off-diagonal element of the density matrix. The paths that we have been discussing up to this point, each ending at the start of another particle's path, cannot be used to calculate the momentum distribution. Simply put, to get an observable in momentum space we cannot do the simulation entirely in the position representation. The method by which to calculate the single-particle density matrix is quite simple: one samples paths from the probability distribution,

$$\pi_n(R, \mathbf{r}'_1) = \frac{1}{Z'} \rho(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, \mathbf{r}'_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; \beta), \tag{108}$$

where Z' is a new normalization constant and $\mathbf{r}$ and $\mathbf{r}'$ are independent variables. This density matrix is expanded into a path. We were careful when we defined the path-integrals to do it for a general (off-diagonal) matrix element. Then the distribution of $\mathbf{r}_1$ and $\mathbf{r}'_1$ is given by

$$n(\mathbf{r}, \mathbf{r}') \propto \left\langle \delta(\mathbf{r}_1 - \mathbf{r}) \delta(\mathbf{r}'_1 - \mathbf{r}') \right\rangle_{\pi_n} \tag{109}$$

where the brackets denote an average over π_n. The classical simulation to be performed is of $(N-1)$ ring polymers and 1 linear polymer.

At high temperature there is no particle exchange and the distance between the polymers is much greater than the size of a given polymer, so the internal coordinates of the single linear polymer will be almost free-particle like and its end-to-end distribution Gaussian: $n(\mathbf{r}, \mathbf{r}') \propto \exp[-(\mathbf{r} - \mathbf{r}')^2/(4\lambda\beta)]$. Taking the Fourier transform, we end up with the Maxwellian momentum distribution with a width $k_B T$.

Now we have to consider how Bose statistics affects the types of paths that are allowed. Care must be taken to understand the imaginary-time boundary conditions once permutations are present. Suppose particle 1 is involved in a three-body cyclic permutation with particles 2 and 3. We know that particle 1 begins at $\mathbf{r}$ and ends at $\mathbf{r}'$. That means one has the following boundary conditions on the paths:

$$\begin{aligned} \mathbf{r} &= \mathbf{r}_1(0), \\ \mathbf{r}_1(\beta) &= \mathbf{r}_2(0), \\ \mathbf{r}_2(\beta) &= \mathbf{r}_3(0), \\ \mathbf{r}_3(\beta) &= \mathbf{r}'. \end{aligned} \tag{110}$$

It is simpler to state the conditions physically. There are two cut ends in the path space, but it does not matter which particle labels are attached to the ends. If a macroscopic exchange is present, as is usually the case in the superfluid state, the two ends can become separated by much more than a thermal wavelength if they are attached to a macroscopic exchange. How far they become separated depends on the statistical mechanics of the polymer system and is different for bulk ^{4}He and for ^{4}He films (*i.e.*, in 2D or 3D).

For a 3D bulk liquid the single particle density matrix in the superfluid state goes to a constant at large r. The momentum distribution, its Fourier transform, will then have a delta function at the origin. We define the condensate fraction as the probability of finding an atom with precisely zero momentum. This will equal

$$\begin{aligned} \tilde{n}_0 = \frac{(2\pi)^3}{\Omega} n_0 &= \frac{1}{\Omega^2} \int d\mathbf{r} d\mathbf{r}' n(\mathbf{r}, \mathbf{r}') \\ &= \frac{1}{\Omega} \int d\mathbf{r} n(r). \end{aligned} \tag{111}$$

The factor $(2\pi)^3 \Omega^{-1}$ comes about because n_0 is a probability density, while $\tilde{n}_0$ is a probability. The last equation holds for a homogeneous liquid. If we take the volume of the box to infinity, the condensate fraction is the large- distance limit of the single-particle density matrix,

$$\tilde{n}_0 = \lim_{r \to \infty} n(r). \tag{112}$$

The condensate fraction is essentially the probability of the two cut ends attaching themselves to a macroscopic exchange.

In ^{4}He films the momentum distribution is quite different at small momentum. At a nonzero temperature the two cut ends never lose sight of each other. They feel an attraction to each other which varies like $\eta ln(|\mathbf{r} - \mathbf{r}'|)$ at large separations. Hence the single-particle density matrix decays to zero algebraically: $n(r) \propto r^{-\eta}$. The strength of this interaction depends on the temperature through the Kosterlitz-Thouless relation. $\eta^{-1} = 4\pi\lambda\beta\rho_s$, where ρ_s is the superfluid density. Hence a nonzero condensate only appears at zero temperature. Nonetheless the system is superfluid below its transition temperature.

4.3 Response to rotation and the superfluid density

Superfluidity is experimentally characterized by the response of a system to movements of its boundaries. The rotating bucket experiment was first discussed by Landau on the basis of his theory of superfluidity. He predicted that superfluid helium would show an abnormal relation between the energy it takes to spin a bucket and its moment of inertia. Suppose one measures the work needed to bring a container filled with helium to a steady rotation rate. A normal fluid in equilibrium will rotate rigidly with the walls. The work done is $E = \frac{1}{2} I \omega^2$, where I is the momentum of inertia and ω is the angular rotation rate. On the other hand, a superfluid will stay at rest if the walls rotate slowly, so that a smaller energy is needed to spin up the container. The liquid that stays at rest is the superfluid.

The effective moment of inertia is defined as the work done for an infinitesimally small rotation rate,

$$I = \left. \frac{dF}{d\omega^2} \right|_{\omega=0} = \left. \frac{d < \mathcal{L}_z >}{d\omega} \right|_{\omega=0} \tag{113}$$

where F is the free energy, $\mathcal{L}_z$ is the total angular momentum operator in the $\hat{z}$ direction,

$$\mathcal{L}_z = i\hbar \sum_{i=1}^{N} \frac{\partial}{\partial \theta_i} \tag{114}$$

and θ_i is the angle of the ith particle in cylindrical coordinates. On the other hand, the classical moment of inertia is given by

$$I_c = < \sum_{i=1}^{N} m_i (\mathbf{r}_i^{\perp} \times \hat{z})^2 > . \tag{115}$$

The ratio of the two moments is defined as the normal density; what is missing is the superfluid density:

$$\frac{\rho_n}{\rho} = 1 - \frac{\rho_s}{\rho} = \frac{I}{I_c}. \tag{116}$$

Thus the superfluid density is the linear response to an imposed rotation, just as the electrical conductivity is the response to an imposed voltage.

One might not think that imaginary-time path-integrals would be appropriate to calculate the superfluid density, since motion in real time is involved. This is not so. Statistical mechanics does not require the use of an inertial reference frame. We can transform to the frame rotating with the bucket to determine the free energy of rotation. The Hamiltonian in the rotating coordinate system is simply given by

$$\mathcal{H}_\omega = \mathcal{H}_0 - \omega \mathcal{L}_z. \tag{117}$$

Here $\mathcal{H}_0$ is the Hamiltonian at rest. We pick up the extra term in transforming the Schroedinger equation from the laboratory frame to the rotating frame, since the new angle is given by $\theta' = \theta - \omega t$. Now we have to find a path-integral expression for the effective moment of inertia defined in Eq. (113). The following identity allows us to take the derivative of an exponential operator that contains a parameter ω. First we break up the exponential into M pieces:

$$\frac{de^A}{d\omega} = \sum_{k=1}^{M} e^{(k-1)A/M} \frac{de^{A/M}}{d\omega} e^{(M-k)A/M}. \tag{118}$$

Now we take the limit $M \to \infty$:

$$\frac{de^A}{d\omega} = \int_0^1 dt e^{tA} \, \frac{dA}{d\omega} \, e^{(1-t)A}. \tag{119}$$

The first equation is appropriate to discrete-time path-integrals, the second should be familiar from linear-response theory. Of course, if the derivative $\frac{dA}{d\omega}$ commutes with A, things are much simpler. We do not want to assume that the potential is invariant with respect to rotations so that the angular momentum operator does not commute with the Hamiltonian.

Now let us take the derivative of the rotating density matrix with respect to ω, as required by Eq. (113). We get

$$\frac{\rho_n}{\rho} = \frac{I}{I_c Z} tr \left[\int_0^\beta dt \mathcal{L} e^{-(\beta - t)\mathcal{H}} \mathcal{L} e^{-t\mathcal{H}} \right], \tag{120}$$

We have expressed the normal fluid density in terms of the matrix elements involving the system at rest. Now we explicitly evaluate this in terms of discrete path-integrals by having the angular momentum operate on the action. Since angular momentum commutes with the internal potential energy, that term will not contribute. After some algebra we get

$$\frac{\rho_s}{\rho} = \frac{2m < A_z^2 >}{\beta \lambda I_c}, \tag{121}$$

where we have defined two functions of a given path, namely the projected area

$$\mathbf{A} = \frac{1}{2} \sum_{i,j} \mathbf{r}_{i,j} \times \mathbf{r}_{i,j+1} \tag{122}$$

and the moment of inertia (this is a better definition than given previously)

$$I_c = \left\langle \sum_{i,j} m_i \mathbf{r}_{ij}^{\perp} \cdot \mathbf{r}_{i,j+1}^{\perp} \right\rangle. \tag{123}$$

The superfluid density is proportional to the mean-squared area of paths sampled for a container at rest divided by the classical moment of inertia.

