67

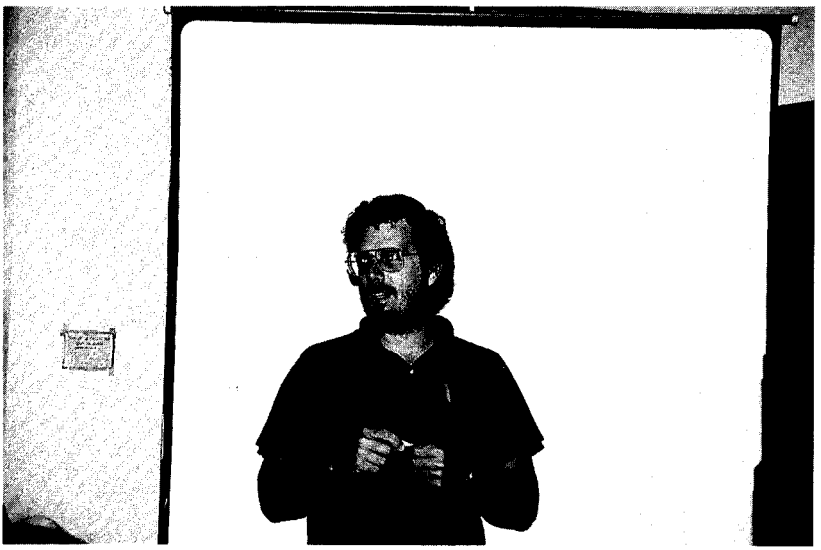
COURSE 9

QUANTUM MONTE CARLO METHODS FOR FERMIONS

D.M. Ceperley

Department of Physics,
National Center for Supercomputing Applications,
Beckman Institute,
University of Illinois at Urbana-Champaign,
1110 W. Green St., Urbana, IL 61801, USA

B. Douçot and J. Zinn-Justin, eds.
Les Houches, Session LVI, 1991
Strongly Interacting Fermions and
High T_c Superconductivity
© 1995 Elsevier Science B.V. All rights reserved



Contents

1. Introduction 431
   1.1. Random walks 432
2. Variational Monte Carlo 432
   2.1. The pair-product trial function 434
   2.2. Details 434
   2.3. Optimization of trial functions 435
   2.4. Beyond the pair-product trial function 437
   2.5. Problems with variational methods 439
3. Projector Monte Carlo 440
   3.1. Importance sampling 442
   3.2. The fixed-node method 444
   3.3. Exact fermion methods 446
   3.4. Lattice models 449
   3.5. Problems with projection methods 450

References 452

1. Introduction

In this chapter, I briefly review some of the quantum Monte Carlo methods that have been used to calculate properties of many-fermion systems. It is hoped that QMC will be useful in providing exact results, or at least exact constraints, on properties of many-body fermion 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 discussed at this school.

This will not be an exhaustive review. In variance with the other lectures, I will primarily discuss continuum models, not lattice models, although most of the techniques can be carried over directly. Monte Carlo techniques appropriate to spin systems are discussed by Young. As an example I will discuss applications of these methods to liquid helium. More extensive discussion of these topics can be found in refs. [1–3].

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:

$$\mathcal{H} = -\lambda \sum_{i=1}^{N} \nabla_i^2 + \sum_{i<j} v(r_{ij}), \tag{1.1}$$

where $\lambda = \hbar^2/2m$ and $v(r)$ is a two-body pair potential. I will stick to the first-quantized notation in the canonical ensemble. A boson wavefunction is then totally symmetrical under particle exchange and a fermion function is antisymmetrical. The permutation operator acting on particle labels is denoted by 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 three spatial and one 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 wavefunction will be denoted by $\Psi(R)$. The symbol $\int$ will imply an integral over the full $3N$-dimensional configuration space of the particles.

1.1. Random walks

Monte Carlo methods for many-body systems are exclusively examples of Markov processes. Let us briefly review the basic concepts of random walks. Let $\mathcal{S}$ be a state space and let $(s_0, s_1, \ldots)$ be a random walk in that space. The choice of the state space will vary depending on the method, but for now let s_i represent the $3N$ coordinates of all the particles. In the simplest Markov process, there is a constant transition probability for generating state b given that the walk is presently in state a, which we will denote by P_{ba}. The transition probability, or moving rule, generates the walk. Under very general conditions [4], the asymptotic probability distribution of the walk converges exponentially fast to a unique distribution:

$$\mathcal{P}(s_n) \to \Pi(s_n). \tag{1.2}$$

In projector Monte Carlo, to be discussed in section 3, the transition rules are set up so that the asymptotic population is the ground-state wavefunction for a given Hamiltonian. But, let us first review the Metropolis [5,6] rejection method, where moving the particles is a two step process; first one samples a trial position (state a') from a transition probability $T_{a'a}$, then this trial move is either accepted (i.e. $b = a'$) or rejected (i.e. $b = a$), with a probability given by:

$$\min\left(1, \frac{\Pi_{a'} T_{aa'}}{\Pi_a T_{a'a}}\right). \tag{1.3}$$

The acceptance probability has been chosen to satisfy the detailed-balance relation, which implies that its asymptotic probability will converge to Π_a independent of the transition rules $T_{aa'}$. The rate of convergence or the efficiency of the walk in sampling-state space is very much determined by the transition rules. The rejection method is appropriate when one wants to sample a known, computable function. If one had an exact analytic expression for the many-body wavefunction, then it would be straightforward to use this method to determine quantum expectation values in that state. However, such is not the case, and one is forced to resort to either more complicated or more approximate methods. The Metropolis method is discussed in more detail by Young in this volume.

