12

Quantum Monte Carlo methods

12.1 Introduction

In Chapters 1 to 4 we studied methods for solving the Schrödinger equation for many-electron systems. Many of the techniques described there carry over to other quantum many-particle systems, such as liquid helium, and the protons and neutrons in a nucleus. The techniques which we discussed there were, however, all of a mean-field type and therefore correlation effects could not be taken into account without introducing approximations. In this chapter, we consider more accurate techniques, which are similar to those studied in Chapter 10 and are based on using (pseudo-)random numbers – hence the name 'Monte Carlo' for these methods. In Chapter 10 we applied Monte Carlo techniques to classical many-particle systems; here we use these techniques for studying quantum problems involving many particles. In the next section we shall see how we can apply Monte Carlo techniques to the problem of calculating the quantum mechanical expectation value of the ground state energy. This is used in order to optimise this expectation value by adjusting a trial wave function in a variational type of approach, hence the name *variational Monte Carlo* (VMC).

In the following section we use the similarity between the Schrödinger equation and the diffusion equation in order to calculate the properties of a collection of interacting quantum mechanical particles by simulating a classical particle diffusion process. The resulting method is called *diffusion Monte Carlo* (DMC).

Then we describe the path-integral formalism of quantum mechanics, which is a formulation elaborated by Feynman, based on ideas put forward by Dirac [1], in which a quantum mechanical problem is mapped onto a classical mechanical system (at the expense of increasing the number of degrees of freedom). This classical many-particle system can then be analysed using methods similar to those used in Chapter 10. This is called the *path-integral Monte Carlo* method (PIMC).

The last section of this chapter is dedicated to a stochastic technique, based on diffusion Monte Carlo, for diagonalising the transfer matrix of a lattice spin model

on a strip, for cases where the matrix size renders even sparse matrix diagonalisation methods unusable.

Some important applications of quantum Monte Carlo methods are to the electronic structure of molecules [2], to dense helium-4 [3, 4], and to lattice spin-systems [5]. The cited literature also contains detailed accounts of the various methods.

12.2 The variational Monte Carlo method

12.2.1 Description of the method

In Chapter 3 we studied the variational method for finding the ground state and the first few excited states of the quantum Hamiltonian. This was done by parametrising the wave function – in a linear or nonlinear fashion – and then finding the minimum of the expectation value of the energy in the space of parameters occurring in the parametrised (trial) wave function. We described in some detail how this calculation can be carried out if the parametrisation is linear, and we have seen in Chapters 4 to 6 that the choice of basis functions in the linear parametrisation is crucial for the feasibility of the method. Calculating the expectation value of the energy involves integrals over the degrees of freedom of the collection of particles, which can only be carried out if the basis does not include correlations (single-particle picture) and if parts of the integration can be done analytically, for example by using Gaussian basis functions.

In this section we consider the variational method again, but we want to relax some of the above-mentioned restrictions on the trial wave functions and calculate the high-dimensional integrals using Monte Carlo methods, which are very efficient for this purpose as we have seen in Chapter 10. This is called the variational Monte Carlo approach. It should be noted that for some simple atoms, such as hydrogen and helium, the integrations can often be carried out analytically or using direct numerical integration (as opposed to MC integration). However, if there are many more electrons, these methods are no longer applicable.

Let us briefly recall the variational method in the form of an algorithm:

1. Construct the trial many-particle wave function $\psi_{\boldsymbol{\alpha}}(R)$, depending on the S variational parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_S)$. $\psi_{\boldsymbol{\alpha}}$ depends on the combined position coordinate R of all the N particles $R = \mathbf{r}_1, \ldots, \mathbf{r}_N$.
2. Evaluate the expectation value of the energy

$$\langle E \rangle = \frac{\langle \psi_{\boldsymbol{\alpha}} | H | \psi_{\boldsymbol{\alpha}} \rangle}{\langle \psi_{\boldsymbol{\alpha}} | \psi_{\boldsymbol{\alpha}} \rangle}. \tag{12.1}$$

3. Vary the parameters $\boldsymbol{\alpha}$ according to some minimisation algorithm and return to step (1).

The loop stops when the minimum energy is reached according to some criterion. It is the second step in this algorithm that we consider in this section. However, below, we shall describe a variational method in which the parameters $\boldsymbol{\alpha}$ are adjusted according to some numerical scheme within the Monte Carlo simulation.

It turns out that in realistic systems the many-body wave function assumes very small values in large parts of configuration space, so a straightforward procedure using homogeneously distributed random points in configuration space is bound to fail. This suggests that it might be efficient to use a Metropolis algorithm in which a collection of random walkers is pushed towards those regions of configuration space where the wave function assumes appreciable values. Suppose that we can evaluate $H\psi_T$ for any trial function ψ_T, which we shall always assume to be real, and let us define

$$E_L(R) = \frac{H\psi_T(R)}{\psi_T(R)} \tag{12.2}$$

(we omit the $\boldsymbol{\alpha}$-dependence of ψ_T). $E_L(R)$ is called the *local energy*: it is a function that depends on the positions of the particles and it is constant if ψ_T is the exact eigenfunction of the Hamiltonian. The more closely ψ_T approaches the exact wave function (apart from a multiplicative constant), the less strongly will E_L vary with R.

The expectation value of the energy can now be written as

$$\langle E \rangle = \frac{\int dR\ \psi_T^2(R) E_L(R)}{\int dR\ \psi_T^2(R)}. \tag{12.3}$$

Let us now construct a Metropolis-walk in the same spirit as in ordinary Monte Carlo calculations, but now with a stationary distribution $\rho(R)$ given by

$$\rho(R) = \frac{\psi_T^2(R)}{\int dR'\ \psi_T^2(R')}. \tag{12.4}$$

The procedure is now as follows.

Put N walkers at random positions;
REPEAT
  Select next walker;
  Shift that walker to a new position, for example by moving one
    of the particles in the system within a cube with a suitably
    chosen size d;
  Calculate the fraction $p = [\psi_T(R')/\psi_T(R)]^2$, where R' is the new and
    R the old configuration;
  If $p < 1$ the new position is accepted with probability p;
  If $p \geq 1$ the new position is accepted;
UNTIL finished.

The expectation value of the local energy is now calculated as an average over the samples generated in this procedure, excluding a number of steps at the beginning, necessary to reach equilibrium. The decision to stop the simulation is based on the precision achieved and on the available processor time.

The algorithm should work in principle with a single walker. However, chances are that this walker gets stuck in one favourable region surrounded by barriers which are difficult to overcome. Using a large collection of walkers reduces this effect.

12.2.2 Sample programs and results

We demonstrate the VMC approach with some simple programs. Here and in the rest of this chapter, when dealing with many-particle systems, we shall assume units of mass, distance and energy to be such that the kinetic energy operator occurs in the Schrödinger equation as $-\nabla^2/2$.

We start with the harmonic oscillator in one dimension, described by the Hamiltonian (in dimensionless units):

$$H\psi(x) = \left[-\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2\right]\psi(x). \tag{12.5}$$

The exact solution for the ground state is given by $\exp(-x^2/2)$ with energy $E_G = 1/2$; we shall use the trial function $\exp(-\alpha x^2)$. The exact solution lies therefore in the variational subspace. The local energy is given by

$$E_L = \alpha + x^2\left(\frac{1}{2} - 2\alpha^2\right). \tag{12.6}$$

For $\alpha = 1/2$ the local energy is $1/2$, independent of the position, and we shall certainly find an energy expectation value $1/2$ in that case (this might happen even when the program contains errors!). The crucial test is whether this energy expectation value is a minimum as a function of α. In Table 12.1 we show that this is indeed the case. We also show the variance of the energy. This quantity will be small if E_L is rather flat, and this will be the case when ψ_T is close to the exact ground state: the closer ψ_T is to the ground state wave, the smaller the variance, and this quantity reaches its minimum value at the variational minimum of the energy itself. Again, in this particular case where the trial wave function can become equal to the exact ground state, the variance becomes zero. From the table we see that the variance does indeed decrease to 0 when the ground state is approached. Interestingly, for this simple case, it is possible to calculate the expectation value of the energy as a function of α by integrating the local energy weighted by ψ_T^2. The Gaussian form of the trial wave function makes the integral solvable with the result

$$E_v = \frac{1}{2}\alpha + \frac{1}{8\alpha}. \tag{12.7}$$

Table 12.1. *Variational Monte Carlo energies.*

	Harmonic oscillator				Hydrogen atom			Helium atom		
α	$\langle E\rangle$	var($\langle E\rangle$)	E_v	var$(E)_v$	α	$\langle E\rangle$	var($\langle E\rangle$)	α	$\langle E\rangle$	var($\langle E\rangle$)
0.4	0.5124(1)	0.02521(5)	0.5125	0.0253125	0.8	−0.4796(2)	0.0243(6)	0.05	−2.8713(4)	0.1749(2)
0.45	0.50276(4)	0.00556(2)	0.50278	0.00557	0.9	−0.4949(1)	0.0078(2)	0.075	−2.8753(4)	0.1531(2)
1/2	1/2	0	1/2	0	1.0	−1/2	0	0.10	−2.8770(3)	0.1360(2)
0.55	0.50232(6)	0.00454(1)(1)	0.5022727	0.0045558	1.1	−0.4951(2)	0.0121(4)	0.125	−2.8780(4)	0.1223(2)
0.6	0.5084(1)	0.0168(4)	0.508333	0.0168056	1.2	−0.4801(3)	0.058(2)	0.15	−2.8778(3)	0.1114(2)
								0.175	−2.8781(3)	0.1028(2)
								0.20	−2.8767(4)	0.0968(2)
								0.25	−2.8746(10)	0.0883(2)

VMC energies are given for the harmonic oscillator, the hydrogen atom and the helium atom for various values of the variational parameters. In each case, 400 walkers have been used and 30 000 displacements per walker were attempted. The first 4000 of these were removed from the data to ensure equilibrium. The expectation value $\langle E\rangle$ of the ground state energy is given, together with the variance in this quantity, var($\langle E\rangle$). For the harmonic oscillator, the analytical values for the energies and variance are also given (E_v and var$(E)_v$).

The same can be done for the variance with the result

$$\mathrm{var}(E)_v = \frac{(1-4\alpha^2)^2}{32\alpha^2}. \tag{12.8}$$

The Monte Carlo results match the analytical values as is clear from the table. Also in Table 12.1 we show results for the hydrogen atom with the Hamiltonian

$$H = -\frac{1}{2}\nabla^2 - \frac{1}{r}. \tag{12.9}$$

The exact ground state with energy $E = -1/2$ is given as e^{-r}; we take variational trial functions of the form $e^{-\alpha r}$, so that the ground state is again incorporated in the variational subspace. Although we could consider the present problem as a one-dimensional one by using the spherical symmetry of the potential and the ground state wave function, we shall treat it here as a fully three-dimensional problem to illustrate the general approach. For this case, the analytical values of the average local energy and variance can also be calculated. This is left as an exercise for the reader.

The local energy is given by

$$E_L(r) = -\frac{1}{r} - \frac{1}{2}\alpha\left(\alpha - \frac{2}{r}\right). \tag{12.10}$$

It is seen from Table 12.1 that the energy is minimal at the ground state and that its variance vanishes there too.

Finally we consider the helium atom, which we have already studied extensively in Chapters 4 and 5. Constructing good trial functions is a problem on its own – here we shall use the form:

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = e^{-2r_1}e^{-2r_2}e^{r_{12}/[2(1+\alpha r_{12})]} \tag{12.11}$$

where $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$. This function consists of a product of two atomic one-electron orbitals and a correlation term. The local energy now has the form:

$$E_L(\mathbf{r}_1, \mathbf{r}_2) = -4 + (\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2)\cdot(\mathbf{r}_1 - \mathbf{r}_2)\frac{1}{r_{12}(1+\alpha r_{12})^2} - \frac{1}{r_{12}(1+\alpha r_{12})^3} - \frac{1}{4(1+\alpha r_{12})^4} + \frac{1}{r_{12}} \tag{12.12}$$

With $\hat{\mathbf{r}}$ we denote a unit vector along $\mathbf{r}$, and r_{12} is the distance between the two electrons. Energies and variances are also displayed in Table 12.1. The variance does not have a sharp minimum for the same value of α as the energy. The reason is that most of the variance is due to the trial wave function not being exact, even for the best value of α. The optimum value of the energy, $-2.878\,1 \pm 0.000\,5$, should be compared with the Hartree-Fock value of $-2.861\,7$ a.u. and the DFT value

of -2.83 a.u., and with the exact value of $-2.903\,7$ a.u. The VMC value can obviously be improved by including more parameters in the wave function. The wave function is apparently not perfect. One of its deficits can be appreciated by considering the case where one of the electrons is far away from the nucleus and the other electron. Then the trial wave function depends on the position of this particle like the wave function of the helium ion, i.e. it is the asymptotic wave function for an electron in the field of a $Z = 2$ nucleus. In reality, however, the wave function should 'see' a charge $Z = 1$ as the other electron shields off one unit charge.

It is possible to adjust the value of the parameters α in these simulations 'on the fly' [6]. To this end, we need a minimum finder. The most efficient minimum finders use the gradient of the function to be minimised (see Appendix A). This is a problem, as a finite difference calculation of the gradient is bound to fail: the derivatives of stochastic variables are subject to large numerical errors. However, from the analytic derivative of the wave function with respect to α, we can sample this derivative over the population of walkers. From (12.3) we see that

$$\frac{\mathrm{d}E}{\mathrm{d}\alpha} = 2\left(\left\langle E_{\mathrm{L}}\frac{\mathrm{d}\ln\psi_{\mathrm{T}}}{\mathrm{d}\alpha}\right\rangle - E\left\langle\frac{\mathrm{d}\ln\psi_{\mathrm{T}}}{\mathrm{d}\alpha}\right\rangle\right). \tag{12.13}$$

Using a simple damped steepest decent method:

$$\alpha_{\mathrm{new}} = \alpha_{\mathrm{old}} - \gamma\left(\frac{\mathrm{d}E}{\mathrm{d}\alpha}\right)_{\mathrm{old}}, \tag{12.14}$$

the method then finds the optimal value (and therefore also the energy) for α. This method works remarkably well for the harmonic oscillator, where, starting from $\alpha = 1.2$, the correct value $\alpha = 0.5$ is found in a small fraction of the time needed for accurately evaluating one of the points in Table 12.1. However, the success in this particular case is partly due to the exact solution being in the family of solutions considered. The method is generalised straightforwardly to more parameters. It has been applied successfully to electrons in quantum dots [6].

The reader is invited to write the programs described and check the results with those given in Table 12.1.