At high temperature the mean-squared area will be the sum of the mean-squared areas for each atom's path, since we can assume that the areas will be uncorrelated with each other. Hence the superfluid density will be $\rho_s/\rho = 2\lambda\beta/(3 < r^2 >)$. It will be negligible once the size of the cylinder is greater than the thermal wavelength.

But for a superfluid, the mean-squared area can be much greater. One finds that the superfluid density approaches unity at low temperature. Superfluidity is a microscopic property that can be defined in a finite system. It is not necessary to take the thermodynamic limit or to have a phase transition to see its effect. The effect of Bose statistics in a Bose liquid is to reduce the number of excited states and hence the coupling to an external potential. This can happen in a finite system as well as in an infinite system.

Now let us change the geometry of the rotating cylinder, so we can see how superfluidity manifests itself in periodic boundary conditions. Periodic boundary conditions are more convenient for simulations, since no surfaces appear and there is no curvature in making a loop around the boundaries. Instead of using a filled cylinder, we enclose the helium between two cylinders of mean radius R and spacing d, where $d \ll R$. The classical moment of inertia will be mNR^2 and the area can be written as $WR/2$ where W is the *winding number*, defined as the flux of paths winding around the torus times the circumference of the torus. Here we have ignored all nonwinding paths, those paths which do not make a complete circuit around the cylinder, since their contribution is $\mathcal{O}(R^{-2})$ and negligible at large R. Now substituting these values of A and I_c into Eq. (121) for the superfluid density we get

$$\frac{\rho_s}{\rho} = \frac{< W^2 >}{2\lambda\beta N} \tag{124}$$

where the winding number is defined by

$$\mathbf{W} = \sum_{i=1}^{N} \int_0^\beta dt \left[\frac{d\mathbf{r}_i(t)}{dt}\right]. \tag{125}$$

In contrast to the area, tThe winding number is a topological invariant of a given path; one can determine the winding number by counting the flux of paths across any plane; it does not matter where the plane is inserted. We can think of these winding paths as the imaginary-time version of circulating currents. Paths with a nonzero winding are the signal for superfluidity. This justifies the claim made earlier, that the identification of a Bose superfluid requires the full imaginary-time paths. Static correlation functions are not enough; one needs to know how the paths are connected up. Macroscopic exchange is necessary to have both superfluidity and momentum condensation. However, neither property is simply proportional to the number of macroscopic exchanges. In 3D systems they go together; in 2D there is no condensation but the system is still superfluid.

4.4 Constructing the Action

It is clearly desirable to make a good but cheap approximation to the exact link action. The better we can make the individual link action, the fewer the number of time slices and the shorter the "polymer." The sampling becomes much easier as the paths have fewer links and the estimation of various quantities such as the kinetic energy have smaller statistical fluctuations. We have found that accurate simulations of liquid helium using the primitive approximation for the action, Eq. (92), would require an $M \approx 1000$ to reach the temperature of the superfluid transition while using a more accurate action uses only about $M \approx 20$ slices. Without improved actions the simulation of the superfluid transition[58] (Ceperley and Pollock, 1986) in ^{4}He would not have been possible, given the available resources. The task of finding a good action is different from that of finding a good integrator for an ordinary differential equation, for example, Newton's equation, because of the fractal nature of the paths. Since paths do not have continuous derivatives, predictor-corrector or leapfrog methods are not as useful. The exact action is a many-body function. If the interaction is a pair potential, the exact action will have not only renormalized pair terms, but also three-body terms, four-body terms, etc. Finding a good action is analogous to averaging out solvent degrees of freedom in a liquid or of renormalizing out small-scale motions. We want to integrate analytically over all the intermediate time steps so we can leave them out. The action is not a tremendously sensitive function of the end points, since averaging over paths acts to smooth out the potential.

The "traditional" way in simulations of deciding that a time step is small enough is to study the convergence of interesting properties with a series of long simulations with smaller and smaller time steps. A better action will give the exact result with a larger time step. The primary quantity to look at is the energy, since it is related to the partition function. But other static quantities such as the kinetic energy, potential energy, and pair-correlation function should also be studied. The main problem with convergence studies is that the convergence of the energy does not establish how other quantities converge. For example, it is often found that the potential energy converges much quicker that the kinetic energy. This means that the primitive action may correctly describe static correlations, but not the imaginary-time dynamics, which as we have seen, are directly related to superfluidity. Another practical annoyance of convergence studies is that one would need a new one for every density and temperature; that is very costly. The computer time to converge the statistical error becomes much longer as the time step is decreased. This is because paths with smaller time steps move much more slowly through phase space and because the statistical error of the standard estimator for the energy blows up at small τ.

Of course it is best to put in as many exact properties of the action as are known. The simplest is the Hermitian property, namely, that $U(R, R') = U(R', R)$. Without this property, paths will not have "time-reversal" invariance. This property is easy to put in: one simply symmetrizes any unsymmetrical form by using the action,

$$U_S(R, R') = \frac{1}{2}[U(R, R') + U(R', R)]. \tag{126}$$

This can have the effect of making the action good to one higher-order if the unsymmetrical components are the lowest-order errors. Another exact property is the behavior of the action as two particles approach each other, the other particles remaining a constant distance apart. It can be shown that the divergent part of the action should approach a two-particle form. For a Coulomb interaction, this condition leads to a cusp condition on the action at $r = 0$:

$$\lim_{r_{ij}\to 0} \frac{dU(R, R'; \tau)}{dr_{ij}} = -\frac{e_i e_j}{(d-1)(\lambda_i + \lambda_j)}. \tag{127}$$

Here e_i is the charge on particle i and $\lambda_i = \hbar^2/2m_i$. The path averaging in the FK formula smooths the potential. The smoothing makes the action finite at the origin instead of having a r^{-1} singularity.

For a Lennard-Jones, r^{-12} potential, one can show that the action at small r must diverge as r^{-5}. These small r conditions can be established by looking at the residual energy of the action, which we shall define in a moment. ne can also derive exact properties of the action at large distances by considering the action as a function of the fourier transform of the density and then going to the long-wavelength limit.

Now let us begin the task of finding improvements to the primitive action. Semiclassical methods rely on the fact that, at very high temperatures, the major contribution to the FK path-integral comes from paths neighboring a single "classical" path. The most probable path connecting the end points is obtained by optimizing the action in the Feynman-Kacs formula. Suppose it begins at R_0 and ends at R_F. Then the classical path will satisfy an equation of "motion"

$$\frac{d^2R}{dt^2} = 2\lambda\nabla V(R). \tag{128}$$

This is Newton's equation of motion in the inverted potential $-V(R)$. At sufficiently small "time," the action is dominated by the contribution from this one trajectory. This contribution can be written in the familiar WKB form as an integral over the potential,

$$S_{SC}(R_0, R_F; \tau) = -\tau E + \int_{R_0}^{R_F} dx\sqrt{\frac{V(R)+E}{\lambda}} \tag{129}$$

The energy $E = -V(R) + \frac{1}{4\lambda}\left(\frac{dR}{dt}\right)^2$ is a constant of "motion," and the integration variable x is the distance along the path; it has units of length. This formula is not very useful until we determine how the energy depends on R_0, R_F and τ. The initial "velocity" of the classical path must be chosen so that the path will end up at the final position at the right "time." For small imaginary-times τ, we can neglect the "acceleration." The optimal path is then a straight line connecting R_0 and R_F, and the action is the integral of the potential energy along this straight-line path,

$$U_{SC}(R_0, R_F; \tau) = \tau \int_0^1 ds V\left(R_0 + (R_F - R_0)s\right). \tag{130}$$

This is a better approximation than the primitive approximation, since there is some contribution from the entire region between R_0 and R_F. Higher-order terms will both have to improve the trajectory and have to average locally around the semiclassical path. It is difficult to make further corrections to this formula in the general many-body case without ending up with an expression that is too slow to evaluate at each step of the PIMC.

For small "times" the Gaussian paths in the FK formula sample only a small region around the initial and final points, with a size determined by the thermal wavelength Λ_τ. Suppose we assume that the potential is quadratic in this region. Then the potential can be specified by giving $V(R^*), \nabla V(R^*)$, and $\nabla\nabla V(R^*)$ where R^* is some point in the neighborhood such as R_0 or R_F. The vectors and tensors have dimension $3N$. The exact density matrix for a quadratic potential is a Gaussian [54] Let us expand that exact density matrix for the harmonic potential in powers of Λ_τ keeping in mind that the distance between the two "legs" of the density matrix, $(R_F - R_0)$, will be of order Λ_τ. The result to order λ is

$$\begin{aligned} U_H(R_0, R_F; \tau) = {} & \tau V(R^*) + \frac{\tau^2\lambda}{6}\nabla^2 V(R^*) \\ & -\frac{\tau}{12}(R_F - R_0)\nabla\nabla V(R^*)(R_F - R_0) - \frac{\tau^3\lambda}{12}[\nabla V(R^*)]^2. \end{aligned} \tag{131}$$

This expansion is equivalent to the Wigner-Kirkwood or $\hbar$ expansion of the action if we make the choice of $R^* = R_0$.