2. Variational Monte Carlo

The variational Monte Carlo method was first used by McMillan [7] to calculate the ground-state properties of liquid ^{4}He and then generalized to fermion systems by Ceperley et al. [8]. 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:

$$E_V = \frac{\int \Psi^*(R)\mathcal{H}\Psi(R)}{\int \Psi^*(R)\Psi(R)} \geqslant E_0. \tag{2.1}$$

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

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

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

(3) The integrals $\int \Psi^2$, $\int \Psi^2\mathcal{H}\Psi$, 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 in the energy will be infinite.

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

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

$$\Pi(R) = \frac{|\Psi(R)|^2}{\int |\Psi(R)|^2}. \tag{2.2}$$

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

$$E_V = \int \Pi(R)E_L(R) = \langle E_L(R)\rangle, \tag{2.3}$$

where the local energy of Ψ is defined as:

$$E_L(R) = \Psi^{-1}\mathcal{H}\Psi(R). \tag{2.4}$$

Variational Monte Carlo (VMC) has a very important "zero-variance property": as the trial function approaches an exact eigenfunction, $\Psi \to \phi_\alpha$, the local energy approaches the eigenvalue everywhere, $E_L(R) \to E_\alpha$, and the Monte Carlo estimate of the variational energy converges more rapidly with the number of steps 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*

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

$$\Psi(R,\sigma) = \exp\left[-\sum_{i<j} u(r_{ij})\right] \det[\theta_k(r_i,\sigma_i)], \tag{2.5}$$

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 Jastrow [9] wavefunction, although Bijl [10] much earlier described the motivation for its use in liquid ^{4}He. Closely related forms are the Gutzwiller function for a lattice and 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 it best to move the particles one at a time, by adding a random vector to the particle's coordinates, where the vector is either uniform inside a cube centered about the old coordinates, or is a normally distributed random vector. Assuming the first kind of move for the ith particle, the trial move is accepted with probability:

$$|\Psi(R')/\Psi(R)|^2 = \exp\left\{-2\sum_{j\neq i}[u(r_i'-r_j)-u(r_i-r_j)]\right\} \times\left|\sum_k \theta_k(r_i')C_{ki}\right|^2, \tag{2.6}$$

where the matrix C is the transposed inverse of the Slater matrix. 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 [8], which

takes $O(N^2)$ operations. Hence, to attempt moves for all N particles (a pass) takes $O(N^3)$ operations.

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

$$E_L(R) = V(R) + \lambda \sum_i \left[\nabla_i^2 U - \sum_k \nabla_i^2 \theta_k(r_i) C_{ki} - G_i^2 \right], \tag{2.7}$$

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 of 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 [8] 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 estimation of eq. (2.7) 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.

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 for the quantity to optimize, and it is not yet clear which is best.

– The variational energy: E_V. Clearly, one minimizes E_V if the objective 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 [11].

– The dispersion of the local energy: $\int [(\mathcal{H} - E_V)\Psi]^2$. If we assume that every step of 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 that of the variational energy, since it is a positive definite quantity.

– The overlap with the exact wavefunction: $\int \Psi\phi$. This is equivalent to finding the trial function which is closest to the exact wavefunction 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. Assume that the spin orbits come from an exact solution for some model potential.

Consider bringing two particles together and let us examine the dominant terms in the local energy. In a good trial function the singularities in the kinetic energy should cancel the potential. The local energy will have the form:

$$E_L(R) = v(r) + 2\lambda \nabla^2 u(r) - 2\lambda[\nabla u(r)]^2 + \cdots, \tag{2.8}$$

where r is the distance separating the particles. An intuitive result emerges: $e^{-u(r)}$ will equal the solution to the two-body Schrödinger equation. For He atoms interacting with the 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.

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:

$$E_V = E_F + \sum_k (v_k - \lambda k^2 u_k)(S_k - 1), \tag{2.9}$$

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} the structure factor for uncorrelated fermions, we obtain [12] the optimal pseudopotential at long wavelengths:

$$2\rho u_k = -\frac{1}{S_{0k}} + \left(\frac{1}{S_{0k}} + \frac{2\rho v_k}{\lambda k^2}\right)^{1/2}. \tag{2.10}$$

For a short-ranged potential (e.g. liquid helium), v_k can be replaced by a constant and we find the Reatto–Chester [13] 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 the correlation will extend beyond the simulation box. The ground-state energy is little affected by this tail in the wavefunction, 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 [14,12] 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 of such methods 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, etc., 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 the trial-function parameters is very poorly calculated. Also, it is difficult to optimize by hand functions involving more than three variational parameters. A correlated sampling method, known as reweighting [1,8], solves this problem.