12.2.3 *Trial functions*

The trial wave function for helium, Eq. (12.11), is the two-particle version of the general ground state trial wave function used in quantum Monte Carlo (QMC) calculations of fermionic systems:

$$\psi(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \Psi_{\mathrm{AS}}(\mathbf{x}_1,\ldots,\mathbf{x}_N)\exp\left[\frac{1}{2}\sum_{i,j=1}^{N}\phi(r_{ij})\right]. \tag{12.15}$$

Ψ_{AS} is the Slater determinant (see Chapter 4) and ϕ is a function which contains the two-particle correlation effects. For identical bosons, all the minus-signs in the determinant are replaced by pluses. The particular form we chose in the helium case is a simple form of a class called Padé–Jastrow wave functions [7]. Inclusion of three and four point correlations is obviously possible. We shall not go into the problem of finding the best Slater determinants and ϕ-functions but restrict ourselves to a short discussion of the requirements which we can derive for special particle configurations – these are the 'cusp conditions': boundary conditions satisfied at the points where the potential diverges. Near these points the kinetic and potential energy contributions of the Hamiltonian are both very large, and they should cancel out for a large part. This leads to large statistical fluctuations which are avoided by respecting the cusp conditions. In the next section we shall see that these cusp conditions are essential for trial wave functions used in the DMC method. We have already dealt with a similar problem in Chapter 2 of this book, when we found appropriate boundary conditions for the numerical solution of the radial Schrödinger equation with a Lennard–Jones potential, which diverges strongly at $r = 0$. Now we consider singularities in the Coulomb potential.

In the helium atom, the potential diverges when one of the electrons approaches the nucleus, or when the electrons are close to each other. The Schrödinger equation can be solved analytically for these configurations since the Coulomb potential dominates all other terms except the kinetic one. Suppose that one of the electrons, labelled i, is very close to a nucleus (which we take at the origin) with charge Z. In that case the Schrödinger equation becomes approximately

$$\left[-\frac{1}{2}\nabla_i^2 - \frac{Z}{r_i}\right]\psi(\mathbf{r}_1,\ldots,\mathbf{r}_N) = 0. \tag{12.16}$$

Writing out the kinetic energy in spherical coordinates of particle i, we arrive at a radial Schrödinger equation of the form ($r = r_i$)

$$\left[\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}}{\mathrm{d}r} + \frac{2Z}{r} - \frac{l(l+1)}{r^2}\right]R(r) = 0. \tag{12.17}$$

If, as is usually the case, the wave function is radially symmetric in $\mathbf{r}_i$ for r_i small, we have exclusively an $l = 0$ contribution, and the two terms containing the factor $1/r$ must cancel (the first term does not contribute for a function which is regular at the origin). For $R(0) \neq 0$ this leads to

$$\frac{1}{R}\frac{\mathrm{d}R}{\mathrm{d}r} = -Z, \qquad r = 0; \tag{12.18}$$

so that $R(r) = \exp(-Zr)$.

For $l > 0$, the radial wave function is written in the form $r^l\rho(r)$ where ρ does not vanish at $r = 0$. Analysing this in a way similar to the $l = 0$ case leads to the cusp condition

$$\frac{1}{\rho(r)}\frac{\mathrm{d}\rho(r)}{\mathrm{d}r} = -\frac{Z}{l+1}. \tag{12.19}$$

Note that this form is the same as (12.18) if we put $l = 0$.

Another cusp condition is found for two electrons approaching each other. Considering the trial wave function of the helium atom, Eq. (12.11), we see that it is the dependence on the separation r between the two electrons which must incorporate the correct behaviour in this limit. The resulting radial equation for the r dependence is the same as for the electron–nucleus cusp except for the $-Z/r$ potential being replaced by $1/r$ (the Coulomb repulsion between the two electrons), and the kinetic term being twice as large (because the reduced mass of the two electrons is half the electron mass):

$$\left[2\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{4}{r}\frac{\mathrm{d}}{\mathrm{d}r} - \frac{2}{r} - \frac{l(l+1)}{r^2}\right]R(r) = 0. \tag{12.20}$$

The cusp condition, written in terms of $\rho(r) = r^{-l}R(r)$, is therefore

$$\frac{1}{\rho(r)}\frac{\mathrm{d}\rho(r)}{\mathrm{d}r} = \frac{1}{2(l+1)}. \tag{12.21}$$

The right hand side reduces to $1/2$ in the usual case of an s-wave function ($l = 0$). For like spins, the value of the wave function must vanish if the particles approach each other; therefore the wave function with lowest energy is a p-state and the right hand side will reduce to 1/4. For a general system, containing more than two electrons, we have this cusp condition for each electron pair ij. It is recommended to have a look at Problem 12.5 to see how cusp conditions are implemented in practice.

12.2.4 *Diffusion equations, Green functions and Langevin equations*

In the following sections we shall discuss several QMC methods in which the ground state of a quantum Hamiltonian is found by simulating a diffusion process. In the next section for example, we shall use such a simulation to improve on the variational method described above. In this section, we give a brief overview of diffusion and the related equations.

Consider a one-dimensional discrete axis with sites located at na, with integer n. We place a random walker on a site, and this walker jumps from site to site with time intervals h. The walker can only jump from a site to its left or right neighbour. Both jumps have a probability α, and the walker remains at the current position with probability $1 - 2\alpha$. This is clearly a Markov process as described in Section 10.3. We are interested in the probability $\rho(x,t)$ to find the walker at site $x = na$ at time $t = mh$, where n and m are both integer. This probability satisfies the master equation of the Markov process:

$$\rho(x,t+h) - \rho(x,t) = \alpha[\rho(x+a,t) + \rho(x-a,t) - 2\rho(x,t)] \approx \alpha a^2\frac{\partial^2\rho(x,t)}{\partial x^2}. \tag{12.22}$$

For small h, the left hand side can be written as $h(\partial\rho/\partial t)$, and defining $\gamma = a^2\alpha/h$, we can write the continuum form of the master equation (for small a) as

$$\frac{\partial\rho(x,t)}{\partial t} = \gamma\frac{\partial^2\rho(x,t)}{\partial x^2}. \tag{12.23}$$

This equation is called the *diffusion equation*: it describes how the probability distribution of a walker evolves in time. It may equivalently be interpreted as the density distribution for a large collection of independent walkers.

Consider the following function:

$$G(x,y;t) = \frac{1}{\sqrt{4\pi\gamma t}}\mathrm{e}^{-(x-y)^2/(4\gamma t)}. \tag{12.24}$$

This function has the following properties:

- Considered as a function of y and t, keeping x fixed, it is a solution of the diffusion equation for $t > 0$.
- For $t \to 0$, G reduces to a delta-function:

$$G(x,y;t) \to \delta(x-y) \text{ for } t \to 0. \tag{12.25}$$

G is called the *Green's function* of the diffusion equation. This function can be used to write the time evolution of any initial distribution $\rho(x,0)$ of this equation in integral form:

$$\rho(y,t) = \int \mathrm{d}x\, G(x,y;t)\rho(x,0), \tag{12.26}$$

which can easily be checked using the properties of G. Inspection of the Green's function shows that it is normalised, that is, $\int \mathrm{d}y\, G(x,y;t) = 1$, independent of x and t.

The Green's function can be interpreted as the probability distribution of a single walker which starts off at position x at $t = 0$. We can use G to construct a new Markov process corresponding to the diffusion equation. We discretise the time in steps Δt. We start with a walker localised at x at $t = 0$. Then we move this walker to a new position y at time Δt with probability distribution $G(x,y;\Delta t)$. From this,

we move the walker to a new position z at time $2\Delta t$ with probability distribution $G(y, z; \Delta t)$. We have therefore a Markov process with transition probability given by G:

$$T_{\Delta t}(x \to y) = G(x, y; \Delta t). \tag{12.27}$$

Using the properties of the Green's function it can be shown that the detailed balance condition for the master equation for the Markov process leads to the integral form (12.26), so that the Markov process does indeed model the diffusion process described by (12.23) (check this). The difference between this process and the previous one on the discrete lattice is that we now use the continuum solution of the former version, which should be much more efficient, as a single step in the continuum diffusion process represents a large number of steps in the discrete diffusion process. The Markov process described by (12.27) can be summarised by the equation

$$x(t + \Delta t) = x(t) + \eta\sqrt{\Delta t}, \tag{12.28}$$

where η is a Gaussian random variable with variance 2γ:

$$P(\eta) = \frac{1}{\sqrt{4\pi\gamma}} e^{-\eta^2/(4\gamma)}. \tag{12.29}$$

This result can be understood by realising that a step in the Markov process (12.27) is distributed according to a Gaussian with width $\sqrt{2\gamma\Delta t}$. In this form, the process is recognised as a Langevin equation for discrete time. Note that a random momentum rather than a random force is added at each step, in contrast to the Langevin equation discussed in Section 8.8.

The general form of the diffusion equation is

$$\frac{\partial \rho}{\partial t} = \mathcal{L}\rho(x, t), \tag{12.30}$$

where $\mathcal{L}$ is a second order differential operator. The formal solution of this equation with a given initial distribution $\rho(x, 0)$ can be written down immediately:

$$\rho(x, t) = e^{t\mathcal{L}}\rho(x, 0) \tag{12.31}$$

but as this involves the exponential of an operator (which is to be considered as an infinite power series), it is not directly useful. Using Dirac notation, the Green's function can formally be written as

$$G(x, y; t) = \langle x|e^{t\mathcal{L}}|y\rangle, \tag{12.32}$$

which indeed satisfies the equation (12.31) as a function of y and t, and which reduces to $\delta(x - y)$ for $t = 0$. The diffusion equation can only be used to construct a Markov chain if the Green's function is normalised, in the sense that $\int dy\ G(x, y; t) = 1$, independent of t. This is not always the case, as we shall now see.

A particular diffusion equation which we shall encounter later in this chapter is

$$\frac{\partial \rho}{\partial \tau} = \frac{1}{2}\frac{\partial^2 \rho(x, \tau)}{\partial x^2} - V(x)\rho(x, \tau). \tag{12.33}$$

This looks very much like the one-dimensional time-dependent Schrödinger equation for a zero-mass particle; in fact, this equation is recovered when we continue the time analytically into imaginary time $\tau = it$ (we use τ for imaginary time). Using (12.31), we can write the solution as

$$\rho(x, \tau) = e^{\tau(-K-V)}\rho(x, 0) \tag{12.34}$$

where K is the kinetic energy operator $K = p^2/2 = -1/2(\partial^2/\partial x^2)$ (p is the momentum operator $p = -i(\partial/\partial x)$ of quantum mechanics). The exponent cannot be evaluated because the operators K and V do not commute. However, we might neglect Campbell–Baker–Hausdorff (CBH) commutators – this is only justified when τ is small. To emphasise that the following is only valid for small τ, we shall use the notation $\Delta\tau$ instead of τ. We have

$$e^{-\Delta\tau(K+V)} = e^{-\Delta\tau K}e^{-\Delta\tau V} + \mathcal{O}(\Delta\tau^2) \tag{12.35}$$

where the order $\Delta\tau^2$ error term results from the neglect of CBH commutators. To find the Green's function explicitly, we must find the matrix element of the exponential operator on the right hand side. The term involving the potential is not a problem as this is simply a function of x. It remains then to find the matrix elements of the kinetic operator:

$$G_{\text{Kin}}(x, y; \Delta\tau) = \langle x|e^{-\Delta\tau\hat{p}^2/2}|y\rangle \tag{12.36}$$

where $\hat{p}$ is the momentum operator – we have used the caret $\hat{}$ to distinguish the operator from its eigenvalue.

The Green's function can be evaluated explicitly by inserting two resolutions of the unit operator of the form $\int dp\ |p\rangle\langle p|$ and using the fact that

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi}} e^{ipx} \quad (\hbar \equiv 1). \tag{12.37}$$

As the kinetic operator is diagonal in the p-representation, the matrix element is then found simply by performing a Gaussian integral. The result is

$$G_{\text{Kin}}(x, y; \Delta\tau) = \frac{1}{\sqrt{2\pi\Delta\tau}} e^{-(x-x')^2/(2\Delta\tau)}. \tag{12.38}$$

This form is recognised as the Green's function of the simple diffusion equation; indeed our imaginary-time Schrödinger equation reduces to this equation for $V \equiv 0$, and therefore the kinetic part of our Green's function should precisely be equal to the Green's function of the simple diffusion equation. We have derived this form

explicitly here, because we need to find the Green's function for a more complicated type of diffusion equation along the same lines below.

The full Green's function for the diffusion equation (12.33) reads:

$$G(x, y; \Delta\tau) = G_{\mathrm{Kin}}(x, y; \Delta\tau)\mathrm{e}^{-\Delta\tau V(y)} + \mathcal{O}(\Delta\tau^2). \tag{12.39}$$

Unfortunately, the term involving the potential destroys the normalisation of the full Green's function, and this prevents us from using it to construct a Markov chain evolution, which is convenient, if not essential, for a successful simulation as we shall see later. We can make the transition rate Markovian by normalising it, which can be done by multiplying the Green's function by a suitable prefactor $\exp(\tau E_{\mathrm{T}})$. Of course we do not know beforehand what the value of this prefactor is, but we shall describe methods for sampling its value in Section 12.3. The new, normalised, Green's function is no longer the proper Green's function for Eq. (12.33), but for a modified form of this equation, in which the potential has been shifted by an amount E_{T}:

$$\frac{\partial\rho}{\partial\tau} = \frac{1}{2}\frac{\partial^2\rho(x,\tau)}{\partial x^2} - [V(x) - E_{\mathrm{T}}]\rho(x,\tau). \tag{12.40}$$

If we choose E_{T} such that the Green's function is normalised, it describes a Markov process, hence there will be an invariant distribution. This invariant distribution is determined by Eq. (12.40), which for stationary distributions reduces to

$$-\frac{1}{2}\frac{\partial^2\rho(x)}{\partial x^2} + V(x)\rho(x) = E_{\mathrm{T}}\rho(x), \tag{12.41}$$

which is the stationary Schrödinger equation.

For many problems, it is convenient to construct some Markovian diffusion process which has a predefined distribution as its invariant distribution. This turns out to be possible, and the equation is called the Fokker–Planck (FP) equation. It has the form

$$\frac{\partial\rho(x,t)}{\partial t} = \frac{1}{2}\frac{\partial}{\partial x}\left[\frac{\partial}{\partial x} - F(x)\right]\rho(x,t). \tag{12.42}$$

The 'force' $F(x)$ is related to the invariant distribution $\rho(x)$: the relation is given by

$$F(x) = \frac{1}{\rho(x)}\frac{\mathrm{d}\rho(x)}{\mathrm{d}x}. \tag{12.43}$$

It can easily be checked that $\rho(x)$ satisfies (12.42) when the time derivative occurring in the left hand side of this equation is put equal to zero.

The Green's function can be found along the same lines as that of the kinetic part of the Green's function for the imaginary-time Schrödinger equation. We must work out

$$G(x, y; t) = \langle x|\mathrm{e}^{-\Delta t\hat{p}[\hat{p} - iF(\hat{x})]/2}|y\rangle. \tag{12.44}$$

We again separate the exponent into two terms, one containing $\hat{x}$ and the other $\hat{p}$, at the expense of an $\mathcal{O}(\Delta t^2)$ error. Calculating Gaussian Fourier transforms as before, we obtain the result:

$$G(x, y; \Delta t) = \frac{1}{\sqrt{2\pi\,\Delta t}}\mathrm{e}^{-[y - x - F(x)\Delta t/2]^2/(2\Delta t)}. \tag{12.45}$$

Note that this expression is a first order approximation in Δt of the exact Green's function. This is normalised, and we can therefore use it again for constructing a Markov chain. This is done by moving the random walker first from its old position x to the position $x + F(x)\Delta t/2$ and then adding a random displacement $\eta\sqrt{\Delta t}$, where η is drawn from a Gaussian distribution with a variance 1 (see Eq. (12.29)). In formula, the method reads

$$x(t + \Delta t) = x(t) + \Delta t F[x(t)]/2 + \eta\sqrt{\Delta t}, \tag{12.46}$$

so it is a discrete Langevin equation with 'force' F.