If the potential energy is a sum of pair interactions, all of the terms except the last one are also pair terms, since they are linear in V. Hence the effect of the second and third terms is to renormalize a pair interaction. But one can show that the average of the third term over normally distributed values of $(R_F - R_0)$ (those arising from a free-particle path) equals the negative of the second term. As a result, the second and third terms together have a much smaller average effect on the probability distribution of a path. This effect pushes their combined effect to higher-order in τ. We shall call the last term the ***polarization action***, since it is similar to the energy of a polarizable atom in an electric field. This term is not a pair sum, but the first genuine many-body contribution to the action.

Although one is picking up higher-order contributions with the harmonic expansion, it is not uniformly convergent for a hard potential. At large r, where the potential is small, the expansion is adequate, but at small r, where quantum effects are very important, all terms in the expansion are large. Suppose the potential goes as r^{-12} at small r. Then the second and third terms will diverge as r^{-14} while the last term will diverge as r^{-26} at small r. In fact, quantum diffraction causes the exact action to diverge only as r^{-5}. Clearly the expansion does not converge at small r.

The helium-helium interaction is better thought of as a hard-sphere interaction, ***i.e.***, having an infinite strength, for which the $\hbar$ expansion does not converge since the gradients of the potential do not exist. This expansion (where one does a Taylor expansion of the potential about a nearby point) can only be trusted if the higher-order terms are much less than one.

A better approach for a hard-sphere-like system is to determine the exact action for two atoms and then to use that to construct a many-body action. To justify this approach, first assume that the potential energy can be broken into a pairwise sum of terms,

$$V(R) = \sum_{i<j} v(r_i - r_j). \tag{132}$$

Now apply the Feynman-Kacs formula. What enters is the integral of the potential energy along a path. Let x_{ij} be the exponentiated integral of the pair energy along a random walk,

$$x_{ij} = \exp\left[-\int_0^\tau dt v(\mathbf{r}_{ij}(t)\right]. \tag{133}$$

Then x_{ij} is a random variable drawn from some distribution function that depends on the end points (R_0, R_F). In terms of these random variables the FK formula for a pair potential reads,

$$e^{-U} =< \prod_{i<j} x_{ij} > . \tag{134}$$

If the variables x_{ij} are uncorrelated with each other, we can interchange the product and averaging operation,

$$e^{-U} \approx \prod_{i<j} < x_{ij} > . \tag{135}$$

But the average on the RHS is exactly the interacting part of the exact action for a pair of atoms. The ***pair-product*** action is

$$U_2(R, R'; \tau) = \sum_{i<j} u_2(\mathbf{r}_{ij}, \mathbf{r}'_{ij}; \tau), \tag{136}$$

where $u_2(\mathbf{r}_{ij}, \mathbf{r}'_{ij}; \tau)$ is the exact action for a pair of atoms.

This approximation has several advantages over the other approaches. First, it is exact for a pair of particles by definition. Since most collisions occur between atoms two at a time, they are described correctly. The errors of U_2 come from three- and higher-body correlations. As an example, consider particle 1 interacting with two other particles, say 2 and 3. If the path goes toward particle 2, then v_{12} is larger and v_{13} is smaller than average, and vice versa if it goes toward particle 3. This correlation effect is not large in a homogeneous system, since there are other particles in other directions which will have the opposite correlations, so that most of the many-body effects tend to cancel. Considerations like this suggest that the pair product will be correct to lowest order in a density expansion of the action, since it is only when we have three atoms in close proximity that we make a substantial error.

The exact pair action can be calculated efficiently by the matrix-squaring method introduced by Storer[59]. First, the pair density matrix is factorized into a center-of-mass term that is free-particle like and a term that is a function of the relative coordinates. Without loss of generality one can consider only the density matrix for a single particle in a spherical external potential. One now expands the relative pair density matrix in partial-waves:

$$\rho(\mathbf{r},\mathbf{r}';\tau) = \begin{cases} \frac{1}{2\pi\sqrt{rr'}} \sum_{l=-\infty}^{\infty} \rho_l(r,r';\tau) e^{il\theta}, & \text{2D} \\ \frac{1}{4\pi rr'} \sum_{l=0}^{\infty} (2l+1)\rho_l(r,r';\tau) P_l(\cos\theta), & \text{3D} \end{cases} \tag{137}$$

where θ is the angle between $\mathbf{r}$ and $\mathbf{r}'$. Each partial-wave component is the density matrix for a 1D particle in a potential with an additional centrifugal term and satisfies the Bloch equation:

$$-\frac{\partial \rho(r,r';t)}{\partial t} = [-\lambda \frac{d^2}{dr^2} + \tilde{v}_l(r)]\rho(r,r';t) \tag{138}$$

with boundary conditions $\rho_l(r,r';0) = \delta(r-r')$ and $\rho_l(0,r';t) = 0$. The effective potential is defined as

$$\tilde{v}_l(r) = v(r) + \frac{\lambda}{4r^2} \begin{cases} (4l^2-1), & \text{2D} \\ 4l(l+1), & \text{3D} \end{cases} \tag{139}$$

Since each partial-wave is a Green's function, they satisfy the convolution equation,

$$\rho_l(r,r';\tau) = \int_0^\infty dr'' \rho_l(r,r'';\tau/2)\rho_l(r'',r';\tau/2) \,. \tag{140}$$

This is the basic equation of the ***matrix-squaring method.*** If we square the density matrix k times, it will result in a lowering of the temperature by a factor of 2^k. Each squaring involves a one-dimensional integral for each value of r, r', and l. If a uniform grid in r and r' is used to tabulate the density matrix and the trapezoidal rule is used for integration, one literally can square the matrix, $\rho_l(r,r')$.

Once the pair density matrix is computed for some value of τ, we must reexpress it in a form such that it can be quickly evaluated during the Monte Carlo simulation. Summation over partial-waves is too slow, particularly at large r and small τ; one would need on the order of $r/\sqrt{\lambda\tau}$ partial-waves. It is convenient to use the three distances

$$q = (|\mathbf{r}| + |\mathbf{r}'|)/2 \,, \quad s = |\mathbf{r} - \mathbf{r}'| \,, \quad z = |\mathbf{r}| - |\mathbf{r}'| \,, \tag{141}$$

where $\mathbf{r} = \mathbf{r}_i - \mathbf{r}_j$ and $\mathbf{r}' = \mathbf{r}_i' - \mathbf{r}_j'$. The variables s and z are small, on the order of the thermal de Broglie wavelength Λ_τ, and so we can expand the action in a power series:

$$\begin{aligned} u(\mathbf{r},\mathbf{r}';\tau) &= \\ &= \frac{u_0(r;\tau) + u_0(r';\tau)}{2} + \sum_{k=1}^{n}\sum_{j=0}^{k} u_{kj}(q;\tau) z^{2j} s^{2(k-j)}. \end{aligned} \tag{142}$$

The first term is the ***end-point*** action. The following terms are purely off-diagonal contributions. The functions $u_{kj}(q)$ can be determined by a least-squares fit to the partial-wave expansion and tabulated for use in the subsequent Monte Carlo calculations.

4.5 Path Sampling Methods

Here we consider how to do the multidimensional integrations and summations that path-integrals require. The total configuration space to be integrated and summed over is made of elements: $s = [P, R_1, ..., R_M]$ where $R_k = \{\mathbf{r}_{1k} \ldots \mathbf{r}_{Nk}\}$ are the path variables and P is the permutation that closes the path, $R_{M+1} \equiv PR_1$. We wish to sample these elements in the simulation from the probability distribution

$$\pi(s) = \frac{exp[-\sum_{k=1}^{M} S^k]}{Z}, \tag{143}$$

where S^k is the action of the k^{th} link. The partition function Z normalizes the function π in this space. This distribution is different from that of a simple liquid because the points on the path are linked together by the kinetic springs, which can cause the convergence of simple simulation techniques to become exceedingly slow. Ways of speeding up the convergence have been addressed by several methods, which we shall discuss.

Before we begin the discussion of Monte Carlo methods, the reader may be asking whether the molecular-dynamics (MD) method may be used instead. Indeed, such methods are useful for some path-integral applications, and several results have appeared[60]. The chief difficulty with dynamical methods, by which we mean those in which the paths variables change continuously with an artificial dynamics, is that it is not possible for the permutation to change continuously, since it is a discrete variable. Hence dynamical methods by themselves cannot treat problems in which quantum statistics are important. But even for systems of distinguishable particles, there are particular problems in applying MD methods to path-integrals.