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.

(1) One can replace the spin orbitals with pairing functions. For example, if the particles are paired in a spin singlet state, one obtains a determinant of the form:

$$\det[\chi_s(r_i \uparrow, r_j \downarrow)], \tag{2.11}$$

or, if the particles are paired in a spin triplet:

$$\mathcal{A} \prod_{i=1}^{N/2} \chi_p(r_{2i}, r_{2i-1}), \tag{2.12}$$

where $\mathcal{A}$ is an antisymmetrizer. Bouchaud and Lhuiller [15] have pointed out that this object is a Pfaffian, and hence one can use the theorem that the square of a Pfaffian is a determinant, to sample it with VMC.

(2) The dominant term missing in the trial function for liquid ^{4}He, is a three-body term with the functional form of a squared force:

$$U_3(R) = -\sum_i \left[\sum_j \xi(r_{ij}) \boldsymbol{r}_{ij} \right]^2. \tag{2.13}$$

The form makes it particularly rapid to compute. It is no slower than a two-body function.

Table I

The energies of liquid ^{3}He in K/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 that back-flow terms were included. BCS refers to the spin-paired trial functions in eqs. (2.11), (2.12). 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 function Monte Carlo method, which will be described in section 3.2. The numbers in parenthesis represent the statistical error in units of 0.01 K.

Terms	E_V (K/atom)	$(E_V - E_0)/(2T)$ (%)	E_{FN} (K/atom)	Ref.
u	−1.08(3)	5.7	−1.95(3)	[16]
u, ξ	−1.61(3)	4.6	−1.95(3)	[16]
u, η	−1.55(4)	3.7	−2.37(1)	[16]
u, ξ, η	−2.15(3)	1.3	−2.37(1)	[17]
BCS, $s = 0$	−1.2			[18]
BCS, $s = 1$	−2.05			[18]
Exp.	−2.47	0.0	−2.47	

(3) For liquid ^{3}He the dominant correction for the pair-product trial function is the modification of the spin orbitals to include back-flow correlations. The particle coordinates in the Slater determinant become "quasiparticle" coordinates:

$$\det[\theta_k(\boldsymbol{x}_i, \sigma_i)], \tag{2.14}$$

where the "quasiparticle" coordinates are defined by: $\boldsymbol{x}_i = \boldsymbol{r}_i + \sum_j \eta(r_{ij})\boldsymbol{r}_{ij}$. Back-flow is needed to satisfy local current conservation.

Table I gives VMC energies for a variety of trial functions. It is important to realize that the kinetic and potential energies are almost completely canceling out, liquid ^{3}He is very weakly bound. The third column, $(E_V - E_0)/(2T)$, is a measure of the accuracy of the trial function, where $T = 12.3$ K is the kinetic energy and $E_0 = -2.47$ K 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 interatomic 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 back-flow terms. The results with BCS paired trial functions should be compared with

the other results with caution, since a different interatomic potential was used and the corrections for size effects are very large.

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. By 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 should be computable in a reasonable amount of time. 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 the 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 for helium at zero temperature. The solid wavefunction is simpler than the liquid wavefunction since in the solid the particles are localized, so that the phase space that the atoms explore is much reduced. This means that if you compare 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 energy, 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, shown in fig. 1. The wavefunction 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. 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 one range of 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 function completely incorrect.

– 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

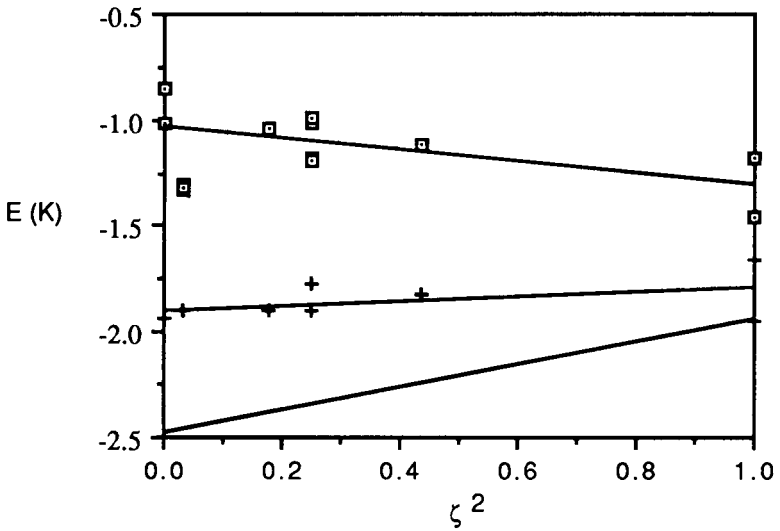


Fig. 1. The energy of liquid ^{3}He at zero pressure and zero temperature as a function of the square of the spin polarization. Zero is unpolarized and unity fully polarized. The boxes with a linear fit are variational Monte Carlo calculations with the pair-product trial function. The plusses (+) and fitted line are from the fixed-node diffusion Monte Carlo calculation with the same trial function. The scatter results from the fact that no size-dependence corrections have been made to the Monte Carlo data. The lowest line is the extrapolation of a low-polarization experimental measurement of the spin polarization energy. Recent experiments at high polarization indicate that the ground-state energy is somewhat below this line.