We end this section with a few remarks. First, all results can be extended straightforwardly to higher dimensions. Using a $3N$-dimensional variable R instead of the one-dimensional variable x (R denotes the positions of a set of particles in three dimensions as usual), the Green's function of the simple diffusion equation Eq. (12.23) with $\gamma = 1/2$ is

$$G_{3N}(R, R'; t) = \frac{1}{(2\pi t)^{3N/2}}\mathrm{e}^{-(R' - R)^2/(2t)}. \tag{12.47}$$

The Green's function of the Fokker–Planck equation (12.42) becomes

$$G_{3N}(R, R'; \Delta t) = \frac{1}{(2\pi\,\Delta t)^{3N/2}}\mathrm{e}^{-[R' - R - \Delta t\mathbf{F}(R)/2]^2/(2\Delta t)}, \tag{12.48}$$

where $\mathbf{F}(R)$ is a three-dimensional vector, given by

$$\mathbf{F}(R) = \nabla_R\rho(R)/\rho(R). \tag{12.49}$$

You might have been surprised by the way in which the exponential containing noncommuting operators was split in Eq. (12.35). After all, the following splitting

$$\mathrm{e}^{-\Delta\tau(V+K)} = \mathrm{e}^{-\Delta\tau V/2}\mathrm{e}^{-\Delta\tau K}\mathrm{e}^{-\Delta\tau V/2} + \mathcal{O}(\Delta\tau^3) \tag{12.50}$$

is more accurate: you can check that the first order CBH commutator vanishes, hence the $\mathcal{O}(\Delta\tau^3)$ error. The reason we use the simpler splitting (12.35) is that diffusion steps are carried out *successively*, hence the rightmost term in the right hand side of (12.50) at one step combines with the leftmost term at the next step, so that the total effect of the more accurate splitting is reduced to a different first and final step. This difference is, however, of the same order of magnitude as the accumulated error of the sequence of steps, and therefore it does not pay to use (12.50).

12.2.5 The Fokker–Planck equation approach to VMC

The VMC method described in Sections 12.2.1 and 12.2.2 has an important disadvantage: typical many-particle wave functions are very small in large parts of configuration space and very large in small parts of configuration space. This means, first, that we might have difficulty in finding the regions where the wave function is large, and second, that attempted moves of walkers from a favourable region (where the wave function is large) will be rejected when they move out of that region. Having a substantial fraction of rejected moves is part of any Metropolis Monte Carlo scheme, and we could live with that if there did not exist a more efficient approach, based on the Fokker–Planck equation described in the previous section.

In this method we try to sample the function $\rho(R) = \psi_T^2(R)$ rather than the trial function $\psi_T(R)$ itself: that is, we use

$$\mathbf{F} = 2\nabla_R \psi_T(R)/\psi_T(R) \tag{12.51}$$

in the FP equation.

The distribution $\rho(R,t)$ can be sampled by simulating a diffusion process. The algorithm is close to that of ordinary VMC. Now we let a collection of walkers diffuse with probabilities given by the Green's function (12.45):

Put N walkers at random positions;
REPEAT
  Select next walker;
  Shift that walker from its current position R to $R + \mathbf{F}(R)\Delta t/2$;
  Displace that walker by an amount $\eta\sqrt{\Delta t}$, where η is a
    random vector with a Gaussian distribution (see (12.29) and (12.28));
UNTIL finished.

We see that there is no acceptance/rejection step; this causes the gain in efficiency when using the FP approach.

Note that we have made a time-step error of order $(\Delta t)^2$. It is possible to eliminate this error by combining this Langevin approach with a Metropolis procedure. The point is that we know the form of stationary distribution ρ (it is the square of the trial function ψ_T), and the Langevin process leads to a distribution which is close to but not exactly equal to this distribution. The Metropolis algorithm can give us the desired distribution ρ by acceptance/rejection of the Langevin steps, which themselves are considered as trial moves in the Metropolis algorithm. Referring back to Section 10.3, we call the transition probability of the Langevin equation $\omega_{RR'} = G(R,R';\Delta t)$, where G is given in (12.48). This is not symmetric in R and R' as F depends only on R, and therefore we have to use the generalised Metropolis algorithm, described at the very end of Section 10.3. The Langevin trial move is

accepted with probability $\min(1, q_{RR'})$, where

$$q_{RR'} = \frac{\omega_{R'R}\rho(R')}{\omega_{RR'}\rho(R)}. \tag{12.52}$$

Note that the fraction $\omega_{R'R}/\omega_{RR'}$ is in equilibrium approximately equal to the ratio $\rho(R)/\rho(R')$ – if no time step error was made in constructing $\omega_{RR'}$, they would have been exactly equal – so $q_{RR'}$ is always close to 1. The acceptance rate is therefore always high when Δt is taken small, and the method is very efficient. The Metropolis acceptance/rejection step is merely a correction for the time step discretisation error made in the Langevin procedure.

The implementation of the algorithm is straightforward. The resulting energies must be the same as for the standard VMC method, but the error bars are smaller. As an example, an MC simulation for the harmonic oscillator using 300 walkers which perform 3000 steps and $\alpha = 0.4$ yields for the energy expectation in the ordinary VMC program value $E = 0.51 \pm 0.03$, to be compared with $E = 0.515 \pm 0.006$ in the Fokker–Planck program.

Variational Monte Carlo has the advantage that it is simple and straightforward. An important disadvantage is that it relies on the quality of the trial function, hence subtle but important physical effects are sometimes neglected when they are not taken into account when constructing the trial function.

12.3 Diffusion Monte Carlo

12.3.1 Simple diffusion Monte Carlo

The second quantum Monte Carlo method that we consider is the so-called *diffusion* or *projector* Monte Carlo method, abbreviated as DMC. This method does not use variational principles for obtaining ground state properties, but as we shall see, the convergence rate of the practical version of this method relies heavily on the accuracy of the trial functions. The idea of this method has already been sketched in Section 12.2.4. We use the imaginary time form of the time-dependent Schrödinger equation. This is a diffusion equation with a potential. We use the Green's function in the 'normalised' form, i.e. with the normalisation factor $\exp(-\Delta\tau E_T)$ present:

$$G(R,R';\Delta\tau) = \mathrm{e}^{-\Delta\tau[V(R)-E_T]}\frac{1}{\sqrt{2\pi\Delta\tau}}\mathrm{e}^{-(R-R')^2/(2\pi\Delta\tau)} + \mathcal{O}(\Delta\tau^2). \tag{12.53}$$

This Green's function is a short-time approximation of the imaginary-time operator $\exp[-\tau(H+E_T)]$. If we resolve this operator in its eigenstates $|\phi_n\rangle$, we obtain

$$\mathrm{e}^{-\tau(H-E_T)} = \sum_n |\phi_n\rangle \mathrm{e}^{-\tau(E_n-E_T)}\langle\phi_n|. \tag{12.54}$$

For large τ the ground state energy E_G dominates in the sum by a factor $\exp[-\tau(E_1 - E_G)]$; therefore it acts as a projector onto the ground state (for large enough times).

As we have the explicit form of the time-evolution operator at our disposal only in a short-time approximation, we have to perform many short time steps before the distribution will approach the ground state wave function.

In the simulation, a collection of walkers diffuses through configuration space. Every diffusion step consists of two stages: a diffusion step and a *branching step*. In the diffusion step, the walkers are moved to a new position with a transition rate given by the diffusive part of the Green's function, i.e. the part due to the kinetic energy. The term involving the potential is dealt with in the second stage. Suppose we were to assign a weight to each walker, then the effect of the potential term could be taken into account by multiplying this weight for a walker which has arrived at a position R' by a factor $\exp\{-\Delta\tau[V(R') - E_T]\}$.[1] It turns out that this procedure is not very efficient. In the end quite a few walkers might have moved to unfavourable regions and represent small weight, but they require similar computational effort to the more favourable ones. This problem was previously encountered in Section 10.6. It would be more efficient to use computational effort proportional to the significance of the region probed by a particular walker. This is possible by a 'birth and death', or 'pruning and enrichment' (Section 10.6) or *branching* process: poor walkers die, favourable ones give rise to new walkers. More precisely, if a walker moves from a point R to a new point R', we calculate $q = \exp\{-\Delta\tau[V(R') - E_T]\}$. If $q < 1$, the walker survives with a probability q and dies with probability $1-q$. If $q > 1$, the walker gives birth to either $[q-1]$ or $[q]$ new ones at R, where $[q]$ represents the integer part (truncation) of q. The probability for having $[q]$ new walkers is given by $q - [q]$, and $[q-1]$ new walkers will come into existence with the complementary probability $1 + [q] - q$. An efficient way of coding this is to add a uniform random number r between 0 and 1 to q: for $s = q + r$, $[s]$ new walkers are created; if $[s] = 0$ then the walker is deleted.

Finally, we must specify how E_T is found. Remember that this value is ideally chosen such as to normalise the overall transition rate in the process. This is necessary to prevent the population from growing or decreasing steadily. A growing population would cause a steady increase in the computer time per diffusion step, whereas a decrease leads to bad statistics, if not a vanishing population! The energy E_T is in fact determined by keeping track of the change in population and adjusting it at each step in order to keep the population stable. The average value of E_T after many steps will then converge to the ground state energy as we have already seen in Section 12.2.4. Suppose we have a target number of $\tilde{M}$ walkers in our simulation

[1] It is also possible to multiply the weight by $\exp\{-\tau[(V(R') + V(R))/2 - E_T]\}$, which corresponds to the symmetric distribution of the potential terms in the Green's function as in (12.50).

and that after the last branching step their actual number is M, then we adjust E_T as

$$E_T = E_0 + \alpha \ln\left(\frac{\tilde{M}}{M}\right) \tag{12.55}$$

where E_0 is close to the ground state energy (our 'best estimate'), and α is some small parameter.

In an algorithmic form, the resulting procedure can be presented as follows:

```
Put the walkers at random positions in configurational space;
REPEAT
   FOR all walkers DO
      Shift walker from its position R to a new position R'
            according to the Gaussian transition probability (12.24);
      Evaluate q = exp{-Δτ[V(R') - Δτ E_T]};
      Eliminate the walker or create new ones at R',
            depending on s = q + r, where r is random,
            uniform between 0 and 1;
   END FOR;
   Update E_T;
UNTIL finished.
```

The major difference with the variational Monte Carlo method described in the previous section is that the present method does not rely on a trial function and therefore the results have no systematic error due to the trial function being (in general) not exact. There is, however, an error due to the fact that we have split the time-evolution operator into two parts, one depending on the kinetic energy and the other on the potential, by neglecting CBH commutators. By reducing $\Delta\tau$ we can make this error arbitrarily small, but the convergence speed will be reduced accordingly. In Section 12.3.3, we shall describe a Metropolis algorithm to correct for the discretisation error.

The population itself should represent the ground state wave function. For a one-dimensional problem (or a radially symmetric three-dimensional problem) this can be checked by constructing a histogram in which we record the frequencies with which the various positions are occupied. Below we shall give some results of DMC simulations for the harmonic oscillator and the helium atom.

The DMC procedure outlined here might fail in some cases. The distribution of walkers can only represent a density which is positive everywhere. Therefore, it can sample the ground wave function only if the latter is everywhere positive. Fortunately, the ground state of a boson system is indeed everywhere positive. However, for fermions this is no longer the case. Moreover, the Green's function is no longer positive in that case and it is not clear how to perform the diffusion, as

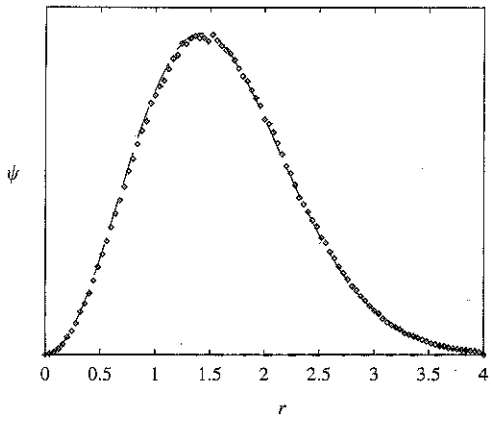

Figure 12.1. Ground state wave function (times r^2) for the three-dimensional harmonic oscillator as resulting from the DMC calculation (dots) compared with the exact form, scaled to match the numerical solution best.

the transition probability should be positive. This is called the *fermion problem*. We shall come back to this later. Another problem arises when the interaction potential assumes strongly negative values. This will be discussed in some detail in the next section and then we shall consider a refinement of the DMC which is not susceptible to this problem.

12.3.2 Applications

We apply the DMC procedure first to the three-dimensional harmonic oscillator. The exact ground state wave function is given by

$$\psi(r) = \frac{1}{(2\pi)^{3/2}} \mathrm{e}^{-r^2/2}; \tag{12.56}$$

the energy is 3/2 (in dimensionless units). It should be noted that the probability distribution for finding a walker at a distance r from the origin is given by the wave function times r^2, because the volume of a spherical shell of thickness $\mathrm{d}r$ is $2\pi r^2 \mathrm{d}r$. For an average population of 300 walkers executing 4000 steps and a time step $\tau = 0.05$, we find $E_\mathrm{G} = 1.506 \pm 0.015$, to be compared with the exact value $1\frac{1}{2}$. The distribution histogram is shown in Figure 12.1, together with the exact wave function, multiplied by r^2 and scaled in amplitude to fit the DMC results best. Ground state energy and wave function are calculated with reasonable accuracy. Note that these results are obtained without using any knowledge of the exact solution: the diffusion process 'finds' the ground state by itself.

Next we analyse the helium atom using the diffusion Monte Carlo method. This is less successful. The reason is that writing the time-evolution operator as a product of a kinetic and potential energy evolution operator

$$\mathrm{e}^{-\Delta\tau(K+V-E_\mathrm{T})} = \mathrm{e}^{-\Delta\tau K}\mathrm{e}^{-\Delta\tau(V-E_\mathrm{T})} + \mathcal{O}(\Delta\tau^2) \tag{12.57}$$

is not justified when the potential diverges, as is the case with the Coulomb potential at $r = 0$. Formally, this equation is still true, but the prefactor of the $\mathcal{O}(\Delta\tau^2)$ term diverges. However, even if the potential does not diverge but varies strongly, the statistical efficiency of the simulation is low. This is due to the fact that if a walker moves to a very favourable region, it will branch into many copies. But these are all the same, and together they form a rather biased sample of the distribution in that region. It requires some time before they have diffused and branched in order to form a representative ensemble. Frequent occurrence of such strong branching events will degrade the efficiency considerably. Quite generally one can say that the efficiency increases with the flatness of the potential.