There are two major concerns with MD methods: ergodicity and efficiency. It is easy to see that a free-particle path-integral system will never come into equilibrium. The classical analog is a collection of uncoupled harmonic oscillators which will never exchange energy with each other. Then the "time" averages will be different from phase-space averages. But even with an interparticle interaction, if the time step τ is small enough, ergodicity is a major worry. One can break up the long-term correlations by periodically resampling the momentum. This can be done in a continuous fashion by using a Nose thermostat, or one can resample the velocities (hybrid Monte Carlo) each dynamical step and accept or reject the changed velocities. Using these methods one is guaranteed to get convergence to the right distribution.

Once ergodicity is ensured, the major concern with molecular dynamics is the efficiency of sampling phase space. For small τ one needs dynamical steps small enough to capture the oscillations of the springs.One finds that the paths move very slowly through phase space. Tuckerman *et al.*[61] have introduced novel methods for speeding convergence by separating the slow and fast dynamical scales. In fact, the methods for separating these motions are an imitation of how one solves the equivalent problem in Metropolis Monte Carlo simulation. They have shown in some cases that path-integral molecular dynamics (PIMD) can be almost as efficient as PIMC.

We have already discussed the overall strategy for Metropolis Monte Carlo. Now consider the problem of how best to sample a single point on the path. This is an elementary operation of the path-integral algorithm.The task is to sample a point R at time τ which is to be connected to two fixed end points, R_1 and R_2, with imaginary-time coordinates, 0 and 2τ, respectively. Usually we want to resample only a few coordinates, say only $n \ll N$ particles are allowed to move.

In the simplest choice for the transition probability, the *classic* rule, a single atom at a single time slice is displaced uniformly inside a cube of side Δ, adjusted to achieve 50% acceptance. t is clear that Δ must be on the order of, or smaller than, the thermal de Broglie wavelength for a slice, $\Delta \approx \Lambda_\tau = \sqrt{\lambda\tau}$. The heat-bath transition rule will have the smallest correlation time among all transition rules. The neighborhood of this move is the subspace obtained by fixing $3(NM - n)$

variables and the permutation, but allowing n atoms at one time slice to vary throughout the box. The optimal sampling distribution for a point $\mathcal{R}$, conditional on the path's having earlier visited $\mathcal{R}_1$ and later visiting $\mathcal{R}_2$, is then proportional to

$$T^*(R) \propto \rho(\mathcal{R}_1, \mathcal{R})\rho(\mathcal{R}, \mathcal{R}_2). \tag{144}$$

Dropping factors independent of R and factoring out the free-particle action,

$$T^*(R) \propto \exp\left[-\frac{(R-R_m)^2}{2\sigma} - U(\mathcal{R}, \mathcal{R}_1) - U(\mathcal{R}, \mathcal{R}_2),\right] \tag{145}$$

where the ***midpoint*** is $R_m = (R_1 + R_2)/2$, the squared width is $\sigma = \lambda\tau$, and n is the number of moving particles. The noninteracting density matrix gives a Gaussian centered at R_m and width $\sqrt{\sigma}$. This distribution can be easily sampled and is called ***free-particle sampling.*** Free-particle sampling is already an improvement over classic sampling, because it leads to 100% acceptances in the absence of the potential or in the high-temperature limit and because the step size Δ is automatically set to be the width of the kinetic action.

A repulsive potential will cut holes in the free-particle Gaussian distribution where a nonmoving atom is present or where two moving atoms overlap. To go beyond free-particle sampling, we can choose for a transition probability the most general correlated Gaussian in $3n$ variables,

$$T_S(R) = \sqrt{(2\pi)^{3m} det(\mathbf{A})} e^{-(R-\bar{R})(2\mathbf{A})^{-1}(R-\bar{R})}, \tag{146}$$

where the $3n \times 3n$ positive-definite covariance matrix $\mathbf{A}$ and the mean position vector $\bar{R}$ are free parameters of the sampling. We shall choose the mean and covariance to approximate the moments of $T^*(R)$.

No matter how well single-bead sampling has been optimized, as the value of τ decreases, the random walk will diffuse through configuration space more and more slowly. In this subsection we examine how the paths diffuse through phase space if the random walk consists of only single- bead moves. We assume that the temperature is held fixed, but τ and hence the number of time slices varies. The largest displacement allowed by the free- particle density matrix is order $\Lambda_\tau = \sqrt{\lambda\tau}$. Interactions or poor sampling can reduce this displacement, but it is impossible for the average displacement to become much greater than Λ_τ, since this is fixed by the kinetic springs.

We can calculate how fast a free-particle path will move through path space. Consider a Monte Carlo procedure in which each bead is moved in turn. After a move of a single bead, the mean-squared center of mass will change as

$$< (\delta\mathbf{c})^2 > = \frac{1}{M^2} < (\delta\mathbf{r})^2 > \leq \frac{3\lambda\beta}{M^3}, \tag{147}$$

so the computer time needed to get the center of mass to diffuse a fixed distance will also scale as M^3. Hence the efficiency of any Markov process that has single-time-slice moves will have a correlation time that scales as $M^{-3} \propto \tau^3$ for large M. Entanglement effects coming from the interaction of several atoms will slow the relaxation further. This scaling law, in conjunction with the use of the primitive action, which necessitates very large values of M, has ruined many path-integral studies. For example, using the primitive action requires a time step twenty times smaller than the pair action. If one is also using single-slice sampling, this will slow convergence by a factor of 8000.

The simplest multiple-slice move is a ***displacement*** move, in which the entire chain is translated by an amount δ. One could say that we are treating the center of mass of a chain as a classical degree of freedom. The size of the displacement δ can be sampled from a uniform distribution inside a cube with side Δ, with Δ chosen to maximize the mean-squared diffusion of the center of mass. Usually this is done by making sure that the acceptance ratio is between 25% and 75%. The kinetic action is unchanged by the displacement, assuming that the atom is not permuting with another atom, otherwise all members of an exchange cycle must be displaced. Displacements will be rejected if the chain ends up overlapping with another chain. If the temperature is somewhat higher than the degeneracy temperature, so that the size of the path is less than the interparticle spacing, these moves are very useful. The displacement will not change the internal shape of the path; there has to be a different kind of move to do that. Since the move will take $\mathcal{O}(M)$ operations, a displacement should be attempted much less frequently than other kinds of quicker moves.

To generalize the displacement move to the internal degrees of freedom of the paths we use the normal modes of the kinetic action. These are obtained by a discrete Fourier transform along the "time direction" [62]. We define the normal-mode coordinate by:

$$Q_k = \sum_{l=1}^{M} R_l e^{2\pi i k l/M}. \tag{148}$$

The total kinetic action is decoupled in normal modes,

$$\begin{aligned} K &= \frac{1}{4\lambda\tau}\sum_{i=1}^{M}(R_i - R_{i-1})^2 \\ &= \frac{1}{\lambda\beta}\sum_{k}\sin^2(\pi k/M)|Q_k|^2. \end{aligned} \tag{149}$$

Each of the $3NM$ normal-mode variables Q_k is independent of the others and has a Gaussian distribution.

There are two quite different ways of using normal modes. First, in ***normal-mode sampling*** one uses this form of the kinetic action to construct a transition move[63]. One samples one or more Q_k from some transition probability, for example a Gaussian distribution with squared width, $\lambda\beta/[2\sin^2(\pi k/m)]$. Then the new path coordinates are determined by the inverse Fourier transform and the move is accepted or rejected based on the change in action and the ratios of transition probabilities. In the absence of a pair potential, all moves would be accepted. When a potential is present, only the large k modes can be sampled directly from the free-particle Gaussian, since they cause a small movement of the path. In contrast, the low k modes are moved only a small amount, say $|Q'_k - Q_k| < \gamma_k$, with γ_k adjusted to get 50% acceptances. The center-of-mass mode ($k = 0$) is just the displacement move that we already described. These moves are much slower than single-slice moves, since they take

The second, much more radical, approach is to work directly with the normal- mode variables by rewriting the path integrals as integrals over Q_k instead of R_k. This is called the method of ***Fourier path integrals***, [64]. In using the Feynman-Kacs formula, rather than discretizing the random walk in M time steps, one instead discretizes in M normal modes. Either discretization is valid, but the two truncations have different convergences. Once one limits the number of modes, then the coordinate-space path is differentiable to all orders in imaginary-time, and thus one can use higher-order integration formulas for the action, $\int_0^\beta dt V(R(t))$, where $R(t)$ is now defined by

$$R(t) = R_0 + \sum_{k=1}^{M} Q_k e^{-2\pi i t k/\beta}. \tag{150}$$

Multilevel Monte Carlo is a general sampling method (Ceperley and Pollock, 1986, 1990) which can efficiently make multislice, many-particle moves. It gains in efficiency because the coarsest movements are sampled and accepted or rejected before the finer movements are even constructed. Thus the number of moves/second is much higher, because time is not wasted on moves that will eventually be rejected. Suppose the full configuration s is dynamically partitioned at the beginning of a Monte Carlo step into $l+1$ levels $s = (s_0, s_1, \ldots, s_l)$, where the coordinates s_0 are to be unchanged by the move, s_1 are sampled in the first level, s_2 are sampled in the second level, etc. The primed coordinates $(s'_1, \ldots, s'_l)$ are the new trial positions in the sense of a Metropolis rejection method; the unprimed ones are the corresponding old positions with $s_0 = s'_0$.

