CHAPTER SIX

Computer Simulation Methods

6.1 Introduction

Energy minimisation generates individual minimum energy configurations of a system. In some cases the information provided by energy minimisation can be sufficient to predict accurately the properties of a system. If all minimum configurations on an energy surface can be identified then statistical mechanical formulae can be used to derive a partition function from which thermodynamic properties can be calculated. However, this is possible only for relatively small molecules or small molecular assemblies in the gas phase. The molecular modeller more often wants to understand and to predict the properties of liquids, solutions and solids, to study complex processes such as the adsorption of molecules onto surfaces and into solids and to investigate the behaviour of macromolecules which have many closely separated minima. In such systems the experimental measurements are made on macroscopic samples that contain extremely large numbers of atoms or molecules, with an enormous number of minima on their energy surfaces. A full quantification of the energy surfaces of such systems is not possible, nor is it ever likely to be. Computer simulation methods enable us to study such systems and predict their properties through the use of techniques that consider small replications of the macroscopic system with manageable numbers of atoms or molecules. A simulation generates representative configurations of these small replications in such a way that accurate values of structural and thermodynamic properties can be obtained with a feasible amount of computation. Simulation techniques also enable the time-dependent behaviour of atomic and molecular systems to be determined, providing a detailed picture of the way in which a system changes from one conformation or configuration to another. Simulation techniques are also widely used in some experimental procedures, such as the determination of protein structures from X-ray crystallography.

In this chapter we shall discuss some of the general principles involved in the two most common simulation techniques used in molecular modelling: the molecular dynamics and the Monte Carlo methods. We shall also discuss several concepts that are common to both of these methods. A more detailed discussion of the two simulation methods can be found in Chapters 7 and 8.

6.1.1 Time Averages, Ensemble Averages and Some Historical Background

Suppose we wish to determine experimentally the value of a property of a system such as the pressure or the heat capacity. In general, such properties will depend upon the positions and
momenta of the $N$ particles that comprise the system. The instantaneous value of the property $A$ can thus be written as $A(p^N(t), r^N(t))$, where $p^N(t)$ and $r^N(t)$ represent the $N$ momenta and positions respectively at time $t$ (i.e. $A(p^N(t), r^N(t)) = A(p_{1x}, p_{1y}, p_{1z}, p_{2x}, \ldots, x_1, y_1, z_1, x_2, \ldots, t)$ where $p_{1x}$ is the momentum of particle 1 in the $x$ direction and $x_1$ is its $x$ coordinate). Over time, the instantaneous value of the property $A$ fluctuates as a result of interactions between the particles. The value that we measure experimentally is an average of $A$ over the time of the measurement and is therefore known as a time average. As the time over which the measurement is made increases to infinity, so the value of the following integral approaches the 'true' average value of the property:

$$A_{\text{ave}} = \lim_{T \to \infty} \frac{1}{T} \int_{t=0}^{T} A(p^N(t), r^N(t)) \, dt$$  \hfill (6.1)

To calculate average values of the properties of the system, it would therefore appear to be necessary to simulate the dynamic behaviour of the system (i.e. to determine values of $A(p^N(t), r^N(t))$, based upon a model of the intra- and intermolecular interactions present). In principle, this is relatively straightforward to do. For any arrangement of the atoms in the system, the force acting on each atom due to interactions with other atoms can be calculated by differentiating the energy function. From the force on each atom it is possible to determine its acceleration via Newton's second law. Integration of the equations of motion should then yield a trajectory that describes how the positions, velocities and accelerations of the particles vary with time, and from which the average values of properties can be determined using the numerical equivalent of Equation (6.1). The difficulty is that for 'macroscopic' numbers of atoms or molecules (of the order of $10^{23}$) it is not even feasible to determine an initial configuration of the system, let alone integrate the equations of motion and calculate a trajectory. Recognising this problem, Boltzmann and Gibbs developed statistical mechanics, in which a single system evolving in time is replaced by a large number of replications of the system that are considered simultaneously. The time average is then replaced by an ensemble average:

$$\langle A \rangle = \int \, dp^N \, dr^N A(p^N, r^N) \rho(p^N, r^N)$$  \hfill (6.2)