There exist, in principle, two ways to solve the divergent potential problem. The first one consists of finding a better alternative to the simple approximation to the time-evolution operator than in (12.57). Such approximations have been devised and we shall consider these in the context of path-integral Monte Carlo (see Section 12.4). The common procedure, however, is to use a *guide function*, which transforms the original Schrödinger equation into a new one with a flatter potential, just as in the case of the Fokker–Planck variational Monte Carlo method. This method will be described in the next section.

12.3.3 Guide function for diffusion Monte Carlo

As we have just seen, the diffusion Monte Carlo method causes problems if the potential is unbounded, and this is the case in almost every many-particle system. Sampling some other function instead of the ground state wave function ψ might cure this problem.

A suitable function is $\rho(R,\tau) = \psi(R,\tau)\Psi_\mathrm{T}(R)$ where $\Psi_\mathrm{T}(R)$ is some trial function which models the exact wave function in a reasonable way. It turns out that ρ satisfies a Fokker–Planck type of equation:

$$\frac{\partial\rho(R,\tau)}{\partial\tau} = \frac{1}{2}\nabla_R[\nabla_R - \mathbf{F}(R)]\rho(R,\tau) - [E_\mathrm{L}(R) - E_\mathrm{T}]\rho(R,\tau). \tag{12.58}$$

Here, the 'force' $\mathbf{F}(R)$ is again given as $2\nabla_R\Psi_\mathrm{T}(R)/\Psi_\mathrm{T}(R)$. This form differs from (12.49) because (12.58) is not a 'pure' Fokker–Planck equation: it contains a 'potential term' $E_\mathrm{L}(R) - E_\mathrm{T}$. The 'local energy' $E_\mathrm{L}(R)$ is given as usual by

$$E_\mathrm{L}(R) = \frac{H\Psi_\mathrm{T}(R)}{\Psi_\mathrm{T}(R)} = \frac{-\nabla^2\Psi_\mathrm{T}(R)/2 + V(R)\Psi_\mathrm{T}}{\Psi_\mathrm{T}(R)}. \tag{12.59}$$

The FP-diffusion term will be used to diffuse the walkers, whereas the 'potential' $E_L(R) - E_T$ is used in a branching process. By writing out all the terms on the left and right hand sides of Eq. (12.58), it can be checked that this equation reduces to the imaginary time-dependent Schrödinger equation (12.33).

The procedure is now a combination of the Fokker–Planck VMC and of the DMC method without guide function: we let the walkers diffuse just as in the Fokker–Planck VMC method, with a transition probability

$$T_{\Delta\tau}(R_n \to R_{n+1}) = \frac{1}{\sqrt{2\pi\,\Delta\tau}} \exp\{-[R_{n+1} - R_n - \mathbf{F}(R_n)\Delta\tau/2]^2/(2\Delta\tau)\}. \tag{12.60}$$

Then branching is performed, according to the value $q = \exp\{-\Delta\tau[E_L(R) - E_T]\}$. What do we gain by this method? We avoid problems of the kind encountered above with strongly varying potentials. The role of V in standard DMC is now taken over by $E_L(R)$, which is (we hope) rather flat. If $\Psi_T(R)$ were an *exact* eigenstate, then E_L would be independent of R. If Ψ_T is a reasonable approximation to the ground state, then $E_L(R)$ is reasonably flat, and the method will be reliable. It is clear now why the cusp conditions are so important: they guarantee that the trial function converges to the exact solution in those regions where the potential diverges strongly. These are the points that cause problems. The method using trial – or guide – functions was introduced by Kalos [8] and is commonly called *importance sampling Monte Carlo*.

We can again correct for the time step error using a Metropolis procedure, just as we did for VMC in Section 12.2.5. Note that G is not symmetric, so we must use the generalised Metropolis method in order to guarantee detailed balance (see also the variational Fokker–Planck simulation). A trial displacement is accepted with probability

$$\min\left(1, \frac{T_{\Delta\tau}(R' \to R)\rho(R')}{T_{\Delta\tau}(R \to R')\rho(R)}\right) \tag{12.61}$$

and rejected otherwise.

With importance sampling, the algorithm reads:

```
Put the walkers at random positions in configurational space;
REPEAT
    FOR all walkers DO
        Shift walker from its position R to a new position R'
            by first moving it over a distance F Δτ/2 and then
            adding a random displacement according to the
            transition probability (12.24);
        Accept the move with a probability given by (12.61);
        IF Accepted THEN
```



```
            Evaluate q = exp{−Δτ[E_Local(R') + E_Local(R)]/2 − E_T};
            Eliminate the walker or create new ones at R',
                depending on s = q + r, where r is random,
                uniform between 0 and 1;
        END IF;
    END FOR
    Update E_T using (12.55);
UNTIL finished.
```

Let us first apply the importance sampling method to the one-dimensional harmonic oscillator. We use the same trial (or guide) function $\Psi_T(x) = e^{-\alpha x^2}$ as in the VMC simulation. In that case the quantum force is given by

$$F(x) = -4\alpha x, \tag{12.62}$$

and the local energy by Eq. (12.6). Indeed, the local energy is a constant if $\alpha = 1/2$ and it will be slowly varying if α is close to $1/2$. For $\alpha = 0.4$, a target number of 6000 walkers and 4000 steps, we find for the ground state energy $E = 0.5002 \pm 0.0003$ and with $\alpha = 0.6$, $E = 0.4998 \pm 0.0003$.

We can now do the hydrogen and the helium atom problems. For hydrogen we use a guide function $\exp(-\alpha r)$ and a target number of 2000 walkers performing 4000 steps. The local energy is given by (12.10). Obviously, for $\alpha = 1$ we find the exact ground state energy of -0.5 Hartree as the local energy is constant and equal to this value. For $\alpha = 0.9$, we find a ground state energy of $-0.4967(5)$ and for $\alpha = 1.1$ we find $E_G = 0.5035(5)$. Neither of these values agrees with the exact value. The reason is that the guide function should solve the divergence problem at $r = 0$, but it can do this only if the cusp conditions are satisfied. For $\alpha \neq 1$ this is not the case. This shows the importance of the cusp conditions being satisfied for the trial function.

Finally we present results for the helium atom. We use the Padé–Jastrow wave function (12.11). Varying the parameter α gives values above and below the exact energy. If we monitor the variance of the energy, we find a minimum at $\alpha \approx 0.15$ and an energy $E_G = -2.9029(2)$ for 1000 walkers performing 4000 steps. Remember the exact energy is -2.903 and the variational energy for the uncorrelated wave function (the Hartree–Fock energy) is -2.8617 atomic units.

PROGRAMMING EXERCISE

Modify the DMC programs of the previous section to include a guide function and compare the results with those given in this section.

12.3.4 Problems with fermion calculations

We have described how the simulation of a diffusion process can generate an average distribution of random walkers which is proportional to the ground state wave function or (in the case of guide function DMC) to the product of this function and a trial function. But a distribution of walkers can only represent wave functions which are positive everywhere. For bosons, this property is satisfied by the ground state, but the same does not hold in the case of fermions. The difficulties associated with treating fermions in quantum Monte Carlo are generally denoted as 'the fermion problem'. It should be noted that there is no fermion problem in VMC.

The fixed-node method

There are several approaches to the fermion problem. The simplest approximation is the *fixed-node* method, in which the diffusion process is simulated as before, except for steps crossing a node of the trial function being forbidden. The nodes of the trial function divide the configuration space up into simply connected volumes in which the trial wave function has a unique sign. These volumes are separated from each other by nodal surfaces: hypersurfaces on which the wave function vanishes. To understand why the fixed-node method is useful, suppose that we know the nodes of the exact ground state wave function. If we could solve the ground state of the Schrödinger equation in each simply connected region bounded by the nodal surfaces of the ground state wave function with vanishing boundary conditions on these surfaces, this solution would be proportional to the exact ground state of the full Hamiltonian in each region. In the fixed-node solution, we solve the Schrödinger equation in connected regions bounded by the nodal surfaces of the trial function instead of the exact function, and therefore the quality of the solution depends on how close these surfaces are to those of the exact ground state. It can be shown that the resulting energy is a variational upper bound to the exact ground state energy [2]. It should be noted that the fixed-node method often gives a substantial improvement over the variational Monte Carlo method (which does not suffer from the fermion problem).

An additional problem with the fixed-node method is the fact that moves in which two (or any even number of) nodal surfaces are crossed are accepted. This introduces an error as the number of walkers in two regions separated by an even number of node crossings does not necessarily represent the norm of the wave functions on those regions. The degree to which we suffer from this increases with the time step, as a larger time step will result in larger steps to be taken. It introduces an extra time-step bias error which goes by the name *cross–recross error*.

Let us study the nodes more carefully. The requirement that $\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) = 0$ ($\mathbf{x}_i$ denotes the spin-orbit coordinate of electron i) defines the nodal surfaces. If we assume the spins of the N fermions to be given, then the nodes form $(3N - 1)$-dimensional hypersurfaces in the $3N$-dimensional configurational space. The obvious zeroes of ψ whenever $\mathbf{x}_i = \mathbf{x}_j$ for any pair $i \neq j$ define a $(3N - 3)$-dimensional scaffolding for the nodal surface structure. This scaffolding does not depend on the particular form of the trial function. A node of a one-electron orbital in the Slater determinant occurring in the wave function should not be confused with a 'fermionic zero', as such an orbital node does not force the many-electron wave function to vanish: one of the electrons, say i, might be at a zero of some orbital, but the wave function also contains contributions with the coordinates of the electrons permuted, and in general the coordinates of the other electrons are different from those of electron i.

Changing the diffusion Monte Carlo method to a fixed-node simulation is easy. Simply add the following step just after having generated a new trial position of a particle, say i. Check whether the trial wave function changes sign for this displacement. If this is the case, the move is not accepted, otherwise proceed as in the boson case. The interested reader can implement the fixed-node extension and test it, for example, for the lithium atom, taking an appropriate Slater determinant for the guide function. More details can be found in Ref. [9].

**The transient estimator method*

In view of the variational error present in the fixed-node method it is worthwhile to devise other methods. A method which does not depend on fixed nodal surfaces is the *transient estimator* method. To understand how and why this method works, it is important to realise that the Hamiltonian and hence the time-evolution operator are the same for fermions and for bosons. However, because the time-evolution operator is symmetric with respect to particle permutations, an antisymmetric (fermionic) initial state will remain antisymmetric and a symmetric (bosonic) state remains symmetric.

Let us split an arbitrary fermion wave function ϕ into two parts, ϕ_- and ϕ_+, which contain the negative and positive parts of ϕ respectively (all wave functions depend on all the spin-orbit coordinates $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, and on imaginary time τ):

$$\phi_+ = \tfrac{1}{2}(|\phi| + \phi) \tag{12.63a}$$

$$\phi_- = \tfrac{1}{2}(|\phi| - \phi), \tag{12,63b}$$

so that

$$\phi = \phi_+ - \phi_-. \tag{12.64}$$

Now perform two independent DMC calculations, one with ϕ_- and the other with ϕ_+ as a starting distribution, where ϕ is a trial fermion wave function. What will happen? Applying the (exact) imaginary-time evolution operator $T(X \to Y;\tau)$ to ϕ we obtain

$$\begin{aligned}\phi(Y;\tau) &= \int \mathrm{d}X\, T(X \to Y;\tau)\phi(X;0) \\ &= \int \mathrm{d}X\, T(X \to Y;\tau)\phi_+(X,0) - \int \mathrm{d}X\, T(X \to Y;\tau)\phi_-(X,0) \\ &= \phi_+(Y,t) - \phi_-(Y,t). \end{aligned} \tag{12.65}$$

This suggests that we can follow the time evolution of ϕ by subtracting $\phi_+(t)$ and $\phi_-(t)$ as produced in the two simulations. As $\phi_-(0)$ and $\phi_+(0)$ are both positive, and as the imaginary time-evolution operator is always positive, the application of the DMC approach causes no problems. In fact, one could also say that if the initial wave function is positive everywhere, it contains no fermion character and hence we have an unambiguous bosonic time evolution for such an initial state. A guide function approach can be used in the two boson simulations.

As the time-evolution operator contains no fermion-like features (see above), both simulations will tend to the bosonic ground state solution for long times. The fermion ground state wave function is an excited state solution of the many-particle Hamiltonian, so the boson ground state contribution to the solution at imaginary time τ will dominate the fermion contribution by a factor $\exp[\tau(E_F - E_B)]$, where E_B and E_F are the fermion and boson ground state energies respectively. Note that this factor grows exponentially with time. The fermion ground state wave function is the *difference* between the two distributions resulting from ϕ_- and ϕ_+, which because of the foregoing analysis are both essentially boson-like. If we are to find a fermion wave function as a small difference of two large, essentially boson wave function distributions we must be prepared for large statistical errors. The analysis given here is represented pictorially in Figure 12.2.

The analysis so far leads to the conclusion that, at the beginning, the difference between the distributions is equal to the trial function ϕ, and for large times it converges to the exact fermion wave function, but it will be buried in the noise of the boson solutions forming the bulk of the two distributions. We might be lucky: if the trial function relaxes to the exact Fermi wave function quickly enough, before the latter is buried in the 'boson noise', then we have an intermediate ('transient') regime in imaginary time during which we might extract useful data from the simulation. The trial energy which is adjusted to keep the respective population sizes stable is no longer a suitable energy estimator as this will converge to the

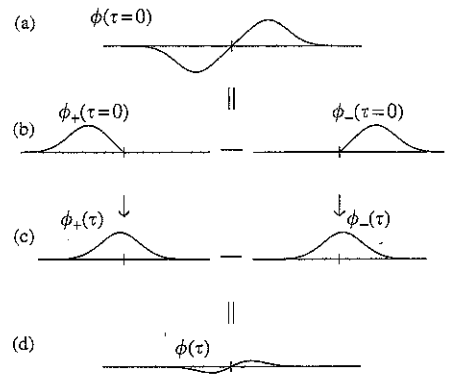

Figure 12.2. Evolution of the distributions in the transient energy estimator method. The wave function $\phi(\tau = 0)$ is shown in (a); it can be written as the difference of the ϕ_+ and ϕ_-. These two functions evolve separately and tend therefore to the same boson ground state solution, as shown in (c). Subtracting the two wave functions in (c) gives the small difference in (d), and this will be soon buried in the noise in the solutions in (c).

boson energy. Therefore we use the 'transient estimator':

$$\begin{aligned} E_{\mathrm{TE}}(\tau) &= \frac{\int \mathrm{d}X\, \phi(\tau) H\phi(\tau=0)}{\int \mathrm{d}X\, \phi(\tau)\phi(\tau=0)} \\ &= \frac{\int \mathrm{d}X\, \phi_-(\tau) H\phi(\tau=0)}{\int \mathrm{d}X\, [\phi_+(\tau) - \phi_-(\tau)]\phi(\tau=0)} - \frac{\int \mathrm{d}X\, \phi_+(\tau) H\phi(\tau=0)}{\int \mathrm{d}X\, [\phi_+(\tau) - \phi_-(\tau)]\phi(\tau=0)}. \end{aligned} \tag{12.66}$$

As the wave function $\phi(\tau)$ converges to the exact fermion ground state, this estimator will indeed relax to the exact fermion energy. As mentioned already, the problem resides in $\phi(\tau)$ to be extracted as the small difference between two large distributions.