We now make an approximation to the action as a function of variables in that level and in previous levels. This approximate action will help in deciding whether the sampling of the path should continue beyond the current level. [We shall call π_k the *level action*; properly speaking, the action is $-\ln(\pi_k)$.] Now, we choose a sampling rule for s_k contingent on the levels already sampled. $T_k(s'_k)$ can depend on $s'_0, s'_1, ..., s'_{k-1}$. Once the partitioning and the sampling rule T_k are chosen, the sampling proceeds past level k with probability

$$A_k(s') = \min\left[1, \frac{T_k(s_k)\pi_k(s')\pi_{k-1}(s)}{T_k(s'_k)\pi_k(s)\pi_{k-1}(s')}\right]. \tag{151}$$

That is, we compare A_k with a uniformly distributed random number in $(0,1)$, and if A_k is larger we go on to sample the next level. If A_k is smaller, we go back to the beginning and make a new partitioning. This acceptance probability has been constructed so that it satisfies a form of detailed balance for each level k:

$$\frac{\pi_k(s)}{\pi_{k-1}(s)} T_k(s'_k) A_k(s') = \frac{\pi_k(s')}{\pi_{k-1}(s')} T_k(s_k) A_k(s). \tag{152}$$

Having sampled $R = R_{\beta/2}$, one now bisects the two new intervals $(0, \beta/2)$ and $(\beta/2, \beta)$, generating points $R_{\beta/4}$ and $R_{3\beta/4}$ with the same algorithm. One continues recursively, doubling the number of sampled points at each level, stopping only when the "time" difference of the intervals is τ.

This is a simpler, but more powerful, sampling method for free-particles than the normal-mode method. It is simpler in that there are no Fourier transforms. It is more powerful because it generalizes to fully interacting paths and can be used in combination with the multilevel method to accomplish early rejection.

4.5.1 The bisection method

Let us now combine the Lévy construction of the path with the multilevel Metropolis method. Suppose a single-particle or many-particle path consisting of $m = 2^l - 1$ time slices is "clipped out" where l is the *level.* The fixed end points are R_i and R_{i+m}. The new points to be sampled will have the coordinates: $R_{i+1}, \ldots, R_{i+m-1}$. The places that pose the greatest difficulty for finding a new path are in the middle of the interval $R_{i+m/2}$, simply because the middle is the farthest from the end points, which are known to have acceptable potential energies. The coordinates are partitioned into levels as in the Lévy construction. By bisecting the interval rather than working from one end, one discovers the blockages quickly. If an overlap is found, the construction of the paths comes to a halt.

The bisection algorithm is recursive. First the midpoint is sampled. Then the same algorithm is used to find the midpoints of the two remaining intervals, etc. The coordinates to be moved are partitioned as

- s_0 = atom positions outside of time slices in consideration and atoms not being moved.
- s_1 = coordinates of atoms being moved at the middle time slice $i + m/2$.
- s_2 = coordinates of atoms being moved at $i + m/4, i + 3m/4$.
- ...
- s_l = coordinates of atoms being moved at $i + 1, i + 3, ..., i + m - 1$.

Now we need to define an action at the k^{th} level. The optimal level action would be simply the product of density matrices with the appropriate time argument. For the first level we get

$$\pi_1^*(R_{i+m/2}) = \rho(\mathcal{R}_i, \mathcal{R}_{i+m/2})\rho(\mathcal{R}_{i+m/2}, \mathcal{R}_{i+m}). \tag{153}$$

We are free to choose any convenient approximation, since it only affects the convergence. One can use the same approximations to the action at $(m/2)\tau$ that were developed earlier, but now accuracy is less important than speed. t the final level, the exact action must be calculated but rejections are less frequent at the final level because the sampling methods work better the smaller the "time" difference, so the extra work is less likely to be wasted.

Once the level action has been chosen, we must choose the transition probability. But this is exactly the problem that we already considered. The only difference is that the time step is some multiple of τ instead of τ. Rejections are due to the combined effect of using approximate sampling functions and using approximate level actions.

4.5.2 The necessity of joint permutation-path moves

We now take up the problem of permutation-space sampling. Here the problem of ergodicity is particularly acute. Path coordinates will eventually reach equilibrium if the calculation is sufficiently long, since the paths can slowly diffuse through phase space. However, permutation space is discrete, and it can easily occur that all the attempted permutation moves of a (finite) random walk are rejected.

Now let us reapply the heat-bath and multilevel Metropolis methods to the joint sampling of permutations and path moves. As we discussed with regard to bisection, a set of $m-1$ time slices are selected for the path move with end points $\mathcal{R}_i$ and $\mathcal{R}_{i+m}$. A local permutation move consists of applying a cyclic exchange of n atoms to an existing path. What we now describe is how to pick the permutation. Once the permutation is picked, the bisection algorithm is used to sample a path corresponding to that permutation exactly as before. We can regard the permutation change as the first level in the multilevel sampling method. The second level will be the midpoint of the interval, $R_{i+m/2}$, and so forth.

Since permutation space is discrete, we can directly use the optimal algorithm, the heat-bath transition probability. The heat bath transition probability for a permutational change is $T^*(\mathcal{P}) \propto \rho(\mathcal{R}_i, \mathcal{P}R_{i+m})$ where $\mathcal{P}$ ranges over all cyclic permutations involving n atoms. The neighborhood for pair permutations has $N(N-1)/2$ elements, for three-body permutations $N(N-1)(N-2)/3$ elements, *etc.* If we make the end-point approximation for the density matrix, terms involving the interaction will drop out since they are symmetric under particle interchange. Hence T^* depends only on the free-particle kinetic action,

$$T^*(\mathcal{P}) = \frac{1}{C_I} \exp\left[-\sum_{j=1}^{n}(\mathbf{r}_{j,i} - \mathbf{r}_{\mathcal{P}j,i+m})^2/4n\lambda^*\tau\right] \tag{154}$$

where C_I is a normalization factor, Eq. (21), defined so that the probability of making some permutation move is one. The λ^* in this expression is an effective mass to take into account off-diagonal contributions that we dropped. This transition probability can be used in two different ways. The cyclic permutation can either be explicitly sampled from a precomputed table or it can be implicitly sampled with a walk through particle labels.

In the first method, a table of all transition probabilities within the neighborhood is constructed. The table can be constructed rather rapidly, since it involves only particle distances between the end points,

$$t_{kj} = \exp[-(\mathbf{r}_{k,i} - \mathbf{r}_{j,i+m})^2/(4m\lambda\tau)]. \tag{155}$$

The probability for trying a cyclic exchange of l atoms with labels $\{k_1, \ldots, k_l\}$ is

$$T^*(\mathcal{P}) = \frac{1}{C_I} t_{k_1,k_2} t_{k_2,k_3} \ldots t_{k_l,k_1}. \tag{156}$$

It is best to put in the table only permutations that have a probability of being chosen greater than some threshold. The total number of possible permutations grows rapidly with the size of the maximum cyclic exchange being considered and the number of particles. But the number of permutations with a probability greater than some threshold does not grow rapidly, since all of the atoms need to be within a thermal wavelength of their exchanging partner. Those permutations can be found quickly using a tree search. One then constructs a list of the likely permutations and of the probability of choosing a given permutation. The permutation is sampled with the usual method of sampling a discrete distribution. Having set up this permutation table, one amortizes its computational cost by attempting many permutation moves before moving on to a new interval of time slices.

An alternative way of sampling a permutation solves these two problems. One constructs the t_{kj} matrix as before, but then walks through the table at random, trying to make a cyclic permutation of l atoms. The initial atom of the cyclic exchange, k_1, is chosen randomly from the list of all the atoms. The second atom, k_2, is then selected with probability proportional to $t_{k_1,k_2}/h_{k_1}$, where $h_{k_1} = \sum_k t_{k_1,k}$, and so forth. After all of the l labels are selected (and a check is done to make sure they are all different), the trial permutation is accepted or rejected with probability:

$$A = \min\left[1, \frac{\frac{h_{k_1}}{t_{k_1,k_1}} + \ldots + \frac{h_{k_l}}{t_{k_l,k_l}}}{\frac{h_{k_1}}{t_{k_1,k_2}} + \ldots + \frac{h_{k_l}}{t_{k_l,k_1}}}\right]. \tag{157}$$

One gets a sum of terms in the numerator and denominator because a cyclic permutation can be generated by starting at any one of the members of the cycle. If it is accepted, the bisection algorithm to sample the path variables begins. Acceptances are rare. Essentially there is only 1 chance in N that the cycle will close on itself with a large value of the last link t_{k_l,k_1}. But the process of constructing each loop is very rapid so the overall efficiency is not bad.

