Path integrals in the theory of condensed helium

D. M. Ceperley*

National Center for Supercomputer Applications, Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

One of Feynman's early applications of path integrals was to superfluid ^{4}He. He showed that the thermodynamic properties of Bose systems are exactly equivalent to those of a peculiar type of interacting classical "ring polymer." Using this mapping, one can generalize Monte Carlo simulation techniques commonly used for classical systems to simulate boson systems. In this review, the author introduces this picture of a boson superfluid and shows how superfluidity and Bose condensation manifest themselves. He shows the excellent agreement between simulations and experimental measurements on liquid and solid helium for such quantities as pair correlations, the superfluid density, the energy, and the momentum distribution. Major aspects of computational techniques developed for a boson superfluid are discussed: the construction of more accurate approximate density matrices to reduce the number of points on the path integral, sampling techniques to move through the space of exchanges and paths quickly, and the construction of estimators for various properties such as the energy, the momentum distribution, the superfluid density, and the exchange frequency in a quantum crystal. Finally the path-integral Monte Carlo method is compared to other quantum Monte Carlo methods.

CONTENTS

*Electronic address: ceperley@ncsa.uiuc.edu

I. INTRODUCTION

The quantum mechanics of many-body systems is usually presented as a difficult subject, and the phenomena of Bose condensation and superfluidity are often characterized as ill understood. One of Feynman's early suc-

cesses (1953) with path integrals is often neglected, his mapping with path integrals of a quantum system onto a classical model of interacting "polymers." The polymers are ring exchanges of bosons in imaginary time. This gives us a simple classical picture for a superfluid. Not only is it simple, but it is exact for all thermodynamic properties. Feynman showed that, even though the atomic interaction between helium atoms is strong, Bose condensation still occurs at a lower temperature than the free-particle transition. This review article is an update of Feynman's imaginary-time paths applied to low-temperature helium.

Besides providing a physical picture, the path-integral approach translates directly into a computational technique that allows us to compute all sorts of detailed microscopic properties of helium: the energy, the pressure, the momentum distribution, the superfluid density, and the interatomic correlations, all without any essential approximations. While the intuitive picture of path integrals helps us to understand superfluidity, it also makes path integrals into a powerful computational tool. With recent advances in computer hardware and algorithms, it is now possible to do respectable exact simulations of Bose superfluids on modest work stations. In this introductory review, we hope to convince the reader of the simplicity and elegance of the approach and to explain the details of how the method is used.

First, let us digress a bit to review the basic facts concerning helium. In the realm of low-temperature physics, a helium atom can be considered an elementary particle. The lowest electronic excitation of a helium atom has an energy of 2.3×10^5 K above the ground state. Hence, for temperatures on the order of 1 K, where quantum effects of the atoms are important, it is a *very* good approximation to assume that the helium atoms are in their electronic ground state and, except for nuclear spin, we can ignore the fact that a helium atom is made of a nucleus and two electrons. A ^{4}He atom has the statistical properties of a boson, in contrast to a ^{3}He atom, which is a spin-$\frac{1}{2}$ fermion.

The interaction between helium atoms is well described by a nonrelativistic Hamiltonian of atoms interacting by a pair potential,

$$\mathcal{H} = -\lambda \sum_{i=1}^{N} \nabla_i^2 + \sum_{i<j} v(|\mathbf{r_i} - \mathbf{r_j}|), \qquad (1.1)$$

where N is the number of particles and $\lambda \equiv \hbar^2/2m = 6.0596$ Å^2 K for ^{4}He. The pair interaction $v(r)$ is known (Aziz *et al.*, 1992) both from theoretical calculations and from interpretation of experiments in the gas phase and is shown in Fig. 1. The salient features are a very strong repulsion for distances less than 2.5 Å and a very weak van der Waals attraction at larger distances, with a well depth of 10.9 K±0.1 K. Also shown is the "standard" Lennard-Jones potential. Contributions to the potential of explicit three-body and higher-order interactions are non-negligible, particularly as the density is increased.
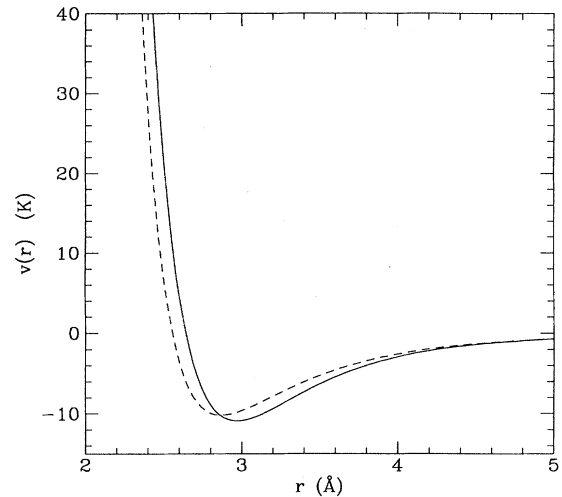
FIG. 1. The semiempirical pair potential between two helium atoms: solid line, Aziz *et al.* (1992); dashed line, Lennard-Jones 6-12 potential with $\epsilon = 10.22$ K and $\sigma = 2.556$ Å.

To judge the accuracy of these potentials we might consider how the calculations of the ground-state energy at zero pressure have evolved in time. The experimental binding energy at zero temperature and zero pressure is -7.17 K/atom. This relatively small binding energy is a result of a cancellation between a kinetic energy of about 14.5 K/atom and a potential energy of -21.7 K/atom. Green's-function quantum Monte Carlo calculations with the Aziz-79 potential (Aziz *et al.*, 1979) give a binding energy of -7.12 K/atom (Kalos *et al.*, 1981), while calculations with the Aziz-87 potential (Aziz *et al.*, 1987) which is based on better scattering experiments and quantum chemistry calculations, give a binding energy of -7.27 K/atom (Boronat and Casulleras, 1994). This 1% overbinding (i.e, 0.15 K out of 15 K) is probably due to the neglect of a weak repulsion coming from the overlaps of triplets of atoms. At high densities, many-body forces give rise to a significant attraction (Boninsegni, Pierleoni, and Ceperley, 1994). However, these many-body forces are almost entirely a density-dependent background energy; they shift the pressure and energy but do not affect any microscopic properties in a homogeneous system substantially.

In this article we shall take the point of view that, even though the helium potential is not exactly known, an empirical pair potential is close enough to the true potential that the physical properties we consider should be predicted correctly by the model. The helium-helium potential is better known than almost any other atomic or molecular potential. The numerical work that we shall report here has used various semiempirical potentials (Aziz *et al.*, 1979, 1992). The differences between these potentials are not significant for our purposes. In fact, many of the properties of helium can be understood by treating the potential as a hard-sphere interaction with a radius

of 2.14 Å (Kalos, Levesque, and Verlet, 1974). We shall occasionally use the term "hard-sphere" interaction, but we do not mean to imply that the calculations have used this approximation.

Helium, under its own vapor pressure, is a liquid down to the absolute zero of temperature because the potential is so weak that the zero-point motion of the atoms disrupts the formation of a crystal; see Fig. 2. It does form a solid if it is pressurized to 25.3 bars (^{4}He) and 34.4 bars (^{3}He). The fact that it is liquid down to zero temperature makes it a nearly perfect system for studying macroscopic quantum effects. Since the atoms are delocalized, the indistinguishability of the particles becomes very important. The most dramatic effect is superfluidity, which occurs when ^{4}He is cooled below 2.17 K. If a cylinder containing helium at sufficiently low temperature is rotated slowly, the helium inside will remain at rest, or in whatever state it was initially prepared. There are other related properties: superfluid helium can flow through capillaries without apparent friction, its thermal conductivity is effectively infinite, and so forth. The reader can consult Wilks (1967) for details of the experimental measurements and for the basic theoretical models describing superfluidity.

The properties of helium are among the best-measured in experimental physics. Liquid helium has a natural ability to clean itself of impurities, so that one does not have the material complications of superconductivity, for example. Since helium has two isotopes, ^{3}He and ^{4}He, one can experimentally distinguish the effects of correlation from those of statistics. One can see that the phase diagram of bulk ^{4}He has gas, normal-liquid, superfluid, and several crystal phases. Many additional phases occur with ^{4}He on substrates. The phase diagram of ^{3}He is even richer; there are several types of superfluids and magnetically ordered solid phases.

In this review, we shall concentrate on calculations of liquid ^{4}He. We want to understand what happens

to a box of "hard-sphere" bosons at a temperature low enough that quantum effects are important. Liquid ^{4}He is a prototypic Bose system, the only simple bosonic superfluid, and is one of the basic models of condensed-matter physics, only slightly less ubiquitous than a magnetic model like the Hubbard model or the Ising model. It is one of the simplest examples of a phase transition involving broken symmetry. Bose superfluids are a model related to the quantum Hall effect (Zhang, Hansson, and Kivelson, 1989) and to the melting of vortices in high-temperature superconductors (Nelson, 1988) and are closely related to models of magnets (Matsubara and Matsuda, 1956). They are also a model for preformed bosons in high-temperature superconductors (Alder and Peters, 1989; Mott, 1991). Understanding a Bose superfluid to the point where we can calculate detailed properties is probably a necessary first step to being able to do the same thing for fermion superfluids.

Feynman's path integrals, which map quantum mechanics onto a peculiar type of classical "polymers," provide a direct way of calculating its properties. These classical systems can be simulated with either path-integral Monte Carlo (PIMC) or path-integral molecular-dynamics (PIMD) techniques and can be understood with methods from classical statistical mechanics. This implies that the thermodynamic properties of a Bose superfluid can be calculated, without approximation, on a computer.

One might ask why simulation, as opposed to other theoretical techniques, is called for. To answer this question, consider the following syllogism. General classical systems can be calculated rigorously only by simulation techniques. (We do not mean to imply that high-quality classical simulations are easy to do, just that it is possible to get very good results given enough effort.) Much of the progress in classical statistical mechanics, such as our improved understanding of critical properties, has been crucially aided by numerical experiments. But quantum systems reduce to classical systems in the high-temperature limit. In addition, as we shall see, bosons at any temperature are isomorphic to a classical system of "polymers." Thus, accepting the premise, we conclude that exact properties of general Bose systems at nonzero temperature can only be calculated by simulation.

Although one may develop accurate approximate methods for classes of quantum systems, and nice theoretical models to understand the relationship between various properties, it seems almost certain that simulations are needed for high-accuracy calculations on general quantum systems and to provide the same sort of support for theory and experiment as simulation does for classical systems. One finds, in general, that the more complex a system is, the more one needs simulation as a tool. Although this article will discuss mainly bulk helium, frankly simulations for bulk systems are not crucial, since they are well characterized experimentally. But for more complicated systems, for example, multicomponent or inhomogeneous systems, quantum simulation methods
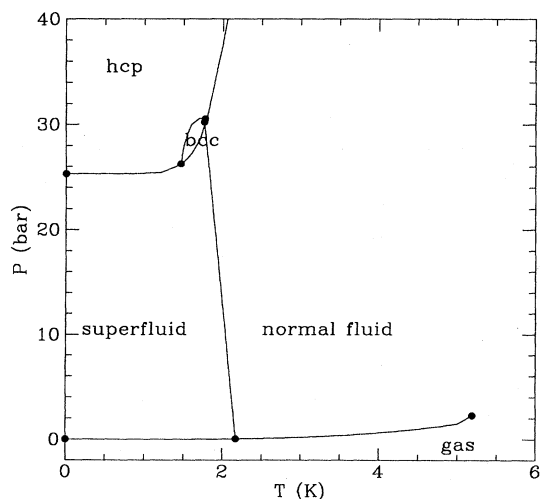

FIG. 2. The phase diagram of ^{4}He.

can be very useful. While other theoretical treatments become much more complicated and approximate, and those systems are harder to characterize experimentally, PIMC simulation is essentially no more difficult or approximate for complex systems.

It has taken forty years for simulation methods and computational resources to be developed to the point where calculations of Bose superfluids are routine. Commonly available work stations are sufficient to do most of the calculations we shall discuss here. Among the results that have been calculated are the energy, the specific heat, the radial distribution function, the momentum distribution, the condensate fraction, and the superfluid density of bulk liquid ^{4}He through the superfluid transition in both two and three dimensions; properties of solid ^{4}He; atomic exchange frequencies in solid ^{3}He and on graphite substrates; superfluid densities and energies of ^{4}He droplets; and energies and superfluid densities in two-dimensional charged Bose liquids. We are not going to review all of the results of those calculations here. Other reviews are available (Schmidt and Ceperley, 1992).

There are many other theoretical approaches to understanding superfluidity. The best known are field theory and the two-fluid model. We shall also not review the enormous literature on the theory of, and experiments on, liquid helium, except where necessary to relate this literature to the path-integral description. It would be desirable if the PIMC method could determine when these theories are applicable and estimate their parameters, but to date there has been little work on that subject. Since PIMC is an exact method, its predictions do not necessarily fall out of simple equations. Therefore it is complementary to other theoretical approaches, which will lead to more simplified models. PIMC is unique in being able to make quantitative predictions about quantities such as the superfluid transition temperature, the condensate fraction, the superfluid density, the freezing density, and detailed properties of helium films on substrates. All of this can be done with only the assumption of the interatomic potential.

On the other hand, path integrals do not naturally describe the quasiparticle picture of liquid helium, as a gas of interacting phonons and rotons. Feynman switched away from the path-integral theory of helium in favor of the excitation picture: phonons, rotons, and vortices. How to go reliably, efficiently, and simply from the imaginary time of path integrals into the real time necessary to describe excitations is an area of active research.

There are other computational methods by which to calculate the properties of liquid ^{4}He. McMillan (1965) introduced the variational Monte Carlo method, calculating energies and condensate fractions using a pair product or Jastrow trial wave function. Later Kalos, Levesque, and Verlet (1974) used the Green's-function Monte Carlo (GFMC) method to calculate exact energies for a system of hard-sphere bosons. This was extended to continuous potentials by Kalos *et al.* (1981). The

Green's-function Monte Carlo method is closely related to PIMC simulation but is restricted to zero temperature. In practice, path integrals give more insight and physical intuition, while being about equally efficient numerically, even for calculating ground-state properties. As an example, it is not easy to calculate the superfluid density in the Green's-function Monte Carlo method, since particle statistics do not enter directly, but it is quite straightforward in the PIMC method. We shall return to a comparison between these methods at the end of the paper.

There has been an enormous amount of work using PIMC for lattice models. When applied to lattice models, PIMC is called the "world-line" method. Applications have been in the lattice gauge theory of particle physics or for models of high-temperature superconductors. This review will discuss only continuum models, since the techniques for lattice models, although mathematically related, appear rather different. Most of the methods discussed here are applicable to bosonic lattice models with straightforward modifications (Krauth, Trivedi, and Ceperley, 1991). In addition, there are very efficient "cluster" algorithms (Kawashima and Gubernatis, 1995) for lattice models that we shall not discuss here since they have not yet been applied to continuum models.

There have also been many PIMC calculations for continuum systems in which particle statistics are not important. Examples are electron transfer in liquids and quantum corrections to classical liquids (Berne and Thirumalai, 1986). We shall not specifically talk about these applications but concentrate on the cases in which many-body exchange can be important. However, the discussion of PIMC techniques applies to simulation of single quantum particles.

A brief account of this work has appeared in Ceperley and Pollock (1990). The techniques for classical liquids are discussed in the book by Allen and Tildesly (1987). Two general reviews of Monte Carlo methods in statistical physics have been provided by Binder (1979, 1992). A general description of path integrals may be found in Feynman and Hibbs (1965), Schulman (1981), and Kleinert (1990). Applications to polymers and to quantum systems are discussed by Wiegel (1986).

This review explores in detail the path-integral picture, how it relates to experiments and other ways of thinking about quantum systems, and how one can use it in practice to calculate properties of quantum systems. Most of the paper will be at a very elementary level. Only the basics of quantum mechanics, probability theory, and statistical mechanics will be assumed. We shall try to provide clear, practical guidelines for developing, judging, and testing new algorithms for path integrals.

The first part of this paper discusses the mathematical basis of path integrals and the relationship between path integrals and physical properties. We then make explicit comparisons between experimental measurements of liquid helium and path-integral calculations, particu-

larly those that characterize the microscopic structure. Hopefully we shall convince the reader that path integrals truly are a reliable way of calculating properties of superfluids, so that the methods can be applied to more complex systems with confidence. We shall indicate which kinds of physical properties have been calculated or could be calculated with the path-integral Monte Carlo method and the accuracies that can be obtained. We also show how path integrals can calculate the exchange frequencies in solid helium and how these results compare with experimental measurements on solid ^{3}He.

The rest of this review is concerned with the details of the path-integral Monte Carlo approach. The computational task of simulating a quantum system can be broken into three parts. First, it is necessary to optimize the high-temperature action to reduce the number of time steps needed to reach superfluid temperatures, as discussed in Sec. IV. Second, it is necessary to move quickly through "path space." The Metropolis Monte Carlo method is used for this, but specialized techniques are needed to move the paths, since atomic paths must be exchanged many times to achieve reliable results when quantum statistics are important. A straightforward approach can run into difficulty, since the time scales for relaxation of the entangled "polymers" can become exceedingly long. This happens when the thermal de Broglie wavelength is comparable to the interparticle spacing, which is precisely when quantum many-body effects are important. Sampling methods are discussed in Sec. V. We discuss these aspects in some detail, since they are expected to have applicability far beyond the physics of helium and many of the issues have not been previously reviewed. Third, it is not always obvious how best to calculate a given quantum property from a well-converged path. We discuss the calculation of properties such as the energy and momentum distribution in Sec. VI. Finally we compare PIMC with other methods of simulating quantum systems and discuss the computational complexity of quantum simulations.

II. IMAGINARY-TIME PATH INTEGRALS

All static properties and, in principle, dynamical properties 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 (1972). 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.

A. The thermal density matrix

Suppose the exact eigenvalues and eigenfunctions of a Hamiltonian $\mathcal{H}$ are ϕ_i and E_i. 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

$$\langle O \rangle = Z^{-1} \sum_i \langle \phi_i | O | \phi_i \rangle e^{-\beta E_i} \tag{2.1}$$

where the partition function is

$$Z = \sum_i e^{-\beta E_i} \tag{2.2}$$

and $\beta = 1/k_B T$. In operator notation we write these equations more simply as $Z = \text{tr}(e^{-\beta \mathcal{H}})$ and $\langle O \rangle = \text{tr}(O e^{-\beta \mathcal{H}})/Z$. The operator $e^{-\beta \mathcal{H}}$ is the density matrix. In this article β will always be real, by *density matrix* we shall always mean $e^{-\beta \mathcal{H}}$, and the density matrix will *not* be normalized by the partition function. For simplicity, our statistical-mechanical ensemble always has fixed particle number N, temperature T, and volume Ω.

Although the above traces can be carried out in any complete basis, we shall work exclusively in a *position basis* where the *particles are labeled*. We work in a position basis because then all of the elements of the density matrix are non-negative and can be interpreted as a probability. Monte Carlo methods are usually much less efficient in a basis where some of the matrix elements are negative. It will become apparent at the end of this section why the particles are labeled. The position-space density matrix is

$$\rho(R, R'; \beta) = \langle R | e^{-\beta \mathcal{H}} | R' \rangle$$
$$= \sum_i \phi_i^*(R) \phi_i(R') e^{-\beta E_i}, \tag{2.3}$$

where $R = \{\mathbf{r}_1, \ldots, \mathbf{r}_N\}$ and $\mathbf{r}_i$ is the position of the ith particle. Assuming space has dimension $d = 3$ (just to simplify notation, everything that we discuss can be done in any dimension), then $\rho(R, R'; \beta)$ is, in general, a function of $6N + 1$ variables. In the position representation, the expectation of O becomes

$$\langle O \rangle = Z^{-1} \int dR dR' \rho(R, R'; \beta) \langle R | O | R' \rangle \tag{2.4}$$

and the partition function is given by

$$Z = \int dR \rho(R, R; \beta). \tag{2.5}$$

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:

$$e^{-(\beta_1 + \beta_2)\mathcal{H}} = e^{-\beta_1 \mathcal{H}} e^{-\beta_2 \mathcal{H}}. \tag{2.6}$$

Written for positions, one has a convolution,

$$\rho(R_1, R_3; \beta_1 + \beta_2) = \int dR_2 \rho(R_1, R_2; \beta_1) \rho(R_2, R_3; \beta_2). \tag{2.7}$$

B. Discrete path integrals

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,

$$e^{-\beta\mathcal{H}} = \left(e^{-\tau\mathcal{H}}\right)^M, \tag{2.8}$$

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

$$\rho(R_0, R_M; \beta) = \int \cdots \int dR_1 dR_2 \cdots dR_{M-1} \rho(R_0, R_1; \tau) \rho(R_1, R_2; \tau) \cdots \rho(R_{M-1}, R_M; \tau). \tag{2.9}$$

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 $0 \le t \le \beta$. But note that Eq. (2.9) is *exact for any* $M \ge 1$.

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. Recall the exact operator identity,

$$e^{-\tau(\mathcal{T}+\mathcal{V}) + \frac{\tau^2}{2}[\mathcal{T},\mathcal{V}]} = e^{-\tau\mathcal{T}}e^{-\tau\mathcal{V}}. \tag{2.10}$$

As $\tau \to 0$ the commutator term on the left-hand side, which is of order τ^2, becomes smaller than the other terms and thus can be neglected. This is known as the *primitive* approximation:

$$e^{-\tau(\mathcal{T}+\mathcal{V})} \approx e^{-\tau\mathcal{T}}e^{-\tau\mathcal{V}}. \tag{2.11}$$

Hence 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 Trotter (1959) formula, one does not have to worry:

$$e^{-\beta(\mathcal{T}+\mathcal{V})} = \lim_{M \to \infty} \left[e^{-\tau\mathcal{T}}e^{-\tau\mathcal{V}}\right]^M. \tag{2.12}$$

The Trotter formula holds if the three operators $\mathcal{T}$, $\mathcal{V}$, and $\mathcal{T} + \mathcal{V}$ are self-adjoint and make sense separately, for example, if their spectrum is bounded below (Simon, 1979). This is the case for the Hamiltonian describing helium.

Let us now write the primitive approximation in position space,

$$\rho(R_0, R_2; \tau) \approx \int dR_1 \left\langle R_0 | e^{-\tau\mathcal{T}} | R_1 \right\rangle \left\langle R_1 | e^{-\tau\mathcal{V}} | R_2 \right\rangle, \tag{2.13}$$

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

$$\left\langle R_1 | e^{-\tau\mathcal{V}} | R_2 \right\rangle = e^{-\tau V(R_1)} \delta(R_2 - R_1). \tag{2.14}$$

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 $L^{-3N/2}e^{iK_n R}$ and λK_n^2, with $K_\mathbf{n} = 2\pi\mathbf{n}/L$ and $\mathbf{n}$ a $3N$-dimensional integer vector. Then

$$\left\langle R_0 | e^{-\tau\mathcal{T}} | R_1 \right\rangle = \sum_\mathbf{n} L^{-3N} e^{-\tau\lambda K_n^2 - iK_n(R_0 - R_1)} \tag{2.15}$$

$$= (4\pi\lambda\tau)^{-3N/2} \exp\left[-\frac{(R_0 - R_1)^2}{4\lambda\tau}\right]. \tag{2.16}$$

Equation (2.16) 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,

$$\lambda\tau \ll L^2. \tag{2.17}$$

In some special situations this condition could be violated, in which case one should use Eq. (2.15) or add periodic "images" to Eq. (2.16). The exact kinetic density matrix in periodic boundary conditions is a theta function, $\theta_3(z, q)$, where $z = \pi\delta x/L$ and $q = e^{-\lambda\tau(2\pi/L)^2}$. See Abramowitz and Stegan (1964), Chapter 16, for its properties. Just to simplify the equations we shall always assume Eq. (2.17) holds.

Using Eqs. (2.9), (2.13), (2.14), and (2.16) we arrive at the discrete path-integral expression for the density matrix in the primitive approximation:

$$\rho(R_0, R_M; \beta) = \int dR_1 \ldots dR_{M-1}(4\pi\lambda\tau)^{-3NM/2}$$
$$\times \exp\left(-\sum_{m=1}^{M}\left[\frac{(R_{m-1} - R_m)^2}{4\lambda\tau} + \tau V(R_m)\right]\right). \tag{2.18}$$

This expression relates the quantum density matrix at

any temperature to integrals over the path $R_1 \cdots R_{M-1}$ 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.

Let us recap the various restrictions and approximations that have been made along the way in deriving the final result:

(i) The Hamiltonian is the sum of a nonrelativistic kinetic energy (without a magnetic field) and a real potential energy that depends only on position.

(ii) We can neglect the commutator between the kinetic and potential operators. Trotter showed that this is mathematically rigorous in the limit of large M. In Sec. IV we shall explore approximations that converge faster to the limit.

(iii) We neglected the periodic boundary conditions in evaluating the kinetic operator. This was only to keep the formulas simple. To get rid of this approximation one can either use the exact periodic density matrix or estimate errors with an image expansion. Errors from ignoring the boundary conditions are $\mathcal{O}(q)$, exponentially small at large M. Other boundary conditions, for example, hard walls, are easy to use.

(iv) We assumed distinguishable particles. We shall get rid of this assumption in the next subsection by symmetrizing over particle labels.

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. (2.18)], 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. We shall discuss this in more detail in Sec. VIII.

C. Our path-integral notation

Before we further develop the ideas of path integrals, let us specify the notation that we shall use in the rest of the article.

The *time step* is defined as

$$\tau \equiv \beta/M, \tag{2.19}$$

and a single R_k is referred to as the kth *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 ith particle in the kth 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 $t_k = k\tau$. We occasionally use a *space-time* notation where we define $\mathcal{R}_k = (R_k, t_k)$. Then we can drop the time argument of the density matrix, $\rho(\mathcal{R}_1, \mathcal{R}_2) \equiv \rho(R_1, R_2; |t_1 - t_2|)$.

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:

$$S^m \equiv S(R_{m-1}, R_m; \tau) \equiv -\ln[\rho(R_{m-1}, R_m; \tau)]. \tag{2.20}$$

Then the (exact) path-integral expression becomes

$$\rho(R_0, R_M; \beta) = \int dR_1 \cdots dR_{M-1} \exp\left[-\sum_{m=1}^{M} S^m\right]. \tag{2.21}$$

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,

$$K^m = \frac{3N}{2} \ln(4\pi\lambda\tau) + \frac{(R_{m-1} - R_m)^2}{4\lambda\tau}. \tag{2.22}$$

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

$$U^m = U(R_{m-1}, R_m; \tau) = S^m - K^m. \tag{2.23}$$

We shall frequently refer to U as the action, but of course the complete action also includes the kinetic action.

The approximation [Eq. (2.13)] of allowing the kinetic and potential energies to commute will be called the *primitive approximation*. In the primitive approximation, the inter-action is

$$U_1^m = \frac{\tau}{2}[V(R_{m-1}) + V(R_m)]. \tag{2.24}$$

We have symmetrized U_1^m with respect to R_m and R_{m-1}, since one knows the exact density matrix is symmetric and thus the symmetrized form is more accurate. If a subscript is present on the inter-action, it indicates the order of approximation; the primitive approximation is only correct to order τ. No subscript implies the exact action.

A capital letter U refers to the total link inter-action, while a small letter $u(\mathbf{r}_i, \mathbf{r}_j; \tau)$ refers to a contribution to the action from a single atom or pair of particles (i, j). One should not think of the exact U as being strictly the potential action. That is true for the primitive action but, in general, is only correct in the small-τ limit. The exact U also contains kinetic contributions of higher order in τ.

D. The classical isomorphism

We can interpret the path-integral expression, Eq. (2.18), as a classical configuration integral; the action is analogous to a classical potential-energy function divided by $k_B T$. 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 interpolymeric 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. (2.18) 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*.

We cannot overemphasize the importance of the quantum-classical isomorphism (Chandler and Wolynes, 1981). 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. Anything about helium that can be written in terms of the partition function, or more generally as matrix elements of the density matrix, has a classical statistical-mechanical analog.

There is a curious shift of vocabulary in going from the quantum system to the polymer model. Scientists discussing liquid helium with the aid of path integrals sometimes resemble children playing the game of "opposites," where the child says the opposite of what is intended. (I do not want a cookie.) Usually, the children's game degenerates quickly into confusion because common language is ambiguous and not entirely logical. Discussions of path integrals should be clearer, since path integrals are based on mathematics, but the "translation" is complicated by several features.

The same word applied to the quantum system and the classical system can mean quite different things. In cases where confusion of terms is possible, we shall put the term referring to polymers in quotes. 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

$$\langle \mathcal{T} \rangle = \frac{3N}{2\tau} - \frac{m}{2} \left\langle \left(\frac{R_i - R_{i-1}}{\hbar\tau} \right)^2 \right\rangle. \qquad (2.25)$$

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. There is a large cancellation between these two terms as τ gets small, since they both diverge as $1/\tau$. We shall find other ways to estimate the kinetic energy in Sec. VI.

Any observable corresponding to a scalar function of coordinates maps trivially from the quantum system into the polymer model. For example, the particle density is simply the average density of the beads,

$$\langle \rho(\mathbf{r}) \rangle = \left\langle \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_{im}) \right\rangle. \qquad (2.26)$$

Here $\langle \cdots \rangle$ denotes the configurational average over the polymer configurations. Since the ring polymers are periodic in imaginary time, we can average over time slices m as well as over paths.

There are often several different ways of mapping a quantum concept onto the classical system. A concept such as superfluidity is very general and related to many

quantum-mechanical observables. In a few words, superfluidity is equivalent to the presence of macroscopic polymers in the classical model. Further explanations and qualifications of this relation will be given in Sec. III.

A simple but very important quantity is the internal energy, the sum of the potential and kinetic energies. As discussed above, the potential energy is identical in the quantum and classical systems, since it is a function of coordinates alone. We have also given an expression for the kinetic energy. Hence we know how to compute the total energy. But by using Green's theorem we can write the internal energy in various other forms having different fluctuations and different systematic errors. We call these various forms *estimators* and discuss them in Sec. VI.

This aspect of figuring out different ways of calculating quantum properties in some ways resembles experimental physics. The theoretical concept may be perfectly well defined, but it is up to the ingenuity of the experimentalist to find the best way of doing the measurement. Even what is meant by "best" is subject to debate. Although mapping the quantum system onto a classical system is a big step forward, there may still be severe problems in calculating properties. One limitation is that simulation methods usually calculate only ratios of integrals, as in Eq. (2.1). Specialized techniques are required for such quantities as the free energy or entropy. This is a well-known and well-studied problem in classical statistical mechanics. Another problem is that the variance of an estimator may be too high. This is often the case when the integrand is both positive and negative. It is most severe for real-time or fermion path integrals but can occur in other contexts as well, for example, any excited state of a quantum system. Third, many important quantities of quantum systems are really defined as dynamical quantities, while the quantum-classical correspondence is restricted to imaginary time. Often, one can reformulate the quantum property in imaginary time, but not always. We shall give some examples in the next sections. There is still much to be done in learning how to exploit the quantum-classical correspondence.

E. Visualizing the paths

If path integrals are to be a useful, intuitive tool for understanding quantum systems, one needs to develop convenient ways of depicting paths. Here we give some pictures of two-dimensional path integrals (two dimensions because we do not want to consider the additional problem of projecting out the z dimension). What is usually done is to plot the *trace* of the paths in the X-Y plane, projecting out the "time" coordinate. What is shown in Fig. 3 are six distinguishable He atoms ($N=6$) in a periodic square at 2 K with 80 time slices. This is four times as many slices as are needed for a real simulation. The filled circles shown are markers for the beginning of imaginary time, i.e., R_1. There is nothing special about that value of imaginary time, but it is useful to
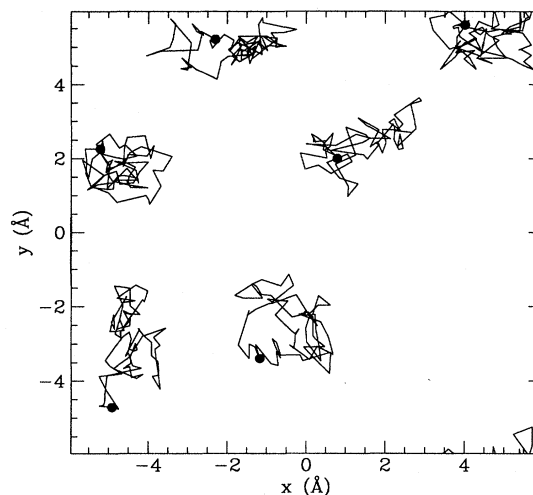
FIG. 3. The trace of the paths of six helium atoms at a temperature of 2 K with 80 steps on the paths. The filled circles are markers for the (arbitrary) beginning of the path. Paths that exit on one side of the square reenter on the other side. Successive beads are connected with straight lines.

place a mark at a common bead for each polymer, since the potential acts only at equal "times." It also helps the eye to pick out individual atoms when their paths overlap.

Shown in Fig. 4 are the same paths projected onto the (x,t) and (y,t) planes; the vertical axis is imaginary time. This is the *world-line* perspective. Clearly the trace perspective is more useful, since the spatial, two-dimensional relationship of the atoms is very important.

What is striking about the trace perspective is the
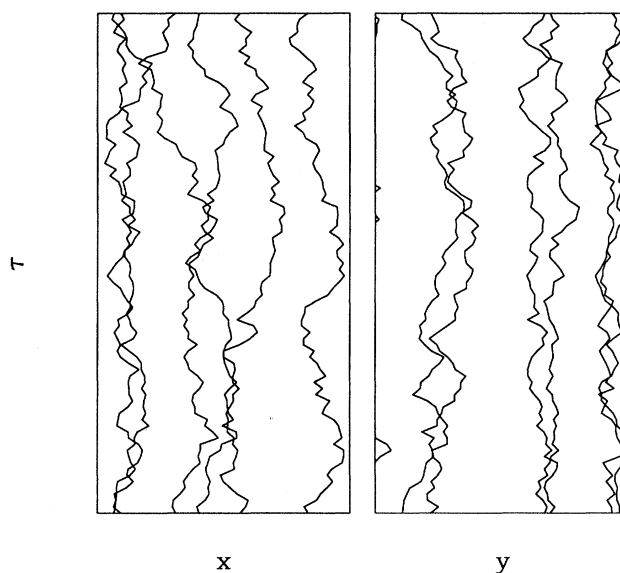
FIG. 4. A world-line diagram of the paths shown in Fig. 3. The vertical axis is imaginary time.

messy nature of the paths. They do not look like our idealized picture of ring polymers. In fact real polymers are quite different. On the length scale of individual atoms, chemical bonds keep the polymer rigid, but on a much larger length scale, real polymers are also messy looking. Because they cannot self-intersect, their paths are more spread than those of our "ring polymers."

Both real polymers and these paths are "fractals." Suppose we take any one of the line segments in the trace perspective and blow it up by making the time step many times smaller. To generate Fig. 5 we have used 8192 time slices. We can see that the line segment that was used in Fig. 3 to connect the beads is a poor approximation to this "fuzzy" object. As the number of time steps tends to infinity, the set of points visited by the paths will occupy a finite area; it has a Hausdorf dimension 2, while any smooth curve will have dimension 1. (A real polymer would have a fractal dimension between 1 and 2, but that does not concern us.) In the small-τ limit the kinetic action always dominates. This means that the fractal character seen in Fig. 5 is universal, independent of the potential. In the world-line picture, we see that the slope dr/dt is almost nowhere continuous. That curve would appear more and more ragged if the graph were expanded. For a detailed discussion of these issues see Mandelbrot (1977).

This fractal character of paths results in a very basic difference between viewing trajectories from a molecular-dynamics simulation and from path integrals. In a "movie" of a path (where viewing time is along the imaginary time direction) one's eye is distracted from seeing any pattern by the continual jerking of the particles, so "movies" are not a satisfactory way of getting insight into path integrals.

We are seriously misrepresenting the character of the paths by simply connecting the neighboring beads with straight lines. It is far better to shade (or even better color) the area in between the beads, to indicate which points could have been visited. This is closely related to a way of calculating a better approximation to the action, the cumulant action, which we shall discuss later. Shown in Fig. 6 are the same paths with all points that are within a distance α of the beads shaded. For a detailed discussion of how this figure was generated, see Moran

FIG. 6. The same paths as in Fig. 3. The shaded area are points within 0.38 Å of a visited point.

and Wagner (1994). On a raster device it is very easy to generate this type of plot; all of the pixels within a thermal wavelength Λ_τ of a bead are turned on.

In common with classical systems, periodic boundary conditions pose another problem for visualization. The paths really live on the surface of a 2D or 3D torus. By showing just the central unit cell, one introduces some unphysical features: a path can disappear from one side of the square and reappear on the other side. It is more satisfactory to repeat the unit cell, so that each polymer appears uncut at least once in the diagram. The same paths are plotted in this extended view in Fig. 7. The central cell is outlined with dashes. The extended cell view is especially important for superfluids, since it is the topological difference of paths that wind around the

FIG. 7. This is the same trace as in Fig. 3, but the paths have been replicated four times to eliminate the artificial boundaries.

FIG. 5. The trace of a free-particle path going from one circle to the other with 8192 steps.

boundaries that counts.

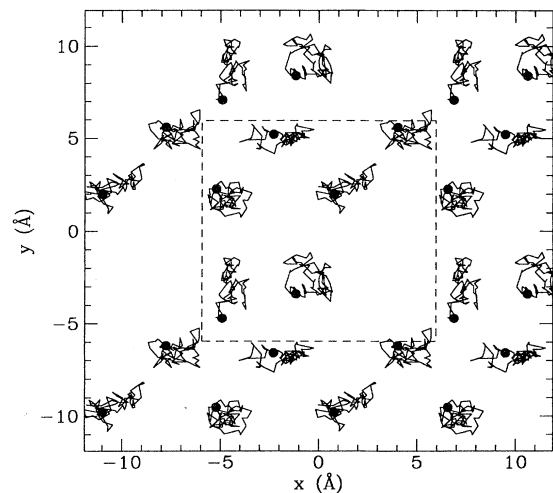We have already explained how "true" paths are messy objects. The fine-scale details are important for getting the calculation right, but not for viewing the paths. Every little detail is not important, whereas the overall shape and connectivity of a path is important. It is much more pleasing to the eye to smooth the paths. In Fig. 8 we have zeroed out all Fourier components above some threshold. In fact, Fourier smoothing represents a different way of doing discrete path integrals. Instead of working with a finite number of time steps, it is equivalent to keep a finite number of Fourier components. We shall discuss the Fourier path-integral method in Sec. V.E.

There is another way in which path integrals are unlike polymers: the paths have a direction. They correspond to directed polymers. One can reverse the direction of "time" for a whole system and one will get an equally probable path, since the action is invariant on changing t to $\beta - t$. But one cannot reverse the time on one atom's path and not on another atom's path. Neighboring paths may prefer to have their "velocities" parallel rather than antiparallel. Thus it may be important to place an arrow on each particle's path of a trace to indicate its "velocity."

One should always keep in mind that all these pictures show are points sampled from a product of thermal density matrices. They have only an indirect relationship to real-time dynamics or paths. They tell us mathematically precise information, but it must be correctly interpreted. Correlations along the paths are the Laplace transform of real-time linear-response functions. This will be considered in more detail in Sec. VI.F. Even though the imaginary-time dynamics is not directly real-time dynamics, it is very important. For example, it is impossible to tell from a single time slice whether or not a liquid is a superfluid, but one can recognize a superfluid by examining the connection of the paths in imaginary time.
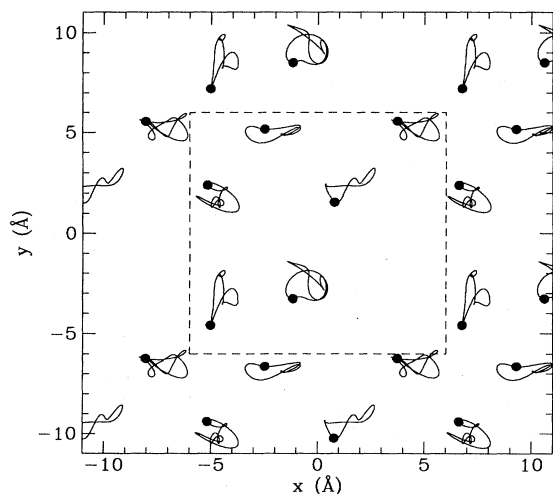

FIG. 8. The same trace as shown in Fig. 3, but the paths have been Fourier smoothed by taking only the lowest 11 Fourier components.

F. Bose symmetry

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

$$\mathcal{P}\phi(R) = \frac{1}{N!} \sum_P \phi(PR). \tag{2.27}$$

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

$$\rho_B(R_0, R_1; \beta) = \frac{1}{N!} \sum_P \rho(R_0, PR_1; \beta) \tag{2.28}$$

where ρ_B is the boson density matrix and ρ is the boltzmannon density matrix. Note that we can apply the permutation to the first argument of ρ, the last argument, or both. We will get the same result.

A straightforward evaluation of the permutation sum is out of the question once N gets large, since there will be $N!$ terms. The bosonic density matrix for free particles is an object known in mathematics as a permanent (a determinant with all the minus signs removed). A permanent takes on the order of $N \times 2^{N-1}$ operations to compute explicitly (Minc, 1978), as compared to a determinant, which takes only N^3 operations. 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.