The estimator (12.66) is evaluated as follows. At time τ, the walkers occupy points in configuration space which are distributed according to $\phi_\pm(\tau)$. For a walker at the point X in the ϕ_+-simulation we evaluate $H\phi(X, \tau = 0)$ (for the numerator) and $\phi(X, \tau = 0)$ (for the denominator), and sum over walkers. We do the same with the ϕ_- simulation, but now give the contributions a minus sign. The quantity $H\phi(X, \tau = 0)$ can be evaluated because $\phi(X, \tau = 0)$ is a trial function, given in analytic form. The sum is divided by the sum of $\phi(X, \tau = 0)$ over all the walkers.

There exist several extensions to and refinements of the transient estimator method, which are beyond the scope of this book. A common characteristic of these methods is that they are subject to instability in the errors for large τ.

12.4 Path-integral Monte Carlo

In Chapter 11 we saw that the partition function of a classical lattice spin system on a strip can be evaluated by diagonalising the transfer matrix. The transfer matrix can be considered as a kind of 'time-evolution operator', which projects out the eigenvector belonging to the largest eigenvalue (in absolute value). The relation with the time-evolution process described in the previous section is evident. The transfer matrix effectively reduces the dimension of the classical system by one, but the price we pay for this reduction is that the diagonalisation of the transfer matrix is an expensive operation. In this section we consider the reverse transformation: we shall transform a quantum mechanical system in d dimensions, which can be solved by diagonalising the Hamiltonian matrix, to a classical system in $d+1$ dimensions. This system can then be simulated with the Monte Carlo procedures described in Chapter 10. The new formulation enables us to obtain time-dependent properties, or physical quantities of the system at finite temperature. For a very clear discussion of the path-integral concept, see the book by Feynman and Hibbs [10].

12.4.1 Path-integral fundamentals

The path-integral method provides a way to calculate matrix elements and traces of the time-evolution operator of a quantum system in imaginary time:

$$\mathcal{T}(\tau) = \mathrm{e}^{-\tau H} \tag{12.67}$$

which we have encountered in the previous section. If we interpret the imaginary time as an inverse temperature $\tau \leftrightarrow \beta$ and take the trace of the time-evolution operator, we obtain the partition function Z of the quantum system at a finite temperature T:

$$Z(\beta) = \mathrm{Tr}(\mathrm{e}^{-\beta H}) = \int \mathrm{d}R \, \langle R|\mathrm{e}^{-\beta H}|R\rangle. \tag{12.68}$$

R denotes the coordinates of N particles. The path-integral method enables us to sample system configurations with the appropriate Boltzmann factor, so that expectation values for a quantum system at a finite temperature can be evaluated.

The problem with expression (12.68) is that it contains the exponential of the Hamiltonian, which, as mentioned in Section 12.2.4, makes the trace of the time-evolution operator difficult to evaluate. For short times τ (or β), this is not a problem as we can write the Hamiltonian as a sum of several terms (e.g. kinetic and potential energy) which themselves are easily tractable in an exponential – the neglected CBH commutators yield systematic errors of order τ^2. What can we do if τ is not small? In that case, we divide the time τ up into many (say M) small segments $\Delta\tau = \tau/M$ which can be treated in the short-time approximation. For a system consisting of

N spinless particles with coordinates R_i, the partition function can be written as

$$\int \mathrm{d}R_0 \langle R_0|\mathrm{e}^{-\tau H}|R_0\rangle = \int \mathrm{d}R_0 \, \mathrm{d}R_1 \ldots \mathrm{d}R_{M-1}$$
$$\langle R_0|\mathrm{e}^{-\Delta\tau H}|R_1\rangle\langle R_1|\mathrm{e}^{-\Delta\tau H}|R_2\rangle \cdots \langle R_{M-1}|\mathrm{e}^{-\Delta\tau H}|R_0\rangle. \tag{12.69}$$

We have inserted $M-1$ unit-operators $\int \mathrm{d}R_i \, |R_i\rangle\langle R_i|$ between the short-time evolution operators. The procedure in which time is divided up into many short segments is called *time-slicing*. The fact that the first and the last state in the product of matrix elements are identical ($|R_0\rangle$) implies that we have periodic boundary conditions in the τ-direction.

We know the matrix elements of the short-time evolution operator: it has been derived in Section 12.2.4:

$$T(R, R'; \Delta\tau) = \langle R|\mathrm{e}^{-\Delta\tau H}|R'\rangle = \frac{1}{(2\pi\,\Delta\tau)^{3N/2}} \mathrm{e}^{-\Delta\tau V(R)} \mathrm{e}^{-(R-R')^2/(2\Delta\tau)}. \tag{12.70}$$

The potential could have been distributed symmetrically over R and R', but we shall see that the final result does not depend on this distribution. The first order CBH commutator can be shown to vanish in this case, so that this short-time approximation is accurate to order $\Delta\tau^2$. Substituting this result into (12.69), we obtain

$$\int \mathrm{d}R_0 \langle R_0|\mathrm{e}^{-\tau H}|R_0\rangle \approx \frac{1}{(2\pi\,\Delta\tau)^{3NM/2}} \int \mathrm{d}R_0 \, \mathrm{d}R_1 \, \mathrm{d}R_2 \ldots \mathrm{d}R_{M-1}$$
$$\exp\left\{-\Delta\tau \sum_{m=0}^{M-1} \left[\frac{1}{2}\left(\frac{R_{m+1}-R_m}{\Delta\tau}\right)^2 + V(R_m)\right]\right\}. \tag{12.71}$$

In this expression, $R_M = R_0$. The prefactor before the integral seems dangerous in the sense that it explodes when we take the limit $\Delta\tau \to 0$. However, this is balanced by the fact that, of the huge integration volume, only a tiny part gives significant contributions to the integrand – in fact, the smaller we take $\Delta\tau$, the narrower the Gaussian kinetic energy integrands will be and the limit for large M therefore still exists.

You might recognise the summand in the exponent as the Lagrangian (in discrete imaginary time) of the classical many-particle system with coordinates R_i if we take $\Delta\tau \to 0$. The sum is then the *action*, which assumes its minimum for the classical trajectory. The integral is a sum over *all* possible sets of coordinates $R_0, \ldots, R_M$. Such a set denotes a *path* in configuration space. We see that the trace of the time-evolution operator is written as a sum, or rather an integral, over all possible paths. It is important to realise what the classical system represents. The quantum many-particle system we are describing contains N particles, interacting with each other and with an external potential through the potential $V(R)$. We have M copies

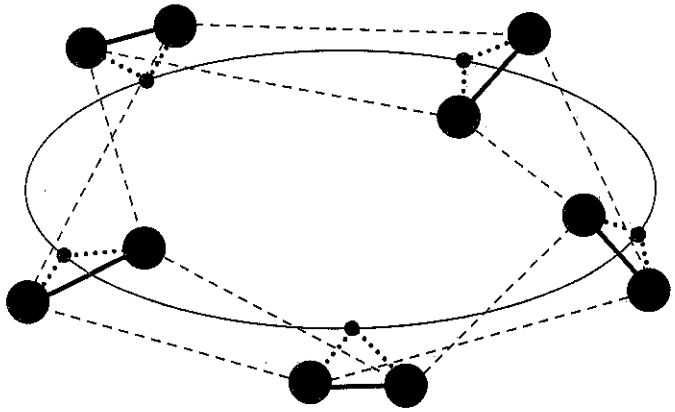

Figure 12.3. Classical system described by the path integral of the two electrons in the helium atom. Periodic boundary conditions are imposed along the quantum imaginary time (the circle). The small full circles denote the helium nuclei, the heavy ones the electrons. The circle is the time axis with periodic boundary conditions. The dashed lines represent harmonic couplings between the electrons of adjacent copies (along the time axis). The heavy solid lines denote the electron–electron interaction, and the heavy dotted lines the electron–nucleus interactions.

of this many-particle system along the quantum imaginary-time direction, so that the classical system consists of NM particles. The first term in the sum in (12.71) derives from the kinetic part of the quantum Hamiltonian, but in the classical system it denotes a harmonic coupling between corresponding particles in adjacent copies: they are connected by springs. Figure 12.3 shows the classical particle system and couplings for the two electrons in helium with $M = 5$.

The quantum partition function for a system of N three-dimensional particles is given as $\mathrm{Tr}\exp(-\beta H)$. The right hand side of Eq. (12.71) can be interpreted as the *classical* partition function of NM particles in three dimensions (without momentum degrees of freedom – these can be thought of as being integrated over), because it is an integral over all the configurations of the coordinates R_i with an appropriate Boltzmann factor. The energy $\mathcal{H}$ of the classical system is identified with the Lagrangian associated with the quantum Hamiltonian H. An unusual feature is the inverse temperature occurring in the denominator of the harmonic interactions of the classical Hamiltonian $\mathcal{H}$ (remember $\Delta\tau = \beta/M$). We see that the path integral maps the partition function of a $3N$-dimensional system onto a $(3N+1)$-dimensional system where the extra dimension can be interpreted either as an imaginary-time or as an inverse-temperature axis – it corresponds to the sub-index i of the R_i.

The path integral provides a very clear insight into the nature of quantum mechanics. Up to now, we have put $\hbar \equiv 1$. Had we kept $\hbar$ in the problem, we would have

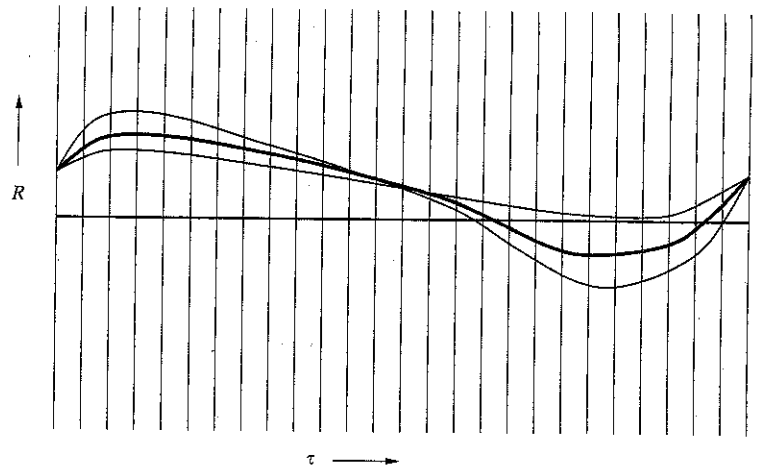

Figure 12.4. The path integral for a one-dimensional system. The vertical axes are R-axes at different times. A path is a set of points given on these axes. The heavy drawn path is the stationary path of the action, which is the solution to the classical equations of motion. The thin lines represent neighbouring paths. For these paths, the action is not stationary, but they are taken into account in the quantum mechanical path integral.

seen that the prefactor in the exponent occurring before the sum was $\Delta\tau/\hbar$ instead of $\Delta\tau$. The classical limit corresponds to $\hbar = 0$, which implies that the path with minimal action dominates all the other paths. This is in fact Hamilton's principle: the classical path corresponds to the minimal action. If we 'switch Planck's constant on', we see a contribution from the nonminimal paths emerging. If we had not identified R_0 with R_M and if we had not integrated over this coordinate, we would have a system with fixed end points, which brings the analogy with classical mechanics even closer. Figure 12.4 gives a pictorial representation of the idea of the path integral.

In this section and in the previous one, we have assumed that the errors in the individual short-time approximations do not add up to significant errors for large times. The justification of this assumption is a theorem, which is usually denoted as the Lie–Trotter–Suzuki formula, which says that for a Hamiltonian H which can be written as the sum of K operators:

$$H = \sum_{k=1}^{K} H_k \tag{12.72}$$

it holds that

$$\mathrm{e}^{-\alpha H} \to (\mathrm{e}^{-\alpha H_1/M}\mathrm{e}^{-\alpha H_2/M}\ldots\mathrm{e}^{-\alpha H_K/M})^M \tag{12.73}$$

for large M. The error is then given by [11, 12]

$$\frac{\alpha^2}{M} \sum_{m>m'} |[H_m, H_{m'}]| \mathrm{e}^{-\alpha \sum_m |H_m|}, \tag{12.74}$$

where $|\ldots|$ denotes the norm of an operator.

It is very easy to get confused with many physical quantities having different meaning according to whether we address the time-evolution operator, the quantum partition function, or the classical partition function. Therefore we summarise the different interpretations in Table 12.2. The classical time in the last row of Table 12.2 is the time that elapses in the classical system and is analogous to the time in a Monte Carlo simulation. This quantity has no counterpart in quantum mechanics or in the statistical partition function.

The quantum partition function is now simulated simply by performing a standard Monte Carlo simulation on the classical system. The PIMC algorithm is

```
Put the NM particles at random positions;
REPEAT
   FOR m = 1 TO M DO
      Select a time slice m̃ at random;
      Select one of the N particles at time slice m̃ at random;
      Generate a random displacement of that particle;
      Calculate r = exp[−Δτ(H_new − H_old)];
      Accept the displacement with probability min(1, r);
   END FOR;
UNTIL finished.
```

In this algorithm we have used $\mathcal{H}$ to denote the Hamiltonian of the classical system, which is equal to the Lagrangian occurring in the exponent of the path integral – see Eq. (12.71).

Let us compare the path integral method with the diffusion Monte Carlo approach. In the latter we start with a given distribution and let time elapse. At the end of the simulation the distribution of walkers reflects the wave function at imaginary time τ. Information about the history is lost: physical time increases with simulation time. The longer our simulation runs, the more strongly will the distribution be projected onto the ground state. In the path integral method, we change the positions of the particles along the imaginary-time (inverse-temperature) axis. Letting the simulation run for a longer time does not project the system more strongly onto the ground state – the extent to which the ground state dominates in the distribution is determined by the temperature $\beta = M\Delta\tau$, i.e. for fixed $\Delta\tau$, it is determined by the length of the chain. The PIMC method is not necessarily carried out in imaginary

Table 12.2. *Meaning of several physical quantities in different interpretations of the path integral.*

Quantum mechanics	Quantum statistical mechanics	Classical mechanics	Statistical physics
d-dimensional configuration space	d-dimensional configuration space	d-dimensional subspace of configuration space	d-dimensional configuration space
imaginary time τ	inverse temperature $\beta = 1/k_B T$	1-dimensional axis in configuration space	inverse temperature $\beta = 1/k_B T$
time-evolution operator	Boltzmann operator $\mathrm{e}^{-\beta H}$	–	transfer matrix
kinetic energy	kinetic energy	harmonic interparticle potential	inter-row coupling of transfer matrix
Lagrangian	Lagrangian	Lagrangian	Hamiltonian
path integral	quantum partition function of d-dimensional system	–	partition function of $(d+1)$-dimensional system
classical limit	zero temperature	stationary path	zero temperature

time – there exist versions with real time, which are used to study the dynamics of quantum systems [13–15].