4.6 Calculating Properties

Once the action is chosen and sampling is accomplished, we are ready to calculate expectation values. It is straightforward to calculate scalar operators, such as the density, the potential energy, and the pair-correlation function; they are simply averages over the paths. Use can be made of the symmetry in imaginary time, since all time slices are equivalent. Thus the average density is

$$\rho(\mathbf{r}) = \frac{1}{M} \sum_{i,t} \langle \delta(\mathbf{r} - \mathbf{r}_{it}) \rangle \tag{158}$$

We shall use $\langle\ldots\rangle$ to indicate an average over the paths *and* over links t.

The internal energy is one of the main properties that one wants to get out of a simulation. There are a variety of ways of estimating the energy, but surprisingly, the problem of finding the best estimator has not yet been resolved. Let us for the moment split the energy into a calculation of the potential energy and the kinetic energy $\mathcal{K}$. The potential energy is easy to calculate, since it is diagonal in configuration space, although we shall discuss an alternative estimator in terms of a free-energy derivative. The simplest way to think of the kinetic energy is in terms of the stretching of the polymers, or equivalently, the single-particle imaginary-time "diffusion." We define the diffusion distance as

$$D(t) = \left\langle (\mathbf{r}_i(t) - \mathbf{r}_i(0))^2 \right\rangle . \tag{159}$$

Then $\mathcal{K}_D$ is the *diffusion* estimate of the kinetic energy. The kinetic energy is the initial slowing down of the dynamics of the paths, due to the interaction and due to the periodic boundary conditions on the paths in imaginary time. The use of this equation is not convenient for calculating the kinetic energy, because it is hard to estimate the second derivative at zero time, since even the first derivative is fluctuating.

The *thermodynamic* estimator of the energy is obtained by differentiating the partition function with respect to the inverse temperature,

$$E_T = -\frac{1}{Z}\frac{dZ}{d\beta}. \tag{160}$$

Interpreting the ratio as an average over imaginary-time paths, applying the derivative to link i alone, and writing in terms of the action, we get

$$E_T = \left\langle \frac{3N}{2\tau} - \frac{(R_i - R_{i-1})^2}{4\lambda\tau^2} + \frac{dU^i}{d\tau} \right\rangle . \tag{161}$$

At sufficiently small τ, U reduces to τV. In the high-temperature limit, the first two terms are the kinetic energy and the last is the potential energy. For larger τ, the last term also contain a kinetic contribution.

The error behaves very poorly at small τ; they grow as τ^{-1}. The first two terms are of order τ^{-1} but since kinetic energy is independent of τ, there is a cancellation between these terms. As τ becomes small we are trying to find a small difference between the constant first term and an almost equally large but fluctuating second term. This is exactly the same problem that we mentioned with the diffusion estimator. The problem is independent of the temperature but depends on the time step. If the additional variance caused by autocorrelation is ignored, the error will be proportional to $\tau^{-1}(1 - 2\tau\mathcal{K}/3) \approx \tau^{-1}(1 - 1/M)$, where the second expression uses the classical expression for the kinetic energy and M is the number of time slices. If one goes to the classical limit by fixing M and letting τ get small, the absolute error of the kinetic energy will grow. In the classical limit, the kinetic energy approaches $3k_BT/2$. Hence using this estimator, it is very difficult to estimate quantum corrections to the kinetic energy in the classical limit.

It is possible to eliminate the troublesome kinetic-energy terms, which cause the large variance at small τ, by integrating by parts over the path variables [65]. One ends up with an estimator similar to the virial expression for the pressure. The *virial* energy estimator is

$$\begin{aligned} E_V \quad = \quad & \left\langle \frac{3N}{2L\tau} - \frac{1}{4L\tau^2\lambda}(R_{L+i} - R_i)(R_{i+1} - R_i) \right. \\ & \left. -\frac{1}{2}F^i\Delta_i + \frac{dU^i}{d\tau} \right\rangle \end{aligned} \tag{162}$$

where F_i is a generalization of the classical force,

$$F^i = -\frac{1}{\tau}\nabla_i(U^{i-1} + U^i), \tag{163}$$

and Δ_i is the deviation of a particle's position from its average position,

$$\Delta_i = \frac{1}{2L}\sum_{j=-L+1}^{L-1}(R_i - R_{i+j}). \tag{164}$$

The parameter L, with $(1 \leq L \leq M)$, is the ***window size*** for averaging. If it is chosen to be unity then by inspection the virial estimator reduces to the thermodynamic estimator. Its maximum value is $L = M$; this is the conventional choice. If there are no exchanges or windings, the second term will drop out, since $R_{i+M} = R_i$. The virial estimator is very effective at computing quantum corrections to a nearly classical system, since the first term does not fluctuate and is the classical kinetic energy, the second term vanishes, and the last term is approximately the classical potential energy.

4.7 Comparison with other Quantum Monte Carlo Methods

In this section we make some brief comparisons with other Quantum Monte Carlo methods.

One of the advantages of the VMC method is that it is simple both to understand and to program. The calculations are perhaps an order of magnitude faster than for PIMC. States that are a ground-state of a given symmetry, such as Fermions, phonons, rotons and vortices can be treated by making an appropriate trial function. With VMC one can tell energetically how important a given correlation is by systematically adding terms to the trial function. One ends up with an explicit trial function which helps in understanding the quantum system.

But VMC is hardly a black box. To get reliable results one must very carefully optimize trial functions and systematically add more complicated effects. There is nothing internal to the method that tells you when to stop introducing more correlations. This variational bias (*i.e.*, the amount of energy missed by a given class of trial functions) depends on the phase; it is smaller in the solid than in the liquid. Thus variational calculations of the liquid-solid transition will put the transition density too low.

In PIMC, the entire path is held in the computer memory and one jiggles the path with the Metropolis Monte Carlo method: the random walk is an artificial process used to sample path space. The walk continues for an indefinite number of steps to reduce the statistical errors. The number of slices on the path is held fixed. Changing β involves a new run.

In GFMC, the implementation is entirely different from that of variational path integrals. In GFMC, the evolution in imaginary time is also the evolution of the Markov process; the dynamics of the random walk is given by the density matrix. The evolution in imaginary time continues indefinitely until a steady state is reached, simultaneously reaching convergence in β and reducing the statistical errors.

In GFMC (without importance sampling), the probability of sampling R' contingent on R is proportional to $< R'|e^{-\tau\mathcal{H}}|R >$. The normalization (the integral over R') is not unity and depends on R. Hence the the number of sampled points R', or their weight, must depend on R. For systems of many atoms, one has to use branching in order to interpret the projection as a random walk. The state space of the stochastic process is an ensemble of configurations $\{R_i\}$. A step consists of a diffusion for each member of the ensemble and a branching step, in which some configurations are deleted and some are duplicated.

Of course the major advantage of PIMC is the ability to calculate properties at temperatures greater than zero. This could be a disadvantage for calculating purely zero-temperature properties, but it is generally an advantage in comparing with experimental data. In a bulk superfluid, there are very few excited states; so that below 1 K, ^{4}He is essentially in the ground-state, so in practice even the restriction to nonzero temperature is not always important. One of the main advantages of PIMC is that order parameters, such as the superfluid density and the tunneling frequency in solid ^{3}He, are more simply expressed in terms of path-integrals. In GFMC it is much less obvious how Bose symmetry is expressed. In a superfluid system, the GFMC walks diffuse through phase space, they are not trapped.

It would seem that, because of the zero-variance principle, the GFMC method would be more efficient at computing the ground-state energy. However, in practice GFMC and PIMC give similar error bars on the energy for similar amounts of computer time. Other properties, such as the pair distribution function, are more difficult to estimate with GFMC, since the simulation calculates averages with the "mixed" estimator, the product of the ground-state wave function and the trial wave function. Removal of the effect of the trial function is biased and adds to the difficulty of the method. It is difficult for GFMC to break away from the long-range order of a trial function. PIMC does not have this difficulty, giving exact thermal averages. For an efficient GFMC calculation one needs to have a good trial function. Usually a preliminary step is a good variational Monte Carlo optimization, at least at the pair-product level. Once a good trial function has been found, then the machinery of GFMC takes over. But this first step can take a lot of graduate student time. PIMC is much more of a "black box." One puts in the action, allows the code to run for a long time, and measures the observables. It is much more likely in PIMC that the paths by themselves will make the transition to a new, unexpected state. To balance this, PIMC has more problems with ergodicity. We have seen the difficulty with constructing moves that change the winding number and permutation cycles. Those issues do not arise in GFMC, where the dynamics is fixed by the density matrix and the trial function. Since the random walks need not close, they can move through phase space more easily. Another problem with GFMC is its efficiency as the number of atoms gets large. Two problems arise. First the branching factor grows exponentially with the number of atoms. To keep the branching fixed requires the time step to go as $N^{-1/2}$, which increases the computational effort. Second, members of the ensemble get more correlated with each other. To keep the algorithm unbiased, the size of the ensemble must grow with N. These scaling difficulties with GFMC have not been investigated in detail, and it is not known how serious they are. They do not arise in PIMC; classical statistical mechanics assures us that nothing strange happens as we add more polymers. Correlation times can be longer, but they can be reduced with classical sampling techniques.