The partition function for a Bose system has the form

$$Z_B = \frac{1}{N!} \sum_{\mathcal{P}} \int dR_0 \cdots dR_{M-1} \exp\left(-\sum_{m=1}^{M} S^m\right), \tag{2.29}$$

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.

Figure 9 shows the world-line view of a typical path of a 2D system of six ^{4}He atoms in the superfluid state. Three of the atoms are involved in a cyclic exchange, which wraps around the periodic boundary conditions. This path wrapping around the boundaries is called a *winding* path and is a direct manifestation of superfluidity. In the world-line view, superfluidity appears as a barber-pole or candy-cane design of the paths. The next figure, Fig. 10, shows the trace perspective (the projection onto the X-Y plane) of the same paths. Here the connection of the paths and of the winding is much more evident. The paths have a net "velocity" to the right.

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, 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.

It is easy to determine when quantum statistics will be important. In the absence of interaction, the size of a path (or polymer) is its thermal wavelength,
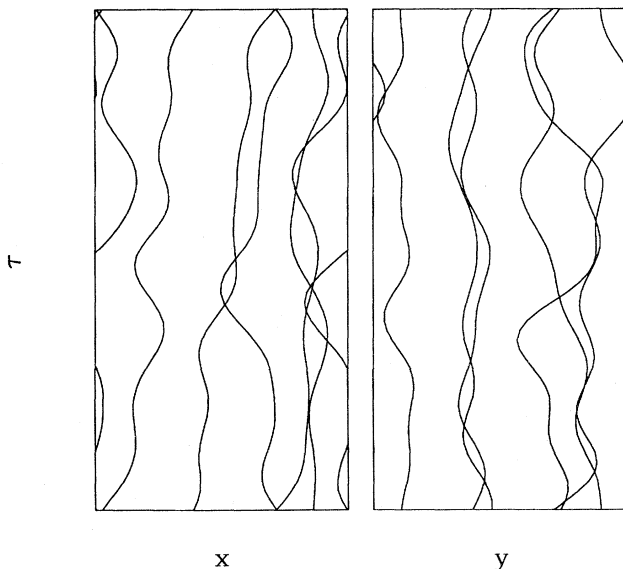

FIG. 9. The world-line view of same paths as in Fig. 10. Three of the atoms are involved in an exchange which winds around the boundary in the x direction.
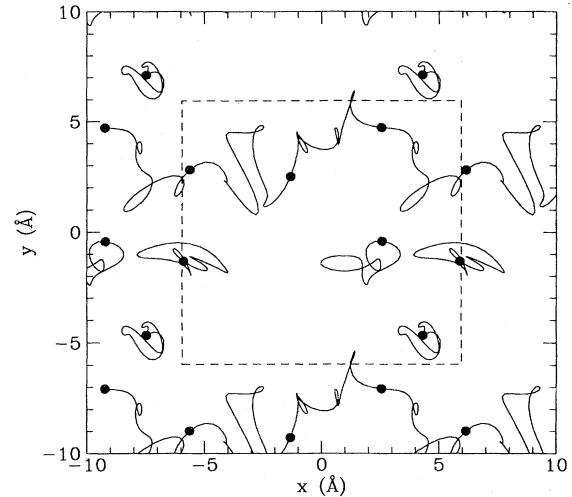
FIG. 10. The extended trace of six ^{4}He atoms at a temperature of 0.75 K and with 53 time slices. The dashed square represents the periodic boundary conditions. Three of the atoms are involved in an exchange which winds around the boundary in the x direction.

$$\Lambda_\beta = (2\beta\lambda)^{1/2}. \qquad (2.30)$$

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

$$T_D = \frac{\rho^{2/d}\hbar^2}{mk_B}. \qquad (2.31)$$

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), $T_c/T_D = 2.32$. Particle localization can cause T_c to become reduced. For example, as liquid ^{4}He is compressed, T_c decreases somewhat, while T_D increases, so that by the time liquid ^{4}He freezes $T_c/T_D = 1.64$. To explain this change, Feynman (1953) argued that the effective mass of the helium atoms is larger than the bare mass and increases with density. Later, we shall determine this effective mass with path integrals. Putting particles in a quantum solid suppresses exchange to such an extent that it is no longer a superfluid. In ^{4}He films, the transition is of the Kosterlitz-Thouless type, but nonetheless the estimate of the superfluid transition is reasonable: $T_c/T_D = 1.8$ at the minimum liquid density (Ceperley and Pollock, 1989). Note that the Fermi energy for an ideal spin-$\frac{1}{2}$ Fermi liquid is given in terms of the degeneracy temperature by $E_F/(k_BT_D) = 7.6$ in 3D and 6.3 in 2D.

In Eq. (2.28), it appeared that R_0 was singled out to

receive the permutation. Examination of Fig. 10 shows that the time slice where one relabels the particles is arbitrary. By change of variables one can make the permutation occur at any time slice. This is analogous to the international date line on the Earth: by convention, we change the calendar date in the middle of the Pacific Ocean, but the date line could have been placed anywhere. We could also insert a permutation on every link, but this would be a useless relabeling of the particles. As on the Earth, it is convenient to label particles in neighboring time slices the same, but somewhere along the path the labels must change if the permutation is nontrivial. All time slices are still equivalent, and they can all be used to calculate averages. Whenever one constructs an estimator for an observable, one should consider whether Bose symmetry has been taken into account properly.

Molecular hydrogen is another example of how permutational symmetry is treated with path integrals. In liquid or solid molecular hydrogen, two molecules exchange places very rarely, but the protons within a single molecule exchange very frequently. It is found that, in the absence of magnetic impurities, the total nuclear spin of the molecule is conserved for very long periods of time, so there are effectively two chemical species, para-hydrogen and ortho-hydrogen. The para-hydrogen nuclear spatial wave function must be symmetric under the exchange of the two protons, i.e., bosonlike. To calculate the partition function of para-hydrogen with path integrals, we must allow paths to close on themselves with a possible permutation, where atoms within the same molecule may exchange. There are 2^N such permutations where N is the number of hydrogen molecules. These pair permutations will not cause superfluidity, since they only modify the local wave function, but they can affect the thermodynamic properties. See, for example, Runge et al. (1992) or Marx et al. (1993). If we further assume that the molecule is localized at a crystal site and the molecular bond length is fixed, the only dynamical variable left is the bond angle. Then the path for each molecule is a path on the surface of a sphere. For para-hydrogen, points related to each other by inversion symmetry are considered equivalent. For ortho-hydrogen, the molecular wave functions are antisymmetric under particle exchange. Each pair exchange brings in an additional minus sign. Hence the integrand of the partition function is both positive and negative, so that it cannot map onto a classical distribution function.

For any discrete or continuous symmetry one can construct a projection operator for states restricted to a particular symmetry. Whether this is computationally useful depends on whether one has introduced too many minus signs in the process. Luckily for bosons there are no minus signs.

III. PATH INTEGRALS AND HELIUM

In this section we systematically go through the properties of liquid helium, explain how they are related to

path integrals, and compare the results with experimental data. In particular, we focus on the properties directly related to superfluidity—the specific heat, the momentum distribution, and the superfluid density. Finally, we discuss exchange in solid helium and attempt to calculate the excitation spectrum with path integrals.

A. The lambda phase transition

It was the shape of the specific-heat curve and its singularity at 2.17 K that gave rise to the name of the lambda transition. See Fig. 11. Feynman (1953) explained how macroscopic bosonic exchange gives rise to this peak. First he argued that the primitive approximation for the action would be qualitatively correct for temperatures near the critical temperature if we allowed the mass to be an effective mass even with one time slice. He arrived at the simple form for the partition function

$$Z_{MF} \approx K_\beta \int dR f(R) \sum_P \exp\left(-\frac{(R - PR)^2}{4\lambda^*\beta}\right), \quad (3.1)$$

where $f(R)$ is a normalized configurational distribution that at low temperature equals the ground-state density: $f(R) \approx \|\phi_0(R)\|^2$, K_β is a normalization factor, and λ^* accounts for the effective mass. This partition function was earlier proposed by Matsubara (1951) based on the symmetry of the boson ground state.

The Matsubara-Feynman approximate partition function captures the physics of the lambda transition. According to Feynman (1953), "It is not hard to understand that Eq. (3.1) gives a transition. If f were a constant it would be the same as a partition function for an ideal gas. The fact that f is not perfectly uniform cannot change this much."

One crucial difference is that the density of liquid helium is uniform, while ideal bosons tend to attract each other to maximize exchange and create pockets of high density. The interatomic potential between helium atoms will not allow regions of high density. To derive a simpler lattice model, Feynman neglected the temperature dependence of $f(R)$ and assumed that all configurations R are more or less equivalent, as long as the atoms are well separated. The neglect of the temperature dependence is justified empirically by noting that the pair-correlation function does not change much in the region of the lambda transition (Ceperley and Pollock, 1986). We can assume $f(R) = \frac{1}{N!} \sum_P \delta(R - PR_0)$ where R_0 is a typical configuration of atoms. For simplicity, Feynman (and later Kikuchi) took R_0 to be a perfect cubic lattice. The hard-core repulsion between helium atoms is then taken into account by allowing precisely one atom per lattice site. These approximations do not concern us now, since the PIMC computations will not use them. Taking a reference configuration, R_0, let us define the density of exchange distances,

$$\eta(x) = \frac{1}{N!} \sum_P \delta[x - (R_0 - PR_0)^2/(4\lambda^*)]. \quad (3.2)$$

Then the partition function reduces to

$$Z_{FK} = K_\beta \int dx \eta(x) \exp\left(-\frac{x}{\beta}\right). \qquad (3.3)$$

On a perfect cubic lattice with spacing d, x will be a multiple of $d^2/(4\lambda^*)$. Then $y = \exp[-d^2/(4\lambda^*\beta)]$ is the link probability. Feynman argued that in the critical region only permutations that send an atom back to its original site or to nearest-neighbor sites are important. If the permutation P is broken into cycles, a 4-cycle corresponds to a square on the lattice. A valid permutation consists of a nonintersecting collection of polygons on the lattice, nonintersecting because each atom can be a member of only a single cycle. Thus the partition function is obtained by finding the number of ways of drawing nonintersecting polygons on a lattice. In general, this combinatorical problem does not have a known solution. But on a two-dimensional honeycomb lattice Nienhuis (1984) has managed to evaluate the lattice sums exactly for the critical value of y, showing that the lambda transition is of the Kosterlitz-Thouless type.

Qualitatively, one can understand why there will be a phase transition when the temperature is low enough. From Feynman (1953a): "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."

In Feynman's original calculation, he got a third-order transition because of approximations made in calculating the lattice sum. Kikuchi, Denman, and Schreiber (1960) have refined these calculations using more elaborate methods. Chester (1955) showed, using cluster expansion methods directly on the partition function of Eq. (3.1), that introducing any correlation at all between the loops will change the transition from third order (ideal bosons) to second order.

Differentiating Eq. (3.3) with respect to β, we obtain for the energy

$$e = e_0 + \frac{3T}{2} - \frac{T^2\langle x\rangle}{N} \qquad (3.4)$$

and for the specific heat

$$C_v = \frac{3}{2} - 2\frac{T\langle x\rangle}{N} + \frac{T^2\langle(x - \langle x\rangle)^2\rangle}{N}. \qquad (3.5)$$

The specific heat is proportional to the mean-squared fluctuation of the exchange distance. At the critical point the specific heat diverges because there are both long and short exchanges present. Because it has to do with the stretching of paths, this is a purely kinetic contribution. We shall see in, Sec. VI that there is an exact equivalent to Eq. (3.4), one not based on an approximate partition function.

Elser (1984) has performed a Monte Carlo evaluation of the sum on a cubic lattice, obtaining the specific heat shown in Fig. 11. To get the transition temperatures to match, we adjusted the effective mass of the helium atoms to be 1.30 times the bare mass. While this is a somewhat small value for the effective mass, scaling the temperature also takes into account the approximation of a cubic lattice and the restriction to nearest-neighbor exchanges. [When Elser calculated the specific heat, he dropped the first two terms of Eq. (3.5) and plotted only the fluctuation term, which is the one that diverges at T_c.] There is a remarkable agreement between the lattice specific heat and the experimental measurement, thus verifying Feynman's conjecture that the Feynman-Kikuchi model is a correct description in the critical region.

Simulations of the lattice model of course run much faster than those of the detailed microscopic model. Below T_c one has to make modifications to the Matsubara-Feynman partition function to get the right specific heat, as explained in Feynman (1953b). This is because the number of permutations depends on the local density. But density fluctuations (phonons) have been left out of the lattice model.

Shown in Fig. 12 is the probability P_m that a given particle belongs to a cycle of length m (hence $\sum_m P_m = 1$) as a function of temperature. These results were obtained with PIMC simulation and a realistic interatomic potential. One sees that above T_c very few particles are involved in any exchanges, but at the transition there are many exchanges. The monomer density, P_1 is related to the energy of placing an impurity in ^{4}He, as we discuss

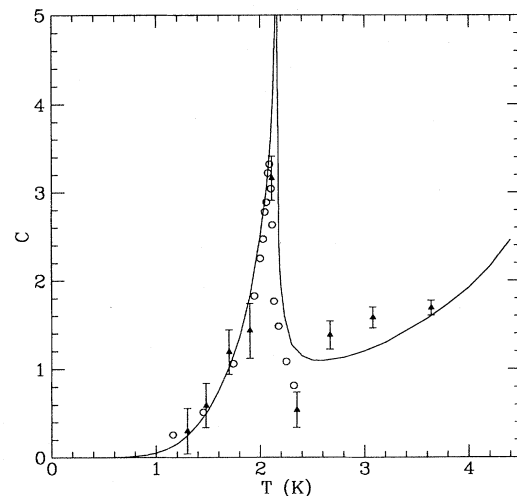FIG. 11. The specific heat of ^{4}He: solid line, experiment at saturated vapor pressure (Wilks, 1967); triangles with error bars, PIMC calculations (Ceperley and Pollock, 1986); open circles, Feynman-Kikuchi model with 20^3 sites (Elser, 1984). In Elser's calculation, only the fluctuation term in the specific heat has been included, and the temperature has been scaled to match the experimental transition temperature.
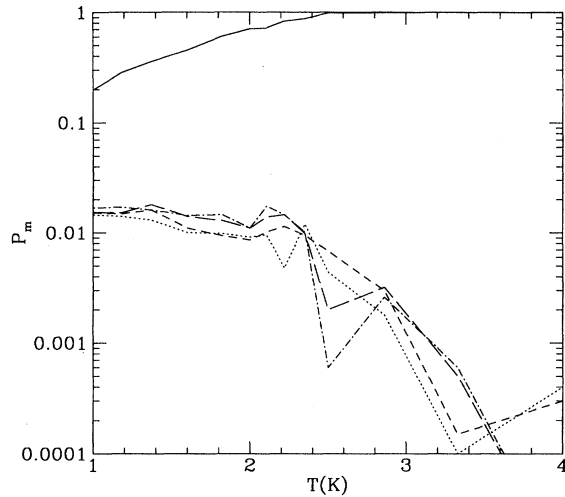
FIG. 12. The probability that an atom belongs to a permutation of length m as a function of temperature in liquid ^{4}He at saturated vapor pressure. The solid line represents the probability of a nonexchanging atom. The dashed lines are for $m = 2, 3, 4, 5$. All curves approach the zero-temperature value of $1/N = 1/64$.

FIG. 13. The energy/atom of ^{4}He: solid line, experiment at saturated vapor pressure; symbols with error bars, PIMC calculations (Ceperley and Pollock, 1986). The upper panel is the potential energy (solid line and left scale) and kinetic energy (dashed line and right scale).

below.

The phase transition for liquid helium belongs to the universality class characterized by a two-component order parameter, usually represented by a single-particle wave function $\phi(r)$ and a free-energy functional (Ginzburg and Pitaevshii, 1958; Pitaevskii, 1961) appropriate for long-wavelength excitations:

$$F = \int dr \left[\lambda |\nabla \phi(r)|^2 + \mu |\phi(r)|^2 + \frac{V}{2} \phi(r)^4 \right], \quad (3.6)$$

where μ is the chemical potential and V represents the short-ranged helium-helium potential. The chemical potential is negative and vanishes at the transition temperature, $\mu = (1 - T/T_c)\mu_o$. The free energy vanishes above T_c, where the minimum energy is attained when $\phi(r) = 0$. But below T_c the ground state has a broken symmetry because the phase of ϕ is undetermined: $\phi(r) = e^{i\gamma}\sqrt{-\mu/V}$.

Elser (1984) has shown in the dilute-gas limit how the ϕ^4 theory arises from bosonic exchanges in the path-integral description. A two-component field theory results from the fact that a permutation cycle has two possible senses (except for 1 and 2 cycles). Such a derivation has to make some assumptions about the existence of long exchanges, since application of pressure to ^{4}He produces a solid phase, which has completely different properties and a different effective action.

B. The energy of liquid ^{4}He

Switching now to computations that use the best interatomic and path-integral techniques, we see in Fig. 13

the energy of liquid ^{4}He as a function of temperature at saturated-vapor-pressure (SVP) conditions. One sees that there is a good agreement with experiment. But the effects of superfluidity are hardly evident in this curve; there is just a small point of infinite derivative at $T_c = 2.17$ K. Figure 11 shows the specific heat as obtained by taking finite differences of the path-integral energies. We shall discuss alternate ways of computing the specific heat in Sec. VI.B. Although the agreement with experiment is not bad, the statistical error of C_V is large. Lattice models such as the Feynman-Kikuchi model are much more convenient for determining critical properties such as critical exponents, since they are so much faster than continuum calculations. However, the continuum calculations are necessary for calculation of detailed microscopic properties and away from the critical point.

The effect of Bose condensation on the kinetic energy is more pronounced. At the transition, the kinetic energy drops by -1.2 K/atom, but the potential energy increases by only 0.25 K. Thus the atoms are able to delocalize with hardly any change in their spatial distribution. It is difficult with neutron scattering measurements to see the small change in kinetic energy. It is possible to measure the much larger dependence of kinetic energy on density, which is shown in Fig. 14, both for liquid and solid helium. The evident agreement means that the core of the He-He interaction has been chosen to have the correct size, since it is the core size which controls the kinetic energy.

Pollock and Runge (1992b) have directly calculated the transition temperature for liquid helium at saturated vapor pressure using PIMC and a realistic interatomic potential. To estimate the critical temperature they used ideas of finite size scaling, which have been extensively

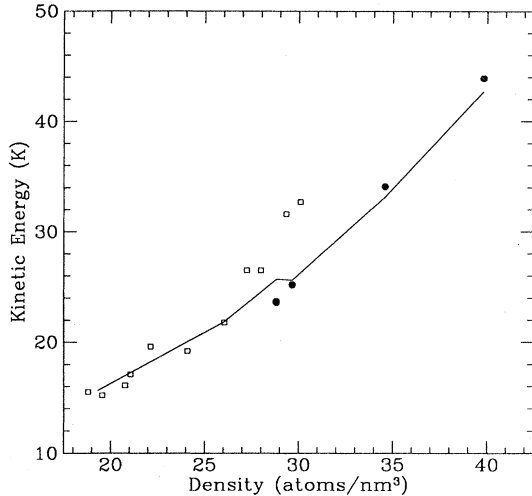FIG. 14. The kinetic energy in ^{4}He as estimated from PIMC at SVP as a function of density (solid line). Most of the calculations were done at 4 K. The jog in the curve near 30 nm^{-3} is the reduction in kinetic energy upon solidification. The points are estimates from deep-inelastic neutron scattering. Filled circles are the measurements in the solid phase (Blasdell, Ceperley, and Simmons, 1993) at various temperatures between 1 and 2 K. Open squares are measurements in the normal liquid at 4.25 K by Herwig *et al.* (1990).

FIG. 15. Free-energy difference between periodic and antiperiodic boundary conditions for the indicated temperatures (in K) as a function of the number of atoms. An estimate of T_c, based on extrapolating the slopes of the curves to zero, gives 2.19 ± 0.02 K (Pollock and Runge, 1992). The experimental transition temperature is 2.172 K.

applied to lattice models. The very simplest, but surprisingly accurate, way of determining the transition temperature is to calculate the free-energy difference from periodic to antiperiodic boundary conditions. If $\phi(r)$ is the wave function of the order parameter, then antiperiodic boundary conditions are defined by $\phi(\mathbf{r} + \mathbf{L}) = -\phi(\mathbf{r})$ with L the size of the simulation box. Using the energy functional of Eq. (3.6) we determine that the energy to make this twist equals $-\mu\lambda\pi^2 L^{d-2}/V$ in the superfluid phase. Thus for $d = 3$ the free-energy change will be proportional to L. In the normal phase, the free energy to make the twist will decrease exponentially with the correlation length, ξ, $\Delta F \sim e^{-L/\xi}$. As we shall discuss in the following sections, this free-energy change can be directly calculated by path integrals: It is

$$\Delta F = -k_B T \ln\langle e^{iW_x\pi/L}\rangle \qquad (3.7)$$

where W_x is the instantaneous winding number in the $\hat{x}$ direction [see Eq. (3.32)] and the brackets represent an average over the distribution of paths. The twist free energy is related to the difference in probability of an even winding and of an odd winding.

Figure 15 shows the result of this calculation for various sized systems and as a function of temperature. One can see very clearly that there is a change in behavior at the transition temperature. Pollock and Runge estimate the transition from this plot to be $T_c = 2.19 \pm 0.02$ K. The experimental value is 2.172 K. This remarkable agreement (along with others that we shall discuss) removes

any doubt of the underlying mechanism of the superfluid transition. Path-integral Monte Carlo simulation has succeeded in going directly from a parameter-free microscopic Hamiltonian to the experimental transition of a quantum many-body system.

C. The pair-correlation function and the structure function

The pair-correlation function $g(r)$ and its Fourier transform, the structure factor $S(k)$, describe the microscopic arrangement of atoms in liquid helium. There have been extensive and systematic measurements of these functions with neutron and x-ray scattering experiments. The calculated radial distribution function is shown in Fig. 16 and compared with x-ray and neutron scattering experiments. The largest disagreement, of about 2%, is at the nearest-neighbor peak. One expects errors in the peak height of order $N^{-1} \approx 1\%$ because of finite size effects. In fact, the disagreement between the two scattering measurements is larger than between theory and experiment. The comparison with the PIMC simulations gives us confidence in both the numerical methods and the assumed intermolecular potential.

The pair-correlation function shows very little effect of Bose condensation. Shown in Fig. 17 is the difference between the pair correlations of a boson system and a system with no bosonic exchange, i.e., "boltzmannon statistics" at the same temperature, 2 K. This is an example of a computer "experiment" that is easy to perform on the computer but impossible in nature. The maximum difference is only 2%. This temperature, just below
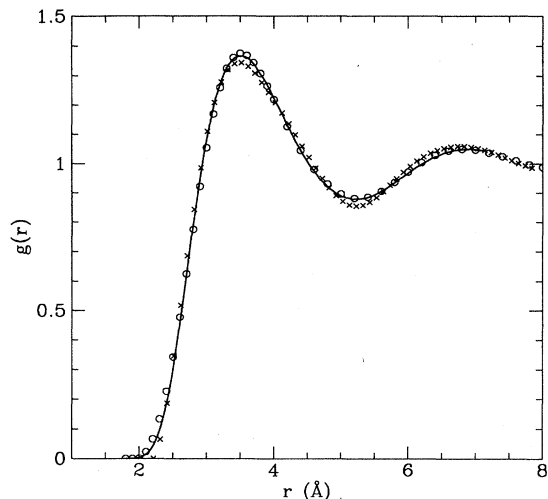
FIG. 16. The low-temperature pair-correlation function of liquid helium at the saturated vapor pressure density of 0.02182 atoms Å^{-3}: solid line, PIMC calculation at 1.21 K; $\bigcirc$, neutron-scattering measurement of Sears (1979) at 1.38 K; $\times$, x-ray scattering of Robkoff and Hallock (1981) at 1.38 K.

T_c, was chosen because one expects that the effects of Bose statistics on the structure will be largest there. At higher temperatures, bosonic exchange is unimportant; at lower temperatures the thermal wavelength of even a distinguishable atom becomes greater than the interparticle spacing. The insensitivity of $g(r)$ to exchange justifies the assumption that Feynman made, that the spatial distribution $f(R)$ is independent of temperature.
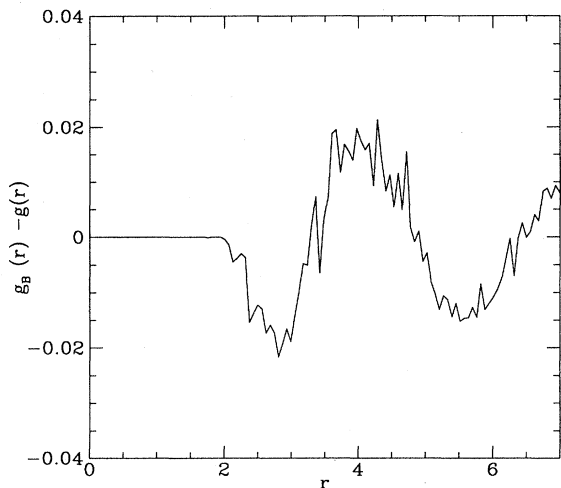
FIG. 17. The difference in the pair-correlation functions of "distinguishable" and boson-liquid helium at a temperature of 2.0 K and a saturated vapor pressure density of 0.02182 atoms/Å^{-3}. The peak height of $g(r)$ is reduced by about 2% because of bosonic exchange.
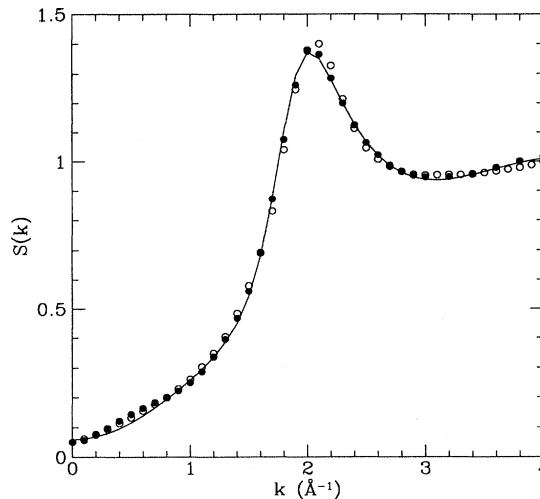
FIG. 18. The structure factor at 1.38 K and saturated vapor pressure: solid line, as calculated by PIMC; $\bullet$, measured by neutron scattering (Sears et al., 1979); $\bigcirc$, measured by x-ray scattering (Robkoff and Hallock, 1981).

The next three figures compare the PIMC calculated structure factors with neutron and x-ray scattering measurements. Figure 18 is in the superfluid phase (1.38 K). The agreement is superb. Figure 19 shows this comparison in the normal liquid, near the liquid-gas critical point. Except at long wavelengths the comparison is very good.
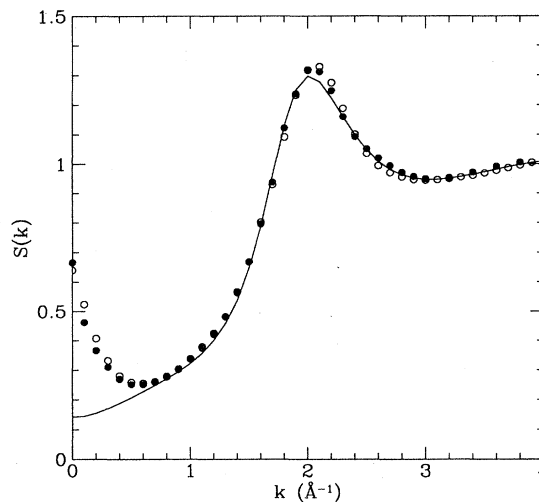
FIG. 19. The structure factor at 4.25 K and saturated vapor pressure; solid line, as calculated by PIMC at a density of 0.01873 Å^{-3}; $\bullet$, as measured by neutron scattering (Sears et al., 1979); $\bigcirc$, as measured by x-ray scattering (Robkoff and Hallock, 1981). The lowest wave vector actually calculated by PIMC is at 0.42 Å^{-1}. The PIMC curve shown is the Fourier transform of the pair-correlation function, which has been extrapolated to values for r greater than the size of the simulation cell. Hence the disagreement for $k \leq 0.5$ Å^{-1} arises because the small size of the simulation cell could not contain longer-wavelength density fluctuations.
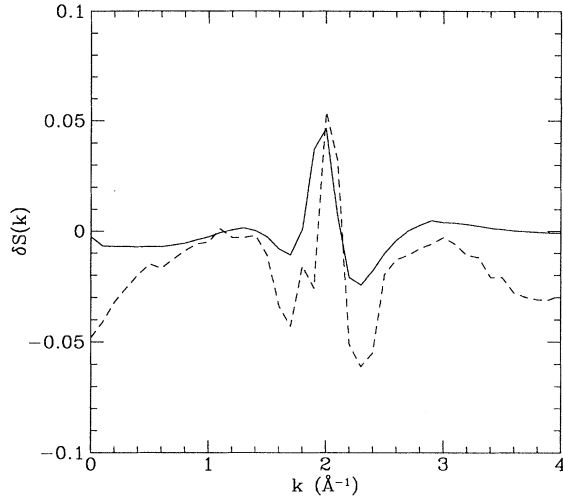
FIG. 20. The difference between the structure factor at 4 and 1.67 K at a density 0.0256 Å^{-3}: solid line, as computed with PIMC; dashed line, as measured by x-ray scattering (Robkoff and Hallock, 1982).

The smallest wave vector in the simulation is at 0.42 Å^{-1} so that the small cells with a fixed number of particles do not have the density fluctuations caused by the growing compressibility. Finally Fig. 20 shows the difference in the structure factor at high density in going through the superfluid transition, as calculated by PIMC and as measured by experiment. The calculated and measured changes are rather similar. This last plot shows that not only is PIMC getting the structure correct, it is also getting the change in structure due to Bose condensation correct.

D. The momentum distribution

London (1938) 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. How can they remain in a zero-momentum state, which means that they are uniformly occupying the entire box, while at the same time keeping out of each other's way?

Penrose and Onsager (1956) 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. Since that time the ground-state calculations have been improved considerably, for example, with the variational Monte Carlo calculations of McMillan (1965) and the

Green's-function Monte Carlo calculations of Whitlock and Panoff (1987), but the estimate of the ground-state condensate is still close to 8%.

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

$$n_{\mathbf{k_1}} = (2\pi)^{-Nd} \int d\mathbf{k}_2 \cdots d\mathbf{k}_N \left| \int dR \phi(R) e^{-iKR} \right|^2, \quad (3.8)$$

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

$$n_{\mathbf{k}} = \frac{1}{\Omega(2\pi)^d} \int d\mathbf{r}_1 d\mathbf{r}_1' e^{-i\mathbf{k}(\mathbf{r}_1 - \mathbf{r}_1')} n(\mathbf{r}_1, \mathbf{r}_1'), \quad (3.9)$$

where the single-particle density matrix is

$$n(\mathbf{r}_1, \mathbf{r}_1')$$

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

$$(3.10)$$

We have assumed periodic boundary conditions in a cell with volume Ω. Note that $n(\mathbf{r}, \mathbf{r}')$ and $n_{\mathbf{k}}$ are normalized as $\int d\mathbf{r} n(\mathbf{r}, \mathbf{r}) = \Omega$ and $\int d\mathbf{k} n_{\mathbf{k}} = 1$. For a homogeneous isotropic liquid we have $n(\mathbf{r}, \mathbf{r}') = n(|\mathbf{r} - \mathbf{r}'|)$, in which case $n(0) = 1$. The kinetic energy is proportional to the second moment of n_k and to the curvature of $n(\mathbf{r})$ at $\mathbf{r} = 0$.

According to Eq. (3.9), 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. All that is needed to get the momentum distribution is to remove the restriction on one of the atoms that it return to its starting position.

The method by which to calculate the single-particle density matrix is quite simple: one samples paths from the probability distribution,

$$\pi_n(R, \mathbf{r}_1') = \frac{1}{Z'} \rho(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N, \mathbf{r}_1', \mathbf{r}_2, ..., \mathbf{r}_N; \beta), \quad (3.11)$$

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

$$n(\mathbf{r}, \mathbf{r}') \propto \langle \delta(\mathbf{r}_1 - \mathbf{r}) \delta(\mathbf{r}_1' - \mathbf{r}') \rangle_{\pi_n}, \quad (3.12)$$

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

mers and 1 linear polymer. An example of such a path is shown in Fig. 21.

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:

$$\begin{aligned}
\mathbf{r} &= \mathbf{r}_1(0), \\
\mathbf{r}_1(\beta) &= \mathbf{r}_2(0), \\
\mathbf{r}_2(\beta) &= \mathbf{r}_3(0), \\
\mathbf{r}_3(\beta) &= \mathbf{r}'.
\end{aligned} \tag{3.13}$$

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; see Fig. 22. 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

$$\tilde{n}_0 = \frac{(2\pi)^3}{\Omega} n_0 = \frac{1}{\Omega^2} \int d\mathbf{r} d\mathbf{r}' n(\mathbf{r},\mathbf{r}')$$

$$= \frac{1}{\Omega} \int d\mathbf{r}\, n(r). \tag{3.14}$$

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,

$$\tilde{n}_0 = \lim_{r\to\infty} n(r). \tag{3.15}$$

The condensate fraction is essentially the probability of the two cut ends attaching themselves to a macroscopic exchange. Figure 23 shows the fraction of atoms in the zero-momentum state, and Fig. 24 the momentum distribution of the noncondensed atoms.

One often hears the question: why is the condensate fraction at low temperature only 10% while the system

FIG. 21. The extended trace of five ^{4}He atoms at a temperature 0.75 K. The dotted path is that of the cut polymer, the one that is not periodic in imaginary time. Its end-to-end distribution is used to calculate the momentum distribution. The other four atoms are involved in an exchange which winds around the boundary in the x and y directions. The dashed square represents the periodic boundary conditions. The paths have been Fourier smoothed.

FIG. 22. The single-particle density matrix of ^{4}He above and below the lambda transition at temperatures 1.18, 2.22, and 4 K (from top to bottom). The calculations (Ceperley and Pollock, 1987) were done at the density corresponding to saturated vapor pressure. Note that the dotted line, corresponding to a temperature of 2.22 K, which is above T_c, does not approach zero at the edge of the box because the finite-sized sample (64 atoms) has a higher transition temperature than bulk helium.
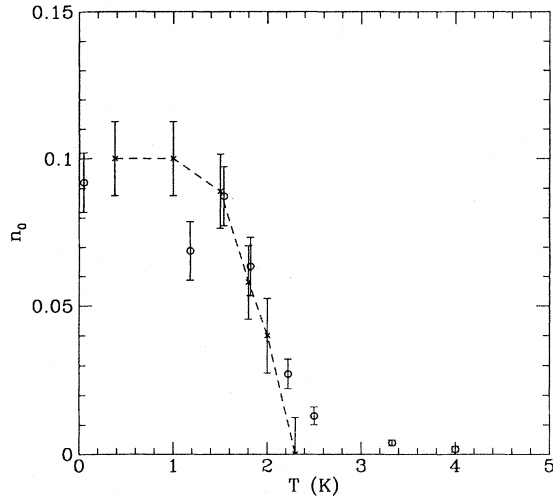
FIG. 23. The condensate fraction in ^{4}He as estimated from PIMC at SVP as a function of temperature. The estimates are obtained by averaging $n(r)$ in the range 2.5 Å$\le r \le$ 6.0 Å. Thus they are not zero in the normal fluid. The value at 0.1 K is from the zero-temperature Green's-function Monte Carlo calculations of Whitlock and Panoff (1987). The dashed line with crosses and error bars is from an analysis of neutron-scattering data by Snow, Wang, and Sokol (1992).

is 100% superfluid? We shall discuss the superfluidity in the next subsection, but it is easy to understand the fact that $\tilde{n}_0$ is much less than unity. The two ends of the "cut" polymer are attracted to each other because they can share the same correlation hole and thus minimize the action. In contrast to real polymers, there is no interaction between the two ends, since they represent two halves of the same particle that would push them

apart. "Entropy" counteracts this tendency to localize the two ends. If the attractive force between the ends drops sufficiently quickly, as happens in 3D, then $\tilde{n}_0$ will be finite. But the delocalization can only happen if there are macroscopic exchanges.

The explanation of the condensate fraction in terms of wave functions is quite different. Clearly the many-body wave function must have lots of curvature so that it can go to zero whenever any pair of atoms overlap. Hence the kinetic energy is nonzero and high momentum states must be occupied. In spite of this curvature, a "condensed" wave function manages to stay delocalized and uniformly fills up the low-potential-energy part of configuration space. One can pick an arbitrary atom, displace it a long distance and reinsert it, and with probability $\tilde{n}_0$, find that the wave function is still large.

In neutron-scattering experiments, the distribution of angles and energies of a neutron beam passing through a sample of helium is recorded. The measured cross section is proportional to the dynamic structure factor $S(\mathbf{k}, \omega)$. To relate the dynamic structure factor to the momentum distribution, one must take into account the quantum states the liquid can end up in after the scattering, "the final-state effect." At sufficiently high momentum transfer, the impulse approximation holds because the neutron scatters from a single atom. In this case one finds that

$$S(k, \omega) = \frac{m}{k} J(Y), \qquad (3.16)$$

where the scaling variable is defined by $Y = (m/k)(\omega - \lambda k^2)$ and

$$J(k_z) = \int dk_x dk_y n_\mathbf{k} = \frac{1}{\pi} \int dr n(r) \cos(k_z r). \qquad (3.17)$$

A comparison between the neutron-scattering measurements (Sokol, Soosnick, and Snow, 1989) and PIMC (Ceperley, 1989) for liquid helium is shown in Fig. 25. Note the good agreement with respect to experiment, at various temperatures. The experimental results are relatively insensitive to the condensate, thus the rather large errors on the condensate in Fig. 23.

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 (Nelson, 1983), $\eta^{-1} = 4\pi\lambda\beta\rho_s$, 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. PIMC calculations (Ceperley and Pollock, 1989) on 2D helium are in good agreement both with the Kosterlitz-Thouless theory and with experiments on helium films (Greywall and Busch, 1991).
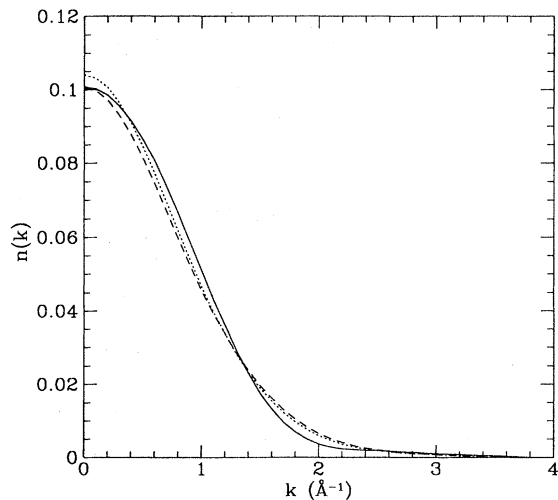

FIG. 24. The noncondensed momentum distribution of ^{4}He from PIMC calculations with the same conditions as in the previous figure.
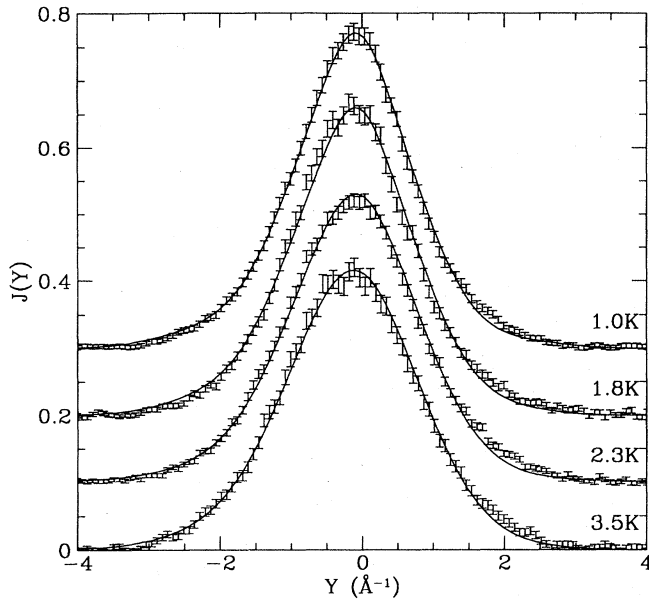
FIG. 25. Observed neutron scattering (Sokol, Sosnick, and Snow, 1989) at temperatures of 1.0, 1.8, 2.3, and 3.5 K. The solid lines are the PIMC predictions (Ceperley and Pollock, 1987) with instrumental resolution and finite-size effects included.

E. 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 (1941) 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. Experiments by Andronikashvili (1946) confirmed this prediction.