The analysis so far is correct for distinguishable particles. In fact, we have simply denoted a coordinate representation state by $|R\rangle$. For indistinguishable bosons, we should read for this state:

$$|R\rangle = \frac{1}{N!}\sum_P |\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\rangle, \tag{12.75}$$

where the sum is over all permutations of the positions. The boson character is noticeable when we impose the periodic boundary conditions along the τ-axis, where we should not merely identify $\mathbf{r}_k$ in the last coordinate $|R_M\rangle$ with the corresponding position in $|R_0\rangle$, but also allow for permutations of the individual particle positions in both coordinates to be connected.

This feature introduces a boson entropy contribution, which is particularly noticeable at low temperatures. To see this, let us consider the particles as diffusing from left (R_0) to right (R_M). On the right hand side we must connect the particles to their counterparts on the left hand sides, taking all permutations into account. If the Boltzmann factor forbids large steps when going from left to right, it is unlikely that we can connect the particles on the right hand side to the permuted leftmost positions without introducing a high energy penalty. This is the case when $\tau = \beta$ is small, or equivalently when the temperature is high. This can be seen by noticing that, keeping $\Delta\tau = \beta/N$ fixed, a decrease β must be accompanied by a decrease in the number of segments N. Fewer segments mean less opportunity for the path to wander away from its initial position. On the other hand, we might keep the number of segments constant, but decrease $\Delta\tau$. As the spring constants are inversely proportional to $\Delta\tau$ (see Eq. (12.71)), they do not allow, in that case, for large differences in position on adjacent time slices; hence permutations are quite unlikely. When the temperature is high ($\tau = \beta$ small), large diffusion steps are allowed and there is a lot of entropy to be gained from connecting the particles to their starting positions in a permuted fashion. This entropy effect is responsible for the superfluid transition in ^{4}He [14–16]. Path-integral methods also exist for fermion systems. A review can be found in Ref. [19].

What type of information can we obtain from the path integral? First of all, we can calculate ground state properties by taking β very large (temperature very small). The system will then be in its quantum ground state. The particles will be distributed according to the quantum ground state wave function. This can be seen by considering the expectation value for particle 0 to be at position R_0. This is given by

$$P(R_0) = \frac{1}{Z}\int \mathrm{d}R_1 \mathrm{d}R_2 \ldots \mathrm{d}R_{M-1} \langle R_0|\mathrm{e}^{-\Delta\tau H}|R_1\rangle\langle R_1|\mathrm{e}^{-\Delta\tau H}|R_2\rangle \ldots \langle R_{M-1}|\mathrm{e}^{-\Delta\tau H}|R_0\rangle. \tag{12.76}$$

Note that the numerator differs from the path integral (which occurs in the denominator) in the absence of the integration over R_0. Removing all the unit operators we obtain

$$P(R_0) = \frac{\langle R_0|\mathrm{e}^{-\tau H}|R_0\rangle}{\int \mathrm{d}R_0 \langle R_0|\mathrm{e}^{-\tau H}|R_0\rangle}. \tag{12.77}$$

Large τ is equivalent to low temperature. But if τ is large indeed, then the operator $\exp(-\tau H)$ projects out the ground state ϕ_G:

$$\mathrm{e}^{-\tau H} \approx |\phi_G\rangle \mathrm{e}^{-\tau E_G} \langle \phi_G|, \quad \text{large} \tau. \tag{12.78}$$

Therefore we have

$$P(R_0) = \frac{1}{Z}\mathrm{e}^{-\tau E_G}|\langle \phi_G|R_0\rangle|^2, \text{large} \tau. \tag{12.79}$$

Because of the periodic boundary conditions in the τ direction we obtain the same result for each time slice m. To reduce statistical errors, the ground state can be therefore obtained from the *average* distribution over the time slices via a histogram method.

The expectation value of a physical quantity A for a quantum system at a finite temperature is found as

$$\langle A\rangle_\beta = \frac{\mathrm{Tr}(A\mathrm{e}^{-\beta H})}{\mathrm{Tr}\,\mathrm{e}^{-\beta H}}. \tag{12.80}$$

The denominator is the partition function Z. We can use this function to determine the expectation value of the energy

$$\langle E\rangle_\beta = \frac{\mathrm{Tr}(H\mathrm{e}^{-\beta H})}{Z} = -\frac{\partial}{\partial\beta}\ln Z(\beta). \tag{12.81}$$

If we apply this to the path-integral form of Z, we obtain for the energy per particle (in one dimension):

$$\langle \frac{E}{N}\rangle_\beta = \frac{M}{2\beta} - \frac{1}{N}(\langle K\rangle - \langle V\rangle). \tag{12.82}$$

with

$$K = \sum_{m=0}^{M-1} \frac{(R_m - R_{m+1})^2}{2\beta^2} \tag{12.83}$$

and V is the potential energy (see also Problem 12.1). The first term in (12.82) derives from the prefactor $1/\sqrt{2\pi\Delta\beta}$ of the kinetic Green's function. The angular brackets in the second and third term denote expectation values evaluated in the classical statistical many-particle system. It turns out that this expression for the energy is subject to large statistical errors in a Monte Carlo simulation. The reason

is that $1/\beta$ and $\langle K\rangle/(NM)$ are both large, but their difference is small. Herman *et al.* [20] have proposed a different estimator for the energy, given by

$$\left\langle \frac{E}{N} \right\rangle_\beta = \left\langle \frac{1}{M} \sum_{m=0}^{M-1} \left[V(R_m) + \frac{1}{2} R_m \cdot \nabla_{R_m} V(R_m) \right] \right\rangle. \tag{12.84}$$

This is called the *virial energy estimator*, and it will be considered in Problem 12.6.

The virial estimator is not always superior to the direct expression, as was observed by Singer and Smith for Lennard–Jones systems [21]; this is presumably due to the steepness of the Lennard–Jones potential causing large fluctuations in the virial.

12.4.2 Applications

We check the PIMC method for the harmonic oscillator in one dimension. We have only one particle per time slice. The particles all move in a 'background potential', which is the harmonic oscillator potential, and particles in neighbouring slices are coupled by the kinetic, harmonic coupling. The partition function reads

$$Z = \int \mathrm{d}x_0 \ldots \mathrm{d}x_{M-1} \exp\left\{ -\frac{\beta}{M} \sum_{m=0}^{M-1} \left[\frac{(x_m - x_{m+1})^2}{2\Delta\beta^2} + \frac{1}{2} x_m^2 \right] \right\}. \tag{12.85}$$

We have used $\beta = 10$ and $M = 100$. Thirty thousand MCS were performed, of which the first two thousand were deleted to reach equilibrium. The maximum displacement was tuned to yield an acceptance rate of about 0.5. The spacing between the energy levels of the harmonic oscillator is 1; therefore $\beta = 10$ corresponds to large temperature. We find for the energy $E = 0.51 \pm 0.02$, in agreement with the exact ground state energy of 1/2. The ground state amplitude can also be determined, and it is found to match the exact form $|\psi(x)|^2 = \mathrm{e}^{-x^2}$ very well.

The next application is the hydrogen atom. This turns out to be less successful, just as in the case of the diffusion MC method. The reason is again that writing the time-evolution operator as the product of the exponentials of the kinetic and potential energies is not justified when the electron approaches the nucleus, as the Coulomb potential diverges there – CBH commutators therefore diverge too. The use of guide functions is not possible in PIMC, so we have to think of something else. The solution lies in the fact that the *exact* time-evolution operator over a time slice Δt does not diverge at $r = 0$; we suffer from divergences because we have used the so-called *primitive approximation*

$$T(\mathbf{r} \to \mathbf{r}'; \Delta\tau) = \frac{1}{(2\pi\,\Delta\tau)^{3/2}} \exp[-(\mathbf{r} - \mathbf{r}')/(2\Delta\tau)] \exp\{-\Delta\tau[V(\mathbf{r}) + V(\mathbf{r}')]/2\} \tag{12.86}$$

to the time-evolution operator. The effect of averaging over all the continuous paths from $(\mathbf{r}, \tau)$ to $(\mathbf{r}', \tau + \Delta\tau)$, as is to be done when calculating the exact time evolution, is that the divergences at $\mathbf{r}, \mathbf{r}' = 0$ are rounded off. So if we could find a better approximation to this exact time evolution than the primitive one, we would not suffer from the divergences any longer. Several such approximations have been developed [22, 23]. They are based either on exact Coulomb potential solutions (hydrogen atom) or on the cumulant expansion. We consider the latter approximation in some detail in Problems 12.2 and 12.3; here we shall simply quote the result:

$$V_{\text{cumulant}}(\mathbf{r}, \mathbf{r}'; \Delta\tau) = \int_0^{\Delta\tau} \mathrm{d}\tau' \frac{\operatorname{erf}[r(\tau')/\sqrt{2\sigma_{\tau'}}]}{r(\tau')}, \tag{12.87a}$$

where

$$\mathbf{r}(\tau') = \mathbf{r} + \frac{\tau'}{\Delta\tau}(\mathbf{r}' - \mathbf{r}) \quad \text{and} \quad \sigma(\tau') = \frac{(\Delta\tau - \tau')\tau'}{\Delta\tau}. \tag{12.87b}$$

The cumulant approximation for V can be calculated and saved in a tabular form, so that we can read it into an array at the beginning of the program, and then obtain the potential for the values needed from this array by interpolation. In fact, for $\Delta\tau$ fixed, V_{cumulant} depends on the norms of the vectors $\mathbf{r}$ and $\mathbf{r}'$ and on the angle between them. Therefore the table is three-dimensional. We discretise r in, say, 50 steps Δr between 0 and some upper limit $r_{\max}$ (which we take equal to 4), and similarly for r'. For values larger than $r_{\max}$ we simply use the primitive approximation, which is sufficiently accurate in that case. For the angle θ in between $\mathbf{r}$ and $\mathbf{r}'$ we store $\cos\theta$, discretised in 20 steps between -1 and 1 in our table. For actual values r, r' and $u = \cos\theta$ we interpolate linearly from the table – see Problem 12.4. Figure 12.5 shows the cumulant potential $V(r = r', \theta = 0; \Delta\tau = 0.2)$, together with the Coulomb potential; the rounding effect of the cumulant approximation is clear. In a path-integral simulation for the hydrogen atom we find a good ground state distribution, shown in Figure 12.6. For the energy, using the virial estimator with the original Coulomb potential (which is of course not entirely correct), we find $E_{\mathrm{G}} = -0.494 \pm 0.014$, using $\Delta\tau = 0.2$, 100 time slices and 60 000 MC steps per particle, of which the first 20 000 were removed for equilibration.

Applying the method to helium is done in the same way. Using 150 000 steps with a chain length of 50 and $\tau = 0.2$, the ground state energy is found as 2.93 ± 0.06 atomic units. Comparing the error with the DMC method, the path-integral method does not seem to be very efficient, but this is due to the straightforward implementation. It is possible to improve the PIMC method considerably as will be described in the next section.

The classical example of a system with interesting behaviour at finite temperature is dense helium-4. In this case the electrons are not taken into account as independent

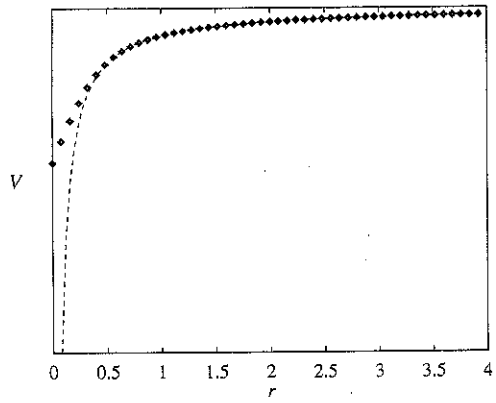

Figure 12.5. The cumulant potential for $\Delta\tau = 0.2$ (diamonds) and the Coulomb potential. It is clearly seen that the cumulant potential is rounded off at $r = 0$.

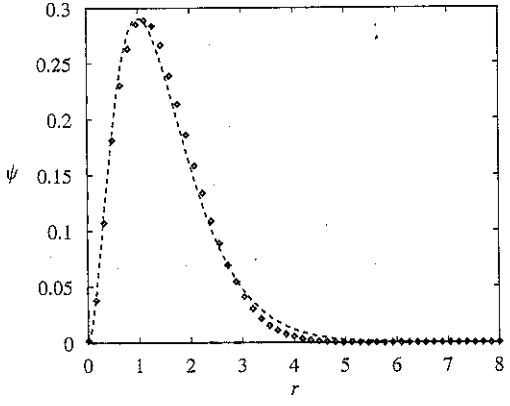

Figure 12.6. PIMC ground state amplitude $|\psi(r)|^2$ (diamonds) and the exact result. Sixty thousand Monte Carlo sweeps with a chain length of 100 and $\tau = 0.2$ were used.

particles; rather, a collection of atoms is considered, interacting through Lennard–Jones potentials. We shall not go into details of implementation and phase diagram, but refer to the work by Ceperley and Pollock [3, 4].

12.4.3 Increasing the efficiency

The local structure of the action enables us to use the heat-bath algorithm instead of the classical sampling rule, in which particles are displaced at random uniformly

within a cube (or a sphere). If we update the coordinate R_m, keeping R_{m-1} and R_{m+1} fixed, then in the heat-bath algorithm, the new value R'_m must be generated with distribution

$$\rho(R'_m) = \exp\left[-\Delta\tau\frac{(R'_m - \overline{R}_m)^2}{2\Delta\tau^2} - \Delta\tau V(R'_m)\right] \tag{12.88}$$

where $\overline{R}_m = (R_{m+1} + R_{m-1})/2$. We may sample the new position directly from this distribution by first generating a new position using a Gaussian random generator with width $1/(2\Delta\tau)$ and centred around $\overline{R}_m$, and then accepting or rejecting the new position with a probability proportional to $\exp[-\Delta\tau V(R'_m)]$. This procedure guarantees 100% acceptance for zero potential. If there are hard-core interactions between the particles, the Gaussian distribution might be replaced by a more complicated form to take this into account [4].

A major drawback of the algorithm presented so far is that only one atom is displaced at a time. To obtain a decent acceptance rate the maximal distance over which the atom can be displaced is restricted by the harmonic interaction between successive 'beads' on the imaginary time-chain to $\sim\sqrt{\Delta\tau}$. The presence of the potential V can force us to decrease this step size even further. It will be clear that our local update algorithm will cause the correlation time to be long, as this time is determined by the long-wavelength modes of the chain. As it is estimated that equilibration of the slowest modes takes roughly $\mathcal{O}(M^2)$ Monte Carlo sweeps (see the next chapter), the relaxation time will scale as M^3 single-update steps. This unfavourable time scaling behaviour is well known in computational field theory, and a large part of the next chapter will be dedicated to methods for enhancing the efficiency of Monte Carlo simulations on lattices. An important example of such methods is *normal mode sampling* in which, instead of single particle moves, one changes the configuration via its Fourier modes [24, 25]. If one changes for example the $k = 0$ mode, all particles are shifted over the same distance. The transition probability is calculated either through the Fourier-transformed kinetic (harmonic interaction) term, followed by an acceptance/rejection based on the change in potential, or by using the Fourier transform of the full action. We shall not treat these methods in detail here; in the next chapter, we shall discuss similar methods for field theory.