will also have a Fermi surface, although it will be renormalized. This does not imply that the true wavefunction has a sharp Fermi surface. Only if localized spin-orbitals are used will a gap appear.

3. Projector Monte Carlo

In the last section, I discussed the variational Monte Carlo method. Now I will turn to a potentially more powerful method, where a function of the Hamiltonian projects out the ground state, hence the name, projector Monte Carlo. In fact, the nomenclature of the various quantum Monte Carlo methods is not at all standardized. Table 2 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.

Table 2

The Green functions for various projection methods. Here τ is the time step and E_T is the trial energy.

Method	$G(R, R')$	Refs.
Diffusion (DMC)	$\exp[-\tau(\mathcal{H} - E_T)]$	[19,22,23]
Green's function (GFMC)	$[1 + \tau(\mathcal{H} - E_T)]^{-1}$	[20,21]
Power (PMC)	$[1 - \tau(\mathcal{H} - E_T)]$	[24]

A sequence of trial functions is defined by applying the projector, $G(R, R')$:

$$\psi_{n+1}(R) = \mathrm{e}^{-\tau(\mathcal{H}-E_T)}\,\psi_n(R) = \int \mathrm{d}R'\, G(R, R')\psi_n(R'), \tag{3.1}$$

with the initial condition $\psi_0(R) = \Psi(R)$. The effect on the trial function of the Green function is seen by expanding the trial function in the exact eigenfunctions of the Hamiltonian:

$$\psi_n(R) = \sum_\alpha \phi_\alpha(R)\,\langle\phi_\alpha|\Psi\rangle\,\mathrm{e}^{-n\tau(E_\alpha - E_T)}. \tag{3.2}$$

The Green function will project out the state of lowest energy having non-zero overlap with the initial trial function.

$$\lim_{n\to\infty} \psi_n(R) = \phi_0(R)\,\langle\phi_0|\Psi\rangle\,\mathrm{e}^{-n\tau(E_0 - E_T)}. \tag{3.3}$$

The role of the trial energy is to keep the overall normalization of ψ_n fixed, which implies $E_T \approx E_0$. The time step, τ, controls the rate of convergence to the ground state.

The projection can only be done directly for few-body systems. For many-body systems, the trial function and the Green function are sampled. For the moment, let us discuss the case where Ψ is non-negative, the boson case. In the limit that the time step approaches zero, a coordinate-space representation of the Green function is:

$$\langle R|\,\mathrm{e}^{-\tau(\mathcal{H}-E_T)}\,|R'\rangle = (4\pi\lambda\tau)^{-3N/2}\exp\left(-\frac{(R-R')^2}{4\lambda\tau}\right)\exp[-\tau(V(R) - E_T)]. \tag{3.4}$$

The iteration equation, eq. (3.1), has a simple interpretation in terms of branching

random walks, since the first factor is the Green function for diffusion and the second is multiplication of the distribution by a positive scalar.

An ensemble of configurations is constructed with a Metropolis sampling procedure for $\Psi(R)$. This is called the zeroth generation, i.e. $n = 0$, and 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. (3.4) and then branching. The number of copies of R' in the next generation is the integer part of $u + \exp[-\tau(V(R) - E_\mathrm{T})]$, 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 future generations, these walks propagate independently of each other. In places of high potential energy random walks are terminated.

The above procedure, depicted in fig. 2, is a Markov process where the state of the walk in the nth generation is $\{P_n; R_1, R_2, \ldots, R_{P_n}\}$. Hence, it has a unique stationary distribution, constructed to be the ground-state wavefunction. The number of walkers fluctuates from step to step. The trial energy, E_T, must be adjusted to keep the population within computationally acceptable limits.