We do not assume that the bucket has cylindrical symmetry, so there will be some coupling between the walls of the bucket and the liquid helium, allowing the liquid to come to thermal equilibrium with the walls. The effective moment of inertia is defined as the work done for an infinitesimally small rotation rate,

$$I = \left.\frac{dF}{d\omega^2}\right|_{\omega=0} = \left.\frac{d\langle \mathcal{L}_z \rangle}{d\omega}\right|_{\omega=0}, \tag{3.18}$$

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

$$\mathcal{L}_z = i\hbar \sum_{i=1}^{N} \frac{\partial}{\partial \theta_i}, \tag{3.19}$$

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

$$I_c = \left\langle \sum_{i=1}^{N} m_i (\mathbf{r}_i^{\perp} \times \hat{z})^2 \right\rangle. \tag{3.20}$$

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

$$\frac{\rho_n}{\rho} = 1 - \frac{\rho_s}{\rho} = \frac{I}{I_c}. \tag{3.21}$$

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

$$\mathcal{H}_\omega = \mathcal{H}_0 - \omega \mathcal{L}_z. \tag{3.22}$$

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. (3.18). 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:

$$\frac{de^A}{d\omega} = \sum_{k=1}^{M} e^{(k-1)A/M} \frac{de^{A/M}}{d\omega} e^{(M-k)A/M}. \tag{3.23}$$

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

$$\frac{de^A}{d\omega} = \int_0^1 dt\, e^{tA} \frac{dA}{d\omega}\, e^{(1-t)A}. \tag{3.24}$$

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. (3.18). We get

$$\frac{\rho_n}{\rho} = \frac{I}{I_c Z} \text{tr} \left[\int_0^\beta dt \mathcal{L} e^{-(\beta-t)\mathcal{H}} \mathcal{L} e^{-t\mathcal{H}} \right]. \qquad (3.25)$$

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. One can show (Pollock and Ceperley, 1987) that an external potential also does not contribute in the limit as $\tau \to 0$. In evaluating the sum over k in Eq. (3.23) there is one tricky point. The $k = 1$ term must be treated separately, since $\mathcal{L}$ operates twice on one link. That term gives rise to the classical response. After some algebra we get

$$\frac{\rho_s}{\rho} = \frac{2m\langle A_z^2 \rangle}{\beta \lambda I_c}, \qquad (3.26)$$

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

$$\mathbf{A} = \frac{1}{2} \sum_{i,j} \mathbf{r}_{i,j} \times \mathbf{r}_{i,j+1} \qquad (3.27)$$

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

$$I_c = \left\langle \sum_{i,j} m_i \mathbf{r}_{ij}^\perp \cdot \mathbf{r}_{i,j+1}^\perp \right\rangle. \qquad (3.28)$$

Note that the area of a path is a vector. For rotations about the $\hat{z}$ axis we need only the $\hat{z}$ component of the area. By symmetry the average value of $\mathbf{A}$ vanishes. Equation (3.26) is the main result of this section and is an exact fluctuation-dissipation formula. 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, $\langle A^2 \rangle = N\langle a^2 \rangle$, where the mean-squared area of a free particle is $\langle a^2 \rangle = (\lambda\beta)^2/3$ (Pollock and Runge, 1992a). The classical moment of inertia will be $mN\langle r^2 \rangle$. Hence the superfluid density will be $\rho_s/\rho = 2\lambda\beta/(3\langle r^2 \rangle)$. 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. See Fig. 26 for the path-integral estimates of the superfluid density of ^{4}He droplets (Sindzingre et al., 1989). The superfluid density of a small droplet is not very different from that of the bulk liquid. 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
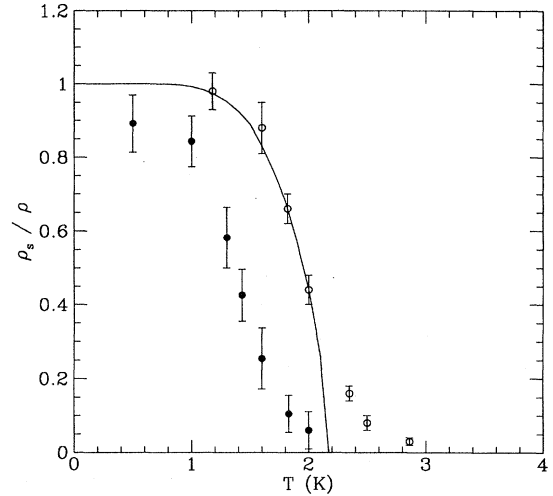
FIG. 26. The ratio of superfluid density to total density: solid line, measured value for bulk ^{4}He at SVP (Donnelly, 1967); $\bigcirc$, PIMC calculations with 64 atoms in periodic boundary conditions and using the winding number formula (Pollock and Ceperley, 1987); $\bullet$, calculations for a droplet of 64 ^{4}He atoms and using the area formula (Sindzingre, Klein, and Ceperley, 1989).

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.

A uniform magnetic field acting on charged particles is equivalent to a rotation, and we can directly apply all that we have learned about superfluidity to charged particles in magnetic fields. The analogous phenomenon to superfluidity is the Meisner effect of superconductors. The Hamiltonian in a magnetic field is

$$\mathcal{H} = \frac{1}{2m}[p - e/c\mathcal{A}]^2 + V(R). \qquad (3.29)$$

In a constant magnetic field in the Coulomb gauge, the lowest-order change in the Hamiltonian due to the magnetic field is proportional to the angular momentum operator, $\frac{e}{mc}\mathbf{p} \cdot \mathcal{A} = \frac{eB}{2mc}\mathcal{L}_z$, hence the previous discussion applies if we replace the rotation rate with $\frac{eB}{2mc}$. Then the zero-field diamagnetic susceptibility is

$$\chi = -\frac{\langle A^2 \rangle e^2}{\beta \hbar^2 c^2}, \qquad (3.30)$$

where A is the area of the path. For details see Parrinello and Rahman (1984) and Pollock and Runge (1992a).

There are serious differences which prevent one from applying path integrals to superconductors. First (and less serious), periodic boundary conditions complicate matters, even in a uniform field. Secondly, most charged particles are fermions, and so one cannot use these formulas when there are more than two fermions without taking antisymmetry into account. The minus signs, and/or

phases that are introduced, completely change the classical picture of quantum systems that we have been developing here. We will not discuss the application to superconductors any further.

F. Superfluidity in periodic boundary conditions

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. (3.26) for the superfluid density we get

$$\frac{\rho_s}{\rho} = \frac{\langle W^2 \rangle}{2\lambda\beta N}, \qquad (3.31)$$

where the winding number is defined by

$$\mathbf{W} = \sum_{i=1}^{N} \int_0^\beta dt \left[\frac{d\mathbf{r}_i(t)}{dt} \right]. \qquad (3.32)$$

Let us review the steps we took to get this new formula for the superfluid density: we first derived the formula for a general rotating system, then applied it to concentric cylinders. At the end, we remarked that a torus is topologically equivalent to the usual periodic boundary conditions. We could have made the whole derivation directly in the periodic space by calculating the response of a periodic system to a linear velocity of its walls. What appears in Eq. (3.25) in place of $\mathcal{L}$ is the total linear momentum operator $\mathcal{P}$ (Pollock and Ceperley, 1987).

Usually one applies periodic boundary conditions in all three spatial directions. Then the winding number becomes a vector, just as the area was a vector. But in contrast to the area, it is "quantized" in units of the box length. The 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. Figure 10 shows a path with winding $(L, 0)$. 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. Often one sees an extra factor of 1/3 in front of the winding number formula. This comes from symmetrizing over the three equivalent axes in the case of a cubic unit cell.

We can now come back to the relationship between superfluidity and momentum condensation. 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.

In the above derivation we considered the linear response of the system to a movement of the walls. One can also determine how the free energy depends on an arbitrary velocity $\mathbf{u}$ of the walls. One finds (Pollock and Ceperley, 1987)

$$e^{-\beta(F_\mathbf{u} - F_0)} = \left\langle e^{im\mathbf{u}\cdot\mathbf{W}/\hbar} \right\rangle, \qquad (3.33)$$

where the averages are over paths in the system with the walls at rest and $F_\mathbf{u}$ is the free energy in the system with moving walls (where the energy is measured in the moving coordinates). We see that the partition function is the Fourier transform of the winding number distribution. Since the winding numbers are quantized in units of L, the free energy is periodic: increasing the velocity by an amount $2\pi\hbar/ML$ leaves the free energy invariant. Moving the walls has the effect of changing the boundary conditions on the phase of the wave function. If a particle makes a circuit around the periodic box, the wave function picks up a phase, $2\pi muL/\hbar$. One can apply antiperiodic boundary conditions, instead of periodic boundary conditions, by applying a velocity to the walls of $\pi\hbar/ML$. This is known in field theory as "twisting the order parameter." A superfluid and a normal fluid respond quite differently to such a twist. A normal fluid can isolate the twist so the free energy is independent of the boundary conditions for a large enough system, while the superfluid's energy increases with the size of the system. We already used this idea to determine the superfluid transition temperature. See Fig. 15.

The winding number formula describes how the entire system will respond to motion of an external potential. We would also like to know how microscopic correlations differ between the superfluid phase and the normal phase of helium. For example, in the Ising model for a ferromagnet one can talk about the total spin susceptibility of the magnet or about the spin-spin correlation function. The quantity that is related to superfluidity is the correlation of the momentum density, the correlation of the velocity of the fluid at $\mathbf{r}$ with the velocity at $\mathbf{r}'$. One can define it either as an equilibrium velocity-velocity correlation function or as the linear response of the liquid to an imposed velocity perturbation. Let us define the local momentum as the expectation of the operator $\mathbf{p}(\mathbf{r}) = -i\hbar \sum_{j=1}^{N} \delta(\mathbf{r} - \mathbf{r}_j)\nabla_j$. Then the momentum density correlation function is

$$G(\mathbf{r} - \mathbf{r}') = \langle \mathbf{p}^\dagger(\mathbf{r}')\mathbf{p}(\mathbf{r}) \rangle \tag{3.34}$$

$$= G_N(r)I + G_s(r)[3\hat{\mathbf{r}}\hat{\mathbf{r}} - I] \tag{3.35}$$

$$= -\frac{2}{\Omega\lambda} \left\langle \sum_{i,j}^{N} \frac{d\mathbf{r}_i}{dt}\frac{d\mathbf{r}_j}{dt}\delta(\mathbf{r} - \mathbf{r}_i + \mathbf{r}_j) \right\rangle. \tag{3.36}$$

The first equation is the definition of the correlation tensor. In a homogeneous liquid a tensor can always be broken into a longitudinal and a transverse part. G_N has the short-range correlations of the particles as they try to avoid each other, while G_s has the long-range transverse superfluid velocity correlations (Baym, 1969),

$$G_s(r) \rightarrow -\frac{\rho_s m}{4\pi\beta r^3}. \tag{3.37}$$

The superfluid response has the following physical interpretation. Suppose we insert a moving particle at the origin. The superfluid will screen out this motion by constructing a long-range dipolar backflow field so that it isolates the rest of the system from the moving impurity.

Equation (3.36) relates the momentum correlation function to path integrals. If we put arrows on our paths indicating the direction of imaginary time, and we take these to be "velocities," the momentum correlation function is minus the correlation between atoms displaced by a distance $\mathbf{r}$. The minus sign comes in because the path integrals are in imaginary time. The macroscopic exchanges give rise to a long-ranged "velocity" field. The correlation in velocity is only in the direction parallel to the separation. G_s is simply the probability that two well-separated particles are members of the same macroscopic exchange.

The imaginary-time "velocity" is the order parameter of the superfluidity. The "velocity" can orient itself in any direction, but it is advantageous to break the symmetry and exchange in one particular direction. Here again we see how the quantum language and classical language are reversed. A superfluid isolates itself from the boundary, while in the corresponding classical system one can pull a macroscopic exchange and feel the effect all the way across the sample.

G. Impurities in superfluid helium

Superfluid ^{4}He is unique in its ability to purify itself and expel impurities. In fact, ^{3}He is the only impurity that will dissolve in ^{4}He at very low temperatures. Feynman discussed this problem in one of his original path-integral papers (Feynman, 1953c). The impurities allow us to define formally the exchange energy, how it relates to the exchange probability, the effective mass, and the chemical potential of an atomic impurity.

We will add the impurity in two steps. In the first step we determine the free energy of tagging one of the ^{4}He atoms in a superfluid, while in the second step we change the mass of the tagged atom from 4 to 3. The Hamiltonian of the tagged system is the same as that for

the untagged system, but one of the atoms is no longer symmetrized with respect to the other atoms. A little thought shows that we do not need to do a new calculation. Using Eq. (2.28) we see immediately that the free energy to tag an atom is

$$\mu = -k_B T \ln(P_1), \tag{3.38}$$

where P_1 is the monomer probability. Here we have dropped the ideal solution entropy: $-k_B T \ln(N)$.

Hence the distribution of cycle lengths is something that can, in principle, be measured. The tagging energy or exchange energy is shown in Fig. 27. It reaches a maximum of about $1.5K$ at a temperature of about 1 K and then decreases. If anything is done to a helium atom to inhibit its ability to exchange with its neighbors, one pays this energetic price. The tagging free energy is due to the extra thermal excitations around the tagged atom, which a homogeneous Bose superfluid does not have.

We can calculate the low-temperature limit of $\mu(T)$ in another way. By translation invariance, the tagged particle has a dispersion relation: $E_k = E_0 + \lambda^* k^2$ where λ^* is defined here. At a sufficiently low temperature, the excitations of a superfluid having an energy ck will be thermally damped and can be neglected. Integrating over the Boltzmann distribution of the tagged system, we find at low temperature

$$\lim_{T \to 0} \mu = 1.5 k_B T \ln(4\pi\beta\lambda^*\rho^{2/3}). \tag{3.39}$$

One can calculate the tagged particle's effective mass (i.e., λ^*) by looking at how an individual atom "diffuses" in imaginary time. Figure 28 shows the single-particle diffusion constant as a function of imaginary time for liquid ^{4}He, with and without bosonic exchange. The ef-
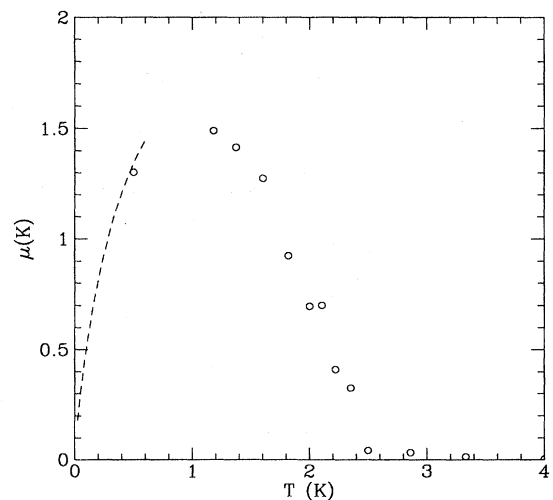

FIG. 27. The excess free energy to tag a ^{4}He atom at SVP as a function of temperature. The circles are computed from the nonexchange probability. The dashed line is the effective-mass formula.
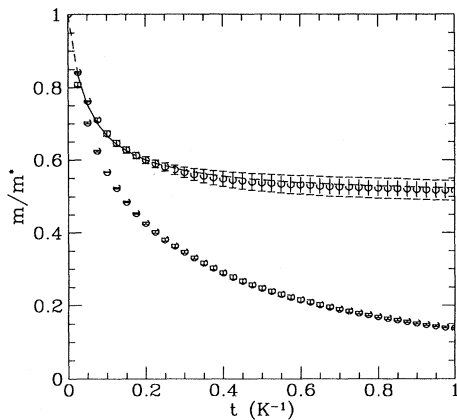
FIG. 28. The diffusion of ^{4}He atoms as a function of imaginary time. The vertical axis is $\langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle / (6\lambda t)$. The simulation was done for 64 atoms at a density of $0.02191/\text{Å}^3$ and at a temperature of 0.5 K. The upper curve is for bosons, the lower for distinguishable particles. Distinguishable particles diffuse less because they must return to their starting places after a "time" of 2 K^{-1}.

fective mass is proportional to the diffusion relative to the free-particle diffusion. For very short times, kinetic energy dominates over potential energy, so that the diffusion equals the free-particle diffusion. But at larger times, the cage of surrounding atoms in the liquid slows down the diffusion, by a factor of 2. In a liquid, this diffusion in imaginary time is Markovian, so the probability density that a particle, after a time β, has diffused a distance $\mathbf{r}$ is Gaussian:

$$G(\mathbf{r}) = (4\pi\lambda^*\beta)^{-1.5} \exp\left(-\frac{r^2}{4\lambda^*\beta}\right). \qquad (3.40)$$

These two ways of defining λ^* (from the excitation energy and from the diffusion constant) are equal only if we assume that the probability of a random walk closing on itself is the value of this distribution function at the origin, $P_1 = G(0)/\rho$. Hence we can compute the low-temperature value of P_1 by calculating the diffusion constant in imaginary time. The dashed line in Fig. 27 shows this low-temperature behavior obtained taking $\lambda^* = 0.5\lambda$.

Basile (1992) introduced a third way of calculating the effective mass. The partition function for a system with momentum $\mathbf{k}$ is obtained by using a momentum projection operator,

$$Z_{\mathbf{K}} = \frac{1}{\Omega} \int d\mathbf{r} e^{-i\mathbf{K}\cdot\mathbf{r}} \rho(R + \mathbf{r}, R; \beta) = Z_0 e^{-\lambda^* k^2}. \qquad (3.41)$$

In the small-k limit one can evaluate this effective mass to find that

$$\lambda^* = \frac{\lambda}{N}\left[1 + \frac{\langle\langle W^2 \rangle\rangle}{6\lambda\beta N}\right], \qquad (3.42)$$

where W is the winding number and $\langle\langle \cdots \rangle\rangle$ denotes an average over an ensemble in which one particle is excluded from exchanging and the action has an additional winding number bias of $-\frac{\mathbf{W}^2}{4\lambda\beta\mathbf{N}}$. Very large windings are encouraged. Because one particle cannot exchange, the winding number cannot reach its full superfluid value, so that the effective mass is greater than the bare mass. Numerically the calculations have not been carried far enough to see if precise values for the effective mass can be obtained.

Now we must do the second step in which we change the tagged atom into a real impurity. ^{3}He is a particularly nice impurity, since the interaction potential between helium atoms does not depend on the mass or statistics of the nucleus, at least within the Born-Oppenheimer approximation. With a single impurity, one does not have to worry about the Fermi statistics. One need only change the mass of the tagged particle from 4 to 3 to determine its free energy. The most convenient way to do this is to differentiate the partition function with respect to the tagged particle's mass; the derivative of the free energy with respect to the mass is the kinetic energy. Then the free energy necessary to tag a particle and change its mass is

$$\mu_3 = \mu_4 - k_B T \ln(P_1) + \int_{\ln(3)}^{\ln(4)} d\ln(m) K_m, \qquad (3.43)$$

where K_m is the kinetic energy of a tagged particle with mass m. The method of calculating free energies by integrating the derivative is a standard method used for classical simulations, where it is known as thermodynamic integration (Frenkel, 1986). It avoids the direct calculation of the partition function.

At low temperature the kinetic energy of pure ^{4}He is 14.3 K. Calculations (Bonisegni and Ceperley, 1995) for a ^{3}He impurity in ^{4}He give a kinetic energy of 17.1 K. This kinetic energy is higher because of its lower mass, while the surrounding cage in the liquid is about the same size. Recent (Wang and Sokol, 1994) deep-inelastic neutron-scattering experiments give a significantly lower kinetic energy of the impurity of 11±3 K. The end-point estimate for the mass integral gives a kinetic contribution of about 4.5 K. Thus the chemical potential at low temperature is about -7.14 K+4.5 K$\approx -2.6 \pm 0.2$ K. This is very close to, but lower than, the chemical potential of pure ^{3}He of -2.47 K. Thus the impurity can dissolve even at zero temperature. The experimental energy difference is about 0.31 K (Yorozu et al., 1992).

Other types of impurities will require changing the interatomic potential. Most other atoms are much more strongly bound to themselves than to a helium atom; thus they will not dissolve at very low temperatures. Few path-integral calculations on impurity systems have been reported, but there are no reasons why such calculations could not be carried out.

H. Exchange in quantum crystals

Up to now we have been discussing only liquid ^{4}He. One of the fantastic aspects of path integrals is that they are able also to improve our understanding of solid helium and calculate its properties on the same footing as liquid helium. No other method has been able to do this accurately. Pollock and Ceperley (1984) calculated the particle density, the structure factor, and the sound velocity in solid helium with PIMC simulation. In contrast to ground-state simulations, the solid symmetry was not put in with a trial wave function. Spontaneous melting was observed upon lowering the density. Calculated kinetic energies are in agreement with experiment (Blasdell, Ceperley, and Simmons, 1993). The thermodynamic integration method described in the previous subsection was used to determine the difference in melting pressure between ^{3}He and ^{4}He, resolving a discrepancy between theory and experiment (Boninsegni and Ceperley, 1994). An extensive study (Runge and Chester, 1988) of quantum hard spheres was made with PIMC, determining the solid-fluid phase transition, the elastic moduli, and the Lindemann melting criterion. In all of the above studies, the agreement with experiment was very good but particle statistics was not important.

Crystal ^{3}He at millikelvin temperatures is one of the simplest and cleanest examples in nature of a lattice-spin system. It is unlike any other solid, except solid hydrogen, in that the atoms have a large enough zero-point motion that they frequently change places. This has little effect in solid ^{4}He, since without spin to "label" the atoms there is no direct consequence of exchange. But in ^{3}He, exchange has the effect of permuting the spins and changing its magnetic properties. In this section, we shall describe how path integrals are related to quantum exchange and discuss the agreement between experiment and the calculations of exchange frequencies.

Solid ^{3}He at low pressure forms a bcc lattice with a Debye energy $\theta \approx 20$ K. A magnetic ordering occurs at a much lower temperature, at 1.5 mK. This is the characteristic exchange energy in solid helium. If exchange were restricted to nearest-neighbor spins, the spin Hamiltonian would be the antiferromagnetic Heisenberg model, which is rather well understood theoretically. Experimentally the phase diagram is completely different. The ground state is not antiferromagnetic, but the more complicated u^2d^2 phase with a unit cell consisting of two planes of up spins followed by two planes of down spins. Careful analysis of experimental data made plausible the model that the frequency of exchange of two, three, and four atoms is approximately equal (Roger, Hetherington, and Delrieu, 1983), so there is a competition between ferromagnetism and antiferromagnetism. Strong objections were raised against this model (Cross and Fisher, 1985), since it seemed a priori unlikely, based on estimates of exchange frequencies using the WKB method, that one would have several different exchange frequencies of the same order of magnitude. Calculation of exchange frequencies by path integrals played a crucial role in getting

this multiple-spin model accepted.

Shown in Fig. 29 is a simple model for solid ^{3}He. In this model, a single particle is confined to be in the interior of the union of two spheres. Because of the mirror symmetry, states can be classified in terms of parity. The splitting between the lowest even and odd states is the exchange frequency, $2J = E_1 - E_0 > 0$. A wave function initially localized in one of the spheres will oscillate back and forth with a frequency given by $J/\hbar$. We shall henceforth assume that the splitting energy is much less than the zero-point energy, $J \ll k_B T \ll \theta$, so higher excitations can be neglected. This is very well justified in solid helium, where $J/\theta < 10^{-5}$. Moreover, $J \ll$ phonon energies, so that solid helium at millikelvin temperatures is really in the spatial ground state. This is why solid ^{3}He is almost perfectly described by a lattice spin model.

Our goal is to calculate the exchange frequencies with path integrals. Shown in Fig. 30 is a typical world-line diagram of the imaginary-time paths in the double-sphere model. One sees that the path spends a long time in a single sphere, but occasionally it "tunnels" across to the other sphere. (The "..." is necessary because we are in imaginary time.) The tunneling is rapid, since it costs kinetic energy (not potential energy) for the system to stay very long in between the two spheres. The imaginary-time transversal of the barrier is called an "instanton" because it takes place so quickly.

The steady-state rate at which paths cross from one sphere to the other is J. For reasons that will become obvious shortly, we shall denote the centers of the two spheres by Z and PZ. Consider the ratio of the imaginary-time matrix element connecting Z to PZ to that connecting Z to itself:

$$f_P(\beta) = \frac{\rho(Z, PZ; \beta)}{\rho(Z, Z; \beta)}. \tag{3.44}$$
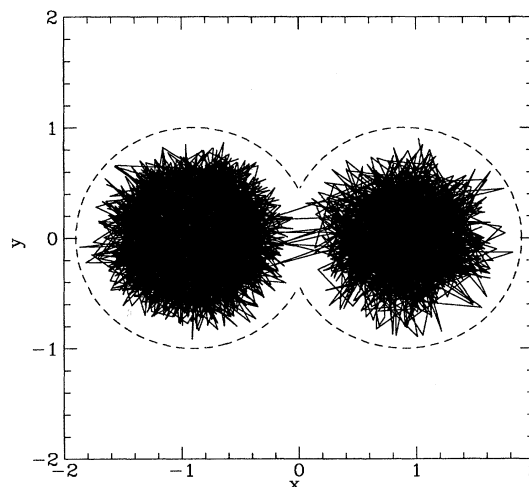

FIG. 29. The trace of a path confined to the interior of two spheres (shown with dashed lines) of radius 1, with centers separated by a distance of 1.8.
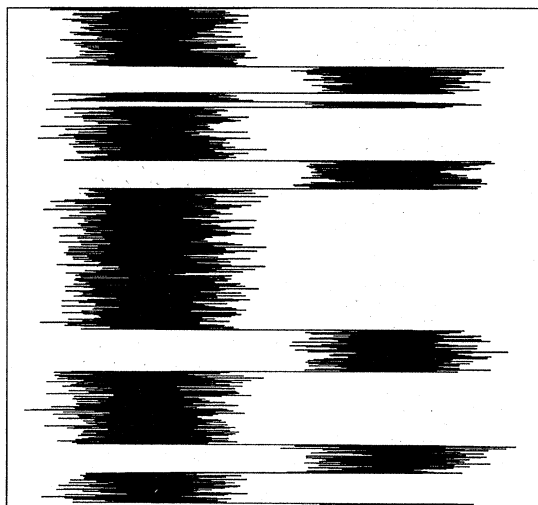
FIG. 30. The world-line view of the same path shown in the previous figure. One can see that the path stays inside one sphere for a long period, until it finds the duct to the other sphere. The actual tunneling occurs very rapidly; these transitions are called instantons.

If we now assume that β is large enough that only the lowest two states contribute to the density matrix, then we can show

$$f(\beta) = \begin{cases} 0 & \text{if } \beta < \beta_0 , \\ \tanh[J(\beta - \beta_0)] & \text{if } \beta > \beta_0 . \end{cases} \quad (3.45)$$

Here $\beta_0 = \ln[\phi_1(Z)/\phi_0(Z)]/J$. It has the interpretation of the "tunneling time." In what follows we shall assume that $\beta_0 < \beta \ll 1/J$, so that $f(\beta) = J(\beta - \beta_0)$. This justifies the claim made earlier that J is the imaginary-time diffusion rate across the barrier.

In the polymer language, J is related to the free energy it takes to pull a single end of a "linear polymer" from one sphere to the other. The difference of classical free energies can be estimated with special techniques developed for classical simulations (Bennett, 1976). We shall discuss these methods in Sec. VI.E.

Now let us generalize from the two-sphere model to a quantum crystal. We follow the theory of Thouless (1965), which is based on the earlier work of Herring (1962) on electronic exchange. Because exchange is so rare in quantum crystals, we can take as our unperturbed basis a distinguishable basis, where labeled particles are assigned lattice sites. Z denotes one such assignment of particles to lattice sites, PZ the effect of applying the permutation P to that assignment. In a crystal with N atoms and N lattice sites there are $N!$ such states, so there is an $N!$ degeneracy of the ground state in the absence of particle exchange. The splitting induced by tunneling between states Z and PZ is defined to be $2J_P$. All of the discussion concerning how to calculate J_P with

path integrals applies unchanged. The ratio of the pair-exchanged partition function to the nonexchanged one for solid ^{3}He is shown in Fig. 31.

Atomic exchange couples the nuclear spins on different lattices sites. One can show (Thouless, 1965) that the effective spin Hamiltonian is

$$\mathcal{H} = -\sum_P J_P (-1)^P P_\sigma, \quad (3.46)$$

where P ranges over all $N!$ permutations, $(-1)^P$ is the sign of the permutation, and P_σ permutes spins. One need only consider cyclic permutations of a few neighboring atoms, otherwise J_P will be vanishingly small. Two- and three-body permutations of spin-1/2 particles can be written as a nearest-neighbor Heisenberg Hamiltonian: $-J \sum_{(i,j)} \sigma_i \cdot \sigma_j$. It is not particularly convenient to transform higher-body exchanges into this form. Thouless (1965) showed that exchange of an even number of spins favors antiferromagnetism; exchange of an odd number of spins favors ferromagnetism.

It is quite difficult to calculate these exchange frequencies for solid helium with other methods. The traditional WKB approach completely fails because the exchange is not a potential barrier. The PIMC simulations show that the potential energy actually becomes *lower* as the particles exchange. This is because, in solid helium, the atoms do not sit at the bottom of the potential well, but slightly outside it. When they exchange, they move into the well and thus lower their potential energy. The physics of solid helium is much better understood by considering a hard-sphere model; the potential energy is either zero
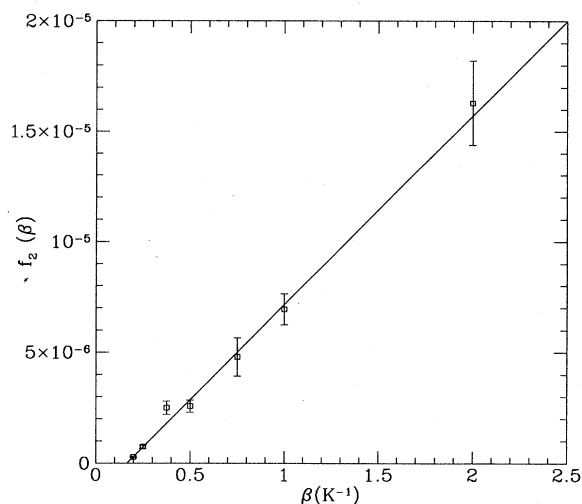
FIG. 31. The ratio of the nearest-neighbor pair-exchange density matrix to the non-exchanged density matrix as a function of inverse temperature. The calculations were done for 16 ^{4}He atoms in a bcc structure at a molar volume of 20.07 cc/mole. The straight line is a fit through the data; the slope $J_2 = 8.6 \pm 0.5$ μK is the pair-exchange frequency. The calculations shown in Table I used 54 atoms and a direct method for estimating the slope.

if no spheres are overlapping or infinite if a pair overlaps. The solid forms in order to minimize the kinetic energy, not the potential energy (see Fig. 14). Atoms are prevented from exchanging in the solid because of a kinetic-energy barrier. As they are exchanging, the wave function in the transverse direction is compressed and the kinetic energy increases. Calculations of the exchange frequency of a hard-sphere model of helium give roughly the same results as those for a realistic interatomic potential.

It is difficult to calculate the exchange frequency because it is a collective phenomenon: other atoms move out of the way of the exchanging pair. Hence it is difficult to estimate the transverse kinetic energy. Other methods of calculating J_P rely on variational wave functions. Unfortunately one needs values of the wave functions in regions where they are very small, far away from where the wave functions have been optimized, so the variationally calculated exchange frequencies are not reliable.

As we have explained above, the quantum exchange rate is related to the free energy necessary to cross-link two or more "polymers" in a "polymer crystal." It is then possible to take the computer method developed to simulate superfluid ^{4}He and find these magnetic coupling constants in the crystal ^{3}He. The exchange frequencies of ^{3}He and ^{4}He atoms in the bulk solid have been calculated by PIMC simulation (Ceperley and Jacucci, 1987) and are shown in Tables I and II. These authors found that pair interchange is most frequent, but that is followed closely by three- and four-atom exchange. So the basic conclusion of the multiple-exchange model is verified: there are several competing exchanges which give rise to the observed magnetic phenomena. But the original multiple-exchange model did not go far enough. The

PIMC calculations also show a non-negligible influence of five- and six-atom exchange.

One might wonder how we can calculate the exchange couplings of spins in a fermion system by doing a calculation of distinguishable helium atoms. Such an approach would certainly fail for liquid ^{3}He. The Thouless approach works because exchange is very rare, so that between exchanges one can label the helium atoms. Antisymmetrization is done only on the spin wave function. The spatial wave function is essentially bosonic.

Having calculated the lattice Hamiltonian, one must now solve the spin Hamiltonian to compare with experiment. This is by no means trivial; the ground state is frustrated, so quantum Monte Carlo techniques will have a "sign problem." However, other techniques—high-temperature series expansions, exact diagonalization, variational Monte Carlo, and spin-wave methods—can be used to understand the magnetic phase diagram.

When comparison can be made, the PIMC calculated frequencies are in agreement with experiment. For example, if one expands the inverse magnetic susceptibility in powers of β, the first nontrivial coefficient is conventionally denoted θ and has been measured to be -1.7 ± 0.1 K. This expansion should be valid for temperatures greater than the magnetic ordering temperature but much less than the Debye temperature. This coefficient is simply a linear combination of the exchange frequencies, $\theta = 4J_{NN} - 36J_t + 18J_f + 18J_f + \cdots$. Using the PIMC-determined frequencies we obtain the value -1.8 ± 0.5 K. The large inaccuracy occurs because of the cancellation of even and odd exchanges. The predicted value of θ if we stopped at pair exchanges would be 2.0, if we include the three largest exchange frequencies 0.2 ± 0.5 K. A less sensitive comparison of theory and experiment is given

TABLE I. Calculated exchange frequencies (in μK) in bcc ^{3}He at two densities with the statistical error in % next to it. The density (in cc/mole) is shown at the top of the column. The exchange notation gives the set of $p(p-1)/2$ pair distances among the p atoms exchanging, where 1 is a nearest neighbor, 2 a next-nearest neighbor, etc. The first set of numbers specifies the distances of adjacent atoms on the cycle, the next set the second neighbors, etc. Thus the planar and folded exchanges are both nearest-neighbor four-body exchanges, but differ in the distance between the second and fourth atoms of the exchange.

p	Exchange	Name	20.07		24.12	
2	(11)	nn	13.8	5%	453	3%
	(22)	nnn	1.0	6%	62	7%
3	(112)	t	4.8	6%	182	5%
	(113)				5.3	15%
4	$(1^4; 23)$	planar	9.8	7%	250	6%
	$(1^4; 22)$	folded	0.45	10%	32	11%
	(1122; 31)	diamond			6	25%
	(1212; 11)	eight			0.5	45%
	(1212; 14)	para			11	30%
	$(2^4; 33)$	square			1.9	30%
5	$(1^42; 52341)$	planar			5.3	15%
6	$(1^6; 3^6; 4^3)$	crown			34	22%
	$(1^6; 123^2; 417)$	planar			10	10%

by the first term in the high-temperature expansion of the specific heat, e_2, which is measured to have the value of 5.9 mK^2. With the largest three exchanges one gets a value of 4.5±0.5 mK^2 and with all the exchanges 5.0±0.7 mK^2. Some of the difference is undoubted due to the 5-, 6-, and higher-body exchanges which were not calculated.

One of the striking experimental results is the fact that the exchange frequencies scale with the helium density with a common exponent, $\eta = 18$. This is hard to understand from a traditional tunneling picture, since the exchange frequency would depend exponentially on the square root of the barrier height, and there is no reason to expect that the various exchanges would share a common barrier, especially as a function of density. In fact the common scaling with density does come out of the PIMC calculations, though it should be kept in mind that there is nothing magical about the number 18. For the three dominant exchanges, $\eta_{nn} = 19.0$, $\eta_t = 19.8$, and $\eta_p = 17.6$.

The fact that this exponent is roughly independent of the number of atoms exchanging can be understood if we assume that exchange is due to virtual vacancy-interstitial fluctuations. Once a vacancy is created it can move very rapidly, with an exchange constant $J_v \approx 2$ K more or less independent of density. (This has also been calculated with PIMC by a very similar method. One eliminates one of the atoms in the solid and cuts the "polymer" for a neighboring atom. Then βJ_v is the probability that the two ends of the cut polymer are on different sites.) The vacancy moves through the crystal, causing spins to be interchanged, and eventually recombines with the interstitial. The pair must be bound, otherwise zero-point vacancies would exist, contrary to Thouless' theory and experiment. So the exponent η reflects the probability of creating such a pair. The various exchange frequencies are a consequence of the probabilities of paths that the vacancy takes before it recombines. Such a theory has not been worked out in detail but qualitatively explains the experimental and theoretical findings.

Let us briefly return to hcp ^{4}He, the stable low-temperature solid structure. The same calculations that were done on solid ^{3}He can also be done for solid ^{4}He. The results are shown in Table II. Here there are no experiments to compare with, since there is no magnetism. One notices that exchange frequencies in solid ^{4}He are

TABLE II. Calculated exchange frequencies in hcp ^{4}He at a molar volume of 21.04 cc. The notation of the exchange is from Roger (1984).

p	Name	J (μK)	% error
2	nn	3.2	13%
	nn'	3.4	13%
3	T	2.3	12%
	$T*$	0.5	13%
	T'	2.3	12%
4	K'	1.4	20%

on the order of 1 μK, presumably because of the heavier mass, the different structure, and the higher density. There have been speculations over the years that solid ^{4}He might be a supersolid. The results of Table II suggest that the transition temperature, if it exists at all, would be at very low temperatures. The superfluid (solid) density would be proportional to the mean-squared winding number, since in the derivation of the winding number formula we made no assumptions about the arrangements of the atoms. The supersolid density would be

$$\frac{\rho_s}{\rho} = \frac{1}{6\lambda N} \sum_P J_P \mathbf{W}_P^2, \qquad (3.47)$$

where $\mathbf{W}_P$ is the winding number of permutation P. We have assumed that $\beta_{0P} < \beta << J_P$ for all significant P. It is not at all obvious that one can find such a β satisfying this inequality for all P, since the permutations which lead to winding will likely take a much longer time to tunnel. No long-range superfluidity is possible unless whole rows of atoms move in a collective fashion, since only macroscopic exchange leads to mass transport. Microscopic ring exchanges do not contribute to the superfluidity. Clearly a supersolid would exist if ground-state vacancies were present. (One has ground-state vacancies if the number of lattice sites is greater than the number of atoms.) But the experimental evidence supporting the multiple-exchange model of ^{3}He is very strong evidence against the existence of such macroscopic exchanges or vacancies. Thus the path-integral calculations in conjunction with experimental ^{3}He data strongly suggest that there will be no supersolid behavior in ^{4}He at very low temperatures.

I. The excitation spectrum

In this section we shall discuss the relation between path integrals and the excitation spectrum, an area in which progress has begun only recently. A breakthrough in understanding liquid ^{4}He came when Landau (1941) guessed that the excitation spectrum was unusual. He postulated the now-familiar picture of superfluidity based on the properties of the weakly interacting gas of phonons and rotons.

The dynamic structure factor $S_k(\omega)$ is a crucial correlation function to calculate or observe, since through it one can see the excitations. With neutron scattering one can directly measure $S_k(\omega)$. Feynman (1954) introduced a variational trial function to calculate the phonon-roton spectrum. He determined an upper bound to the excitation energy, $\omega_k \leq \lambda k^2 / S_k$, which becomes precise in the long-wavelength limit. Feynman and Cohen (1956) improved these estimates by including backflow effects. The variational approach leaves a number of questions unanswered. One is guaranteed to find a single excitation energy, not the entire excitation spectrum, and not the lifetime of the excitation. It is not clear how superfluidity comes into play in the variational approach, and extending the approach to nonzero temperatures brings in more

approximations. There are theoretical and experimental conjectures (Glyde and Griffin, 1990) that the excitation spectrum also directly reflects the hybridization between density fluctuations (phonons) and single-particle excitations, caused by the presence of the Bose condensate.

The dynamical structure factor is defined as

$$S_k(\omega) = \frac{1}{2\pi N} \int_{-\infty}^{\infty} dt e^{i\omega t} \langle \rho_k(t)\rho_{-k}(0) \rangle \qquad (3.48)$$

$$= \frac{1}{NZ} \sum_{mn} \delta(\omega - E_m + E_n) e^{-\beta E_n} |\langle m|\rho_k|n\rangle|^2. \qquad (3.49)$$

Here $\rho_k = \sum_i e^{ikr_i}$ is the Fourier transform of the density. The sum in the second equation is over all pairs of exact states $|n\rangle$ with energy E_n. The experimental measurement of $S_k(\omega)$ is shown in Fig. 32. It has a very sharp peak associated with the phonon-roton spectrum and broader multiphonon excitations. For a comprehensive review of the microscopic theory and experimental references see Griffin (1993).