For all of these reasons, PIMC is a better "black box" than VMC, or GFMC.

5 Fermion Path Integrals

This section is an abridged version of a longer article[66].

The straightforward application of PIMC to Fermi systems means that odd permutations subtract from the integrand. This is the "fermion sign problem" which we have discussed earlier. Path integral methods as rigorous and successful as those for boson systems are not yet known for fermion systems in spite of the activities of many scientists throughout the last four decades.

Now let us consider how particle statistics are expressed in path integrals. For systems of identical particles, the states can be classified into symmetric and antisymmetric states. The fermion density matrix is defined by restricting the sum to be only over antisymmetric states.

(Similarly for other symmetries such as momentum or spin.) We shall denote the statistics of the particles by subscripts: ρ_F will denote the fermion density matrix, ρ_B the boson density matrix, ρ_D the boltzmannon (distinguishable particle) density matrix, and ρ any of the above density matrices.

Note that for any density matrix the diagonal part is always positive:

$$\rho(R, R; \beta) \geq 0 \tag{165}$$

so that $Z^{-1}\rho(R, R; \beta)$ is a proper probability distribution. It is the diagonal part which we need for many observables, so that probabilistic ways of calculating those observables are, in principle, possible.

Let $\mathcal{P}$ be one of the $N!$ permutations of particle labels. (For the moment we ignore the spin and consider spinless fermions.) Then each of the fermion eigenstates has the following property:

$$\phi(\mathcal{P}R) = (-1)^{\mathcal{P}}\phi(R). \tag{166}$$

The density matrix has the following symmetries:

$$\begin{aligned} \rho(R, R'; \beta) &= \rho(R', R; \beta) \\ \rho_F(R, R'; \beta) &= (-1)^{\mathcal{P}}\rho_F(\mathcal{P}R, R'; \beta) \\ \rho_F(R, R'; \beta) &= (-1)^{\mathcal{P}}\rho_F(R, \mathcal{P}R'; \beta). \end{aligned} \tag{167}$$

One can use the permutation (or relabeling) operator to construct the path integral expression for the boson or fermion density matrix in terms of the Boltzmann density matrix:

$$\rho_{B/F}(R, R'; \beta) = \frac{1}{N!}\sum_{\mathcal{P}}(\pm 1)^{\mathcal{P}}\rho_D(\mathcal{P}R, R'; \beta). \tag{168}$$

More generally, one uses some projection operator to select a desired set of states from the distinguishable particle density matrix which contains all states. In this lecture, except for how paths close, particles are generally considered to be distinguishable. This is in contrast to the second-quantized philosophy, where one always works with an antisymmetric basis.

An alternative definition of the density matrix is by its evolution in imaginary time, the Bloch equation:

$$-\frac{\partial\rho(R, R'; t)}{\partial t} = \mathcal{H}\rho(R, R'; t) \tag{169}$$

which obeys the boundary condition at $t = 0$ for boltzmannon statistics:

$$\rho_D(R, R'; 0) = \delta(R - R') \tag{170}$$

or for Bose or Fermi statistics:

$$\rho_{B/F}(R, R'; 0) = \frac{1}{N!}\sum_{\mathcal{P}}(\pm 1)^{\mathcal{P}}\delta(\mathcal{P}R - R'). \tag{171}$$

The high temperature boundary condition is an (anti)symmetrized delta function.

In the ***direct fermion method*** one sums over permutations just as for bosonic systems. Odd permutations then contribute with a negative weight. The direct method has a major problem because of the cancellation of positive and negative permutations. This was first noted by Feynman and Hibbs (1965).

The efficiency is simply the number of even permutations minus number of odd permutations. In fact, we can show that the efficiency is equal to:

$$\xi = \left[\frac{Z_F}{Z_B}\right]^2 = \exp[-2\beta(F_F - F_B)] \tag{172}$$

where Z_F and Z_B refer to partition function and F_F and F_B to the total free energies for Fermi and Bose statistics respectively. Of course the free energies are proportional to the number of particles. The direct fermion method, while exact, becomes exceedingly inefficient as β and N increase, precisely when the physics becomes interesting.

We now introduce the ***restricted path integral method***; the analog of the fixed-node ground state method. This is based on the restricted path identity: that the nodes of the exact density matrix determine the rule by which one can take only paths with the same sign. For the diagonal density matrix we can arrange things so that we only get positive contributions.

$$\rho_F(R_\beta, R_*; \beta) = \int dR_0 \rho_F(R_0, R_*; 0) \oint_{R_0 \to R_\beta \in \Upsilon(R_*)} dR_t e^{-S[R_t]} \tag{173}$$

where the subscript means that we restrict the path integration to paths starting at R_0, ending at R_β and are node-avoiding (those for which $\rho_F(R_t, R_*; t) \neq 0$ for all $0 < t \leq \beta$.) The weight of the walk is $\rho_F(R_0, R_*; 0)$. It is clear that the contribution of all the paths for a single element of the density matrix will be of the same sign; positive if $\rho_F(R_0, R_*; 0) > 0$, negative otherwise. In particular, on the diagonal all contributions must be positive. Important in this argument is that the random walk is a continuous process (the trajectory is continuous) so we can say definitively that if sign of the density matrix changed, it had to have crossed the node at some point.

The problem we now face is that the unknown density matrix appears both on the left-hand side and on the right-hand side of Eq. (173) since it is used to define the criterion of node-avoiding paths. To apply the formula directly, we would somehow have to self-consistently determine the density matrix. In practice what we need to do is make an ***ansatz***, which we call ρ_T, for the ***nodes*** of the density matrix needed for the restriction. The trial density matrix is used to define trial nodal cells: $\Upsilon_T(R_*)$. Using a trial nodes we generate a better approximation to the density matrix using Eq. (173) with the trial restriction:

$$\tilde{\rho}_T(R_\beta, R_*; \beta) = \int dR_0 \rho_F(R_0, R_*; 0) \oint_{R_0 \to R_\beta \in \Upsilon_T(R_*)} dR_t e^{-S[R_t]}. \tag{174}$$

Hence, $\tilde{\rho}_T(R', R; \beta)$ is a solution to the Bloch equation inside the trial nodal cells, and it obeys the correct initial conditions. It is not an exact solution to the Bloch equation (unless the nodes of ρ_T are correct) because it has possible gradient discontinuities at the trial nodal surfaces.

We call R_* the reference point and it plays a very special role in restricted path integrals since it is the value of the density matrix with respect to the reference point that restricts the paths. Averages such as the density can only be taken at the reference point. By a "time-independent" or "ground-state" restriction is meant that the restriction does not depend on the reference point. This is achieved by using an antisymmetric trial wavefunction $\Psi_T(R)$ and requiring that $\Psi_T(R_t) \neq 0$ throughout the path. This is identical to the ground state fixed-node method (diffusion or Green's Function Monte Carlo). The algorithm is considerably simpler than using time-dependent nodes and time-slice symmetry is restored. However, calculations with ground-state nodes on liquid ^{3}He gave a poorer description of the properties of the liquid at non-zero temperature[67].

To get a feeling for restricted paths let us consider the problem of molecular hydrogen. We will work in the Born-Oppenheimer approximation so the two protons in a single hydrogen molecule are represented by two spin 1/2 particles interacting with an attractive potential. The total spin (S) and total orbital angular momentum (L) are good quantum numbers. The spin 0 state is called para-hydrogen, and must have an even value of L to keep the molecular wavefunction antisymmetric in spin and coordinates. The spin 1 states are called ortho-hydrogen and they must have odd values of L. Often the hydrogen cannot easily change its spin state, so that para- and ortho-hydrogen can be considered as separate chemical species, for a time at least. They can change their angular momentum values with collisions with other molecules but not easily their spin. A third possibility of statistics is if the two nuclei are different particles, *e. g.* a proton and a deuteron, in which case they obey distinguishable particle or Boltzmann statistics.

Let us suppose that the particles are massive enough that the relative coordinate is almost fixed at a given radius $r_0 \approx 0.75$ Å. Hence the relative coordinate $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ is almost fixed on the surface of a sphere.

Now consider what we need to do to calculate the partition function for the three types of statistics. Distinguishable particles are the simplest: allow all paths returning to the starting point (type A in the figure). For ortho- and para-hydrogen, we can use parity to project out the correct states. This generates paths of type B which end up at the opposite pole in relative coordinates. For para-hydrogen the direct method would be to sum over all paths of types A and B. Ortho-hydrogen would be to sum over paths of type A but subtract the contribution from paths of type B.

In the case of ortho-hydrogen it is easy to calculate the exact nodes of the density matrix. In relative coordinates any wavefunction has the angular factor $Y_{lm}(\hat{\mathbf{r}})$. Then in the sum over the quantum states for different m, using the addition formula for spherical harmonics we obtain a factor $P_l(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_*)$. Since all the odd Legendre polynomials vanish when their arguments vanish the ortho-density matrix vanishes when $\mathbf{r} \cdot \mathbf{r}_* = 0$. In general, additional nodes would be possible, but the hydrogen molecule has only this one planar node. Paths of type C are node-crossing as opposed to the node-avoiding path A.