A method introduced by Ceperley and Pollock divides the time slices up in a hierarchical fashion and alters the values of groups of points in various stages [3, 4]. At each stage the step can be discontinued or continued according to some acceptance criterion. It turns out [4] that with this method it is possible to reduce the relaxation time from M^3 to $M^{1.4}$. The method seems close in spirit to the multigrid Monte Carlo method of Goodman and Sokal, which we shall describe in the next chapter.

It will be clear that for a full boson simulation, moving particles is not sufficient: we must also include permutation moves, in which we swap two springs between particles at subsequent beads, for example. However, the configurations are usually equilibrated for a particular permutation, and changing this permutation can be so drastic a move that permutations are never accepted. In that case it is possible to combine a permutation with particle displacements which adjust the positions to the new permutation [4].

12.5 Quantum Monte Carlo on a lattice

There are several interesting quantum systems which are or can be formulated on a lattice. First of all, we may consider quantum spin systems as generalisations of the classical spin systems mentioned in Chapter 7. An example is the Heisenberg model, with Hamiltonian

$$H_{\text{Heisenberg}} = -J \sum_{\langle ij \rangle} \mathbf{s}_i \cdot \mathbf{s}_j \tag{12.89}$$

where the sum is over nearest neighbour sites $\langle ij \rangle$ of a lattice (in any dimensions), and the spins satisfy the standard angular momentum commutation relations on the same site ($\hbar \equiv 1$):

$$[s^x, s^y] = \frac{\mathrm{i}s^z}{2}. \tag{12.90}$$

Another example is the second quantised form of the Schrödinger equation. This uses the 'occupation number representation' in which we have creation and annihilation operators for particles in a particular state. If the Schrödinger equation is discretised on a grid, the basis states are identified with grid points, and the creation and annihilation operators create and annihilate particles on these grid points. These operators are called $c_i^\dagger$ and c_i respectively, and they satisfy the commutation relations

$$[c_i, c_j] = [c_i^\dagger, c_j^\dagger] = 0; \quad [c_i, c_j^\dagger] = \delta_{ij}. \tag{12.91}$$

In terms of these operators, the Schrödinger equation for a one-dimensional, noninteracting system reads [26]

$$\sum_i -t(c_i^\dagger c_{i+1} + c_{i+1}^\dagger c_i) + \sum_i V_i n_i \tag{12.92}$$

where n_i is the number operator $c_i^\dagger c_i$, and where appropriate boundary conditions are to be chosen.

A major advantage of this formulation over the original version of the Schrödinger equation is that the boson character is automatically taken into account: there is no need to permute particles in the Monte Carlo algorithm. A disadvantage is that the lattice will introduce discretisation errors.

Finally, this model may be formulated for interacting fermions. A famous model of this type is the so-called *Hubbard model*, which models the electrons which are tightly bound to the atoms in a crystalline material. The Coulomb repulsion is restricted to an on-site effect; electrons on different sites do not feel it. The creation and annihilation operators are now called $c_{i,\sigma}^\dagger, c_{i,\sigma}$, where $\sigma = \pm$ labels the spin. They anticommute, except for $[c_{i,\sigma}^\dagger, c_{j,\sigma'}]_+ = \delta_{ij}\delta_{\sigma\sigma'}$. The standard form of the Hubbard model in one dimension reads

$$H = \sum_{i,\sigma} -t[c_{i,\sigma}^\dagger c_{i+1,\sigma} + c_{i+1,\sigma}^\dagger c_{i,\sigma}] + U \sum_i n_{i,\sigma} n_{i,-\sigma} \tag{12.93}$$

where $n_{i,\sigma}$ is the number operator which counts the particles with spin σ at site i: $n_i = c_{i,\sigma}^\dagger c_{i,\sigma}$. The first term describes hopping from atom to atom, and the second one represents the Coulomb interaction between fermions at the same site.

We shall outline the quantum path-integral Monte Carlo analysis for one-dimensional lattice quantum systems, taking the Heisenberg method as the principal example. Extensions to other systems will be considered only very briefly. For a review, see Ref. [5]; see also Ref. [27].

The quantum Heisenberg model is formulated on a chain consisting of N sites, which we shall number by the index i. We have already discussed this model in Section 11.5. The Hilbert space has basis states $|S\rangle = |s_1, s_2, \ldots, s_N\rangle$, where the s_i assume values ± 1; they are the eigenstates of the z-component of the spin operator. The Heisenberg Hamiltonian can be written as the sum of operators containing interactions between *two neighbouring* sites. Let us call H_i the operator $-J\mathbf{s}_i \cdot \mathbf{s}_{i+1}$, coupling spins at sites i and $i+1$. Suppose we have N sites and that N is even. We now partition the Hamiltonian as follows:

$$\begin{aligned} H = H_{\text{odd}} + H_{\text{even}} &= (H_1 + H_3 + H_5 + \cdots + H_{N-1}) \\ &\quad + (H_2 + H_4 + H_6 + \cdots + H_N). \end{aligned} \tag{12.94}$$

H_i and H_{i+2} commute as the H_i couple only nearest neighbour sites. This makes the two separate Hamiltonians H_{odd} and H_{even} trivial to deal with in the path integral. However, H_{odd} and H_{even} do not commute. It will therefore be necessary to use the short-time approximation.

The time-evolution operator is split up as follows:

$$\mathrm{e}^{-\tau H} \approx \mathrm{e}^{-\Delta\tau H_{\text{odd}}} \mathrm{e}^{-\Delta\tau H_{\text{even}}} \mathrm{e}^{-\Delta\tau H_{\text{odd}}} \mathrm{e}^{-\Delta\tau H_{\text{even}}} \ldots \mathrm{e}^{-\Delta\tau H_{\text{odd}}} \mathrm{e}^{-\Delta\tau H_{\text{even}}} \tag{12.95}$$

with a total number of $2M$ exponents in the right hand side; $\Delta\tau = \tau/M$. In calculating the partition function, we insert a unit operator of the form $\sum_S |S\rangle\langle S|$ between

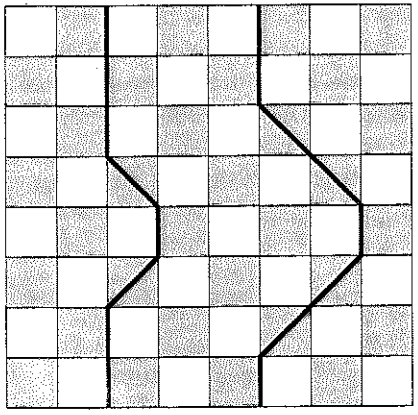

Figure 12.7. The checkerboard decomposition of the space-time lattice. Two world lines are shown.

the exponentials, where $\sum_S$ denotes a sum over all the spins s_i in S:

$$Z = \sum_{S_i,\bar{S}_i} \langle S_0|e^{-\Delta\tau H_{\text{odd}}}|\bar{S}_0\rangle\langle\bar{S}_0|e^{-\Delta\tau H_{\text{even}}}|S_1\rangle\langle S_1|e^{-\Delta\tau H_{\text{odd}}}|\bar{S}_1\rangle$$
$$\times \langle\bar{S}_1|e^{-\Delta\tau H_{\text{even}}}|S_2\rangle \cdots \langle S_{N/2-1}|e^{-\Delta\tau H_{\text{odd}}}|\bar{S}_{N/2-1}\rangle$$
$$\times \langle\bar{S}_{N/2-1}|e^{-\Delta\tau H_{\text{even}}}|S_0\rangle. \tag{12.96}$$

The operators $\exp(\Delta\tau H_{\text{even}})$ and $\exp(\Delta\tau H_{\text{odd}})$ can be expanded as products of terms of the form $\exp(\Delta\tau H_i)$. Each such term couples the spins around a plaquette of the space-time lattice and the resulting picture is that of Figure 12.7, which explains the name 'checkerboard decomposition' for this partitioning of the Hamiltonian. Other decompositions are possible, such as the real-space decomposition [5], but we shall not go into this here.

The simulation of the system seems straightforward: we have a space-time lattice with interactions around the shaded plaquettes in Figure 12.7. At each site of the lattice we have a spin s_{im}, where i denotes the spatial index and m denotes the index along the imaginary-time or inverse-temperature axis. The simulation consists of attempting spin flips, evaluating the Boltzmann weight before and after the change, and then deciding to perform the change or not with a probability determined by the fractions of the Boltzmann weights (before and after). But there is a snake in the grass. The Hamiltonians H_m commute with the total spin operator, $\sum_i s_i^z$; therefore the latter is conserved, i.e.

$$s_{im} + s_{i+1,m} = s_{i,m+1} + s_{i+1,m+1} \tag{12.97}$$

for each plaquette (remember the s_i occurring in this equations are the eigenvalues of the corresponding s_i^z operators). Therefore a single spin flip will never be accepted as it does not respect this requirement. This was already noted in Section 11.5: letting a chain evolve under the Hamiltonian time evolution leaves the system in the 'sector' where it started off. Simple changes in the spin configuration which conserve the total spin from one row to another are spin flips of all the spins at the corners of a nonshaded plaquette.

In the boson and fermion models, where we have particle numbers n_{im} instead of spins, the requirement (12.97) is to be replaced by

$$n_{im} + n_{i+1,m} = n_{i,m+1} + n_{i+1,m+1}. \tag{12.98}$$

In this case the simplest change in the spin configuration consists of an increase (decrease) by one of the numbers at the two left corners of a nonshaded plaquette and a decrease (increase) by one of the numbers at the right hand corners (obviously, the particle numbers must obey $n_{im} \geq 0$ (bosons) or $n_{im} = 0, 1$ (fermions)). Such a step is equivalent to having one particle moving one lattice position to the left (right). The overall particle number along the time direction is conserved in this procedure. The particles can be represented by *world lines*, as depicted in Figure 12.7. The changes presented here preserve particle numbers from row to row, so for a simulation of the full system, one should consider also removals and additions of entire world lines as possible Monte Carlo moves.

Returning to the Heisenberg model, we note that the operator $\exp(-\Delta\tau H_i)$ couples only spins at the bottom of a shaded plaquette to those at the top. This means that we can represent this operator as a 4×4 matrix, where the four possible states $|++\rangle$, $|+-\rangle$, $|-+\rangle$ and $|--\rangle$ label the rows and columns. For the Heisenberg model one finds after some calculation

$$\exp\left[-\Delta\tau\frac{J}{4}\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+1}\right]$$
$$= e^{-\Delta J/4}\begin{pmatrix} e^{\Delta\tau J/2} & 0 & 0 & 0 \\ 0 & \cosh(\Delta\tau J/2) & \sinh(\Delta\tau J/2) & 0 \\ 0 & \sinh(\Delta\tau J/2) & \cosh(\Delta\tau J/2) & 0 \\ 0 & 0 & 0 & e^{\Delta\tau J/2} \end{pmatrix} \tag{12.99}$$

($\boldsymbol{\sigma}$ is the vector of Pauli matrices $(\sigma_x, \sigma_y, \sigma_z)$ – we have $\mathbf{s} = \hbar\boldsymbol{\sigma}/2$; $\hbar \equiv 1$). This matrix can be diagonalised (only a diagonalisation of the inner 2×2 block is necessary) and the model can be solved trivially. Some matrix elements become negative when $J < 0$ (Heisenberg antiferromagnet). This minus-sign problem turns out not to be fundamental, as it can be transformed away by a redefinition of the spins on alternating sites [5, 28].

In the case where, instead of spin-1/2 degrees of freedom, we have (boson) numbers on the sites, the matrix H_1 becomes infinite-dimensional. In that case we must expand $\exp(-\Delta\tau H_l)$ in a Taylor series expansion in $\Delta\tau$. We shall not go into details but refer to the literature [5].

If we have fermions, there is again a minus-sign problem. This turns out to be removable for a one-dimensional chain, but not for two and three dimensions. In these cases one uses fixed-node and transient estimator methods as described above [29].

12.6 The Monte Carlo transfer matrix method

In Chapter 11 we have seen that it is possible to calculate the free energy of a discrete lattice spin model on a strip by solving the largest eigenvalue of the transfer matrix. The size of the transfer matrix increases rapidly with the strip width and the calculation soon becomes unfeasible, in particular for models in which the spins can assume more than two different values. The QMC techniques which have been presented in the previous sections can be used to tackle the problem of finding the largest eigenvalues of the very large matrices arising in such models. Here we discuss such a method. It goes by the name of 'Monte Carlo transfer matrix' (MCTM) method and it was pioneered by Nightingale and Blöte [30].

Let us briefly recall the transfer matrix theory. The elements $T(S', S) = \langle S'|T|S\rangle$ of the transfer matrix T are the Boltzmann weights for adding new spins to a semi-infinite system. For example, the transfer matrix might contain the Boltzmann weights for adding an entire row of spins to a semi-infinite lattice model, or a single spin, in which case we take helical boundary conditions so that the transfer matrix is the same for each spin addition (see Figure 12.8). The free energy is given in terms of the largest eigenvalue λ_0 of the transfer matrix:

$$F = -k_B T \ln(\lambda_0). \tag{12.100}$$

From discussions in Chapter 11 and Section 12.4, it is clear that the transfer matrix of a lattice spin model is the analogue of the time-evolution operator in quantum mechanics.

We now apply a technique analogous to diffusion Monte Carlo to sample the eigenvector corresponding to the largest eigenvalue. In the following we use the terms 'ground state' for this eigenvector, because the transfer matrix can be written in the form $\mathcal{T} = \exp(-\tau H)$, so that the ground state of H gives the largest eigenvalue of the transfer matrix. We write the transfer matrix as a product of a normalised transition probability P and a weight factor D. In Dirac notation:

$$\langle S'|T|S\rangle = D(S')\langle S'|P|S\rangle. \tag{12.101}$$

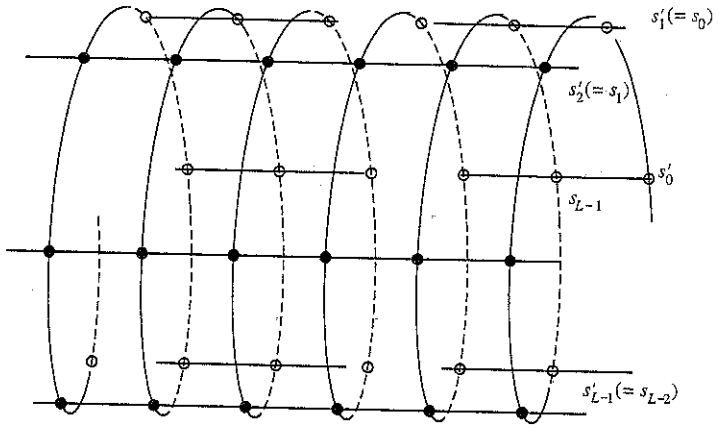

Figure 12.8. Helical boundary conditions for the spin model with nearest neighbour interactions on a strip. A step of the algorithm consists of evolving the 'old' walker S into a new one called S'. This is done by first adding a new 'head' s'_0 of S' according to a probability distribution like (12.104). Then the 'old' components s_{L-2} to s_0 are copied onto s'_{L-1} to s'_1.