Since imaginary-time path integrals are mathematically isomorphic to liquid helium, one would expect that phonons and rotons would show up dramatically in some property of the "polymer system." Such is not the case. Propagating modes in real time become diffusive modes in imaginary time. The imaginary-time density-density response function, defined as

$$F_k(t) = \frac{1}{NZ} \mathrm{Tr} \left\{ \rho_{-k} e^{-t\mathcal{H}} \rho_k e^{-(\beta-t)\mathcal{H}} \right\}, \qquad (3.50)$$

is straightforward to calculate with path integrals and is shown in Fig. 33 for various wavelengths. One can see little in the way of structure in $F_k(t)$; to the eye it is a featureless exponential decay up to $\beta/2$, and then rises
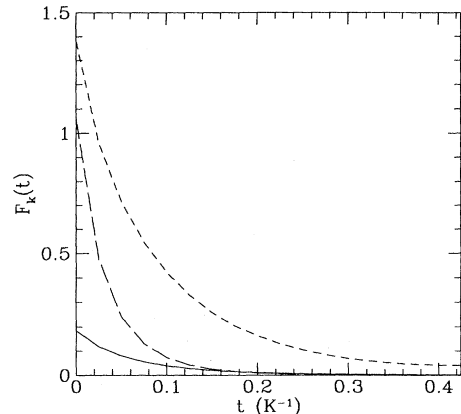
FIG. 33. The density-density response as a function of imaginary time in liquid ^{4}He at a temperature of 1.2 K as computed with PIMC: solid line, for $k = 0.76$ Å^{-1}; short-dashed line, $k = 2.01$ Å^{-1}; long-dashed line, $k = 2.52$ Å^{-1}.

again because of the periodicity in imaginary time (that part is not shown). It is related to $S_k(\omega)$ by a Laplace transform,

$$F_k(t) = \int_{-\infty}^{\infty} d\omega e^{-t\omega} S_k(\omega). \qquad (3.51)$$

Mathematically, $F_k(t)$ and $S_k(\omega)$ are equivalent, since an "inverse Laplace transform" (or equivalently analytic continuation from imaginary to real time) is well defined. The presence of statistical noise, however, rules out a direct inversion, since very small features in $F_k(t)$ come from large features in $S_k(\omega)$. The noise destroys the information needed to do the inversion. The numerical inversion of a Laplace transform is a classic ill-conditioned problem. An early work (Pollock and Ceperley, 1984) illustrates the difficulties with this inversion on phonons using PIMC data from solid ^{4}He; one can almost fit the imaginary-time data with a single excitation energy. If more modes are assumed, the PIMC data do not constrain their position very much.

It has been proposed by Silver, Sivia, and Gubernatis (1990) that Bayesian, or maximum-entropy, techniques could aid in this inversion by making the problem better conditioned. In addition to using the Monte Carlo data for the inverse Laplace transform, one can use *prior* information about $S_k(\omega)$. In some lattice models, the maximum-entropy method has given very satisfactory inversions and it has been shown to give accurate absorption spectra of an electron in liquid helium by Gallicchio and Berne (1994). In Sec. VI.F we shall discuss some numerical details of how this analytic continuation is done.

A maximum-entropy inversion from $F_k(t)$ to $S_k(\omega)$ is shown in Fig. 32 and compared to the experimental neutron-scattering measurements. We are far from a satisfactory calculation of this quantity with path integrals. Although the method reproduces the overall shape of the response function, it does not resolve the two-peak struc-
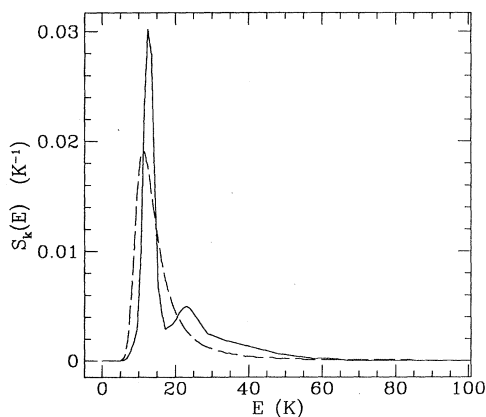

FIG. 32. The dynamical structure factor for liquid ^{4}He at a temperature of 1.2 K and wavelength $k = 0.76$ Å^{-1}: solid line, as measured by neutron scattering (Svensson *et al.*, 1976); dashed curve, as reconstructed from the PIMC imaginary-time response function by the maximum-entropy method (Ceperley and Boninsegni, 1995).

ture and does not tell us how narrow the phonon-roton line is. If we take the position of the peak height as an estimate of the excitation energy, we get the dispersion curve shown in Fig. 34. The agreement on the excitation energy with respect to experiment is better than 10%, but without further studies it is difficult to assign meaningful error bars to the inversion procedure.

The calculations could be improved by at least one order of magnitude by much longer sampling, and the inversion itself could be improved by using sum rules and other prior information. This would lead to a better numerical procedure for determining response functions. What is lacking is the direct connection between the paths and the excitation spectrum. As the temperature is raised through the transition, the excitations dramatically change character; the effect on $F_k(t)$ is hardly noticeable to the eye.

For the other quantities we have discussed, such as the superfluid density, the specific heat, and the condensate fraction, we could directly trace their change in behavior to macroscopic exchanges. There is good reason to believe that this is also the case with the excitation spectrum. When long exchange cycles are present in the paths, $F_k(t)$ contains contributions with periodicity greater than β. Take a path where a p cycle is present; part of the response function has periodicity $p\beta$, and its Fourier transform contains contributions not only at the frequencies $\omega_n = 2\pi n/\beta$ but also at $\omega_n = 2\pi n/p\beta$. These frequencies and permutation cycles are not directly measurable, but nonetheless they can have a profound influence on the analytic form of the imaginary-time correlation function. Perhaps if this information were used, a more direct picture of the excitation spectrum would come out of path integrals.

We now turn from the imaginary-time description of helium to details of how to perform calculations.

IV. CONSTRUCTING THE ACTION

It is clearly desirable to make a good but cheap approximation to the exact link action. Recall that the link action is defined as minus the logarithm of the density matrix between two successive points on a path, Eq. (2.20). 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. Cheap means that the action can be evaluated quickly on a computer, in a time not too much slower than it takes to compute the pair potential.

We have found that accurate simulations of liquid helium using the primitive approximation for the action, Eq. (2.18), 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. See Fig. 35 for an example of how the action converges for three helium atoms in an equilateral triangle separated by 2.8 Å, as a function of the time step. Without improved actions the simulation of the superfluid transition (Ceperley and Pollock, 1986) in ^{4}He would not have been possible, given the available resources. Improved actions were also crucial in calculating the exchange frequencies

FIG. 34. A comparison of the phonon-roton dispersion energies of liquid ^{4}He calculated with PIMC plus maximum-entropy inversion (filled circles) to those determined by neutron scattering and thermodynamical measurements (solid curve; Donnelly, Donnelly, and Hills, 1981).
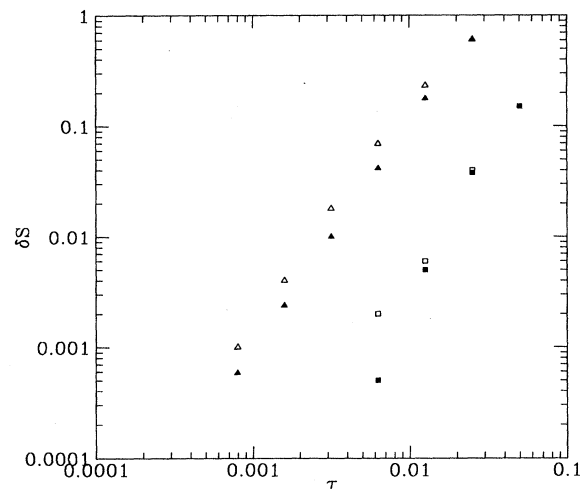
FIG. 35. The error of the action for an equilateral triangle of atoms, separated by 2.88 Å, as computed with the primitive approximation (triangles) and with the pair action (squares). The filled symbols represent the error in the action as a function of τ. The open symbols are the error in the action at $\beta = 0.025$ if intermediate time slices are inserted (with time step τ) and integrated over. This integrated action is worse, presumably because the path can wander into regions closer than 2.88 Å, where the errors are worse. The primitive action converges as τ^2, while the pair action as τ^3. To keep the error in the action less than 0.01 (corresponding to a 1% accuracy in a distribution function or matrix element) would require a primitive step of 0.003 K^{-1} or a pair step of 0.015 K^{-1}.

(Ceperley and Jacucci, 1987) in solid ^{3}He.

We shall use the Feynman-Kac (FK) formula as a guide to better actions. It gives the exact action for an arbitrary system as a path average. One derives the FK formula by taking the $M \to \infty$ limit of our basic discrete path-integral formula, Eq. (2.18). We shall write it as an average over all free-particle paths,

$$e^{-U(R_0,R_F;\tau)} = \left\langle \exp\left[-\int_0^\tau V(R(t))dt\right] \right\rangle_{RW} . \quad (4.1)$$

The notation $\langle \cdots \rangle_{RW}$ means to average over all Gaussian random walks (bridges) from R_0 to R_F in a "time" τ. Here U is the interacting part of the exact action, as defined in Eq. (2.23).

Remember that a typical Gaussian walk is the fractal shown in Fig. 5. The essence of the primitive approximation is that the size of the fractal is small enough that the potential can be assumed constant in the region of configuration space that the path could reach. Then the time integral of the potential is essentially a constant, independent of where the path goes. The fractal is typically furthest from the end points halfway through the path, when the path is on the order of

$$\Lambda_\tau = \sqrt{\lambda \tau N d/2} \quad (4.2)$$

away from the midpoint. Here Nd are the number of degrees of freedom of the path. Hence Λ_τ is the characteristic distance at which the path samples the potential. For ^{4}He with $\tau = 40K^{-1}$ it is 0.48 Å ($N = 1, d = 3$). This wavelength gets smaller only very slowly, as $\sqrt{\tau}$, as τ decreases. Hence the error of the primitive approximation might go to zero only very slowly as $\tau \to 0$. In fact, one can see in Fig. 35 that the error of the primitive approximation goes as τ^{-2}.

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. On the other hand, statistical methods for improving the convergence work very well. Because the path-integral method is based in imaginary time where motion is diffusive, the numerical approximations are much more stable than those for Newton's equation of motion. One does not have chaos in imaginary time, and all distributions at large times always decay to the ground state, independent of how they were started.

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.

We want to caution the reader not to get lost in this section. The primitive approximation contains all the physics and will converge to the correct answer, given a small enough τ, and it is simple and well defined. Using a bad action is equivalent to working with a modified potential energy. There are many cases where it is appropriate to use the primitive approximation. What we are concerned about in this section are more complex forms for the action that will give correct results for much larger τ, thereby saving enormous amounts of computer time and making calculations on large systems possible. Constructing such actions naturally could use all the standard many-body statistical-mechanical and numerical techniques, e.g., perturbation theory, cluster expansions, renormalization techniques, and integral equation methods. Little has been done in applying many-body methods to determine better actions.

Several possible strategies for choosing the high-temperature density matrix will now be examined. All of them converge to the correct answer at sufficiently high temperature, but their rates of convergence differ considerably. Note that we need to evaluate the action off the diagonal. Typically the two arguments of the action will differ by a distance $|R_F - R_0| \approx 2\Lambda_\tau$. For sampling purposes (see Sec. V), it will be necessary to have estimates of the action at times larger than τ. In liquid ^{4}He, we need accurate actions for a range of time steps, $0.1K^{-1} \le \tau \le 0.01K^{-1}$.

The simplest way to calculate the exact density matrix is to use Monte Carlo methods. It is straightforward to sample Brownian paths and to evaluate the exponential in the FK formula; see Sec. V.K. One gets into trouble for many-body systems or at low temperatures, since the exponential can fluctuate wildly. But it is possible to use direct sampling to estimate the exact action in special situations. The results can be used either to suggest new forms or to adjust parameters in an existing form. Naturally, one cannot use a second Monte Carlo procedure to calculate the action during a path-integral Monte Carlo calculation, as that would defeat the purpose of avoiding smaller time steps (although this is essentially the basis of the staging algorithm discussed in the next section). Functions that can be precomputed and tabulated are allowed, but not those that require extensive calculation during the PIMC run.

There exists a Monte Carlo method that does not have time-step errors: the Green's-function Monte Carlo method. Kalos, Levesque, and Verlet (1974) developed this exact method for the ground state of liquid ^{4}He. Whitlock and Kalos (1979) then extended these methods to finite temperature, doing a calculation of the two-particle density matrix with a Monte Carlo method without time-step errors. These methods start from the operator identity

$$e^{-\beta\mathcal{H}} = \rho_T(\beta) - \int_0^\beta dt e^{-(\beta-t)\mathcal{H}} \left(\mathcal{H} + \frac{\partial}{\partial t}\right) \rho_T(t), \quad (4.3)$$

where ρ_T is a trial or model density matrix. If the kernel of the second term is sufficiently well behaved, we can expand by repeated substitutions of the right-hand side (RHS) into the left-hand side (LHS). The kernel is the residual energy and will be discussed below. It is the local error of the trial density matrix. One starts by finding a trial density matrix such that the residual energy can be bounded from above. A method for doing this for hard spheres is discussed in Kalos, Levesque, and Verlet (1974) and for Coulomb interactions in Ceperley (1983).

In this iterative way of determining the density matrix, a new computational element has been introduced into the path-integral method: imaginary time t ranges continuously from 0 to β; it is no longer fixed on a grid. The number of time slices is a dynamical variable, since an indefinite number of iterations of Eq. (4.3) may be required. Whitlock and Kalos mapped this equation onto a diffusion and branching process, rather than onto a classical polymer system. In mapping to a polymer system, the number of beads and the time step would vary during the simulation.

This exact method has not been adopted because it is at least an order of magnitude slower than approximate methods for ground-state calculations, even including the price of several calculations to remove systematic effects. The largest uncertainties in present-day PIMC calculations for liquid helium are the scaling to the thermodynamic limit, the convergence of the Markov process, and the uncertainty in the interatomic potential. If they were ever needed to calculate absolutely accurate energies at finite temperature, Green's-function Monte Carlo methods could prove useful. But the price of the "exact action" label is at present rather high.

A. Criteria for good actions

We begin by discussing criteria that can be used to rate or choose between various proposed actions. This will help as we try to improve the action. It is also good to have a way of telling in advance if a time step is small enough that one can have confidence in the simulation.

1. Convergence studies

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. Although such convergence studies are certainly necessary to establish that the action is good and that the code is correct, one needs

other, more direct ways of improving the action.

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 than 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. In Sec. VI.F we shall describe methods based on Bayesian statistics for calculating real-time dynamics from these imaginary-time correlation functions. Experience with lattice models shows that very good actions are needed to calculate reliable responses. In particular, it is the off-diagonal action which is crucial for the dynamics.

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 (see Sec. VI) blows up at small τ. It is important that a wide range of time steps be used to test convergence, say differing by a factor of 8, since if the studied time steps are too close together, small changes may be masked by statistical fluctuations. This is why convergence studies are limited to simple systems where one can afford this overkill. But the convergence of a realistic system may be quite different from that of a simpler system.

On the other hand, one does know that time-step errors are local in space. Many-body effects only extend to a distance Λ_τ. This implies that errors in the action are independent of the number of particles. Hence it is appropriate to study small systems. One also expects the time-step error to depend primarily on τ, not on β. It is appropriate to study a single temperature and use that to establish convergence for all other temperatures. However, there could well be systematic errors if the character of the path (where it goes in configuration space) changes appreciably at lower temperature or for more particles.

There is one technique, reweighting, which can vastly speed up convergence studies (Ceperley and Kalos, 1979). First, one does a long simulation at the smallest feasible time step τ and saves a large number (at least several thousand) of paths. Using those paths, one can calculate how the energy depends on the assumed time step for time steps that are multiples of the original time step. Suppose one wants to find the systematic error in the energy if the time step is doubled. One simply throws away every second point on the path and recalculates the energy. The difference in energy $E_{2\tau} - E_\tau$ will have a much smaller statistical error than either energy alone, since the same set of paths is being used. In this way one can use a relatively short run to get some idea of the systematic errors coming from the time step before a very long run is initiated.

2. Exact properties

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,

$$U_S(R, R') = \frac{1}{2}[U(R, R') + U(R', R)]. \qquad (4.4)$$

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 $r = 0$:

$$\lim_{r_{ij} \to 0} \frac{dU(R, R'; \tau)}{dr_{ij}} = -\frac{e_i e_j}{(d - 1)(\lambda_i + \lambda_j)}. \qquad (4.5)$$

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.

One 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. We shall not discuss the long-range behavior of the action in this review, since the helium potential is short ranged.

It is useful to establish how a given property should converge as a function of τ, so that one knows how to extrapolate results at a finite time step to a zero time step. In so doing, we must keep in mind that we need the off-diagonal action, $R \neq R'$, since that is what is used in PIMC. A bound on the range of the off-diagonal action is given by the kinetic action. That tells us that we need to look at the action for pairs of configurations for which $|R - R'| \leq \Lambda_\tau \propto \tau^{1/2}$. The important off-diagonal points get closer together as τ gets smaller.

We can often determine the expansion of the action in powers of $\hbar$ (or equivalently λ) about the classical action. This is a powerful method for lattice models because then the potential is a smooth, analytic function and the higher-order terms in an expansion of the action exist and are likely to be well behaved. For singular potentials appropriate to realistic continuum models, the expansion typically diverges. Hence tests of time-step convergence should always be done on the system of interest rather than on an analytic potential such as a harmonic oscillator.

3. The residual energy

The *residual energy* of an approximate density matrix ρ_A is defined as

$$E_A(R, R'; t) = \frac{1}{\rho_A(R, R'; t)} \left[\mathcal{H} + \frac{\partial}{\partial t}\right] \rho_A(R, R'; t). \quad (4.6)$$

This is often called *the local energy*, particularly in ground-state quantum Monte Carlo algorithms. If we set the residual energy to zero, we have the Bloch equation for the density matrix. Thus the residual energy for an exact density matrix vanishes; it is a local measure of the error of an approximate density matrix. The Hamiltonian $\mathcal{H}$ is a function of R; thus the residual energy is not symmetric in R and R'. To evaluate the residual energy, we need to know how the density matrix depends on "time," and we need more than its value at a single time.

It is useful to write the residual energy as a function of the inter-action. We find

$$E_A(R, R'; t) = V(R) - \frac{\partial U_A}{\partial t} - \frac{(R - R') \cdot \nabla U_A}{t} \quad (4.7)$$
$$+ \lambda \nabla^2 U_A - \lambda (\nabla U_A)^2.$$

The terms on the RHS are ordered in powers of τ, keeping in mind that $U(R)$ is of order τ, and $R - R'$ is of order $\tau^{1/2}$. One obtains the primitive action by setting the residual energy to zero and dropping the last three terms on the RHS.

As an example, the residual energy for the primitive approximation is

$$E_1(R, R'; t) = \frac{1}{2}[V(R) - V(R')] \qquad (4.8)$$

$$-\frac{1}{2}(R - R')\nabla V + \frac{\lambda t}{2}\nabla^2 V - \frac{\lambda t^2}{4}(\nabla V)^2.$$

If $V(R')$ is sufficiently smooth around R, we can Taylor expand, and the second term will cancel against the first. If we average over R' using the free-particle distribution, the third term also cancels against the first. If the potential is very smooth, the largest contribution comes from the last term. It is also this last term which is most divergent as two particles approach each other. Suppose the interatomic potential goes as r_{ij}^{-12}; then the last term diverges as r_{ij}^{-24}. The residual energy is far from zero there!

The residual energy can be used in several ways to improve an action. First, one can rather easily establish properties of the action in various limits. For example, as two particles come close to each other, the divergence in the potential must be accompanied by divergences in some of the other terms. This is how one can derive the cusp condition, Eq. (4.5), in a few lines of algebra. Second, the generalized Feynman-Kacs formula (GFK) relates a path average of the residual energy to the exact action (see Sec. IV.F below). The GFK can be used to establish the errors of an existing action or to sug-

gest better forms. Finally, one could minimize the mean-squared residual energy to optimize parameters in the action. This is similar to what is done with ground-state wave functions (Umrigar, Wilson, and Wilkins, 1988).

4. The root-mean-square error of an action

During a PIMC simulation, one needs the difference in the action between a proposed path and the old path in order to decide whether to accept the proposed move. If the difference between the exact action U and an approximate action U_a is always very small, we expect that averages over the paths will be correct. Hence we are led to a χ^2 criterion for an approximate density matrix:

$$\chi^2 = \left\langle [U_a(R_1, R_2) - U(R_1, R_2)]^2 \right\rangle_\mu, \qquad (4.9)$$

where $\langle \cdots \rangle_\mu$ indicates an average over some distribution $\mu(R_1, R_2)$ of multidimensional configurations R_1 and R_2. Ideally one should obtain this average from the paths themselves,

$$\mu(R_1, R_2; \tau; \beta) \propto \rho(R_1, R_2; \tau)\rho(R_2, R_1; \beta - \tau). \qquad (4.10)$$

This distribution will occur in the integrand of any thermodynamic observable at a temperature β and time step τ. In fact, in most cases the β dependence is not important. It is only important that R_1 and R_2 are "typical" points, i.e., likely to be sampled.

The mean-squared error χ^2 is a dimensionless number which gives a global characterization of the quality of an action. If $\chi^2 \ll N$, we have good reason to expect that the paths produced are representative of "exact" paths. The number of particles N appears since χ^2 is an extensive quantity. What is needed in a simulation is an intensive quantity, for example, the change in action on making a local move. We shall give estimates of χ for several approximate actions to assess their ultimate accuracy. The difficulty with using this measure is that U and hence χ^2 can only be obtained with another, more computationally expensive, path-integral Monte Carlo calculation But if it can be calculated, χ is an unambiguous measure of the goodness of an approximate action.

To provide a way of benchmarking actions that we shall consider later in this section, we calculated an estimate of the exact action for points taken from several hundred paths. Using the exact action, we then determined χ for various approximate actions. The paths came from a simulation of 54 ^{4}He atoms interacting with the Aziz et al. (1992) potential with a time step of $\tau = 0.025$, at a temperature of 2 K and at SVP density. The exact action was computed by repeatedly (30 000 times) replacing each link with a new path of eight links (seven beads) and using the FK formula. The Monte Carlo error on the estimate of the exact action is about 0.03. There is also another statistical error because we have estimated χ with only several hundred different end points. Finally there is a systematic error because the action of the subchains may be biased. Nonetheless, all of these errors are much smaller than χ, so the results can usefully discriminate between various approximate actions. Table III gives these estimates of χ. For example, the rms error of the primitive approximation is 1.49. The pair density matrix is 20 times as accurate.

TABLE III. Summary of the various actions discussed. The equation number that defines the action is in column 3, the equation for the residual energy in column 4. Column 5 shows how the error of the action depends on λ and τ. n is the order of the off-diagonal expansion for the pair action as given in Eq. (4.47). The last two columns are the rms error of the action as evaluated with points R and R' taken directly from typical many-body helium paths (54 ^{4}He atoms at 2 K with a time step of $\tau = 0.025$ K^{-1}). χ is defined in Eq. (4.9). χ_D is the diagonal error, where we set $R = R'$, while χ_{OD} is the off-diagonal error. χ_D has a Monte Carlo error estimated at 0.030, while the error of χ_{OD} is 0.014.

Action			Eq.	Residual Eq.	Error	n	χ_D	χ_{OD}
Primitive	U_1		(2.24)	(4.9)	$\lambda\tau^2$		1.49	1.44
Semi-classical	U_{SC}		(4.13)		$\lambda\tau^2$		1.49	1.44
Harmonic	U_H		(4.15)		$\lambda^2\tau^3$		2.97	
Cumulant	U_C		(4.17)	(4.18)	$\lambda\tau^3$			
Pair	U_2		(4.35)	(4.36)	$\lambda\tau^3$	1	0.129	0.138
						2	0.129	0.070
						3	0.129	0.070
Pair/end point	U_{2E}		(4.37)		$\lambda\tau^2$	0	0.129	0.740
Polariz./prim.	U_{PP}		(4.53)		$\lambda\tau^3$	3	0.070	0.117
Polariz./opt.	U_{PO}		(4.54)			3	0.046	0.062
Exact	U		(2.23)	0	0		0.0	0.0

B. Semiclassical improvements

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"

$$\frac{d^2R}{dt^2} = 2\lambda\nabla V(R). \qquad (4.11)$$

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,

$$S_{SC}(R_0, R_F; \tau) = -\tau E + \int_{R_0}^{R_F} dx\sqrt{\frac{V(R) + E}{\lambda}}. \qquad (4.12)$$

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,

$$U_{SC}(R_0, R_F; \tau) = \tau\int_0^1 ds\, V(R_0 + (R_F - R_0)s). \qquad (4.13)$$

This is a better approximation than the primitive approximation since there is some contribution from the entire region between R_0 and R_F. In Fig. 38 below is shown the 1D cumulant distribution of a free-particle path between two points. For small "times" it approaches a straight-line path, justifying this approximation.

Equation (4.13) is the lowest-order approximation to U in powers of τ, assuming the potential is finite and continuous. Note that it is automatically symmetric. 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.

C. The harmonic approximation

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 defined in Eq. (4.2). 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 (see Feynman, 1972). 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

$$U_H(R_0, R_F; \tau) = \tau V(R^*) + \frac{\tau^2\lambda}{6}\nabla^2 V(R^*)$$
$$-\frac{\tau}{12}(R_F - R_0)\nabla\nabla V(R^*)(R_F - R_0)$$
$$-\frac{\tau^3\lambda}{12}[\nabla V(R^*)]^2. \qquad (4.14)$$

This expansion is equivalent to the Wigner-Kirkwood or $\hbar$ expansion of the action if we make the choice of $R^* = R_0$. See, for example, the derivation for the diagonal element in Landau and Lifshitz (1977). But with this choice, off-diagonal elements must be symmetrized with respect to R_0 and R_F as in Eq. (4.4).

If we make the *midpoint* choice, $R^* = (R_0 + R_F)/2$, the action will automatically be symmetrical in R_0 and R_F. The midpoint choice has a very important disadvantage for singular potentials: the potentials and gradients could be highly singular there, even if they are small at the end points. For example, one gets into trouble for a single quantum particle moving around the outside of a hard sphere. It is possible for the end points to be outside the hard sphere, where the potential and gradients are zero, while the midpoint is inside the sphere, where they are infinite.

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. It is higher order in τ but the same order in λ.

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. Figure 36 shows comparisons of the harmonic expansions with the exact action for a helium-helium potential and Fig. 37 for the Coulomb potential. The harmonic effective potential oscillates wildly for $r < 3$ Å. Even in the case of the much softer Coulomb interaction, the harmonic action is not a useful improvement. Usually, the time step is limited by the errors in the action at small r, so one needs an improvement of this cusp region. The harmonic action diverges there. Hence the harmonic expansion used in a straightforward way is useless.

There have been many attempts to improve this method. For example, one can try to choose the harmonic reference system better. One strategy is to choose it to minimize the free energy of the trial action; see, for example, Feynman and Kleinert (1986). One can also evaluate the exact harmonic action instead of keeping only terms of order λ. These improvements could work in a system that was nearly harmonic, for example, a nearly classical solid or a molecule with small zero-point vibrations. However, the improvements will not help much when the potential is highly anharmonic within the radius of a thermal wavelength. In addition, simply computing these higher-order corrections becomes rather time con-

FIG. 37. The diagonal action between an electron and a proton at a temperature of 5 Hartrees computed in different ways: solid line, the primitive action; long dashes, the harmonic expansion; short dashes, the exact action; dotted curve, the cumulant action. r is in units of Bohr radii.

suming in many dimensions. In Sec. VII.D we discuss the effective-potential Monte Carlo method, which is based on just such an approach.

One can make a commutator expansion of the action, which gives a closely related expansion,

$$e^{-\tau(A+B)} = e^{-\frac{\tau}{2}A}e^{-\frac{\tau}{2}B}e^{\tau^3\frac{[[B,A],A]+2B}{24}}e^{-\frac{\tau}{2}B}e^{-\frac{\tau}{2}A} \quad (4.15)$$

$$+\mathcal{O}(\tau^5).$$

See deRaedt and deRaedt (1983) for higher-order expressions. This sort of expansion could be useful on a lattice, since the operators are bounded. In the continuum, the commutators are highly singular at short interparticle distances, so that the expansion is not uniformly convergent in space. In addition, they can also be slow to evaluate.

D. The cumulant approximation

If the potential energy is sufficiently smooth, one can take the averaging process of the FK formula up into the exponent. Let us recall the well-known (van Kampen, 1981) cumulant expansion of the exponential of a random variate,

$$\langle e^x \rangle = \exp[\langle x \rangle + \tfrac{1}{2}(\langle x^2 \rangle - \langle x \rangle^2) + \cdots]. \quad (4.16)$$

Here x is some random variate, the brackets represent an average, and higher-order cumulants have been neglected. If x is normally distributed, all the higher-order cumulants are zero and drop out. This is not usually the case, but it suggests that the expansion might be quickly convergent if x were close to being a Gaussian variable,
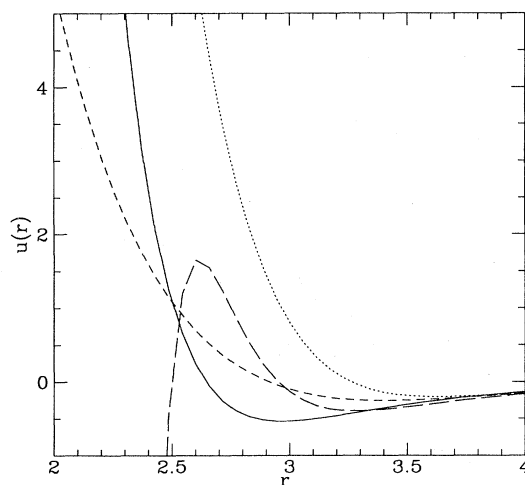

FIG. 36. The diagonal action of two ^{4}He atoms interacting with the Aziz et al. (1987) potential at 20 K computed in different ways: solid line, the primitive action; long dashes, the harmonic expansion; short dashes, the exact action; dotted curve, the cumulant action. r is in units of Å.
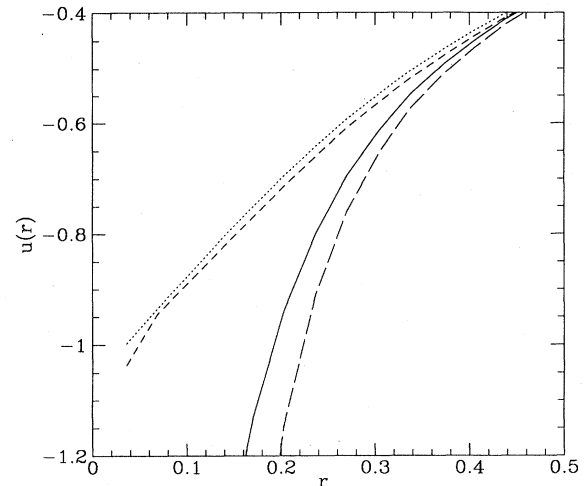
as it would be if it were the sum of uncorrelated terms.

To relate this to the FK formula, we let $x = \int_0^\tau dt V(R(t))$, the total potential of a path, and average over all free-particle paths from R_0 to R_F. Then the first-order cumulant action is

$$U_C(R_0, R_F; \tau) = \int_0^\tau dt \, \langle V(R(t)) \rangle_\mu, \qquad (4.17)$$

where we average over the total density of points, $\mu(R)$, the *cumulant density*, sampled by the random walk in going from R_0 to R_F. The cumulant density and the calculation of the cumulant action are described in the next subsection.

The cumulant action will exist only if the potential energy is integrable, that is, only if $\lim_{r \to 0} r^d V(r) = 0$ (d is the spatial dimensionality). The Lennard-Jones or hard-sphere potentials are not integrable. One might think this is not a serious issue, since most physical potentials are Coulombic at sufficiently small r. But even though a realistic helium-helium potential is integrable, it is sufficiently large at short distances that the cumulant action is a serious overestimate of the helium-helium action; see Fig. 36. On the other hand, for Coulombic systems the cumulant action is quite accurate; see Fig. 37.

The residual energy of the cumulant action has a very simple form:

$$E_C(R, R_0; \tau) = -\sum_i \lambda [\nabla_i U_C(R, R_0; \tau)]^2. \qquad (4.18)$$

The sum is over all particles, and the gradient is with respect to the particle positions at one end point. The residual energy is strictly negative. This can be used to provide a rigorous way of sampling the exact density matrix (Ceperley, 1983) and is related to the fact that the cumulant action is greater than the exact action. Jenson's inequality implies that the first cumulant is an upper bound to the exact average,

$$\langle e^x \rangle \geq e^{\langle x \rangle}. \qquad (4.19)$$

Interestingly enough, we can use Jensen's inequality the other way to get a lower bound to the action, so that

$$\tau \langle V(R) \rangle_S \leq U \leq \tau \langle V(R) \rangle_\mu = U_c. \qquad (4.20)$$

Here the left average, $\langle \cdots \rangle_S$, is over exact paths and the right average, $\langle \cdots \rangle_\mu$, is over the free-particle path distribution.

Since U_c is of order τ the residual energy is of order τ^2. Below we shall see that this implies that corrections to the cumulant action must be of order τ^3. The second-order cumulant is proportional to the correlation between the potential at one time and the potential at another time. Presumably one could develop ways of calculating it, as well. For example, one could use graphical methods, as outlined in Kac (1959).

E. Calculation of the cumulant action

It is not trivial to calculate the cumulant action. Let us write it in the form

$$U_C = \tau \int dR \mu_c(R|R_0, R_F, \tau) V(R). \qquad (4.21)$$

Here $\mu_c(R; R_0, R_F, \tau)$ is the cumulant density, the total density of points that a path visits as it moves from R_0 to R_F. This distribution is shown in one dimension in Fig. 38 and in two dimensions in Fig. 39.

Because we shall need it in the next section, let us first define the *conditional probability distribution*, the probability of sampling a point $\mathcal{R} = (R, t)$ given that the path has already visited $\mathcal{R}_1 = (R_1, t_1)$ and will in the future visit $\mathcal{R}_2 = (R_2, t_2)$ (note the space-time notation):

$$\mu(\mathcal{R}|\mathcal{R}_1, \mathcal{R}_2) = \frac{\rho(\mathcal{R}_1, \mathcal{R})\rho(\mathcal{R}, \mathcal{R}_2)}{\rho(\mathcal{R}_1, \mathcal{R}_2)}. \qquad (4.22)$$

The conditional distribution is well defined for interacting particles, but for free particles it has a very simple analytic form: it is a Gaussian distribution, centered at

$$\bar{R}_t = R_1 + \frac{(t - t_1)}{(t_2 - t_1)}(R_2 - R_1) \qquad (4.23)$$

and with a variance equal to

$$\sigma_t = \frac{2\lambda(t - t_1)(t_2 - t)}{t_2 - t_1}. \qquad (4.24)$$

The mean of the free-particle conditional distribution proceeds in a straight-line path from R_1 to R_2. At the beginning and end it is a delta function and achieves its maximum width at the midpoint. Figure 40 depicts this
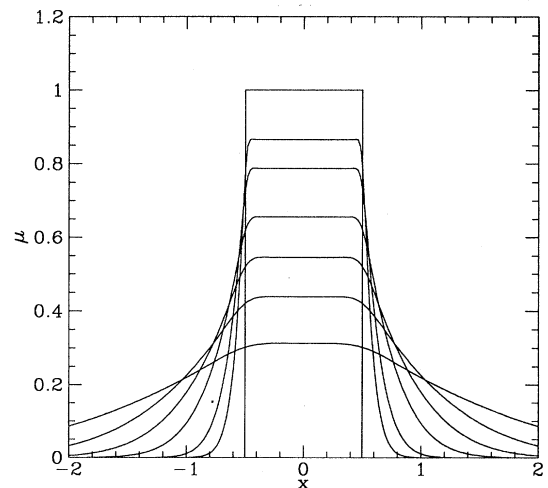

FIG. 38. The cumulative distribution of a 1D path which goes from $1/2$ to $-1/2$. The seven curves are for different amounts of imaginary time. The top curve is a classical distribution in which the path proceeds uniformly from start to end. The others are for the following values of temperature: 0.1, 0.2, 0.5, 1.0, 2.0, 5.0 in units where $\lambda = 1$.
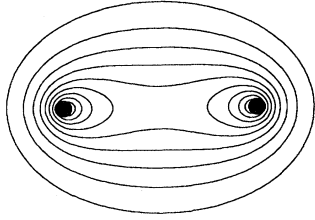
FIG. 39. A contour plot of the 2D cumulative probability distribution of a walk that proceeds from one dark circle to the other. The levels show contours of the probability distribution function (pdf) from 10% to 90% of the maximum. The integral of the pdf over the area of the largest contour contains 87% of the total probability. Figure 38 is the projection on the x axis of this plot. The distance between the end points is equal to the thermal wavelength.

time-dependent flow of probability as the walk progresses from R_0 to R_F.

The cumulant distribution is simply the time average of the conditional distribution,

$$\mu_c(R|R_1, R_2, \tau) = \frac{1}{\tau} \int_0^\tau dt \mu(R, t|\mathcal{R}_1, \mathcal{R}_2). \quad (4.25)$$

Although μ_c has a simple integral representation, it is, in general, not an elementary function.

One can make some progress in performing the cumulant integral with Fourier transforms. First note that the cumulant action is linear in the potential; thus if the potential is a sum of terms, each term can be considered independently. So without loss of generality we can consider the case of a single particle in a spherically symmetrical external potential. This is equivalent to a pair potential if one works with relative coordinates and reduced masses. Define the Fourier transform of the potential as

$$v_{\mathbf{k}} = \frac{1}{(2\pi)^d} \int d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} v(\mathbf{r}). \quad (4.26)$$

Transforming the integral into k space, we find that the cumulant action is
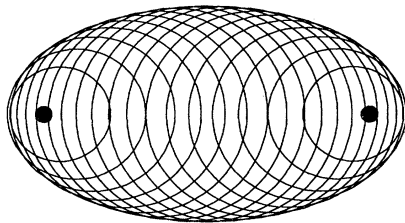
FIG. 40. A representation of the 2D conditional probability distribution as it progresses from one point to another point. The center of the circle is the mean value of the distribution, the radius is its width. Shown are 20 time steps.

$$u_c(\mathbf{r}, \mathbf{r}'; \tau) = \int_0^\tau dt \int d\mathbf{k} \exp\left(-i\mathbf{k}\bar{\mathbf{r}}_t - \frac{\sigma_t}{2}k^2\right) v_{\mathbf{k}}. \quad (4.27)$$

If we first perform the time integral, we get for the general off-diagonal case an error function of complex argument. On the diagonal, where $\bar{\mathbf{r}}_t$ is independent of t, one can write down the Fourier transform in terms of Dawson's integral $F(x)$ (see Chap. 7 of Abramowitz and Stegun, 1970),

$$u_c(\mathbf{r}, \mathbf{r}; \tau) = \tau \int d\mathbf{k} e^{-i\mathbf{k}\mathbf{r}} v_{\mathbf{k}} F(y)/y, \quad (4.28)$$

where $y = \sqrt{\lambda k^2 \tau/2}$. Using the asymptotic limit of Dawson's integral, one can show that the effect of the path averaging is to cut off the high-$\mathbf{k}$ components; they get reduced by a factor $1/(\lambda k^2 \tau)$.

On the other hand, one could first perform the integral over $\mathbf{k}$ in Eq. (4.27). Define the smoothed potential $v_s(r, \sigma)$ as

$$v_s(r, \sigma) = \int d\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{r} - \frac{\sigma}{2}k^2} v_{\mathbf{k}}. \quad (4.29)$$

In the case of a Coulomb potential the smoothed potential has a simple analytic form, $v_s(r, \sigma) = \text{erf}(r/\sqrt{2\sigma})/r$, showing that the smoothing cuts off the potential at small distances to give the correct cusp to the action. Then the cumulant action is the integral along the straight line connecting the two end points of the smoothed potential:

$$u_c(\mathbf{r}, \mathbf{r}'; \tau) = \int_0^\tau dt v_s(\bar{r}_t, \sigma_t). \quad (4.30)$$

This form is reminiscent of the semiclassical action, Eq. (4.13), except now the potential has been smoothed by the quantum fluctuations so the action is more accurate. Tabulations and analytic approximations of the cumulant action for the Coulomb potential are given in Ceperley (1983) and Pollock (1988).

F. The pair-product action

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. This approach was suggested by Barker (1979). To justify this approach, first assume that the potential energy can be broken into a pairwise sum of terms,

$$V(R) = \sum_{i<j} v(r_i - r_j). \quad (4.31)$$

Now apply the Feynman-Kacs formula, Eq. (4.1). 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,

$$x_{ij} = \exp\left[-\int_0^\tau dt v(\mathbf{r}_{ij}(t))\right]. \quad (4.32)$$

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

$$e^{-U} = \left\langle \prod_{i<j} x_{ij} \right\rangle. \qquad (4.33)$$

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

$$e^{-U} \approx \prod_{i<j} \langle x_{ij} \rangle. \qquad (4.34)$$

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

$$U_2(R, R'; \tau) = \sum_{i<j} u_2(\mathbf{r}_{ij}, \mathbf{r}'_{ij}; \tau), \qquad (4.35)$$

where $u_2(\mathbf{r}_{ij}, \mathbf{r}'_{ij}; \tau)$ is the exact action for a pair of atoms. In the next subsection, we discuss how to calculate it.

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 probability for that depends on the density.

Second, the pair density approximation is perfectly well defined for all potentials, either Coulombic or hard-sphere-like, in contrast to the cumulant action, which exists only for integrable potentials. Third, at low temperatures the pair action approaches the solution of the two-particle wave equation. The result is the pair-product or Jastrow ground-state wave function, which is the ubiquitous choice for a correlated wave function because it does such a good job of describing most ground-state correlations.