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 [25]. But it fails on many-body systems since the potential is unbounded. For example, a Coulomb potential can go to infinity both in the positive and negative direction. Even with a bounded potential the method becomes very inefficient as the number of particles increases. But there is a very simple cure, discovered by Kalos [21], for GFMC, but equally applicable to any projector method. In importance sampling the underlying probability distribution is multiplied by a known, approximate solution. Multiply eq. (3.1) by Ψ, the trial function, and let $f_n(R) = \Psi(R)\psi_n(R)$. Then:

$$f_{n+1} = \Psi\, \mathrm{e}^{-\tau(\mathcal{H}-E_\mathrm{T})}\, \psi_n = \int \mathrm{d}R'\, \widetilde{G}(R, R') f_n(R'), \tag{3.5}$$

where $\widetilde{G}(R, R') = \Psi^{-1}\, \mathrm{e}^{-\tau(\mathcal{H}-E_\mathrm{T})}\, \Psi$ is the importance-sampled Green function, and the initial condition is $f_0(R) = \Psi^2(R)$. It is easily shown by differentiating $\widetilde{G}$ with respect to τ, that it satisfies the evolution equation:

$$-\frac{\partial \widetilde{G}(R, R_0; \tau)}{\partial \tau} = -\sum_i \lambda_i\, \boldsymbol{\nabla}_i[\boldsymbol{\nabla}_i \widetilde{G} + 2\widetilde{G}\, \boldsymbol{\nabla}_i \ln \Psi(R)] + [E_\mathrm{L}(R) - E_\mathrm{T}]\widetilde{G}, \tag{3.6}$$

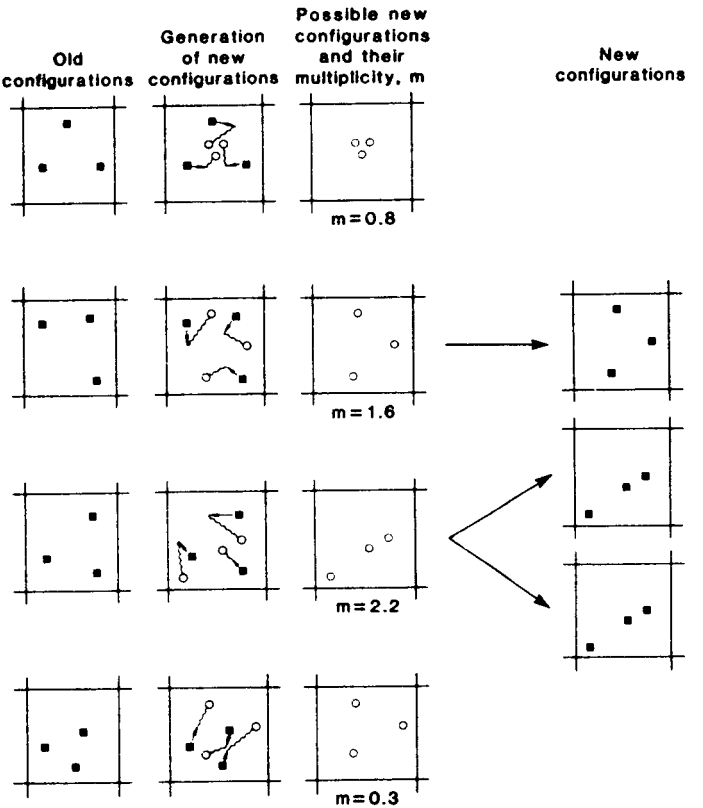


Fig. 2. Schematic of the DMC algorithm, showing the evolution of the random walks which make up a generation. Illustrated is a three-particle system in a box; the squares represent old positions, the circles new ones. The old population has four walkers, the new population has three. The process leading to the new generation consists of drifting of the electron positions with the gradient of the log of the guide function (straight arrows), adding a random diffusion (wiggly lines), and branching with the local energy. With the three new walks the process is repeated.

where E_L is the local energy defined in eq. (2.4). The first three terms on the right-hand side correspond to diffusion, drifting, 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, as illustrated in fig. 2, 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:

$$R_{n+1} = R_n + \chi + \lambda\tau \nabla \ln \Psi(R_n)^2, \tag{3.7}$$

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

$$\exp\{-\tau[E_{\mathrm{L}}(R_n) - E_{\mathrm{T}}]\} + u, \tag{3.8}$$

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

(4) The energy is calculated as the average value of the local energy: $E_0 = \langle E_{\mathrm{L}}(R_n)\rangle$.

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

(6) To obtain ground-state expectation values of quantities other than the energy (e.g. the potential energy), one should correct the average over the DMC walk, the so-called "mixed estimator", $V_{\mathrm{mix}} = \langle\phi_0|V|\Psi\rangle$, by using the variational estimator [1],

$$\langle\phi_0|V|\phi_0\rangle = 2\langle\phi_0|V|\Psi\rangle - \langle\Psi|V|\Psi\rangle + \mathrm{O}[(\phi_0 - \Psi)^2]. \tag{3.9}$$

If the mixed estimator equals the variational estimator, then the trial function has maximum overlap with the ground state.

In the GFMC algorithm there is no error resulting from taking a finite time step, 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 time step. 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.

3.2. *The fixed-node method*

We have not discussed at all the problem posed by Fermi statistics for projector Monte Carlo. Consider the difficulty in implementing the non-importance-sampled algorithm: the initial condition is not a probability distribution, since a fermion trial function will have positive and negative pieces. Hence, we must use the initial sign of the wavefunction as a weight for the random walk. That leads to an exact, but slowly converging, algorithm, which we will discuss in the next subsection.