To summarize, we add the following classes of path for the different statistics of the hydrogen molecule:

1. Distinguishable hydrogen: A+C
2. Para-hydrogen: A+B+C
3. Ortho-hydrogen (direct method): A-B+C
4. Ortho-hydrogen (restricted method): A only ($\mathbf{r}_t \cdot \mathbf{r}_* > 0$)

The reason that restricted path integrals give the same value is that paths of type B and C can be paired together and canceled off against each other. This is because the flux of paths is the gradient of the density matrix at the node and since the gradient is continuous across the node, the positive paths crossing at a given nodal point will precisely cancel against the negative paths.

Hence the restricted paths are limited to be in a half-space. Note that there is no definite location of this half-space. Its position depends on the reference point, even at zero temperature. This is because the ground state of S=1 is three-fold degenerate. Isotropy is restored by averaging over the reference point position.

Now, let us discuss the nodal surfaces of non-interacting fermions. Let $v_e(\mathbf{r})$ be a single-particle external potential. The distinguishable particle density matrix is then a product of solutions of the single-particle Bloch equation:

$$-\frac{dg(\mathbf{r}, t)}{dt} = [-\lambda \nabla^2 + v_e(\mathbf{r})]g(\mathbf{r}, t) \tag{175}$$

with the boundary condition:

$$g(\mathbf{r},\mathbf{r}_*;0)=\delta(\mathbf{r}-\mathbf{r}_*). \tag{176}$$

Then using the antisymmetric projection operator and the definition of a determinant, we find for the spinless case:

$$\rho_F(R,R_*;t)=\frac{1}{N!}det[g(\mathbf{r}_i,\mathbf{r}_{j,*};t]. \tag{177}$$

In the case where the external potential is zero (or a constant), the single-particle density matrix is a Gaussian.

$$g(\mathbf{r},\mathbf{r}_*;t)\propto \exp\left[-\frac{(\mathbf{r}-\mathbf{r}_*)^2}{4\lambda t}\right]. \tag{178}$$

The action for restricted fermions, which we write as $S_F(R_t,R_{t-\tau};\tau;t_r,R_*)$, is more complicated and depends on more variables than the distinguishable action S_D because of the lack of time symmetry in the paths. It is a function of R_* and t in addition to its usual dependence on $R_t, R_{t+\tau}$ and τ. The ***primitive-nodal*** action makes the approximation of checking the restriction only at the M sampled points on the path. We only check to see whether $\rho_T(R_j,R_*;t_j)<0$ for $0\leq j\leq M$. In the primitive approximation the error decreases as $\tau^{1/2}$. This dependence is easy to understand. The energy of a box of size a is $\lambda(\pi/a)^2$. Using the primitive approximation effectively increases the size of the box by an amount proportional to the thermal deBroglie wavelength of a time step: $(\lambda\tau)^{1/2}$. Hence, the energy is decreased by a relative amount $(\lambda\tau)^{1/2}/a$.

Luckily, one can do considerably better, so that fewer time slices are needed to accurately represent the path. The action picks up a contribution from the nodes because walks can wander back and forth across the nodes even though they happen to be on the correct side at the sampled points, R_t and $R_{t+\tau}$. Improved fixed-node sampling methods have been developed for the ground state simulations by Anderson (1976)[26]. The exact nodal action for a particle in a box is easy to calculate. In the case of a particle confined only to the half-space $x>0$, one can solve the Bloch equation by the method of images. The total density matrix is the difference: $\rho(x,x')-\rho(x,-x')$ since the difference satisfies the Bloch equation and the boundary condition at $t=0$ and $x=0$. Using the form for the free particle action the nodal action is:

$$U_N(x_t,x_{t+\tau})=-\ln[1-exp(-\frac{d_t d_{t+\tau}}{\lambda\tau})] \tag{179}$$

where d_t is the distance of x_t to the node at time t and $d_{t+\tau}$ is the distance of $x_{t+\tau}$ from the node at time $t+\tau$. Hence the action diverges logarithmically near the node and is significantly repulsive in a region on the order of $\sqrt{\lambda\tau}$.

The remaining non-trivial problem is how to estimate the distance d to the node. In the many-body time-dependent case we define the distance to the node as $d_t=\min(|R_n-R_t|)$ where R_n varies over all points with $\rho_F(R_n,R_*;t)=0$. To get an estimate, we can use the Newton-Raphson method: given a function (hopefully smooth) $f(R)$ which vanishes when $\rho_T(R)$ does, an estimate of the nodal distance is:

$$d(R)\approx\frac{|f(R)|}{|\nabla f(R)|}. \tag{180}$$

This is good as long as the contribution of higher order derivatives of f is not large within a thermal wavelength.

Let us consider how to calculate the momentum distribution with restricted paths. The momentum distribution is the Fourier transform of an off-diagonal element of the density matrix. Consider the ideal fermi-gas momentum distribution (for spinless fermions).

$$n_{\mathbf{k}} = \begin{cases} 1/(2\pi^3\rho) & \text{for} \quad k < k_F \\ 0 & \text{for} \quad k > k_F \end{cases} \tag{181}$$

$$n(\mathbf{r}) = \frac{3}{(k_F r)^3}[sin(k_F r) - k_F r cos(k_F r)] \tag{182}$$

where the Fermi wavevector for spinless fermion is related to the density by $k_F = (6\pi^2\rho)^{1/3}$. Note that the single particle density matrix is proportional to the spherical Bessel function $j_1(z)$, and slowly decays to zero at large r. It has zeroes at $k_F r = 4.493, 7.725, \ldots$. These zeroes mark the places where the even and odd permutations cancel out. Since $n(r)$ is often negative, even with restricted paths we must have negative weights entering. The momentum distribution has a discontinuity at the Fermi wavevector $\mathbf{k}_F$. As a consequence the single particle density matrix must decay at large distance as r^{-2}. We can get such long-range behavior only if there are macroscopic exchanges. Hence the existence of any kind of non-analytic behavior (we mean a discontinuity in $n_{\mathbf{k}}$ or in any of its derivatives) implies that the restricted paths have important macroscopic permutation cycles.

I acknowledge financial support from NSF-DMR94-224-96 and ONR-N00014-92J-1320. Matthew Jones was helpful in texing some of the notes. Publications from the University of Illinois Quantum Monte Carlo group are available at http://www.ncsa.uiuc.edu/Apps/CMP/cmp-homepage.html

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[15] J.P. Bouchaud and C. Lhuillier, Europhys. Lett. 3, 1273(1987).

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

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

[18] J.P. Bouchaud and C. Lhuillier, in *Spin Polarized Quantum Systems*, ed. S. Stringari, World Scientific, Heidleberg (1989).

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

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

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

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

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

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

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

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

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

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

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

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

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

[37] 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).

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

[39] 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).

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

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

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

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

[44] 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)

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

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

[47] 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)

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

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

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

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

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

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

[54] Feynman, R. P., 1972, *Statistical Mechanics*, (Benjamin, New York).

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

[56] Feynman, R. P., 1953, Phys. Rev. **90**, 1116; **91**, 1291; **91**, 1301.

[57] Penrose, O., and L. Onsager, 1956, Phys. Rev. **104**, 576.

[58] Ceperley, D. M., and E. L. Pollock, 1986, Phys. Rev. Lett. **56**, 351.

[59] Storer, R. G., 1968, J. Math. Phys. **9**, 964; Klemm, A. D., and R. G. Storer, 1973, Aust. J. Phys. **26**, 43.

[60] Parrinello, M., and A. Rahman, 1984, J. Chem. Phys. **80**, 860; deRaedt, B., L. M. Sprik, and M. L. Klein, 1984, J. Chem. Phys. **80**, 5719.

[61] Tuckerman, M. E., B. J. Berne, G. J. Martyna and M. L. Klein, 1993, J. Chem. Phys. **99**, 2796.

[62] Feynman, R. P., and A. R. Hibbs, 1965, *Quantum Mechanics and Path Integrals*, (McGraw-Hill, New York)

[63] Takahashi,M., and M. Imada, 1984, J. Phys. Soc. Jpn. **53**, 963; Runge, K. J., and G. V. Chester, 1988, Phys. Rev. **B 38** , 135.

[64] Coalson, R. D., D. L. Freeman, and J. D. Doll, 1985, J. Chem. Phys., **85** 4567.

[65] Herman, M. F., E. J. Bruskin, and B. J. Berne, 1982, J. Chem. Phys. **76**, 5150.

[66] Ceperley, D. M., "Path Integral Monte Carlo Methods for Fermions" in "Simulation in Condensed Matter Physics and Chemistry", Ed. K. Binder and G. Ciccotti, 1996.

[67] Ceperley, D. M., 1992, Phys. Rev. Lett. **69**, 331.