Finally, the residual energy of this action is less singular than for other forms:

$$E_2(R, R'; \tau) = -\sum_i \lambda_i \left[\sum_{j \neq i} \nabla_i u_2(\mathbf{r}_{ij}, \mathbf{r}'_{ij}; \tau) \right]^2$$
$$+ \sum_{i<j} (\lambda_i + \lambda_j) \left[\nabla_i u_2(\mathbf{r}_{ij}, \mathbf{r}'_{ij}; \tau) \right]^2. \qquad (4.36)$$

The pair residual energy equals the form of the cumulant residual energy, Eq. (4.18), except that all the purely pair terms have been subtracted out. This must be so because the pair density matrix is exact for a single pair. Consider what happens when two particles approach very close to each other, say $r_{12} \to 0$. From Eq. (4.36) the divergent term involves a single factor of $\nabla_1 u(\mathbf{r}_{12}, \mathbf{r}'_{12})$. For an r^{-12} potential, $u(r) \propto r^{-5}$ and thus as $r_{12} \to 0$, $E_2 \propto r_{12}^{-6}$. However, the cumulant residual energy, since it still has pair terms, will diverge as the square of this term, as $E_C \propto r_{12}^{-12}$. From Fig. 36 we see that the pair density matrix for two ^{4}He atoms is only nonzero for two atoms with $r_{ij} < 3$ Å. But looking at the pair-correlation function we see that such pairs are relatively rare. We will pick up a contribution to the residual energy only when a given atom i has *two* atoms closer than 3 Å. This explains why the pair action becomes much better at low density. Because of the form of the residual energy, we expect errors in the action to go as τ^3. This is confirmed in Fig. 35 in an explicit calculation of the action for three helium atoms.

Many authors (Pollock and Ceperley, 1984) use the *end-point* approximation to the pair action,

$$u_{2E}(\mathbf{r}_{ij}, \mathbf{r}'_{ij}; \tau) = \frac{1}{2} \left[u_2(\mathbf{r}_{ij}, \mathbf{r}_{ij}; \tau) + u_2(\mathbf{r}'_{ij}, \mathbf{r}'_{ij}; \tau) \right]. \qquad (4.37)$$

This approximation is exact on the diagonal ($\mathbf{r} = \mathbf{r}'$) and is symmetrical in R and R'. It is very convenient computationally, since the pair potential in the primitive action simply gets replaced by an effective potential, and the action depends only on the radial variables r_{ij}. Hence once $u_2(r; \tau)$ has been computed and saved, its use takes the same amount of computer time as the primitive action (assuming potentials are calculated with look-up tables). However, most of the accuracy of the pair action is lost by the end-point approximation, as we show in Table III. Jacucci and Omerti (1983) have verified that it is primarily the off-diagonal terms that are needed to improve the density matrix, by looking at the convergence of the pair density matrix for two hard spheres. The full pair action contains valuable information about how easily two atoms can circle around each other. This is lost in the end-point approximation. Using the end-point approximation for the action reduces χ in liquid ^{4}He by only a factor of 2. But using the full pair action reduces χ by a factor of 20!

G. Calculation of the pair density matrix

The exact pair action can be calculated efficiently by the matrix-squaring method introduced by Storer (1968) and Klemm and Storer (1973). 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:

$$\rho(\mathbf{r}, \mathbf{r}'; \tau)$$

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

$$(4.38)$$

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

$$-\frac{\partial\rho(r, r'; t)}{\partial t} = \left[-\lambda \frac{d^2}{dr^2} + \tilde{v}_l(r) \right] \rho(r, r'; t) \qquad (4.39)$$

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

$$\tilde{v}_l(r) = v(r) + \frac{\lambda}{4r^2} \begin{cases} 4l^2 - 1, & \text{2D}, \\ 4l(l+1), & \text{3D}, \end{cases} \qquad (4.40)$$

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

$$\rho_l(r, r'; \tau) = \int_0^{\infty} dr'' \rho_l(r, r''; \tau/2) \rho_l(r'', r'; \tau/2) \ . \quad (4.41)$$

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, $\rho_l(r, r')$. This matrix-matrix product is implemented very efficiently on many computers. However, a uniform grid with no interpolation runs into difficulty when high accuracy is required. At high temperature, the density matrix is sharply peaked near the diagonal, so the off-diagonal elements can disappear between the grid points. Also, there may not be a sufficient number of grid points in places where the potential changes quickly. Use of a nonuniform mesh in r can significantly reduce the number of integrals to be performed, since the mesh points can be concentrated in the region where the potential is very steep. Elements far from the diagonal do not need to be computed, reducing the number of required integrals. Finally, the integrand in Eq. (4.41) is Gaussian-like, and thus Hermite integration is a natural choice of numerical integration. (See Table 25.10 in Abramowitz and Stegan, 1970.)

The number of partial waves needed requires some experimentation but is primarily dictated by the free-particle part of the solution and the final temperature; for example, 20–60 partial waves are needed to reach 40

K for ^{4}He. One needs fewer at lower temperatures, more at higher temperatures.

To start off the matrix-squaring iterations, Eq. (4.41), one needs a high-temperature form for the density matrix. This is the only place the potential actually enters. Since it is the solution to a 1D problem, we can use the approximations developed in the previous sections. A convenient approximation is the semiclassical action, Eq. (4.13),

$$\rho_l(r, r'; \tau) = \rho_l^0(r, r'; \tau) \exp\left(-\frac{\tau}{|r - r'|} \int_r^{r'} v(x) dx \right),$$

$$(4.42)$$

where the partial-wave component of the free-particle density matrix is

$$\rho_l^0(r, r'; \tau) = \begin{cases} \frac{2\pi\sqrt{rr'}}{4\pi\lambda\tau} \exp(-\frac{(r^2+r'^2)}{4\lambda\tau}) I_l\left(\frac{rr'}{2\lambda\tau}\right), & \text{2D}, \\ \frac{4\pi rr'}{[4\pi\lambda\tau]^{3/2}} \exp(-\frac{(r^2+r'^2)}{4\lambda\tau}) i_l\left(\frac{rr'}{2\lambda\tau}\right), & \text{3D}, \end{cases}$$

$$(4.43)$$

with I_l and i_l the modified and modified spherical Bessel functions, respectively. [See Secs. 9.6–9.8 and 10.2 in Abramowitz and Stegun, 1970. Note that $i_l(z) = \sqrt{\pi/(2z)} I_{l+1/2}(z)$.]

There are at least two alternative ways of computing the pair density matrix. First, one can directly use the eigenfunction expansion of the density matrix,

$$\rho_l(r, r'; \tau) = \sum_n \phi_n^*(r) e^{-\tau E_n} \phi_n(r')$$

$$+ \int_0^{\infty} dk \phi_k^*(r) e^{\tau\lambda k^2} \phi_k(r'), \qquad (4.44)$$

where ϕ_n and ϕ_k are the bound and continuum stationary solutions of the partial-wave Bloch equation (4.39). This is particularly convenient for Coulomb or hard-sphere potentials, since there are analytic expressions for the continuum wave functions (Pollock, 1988). Second, one can use the Feynman-Kacs formula directly, performing the path integral with Monte Carlo simulation (Pollock and Ceperley, 1984). See Sec. V.K. This is not particularly efficient because each off-diagonal element requires a separate calculation, but, because Monte Carlo simulation is simple, it is a very good way of checking the results of a matrix-squaring evaluation.

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. The pair density matrix between atoms at initial positions $(\mathbf{r}_i, \mathbf{r}_j)$ and final positions $(\mathbf{r}_i', \mathbf{r}_j')$ reduces to a function of four variables— three relative distances and the difference of times. Explicit construction of a four-dimensional table is possible, but unnecessarily wastes computer memory. We can use

the fact that the initial and final positions cannot be too far apart. It is convenient to use the three distances

$$q = (|\mathbf{r}| + |\mathbf{r}'|)/2 , \quad s = |\mathbf{r} - \mathbf{r}'| , \quad z = |\mathbf{r}| - |\mathbf{r}'| , \quad (4.45)$$

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:

$$u(\mathbf{r}, \mathbf{r}'; \tau) = \frac{u_0(r; \tau) + u_0(r'; \tau)}{2} + \sum_{k=1}^{n} \sum_{j=0}^{k} u_{kj}(q; \tau) z^{2j} s^{2(k-j)}. \quad (4.46)$$

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. The first-order terms in this expansion are shown in Fig. 41 for the helium-helium interaction. Notice that the off-diagonal terms are significant in the region 2 Å$< r <$ 3 Å. A similar expansion may be written for the τ derivative of u_{nm}, needed in estimating the internal energy. For the $1/r$ interaction in either two or three dimensions, it can be shown that there is no dependence on z (Pollock, 1988) because of a special symmetry of the Coulomb interaction.

The convergence of this expansion for helium at 40 K in terms of χ, defined earlier, is shown in Fig. 42. We see that, in the important region 2.5 Å$< r <$ 3 Å, including each additional higher-order term reduces the rms error by approximately one order of magnitude. When χ drops
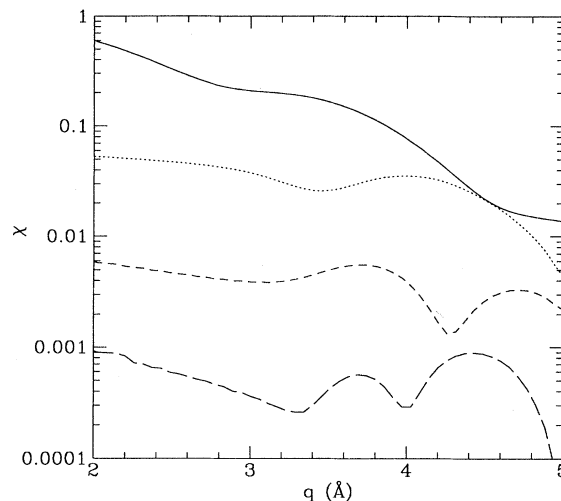
FIG. 42. The error χ of the action for two ^{4}He atoms at 20 K located an average distance of r apart. The variable q is defined in Eq. (4.45). The averaging to determine the error is over s and z. The four curves represent the fits obtained by choosing n in Eq. (4.46) from 1 to 4 (top to bottom). Thus the top curve represents the average error of the end-point approximation.

to a level below 0.01, most thermodynamic properties should be well converged. Our second-order pair action, i.e., $n = 2$, meets this criterion for $r > 2$ Å. Most of the accurate calculations of liquid helium have used $n = 2$. Since there are six terms in the expansion for $n = 2$, the evaluation of the pair action will be at most six times slower than the evaluation of the primitive action. In fact, the terms all share common computation, so it is only about twice as slow. Because one can drastically reduce the number of time slices, using an accurate pair action is much more efficient than using the end-point action, even though it is somewhat slower to evaluate.

H. Beyond the pair action

In the previous subsection we have discussed how to use the exact two-atom density matrix. Naturally the question arises whether it is possible to go further without getting tangled up in very-high-dimensional functions. While the exact pair density matrix is at most a four-dimensional function (with two variables small and τ fixed), the complete three-body density matrix has 13 variables (of which nine are small and τ fixed), so it is clearly much more difficult to calculate and to tabulate the three-body density matrix. It is known from ground-state variational calculations that the dominant terms beyond the pair-product (Jastrow) wave function are not complicated functions but of the simple "polarizability" form. We shall now see how that form arises for the density matrix also.
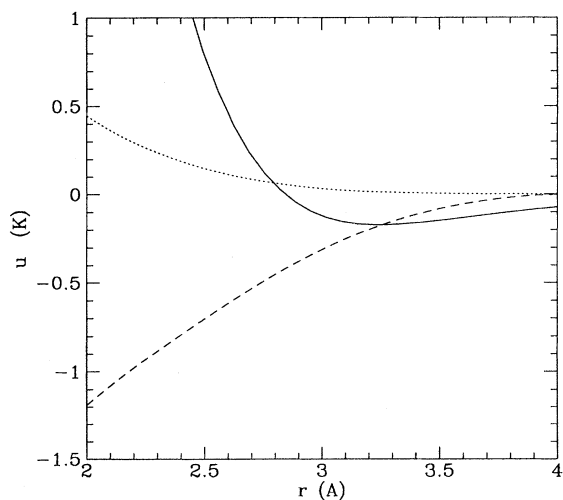

FIG. 41. The diagonal and off-diagonal components of the exact pair density matrix for two ^{4}He atoms vs distance at a temperature of 40 K: solid curve, the diagonal action; dotted curve, u_{10}; dashed curve, u_{11}.

First, we have to establish the connection between the residual energy of any approximate action U_A and the exact action U. Let $\delta U = U - U_A$ be the difference between the exact action and an approximate action. The generalized Feynman-Kacs (GFK) formula expresses δU as a path-integral average over drifting random walks generated by the approximate action,

$$e^{-\delta U(R_0, R_F; \tau)}$$
$$= \left\langle \exp\left[-\int_0^\tau E_A(R(t), R_F; \tau - t)dt\right]\right\rangle_{\text{DRW}},$$
$$(4.47)$$

where $\langle \cdots \rangle_{\text{DRW}}$ denotes the average over all drifting random walks that proceed from R_0 to R_F in time t, and E_A is the residual energy defined earlier, Eq. (4.6). The drifting random walks are defined by the Langevin process,

$$\frac{dR(t)}{dt} = \eta(t) - \frac{R - R_F}{\tau - t} - 2\lambda \nabla U_A(R, R_F; \tau - t) \quad (4.48)$$

with $\eta(t)$ white noise, $\langle \eta(t)\eta(t')\rangle = \lambda\delta(t - t')$, and initial condition $R(0) = R_0$. This process describes Brownian motion in a velocity field that pushes the walk so that it arrives at R_F after a time τ. The derivation of this result is in Pollock and Ceperley (1984).

If the approximate action is taken to be the free-particle action, the original Feynman-Kac formula is obtained, since the residual energy equals the bare potential and the drift merely serves to generate Brownian random walks which begin at R_0 and end at R_F. On the other hand, if U_A is the exact action, then E_A is seen from its definition to be identically zero. Then the velocity field generates fully interacting paths starting from R_0 and ending up at R_F.

The GFK formula gives an exact, nonperturbative expression to correct any approximate action. But to use it to determine a better action we have the difficult problem of how to calculate the averages $\langle \cdots \rangle_{\text{DRW}}$. One can use the techniques already discussed to perform the averages—the primitive approximation, the semiclassical approximation, the cumulant average, or an exact solution of a subsystem. We shall discuss only the primitive approximation to the path average, which assumes that the residual energy is a constant within a thermal wavelength of the initial and final points. But the residual energy will in general depend on the time remaining in the path. Suppose we assume it has a power-law behavior,

$$E_A(R, R_F; t) \approx (t/\tau)^n E_A(R_0, R_F; \tau). \quad (4.49)$$

We get

$$\delta U(R_0, R_F; \tau) \approx \frac{\tau}{(n+1)} E_A(R_0, R_F; \tau). \quad (4.50)$$

Accuracy will be improved if we symmetrize with respect to R_0 and R_F.

Now, let us reexamine the residual energy expressions coming from the approximate actions that we have already considered. First, if we start from the free-particle action, then $n = 0$, and we end up with the primitive action. Second, if we start with the pair action, then $n = 2$, and using Eq. (4.36), the GFK action has the *polarization* form

$$\delta U_{PL}(R_0, R_F; \tau) = -\sum_i \frac{\tau\lambda_i}{3}\left[\sum_{j\neq i}\nabla_i u_2(\mathbf{r}_{ij}, \mathbf{r}'_{ij}; \tau)\right]^2$$
$$+ \sum_{i<j}\frac{\tau(\lambda_i + \lambda_j)}{3}\left[\nabla_i u_2(\mathbf{r}_{ij}, \mathbf{r}'_{ij}; \tau)\right]^2.$$
$$(4.51)$$

It is interesting to note that, if $U_2 = \tau V$, the polarization term exactly equals the $\tau^3\lambda$ term in the harmonic expansion, so this method gives the same result as an $\hbar$ expansion in lowest order. Moreover, the polarization term is identical in functional form to the term that is needed for improving the ground-state variational wave function of liquid ^{4}He helium beyond the pair or Jastrow wave function (Schmidt and Ceperley, 1992). So this form of the action can go over smoothly as temperature is lowered to a very accurate ground-state wave function. This polarization term, even though it is three-body in form, still takes only $\mathcal{O}(N^2)$ computer operations to evaluate. For a short-ranged potential, it can be evaluated using neighbor tables in $\mathcal{O}(N)$ operations.

In the above derivation, we simply took the residual energy as the missing action. This completely neglects the averaging over paths. To approximately include an effect of the averaging, one can assume that the correction to the action is of the above *form* but allow the function that appears inside to be renormalized. To further simplify, let us also make the end-point approximation. Hence

$$\delta U_P(R_0, R_F; \tau) = -\sum_{i=1}^N \lambda_i\left[\sum_{j\neq i}(\mathbf{r_i} - \mathbf{r_j})w(r_{ij})\right]^2$$
$$+ \sum_{i<j}(\lambda_i + \lambda_j)\left[r_{ij}w(r_{ij})\right]^2, \quad (4.52)$$

where $w(r)$ is the adjustable *polarization* function. The derivation above suggests that

$$w_P(r) = \sqrt{\frac{\tau}{12}}\frac{du_2(r)}{rdr}. \quad (4.53)$$

This we call the *unoptimized polarization*. But the total action for the unoptimized polarization can have an instability (the action goes to $-\infty$) if a triplet of atoms collapses to a common point. From Table III we see that putting in an unoptimized polarization function decreases the error on the diagonal by a factor of 2 but actually makes things worse off the diagonal.

To optimize the polarization function $w(r)$, we have computed the exact density matrix for a sequence of equilateral triangles of ^{4}He atoms and use this to fix $w(r)$ by
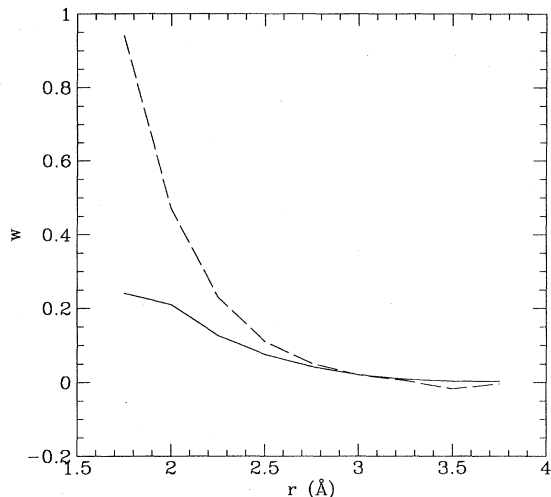
FIG. 43. The polarization function computed either from the residual energy expression, Eq. (4.51) (dashed line) or by Monte Carlo simulation (solid line) from the exact action of three ^{4}He atoms arranged in an equilateral triangle with side r and with a time slice $\tau^{-1} = 20$ K.

$$w_O(r) = [(U - U_2)/(3\lambda r^2)]^{1/2}. \qquad (4.54)$$

This is shown in Fig. 43 and compared with the unoptimized polarization function. It is seen that in the region of $r > 3$ Å where the two functions are small, they agree. But at smaller distances, the optimized w is much smaller and smoother. This is precisely what one expects the averaging would do. The paths coming from the pair density matrix spend a greater fraction of the path in regions where the residual energy is smaller and thus reduce the correction to the density matrix. This optimized polarization function reduces the error even further on the diagonal. Off the diagonal it is an improvement, but not much of one. Probably one needs also to include off-diagonal components of the polarization function. We have already learned that this was important at the pair level, so it is probably true as well for the polarization action.

V. PATH SAMPLING METHODS

The previous sections concerned the physical interpretation of the paths and how to construct accurate high-temperature actions. Here we consider how to do the multidimensional integrations and summations that path integrals require. Without introducing very severe approximations, the only way of doing the integrals is stochastically, that is, by sampling the integrand.

The total configuration space to be integrated and summed over is made of elements: $s = [P, R_1, ..., R_M]$ where $R_k = \{\mathbf{r}_{1k}, ..., \mathbf{r}_{Nk}\}$ 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

$$\pi(s) = \frac{\exp[-\sum_{k=1}^{M} S^k]}{Z}, \qquad (5.1)$$

where S^k is the action of the kth 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. What is less often discussed is the difficulties in exploring permutation space; that is, how the permutations can be changed. Convergence in permutation space is the key to getting reliable results concerning the superfluid transition and other effects of quantum statistics. We shall emphasize methods that can be used both to speed up the sampling of a single chain and to allow sampling of exchanges. The reader should bear in mind that all of the methods discussed here are correct, in the sense that they will eventually converge. What is at stake is how much it will cost in computer time to get a given accuracy.

First, we raise the possibility of doing the integrals with molecular dynamics. We then introduce the general form of the Metropolis Monte Carlo method, define what we mean by optimal sampling, and consider what is the best sampling for a single link of a path. We then discuss the motivation for moving whole sections of a path, using either Fourier-space methods or the bisection method. The bisection method is the only one that has been applied to continuum superfluids. We then describe the problems and methods for sampling permutation space. Finally, we apply these methods to the direct calculation of the density matrix.

A. Molecular dynamics

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 (Parrinello and Rahman, 1984; deRaedt, Sprik, and Klein, 1984). The chief difficulty with dynamical methods, by which we mean those in which the path 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 equi-

librium. 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. Hall and Berne (1984) did simulations of liquid neon and liquid water and found that straight molecular dynamics did not agree with the Monte Carlo results. With the strong harmonic forces in the classical polymer system, the dynamics is in the Kolmogorov-Arnold-Moser (KAM) regime and the flow will be confined to multidimensional tori. 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 (1984) thermostat, or one can resample the velocities (hybrid Monte Carlo) each dynamical step and accept or reject the changed velocities (Duane et al., 1987). 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. (Beware of a new danger of confusion, as there will be a time step associated with this dynamics, distinct from the τ of imaginary time.) One finds that the paths move very slowly through phase space. Tuckerman et al. (1993) 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.

These results are interesting because there are situations in which dynamical techniques are advantageous. In Monte Carlo methods, an elementary move is typically of one or several particles at several time steps, while in molecular dynamics, one moves all particles at all time steps simultaneously. In some cases, this can be computationally much more efficient. For example, one might want to do a PIMD calculation of a system where the Born-Oppenheimer electronic energy is calculated with a dynamical method such as the Car-Parrinello (1986) local-density-functional algorithm. The electronic calculation must be done for all the ions simultaneously, so one may as well move all the path variables as well. Another example of a case in which dynamical methods may prove useful is provided by fermion path integrals. The matrix elements of backflow density matrices (Ceperley, 1991) are all coupled together, and it is convenient to calculate them together. Finally, systems with constrained degrees of freedom, like a molecule with fixed bond lengths, can usually be handled more conveniently with dynamical methods. These applications will not be discussed in this review. Here we are primarily interested in the problem of Bose condensation, and discrete movement through permutation space is crucial. Henceforth we shall discuss only Monte Carlo methods for performing the path sampling.

B. Markov chains and the Metropolis algorithm

The vast majority of path-integral calculations are done with various generalizations of the Metropolis et al. (1953) rejection algorithm, which is a particular type of Markov process (Hammersley and Handscomb, 1964). To construct a Markov chain, one changes the state of the system according to a fixed transition rule, $\mathcal{P}(s \to s')$, thus generating a random walk through state space, $\{s_0, s_1, s_2 \ldots\}$. [Note that $\mathcal{P}(s \to s')$ is a probability distribution.] If the transition probability is ergodic, the distribution of s_n converges to the unique equilibrium state which solves the equation,

$$\sum_s \pi(s)\mathcal{P}(s \to s') = \pi(s'). \tag{5.2}$$

The transition is ergodic if 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 or conserved quantities that restrict the walk to a subset of the full configuration space.

The transition probabilities are usually set up so that they satisfy detailed balance: the transition rate from s to s' equals the reverse rate,

$$\pi(s)\mathcal{P}(s \to s') = \pi(s')\mathcal{P}(s' \to s). \tag{5.3}$$

Assuming ergodicity, detailed balance is sufficient to guarantee that one samples $\pi(s)$ in the limit of many steps.

The Metropolis method is a particular way of ensuring that the transition rules satisfy detailed balance. The transition probability is split into an "a priori" sampling distribution $T(s \to s')$ and an acceptance probability $A(s \to s')$,

$$\mathcal{P}(s \to s') = T(s \to s')A(s \to s'). \tag{5.4}$$

In the original Metropolis procedure, $T(s \to s')$ was chosen to be a constant distribution inside a cube and zero outside. In the generalized Metropolis procedure (Kalos and Whitlock, 1986), more general types of sampling distributions are allowed and trial moves are accepted according to

$$A(s \to s') = \min\left[1, \frac{T(s' \to s)\pi(s')}{T(s \to s')\pi(s)}\right]. \tag{5.5}$$

It is easy to verify detailed balance and hence asymptotic convergence with this procedure. Moves that are not accepted are rejected and remain at the same location for at least one more step. Note that accepted or rejected steps contribute to averages in the same way.

For bosons, it is necessary to have several different kinds of moves, for example, moves that change path

variables and 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 or the time slice to be updated from some distribution. Equation (5.5) still applies if one includes in the definition of $T(s \rightarrow s')$ the probability of selecting that move from the menu. A more common procedure is to go through all possible atoms and time slices systematically. After one *pass*, moves of all coordinates have been attempted once. Equation (5.2) can be used to justify the correctness of this procedure. 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 now 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 long-ranged 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 variance of $\mathcal{O}$ be denoted by $\bar{\mathcal{O}} = \langle \mathcal{O}_k \rangle$ and $\nu_{\mathcal{O}} = \langle (\mathcal{O}_k - \bar{\mathcal{O}})^2 \rangle$, where the averages are over π. These quantities depend only on the distribution, not on the Monte Carlo procedure. On the other hand, the standard error of the estimate, the average of $\bar{\mathcal{O}}$ over a Markov chain with P steps, is

$$\text{error}[\bar{\mathcal{O}}] = \sigma_{\mathcal{O}} = \sqrt{\frac{\kappa_{\mathcal{O}} \nu_{\mathcal{O}}}{P}}. \tag{5.6}$$

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

$$\kappa_{\mathcal{O}} = 1 + 2 \sum_{k=1}^{\infty} \frac{\langle (\mathcal{O}_0 - \bar{\mathcal{O}})(\mathcal{O}_k - \bar{\mathcal{O}}) \rangle}{\nu_{\mathcal{O}}}, \tag{5.7}$$

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 for optimal transition rules.

To determine the true statistical error in a random walk, one needs to estimate this 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*. 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 for all of the time slices and connected with the identity permutation. However, it may take a very large number of steps for the initial solid to melt. The quantum paths are no different in this respect from classical systems; metastability and hysteresis are characteristic near a phase boundary. A random starting place is with a random permutation and placing each path variable randomly in the simulation box. It may be very difficult for the paths to disentangle themselves from this starting place. More physical starting places are well-converged paths 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,

$$\xi_{\mathcal{O}} = \frac{1}{\sigma_{\mathcal{O}}^2 P T} = \frac{1}{\kappa_{\mathcal{O}} \nu_{\mathcal{O}} T}, \tag{5.8}$$

where T is the computer time per step. 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 $\nu_{\mathcal{O}}$ is independent of the sampling algorithm.

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

$$D_R = \left\langle \frac{[\tau \sum_{k=1}^{M} (R_k - R_k')]^2}{T} \right\rangle, \tag{5.9}$$

where $R_k - R_k'$ is the total change in the kth time slice in one Monte Carlo step and T is the CPU time per step. Note that this change is zero if a move is rejected. The time step τ is in the formula so one can make relative comparisons with different values of τ and β. We can introduce a similar measure for movement through per-

mutation space,

$$D_P = \left\langle \frac{|P - P'|}{T} \right\rangle, \qquad (5.10)$$

where $|P - P'|$ counts the number of elements of the permutation that are changed in an (accepted) permutation move. For example, if only pair permutations are being considered, D_P will be twice the number of accepted pair moves per second. Finally the number of winding changes per second, D_W, is an important measure of the speed at which the winding number distribution converges. The values of these diffusion constants depend not only on the computer and the algorithm, but also on the physics. Diffusion of the atoms and change of the permutation 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. (By the way, it is important 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,

$$T^*(s \to s') = \frac{\pi_{s'}}{C_s}, \qquad (5.11)$$

where the normalization constant is

$$C_s = \sum_{s'' \in \mathcal{N}(s)} \pi_{s''}. \qquad (5.12)$$

Then one sees by substitution in Eq. (5.5) that the acceptance probability will be

$$A(s \to s') = \min\left[1, \frac{C_s}{C_{s'}}\right]. \qquad (5.13)$$

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.

This heat-bath rule is frequently used in lattice spin models where one can easily compute the normalization constant. Note that the normalization constant is 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. (5.12) extends over all space. We shall later use the heat-bath rule directly to sample permutation changes, but for path variables we shall try to find a method close to the heat bath, so that the correlation time is small, but able to be executed quickly. We take this up in the next subsection.

C. Single-slice sampling

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 that we shall build upon in the following sections. 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. *All the integrals, vectors, and tensors in this subsection will range only over the n moving-atom coordinates.* The other particles can be regarded as a fixed background.

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. It is clear that Δ must be on the order of, or smaller than, the thermal de Broglie wavelength for a slice, $\Delta \approx \Lambda_\tau = \sqrt{\lambda\tau}$.

As mentioned above, 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

$$T^*(R) \propto \rho(\mathcal{R}_1, \mathcal{R})\rho(\mathcal{R}, \mathcal{R}_2). \qquad (5.14)$$

The ratio of normalization factors in Eq. (5.13) is always one, because the neighborhood does not change after the move. With T^* all moves are accepted.

For a lattice model, it would be straightforward to sample T^*, but in the continuum one cannot quickly compute the needed normalizations. In the rest of this section we shall discuss a way of sampling an approximate T^*. Dropping factors independent of R and factoring out the free-particle action,

$$T^*(R) \propto \exp\left[-\frac{(R - R_m)^2}{2\sigma} - U(\mathcal{R}, \mathcal{R}_1) - U(\mathcal{R}, \mathcal{R}_2)\right], \qquad (5.15)$$

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. Although it would be possible to develop sophisticated ways of sampling T^*, it has been found more efficient to further approximate T^* 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. Note first an important difference between the Monte Carlo sampling of a classical liquid and that of path integrals. In Monte Carlo on a classical system, the new atom could be anywhere in the box, while the effective neighborhood of the path is localized by the kinetic action to a region of the order of Λ_τ. This makes it easier to develop approximate expressions for the optimal path-integral sampling distribution.

To go beyond free-particle sampling, we can choose for a transition probability the most general correlated Gaussian in $3n$ variables,

$$T_S(R) = \sqrt{(2\pi)^{3m}\det(\mathbf{A})}e^{-(R-\bar{R})(2\mathbf{A})^{-1}(R-\bar{R})}, \quad (5.16)$$

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 $T^*(R)$. Basing the parameters of the Gaussian on the moments will not necessarily lead to the best transition probability, but it does give a unique rule for determining them and it works very well if T^* happens to be close to a multivariate Gaussian, as is the case for small τ.

To develop expressions for $\mathbf{A}$ and $\bar{R}$, we define a *sampling potential* by

$$\exp[-\tilde{U}(R_m)] = W = \int dR T^*(R) \quad (5.17)$$

$$= \left\langle e^{-U(\mathcal{R},\mathcal{R}_1)-U(\mathcal{R},\mathcal{R}_2)} \right\rangle_{R_m,\sigma}$$

where the $\langle \cdots \rangle$ simply indicates an average over the free-particle Gaussian centered at the midpoint. The exact moments of T^* can be written in terms of derivatives of $\tilde{U}$,

$$\bar{R} = \frac{\int dR\, R\, T^*(R)}{W} = R_m - \sigma\frac{\partial\tilde{U}}{\partial R_m}, \quad (5.18)$$

$$\mathbf{A} = \frac{\int dR\,(R-\bar{R})\,(R-\bar{R})\,T^*(R)}{W}$$

$$= \sigma\left(\mathbf{I} - \sigma\frac{\partial^2\tilde{U}}{\partial R_m\partial R_m}\right), \quad (5.19)$$

where $\mathbf{I}$ is the unit tensor. We see that $\tilde{U}$ is a generating function for the moments of T^*. If we approximate $\tilde{U}$ (the integral is too slow to evaluate during the PIMC) we can automatically generate expressions for the Gaussian parameters by differentiating.

Now, we make three further approximations. First, we make the end-point approximation for the action in Eq. (5.18),

$$U(\mathcal{R},\mathcal{R}_1) + U(\mathcal{R},\mathcal{R}_2) \approx U(R, R; \tau) + \text{const.} \quad (5.20)$$

Second, we assume the diagonal action is a sum of pair terms so that $U(R,R;\tau) = \sum_{i<j}u(r_{i,j},r_{i,j};\tau)$. Third,

for the average in Eq. (5.18) we make the uncorrelated approximation: we assume the terms in the sum to be uncorrelated so that we can interchange the product and averaging operations,

$$W \approx \prod_{i<j}\left\langle e^{-u(\mathbf{r}_{ij};\tau)}\right\rangle_{R_m,\sigma}. \quad (5.21)$$

Then the averages can be explicitly performed, since they involve two atoms at a time. This is the same trick we used in subsection IV.E to define the pair action. Then the pairwise sampling potential $\tilde{u}$ is

$$e^{-\tilde{u}(\mathbf{r}_{m,ij};\tau)} = \int d\mathbf{r}_i d\mathbf{r}_j \exp\left[-\frac{(\mathbf{r}_i - \mathbf{r}_{m,i})^2}{2\sigma}\right.$$

$$\left. -\frac{(\mathbf{r}_j - \mathbf{r}_{m,j})^2}{2\sigma} - u(\mathbf{r}_{ij}, \mathbf{r}_{ij};\tau)\right]. \quad (5.22)$$

Using spherical coordinates, this integral, a convolution, reduces to two 1D Fourier transforms. The evaluation is done before the Monte Carlo calculation begins and $\tilde{u}$ is tabulated. The total sampling potential is

$$\tilde{U} = \sum_{i<j}\tilde{u}(\mathbf{r}_{m,ij};\tau). \quad (5.23)$$

The above equation is for the case in which both atoms, i and j, are moving. If either atom is fixed, say i is a fixed atom and j is a moving atom, one should replace the Gaussian in $\mathbf{r}_i$ with its actual (old) position.

Correlated sampling proceeds as follows. First the midpoint R_m for the move is determined. Then $\bar{R}$ and $\mathbf{A}$ are determined using Eqs. (5.18) and (5.20) from the tabulated functions $\hat{u}$. To sample the multivariate normal distribution, one Cholesky-factorizes the covariance matrix as $\mathbf{A} = \mathbf{SS}^T$, where S is an upper triangular matrix. Then if χ is a vector of Gaussian random numbers with zero mean and unit variance, $\mathbf{S}\chi + R_m$ has the desired mean and variance. The diagonal divisors in the Cholesky decomposition of $\mathbf{A}$ are needed to find the actual value of $T(R)$ and the acceptance probability for a move in Eq. (5.5).

Because of the uncorrelated approximation used in calculating the covariance matrix, Eq. (5.21), $\mathbf{A}$ may have unphysical negative eigenvalues. This occurs very rarely, typically every few thousand moves. For these rare cases, free-particle sampling, i.e., setting $\tilde{U} = 0$, may be used instead. This will affect only the acceptance ratio. If this happens very frequently, it may be worthwhile to define the sampling potential in some other way, so that its second derivative is smaller.

The effect of atomic interactions on the free sampling is to push the mean position of an atom away from its free-particle mean, if another (nonmoving) particle is nearby or if a moving particle has its free-particle mean nearby. Similarly, the free-particle variance 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. Putting correlation into the transition prob-ability is much more important when several particles are being moved at once, as during a permutation move. Then correlation is essential to keep them out of each other's way.

Finally, consider the high-temperature limit. If the width $\sigma^{1/2}$ is much smaller than variations of the potential, then we see from Eq. (5.18) that $\tilde{U}(R_m) = U(R_m) = \tau V(R_m)$. The mean position is pushed by the classical force. This is equivalent to smart Monte Carlo. See Rossky, Doll, and Friedman (1978) and Allen and Tildesley (1987). For longer time steps, $\tilde{U}(R_m)$ differs from $\tau V(R_m)$ in that it is smoothed out over regions of radius Λ_τ.

Another strategy for improving the sampling is to approximate the potential locally with a harmonic potential and use the density matrix of the harmonic potential for the transition probability. One will have achieved many of the features discussed above, since that density matrix is a correlated Gaussian. This approach will work well if locally the potential is harmonic, e.g., for tightly bound atoms in a molecule or solid. One application for which this has been tried is on protons in water molecules (Friesner and Levy, 1984). On the other hand, it will have difficulties when the potential is nonharmonic. Liquid helium is far from harmonic; the atoms do not like to sit on the bottom of the potential well but rather where the potential has a negative curvature. The harmonic approximation does not solve the main problem, for helium is the hard-core interaction. The sampling potential defined above works for any type of pair interaction. It runs into difficulty if T^* is highly non-Gaussian or when the uncorrelated approximation breaks down.

D. The necessity of multislice moves

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 of 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. *Time in this subsection refers to steps of the Markov process, which moves the system through path space.* Consider a Monte Carlo procedure in which each bead is moved in turn. If we ignore "short-time" details, the evolution of the paths is described by a diffusion equation, the Smoluchowski equation, and the movement of the paths is equivalent to Rouse (1953) dynamics of polymers. A single path has a purely harmonic

action, so one can explicitly solve the Smoluchowski equation and determine the complete relaxation spectrum of the path. The mathematical problem is equivalent to solving for the phonon frequencies of a harmonic crystal. One finds that the inverse relaxation times of a path with M slices are $(n + \frac{1}{2})[1 - \cos(2\pi m/M)]$, where n and m label modes. Thus the slowest mode, $(n = 0, m = 1)$, will take on the order of $(M/\pi)^2$ passes to come to equilibrium. But since the computer time per pass also scales as M, we find that computer time needed to change the overall shape of a single path scales as M^3.

The diffusion of the center of mass $\mathbf{c}$ can be analyzed in the same way. After a move of a single bead, the mean-squared center of mass will change as

$$\langle (\delta\mathbf{c})^2 \rangle = \frac{1}{M^2}\langle (\delta\mathbf{r})^2 \rangle \leq \frac{3\lambda\beta}{M^3}, \qquad (5.24)$$

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.

Our estimate so far has worried only about the internal degrees of freedom of a chain. Entanglement effects coming from the interaction of several atoms will slow the relaxation further. But as M gets large, the kinetic action dominates, so our analysis will give the asymptotic scaling correctly. 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
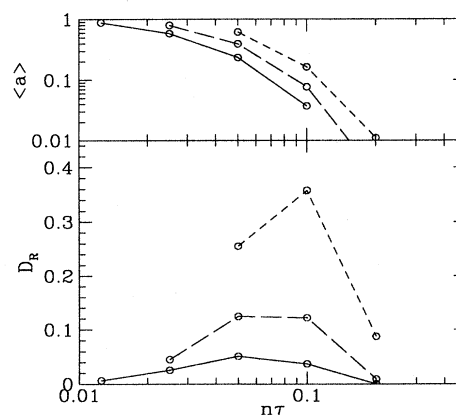
FIG. 44. The diffusion constant and acceptance ratio for 64 ^{4}He atoms at saturated vapor pressure and 2 K: solid line, simulation with $\tau = 0.0125/\mathrm{K}$; long-dashed curve, simulation with $\tau = 0.025/\mathrm{K}$; short-dashed curve, simulation with $t = 0.05/\mathrm{K}$. The horizontal axis refers to the "time" of the move, i.e., $n\tau$. The moves were made with the bisection algorithm with free-particle sampling. All moves were of single atoms, no permutations were done. D_R is the diffusion constant for the center of mass of a single chain in $\text{Å}^2/\mathrm{K}^2\mathrm{s}$ on a SPARC-10 work station. $\langle a \rangle$ is the total acceptance ratio for the same kind of move.

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. A very large price has been paid for simple actions and simple moves! To achieve faster convergence, one must go beyond moves of a single time slice.

Figure 44 shows the diffusion constant as a function of τ and the number of time slices simultaneously moved. Numerically we find that D_R scales as τ^3 if single-slice moves are made. The single-slice moves are the three leftmost points of the three lower curves. We next describe sampling methods with a much more favorable scaling, where several time slices are moved together.

E. Whole-chain and normal-mode sampling

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" (Feynman and Hibbs, 1965). We define the normal-mode coordinate by

$$Q_k = \sum_{l=1}^{M} R_l e^{2\pi i k l/M}. \tag{5.25}$$

The total kinetic action is decoupled in normal modes,

$$\begin{aligned} K &= \frac{1}{4\lambda\tau} \sum_{i=1}^{M} (R_i - R_{i-1})^2 \\ &= \frac{1}{\lambda\beta} \sum_k \sin^2(\pi k/M)|Q_k|^2. \end{aligned} \tag{5.26}$$

Each of the $3NM$ normal-mode variables Q_k is independent of the others and has a Gaussian distribution. This process of transforming into normal modes was used to

make the picture of Fig. 8, where normal modes with $k > 10$ were set to zero.