Consider the effects of using an antisymmetric trial function in the importance-sampled algorithm. The initial distribution is positive, but the Green function, $\widetilde{G}(R, R')$, 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 of all estimators. But there is a simple way to avoid this: 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 gives a zero boundary condition for the evolution equation for the probability. This solves the wave equation with a boundary condition that it vanishes wherever the trial function vanishes. One can easily demonstrate that the resulting energy will be an upper bound to the exact ground-state energy [26]; the best possible upper bound with the given boundary conditions. With the FN method we do not necessarily get the exact fermion energy, but the results are 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 not changed 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 "fermion" pressure on the two sides, unless the nodes are exact. Where comparison has been done between the VMC, the FN-DMC, and the exact answer, 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. The ground-state wavefunction 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 very 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 wavefunction vanishes. These coincident planes, with $r_i = r_j$, 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 one dimension, where the set of nodes is usually equal to the set of coincident hyperplanes. Fermions in one dimension are equivalent to 1D bosons with a no-exchange rule.

Nodal volumes of ground-state wavefunctions have a tiling property [27]. 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 alike 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 wavefunction which is the ground state for some local Hamiltonian, and it can be proved by a simple variational argument. Excited states, ground states of non-local Hamiltonians, or arbitrary antisymmetric functions need not have the tiling property. A more ex-

tensive discussion of fermion nodes and some pictures of cross sections of free particle nodes are given in ref. [27].

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 is only improving the bosonic correlations of the trial function, not the fermion features. There are some fairly simple ways of improving on the FN method, but their use is limited to small systems.

With the transient-estimate (TE) method, one calculates the ratio:

$$E_{\mathrm{TE}}(t) = \frac{\int \Psi \mathcal{H}\, \mathrm{e}^{-t(\mathcal{H}-E_{\mathrm{T}})}\, \Psi}{\int \Psi\, \mathrm{e}^{-t(\mathcal{H}-E_{\mathrm{T}})}\, \Psi}. \tag{3.10}$$

Clearly the variational theorem applies: $E_{\mathrm{TE}}(t) \geqslant E_0$. Also, the energy converges exponentially fast in t:

$$\lim_{t\to\infty} E_{\mathrm{TE}}(t) = E_0 + \mathrm{O}(\mathrm{e}^{-tE_{\mathrm{g}}}), \tag{3.11}$$

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, and would be obtained by creating two particle–hole excitations.

For a method to self-consistently find its own nodes, the walks must be able to go anywhere, and so the drift term in eq. (3.6) 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 the strictly positive guide function, $\Psi_{\mathrm{G}}(R)$, used to guide the walks. The guide function appears in the drift and branching terms of eq. (3.6) and will be assumed to be a reasonable boson ground state trial function, while the trial function appears in eq. (3.10). The Ψ_{G} importance-sampled Green function is:

$$\widetilde{G}(R, R'; t) = \Psi_{\mathrm{G}}(R)\langle R|\, \mathrm{e}^{-t(\mathcal{H}-E_{\mathrm{T}})}\, |R'\rangle \Psi_{\mathrm{G}}^{-1}(R'), \tag{3.12}$$

and we can rewrite eq. (3.10) as:

$$E_{\mathrm{TE}}(t) = \frac{\int \sigma(R) E_{\mathrm{LT}}(R) \widetilde{G}(R, R'; t) \sigma(R') \Psi_{\mathrm{G}}^2(R')}{\int \sigma(R) \widetilde{G}(R, R'; t) \sigma(R') \Psi_{\mathrm{G}}^2(R')}, \tag{3.13}$$

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

The transient-estimate algorithm is:

(1) Sample configuration R' from the square of the guide function with VMC. This corresponds to the rightmost factor in the above integrands.

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

(3) Propagate the walk forward an amount of time, t, with the Green 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:

$$E_{TE}(t) = \frac{\langle [E_{LT}(R) + E_{LT}(R')]\sigma(R)\sigma(R')\rangle}{2\langle \sigma(R)\sigma(R')\rangle}, \tag{3.14}$$

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

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. The trial function nodes are displaced by an unequal diffusion of walks from one side or the other.

The release-node (RN) algorithm [26,28] is an improvement on this TE method. Instead of projecting from the trial function, one begins the projection from the fixed-node solution. There are several advantages. First of all, the 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 (of course this will indirectly affect the bosonic correlation). Second, one can directly calculate the difference between the exact result and the fixed-node solution. It turns out that it is given by the local energy of the 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 introducing a guide function and keeping track of the energy as a function of time since nodal crossing.