The ground state will be represented by a collection of random walkers $\{S_k\}$ which diffuse in configuration space according to the transition probability P. Each diffusion step is followed by a branching step in which the walkers are eliminated or multiplied, i.e. split into a collection of identical walkers, depending on the value of the weight factor $D(S'_k)$.

Let us describe the procedure for a *p-state clock model* with stochastic variables (spins) which assume values

$$\theta = \frac{2\pi n}{p}, n = 0, \ldots, p-1 \tag{12.102}$$

and a nearest neighbour coupling

$$-\frac{\mathcal{H}}{k_B T} = \sum_{\langle ij\rangle} J\cos(\theta_i - \theta_j). \tag{12.103}$$

For $p = 2$ this is equivalent to the Ising model (with zero magnetic field), with J being exactly the same coupling constant as in the standard formulation of this model (Chapter 7). For large p the model is equivalent to the XY model. The XY model will be discussed in Chapter 15 – at this moment it is sufficient to know that this model is critical for all temperatures between 0 and T_{KT}, which corresponds to $\beta J \approx 1.1$ (the subscript KT denotes the Kosterlitz–Thouless phase transition; see Chapter 15). The central charge c (see Section 11.3) is equal to 1 on this critical line.

Table 12.3. *Largest eigenvalues of the transfer matrix of the Ising model on a strip with helical boundary conditions (Figure 12.8) versus strip width L.*

L	$\ln\lambda_0$ (MCTM)	$\ln\lambda_0$ (Lanczos)
6	0.9368(2)	0.9369
7	0.9348(2)	0.9350
8	0.9337(2)	0.9338
9	0.9328(2)	0.9329
10	0.9321(2)	0.9323
11	0.9316(2)	0.9318

The target number of walkers is equal to 5000, and they performed 10 000 diffusion steps. The third column gives the eigenvalues obtained by diagonalising the full transfer matrix using the Lanczos method. These values are determined with high accuracy and are rounded to four significant digits.

The walkers are 'columns' of lattice spins, $(s_0, \ldots, s_{L-1})$, as represented in Figure 12.8. In the diffusion step, a new spin is added to the system, and its value is the s_0-component of the new configuration of the walker. The spin components 1 to $L-1$ of the new configuration are filled with the components 0 to $L-2$ of the old walker – the walker is shifted one position over the cylinder. To sample the new s_0'-value, we use the 'shooting method' in which the interval [0, 1] is divided up into p segments corresponding to the conditional probability $P(s_0'|S)$ which is proportional to the Boltzmann factor for adding a spin $s_0' = 0, \ldots, p-1$ to the existing column S. In our clock model example, we have

$$P(s_0'|S) = \mathrm{e}^{J\cos(s_0'-s_0)+J\cos(s_0'-s_{L-1})}/D(S), \tag{12.104}$$

with normalisation factor

$$D(S) = \sum_{s_0'} \mathrm{e}^{J\cos(s_0'-s_0)+J\cos(s_0'-s_{L-1})}. \tag{12.105}$$

A random number between 0 and 1 is then generated and the new spin value corresponds to the index of the segment in which the random number falls.

The next step is then the assignment of the weight $D(S')$ to the walker with D given in (12.105). Branching is then carried out exactly as in the DMC method. In fact, the weights are also multiplied by a factor $\exp(E_{\mathrm{trial}})$, where E_{trial} is the same for all walkers but varies in time. It is updated as in the DMC method according to

$$E_{\mathrm{trial}} = E_0 - \alpha \ln(N/N_0), \tag{12.106}$$

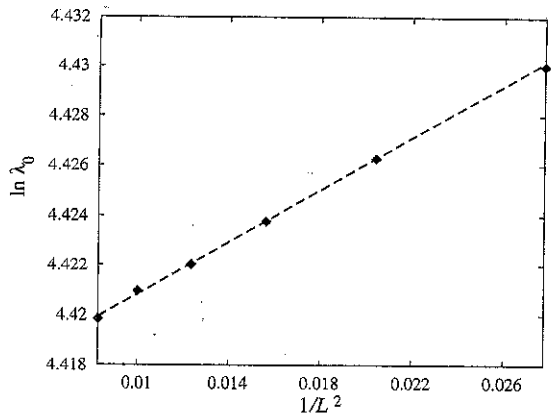

Figure 12.9. The logarithm of the largest eigenvalue of the transfer matrix versus the inverse of the square of the strip width L. The straight line has a slope $\pi/6$ and is adjusted in height to fit the data.

where E_0 is a guess of the trial energy (which should be equal to $-\ln\lambda_0$, λ_0 is the largest eigenvalue), N is the actual number of walkers and N_0 is the target number of walkers. This term aims at stabilising the population size to the target number N_0.

The simplest information we obtain is the largest eigenvalue, which is given as $\exp(E_{\mathrm{trial}})$, where the average value of E_{trial} during the simulation is to be used (with the usual omission of equilibration steps). This can be used to determine central charges. In Table 12.3 we compare the values of this quantity for the Ising model with those obtained by a Lanczos diagonalisation of the transfer matrix. The agreement is seen to be excellent. For the XY model, the eigenvalues cannot be found using direct diagonalisation and we can check the MCTM method only by comparing the central charge obtained with the known value: 1 in the low-temperature phase and 0 at high temperatures. In Figure 12.9 we show the results for $\beta J = 1.25$. The points in a graph of the form $\ln\lambda_0$ vs. $1/L^2$ lie on a straight curve with a slope of $\pi/6$ ($c = 1$).

Exercises

12.1 In this problem we consider the virial expression for the energy [20].

In a path-integral QMC simulation for a particle in one dimension in a potential $V(x)$ we want to find the energy E as a function of temperature $T = 1/(k_{\mathrm{B}}\beta)$. We do this by using the thermodynamic relation

$$E = -\frac{\partial \ln Z}{\partial \beta}$$

(a) Show that for a two-dimensional table containing values of a function $f(x, y)$ for integer x and y, the value $f(x, y)$ for arbitrary x and y within the boundaries set by the table size is given as

$$f(x,y) = (2 - x - y + [x] + [y])f([x],[y]) \tag{12.107}$$
$$+ (1 + x - [x] - y + [y])f([x]+1,[y]) \tag{12.108}$$
$$+ (1 + y - [y] - x + [x])f([x],[y]+1) \tag{12.109}$$
$$+ (x - [x] + y - [y])f([x]+1,[y]+1).$$

Here $[x]$ denotes the largest integer smaller than x.

(b) Find analogous expressions for a table with a noninteger (but equidistant) spacing between the table entries and also for a three-dimensional Table.

12.5 [C]

In this problem we consider applying variational Monte Carlo to the hydrogen molecule. There are two complications in comparison with the helium atom. One is the calculation of the local energy which is quite cumbersome, although straightforward. The second one is the cusp condition.

To specify the trial wave function we take the nuclei at positions $\pm s/2$. A one-particle orbital has the form (in atomic units):

$$\phi(\mathbf{r}) = \mathrm{e}^{-|\mathbf{r}-s\hat{\mathbf{x}}/2|/a} + \mathrm{e}^{-|\mathbf{r}+s\hat{\mathbf{x}}/2|/a}$$

where a is some parameter. The two-electron wave function is given as

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \phi(\mathbf{r}_1)\phi(\mathbf{r}_2)f(r_{12})$$

with f the Jastrow factor

$$f(r) = \exp\left(\frac{r}{\alpha(1+\beta r_{12})}\right).$$

(a) Show that the Coulomb-cusp condition near the nuclei leads to the relation

$$\frac{1}{1+\exp(-s/a)} = a.$$

For a given distance s, this equation should be solved numerically to give you the value a.

(b) Show that the electron–electron cusp condition leads to the requirement $a = 2$. This leaves a single parameter β in the wave function.

(c) Now you can implement the hydrogen molecule in VMC. Calculate the energy as a function of the parameters β and s and find the minimum.

(d) You may also evaluate the ground state by applying the method of Harju *et al.* [6] which was described in Section 12.2, in order to update the values of β and s simultaneously.

(e) What would you need in order to calculate the molecular formation energy from this? Note that this is the difference between the energy of the hydrogen molecule and that of two isolated hydrogen atoms. Consider in particular the contribution arising from the nuclear motion.

References

[1] P. A. M. Dirac, *The Principles of Quantum Mechanics*. Oxford, Oxford University Press, 1958.

[2] B. L. Hammond, W. A. Lester Jr, and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry*. Singapore, World Scientific, 1994.

[3] D. M. Ceperley and E. L. Pollock, 'Path-integral computation of the low-temperature properties of liquid-He-4,' *Phys. Rev. Lett.*, **56** (1986), 351–4.

[4] D. M. Ceperley, 'Path integrals in the theory of condensed helium,' *Rev. Mod. Phys.*, **67** (1995), 279–355.

[5] H. De Raedt and A. Lagendijk, 'Monte Carlo simulations of quantum statistical lattice models,' *Phys. Rep.*, **127** (1985), 233–307.

[6] A. Harju, S. Siljamäki, and R. M. Nieminen, 'Wigner molecules in quantum dots: a quantum Monte Carlo study,' *Phys. Rev. E*, **65** (2002), 075309.

[7] R. J. Jastrow, 'Many-body problem with strong forces,' *Phys. Rev.*, **98** (1955), 1479–84.

[8] M. H. Kalos, D. Levesque, and L. Verlet, 'Helium at zero temperature with hard-sphere and other forces,' *Phys. Rev. A*, **9** (1974), 2178–95.

[9] R. N. Barnett, P. J. Reynolds, and W. A. Lester Jr, 'Monte Carlo determination of the oscillator strength and excited state lifetime for the Li $2^2S \rightarrow 2^2P$ transition,' *Int. J. Quantum Chem.*, **42** (1992), 837–47.

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

[11] M. Suzuki, 'Decomposition formulas of exponential operators and Lie exponents with some applications to quantum-mechanics and statistical physics,' *J. Math. Phys.*, **26** (1985), 601–12.

[12] M. Suzuki, 'Transfer-matrix method and Monte Carlo simulation in quantum spin systems,' *Phys. Rev. B*, **31** (1985), 2957–65.

[13] J. D. Doll, R. D. Coalson, and D. L. Freeman, 'Towards a Monte Carlo theory of quantum dynamics,' *J. Chem. Phys.*, **87** (1987), 1641–7.

[14] V. S. Filinov, 'Calculation of the Feynman integrals by means of the Monte Carlo method,' *Nucl. Phys. B*, **271** (1986), 717–25.

[15] J. Chang and W. H. Miller, 'Monte Carlo path integration in real-time via complex coordinates,' *J. Chem. Phys.*, **87** (1987), 1648–52.

[16] R. P. Feynman, 'The λ-transition in liquid helium,' *Phys. Rev.*, **90** (1953), 1116–17.

[17] R. P. Feynman, 'Atomic theory of the λ-transition in helium,' *Phys. Rev.*, **91** (1953), 1291–301.

[18] R. P. Feynman, 'Atomic theory of liquid helium near absolute zero,' *Phys. Rev.*, **91** (1953), 1301–8.

[19] D. M. Ceperley and E. L. Pollock, 'Path-integral computation techniques for superfluid ^{4}He,' in *Monte Carlo Methods in Theoretical Physics* (S. Caracciolo and A. Fabrocini, eds.). Pisa, Italy, ETS Editrice, 1992, p. 35.

[20] M. F. Herman, E. J. Bruskin, and B. J. Berne, 'On path integral Monte-Carlo simulations,' *J. Chem. Phys.*, **76** (1982), 5150–5.

[21] K. Singer and W. Smith, 'Path integral simulation of condensed phase Lennard–Jones systems,' *Mol. Phys.*, **64** (1988), 1215–31.

[22] D. M. Ceperley, 'The simulation of quantum systems with random walks – a new algorithm for charged systems,' *J. Comp. Phys.*, **51** (1983), 404–22.

[23] E. L. Pollock, 'Properties and computation of the Coulomb pair density matrix,' *Comp. Phys. Comm.*, **52** (1989), 49–60.

[24] M. Takahashi and M. Imada, 'Monte Carlo calculation of quantum-systems,' *J. Phys. Soc. Jpn*, **53** (1984), 963–74.

[25] J. D. Doll, R. D. Coalson, and D. L. Freeman, 'Solid-fluid phase transition of quantum hard-spheres at finite temperatures,' *Phys. Rev. Lett.*, **55** (1985), 1–4.

[26] M. Plischke and H. Bergersen, *Equilibrium Statistical Physics*. Englewood Cliffs, NJ, Prentice-Hall, 1989.
[27] J. E. Hirsch, 'Discrete Hubbard–Stratonovich transformation for fermion lattice models,' *Phys. Rev. B*, **28** (1983), 4059–61.
[28] J. W. Negele and H. Orland, *Quantum Many-Particle Systems*. Redwood City, Addison-Wesley, 1988.
[29] D. B. F. ten Haaf, H. J. M. van Bemmel, J. M. J. van Leeuwen, W. van Saarloos, and D. M. Ceperley, 'Proof for an upper bound in fixed-node Monte Carlo for lattice fermions,' *Phys. Rev. B*, **51** (1995), 13039–45.
[30] M. P. Nightingale and H. W. J. Blöte, 'Monte Carlo calculation for the free energy, critical point and surface critical behaviour of three-dimensional Heisenberg ferromagnets,' *Phys. Rev. Lett.*, **60** (1988), 1562–5.
[31] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*. Amsterdam, North-Holland, 1981.

13

The finite element method for partial differential equations

13.1 Introduction

When we consider a partial differential equation, such as the ubiquitous Laplace equation

$$\nabla^2 \phi(\mathbf{r}) = 0, \tag{13.1}$$

together with some boundary condition(s), the obvious way of solving it that comes to mind is to discretise this equation on a regular grid, hoping that this grid can match the boundary in some way. Then we solve the discretised problem using, for example, iterative methods such as the Gauss–Seidel or conjugate gradients method (see Appendix A7.2). For many problems, this approach is adequate, but if the problem is difficult in the sense that it has a lot of structure on small scales in some region of the domain, or if the boundary has a complicated shape which is difficult to match with a regular grid, it might be useful to apply methods that allow for flexibility of the grid on which the solution is formulated. In this chapter we discuss such a method, the *finite element method*.

One way of looking at the finite element method (FEM) is by realising that many partial differential equations can be viewed as solution methods for variational problems. In the case of the Laplace equation with zero boundary condition, for example, finding the stationary solution of the functional

$$\int_D [\nabla \phi(\mathbf{r})]^2 \, \mathrm{d}^d r, \tag{13.2}$$

where the integral is over the d-dimensional domain D and where we confine ourselves to functions $\phi(\mathbf{r})$ which vanish on the domain boundary, yields the same solution as that of the Laplace equation – in fact, the Laplace equation is the Euler equation for this functional (see the next section).