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 (Takahashi and Imada, 1984; Runge and Chester, 1988). 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 $\mathcal{O}(M)$ operations.

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* (Coalson et al., 1988). 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

$$R(t) = R_0 + \sum_{k=1}^{M} Q_k e^{-2\pi i t k/\beta}. \tag{5.27}$$

The path is no longer a fractal, but a smooth curve, as we have depicted in Fig. 8. The discussion of the previous section concerning improved actions is largely irrelevant. Instead, one has to worry about approximating the effect of the neglected modes, $k > M$. Their effect can be accounted for by using the cumulant action. This is called partial averaging. One works with an effective potential, which is obtained by averaging over the neglected higher modes,

$$V_{\text{eff}}(r, u) = (2\pi\sigma_u)^{-3/2} \int d\mathbf{r}' e^{-r'^2/2\sigma_u} v(r + r'), \tag{5.28}$$

assuming this integral exists where $\sigma_u = 2\lambda\beta[u(1 - u) - \frac{2}{\pi^2} \sum_{k=1}^{M} \sin^2(ku\pi)k^{-2}]$ is the "time-dependent" width.

There have as yet been no systematic comparisons of efficiency of the Fourier-based methods versus the discrete-time path integrals. Some model examples of exchange have been tried with these methods by Chakravarty (1993), but they have not yet been ex-

tended to superfluid systems. It would seem difficult to efficiently construct exchanging paths in Fourier space, while it is not difficult in coordinate space, as we shall describe in Sec. V.J.

F. The multilevel metropolis method

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. We describe the method in general terms in this subsection and apply it to path and permutation moves in the following subsections.

Suppose the full configuration s is dynamically partitioned at the beginning of a Monte Carlo step into $l + 1$ levels $s = (s_0, s_1, ..., s_l)$, 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 $(s'_1, ..., s'_l)$ 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 shall explicitly describe how the levels are constructed in the following subsection.

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 $-\ln(\pi_k)$.] We require only that *feasible paths* (ones that have a nonzero probability) have a nonzero level action and that the action at the last level be exact,

$$\pi_l(s_0, s_1, ..., s_l) = \pi(s). \qquad (5.29)$$

The optimal level action would be one where $\pi_1(s_0, s_1)$ is proportional to the reduced distribution function of s_1 conditional on s_0, i.e., where the finer levels are integrated out. That is, the optimal kth level action is

$$\pi_k^*(s_0, s_1, ..., s_k) = \int ds_{k+1} \cdots ds_l \ \pi(s). \qquad (5.30)$$

This optimal choice is only a guideline. Nonoptimal choices will simply lead to slower movement through phase space.

Now, we choose a sampling rule for s_k contingent on the levels already sampled. $T_k(s'_k)$ can depend on $s'_0, s'_1, ..., s'_{k-1}$. By $T_k(s_k)$ we refer to the same function of the coordinates $s_0, s_1, ..., s_{k-1}$. We require only that T_k be a probability distribution and that T_k be nonzero for feasible paths. The optimal choice is given by the heat-bath rule,

$$T_k^*(s'_k) = \pi_k(s')/\pi_{k-1}(s'). \qquad (5.31)$$

One can show that T_k^* will be a normalized probability if and only if π_k is defined as in Eq. (5.30).

Once the partitioning and the sampling rule T_k are chosen, the algorithm is quite simple. The sampling proceeds past level k with probability

$$A_k(s') = \min \left[1, \frac{T_k(s_k)\pi_k(s')\pi_{k-1}(s)}{T_k(s'_k)\pi_k(s)\pi_{k-1}(s')} \right]. \qquad (5.32)$$

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. Here π_0, needed in the first level, is arbitrary and can be set to one, since it will cancel out of the ratio. This acceptance probability has been constructed so that it satisfies a form of detailed balance for each level k:

$$\frac{\pi_k(s)}{\pi_{k-1}(s)} T_k(s'_k) A_k(s') = \frac{\pi_k(s')}{\pi_{k-1}(s')} T_k(s_k) A_k(s). \qquad (5.33)$$

As is usual in the Metropolis approach, one must compute not only the forward-move action and transition probability, but also the reverse-move action and the probability of sampling the old path. The moves will always be accepted if the transition probabilities and level actions are set to their optimal values.

The total transition probability for a trial move making it through all l levels is

$$\mathcal{P}(s \to s') = \prod_{k=1}^{l} T_k(s') A_k(s'). \qquad (5.34)$$

By multiplying Eq. (5.33) from 1 to l and using Eq. (5.29), one can verify that the total move satisfies the detailed balance condition, Eq. (5.3). Thus the algorithm will asymptotically converge to π independent of T_k and the level actions π_k.

G. The Lévy construction

Before we describe the bisection method, it is useful to describe a simpler and earlier algorithm for sampling a free-particle path, *the Lévy construction*. In the Lévy (1939) construction of a Brownian bridge, one begins with two fixed end points (say R_0 at "time" 0 and R_β at "time" β) and samples a bisecting point at time $\beta/2$, exactly as we have already described with free-particle sampling in Sec. V.C. The bisecting point R of the interval $(\mathcal{R}_1, \mathcal{R}_2)$ is

$$R = \frac{R_1 + R_2}{2} + \eta, \qquad (5.35)$$

where η is a normally distributed random vector, with mean zero and covariance, $\sqrt{\lambda |t_2 - t_1|}$. For free particles, this exactly samples the action except for a minor problem with the periodic boundary conditions, so there

are no rejections. 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.

H. 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 $R_{i+1}, \dots, R_{i+m-1}$. 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

(i) s_0 = atom positions outside of time slices in consideration and atoms not being moved.

(ii) s_1 = coordinates of atoms being moved at the middle time slice $i + m/2$.

(iii) s_2 = coordinates of atoms being moved at $i + m/4, i + 3m/4$.

(iv) ...

(v) s_l = coordinates of atoms being moved at $i+1, i+3, ..., i+m-1$.

Now we need to define an action at the kth level. The optimal choice from Eq. (5.30) is the integral over all the variables in the higher levels. If we assume that all the atoms are being moved, the integral over the time slices yet to be sampled can easily be performed. The optimal level action would be simply the product of density matrices with the appropriate time argument. For the first level we get

$$\pi_1^*(R_{i+m/2}) = \rho(\mathcal{R}_i, \mathcal{R}_{i+m/2})\rho(\mathcal{R}_{i+m/2}, \mathcal{R}_{i+m}). \quad (5.36)$$

In practice, we are moving only a few atoms at once, so even if we could calculate the exact action at $(m/2)\tau$, we make an approximation in assuming all atoms are being moved. The product property of density matrices holds only if all particle coordinates are being integrated over.

In fact, the detailed form of the level action does not matter very much. One is looking for a quick and dirty way of deciding whether to continue the bisection procedure or to reject and start over. Any function that detects overlaps will accomplish this. 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 in Sec. IV, but now accuracy is less important than speed. For example, one can use the end-point approximation for the action, neglecting off-diagonal terms. An effective mass can easily be incorporated to compensate for the off-diagonal contributions. At the final level, the exact action must be calculated [see Eq. (5.29)], 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 in subsection V.C: how to sample a single time slice optimally. 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.

In "classic" Monte Carlo calculations, there is a step-size parameter Δ which is adjusted to make the average acceptance ratio close to 50%. The analogous parameter in this method is the number of levels l and the number n of atoms involved in a move. Figure 44 shows the diffusion of the center of mass as a function of both τ and the number of slices being moved. If the level is too small, diffusion is slowed because of the pinning of the fixed end points of a move. But if the level is chosen to be too large, the acceptance ratio, shown in the upper panel, becomes too small. We find that diffusion through phase space is optimal when the total acceptance ratio is approximately 0.2.

Let us examine now how the efficiency of the bisection algorithm scales with τ. Using the maxima of Fig. 44, we find that the efficiency scales roughly as $\tau^{1.4}$. Any algorithm must scale as τ, since merely calculating the action takes time proportional to the number of time slices. The additional power of 0.4 presumably arises because additional time slices cause rejections, which slow the diffusion through phase space. This scaling is much superior to the τ^3 scaling characteristic of single-slice sampling.

In another type of multiple-time-slice method, *the threading algorithm* (Pollock and Ceperley, 1984), one cuts out a section of m time slices and generates a new path by growing from time slice 0 to slice n with the diffusion algorithm

$$R_{i+1} = R_i - 2\tau\lambda\nabla S_T(R_i, R_n; (n-1)\tau) + \eta_i\sqrt{2\tau\lambda}.$$

$$(5.37)$$

Here η_i is a normally distributed random vector with zero mean and unit variance, and S_T is a trial action. As

in the bisection method, the time argument of S_T will be greater than τ if more than one time slice is being updated. The trial action is used only to guide the walk, and any convenient approximation can be used for it. Inaccuracies will affect only the acceptance ratio, not the converged distribution. After the new path is generated, it is accepted or rejected in the usual Metropolis fashion based on the difference between the old and new actions and on the ratio of the sampling probabilities for the old and new paths. The form of the diffusion can be shown to be optimal in the sense that, if the trial action were exact, with all atoms being moved and τ sufficiently small, the acceptance ratio would be one. In practice, moves of more than a few time slices are often rejected. The disadvantage is that a lot of work has to be done before the move is finally accepted or rejected, so the method has been superseded by the bisection method.

The *staging algorithm*, which has been applied to a single electron in a classical liquid by Sprik, Klein, and Chandler (1985), is based on ideas similar to those for the bisection method. The first level is sampled using the free-particle action, and then a second level is repeatedly sampled (hundreds of times) to find the action of the first level. One moves a primary polymer chain, but to accept or reject the move one introduces secondary chains of P steps. In this way, one computes a level action directly when it is needed. The multilevel sampling method handles this much more effectively, since only a single secondary chain is introduced. Cruzeiro-Hansson, Baum, and Finney (1990) have shown that it is not really necessary to redo the staging many times. Within the Metropolis sampling, it is possible to have a single secondary path for each link and to make a trial secondary move along with the primary move. The multilevel bisection method is a more general and powerful way of setting up a partitioning of path variables.

I. 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. There is no internal indication that permutational equilibrium has not been attained. Of course, at high temperature the rate should be small, since the atoms do not want to exchange. However, once the degeneracy temperature is reached, the atoms should begin to exchange spontaneously. Lack of exchange in an algorithm is occasionally taken as evidence that exchange is not important. To reach this conclusion, one must verify that the types of permutational moves allow the random walk to fully explore permutational space. In this subsection we want to show how this problem arises.

The simplest type of move in permutation space would

be to insert a pair permutation of two atoms without moving the coordinates of the path. We shall call this a *direct permutational move*, a move in permutation space without any relaxation of the path. This type of move can fail to sample the permutation space for a system with hard-core interactions like helium.

Permutation space consists of $N!$ distinct pockets of probability density, each corresponding to a given permutation. To estimate the rate at which the random walk moves from one pocket to another, let us use the transition-state theory of a chemical reaction. The transition state is the point midway between two regions of high probability, the configuration at the top of a potential barrier. The rate at which a system will go from one state to the other depends on the product of the transition-state probability and the attempt frequency. In our case, the transition state for a direct permutational move will be the most favorable arrangement of the path variables midway between two permutations.

The transition state for a path of two helium atoms (in relative coordinates) is shown in Fig. 45. In this state it is equally probable for the path to close to the right (the identity permutation) or to the left (a pair exchange). In the end-point approximation the transition-state probability if the two atoms are a distance r apart is given by

$$T(I \to P_{12}) = T(P_{12} \to I)$$

$$\propto r^2 dr \exp\left[-2[u(r;\tau) + u(r;\beta - \tau)] - \frac{r^2}{4\lambda\tau} \right]. \tag{5.38}$$
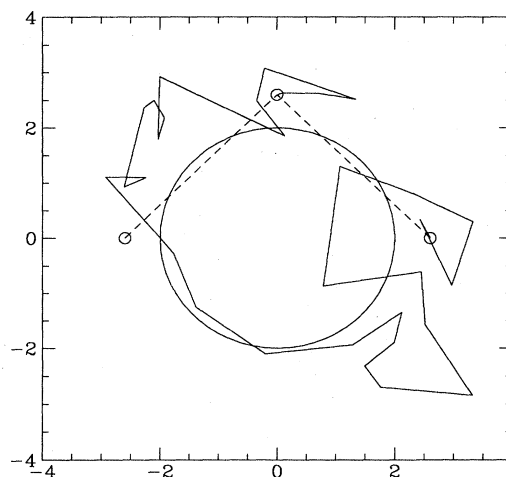
FIG. 45. The transition-state path for pair exchange. For this path the transition rate is maximal going from the direct to the exchange path. The two ways of closing the path, shown as dashed lines, are equally likely. The total path has a temperature of 1.21 K, each link a time step of 0.025 K^{-1}. These are three-dimensional paths projected onto the X-Y plane. That explains why some of the paths appear to go into the excluded region of less than 2 Å.

This rate is shown in Fig. 46 for various values of τ as a function of r. To find the transition state, we maximize T with respect to r. For a reasonable time step of $\tau = 0.025$ K^{-1}, the rate is maximized at 2×10^{-5} when two atoms are a distance 2.75 Å apart and are halfway exchanged.

As can be seen, transition probability depends crucially on the time step. For time steps of 0.1 K it is on the order of 0.1, but it rapidly becomes much smaller. If the time step is small, one will never see any permutational changes with "direct" moves. But we must keep τ from getting too large, otherwise the action will be inaccurate. The length of path that must be moved with the permutation is set by the degeneracy temperature, while the time step of the simulation is determined by the strength of the interaction potential. There is no reason for these two time steps to be the same. The hard-core nature of the He-He interaction is crucial here. It causes the transition rate to have a maximum. If it were possible for atoms to get closer together than 2 Å, two atoms could overlap and thereby exchange. The rate for direct permutation moves would be much larger. However, note that in addition to the "potential" barrier, the rate is also limited by the phase-space factor $r^2 dr$.

A permutation involving three or more atoms will have a smaller "potential" barrier for a direct transition, since three atoms can exchange without going around each other. But one will still find that the transition rate for direct moves is very small because the phase space for direct moves is small.

There is an even more serious problem of ergodicity: to change the winding number requires a global path move. Local moves, i.e., those confined in a small region of the simulation cell, cannot possibly change the winding number. One needs to make a permutation move

that stretches across the cell. For three-dimensional systems, this must involve a cyclic exchange of at least $N^{1/3}$ atoms.

J. Permutation sampling

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 $T^*(\mathcal{P}) \propto \rho(\mathcal{R}_i, \mathcal{P}\mathcal{R}_{i+m})$, 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,

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

(5.39)

where C_I is a normalization factor, Eq. (5.12), 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,

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

(5.40)

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

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

(5.41)

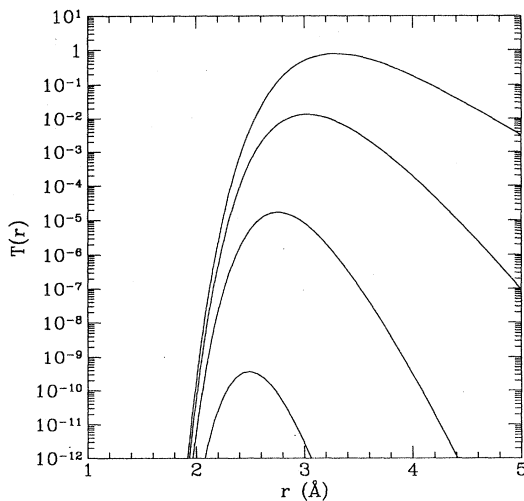It is best to put in the table only permutations that have

FIG. 46. Log of the probability of getting a pair exchange accepted vs the distance between the two particles for four different time steps (from the bottom 0.0125, 0.025, 0.05, and 0.10 K^{-1}).

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.

There are two disadvantages with this explicit construction of the permutation table. First, there is a considerable overhead in building the table, particularly as the size of the permutation move grows. Second, because the normalization constant for the transition probability depends on the initial state, there must be a final test to see if the move will be accepted based on $C_I/C_{\mathcal{P}}$ [see Eq. (5.13)]. To calculate this ratio, the permutation table must be built for the neighborhood of $\mathcal{P}$. Updating the old permutation table to reflect a change in neighborhood is rather involved. The acceptance ratio for the path variables is usually very small, so that the time constructing the permutation table is not dominant, but nonetheless it would be better to avoid this explicit way of selecting the permutation change.

An alternative way of constructing a permutation table 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:

$$A = \min \left[1, \frac{\frac{h_{k_1}}{t_{k_1,k_1}} + \cdots + \frac{h_{k_l}}{t_{k_l,k_l}}}{\frac{h_{k_1}}{t_{k_1,k_2}} + \cdots + \frac{h_{k_l}}{t_{k_l,k_1}}} \right]. \tag{5.42}$$

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.

Now consider the number of atoms involved in a permutation change. It is important to go beyond pair interchanges in a dense liquid, since it is much easier for three or four atoms to permute cyclically than it is for

two atoms. The starting place for the Monte Carlo random walk is typically the identity permutation. We have found that it is particularly difficult to get pair interchanges accepted from the identity permutation. Once triple and quadruple exchanges cause the state of the system to have a number of long exchange loops, pair interchanges can efficiently add and subtract from these longer exchanges. Thus rejection of pair exchanges at the beginning of a calculation is a poor indication of whether or not exchange of particles is an important physical effect.

K. The direct computation of the density matrix

Up to this point, we have been concerned only with how to construct a Markov process, which can then be used to evaluate ratios of integrals over the density matrix. In Sec. IV we explained why it is sometimes useful to determine the precise value of the density matrix for fixed end points. See, for example, the calculation of χ in Table III. We shall encounter another motivation when we introduce the method of estimating the exchange frequencies in quantum crystals. In this subsection we explain how to apply the ideas already developed plus a few new ones to a direct calculation of the many-body density matrix.

The basic idea is quite simple. One samples many paths with fixed end points (R_0 and R_M) and finds the average of the exponentiated action. Suppose we sample a path R_k ($0 < k < M$) from some probability distribution $P(R_1, R_2, \ldots, R_{M-1})$. Then from Eq. (2.21) the density matrix is the average

$$\rho(R_0, R_M; \beta) = \left\langle \frac{\exp[-\sum_{m=1}^M S_A^m]}{P(R_1, R_2, \ldots, R_{M-1})} \right\rangle_P, \tag{5.43}$$

where S_A^m is some approximate link action. The simplest way is to do free-particle sampling, that is, directly sample the free-particle kinetic action. If we also use the primitive action, then this equation is the Feynman-Kacs formula, Eq. (4.1).

The basic difficulty with calculating the density matrix is that we are averaging the exponential of something whose fluctuations grow with the number of particles N and with the inverse temperature β (or M). Physically, the larger the system or the lower the temperature, the larger the chance that somewhere along the path there will be an overlap and the estimator will vanish. The contribution to the integrand will be dominated by rarer and rarer paths that do not have any overlaps. It is because of such problems that one uses Markov chain methods, where one always works with an acceptable path. Markov sampling is the preferred way of calculating expectation values, that is, traces over the density matrix divided by the partition function. The density matrix can be caculated directly, but one must always be careful the estimator is not dominated by a few rare events.

Most of the tricks that we have developed so far can

be used to optimize this evaluation. There are also two new tricks that we shall describe: pruning and multiple sampling.

1. Bisection

Use of the bisection procedure (or the Lévy construction) can help organize the calculation. One samples the midpoint of the path and computes the ratio of the level action to the sampling probability. That average is the estimate of the density matrix for that level. Then, one bisects each of the two subintervals and determines the density matrix in the next level, etc. In this way one determines the dependence of the estimate of the density matrix on τ in a correlated way. As we shall see, one can also sample multiple times and branches within the bisection method. We will call each successive bisection a "level." The first level is the $R_{M/2}$, the second level is $(R_{M/4}, R_{3M/4})$, and so forth.

2. Level action

One can use either the primitive action to estimate the density matrix or any of the improved versions that we discussed in the previous section. The accuracy of the action will control how many levels one needs to obtain convergence to the exact density matrix. The systematic error of the estimate in a given level is controlled by the action one uses in that level alone. Shown in Fig. 36 are estimates using the primitive action and using the pair action. We shall see another reason for using good actions in a moment; they are helpful in predicting whether one should continue a given bisection.

3. Sampling

Normally in direct computations of the action one uses free-particle sampling. Sometimes one will get better results by attempting to optimize single-slice sampling, as was explained in Sec. V.C. However, we must be very careful about undersampling. By undersampling, we mean sampling regions of path space where both the integrand, and P are getting small but where the sampling probability is decreasing faster than the integrand, so that the ratio in Eq. (5.43) diverges. It is quite easy for this to happen. In the worst case, the variance of the estimator does not exist, so the central-limit theorem will not apply. One should always verify analytically that the variance exists with any proposed sampling. Markov sampling does not have this particular problem. That is one of its major advantages. If one undersamples, that merely slows the convergence. It is always best to begin testing with free-particle sampling, since the variance of the density-matrix estimate for free-particle sampling will always exist (assuming the energy spectrum is bounded below).

Now we can state the *zero-variance* principle of esti-

mating the density matrix. If one uses optimal sampling and the exact level action, each sample of the density matrix at each level will give the exact result. Of course, this is not practical, since it would require an analytic form for the exact density matrix at each time step. It is a guide: as one "bootstraps" one's approximations for better actions and better samplings, the variance will be reduced.

4. Pruning

Suppose a given sampled path has an estimate x_l of the density matrix at level l and suppose that x_0 is our best estimate of the density matrix. If x_l is less than x_0, with probability $p = x_l/x_0$, we continue the process of going to finer levels. One saves computer time by not computing the fine details of paths that contribute little to the final average because they already contain an overlap. This is analogous to the rejection procedure we did in multilevel sampling. The probability for the final estimate then includes the additional factor p. Here again we see how accurate level actions can help. (The level actions enter into x_l.) If you discard paths based on bad estimates of their "future" contributions, you can expect to increase the overall variance. If x_l/x_0 is larger than unity, one can use it to make multiple copies; we discuss that next.

5. Multiple sampling

Up to now we have been applying ideas already developed for Markov random walks. Multiple sampling, in which we sample d possible next level points instead of only one, was not discussed in the context of Markov chains. With it we take advantage of two features of path integrals. First, fluctuations in neighboring slices are completely independent, because we take a product of the action to get the total density matrix. The product of terms will fluctuate more than an individual term. If we do some averaging first, we can get rid of most of the nonlinear terms in the variance of the product. Another advantage of multiple sampling is that it takes less than d times as long to sample (and compute the action) for d points, since some of the computation is shared between the samples.

The following illustrates the basic idea of multiple sampling. Suppose we have already sampled the first level and obtained a single sampled midpoint. Now proceeding onto the next level, we need to sample the left subinterval, called "A," and the right subinterval, called "B." Instead of sampling a single A and B, suppose we sample d points of A and d points of B, getting estimates a_i and b_i with $1 \leq i \leq d$. Define the average value on the right and left as a and b (assuming the midpoint is fixed).

Now since the left and right samples are independent of each other, we can show that the best estimator for the product ab is the product of the averages on the left

and right; that is, $d^{-2} \sum_{ij} a_i b_j$. The variance of this estimator (for a given midpoint) is $v_0 + d^{-1}(a^2 v_b + b^2 v_a) + d^{-2} v_a v_b$, where v_a and v_b are the variances of a_i and b_i and v_0 comes from the fact the midpoint itself is sampled; it is $v_0 = \langle a^2 b^2 \rangle - \langle ab \rangle^2$. One can see that the various terms contributing to the variances depend differently on the number of sampled points d. We cannot do much about the first two terms except sample longer, as they represent an inherent variance, but the third term comes from the product of fluctuations. It is nonlinear and it can be reduced by multiple sampling. It will be important to do this if the variance divided by the average value is greater than one, $v_a/a^2 > 1$.

As an illustration, consider the hard-sphere model. The primitive action vanishes if no spheres overlap, otherwise it is infinite. If we use free-particle sampling, then the density-matrix estimator is 1 in the case of no overlap and 0 for an overlap. Let the probability of obtaining a nonoverlapping path in the subinterval A or B be p where $0 < p < 1$. We find that $a = p$ and $v_a = p(1-p)$ and thus the variance of the density-matrix estimator for d samples is $v_0 + p^2(1-p)[2pd^{-1} + (1-p)d^{-2}]$. For small p, the optimal number of sampled points is roughly $d \approx p^{-1}$. We need enough sampled points on the left and the right so that there will be a good chance of having at least one good point on the left and one good point on the right.

We shall only sketch how to proceed to the next level. After the estimator is formed, the points to the left and right are pruned. Then one does multiple sampling on the surviving left and right intervals depending on the individual estimator. All of this information is folded back into the final estimator for the density matrix in a recursive fashion. This multiple sampling trick has not been implemented in Markov sampling, though it would seem to be useful there as well.

For the direct density-matrix calculation it is important to use bisection with very good level actions. Correlated sampling can be dangerous but can also help. It is possible to compute the exact actions with a precision of 0.1 % on a system containing 54 ^{4}He atoms with only 30 000 sampled points. These results were shown in Table III. Pruning and multiple sampling can considerably reduce the variance.

VI. CALCULATING PROPERTIES

Once the action is chosen and sampling is accomplished, we are ready to calculate expectation values. In this section we discuss some of the technical details of constructing estimators for various physical quantities. What we need to do is to express a given quantum expectation of the density matrix as an average over a path.

Properties can often be computed in different ways. A specific formula used to calculate some physical quantity is called an *estimator*. Each estimator is characterized by its *statistical error*, *efficiency* (statistical error for a given length run), *bias* (nonlinear distortion), *time-step error*, and *finite-size error*. In addition, some estimators

are easier physically to interpret or easier to program. These various criteria for what constitutes a good estimator make the choice rather subtle. Generally one wants an estimator that minimizes the maximum of the various errors. There has been very little systematic investigation of these issues for the PIMC method.

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

$$\rho(\mathbf{r}) = \frac{1}{M} \sum_{i,t} \langle \delta(\mathbf{r} - \mathbf{r}_{it}) \rangle , \qquad (6.1)$$

where $\mathbf{r}_{it}$ is the coordinate of particle i at "time" t. To simplify the notation, in this section *we shall not symmetrize over "time"*; we leave that to the reader. We shall use $\langle \cdots \rangle$ to indicate an average over the paths *and* over links t.

The density can also be computed as the functional derivative of the free energy with respect to an external potential. See Takahashi (1986) or Schweizer et al. (1981). This definition has the advantage that the estimator will be consistent with the free energy, so thermodynamic identities will hold between the estimators for any time step instead of only in the limit of zero time step. This might lead to smaller systematic errors, although that has not been demonstrated. The statistical errors for the two estimators should be roughly the same. The thermodynamic way of estimating the density is more complicated and less intuitive than directly applying the density operator to the paths.

Other interesting properties, such as the energy, the free energy, the momentum distribution, and the superfluid density, are not as straightforward to calculate. We discuss how to calculate these other properties below.

A. Energy

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. The situation in variational and Green's-function Monte Carlo and in classical simulations is quite different. In these methods it is obvious how to calculate the energy. For classical Monte Carlo simulation, the temperature is fixed and the kinetic energy is always $k_B T/2$ per degree of freedom. Hence estimating the energy reduces to estimating the potential energy. The statistical error of the potential is then proportional to the root-mean-squared fluctuation of the potential energy, which is related to the specific heat. The principle way to reduce the statistical error is to minimize the autocorrelation of the potential as the random walk proceeds.

On the other hand, in ground-state methods, the sta-

tistical error is proportional to the root-mean-squared fluctuations of the residual (or local) energy of the trial wave function. The statistical error has the zero-variance principle (Ceperley and Kalos, 1979): as the trial wave function approaches an exact eigenstate, the statistical error vanishes. One can control the efficiency by optimizing the trial wave function. At zero temperature there are alternative ways of computing the energy. Since the kinetic energy is a partial derivative, one can use Green's theorem to find different estimators. However, the zero-variance principle applies only to the "direct" estimator.

PIMC is in between the classical and zero-temperature situation. One of the advantages of PIMC simulation is that there is no trial function, but one pays the price in that the variance is more difficult to control. Again, the kinetic energy can be cast in different forms by using Green's theorem. With PIMC it is not clear in general which is the best form. One form is best for small time steps, another for large time steps.

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 will 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

$$D(t) = \left\langle (\mathbf{r}_i(t) - \mathbf{r}_i(0))^2 \right\rangle. \qquad (6.2)$$

The function $D(t)$ for superfluid helium is shown in Fig. 28. Expanding the diffusion distance in a power series about $t = 0$, one can show

$$D(t) = 2\lambda t[3 - 2\mathcal{K}_D t + \mathcal{O}(t^2) \cdots]. \qquad (6.3)$$

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. We will come to this shortly.

The *direct* or *Hamiltonian* energy estimator is defined as

$$E_H = \text{Tr}(\mathcal{H}e^{-\beta\mathcal{H}})/\text{Tr}(e^{-\beta\mathcal{H}}). \qquad (6.4)$$

One interprets the ratio as a path average and then explicitly applies $\mathcal{H}$ to a given link. Clearly it does not matter which link, since $\mathcal{H}$ commutes with $e^{-\tau\mathcal{H}}$. This gives

$$E_H = \left\langle \frac{3N}{2\tau} - \frac{(R_i - R_{i-1})^2}{4\lambda\tau^2} + V(R_i) \right. \qquad (6.5)$$
$$\left. - \frac{(R_i - R_{i-1})\nabla_i U}{\tau} + \lambda\nabla_i^2 U^i - \lambda[\nabla_i U^i]^2 \right\rangle.$$

We remind the reader that U^i is the total action for link i, not including the free-particle kinetic action. We have ordered the terms in increasing powers of τ. Notice that the first two terms diverge as $\tau \to 0$! This is the same fluctuation problem we alluded to with the diffusion estimator of the kinetic energy. This estimator is not often used because one has to compute the first and second spatial derivatives of U. This is not a serious problem, but it can be avoided by changing to a "time" derivative. Again, because all the links are identical, one can average over i.

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

$$E_T = -\frac{1}{Z}\frac{dZ}{d\beta}. \qquad (6.6)$$

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

$$E_T = \left\langle \frac{3N}{2\tau} - \frac{(R_i - R_{i-1})^2}{4\lambda\tau^2} + \frac{dU^i}{d\tau} \right\rangle. \qquad (6.7)$$

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 contains a kinetic contribution.

To see this, let us define the *thermodynamic* estimator of the kinetic energy as the mass derivative of the partition function,

$$\mathcal{K}_T = -\frac{m}{\beta Z}\frac{dZ}{dm}. \qquad (6.8)$$

Applying this to link i alone, we get

$$\mathcal{K}_T = \left\langle \frac{3N}{2\tau} - \frac{(R_i - R_{i-1})^2}{4\lambda\tau^2} + \frac{\lambda}{\tau}\frac{dU^i}{d\lambda} \right\rangle. \qquad (6.9)$$

Then the thermodynamic estimator for the potential energy is

$$\mathcal{V}_T = \left\langle \frac{dU^i}{d\tau} - \frac{\lambda}{\tau}\frac{dU^i}{d\lambda} \right\rangle. \qquad (6.10)$$

Using the pair action, these various expressions are fairly straightforward to compute, since the derivatives are with respect to τ and λ. This simply means that two additional tables need to be constructed for the derivatives.

There is clearly a close relationship between the direct and thermodynamic estimators of the energy: they share the first two terms. But the systematic, time-step errors of these two estimators can be different. Runge and Chester (1988) found the difference between these two estimators, $E_H - E_T$, to be very useful in testing the convergence of the energy with respect to τ. Using the same paths to find both energies, one may estimate the systematic difference between the two with a much

smaller statistical error than an estimate of either energy by itself.

The variance of these two estimators, E_H and E_T, will be equal in the limit of accurate action. This is because an accurate action will satisfy the Bloch equation, so that E_H and E_T will be equal for any particular path sampled. Both estimators have in common that their 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. In a moment we shall look at an estimator without this problem.

On the other hand, in the large-τ limit, the variance of these two estimators vanishes. This follows because at low temperature the action becomes the exact ground-state wave function and so is an eigenfunction of $\mathcal{H}$. It is not obvious a priori whether, for time steps actually in use for liquid helium, one is closer to the small-τ or the large-τ limit.

We call E_T the thermodynamical estimator because it can be integrated to get the free energy,

$$\beta_1 F_1 - \beta_0 F_0 = \int_{\beta_1}^{\beta_0} d\beta E_D, \qquad (6.11)$$

where the free energy is defined as $e^{-\beta F} = Z$. For this to be identically true, E_T has to be computed with constant M, not constant τ as is usually done.

When we take the β derivative, the order of an approximation to the action will be reduced by one, so if the approximate action has an error of order τ^3, the estimator for the energy is correct only to order τ^2. It is for this reason the kinetic energy converges more slowly with respect to the time step than other properties. This is shown in Fig. 47. The largest time-step errors of the energy estimator appear to be fluctuations that average to zero. However, the first contribution beyond the pair action, the polarization action, is a systematic shift that is often important to include.

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 (Herman, Bruskin, and Berne, 1982). One ends up with an estimator similar to the virial expression for the pressure.

The *virial* energy estimator is

$$E_V = \left\langle \frac{3N}{2L\tau} - \frac{1}{4L\tau^2\lambda}(R_{L+i} - R_i)(R_{i+1} - R_i) \right. $$
$$\left. - \frac{1}{2}F^i\Delta_i + \frac{dU^i}{d\tau} \right\rangle \qquad (6.12)$$

where F_i is a generalization of the classical force,

$$F^i = -\frac{1}{\tau}\nabla_i(U^{i-1} + U^i), \qquad (6.13)$$

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

$$\Delta_i = \frac{1}{2L}\sum_{j=-L+1}^{L-1}(R_i - R_{i+j}). \qquad (6.14)$$

The derivation of this estimator is given in Appendix A. Care must be taken in interpreting terms like $(R_i - R_j)$ to ensure that the atoms always have continuous trajectories in the presence of periodic boundary conditions and exchange. 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
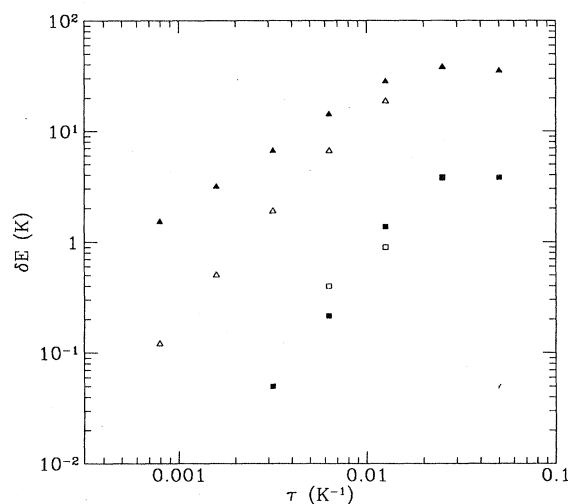

FIG. 47. The error of the estimated energy for an equilateral triangle of atoms, separated by 2.88 Å, as computed with the primitive action (triangles) and with the pair action (squares). The error in the action for this geometry and these time steps is shown in Fig. 35. The filled symbols represent the error in the energy as a function of τ. The open symbols are the error in the action at $\beta = 0.025$ if intermediate time slices are inserted (with time step τ) and integrated over. This integrated energy is worse, presumably because the path can wander into regions closer than 2.88 Å, where the errors are worse. The primitive energy converges as τ^{-2} when it is integrated over, but only as τ^{-1} point-by-point.

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.

The systematic error of E_V is exactly that of E_T and does not depend on L. However, the statistical error will depend on L. By averaging over L links, one reduces the magnitude, and hence the statistical error, of the first two terms in the thermodynamic estimator by a factor L. Shown in Fig. 48 is the energy and its error bars as a function of window size. By windowing over a few time slices we manage to reduce the statistical error by almost 50%. Further increases in window size have little effect. One can see that the value of the energy is independent of L. This is a check that the paths have come into local equilibrium and that the actions and their derivatives are consistent with each other. For example, if the action contains off-diagonal terms, it is rather important to include the gradients of these terms in F^i.

There was a suggestion (Giansanti and Jacucci, 1988) that E_V is much more sluggish in becoming decorrelated in a random walk than is E_T, thus reducing overall efficiency of the virial estimator. Whether this is true or not seems to depend sensitively on the physical system and on the transition rules (Cao and Berne, 1989). It is probably best to compute the energy and its errors (including the effect of autocorrelation) for a range of window sizes and then, after the calculation is over, choose the window size that minimizes the statistical error.

The virial estimator provides a more rigorous justification for the Feynman expression for the energy, Eq. (3.4), if we take $L = M$. The only term that depends explicitly on bosonic exchange is the second term; the rest is just a background kinetic and potential energy. We can approximately write
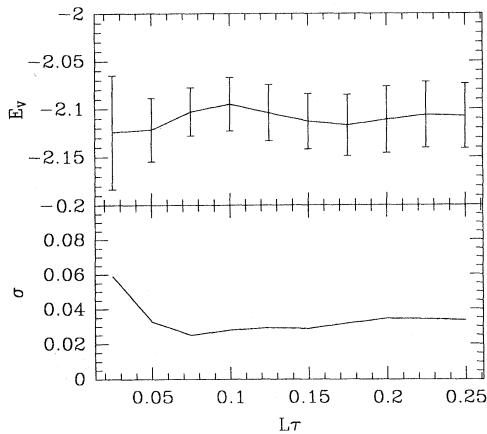

FIG. 48. The energy and the variance of the virial estimator as a function of window size for liquid ^{4}He at 4 K. The error of the leftmost point is equivalent to the direct estimator. The most efficient estimator has error bars roughly 50% lower.

$$E_V = \text{background} - \left\langle \sum_{k=1}^{N} \frac{(\mathbf{r}_{k,i} - \mathbf{r}_{P_k,i})^2}{4\beta^2\lambda} \right\rangle, \quad (6.15)$$

where P is the permutation. The contribution shown is the kinetic energy involved in Bose condensation; it is proportional to the number of exchanges.

There is an estimator related to the virial estimator, the *stretch estimator* of Ceperley and Pollock (1984). It was constructed with the idea of implementing a zero-variance principle. In practice, it has not led to lower variances.

B. Specific heat and pressure

Now we wish to make some brief remarks on two other related quantities, the specific heat and the pressure. The remarks are brief because there is little literature on the subject.

The pressure is the volume derivative of the free energy,

$$P = - \left. \frac{dF}{d\Omega} \right|_{\beta}. \quad (6.16)$$

Expanding now as a path average we get the thermodynamic estimator for the pressure,

$$P_T = \frac{1}{3\tau\Omega} \left\langle 3N - \frac{(R_i - R_{i-1})^2}{2\lambda\tau} + 2R_i\nabla_i U \right\rangle. \quad (6.17)$$

One can also calculate the pressure using the virial theorem. For a pair potential, the formula is

$$P_V = \frac{1}{3\Omega} \left\langle 2\mathcal{K} - \sum_{i<j} r_{ij} \frac{dv(r_{ij})}{dr} \right\rangle. \quad (6.18)$$

This will have a different systematic and statistical error. The errors will depend on how the kinetic energy is calculated. The variance of the thermodynamic pressure can be reduced using the results of Appendix A to get a virial estimator that has the systematic error of the thermodynamic estimator.

The specific heat at constant volume is defined as the derivative of the energy with respect to the temperature,

$$C_v = -\beta^2 \left. \frac{dE}{d\beta} \right|_{\Omega} \quad (6.19)$$

$$= \beta^2 \left\langle (\mathcal{H} - E)^2 \right\rangle, \quad (6.20)$$

where $\langle \cdots \rangle$ is a path average and E is the average energy. The second equation is the fluctuation formula for the specific heat. Because the operator $(\mathcal{H} - E)$ commutes with the density matrix, one can apply the two factors to two different links of the path. One can also replace $\mathcal{H}$ with the τ derivative, as was done with the energy. Hence one can write the fluctuation formula as

$$C_v = \beta^2 \left\langle (E_{i,i+L} - E)(E_{j,j+L} - E) \right\rangle, \quad (6.21)$$

where $E_{i,j}$ is an energy estimator obtained by differenti-

ating slices i through j only,

$$E_{i,j} = -\frac{d\ln\left(\rho(R_i, R_j; (j-i)\tau\right)}{d\tau}. \qquad (6.22)$$

The direct and virial estimators for $E_{i,j}$ are given in Appendix A. The links i and j and the window size L are arbitrary as long as they do not overlap. The estimated specific heat will be independent of the choice of these parameters; that is another good test of the correctness and convergence of the result. Remember that imaginary time is periodic, so that links are numbered modulo M, the number of time slices.

If the intervals of the two windows do overlap, one has additional terms for the specific-heat estimator, since then the operators $(\mathcal{H} - E)$ will act twice on the same link. Marx, Nielaba, and Binder (1992) discuss a total specific-heat estimator for $L = 1$, in which one sums i and j over all possible values including the overlapping cases $i = j$. One might expect that overlapping terms would have a larger variance, since they come from applying the Hamiltonian (or equivalently the τ derivative) to the same link, and the resulting estimator is more singular. One could argue that a low variance estimator would be obtained by choosing the virial window size to minimize the variance of the energy and by avoiding overlapping windows. These estimators have not been applied to helium.