However, there are serious problems with both the TE and the RN method. Let us examine how the statistical error of eq. (3.10) depends on the projection time. Note that the values 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, since 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. One can demonstrate that the signal-to-noise ratio vanishes exponentially fast. This is a general result. In any fermion scheme, as soon as

negative weights are introduced the statistical error will grow as:

$$\epsilon_{\text{stat}} = \mathrm{e}^{-t(E_{\text{F}}-E_{\text{B}})}. \tag{3.15}$$

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 up to 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. (3.11) and (3.15), we find that the total error will decrease as:

$$\epsilon \propto P^{-\eta}, \quad \eta = \frac{E_{\text{g}}}{2(E_{\text{F}} - E_{\text{B}} + E_{\text{g}})}, \tag{3.16}$$

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

One possibility for improving this convergence is to use all of the information given in the function $E_{\text{TE}}(t)$, 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. Recently it has been suggested [29] in the context of quantum Monte Carlo for lattice models, that the proper way to perform such a functional 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. Recently we [30] have applied these ideas to the TE and RN method with simple problems and showed that they do indeed reduce the statistical and systematic errors. When this is done the statistical errors again decrease with $\eta = -\frac{1}{2}$.

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

There is some confusion about the nature of the "fermion" problem in the literature. Note that the TE and RN method do converge to the exact fermion energy. The fermion problem has to do with how long it takes to achieve a given error estimate, and, more precisely, how this scales with the number of fermions: the computational complexity of the calculation. Clearly, one of the important tasks of simulations is to calculate properties of systems near phase transitions, so the ability to do this for large systems is important. 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 a time $O(N)$. Simulations of equilibrium properties of quantum bosons at zero or non-zero temperature are also $O(N)$. A Heisenberg model on a bipartite lattice or any 1D fermion system is $O(N)$. Variational MC calculations of fermions systems are $O(N^3)$ in general, but the exponent would be smaller if localized spin-orbits are used. The Hubbard model at half filling on a bipartite lattice [31] is $O(N^3)$, using the projection Monte Carlo method and auxiliary field techniques. This is the only non-trivial "fermion" problem solved. But known algorithms for general fermion systems are $O(e^{\kappa N})$. Barring a breakthrough, one can still reduce the rate of exponential growth, κ, or use the TE or RN method 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. First, it is convenient to use the power Green function to project out the ground state. The Hamiltonian can directly be used to hop the spins without time step error if the time step is chosen

$$\tau \leqslant \frac{2}{E_{\max} - E_T}, \tag{3.17}$$

where $E_{\max}$ is the maximum energy. Since lattice models have a bounded energy spectrum, the power method is appropriate. 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 calculation of a lattice model. Importance sampling enters in the same way; details can be found in ref. [24].

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. The problem occurs when there are two many-body configurations (s, s') with $\Psi(s)\Psi(s') < 0$ and $\langle s|\mathcal{H}|s'\rangle \neq 0$. If this is possible, the fixed-node energy cannot, in general, be an upper bound to the energy since these unknown "surface terms" can contribute to the energy. If one thinks that the lattice model is a reasonable approximation to a continuum model, the fixed-node approximation should still be useful, and it seems to work well in the one case that has been tried [32]. Going to a lattice does not change the TE and RN method at all. In fact the RN method is a useful way of estimating the surface terms left out in the lattice FN method.

The conditions for a lattice model to be bosonic, and not to have a sign problem, are easy to state. The Green 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:

$$\langle s|\mathcal{H}|s'\rangle \leqslant 0, \quad \forall\, s \neq s'. \tag{3.18}$$

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 all possible local bases to see if some of the other interesting lattices might be solvable. This approach is, of course, equivalent to finding a "Marshall sign rule" for a lattice model.

What I have not discussed are the usual methods for performing simulations of lattice models. These are based on applying the Stratonovitch–Hubbard transformation [31] to $e^{-t\mathcal{H}}$. An auxiliary field is introduced in place of the electron–electron interaction. Several recent reviews are given in ref. [33]. In general the sign problems remain.

3.5. Problems with projection methods

The projection method shares many of the problems with the variational method. In fact it is useful to think of the projection method as a "super-variational" method. Figure 3 shows how the statistical and systematic error decrease with computer time in both the variational and the fixed-node method. The dashed lines show what would happen to the diagram if the trial function were improved. In both methods there is a premium for good trial functions; that is the