What has been done in practice to calculate the specific heat is to perform a sequence of calculations spanning a range of temperatures. The energies are then fit to an appropriate function, e.g., a Padé function. The specific heat is then obtained by analytically differentiating. Ceperley and Pollock (1989) followed this procedure to determine the specific heat of helium films in the region of the superfluid phase transition. The fitting can be constrained to include known high- and low-temperature behavior. However, this brute force method requires simulations at many temperatures and a careful fit to an appropriate function. Clearly systematic errors can be introduced. The specific heat obtained in this way cannot be easily used to determine detailed properties of a phase transition, since one has to assume a given functional form of the fit.

C. Momentum distribution

To determine the momentum distribution and the condensate fraction, one calculates the single-particle density matrix and then takes its Fourier transform. See Sec. III.D. Two complementary algorithms have been used (Ceperley and Pollock, 1987) to calculate the single-particle density matrix, $n(\mathbf{r}_1, \mathbf{r}_1')$.

The first method is an imitation of what one does at zero temperature (McMillan, 1965). During a normal simulation (that is, diagonal, all closed polymers) one occasionally stops and displaces each bead, making a path momentarily open. Then $n(r)$ is the average ratio of the displaced to the undisplaced density matrix. Suppose

atom 1 in time slice 1 is chosen to be cut. We sample a new coordinate $\mathbf{r}_{11}'$ from the free-particle distribution $P(\mathbf{r}_{11}') = \rho_0(\mathbf{r}_{11}', \mathbf{r}_{12}; \tau)$. Note that the new coordinate is connected to atom 1 in the second time slice. The permutation and the coordinates of the other beads, the rest of the path, are held fixed. Then an estimator for the single-particle density matrix is

$$n(\mathbf{r}_{11}, \mathbf{r}_{11}') = \left\langle \frac{\rho(R_1 \leftarrow \mathbf{r}_{11}', R_2)}{\rho(R_1, R_2)P(\mathbf{r}_{11}')} \right\rangle \qquad (6.23)$$

$$= \left\langle \exp\left[\frac{(\mathbf{r}_{11} - \mathbf{r}_{12})^2}{4\lambda\tau} \right.\right. \qquad (6.24)$$

$$\left.\left. -U(R_1 \leftarrow \mathbf{r}_{11}', R_2) + U(R_1, R_2) \right] \right\rangle. \qquad (6.25)$$

The arrow $\leftarrow$ means that a single atomic coordinate is replaced. One sees a possible problem that does not occur in ground-state applications. Springs that were originally stretched out (with large values of $|\mathbf{r}_{11} - \mathbf{r}_{12}|$) contribute exponentially more to the estimator. To reduce the variance one should interpret $\exp[\frac{(\mathbf{r}_{11} - \mathbf{r}_{12})^2}{4\lambda\tau}]$ as the number of samples of $\mathbf{r}_{11}'$ to be taken for that link. One samples more when the expected contribution is larger. One can improve the method still further by displacing both of the cut ends, that is, also sampling $\mathbf{r}_{11}''$, which will connect to $\mathbf{r}_{10}$.

This displacement method is accurate for computing $n(r)$ for r on the order of the thermal wavelength Λ_τ. It is an appropriate method for nearly classical systems and can be carried out simultaneously with computing diagonal properties. One can use the symmetry with respect to particles and time slices. However, for larger displacements from the diagonal, which are needed to compute the condensate fraction, the statistical error grows rapidly because the major contributions come from different arrangements of the neighboring atoms and different permutations. To get reliable results for r greater than a thermal wavelength requires a collective displacement of an entire chain.

In the second method, one atom is allowed to be off the diagonal during the entire random walk. In the polymer language, the simulation is of $N - 1$ ring polymers and 1 linear or open polymer. This is a direct interpretation of the defining equation for the single-particle density matrix, Eq. (3.10). An additional variable, namely, $\mathbf{r}_{11}'$, is introduced into the Monte Carlo simulation. During the random walk this additional variable is moved and accepted or rejected as are the other path and permutation variables. At each step of the random walk, the distance $\mathbf{r}_{11} - \mathbf{r}_{11}'$ is added to a histogram. The final histogram of occurrences of $\mathbf{r}_{11} - \mathbf{r}_{11}'$ is proportional to $n(\mathbf{r}_{11}, \mathbf{r}_{11}')$. In a superfluid the two ends can become well separated because of the addition of other atoms' paths in the middle of the linear polymer. The fraction of atoms having exactly zero momentum (the condensate fraction) is the value of the properly normalized end-to-end distribution at large r.

The difficulty with this second method is that a completely new simulation needs to be done to obtain the momentum distribution. Another difficulty is that only one distance out of the $3NM$ path variables contributes to the single-particle density matrix, so that the efficiency is low. Further, the computer code becomes more complicated, since one atom is now treated differently from the others. However, one can argue that the condensate fraction, which is the superfluid order parameter, is as important a quantity to calculate as the energy, so the extra effort is well justified.

In fact, it is worthwhile to modify the method so as to achieve lower statistical errors on the end-to-end distribution. One easy modification is to add to the action an artificial potential between the two ends, equal to $\ln[n_a(r)r^2]$, so that the end-to-end distance will spend roughly the same amount of time at large and small distances. By Eq. (3.15), the condensate fraction is the ratio of the time the polymer is stretched out to the time it is contracted. Here $n_a(r)$ is any convenient approximation to $n(\mathbf{r})$, for example, $n_0 + (1-n_0)\exp(-cr^2)$ with n_0 a rough guess of the condensate fraction and c set by the kinetic energy. At the end of the calculation, the effect of this importance sampling is divided out of the end-to-end distribution. It serves only to reduce the variance. Finally, one adjusts the normalization of the end-to-end distribution so that $n(0) = 1$. Using the above importance function, this will be easier, since one has controlled how often the two ends are close together. Another way to improve the efficiency of the calculation is to move and permute the disconnected atom more frequently than the other atoms. The atoms not in the immediate neighborhood of the cut ends act as a sluggish background. In this way the efficiency does not decrease so rapidly as more atoms are added to the system.

D. Superfluid density

The superfluid density can be calculated in several different ways. The most elegant approach is to use the relation between the superfluid density and the mean-squared winding number, Eq. (3.31). The winding number $\mathbf{W}$ is the imaginary-time current of particles, defined in Eq. (3.32). Accurate computation of the mean-squared winding number is difficult, since a change in $\mathbf{W}$ involves a global move of the paths. In the $\tau \to 0$ limit, the paths are continuous directed loops, so the winding number equals the flux of paths through any given plane. This implies that to change the winding number in the $\hat{x}$ direction, one must change the flux in all planes orthogonal to the $\hat{x}$ axis. Pieces of paths spanning the entire $\hat{x}$ must change. Hence the number of atoms needed to make a winding number change will, at least, be proportional to the length of the periodic cell. If the simulation is of N atoms in a cube, we should expect to move simultaneously $N^{1/3}$ of them in order to see a winding number change.

In fact, we find that the multiparticle bisection algorithm discussed in the last section will change the winding number very infrequently if only four particle exchanges are included for a system with $N > 100$. Figure 49 shows the instantaneous values of the superfluid density for 100 blocks of a simulation of a 64-atom ^{4}He system at a temperature near the superfluid transition. One can see the system switching back and forth between a superfluid state and a nonsuperfluid state every several blocks. Each of these blocks consisted of about 50 000 attempted paths moves, 300 successful permutation moves, and 50 successful winding number changes, and took 1/2 hour on a SPARC-10 work station. One can improve the efficiency by preferentially making moves that change the winding number. To do this, one enhances the probability for moves that change the winding number during the construction of the permutation table. See Sec. V.J.

There are several ways around this problem. As a first example, one can cut one of the polymers, i.e., allow one atom to be off the diagonal, as was done in the calculation of the single-particle density matrix. The winding number is still defined in Eq. (3.32). If there is a cut end, the winding number is no longer quantized to be a multiple of the box length, so it can change with a sequence of local moves. But the superfluid density is defined on the diagonal. Suppose the two ends of atom 1 are at positions $\mathbf{r}_1$ and $\mathbf{r}_1'$. Define a generalized superfluid density by

$$\frac{\rho_s(a)}{\rho} = \frac{\langle W^2 \delta(\mathbf{r}_1 - \mathbf{r}_1' - \mathbf{a})\rangle}{2\lambda\beta N}. \tag{6.26}$$

[One applies the "minimum image condition" for periodic boundary conditions to the argument of $\delta(\cdots)$.] Then the superfluid density can be obtained by taking $a \to 0$. This is how one ties a knot. The construction uses a linear string and, at the end, one solders the ends to-
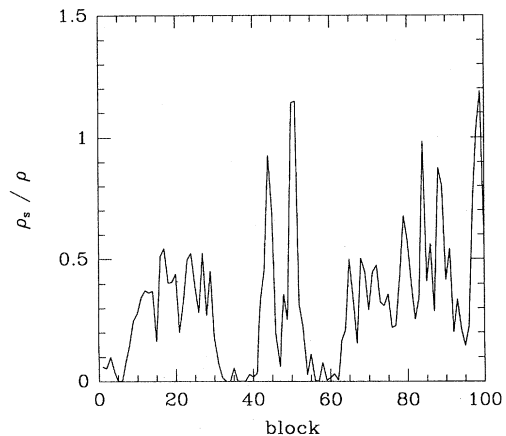

FIG. 49. The block averages of the superfluid density for 100 consecutive blocks for bulk ^{4}He at 2.22 K, very near the superfluid transition. The superfluid density is estimated at 0.32 ± 0.04 from this data. This is higher than the actual superfluid density because of finite-size effects.

gether. The advantage of this method is that it can be done while one is calculating the single-particle density matrix. One expects that the superfluid density will be a smooth function of a, so the extrapolation to zero will not introduce large errors. This method has not yet been applied to superfluid systems.

It is also possible to compute superfluid properties without changing the winding number. After all, there are experimental consequences of superfluidity even if the geometry is not simply connected. If one has a large enough system, the superfluid density can be obtained from the long-range properties of the momentum-momentum correlation function $G(\mathbf{r})$, as defined in Eq. (3.36) and described by Pollock and Ceperley (1987). This method has a larger statistical error and there is some ambiguity in how to determine the asymptotic value of the slowly decaying correlation function in a finite system.

It has been suggested that one can define a "frequency"-dependent winding number (Batrouni, Scalettar, and Zimanyi, 1990). (This is the frequency in imaginary time.) The physical superfluid density is obtained as the zero-"frequency" limit. This method has been used in world-line calculations of bosons on a lattice. Let us define the Fourier transform of the path "velocity" with respect to imaginary time,

$$\mathbf{W}_\omega = \sum_{j,l=1}^{M} (\mathbf{r}_{j,l+1} - \mathbf{r}_{j,l})e^{2\pi i t_l \omega}. \qquad (6.27)$$

Clearly the usual winding number is $\mathbf{W}_0$. Then a "frequency" dependent density is defined as

$$\frac{\rho_s(\omega)}{\rho} = \frac{\langle \mathbf{W}_\omega^2 \rangle}{2\lambda\beta N}. \qquad (6.28)$$

We can exactly calculate the above average in the classical limit. The velocities of the paths are free-particle like, and no windings can occur if the thermal wavelength is smaller than the size of the box. One can write $W_{\omega_k} = [\exp(2\pi i k/M) - 1]Q_k$, where Q_k is the normal-mode coordinate defined in Eq. (5.25). Then the average is trivial, since the action is decoupled in normal modes. The result is $\rho_s(\omega) = \rho(1 - \delta_{\omega,0})$. The new estimator is 100% superfluid at all frequencies except the physically relevant one. Clearly, in a continuum system, this method cannot be used to determine if a system is superfluid, since even a classical system would pass the test for any positive ω! It is mysterious why it should also work for a lattice Hubbard model at positive temperatures, since that model in the low-density limit should approach a continuum model. In using it one is implicitly assuming that the system is at zero temperature, with impurities to break translational symmetry. In Fig. 50 is shown the generalized density as computed for a helium film at $T = 1.17$ K and compared with the results at zero frequency. What is plotted is this "frequency"-dependent density, obtained from a simulation with zero winding number and the equilibrium distribution of winding num-
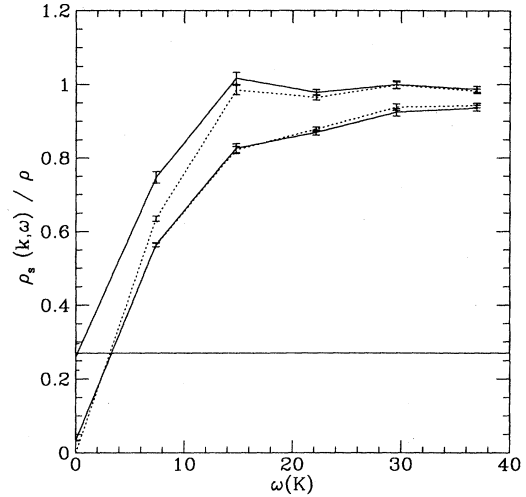

FIG. 50. The "frequency-" and wave-vector-dependent superfluid density for a 2D film of 25 helium atoms at 1.17 K as a function of "frequency." The "true" superfluid density is shown as the horizontal solid line. The upper two curves are for $k = 0$, the lower two curves are for $k = 0.26$ Å^{-1}, the lowest positive wave vector of the periodic box. The two solid lines are from a simulation in which the winding number equilibrated, thus that upper curve ($k = 0$) goes to the true superfluid density at $\omega = 0$. The dotted curves are from a simulation in which the moves constrained the paths to have zero winding number.

bers. These two simulations agree with each other at larger frequencies. It is seen that correlations do indeed reduce the value of the smallest positive frequency from the free-particle value. But it does not seem that there is any way of reliably extrapolating to the zero-frequency limit. Keep in mind that the lines shown in this figure only serve to connect the points. At any positive temperature (as opposed to zero temperature), this response function is only defined for "frequencies" that are integer multiples of $2\pi k_B T$.

Possibly if we Fourier transform with respect to space, defining $\mathbf{W}_{\mathbf{k},\omega} \sum_{j,l=1}^{M} (\mathbf{r}_{j,l+1} - \mathbf{r}_{j,l})e^{2\pi i(t_l\omega + \mathbf{k}\mathbf{r}_j)}$ and taking $\omega \to 0$ before $\mathbf{k} \to 0$ will allow more reliable extrapolations. Figure 50 shows that this extrapolation is also problematical. One ends up with a zero superfluid fraction, instead of a finite but small (0.26) fraction.

E. Exchange frequencies in quantum crystals

In this subsection we describe the algorithm that was used to calculate the exchange frequencies in solid helium referred to in Sec. III.H. Only one algorithm to determine this ratio has been investigated in detail. Remember that the exchange frequency is estimated from the ratio of the partition function of an exchanging path to one that is not exchanging; see Eq. (3.44). Let us call a path that does not exchange an "A" path, and one with a cyclic permutation a "B" path. The respective parti-

tion functions are denoted Z_A and Z_B. To determine the ratio $f_P = Z_B/Z_A$ is nontrivial because the Metropolis method we have been using cannot directly calculate the value of a partition function. In the classical simulation literature there are many specialized techniques for calculating such ratios, and we will adapt one of them.

Bennett (1976) described an optimized method of determining free-energy differences between two related classical systems, sometimes called the two-sided acceptance ratio method. Bennett considers the case in which the action in the two systems differs, while for the exchange frequency it is the topological connection of the paths which differs. As in Sec. V.I, we cannot go directly from system "A" to "B" or vice versa; we have to move path variables also. For this we generalize Bennett's method to allow for sampling the other system. As in the bisection method, we partition any path into two pieces, a fixed background R_0, with action S_0, and an active part of the path where a permutation will be either inserted or removed. The action of the active piece alone will be denoted S_A for an "A" path or S_B for a "B" path.

Now it is not difficult to see that the following integral over a weighting function $W(R_A, R_B, R_0)$ can be expressed either as an "A" average or a "B" average:

$$\int dR_0 dR_A dR_B e^{-S_0 - S_A - S_B} W(R_A, R_B, R_0)$$
$$= Z_A \left\langle \frac{W(R_A, R_B, R_0) e^{-S_B}}{T(R_B | R_0)} \right\rangle_A$$
$$= Z_B \left\langle \frac{W(R_A, R_B, R_0) e^{-S_A}}{T(R_A | R_0)} \right\rangle_B .$$

$$(6.29)$$

The "A" average is constructed with (R_0, R_A) sampled from the distribution $e^{-S_0 - S_A}/Z_A$, by the usual Metropolis method, and then occasionally the R_B path is sampled from $T(R_B | R_0)$. For the "B" average, the roles of "A" and "B" are reversed. We see that the partition function ratio is the ratio of an "A" process estimator to a "B" process estimator. Bennett determined the optimal weighting function, the one which minimizes the variance of the partition function ratio, and found

$$W(R_A, R_B, R_0) = \left[\frac{e^{-S_A}}{T(R_A | R_0)} + \frac{e^{-S_B - c}}{T(R_B | R_0)} \right]^{-1}, \quad (6.30)$$

where c is a constant chosen to minimize the variance. Bennett argues that the value $c = \ln[(N_A Z_B)/(N_B Z_A)]$ is optimal where N_A and N_B are the numbers of uncorrelated estimates in the two runs. We have found that one can do somewhat better by choosing c dynamically to minimize the final error because the correlations are significantly different in the two runs. (Note that Bennett did not sample the other system as we do, hence for him $R_A = R_B$. The algebra for finding the optimal weighting function is not affected by generalizing it to include a sampling function.)

Using this weighting function, we obtain an estimate for the ratio of partition functions,

$$f_P = \frac{Z_B}{Z_A} = e^c \frac{\left\langle (1 + e^{-\delta + c})^{-1} \right\rangle_A}{\left\langle (1 + e^{\delta - c})^{-1} \right\rangle_B}, \quad (6.31)$$

where

$$\delta = S_A - S_B + \ln[T(R_A | R_0)/T(R_B | R_0)]. \quad (6.32)$$

Note that e^δ is exactly what is used to decide whether moves from "A" to "B" will be accepted ($e^{-\delta}$ for moves from "B" to "A") in a generalized Metropolis procedure; see Eq. (5.5). Hence the name "acceptance ratio method." One needs to find both the forward probability and the reverse probability. Bennett showed that it was more efficient and much more robust to do that mapping in both ways, as we have described, and speculated that mapping was superior to actually making the moves.

To summarize, one carries out two PIMC simulations, one in the "A" system and one in the "B" system. During each simulation one tries to map [using $T(B)$ or $T(A)$] into the other system, recording δ for the mapping and the averages of Eq. (6.31). The parameter c and the relative length of the "A" and "B" runs are chosen to minimize the error. Usually it is most efficient to run the "B" system 2–5 times longer than the "A" system, since it is more sluggish. A typical determination of the frequency versus c is shown in Fig. 51. The fact that J is independent of c within statistical errors is a very strong check on the correctness of the algorithm.

It is useful to examine the histogram of occurrences of values of δ. If there is an overlap in the distributions,
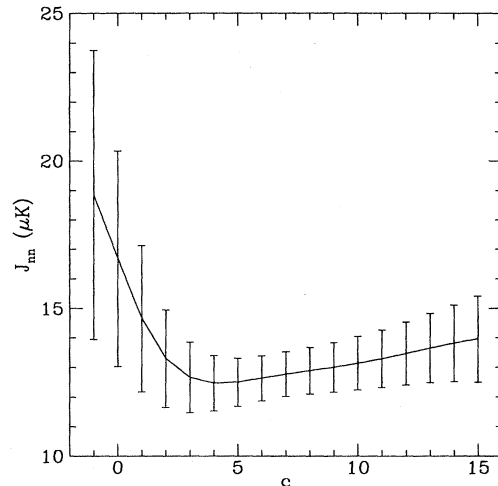

FIG. 51. The estimated nearest-neighbor pair-exchange frequency for various values of the blending parameter c. The curve corresponds to the histograms in Fig. 52. The optimal value of c is found to be 5.5. For a well-converged calculation, the exchange frequency should be independent of c within the statistical errors.

then one can trust the determination of the free-energy estimator. Otherwise the mapping from one system to the other has been unsuccessful; the "B" generated from the Markov chain are in a different region of phase space from those that are mapped directly from "A". An example for nearest-neighbor exchange is shown in Fig. 52; using the bisection procedure we were able to get a significant overlap.

The number of time slices used in the mapping is determined by optimization. If it is too small, the path cannot stretch over the barrier; if it is too large, sampling becomes difficult. It turns out that eight time slices works best for $\tau = 0.025K^{-1}$. Suppose we are able to do optimal sampling over a segment of m time slices, starting at R_i and ending at R_{i+m}. Then for the "A" to "B" mapping we would get the value $\delta = -S(R_i, PR_{i+m}; m\tau) + S(R_i, PR_{i+m}; m\tau)$. An estimate of this can be used to select which intervals from the "B" run attempt sampling. If we attempted to undo the exchange on time intervals where the "B" path is entirely on one side of the barrier, the mapped path would have two exchanges, not zero, and an inappropriate value of δ, one that did not fall in the overlap region. In the "A" run we could use the kinetic action to select the time intervals and the cluster of atoms in which to attempt an exchange, exactly as was done in Sec. V.J.

Note that the error on the exchange frequency in Table I is independent of the magnitude of the exchange frequency. It depends on the ability to map the "A" system into the "B" system and vice versa, not on the magnitude of the exchange frequency. It does depend on the number of exchanging atoms, since long exchanges are harder to sample.
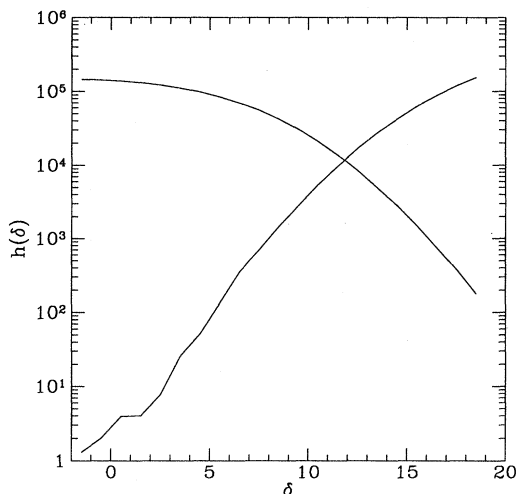
All the techniques described in Sec. V.K can be used to increase the efficiency of the mapping. In particular, multiple sampling is advantageous: at each level several possible bisections are made. The favorable combinations are used to continue the construction of the new path to finer levels. In the "B" walk, the cluster of atoms actually exchanging controls the efficiency. Thus the Metropolis Monte Carlo moves should focus on updating those atoms, and move the other atoms much less frequently. If this is done, the efficiency of calculating the exchange frequency becomes independent of the size of the crystal.

There is a further trick that can speed the calculation. Note that the exchange frequency is the slope of the partition function, and one would have to do calculations for several temperatures to determine both J and β_0 in Eq. (3.45). The offset β_0 has a very intuitive classical interpretation—it is the "tunneling time." The exchange frequency is $J_P = f_P(\beta)/(\beta - \beta_0)$. The denominator counts the total number of ways of placing the exchange in the interval $(0, \beta)$. There is an offset of β_0 because we have constrained the path to form a perfect lattice at the beginning and end. If we keep track of the "crossing time" we can estimate both β_0 and J from a single calculation. To define the *crossing time*, we project the "B" path onto a reaction coordinate: $z(t) = (R_t - Z) \cdot (PZ - z)$. The possible crossing times are the solutions of $z(t_c) = \frac{1}{2}|Pz - z|^2$. There must be an odd number of them; we define the middle one as the crossing time. We record the histograms as before, but all the while keeping track of the crossing time in the "B" system. In Fig. 53 we show the estimated exchange frequency, including only the paths with crossing times satisfying $t \le t_c \le \beta - t$ where $J(t) = f(t, \beta)/(\beta - 2t)$. If

FIG. 52. The histograms for relative actions δ defined in Eq. (6.32) for mapping from the "A" system to the "B" system (upper left curve) and vice versa, which is shown as $h(-\delta)$ (lower left curve). These calculations (Ceperley and Jacucci, 1987) are for nearest-neighbor pair exchange with 54 ^{3}He atoms at a density of 0.03 Å^{-3}.
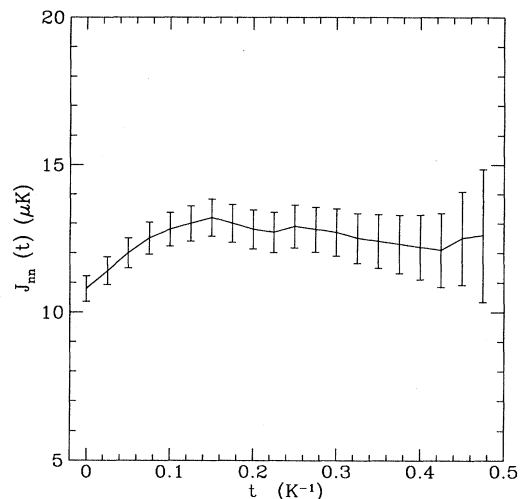
FIG. 53. The exchange frequency versus "crossing time window" t for nearest-neighbor pair exchange as shown in Fig. 52. $J(t)$ is described in the text. $J(0)$ is smaller because the paths are forced to begin and end on perfect lattice sites. If one restricts attention to path exchanging away from the ends, this effect is removed. The error bars increase for $t \approx \beta/2$ because there are fewer exchanging paths with those crossing times.

we pick $t > 0.1 \, K^{-1}$, then $J(t)$ is independent of t within the statistical errors.

F. Bayesian methods for estimating real-time response functions

In Sec. III.I we discussed the relationship between imaginary-time path integrals and the real-time response functions. Here we provide more details on the recent progress in using Bayesian inference to reconstruct the real-time response from the PIMC-determined imaginary-time correlation functions.

The dynamic structure factor $S(\omega)$ is related to the imaginary-time intermediate scattering function $F(t)$ by a Laplace transform, Eq. (3.51). However, the inversion to obtain $S(\omega)$ from $F(t)$ is unstable with respect to noise in $F(t)$. Recently maximum-entropy methods (Silver, Sivia, and Gubernatis, 1990; Gubernatis et al., 1991) have been applied to make the inversion better conditioned, with remarkable success for some lattice models. These methods combine highly accurate Monte Carlo-generated imaginary-time response functions with any theoretical input that may be available, and effectively give the most likely dynamical response function consistent with all the input. Jarrell and Gubernatis (1995) have recently reviewed the fundamentals and practical details of this approach, so we shall only touch on the basic points.

Bayes' theorem can be written in the following form:

$$P(S(\omega)|F(t)) \propto P_L(F(t)|S(\omega))P_P(S(\omega)). \quad (6.33)$$

In words, the probability of $S(\omega)$, given our PIMC data and the theoretical input [the *posterior probability*], equals the probability of the PIMC data given $S(\omega)$ [the *likelihood*] times the prior knowledge of $S(\omega)$ [the *prior function*]. We have neglected a normalization constant. We now discuss assumed forms for P_L and P_P.

The central-limit theorem guarantees that the noise in $F(t)$ from a well-converged PIMC is normally distributed so that the likelihood probability must equal

$$P_L(F(t)|S(\omega)) = \exp\left[-\frac{1}{2}\sum_{t,t'}\delta F(t)\sigma(t,t')^{-1}\delta F(t')\right], \quad (6.34)$$

where $\delta F(t) = F(t) - \langle F(t) \rangle$. There are, however, several practical difficulties. First, we have to ensure that our results are well converged so that a Gaussian distribution is appropriate. This must be done empirically and may require long runs. Second, both the mean value of $F(t)$ and all the correlations, $\sigma(t,t') = \langle \delta F(t)\delta F(t') \rangle$, are obtained from the same data. The relative error of σ can be large, since it is proportional to the fourth moment of δF. In addition, if there are M time slices, then the covariance matrix has about $M^2/4$ entries. To get estimates of all those correlations requires at least that many independent estimates of $F(t)$; in practice many more are

required. Note that we are not assuming that $\delta F(t)$ is independent from $\delta F(t')$ for $t \neq t'$. That assumption is not true, even in the limit of good statistics; fluctuations at one point on the path are positively correlated with fluctuations elsewhere. Very long runs may be needed to determine σ! A practical shortcut is to zero its off-diagonal contributions to σ. For short runs this shortcut generally gives better results, but leads to a bias and may increase the overall error.

The form and even the existence of the prior function is controversial. One knows a few things about any physical $S(\omega)$: it is everywhere non-negative, it satisfies detailed balance and various sum rules, and it has certain asymptotic behaviors at large and small ω. We need to construct a prior function so that any $S(\omega)$ not satisfying these conditions will have a zero prior function. For image reconstruction and many-other applications the entropic prior has worked very well,

$$P_p(S(\omega)) \propto \exp\left[\alpha\sum_\omega S(\omega)\ln(S(\omega)/m(\omega))\right], \quad (6.35)$$

where α is an adjustable parameter and $m(\omega)$ is the *default model*. There is no proof that the entropic prior is correct for the "distribution" that describes $S(\omega)$. But the entropic prior is a convenient function that has worked very well in many applications, and it has a number of nice properties. It keeps $S(\omega)$ positive and it has a maximum when $S(\omega) = m(\omega)$, but decays very slowly around this maximum. This means that as the PIMC data get better, the prior function quickly gets out of the way but, in regions where the data is poor, the default model is important. The "flat" default model is $m(\omega) = 1$. Together the entropic prior and the likelihood function are convex, so that the posterior probability has a unique maximum. The parameter α controls how close S and m should be. It is adjusted using a variety of techniques; see Gubernatis et al., 1991 or Jarrell and Gubernatis, 1995.

Having chosen the likelihood and prior function, one can take either of two approaches to using Bayes' theorem. In the *maximum-entropy* approach one finds the $S(\omega)$ that is most likely, the one which maximizes $P(S(\omega)|F(t))$. This is a good procedure when the probability distribution is narrow. Finding the maximum is very fast. Errors are estimated by computing second derivatives at the maximum. A second, more intuitive and more rigorous approach, *average entropy*, is to sample $S(\omega)$ with a probability equal to the posterior probability. This can easily be done with Metropolis Monte Carlo. This approach is slower (but not nearly as slow as generating the original PIMC data) but does not rely on any assumptions about how narrow or skewed the probability function is. Some features of $S(\omega)$ may be tightly constrained by the PIMC data, while other features are not constrained. To compute errors one simply looks at the fluctuations of $S(\omega)$ coming from the Markov chain which samples $S(\omega)$. An additional advantage of the average-entropy approach is that the model can be

self-consistently defined as $m(\omega) = \langle S(\omega) \rangle$.

There are at least two ways of improving the result of the Bayesian analysis; one can sharpen either the likelihood function or the prior function. One can get more information from the PIMC simulation by running longer, or more efficiently and by using other estimators. One possibility that has been explored by Caffarel and Ceperley (1992) is to estimate the "time" derivative of $F(t)$. Because of strong correlations between $F(t)$ and $\dot{F}(t)$, this constrains $S(\omega)$ more tightly. The second way is to get better prior knowledge, for example, more exact properties. If one has confidence that some theoretical approach generates reasonable families of $S(\omega)$, one can work in that function space instead of in the full $S(\omega)$ space. That will increase the statistical precision at the cost of an additional assumption.

VII. COMPARISON WITH OTHER QUANTUM MONTE CARLO METHODS

In this section we make some brief comparisons with other Monte Carlo methods that have been applied to liquid and solid helium or related quantum many-body systems. In particular we shall examine variational Monte Carlo (VMC), variational path integrals (VPI), shadow wave functions (SWF), projector Monte Carlo [also known as Green's-function Monte Carlo (GFMC) or diffusion Monte Carlo (DMC)], and effective-potential Monte Carlo (EPMC). We hope the reader will be able to see the interrelationships between the various methods, the strengths and weakness of the methods, and how to choose the appropriate method. This will also give us an opportunity to comment on whether some of the tricks and lessons we have discussed in this review apply to the other methods. Finally, we hope to convince the reader that path-integral Monte Carlo is the best general-purpose method for simulations of bosonic systems.

A. Variational Monte Carlo

Variational Monte Carlo (VMC) is the oldest of the simulation techniques that have been applied to helium (McMillan, 1965). One assumes a trial wave function ψ_T, usually of the pair-product (Jastrow) form, $\psi_T = \exp[-\sum_{i<j} u(r_{ij})]$. Then the variational energy, an upper bound to the exact energy, is given by

$$E_V = \frac{\int dR \psi_T^* \mathcal{H} \psi_T}{\int dR \psi_T^* \psi_T}. \qquad (7.1)$$

One performs the ratio of integrals by using the Metropolis Monte Carlo or molecular dynamics method to sample the distribution, $|\psi_T|^2 / \int |\psi_T|^2$. The variational energy is the average value of the residual or local energy of the trial function, $E_L = \psi_T^{-1} \mathcal{H} \psi_T$. One chooses parameters in the trial function to minimize the variational energy or by some other procedure. The zero-variance principle

applies; as the trial function approaches an exact eigenstate, the variance vanishes. For a comprehensive review of the methods and applications to helium see Ceperley and Kalos (1979) and Schmidt and Ceperley (1992).

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. Humans (unlike computers) have a tendency to stop when they like the look of the results. This inevitably introduces a systematic error. Even the best pair-product trial functions miss about 1.2 K energy in liquid ^{4}He. If one includes three-body correlations, one still misses 0.3 K. 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. We discuss other problems with the solid next. VMC becomes much less reliable as the physical system gets more complex because a high-quality trial function has too many possible variational parameters to optimize easily. PIMC and GFMC build this complexity into the algorithm, rather than have the human guess which terms are needed.

Variational calculations of solid helium have an additional possible problem. In order to get the solid to be stable and to have reasonably good energies, one typically uses a trial wave function that breaks both translational and particle symmetry. The atoms are explicitly tied to lattice sites with Gaussians,

$$\psi_T = e^{-\sum u(r_{ij}) - \sum_i C(r_i - Z_i)^2}, \qquad (7.2)$$

where C is a variational parameter and $\{Z_i\}$ is the set of perfect lattice sites. The resulting energy is quite good, better than in the liquid. Although one can argue that it is quite respectable to have a broken-symmetry trial function in the thermodynamic limit, the symmetry of the solid is put into the calculation rather than coming out automatically. One can restore the symmetry of the Hamiltonian by using a translation operator and a permutation operator without increasing the energy significantly. Nonetheless, it is difficult to study such problems as the liquid-solid interface or vacancies in a solid with a trial function that has the crystal symmetry put in from the start.

There is another fundamental problem with VMC; the energy is insensitive to long-range correlations. The en-

ergy is dominated by the local order, i.e., the nearest-neighbor correlation. The long-range order usually has to be put into the trial function using an analytical theory. It does not come out. It is a very tricky matter to use VMC calculations to elucidate the long-range structure of the ground-state wave function, particularly if there are several plausible choices. Of course, liquid and solid helium are well understood, but for systems even a little more complicated, VMC is not really up to giving any more than suggestive results about the order. All VMC rigorously gives is an upper bound to the ground-state energy.

This type of variational calculation is only appropriate at zero temperature. One might think that this would be an advantage if one were interested in helium only at very low temperatures. After all, PIMC has to work its way down from the classical regime. However, at zero temperature there is no difference between a boson system and a boltzmannon system. Particle statistics make a tremendous difference in observed properties, since most of the states near the ground state disappear on symmetrizing, and a gap opens up, resulting in superfluidity. One can easily distinguish between a boltzmannon and a boson system at any positive temperature. This is why the first calculations of the superfluid density were done with PIMC, not with a strictly zero-temperature method.

B. Variational path integrals and shadow wave functions

What we now describe is a cross between the variational method and path integrals. A well-known method to improve a trial function is to apply a function of the Hamiltonian to project out the ground state. Although other choices are possible, for brevity and to make the connection to path integrals, we only consider using the density matrix as a projector. Define

$$\phi(\beta) = e^{-\beta\mathcal{H}}\psi_T. \tag{7.3}$$

Clearly $\phi(\beta)$ converges to the exact ground-state wave function exponentially fast as β increases, assuming ψ_T has some overlap with the ground state. In the next subsection, we shall implement the projection with branching random walks. Here we do the projection with Metropolis Monte Carlo. We shall call this the *variational path-integral* (VPI) method. It allows us formally to make the connection between diffusion Monte Carlo, variational Monte Carlo, shadow wave functions, and PIMC.

In VPI one uses $\phi(\beta)$ as a trial wave function to find a sequence of decreasing upper bounds, $E(\beta)$, to the ground-state energy. The upper bound at $\beta = 0$ will be the usual variational energy of ψ_T, but then E_β will converge monotonically to the exact ground-state energy in β. Figure 54 shows an application of this to liquid helium. After only a single time slice $\tau = 0.025$ K^{-1}, most of the errors of a pair-product trial function are fixed in the energy. In the VPI approach, one corrects a trial function by projecting instead of manually putting new
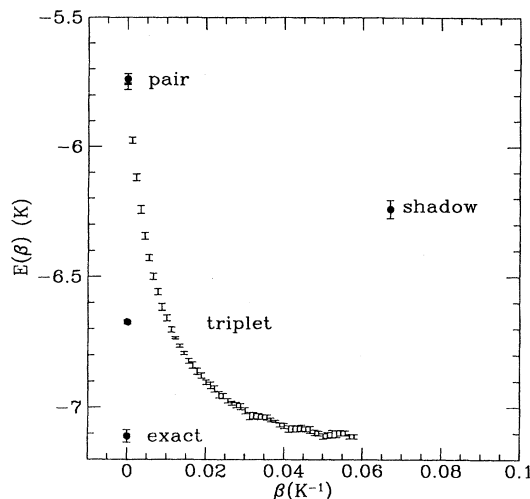

FIG. 54. The energy/atom of liquid ^{4}He versus projection time as calculated by the diffusion Monte Carlo method (Boninsegni, 1994). The initial point ($\beta = 0$) is the variational energy of a McMillan (1965) pair-product trial function. Also shown are the results for a triplet trial function, a shadow trial function (the projection time is defined by the width of its Gaussian), and the exact value as calculated by GFMC (all from Vitiello *et al.*, 1990).

terms into the trial function.

To calculate the VPI energy with Monte Carlo simulation, we need to break up the density-matrix operator into small pieces, each with a time step τ, and approximate the density matrix as before. This leads to a path $\{R_{-M} \ldots R_0 \ldots R_M\}$ having a probability distribution

$$\Pi(R_{-M} \ldots R_0 \ldots R_M) \tag{7.4}$$

$$= \Psi_T^*(R_{-M})e^{-\sum_{i=-M+1}^{M} S(R_{i-1},R_i;\tau)}\Psi_T(R_M), \tag{7.5}$$

where S is the link action. We have $2M$ links because we need integrals over $|\phi(\beta)|^2$.

Except for the different boundary conditions on the ends of the paths, VPI and PIMC are the same: one has linear polymers, not ring polymers, and there is a special action, $-\ln(\Psi_T)$, which acts between the atoms at ends of the polymers. If the trial function is a pair-product (Jastrow) function, this is just an additional pair potential at the ends. One no longer has the symmetry over time of the path integral. The middle slice, R_0, needed to compute observables can be different from the ends. As with PIMC, one has to choose a good form for the action and a method to move the paths around. Since the paths are open, one does not have to worry about permutations; single-atom moves are sufficient.

The energy is calculated by allowing the Hamiltonian to act either on $S(R_0, R_1)$ or on $S(R_0, R_{-1})$. But if the

time step is small enough, the Hamiltonian can commute through density matrices and one can average over which link the Hamiltonian should act, gaining efficiency if there are many links. One can also use the tricks we developed in Sec. VI.A to lower the variance of the energy. But there is a new feature of VPI: one recovers the zero-variance principle of the variational Monte Carlo method by moving the Hamiltonian to one end of the path and letting it act on the trial function instead of on one of the link actions.

The advantage of VPI is that, for strictly ground-state applications, many fewer links will be required than with PIMC, since the number of links is related to deficiencies of the trial function and the needed projection time, not the statistical mechanics of how cold the system must be to achieve the ground state. One must be cautious in determining how many links will be required. Most of the variational energy will be picked up very quickly, as we see in the figure. Long-range correlation functions will be fixed much more slowly. An example is the condensate fraction in the conjectured supersolid helium. It has been proved by Reatto and Masserini (1988) that the VPI wave function for finite M has a nonzero condensate in both the liquid and the solid phase if the trial function does. A short-ranged Jastrow trial function has a condensate, thus the VPI trial function using this will have a condensate for any finite number of links. Hence one cannot use this approach to ask the question whether the solid phase is Bose condensed, without examining the limit $M \to \infty$.

Note that the variational principle allows us to minimize the energy with respect to the action. This is the basic idea of *the shadow wave function* (SWF) as introduced by Vitiello *et al.* (1988, 1990). To date all shadow wave functions have been restricted to the primitive approximation for the action and only a single time step, $M = 1$. The parameters in the shadow trial function are a pair potential between atoms at the ends of the polymer, a different one in the middle, and the spring constant connecting them. They all can be varied to minimize the energy. The energy must be computed by having the Hamiltonian act on the middle time slice. If one had the exact action, one would achieve the lowest energy by making β as large as possible. However, using only the primitive action, it is found that the optimal value of β is 0.068 K^{-1}, more than twice what is used for the PIMC time step. The shadow wave-function energy in liquid helium is about 0.8 K/atom above the exact results, better than the purely pair-product energy (i.e., $M = 0$), but not as good as putting in three-body correlations. The SWF energy is much higher than one would get by using a projection for the same time. The restriction to an action with only simple spring couplings costs a significant energy.

The big advantage VPI and shadow wave functions have over variational Monte Carlo is that one does not have to impose a solid symmetry in the trial function to get a solid (Vitiello, Runge, and Kalos, 1988). Just

going from $M = 0$ to $M = 1$ is enough to stabilize the quantum solid. One can study cases in which the liquid and solid are in equilibrium without imposing the result in advance. However, the result is not a "black box." With shadow wave functions one still has to optimize variational parameters, and the liquid-solid transition does not come out at the experimental density. The code is significantly more complicated than with variational Monte Carlo, since it has many of the sampling problems that PIMC has; polymers can get more entangled than atoms, so convergence can be slow. Error bars on the energy are larger because the zero-variance principle of VMC is lost. The variance would be reduced by using the exact action, since then the Hamiltonian could be moved over to act on the trial function.

VPIs and SWFs are variational wave functions, so they share some of the advantages and disadvantages that we discussed earlier with VMC. By adding more links, one does get much lower energies very quickly, in an automatic fashion. But long-range correlations may converge much more slowly. One can also use VPI to calculate excited-state properties, but by expanding into a path one has made a very significant reduction in efficiency. Take the example of a phonon. The Feynman trial function for a phonon is $\psi_k = \sum_i e^{i\mathbf{k}\cdot\mathbf{r}_i}\psi_0$, where ψ_0 is a ground-state trial function. The distribution to be sampled is no longer positive or even real! The infamous sign problem of quantum Monte Carlo is back. One is forced to sample the modulus of $\psi_{-k}(R_{-M})\psi_k(R_M)$ and use the sign as a weight, but the statistical efficiency will go to zero for long projecting times. One does not have this problem with variational Monte Carlo because the two ends of the path are at the same location, so one has a perfect square. Hence the calculation of excited states, either by PIMC or by variational path integrals, still has major problems.

C. Green's-function and diffusion Monte Carlo

The first exact calculation of liquid helium using the Green's function Monte Carlo (GFMC) method was by Kalos, Levesque, and Verlet in 1974. For reviews of the method and results see Ceperley and Kalos (1979) and Schmidt and Ceperley (1992). Diffusion Monte Carlo (DMC) is a simplified version of GFMC with time-step errors. What we shall be describing here is DMC. The difference between DMC and GFMC is irrelevant in making the comparison with PIMC.

In PIMC or variational path integrals, 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.

GFMC is a technique mathematically very similar to the VPI method. The density matrix $e^{-\beta\mathcal{H}}$ is used to

project out the ground state from some initial state. However, the implementation of this projection 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 $\langle R'|e^{-\tau \mathcal{H}}|R\rangle$. The normalization (the integral over R') is not unity and depends on R. Hence 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. Typically several hundred configurations make up an ensemble. Importance sampling by a trial wave function ψ_T reduces the fluctuations of the branching, so that it is proportional to the error of the trial function. In practice, the number of configurations held in the computer memory is roughly the same in GFMC and PIMC.

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; 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. (Of course this statement refers only to calculations on liquid helium with existing codes.) 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. In Sec. VIII we shall review the complexity of PIMC simulation.

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

D. Effective-potential Monte Carlo

Finally we should like to discuss the relationship between the effective-potential Monte Carlo (EPMC) and PIMC methods. For a nearly classical system, say solid neon at 50 K, one might hope to be able to write down a sufficiently accurate action so that only one link (i.e., $M = 1$) of the path integral would be needed. In a PIMC simulation with a single link, only the diagonal action $U(R)$ is needed. Then a "classical" Monte Carlo or molecular-dynamics simulation of atoms with an effective potential could calculate all thermodynamic properties. For the primitive approximation for the action, $U(R) = \beta V(R)$; the effective potential equals the potential. To get any quantum corrections with $M = 1$ requires one to go beyond the primitive action.

In using the effective-potential Monte Carlo, one chooses $U(R)$ based on a local harmonic approximation

to the full potential. Consider the finite-temperature variational principle for the free energy F (Feynman, 1972),

$$F \leq F_T + \langle V - V_T \rangle_T. \qquad (7.6)$$

Here V_T is a trial potential, V is the true potential, $\langle \cdots \rangle_T$ is an average over the density matrix generated by the trial potential, and F_T is the free energy of the trial potential. In EPMC one chooses the trial potential to be quadratic, so that most of the averages can be performed analytically in terms of normal modes of the potential. The parameters in the quadratic potential are allowed to be general functions of the centroid of the path, and they are determined by minimizing the free energy. If this is done, the second term on the right-hand side vanishes. The self-consistent potential is defined similarly to the sampling potential introduced in Eq. (5.18). For a description of the method see Liu, Horton, and Cowley (1991).

The bound on the free energy will be very good if the true potential is nearly harmonic in important regions of configuration space. However, the equations that determine the effective potential reduce to self-consistently determining the phonon frequencies of an effective dynamical matrix, which involves diagonalizing a $3M \times 3M$ matrix. The effective potential and the phonon frequencies need to be resolved at each step of the EPMC, since the harmonic potential depends on R. To avoid such heavy computation within a classical simulation, further approximations are made, by expanding in terms of the displacement with respect to the lattice sites and truncating terms. This makes the method nonvariational.

There have been claims (Liu et al., 1992; Cuccoli et al., 1992) that EPMC is much more efficient than PIMC when applied to systems such as solid neon. However, the tests used very inefficient PIMC algorithms with only the primitive approximation, the simplest sampling method, and the thermodynamic estimator for the energy and specific heat. Because of the high temperature used for the comparison, the virial estimator of the energy should have been used. The comparisons do not apply to an efficient implementation of PIMC. If one knows in advance that the system is nearly harmonic, then EPMC could give useful results, but only if it is only a few times more expensive than the classical calculation.

PIMC is a general approach that can treat any system, liquid or solid, while EPMC is restricted to nearly harmonic systems and even there requires approximations. The primary justification for using PIMC is its greater generality and accuracy. However, one might wonder in what situations one could gain efficiency by introducing complexity into the trial action. In EPMC the action is quite complex; one has to determine the phonon frequencies at each step of the EPMC, which makes the calculation very slow for large numbers of atoms. The approach in PIMC is to do a good job of constructing the action, by, for example, using the exact pair action. But one uses only actions that can be quickly computed.

Once the pair action is tabulated, looking it up during the PIMC is only a few times slower than using the primitive approximation. One has not increased the complexity of the calculation in so doing but only slowed the evaluation of the action by a factor of about two. To improve the action further, one takes more time slices rather than devoting computation to an improved action. We discuss the complexity of PIMC in the next section. Ultimately, detailed tests are needed to determine where to draw the line between increased complexity of the action, the accuracy of various approximations, and more time slices.

VIII. COMPUTATIONAL COMPLEXITY

Suppose we wish to calculate a given property $\mathcal{O}$ of a quantum system to within an absolute error ϵ. This error can arise from several different sources: the statistical error ϵ_s, the time-step error ϵ_τ, finite-size errors ϵ_N, and possibly extrapolation to zero temperature, ϵ_β. The length of the simulation, the time step, the temperature, and the number of particles need to be adjusted so that the errors are less than ϵ. Let $T_\mathcal{O}(\epsilon)$ be the amount of computer time (or other resources) such a calculation will take. As we shall see, in PIMC the computer time has a power-law dependence on the accuracy,

$$T_\mathcal{O}(\epsilon) \propto \epsilon^{-\delta}. \qquad (8.1)$$

The exponent δ we call the *complexity* of the algorithm for $\mathcal{O}$. In this section we pull together some of the formulas we have discussed earlier to estimate the δ appropriate to calculations of the ground-state energy of bulk liquid ^{4}He. This power-law behavior is quite different from the scaling one gets from an explicit method, one where the wave function or density matrix is completely tabulated throughout many-body configuration space. In that case the computer time will inevitably scale exponentially in the number of particles.

As we discussed in Sec. V, the statistical error of a random walk is related to the computer time by $T = 1/(\xi \epsilon_s^2)$. Here ξ is the efficiency of the Monte Carlo sampling. If we only had to worry about the statistical error, the complexity would be $\delta = 2$. But to make sure that the systematic errors go to zero we need to know how the efficiency depends on the time step, the number of particles, and the temperature. Now the efficiency, in turn, is related to ν, the intrinsic variance of an estimator (the variance assuming all steps of the random walk are uncorrelated), to t, the computer time per step, and to κ, the number of steps over which the random walk is correlated, by

$$T_\mathcal{O}(\epsilon) = \frac{t \kappa_\mathcal{O} \nu_\mathcal{O}}{\epsilon_s^2}. \qquad (8.2)$$

Now let us examine how small the time step τ needs to be in order that the time-step error be equal to the

statistical error. The simplest approach to getting rid of the time-step error is to reduce the time step until the change in the estimated value of $\mathcal{O}$ is independent of τ within the statistical errors. But if one knows the τ dependence, one can extrapolate from larger values of τ. There will be some additional statistical error incurred in extrapolating, but overall one will save computer time. We shall ignore the possibility that one can change the scaling in τ by extrapolation.

Let us assume that the efficiency scales with the time step as $\xi \propto \tau^\alpha$. As mentioned above, there are two parts to the efficiency. The diffusion constant, $D^{-1} = t\kappa$, is roughly independent of the quantity to be estimated. We estimated, in Sec. V.H, that single-slice sampling had an exponent of 3 for D, and the bisection algorithm an exponent of 1.4. The second part of the efficiency is the intrinsic variance. Taking the example of the average internal energy, in Sec. VI.A we showed that for the thermodynamic estimator $\nu \propto \tau^{-2}$, while the virial estimator was independent of τ. Once the time step is small enough, the variance is independent of the number of time slices, since the neighboring time slices are highly correlated in imaginary time. So, for estimating the energy, we see that $1 \le \alpha \le 5$ depending on the sophistication of sampling and constructing the estimator. The value of $\alpha \approx 1.4$ is appropriate for the bisection algorithm with the virial energy estimator.

Let us suppose that the time-step error scales as $\epsilon_\tau \propto \tau^\kappa$ where $\kappa = 2$ in the primitive approximation. Higher exponents are achievable with better actions. Then, when we match these two errors, $\epsilon_s \approx \epsilon_\tau$ implies that $\delta = 2 + \alpha/\kappa$. For very accurate runs, we not only have to beat down the statistical errors but we have to decrease the time step. The effect is that the error bars converge more slowly. With a good sampling method, estimator, and action, $\delta \approx 2.5$, while with the primitive action, single-slice sampling, and the virial estimator, $\delta \approx 4.5$. This shows the importance of optimizing the PIMC algorithm.

Now let us see how the computer time scales with the number of atoms, for the moment ignoring Bose statistics. We assume that the efficiency scales as $\xi \propto N^{-\gamma}$, that the system is not near a phase transition, and that the size of the box is much greater than the correlation length. Then the number of passes (a pass is one move for each particle) to decorrelate some property will be independent of the size of the system, so that $\kappa \propto N$. (It is important to remember here that we have defined κ in terms of elementary local Monte Carlo steps, while physically it is simpler to measure the relaxation in terms of steps in which all the particles get moved.) If we look at a local property, such as the energy/atom or the pressure, what we lose in the slower updating in a larger system, we gain in having more local regions to average over.

We also need to know how the computer time per step depends on the number of atoms. For helium, the potential is short ranged, so that in calculating the potential or action one only needs to sum locally. Using neighbor ta-

bles, the computer time for a single move is independent of the number of atoms. Recently the "fast-multipole" algorithm has been developed for charged systems. In that algorithm, the total potential and force can be computed in a time proportional to the number of atoms. Such an algorithm has not been developed for Monte Carlo methods, where atoms are moved individually; the work for a single-particle move for a Coulomb potential currently goes as $t \propto N^{1/2}$.

With the above assumptions, $\gamma \approx 0$. One can increase the system without incurring any penalty in efficiency and effectively approach the thermodynamic limit. Limitations of computer time and memory, not to speak of slow relaxation of the system, will limit this in practice. If we assume that the error in the energy goes as $\epsilon_N \propto N^{-1}$, then after matching the statistical, time-step, and finite-size errors, we find $\delta = 2 + \alpha/\kappa + \gamma$. Approaching the thermodynamic limit causes a further increase in the exponent, but that increase can be made quite small away from a phase transition.

Now let us return to the question of whether Bose statistics changes the complexity. In the permutational sampling, what is done is to attempt few-body permutations in a local region. Those permutations can be found by local searches which are independent of the number of atoms. Again we assume we are away from the superfluid transition, so the system at large N should settle into either a normal-liquid state or a superfluid state, and the correlation time should be independent of the size of the system. Hence Bose statistics should not change the complexity. Quantities such as the condensate fraction and the superfluid density may not scale as well with the number of particles. As we discussed in Sec. VI, moves that change the mean-squared winding number become harder and harder to construct as the system gets larger. We also discussed new, but untested, ways around these difficulties. How these methods really scale is not known.

Finally what happens as we go lower in temperature? Often one is interested in the zero-temperature values. Clearly the computer time to update the path, scales as the number of time slices. But for a property like the energy, doubling the number of time slices has no effect on the efficiency, since there are twice as many links over which one can collect statistics. This is what is observed in practice. Once one is below the superfluid transition, the errors bars depend only on the duration of the run, not the temperature. In bulk ^{4}He the error of the energy due to thermal fluctuations goes as $\epsilon_\beta \propto \beta^{-4}$. Hence in bulk helium it is quite easy to reduce these statistical fluctuations and get to the ground state. For this, Bose statistics are a distinct advantage over boltzmannon statistics, where $\epsilon_\beta \propto \beta^{-1}$.

The complexity of dynamical properties such as the the roton lifetime or the energy of a vortex are still more difficult to estimate. It is not known how the Bayesian methods we discussed earlier scale with ϵ or N. There is a "leap of faith" required in believing the internal error estimates calculated by the maximum-entropy ana-

lytic continuation. These methods are a necessary tool for PIMC, but more practical and theoretical studies are needed to assess their convergence.

The main goal of this subsection on complexity is to reinforce the conclusion that accurate simulations of the thermodynamics of Bose systems are possible; the complexity of Bose superfluids is not much worse than for the classical Ising model. Of course, the efficiency, i.e., the coefficient in front, is orders of magnitude lower in the case of PIMC calculations of helium. There are many possible ways to increase the efficiency, including waiting for faster computers.

One can use the complexity to estimate crudely when certain calculations will be possible. Looking over the last fifty years, one sees that the speed of the faster computers is exponentially increasing, with a doubling time of between one and two years. Assuming scientists continue to get access to those computers and that this rapid development continues, one can ask when a given PIMC will be possible. As early as 1986 the energy per atom of liquid ^{4}He was calculated to an accuracy of ± 0.04 K/atom (Ceperley and Pollock, 1986). Assuming the performance of computers has a doubling time of $Y = 1$ year and that the complexity has an exponent $\delta = 2.5$, the doubling time for the accuracy will be $Y\delta = 2.5$ years. Hence by the end of this decade, even assuming there are not better methods, we should be able to compute the energy to better than $\approx \pm 0.001$ K/atom. That more accurate energies have not appeared in the literature is because there is little scientific interest in the ground-state energy.

As we indicated above, the complexity of quantum mechanics in general is exponential in the number of particles (Feynman, 1985). By general quantum mechanics, we mean a method analogous to classical molecular dynamics which can evolve an arbitrary initial state in real time. Both specifying the initial state and performing the evolution is exponentially difficult. One sees this effect by watching the computational literature of quantum dynamics. Even with the rapid growth of computational resources, the largest explicit simulations of interacting particles in the continuum are of four-body systems. Path integrals get around this limitation because they are probabilistic, they sample configuration space rather than integrate over it. But known methods of doing this efficiently require a real action which limits their application to bosons in imaginary time. It is not clear how much can be ultimately calculated in imaginary time. We gave a few examples of quantities that one might naively have thought were dynamical but that turned out to be easily obtainable in imaginary time: the superfluid density, the exchange frequency in quantum crystals, and the momentum distribution. One can find many more examples. Computational many-body physics has another challenge, namely to devise an approximation-free algorithm that can treat fermions as well as we can now treat imaginary-time bosons. Without that, the scope of PIMC is quite limited.

IX. SUMMARY AND OUTLOOK

Let us return to an issue raised in the beginning of this article. Does the path-integral picture provide an understanding of superfluidity and Bose condensation? With PIMC we have the ability to make (arbitrarily) precise numerical calculations of boson systems. There is nothing like getting the numbers out and doing precise comparisons with experiment and other theories to prove that the theory and calculational methods are correct. The numerical aspects of path-integral Monte Carlo are still very immature; this is the first comprehensive review, and even the interaction potential between helium atoms is known to only 1% accuracy, which limits the degree to which numerical calculations can be compared to experiment. But so far we have not discovered any major problems in comparing the calculations with experiment. Difficulties to date are readily explained by computational shortcuts, lack of a precise potential-energy function, or experimental difficulties. As we have just explained in the preceding section, there are no insurmountable difficulties with doing much more precise calculations on the equilibrium properties of Bose superfluids if there were a good enough motivation.

But is this an understanding? We assert that PIMC simulation has helped us understand superfluid helium and solid helium. Bose condensation and superfluidity are both seen to result from macroscopic exchange, just as Feynman argued. Path integrals map the original problem of quantum-statistical mechanics into a domain that it is easier to understand. The path-integral picture gives us a new understanding of superfluidity, Bose condensation, the exchange energy, and tunneling in quantum crystals, to mention just a few physical phenomena. The path-integral picture also shows how these various concepts are related in detail, thus unifying our understanding of them.

What is the future outlook? Of course much needs to be done to make PIMC, as applied to bosonic systems, more efficient and better able to calculate a broad range of properties. Vortices are a very important feature of superfluidity, but detailed calculations of vortex properties in superfluid ^{4}He have not yet been attempted. One would like to see vortices come naturally from path integrals and to be able to calculate their properties in an approximation-free way, as we can for quantities like the momentum distribution or the superfluid density.

But the PIMC method has not been extensively developed to do only simulations of liquid and solid ^{4}He. Imaginary-time path integrals are also useful in many other areas of quantum physics. We have not touched at all on the much more extensive uses of path-integral methods for lattice models, for example, those used in lattice gauge theory, in spin models of magnetism or of electrons in disordered media, or in superconductors. There are of many analogies to what we have developed here but also important differences.

There has been some recent progress in extending

bosonic PIMC to continuum fermion systems in an approximate way, but by a method that is practical for hundreds of fermions. The bosonic paths are prevented from crossing nodes of a trial density matrix. If these nodes are chosen correctly, the results are exact. This restricted path-integral method has been applied to liquid ^{3}He (Ceperley, 1992), to mixtures of liquid ^{3}He and ^{4}He (Boninsegni and Ceperley, 1995), and to a hydrogen plasma (Pierleoni et al., 1994). Many more applications are possible. The computational problems posed by fermion path integrals are just those we have reviewed for bosonic systems: finding good actions, sampling paths, constructing permutations, and estimating quantities. Fermions add another level of complexity on top of this. The fermion method promises to give new insights and computational abilities on highly correlated fermion systems.

We hope that eventually PIMC simulation can be a "black box." A nonexpert will specify the temperature, particle masses, spins, interactions, chemical potentials, and boundary conditions of the quantum system. The computer will return estimates of various observables, complete with error bars. It should be a major goal of computational many-body physics to show how this can be done. Of course, much work is needed.

ACKNOWLEDGMENTS

APPENDIX A: THE VIRIAL ESTIMATOR OF THE ENERGY

The derivation of the virial energy estimator for the general case of arbitrary actions, periodic boundary conditions, arbitrary window time, and particle exchange is not in the literature, so we sketch it here (Berne, 1990). We do this by using Green's theorem to transform the highly fluctuating kinetic term to one that fluctuates less.

Consider the path-integral expression for the density matrix with L links,

$$\rho_{0L}(L\tau) = \langle R_0 | e^{-L\tau \mathcal{H}} | R_L \rangle \tag{A1}$$

$$= \int dR_1 \cdots \int dR_{L-1} \exp\left(-\sum_{i=1}^{L} (K^i + U^i) \right),$$

where K and U are defined in Eqs. (2.22) and (2.23). To do things properly, we must worry about periodic boundary conditions. The simplest way to do that is to change the integration variables from R_i to $\delta_i = R_i - R_{i-1}$. We shall always assume that τ is small enough that we can neglect paths that wrap around the boundary within a single step. Note that we are not neglecting paths that wind around the boundaries after L steps. The kinetic action K^i depends only on δ_i. When R_i is needed, in U or elsewhere, we shall define it in terms of the δ's,

$$R_i = R_0 + \sum_{s=1}^{i} \delta_s. \tag{A2}$$

Now we compute the energy as the τ derivative of this path,

$$E_{0L} = -\frac{d\ln(\rho_{0L})}{d\tau}. \tag{A3}$$

Carrying out the derivative of the action as in Eq. (6.7), we obtain

$$E_{0L} = \frac{3LN}{2\tau} + \sum_{i=1}^{L} \left\langle -\frac{\delta_i^2}{4\lambda\tau^2} + \dot{U}^i \right\rangle \tag{A4}$$

where $\langle \cdots \rangle$ indicates an average over $R_1 \cdots R_{L-1}$, and $\dot{U} = dU/d\tau$.

To transform the kinetic term into something that might fluctuate less, consider the integral

$$G = \frac{1}{\rho_{0L}} \sum_{i=1}^{L-1} \int d\delta_1 \cdots d\delta_{L-1} (R_i - R_0) \nabla_i e^{-S}. \tag{A5}$$

Here ∇_i is the $3N$ gradient with respect to R_i. If we apply the ∇_i to the left, we see that

$$G = -3N(L-1), \tag{A6}$$

since all surface terms will vanish for a sufficiently small time step. Applying the gradient to the right gives

$$G = -\left\langle \sum_{i=1}^{L-1} (R_i - R_0) \left[\frac{\delta_i - \delta_{i+1}}{2\lambda\tau} + \nabla_i (U^i + U^{i+1}) \right] \right\rangle. \tag{A7}$$

This expression can be simplified with the identity

$$\sum_{i=1}^{L-1} (R_i - R_0)(\delta_i - \delta_{i+1}) = -\delta_L (R_L - R_0) + \sum_{i=1}^{L} \delta_i^2. \tag{A8}$$

Eliminating the kinetic action with these expressions, we get

$$E_{0L} = \left\langle \frac{3N}{2\tau} - \frac{\delta_L (R_L - R_0)}{4\lambda\tau^2} \right. \tag{A9}$$

$$\left. + \sum_{i=1}^{L} \dot{U}^i + \frac{1}{2\tau} \sum_{i=1}^{L-1} (R_i - R_0) \nabla_i (U^i + U^{i+1}) \right\rangle.$$

The internal energy is an average of E_{0L}; one has to average over the end points as well as R_i. We can also symmetrize both over the initial time-slice label and the "direction of time." We find

$$E_V = \frac{1}{L} \langle E_{k,k+L} \rangle = \frac{1}{L} \langle E_{k,k-L} \rangle. \qquad \text{(A10)}$$

This is the virial estimator. After some rewriting one ends up with Eq. (6.12).

APPENDIX B: LEXICON OF THE QUANTUM-CLASSICAL ISOMORPHISM

In this appendix we summarize the relationship between quantum concepts and the classical polymer language. Where possible, we have included the equation number that gives the precise meaning of this mapping.

Bose condensation: delocalization of ends of an open polymer [Eq. (3.14)].

boson statistics: allowing the possibility that polymers can hook up in any possible way [Eq. (2.29)].

degeneracy temperature: condition in which polymers are dense enough and extended enough that they touch and can exchange [Eq. (2.31)].

density: the bead density [Eq. (2.26)].

exchange energy: logarithm of the fraction of monomers (times $k_B T$) [Eq. (3.38)].

exchange frequency in a crystal: free energy to link polymers in a polymer crystal [Eq. (3.45)].

free energy: free energy of a system of ring polymers [Eq. (2.29)].

imaginary velocity: bond vector.

kinetic energy: negative spring energy [Eq. (2.25)].

moment of inertia: the mean-squared area of ring polymers [Eq. (3.26)].

momentum correlation function: bond-bond correlation [Eq. (3.36)].

momentum distribution: Fourier transform of end-end distribution [Eq. (3.9)].

pair-correlation function: pair-correlation function between beads at the same "time."

particle: ring polymer.

potential energy: iso-"time" potential between beads.

single-particle density matrix: the end-to-end distribution of an open polymer [Eq. (3.10)].

superfluid density: the mean-squared winding number [Eq. (3.31)].

superfluid state: a state in which a finite fraction of polymers are hooked together in polymers of macroscopic size.

temperature: (1) inverse polymer length [Eq. (2.19)]; (2) inverse coupling constant for the interpolymer potential [Eq. (2.24)]; (3) spring constant between neighboring beads [Eq. (2.22)].

thermal wavelength: polymer extension [Eq. (2.30)].

REFERENCES

Alder, B. J., and D. S. Peters, 1989, Euro. Phys. Lett. **10**, 1.
Allen, M. P., and D. J. Tildesley, 1987, *Computer Simulation of Liquids* (Oxford University, New York/London).
Abramowitz, M., and I. A. Stegun, 1970, *Handbook of Mathematical Functions* (Dover, New York).
Andronikashvili, E. L., 1946, J. Phys. USSR, **10**, 201.
Aziz, R. A., F. R. W. McCourt, and C. C. K. Wong, 1987, Mol. Phys. **61**, 1487.
Aziz, R. A., V. P. S. Nain, J. S. Carley, W. L. Taylor, and G. T. McConville, 1979, J. Chem. Phys. **70**, 4330.
Aziz, R. A., M. J. Slaman, A. Koide, A. R. Allnatt, and W. J. Meath, 1992, Mol. Phys. **77**, 321.
Barker, J. A., 1979, J. Chem. Phys. **70**, 2914.
Basile, A. G., 1992, Ph.D. thesis (Cornell University).
Batrouni, G. G., R. T. Scalettar, and G. T. Zimanyi, 1990, Phys. Rev. Lett. **65**, 1765.
Baym, G., 1969, in *Mathematical Methods in Solid State and Superfluid Theory*, edited by R. C. Clark and G. H. Derrick (Oliver and Boyd, Edinburgh), p. 121.
Bennett, C. H., 1976, J. Comput. Phys. **22**, 245.
Berne, B. J., 1990, private communication.
Berne, B. J., and Thirumalai, D., 1986, Annu. Rev. Phys. Chem. **37**, 401.
Binder, K., 1979, *Monte Carlo Methods in Statistical Physics* (Springer, Heidelberg).
Binder, K., 1992, *Monte Carlo Methods in Condensed Matter Physics* (Springer, Heidelberg).
Blasdell, R. C., D. M. Ceperley, and R. O. Simmons, 1993, Z. Naturforsch. **48a**, 433.
Boninsegni, M., 1994, private communication.
Boninsegni, M., and D. M. Ceperley, 1995, Phys. Rev. Lett. **74**, 2288.
Boninsegni, M., C. Pierleoni, and D. M. Ceperley, 1994, Phys. Rev. Lett. **72**, 1854.
Boronat, J., and J. Casulleras, 1994, Phys. Rev. B **49**, 8920.
Caffarel, M., and D. M. Ceperley, 1992, J. Chem. Phys. **97**, 8415.
Cao, J., and B. J. Berne, 1989, J. Chem. Phys. **91**, 6359.

Car, R, and M. Parrinello, 1985, Phys. Rev. Lett. **55**, 2471.

Ceperley, D. M., 1983, J. Comp. Phys. **51**, 404.

Ceperley, D. M., 1985, in *Molecular Dynamics Simulation of Statistical-Mechanical Systems*, edited by G. Ciccotti and W. G. Hoover (North-Holland, Amsterdam), p. 571.

Ceperley, D. M., 1989, in *Momentum Distributions*, edited by R. N. Silver and P. E. Sokol (Plenum, New York), p. 71.

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

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

Ceperley, D. M., and G. Jacucci, 1987, Phys. Rev. Lett. **58**, 1648.

Ceperley, D. M., and M. H. Kalos, 1979, "Quantum Many-Body Problems," in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer, Heidelberg), p. 145.

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

Ceperley, D. M., and E. L. Pollock, 1987, Can. J. Phys. **65**, 1416.

Ceperley, D. M., and E. L. Pollock, 1989, Phys. Rev. B **39**, 2084.

Ceperley, D. M., and E. L. Pollock, 1992, in *Monte Carlo Methods in Theoretical Physics*, edited by S. Caracciolo and A. Fabrocini (ETS Editrice, Pisa, Italy), p. 35.

Chandler, D., and P. G. Wolynes, 1981, J. Chem. Phys. **74**, 4078.

Chakravarty, C., 1993, J. Chem. Phys. **99**, 8038.

Chester, G. V., 1955, Phys. Rev. **100**, 455.

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

Crawford, R. K., 1976, Chap. 11 in *Rare Gas Solids* Vol. 1, edited by M. L. Klein and J. A. Venables (Academic, New York).

Cross, M. C., 1986, Phys. Rev. A **34**, 3531.

Cross, M. C., and D. S. Fisher, 1985, Rev. Mod. Phys. **57**, 1.

Cruzeiro-Hansson, L., J. O. Baum, and J. L. Finney, 1990, Phys. Scr. **T33**, 122.

Cuccoli, A., A. Macchi, M. Neumann, V. Tognetti, and R. Vaia, 1992, Phys. Rev. B **45**, 2088.

deRaedt, H., and B. deRaedt, 1983, Phys. Rev. A **28**, 3575.

deRaedt, B., L. M. Sprik, and M. L. Klein, 1984, J. Chem. Phys. **80**, 5719.

Donnelly, R. J., 1967, *Experimental Superfluidity* (University of Chicago, Chicago).

Donnelly, R. J., J. A. Donnelly, and R. N. Hills, 1981, J. Low Temp. Phys. **44**, 471.

Duane, S., A. D. Kennedy, B. J. Pendleton, and D. Roweth, 1987, Phys. Lett. B **195**, 216.

Edelsbrunner, H., D. G. Kirkpatrick, and R. Seidel, 1983, IEEE Trans. Inf. Theory **IT29**, 551.

Elser, V., 1984, Ph.D. thesis (University of California at Berkeley).

Feynman, R. P., 1953a, Phys. Rev. **90**, 1116.

Feynman, R. P., 1953b, Phys. Rev. **91**, 1291.

Feynman, R. P., 1953c, Phys. Rev. **91**, 1301.

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

Feynman, R. P., 1982, Int. J. Theor. Phys. **21**, 467.

Feynman, R. P., and M. Cohen, 1956, Phys. Rev. **102**, 1189.

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

Feynman, R. P., and H. Kleinert, 1986, Phys. Rev. B **34**, 5080.

Frenkel, D., 1986, in *Molecular-Dynamics Simulation of Statistical-Mechanical Systems*, Proceedings of the Fermi School of Physics, edited by G. Ciccotti and W. G. Hoover (North-Holland, Amsterdam), p. 151.

Friesner, R. A., and R. M. Levy, 1984, J. Chem. Phys. **84**, 4488.

Gallicchio, E., and B. J. Berne, 1994, J. Chem. Phys. **101**, 9909.

Giansanti, A., and G. Jacucci, 1988, J. Chem. Phys. **89**, 7454.

Ginzburg, V. L., and L. P. Pitaevskii, 1958, Zh. Eksp. Teor. Fiz. **34**, 1240 [Sov. Phys. JETP **7**, 858 (1958)].

Glyde, H. R., and A. Griffin, 1990, Phys. Rev. Lett. **65**, 1454.

Greywall, D. S., and P. A. Busch, 1991, Phys. Rev. Lett. **67**, 3535.

Griffin, A., 1993, *Excitations in a Bose-Condensed Liquid* (Cambridge University, Cambridge, England).

Gubernatis, J. E., M. Jarrell, R. N. Silver, and D. S. Sivia, 1991, Phys. Rev. B **44**, 6011.

Hall, R. W., and B. J. Berne, 1984, J. Chem. Phys. **81**, 3641.

Hammersley, J. M., and D. C. Handscomb, 1964, *Monte Carlo Methods* (Chapman and Hall, London).

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

Herring, C., 1962, Rev. Mod. Phys. **34**, 631.

Herwig, K. W., P. E. Sokol, T. R. Sosnick, W. M. Snow, and R. C. Blasdell, 1990, Phys. Rev. B **41**, 103.

Jacucci, G., and E. Omerti, 1983, J. Chem. Phys. **79**, 3051.

Jarrell, M., and J. E. Gubernatis, 1995, "Bayesian Inference and the Analytic Continuation of Quantum Monte Carlo Data," LANL preprint.

Kac, M., 1959, *Probability and Related Topics in Physical Science* (Interscience, New York), p. 165.

Kalos, M. H., M. A. Lee, P. A. Whitlock, and G. V. Chester, 1981, Phys. Rev. B **24**, 115.

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

Kalos, M. H., and P. A. Whitlock, 1986, *Monte Carlo Methods, Volume I: Basics* (Wiley, New York).

Kawashima, N., and J. E. Gubernatis, 1995, Phys. Rev. E **51**, 1547.

Kikuchi, R., H. H. Denman, and C. L. Schreiber, 1960, Phys. Rev. **119**, 1823.

Kleinert, H., 1990, *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics* (World Scientific, Singapore).

Klemm, A. D., and R. G. Storer, 1973, Aust. J. Phys. **26**, 43.

Krauth, W., N. Trivedi, and D. Ceperley, 1991, Phys. Rev. Lett. **67**, 2307.

Landau, L. D., 1941, J. Phys. USSR **5**, 71.

Landau, L. D., and E. M. Lifshitz, 1977, *Statistical Physics*, second edition (Pergamon, New York), p. 99.

Le Doussal, P., and D. R. Nelson, 1991, Europhys. Lett. **15**, 161.

Leggett, A. J., 1973, Phys. Fenn. **8**, 125.

Leggett, A. J., 1984, Prog. Theor. Phys. Suppl. **80**, 10.

Lévy, P., 1939, Compositio Math. **7**, 283.

Liu, S., G. K. Horton, and E. R. Cowley, 1991, Phys. Rev. B **44**, 11 714.

Liu, S., G. K. Horton, E. R. Cowley, A. R. McGurn, A. A. Maradudin, and R. F. Wallis, 1992, Phys. Rev. B **45**, 9716.

London, F., 1938, Phys. Rev. **54**, 947.

Mandelbrot, B. B., 1977, *Fractals, Form, Chance, and Dimension* (Freeman, San Francisco).

Marx, D., P. Nielaba, and K. Binder, 1992, Int. J. Mod. Phys. C **3**, 334.

Marx, D., S. Sengupta, and P. Nielaba, 1993, J. Chem. Phys.

99, 6031.

Matsubara, T., 1951, Prog. Theor. Phys. **6**, 714.

Matsubara, T., and H. Matsuda, 1956, Prog. Theor. Phys. **16**, 569.

McMillan, W. L., 1965, Phys. Rev. A **138**, 442.

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

Minc, H, 1978, *Permanents*, Encylopedia of Mathematics, Vol. 6 (Addison-Wesley, Reading, MA).

Moran, P. J., and M. Wagner, 1994, in *Proceedings of Visualization 94*, edited by R. D. Bergeron and A. E. Kaufman (IEEE Computer Society), p. 52.

Mott, N. F., 1991, Philos. Mag. Lett. **63**, 319.

Nelson, D. R., 1983, Phase Transit. Crit. Phenom. **7**, 1.

Nelson, D. R., 1988, Phys. Rev. Lett. **60**, 1973.

Nienhuis, B., 1984, J. Stat. Phys. **34**, 731.

Nose, S., 1984, J. Chem. Phys. **81**, 511.

Parrinello, M., and A. Rahman, 1984, J. Chem. Phys. **80**, 860.

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

Pierleoni, C., B. Bernu, D. M. Ceperley, and W. R. Magro, 1994, Phys. Rev. Lett. **73**, 2145.

Pitaevskii, L. P., 1961, Zh. Eksp. Teor. Fiz. **40**, 646 [Sov. Phys. JETP **13**, 451 (1961)].

Pollock, E. L., 1988, Comput. Phys. Commun. **52**, 49.

Pollock, E. L., and D. M. Ceperley, 1984, Phys. Rev. B **30**, 2555.

Pollock, E. L., and D. M. Ceperley, 1987, Phys. Rev. B **36**, 8343.

Pollock, E. L., and K. J. Runge, 1992a, J. Chem. Phys. **96**, 674.

Pollock, E. L., and K. J. Runge, 1992b, Phys. Rev. B **46**, 3535.

Reatto, L., and G. L. Masserini, 1988, Phys. Rev. B **38**, 4516.

Robkoff, H. N., and R. B. Hallock, 1981, Phys. Rev. B **24**, 159.

Robkoff, H. N., and R. B. Hallock, 1982, Phys. Rev. B **25**, 1572.

Roger, M., 1984, Phys. Rev. **30**, 6432.

Roger, M., J. H. Hetherington, and J. M. Delrieu, 1983, Rev. Mod. Phys. **55**, 1.

Rossky, P. J., J. D. Doll, and H. L. Friedman, 1978, J. Chem. Phys. **69**, 4628.

Rouse, P. E., 1953, J. Chem. Phys. **21**, 1273.

Runge, K. J., and G. V. Chester, 1988, Phys. Rev. B **38**, 135.

Runge, K. J., M. P. Surh, C. Mailhoit, and and E. L. Pollock, 1992, Phys. Rev. Lett. **69**. 3527.

Schmidt, K. E., and D. M. Ceperley, 1992, in *Monte Carlo Methods in Condensed Matter Physics*, edited by K. Binder, Topics in Applied Physics, No. 71 (Springer, Heidelberg), p. 205.

Schulman, L. S., 1981, *Techniques and Applications of Path Integration* (Wiley, New York).

Schweizer, K. S., R. M. Stratt, D. Chandler, and P. G. Wolynes, 1981, J. Chem. Phys. **75**, 1347.

Sears, V. F., E. C. Svensson, A. D. B. Woods, and P. Martel, 1979, Atomic Energy of Canada Limited Report No. AECL-6779.

Silver, R. N., D. S. Sivia, and J. E. Gubernatis, 1990, Phys. Rev. B **41**, 238 0.

Simon, B., 1979, *Functional Integration and Quantum Physics* (Academic, New York).

Sindzingre, P., D. M. Ceperley, and M. L. Klein, 1991, Phys. Rev. Lett. **67**, 1871.

Sindzingre, P., M. L. Klein, and D. M. Ceperley, 1989, Phys. Rev. Lett. **63**, 1601.

Snow, W. M., Y. Wang, and P. E. Sokol, 1992, Europhys. Lett. **19**, 403.

Sokol, P. E., T. R. Sosnick, and W. M. Snow, 1989, in *Momentum Distributions*, edited by R. N. Silver and P. E. Sokol (Plenum, New York), p. 139.

Sprik, M., M. L. Klein, and D. Chandler, 1985, Phys. Rev. B **31**, 4234.

Storer, R. G., 1968, J. Math. Phys. **9**, 964.

Svensson, E. C., P. Martel, V. F. Sears, and A. D. B. Woods, 1976, J. Can. Phys. **54**, 2178.

Svensson, E. C., V. F. Sears, A. D. B. Woods, and P. Martel, 1980, Phys. Rev. B **21**, 3638.

Takahashi, M., 1986, J. Phys. Soc. Jpn. **55**, 1952.

Takahashi, M., and M. Imada, 1984, J. Phys. Soc. Jpn. **53**, 963.

Thouless, D. J., 1965, Proc. Phys. Soc. London **86**, 893.

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

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

Umrigar, C. J., K. G. Wilson, and J. W. Wilkins, 1988, Phys. Rev. Lett. **60**, 1719.

van Kampen, N. G., 1981, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam).

Vitiello, S. A., K. J. Runge, and M. H. Kalos, 1988, Phys. Rev. Lett. **60**, 1970.

Vitiello, S. A., K. J. Runge, G. V. Chester, and M. H. Kalos, 1990, Phys. Rev. B **42**, 228.

Wang, Y., and P. E. Sokol, 1994, Phys. Rev. Lett. **72**, 1040.

Whitlock, P. A., and M. H. Kalos, 1979, J. Comput. Phys. **30**, 361.

Whitlock, P. A., and R. M. Panoff, 1987, Can. J. Phys. **65**, 1409.

Wiegel, F. W., 1986, *Introduction to Path-Integral Methods in Physics and Polymer Science* (World Scientific, Singapore).

Wilks, J., 1967, *Properties of Liquid and Solid Helium* (Clarendon, Oxford).

Yorozu, S., M. Hiroi, H. Fukuyama, H. Akimoto, H. Ishimoto, and S. Ogawa, 1992, Phys. Rev. B **45**, 12942.

Zhang, S. C., T. H. Hansson, and S. A. Kivelson, 1989, Phys. Rev. Lett. **62**, 82.