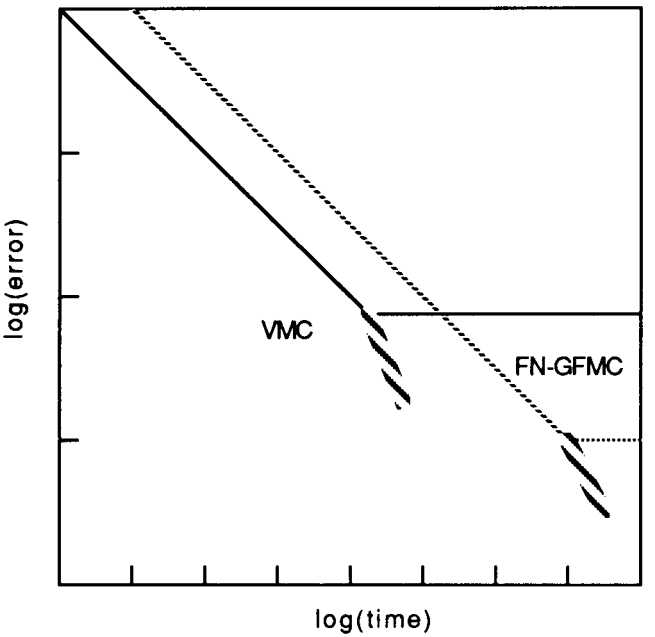


Fig. 3. The rate of decrease of the statistical and systematic error in VMC (solid line) and FN (dotted line) versus computer time. In VMC the total error initially decreases as $P^{-1/2}$ until the statistical error equals the intrinsic error of the variational trial function. The "critical time" is where the line becomes horizontal. A similar history is followed with the FN-DMC, but the method is slower, hence the curve is displaced to the right, and the intrinsic error of the FNA is an order of magnitude smaller. The dashed lines show the effect on the critical time of using an improved trial function.

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 this is a good way of learning what is missing in the trial function. But it is slower than VMC because the time step needs to be smaller. The cost in computer time is typically a factor of two to ten.

– 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 bound-

ary. 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. Finite-temperature Metropolis methods do 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 superior to projector Monte Carlo for calculating correlation functions.

References

[1] D.M. Ceperley and M.H. Kalos, in: *Monte Carlo Methods in Statistical Physics*, ed. K. Binder (Springer Verlag, 1979).
[2] K.E. Schmidt and M.H. Kalos, in: *Monte Carlo Methods in Statistical Physics II, Topics in Current Physics*, ed. K. Binder (Springer Verlag, 1984).
[3] K.E. Schmidt and D.M. Ceperley, in: *Monte Carlo Methods in Condensed Matter Physics*, ed. K. Binder (Springer Verlag, 1992).
[4] J.M. Hammersley and D.C. Handscomb, *Monte Carlo Methods* (Chapman and Hall, 1964).
[5] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, J. Chem. Phys. **21** (1953) 1087.
[6] M.H. Kalos and P.A. Whitlock, *Monte Carlo Methods* (Wiley, 1986).
[7] W.L. McMillan, Phys. Rev. A **138** (1965) 442.
[8] D.M. Ceperley, G.V. Chester and M.H. Kalos, Phys. Rev. B **16** (1977) 3081.
[9] R. Jastrow, Phys. Rev. **98** (1955) 1479.
[10] A. Bijl, Physica **7** (1940) 869.

[11] D. Ceperley, J. Stat. Phys. **43** (1986) 815.
[12] D.M. Ceperley, Phys. Rev. B **18** (1978) 3126.
[13] L. Reatto and G.V. Chester, Phys. Lett. **22** (1966) 276.
[14] M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids* (Oxford, 1987).
[15] J.P. Bouchaud and C. Lhuillier, Europhys. Lett. **3** (1987) 1273.
[16] K.E. Schmidt, M.A. Lee, M.H. Kalos and G.V. Chester, Phys. Rev. Lett. **47** (1981) 807.
[17] R.M. Panoff and J. Carlson, Phys. Rev. Lett. **62** (1989) 1130.
[18] J.P. Bouchaud and C. Lhuillier, in: *Spin Polarized Quantum Systems*, ed. S. Stringari (World Scientific, Heidelberg, 1989).
[19] N. Metropolis and S. Ulam, J. Am. Stat. Assoc. **44** (1949) 247.
[20] M.H. Kalos, Phys. Rev. **128** (1962) 1791; J. Comp. Phys. **2** (1967) 257.
[21] M.H. Kalos, D. Levesque and L. Verlet, Phys. Rev. A **9** (1974) 2178.
[22] J.B. Anderson, J. Chem. Phys. **63** (1975) 1499; J. Chem. Phys. **65** (1976) 4122.
[23] D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. **45** (1980).
[24] N. Trivedi and D.M. Ceperley, Phys. Rev. B **41** (1990) 4552.
[25] M.D. Donsker and M. Kac, J. Res. Natl. Bur. Stand. **44** (1950) 551.
[26] D.M. Ceperley, in: *Recent Progress in Many-Body Theories*, ed. J. Zabolitzky (Springer Verlag, 1981).
[27] D.M. Ceperley, J. Stat. Phys. **63** (1991) 1237.
[28] D.M. Ceperley and B.J. Alder, J. Chem. Phys. **81** (1984) 5833.
[29] R.N. Silver, D.S. Sivia and J.E. Gubernatis, Phys. Rev. B **41** (1990) 2380.
[30] M. Caffarel and D. Ceperley, J. Chem. Phys. **97** (1992) 8415.
[31] J. Hirsch, Phys. Rev. B **28** (1983) 4059; Phys. Rev. Lett. **51** (1983) 1900; Phys. Rev. B **31** (1985) 4403.
[32] G. An and J.M.J. van Leeuwen, Phys. Rev. B **44** (1991) 9140.
[33] H. De Raedt and W. von der Linden, in: *Monte Carlo Methods in Condensed Matter Physics*, ed. K. Binder (Springer Verlag, 1992).