The angle brackets ($\langle \rangle$) indicate an ensemble average, or expectation value; that is, the average value of the property $A$ over all replications of the ensemble generated by the simulation. Equation (6.2) is written as a double integral for convenience but in fact there should be $6N$ integral signs on the integral for the $6N$ positions and momenta of all the particles. $\rho(p^N, r^N)$ is the probability density of the ensemble; that is, the probability of finding a configuration with momenta $p^N$ and positions $r^N$. The ensemble average of the property $A$ is then determined by integrating over all possible configurations of the system. In accordance with the ergodic hypothesis, which is one of the fundamental axioms of statistical mechanics, the ensemble average is equal to the time average. Under conditions of constant number of particles, volume and temperature, the probability density is the familiar Boltzmann distribution:

$$\rho(p^N, r^N) = \exp(-E(p^N, r^N)/k_B T)/Q$$  \hfill (6.3)

In Equation (6.3), $E(p^N, r^N)$ is the energy, $Q$ is the partition function, $k_B$ is Boltzmann's constant and $T$ is the temperature. The partition function is more generally written in terms of the Hamiltonian, $\mathcal{H}$; for a system of $N$ identical particles the partition function
The instantaneous value of the function $\mathcal{H}(t)$ and $r^N(t)$ represent the $N$ particles, and $r^N(t) = (r_1, r_2, r_3, \ldots)$, where $x_i$ is its position in $x$ direction and $x_1$ is its property $A$ fluctuates as a result of thermal motion. To be measured experimentally, it is therefore known as a time average. Thus the time average of the $x_1$ property would therefore appear to be

$$\langle x_1 \rangle = \int dt \langle x_1(t) \rangle$$

(6.1)

This is called a mean average value or a statistical average value, that is, the average position of the particle that is generated by the statistical ensemble. In fact there should be $6N$ equations, one for each particle $\mathcal{H}(p^{N}, r^{N})$, and the one for finding a configuration of the property $A$ is then determined. The statistical mechanics, the ensemble average, is in some sense the most important number of particles, Boltzmann distribution:

$$Q_{NVT} = \frac{1}{N! \hbar^{3N}} \int dp^N dt^N \exp \left[-\frac{\mathcal{H}(p^N, r^N)}{k_B T}\right]$$

(6.4)

The canonical ensemble is the name given to an ensemble for constant temperature, number of particles and volume. For our purposes $\mathcal{H}$ can be considered the same as the total energy, $E(p^N, r^N)$, which equals the sum of the kinetic energy ($\mathcal{H}(p^N)$) of the system, which depends upon the momenta of the particles, and the potential energy ($\mathcal{H}(r^N)$), which depends upon the positions. The factor $N!$ arises from the indistinguishability of the particles and the factor $1/\hbar^{3N}$ is required to ensure that the partition function is equal to the quantum mechanical result for a particle in a box. A short discussion of some of the key results of statistical mechanics is provided in Appendix 6.1 and further details can be found in standard textbooks.

The first computer simulations of fluids were performed in 1952 by Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, who developed a scheme for sampling from the Boltzmann distribution to give ensemble averages. This gave rise to the Monte Carlo simulation method. Not long afterwards (in 1957) Alder recognised that it was, in fact, possible to integrate the equations of motion for a relatively small number of particles, and to mimic the behaviour of a real system using periodic boundary conditions. This led to the first molecular dynamics simulations of molecular systems.

### 6.1.2 A Brief Description of the Molecular Dynamics Method

Molecular dynamics calculates the "real" dynamics of the system, from which time averages of properties can be calculated. Sets of atomic positions are derived in sequence by applying Newton's equations of motion. Molecular dynamics is a deterministic method, by which we mean that the state of the system at any future time can be predicted from its current state. The first molecular dynamics simulations were performed using very simple potentials such as the hard-sphere potential. The behaviour of the particles in this potential is similar to that of billiard or snooker balls: the particles move in straight lines at constant velocity between collisions. The collisions are perfectly elastic and occur when the separation between a pair of spheres equals the sum of their radii. After a collision, the new velocities of the colliding spheres are calculated using the principle of conservation of linear momentum. The hard-sphere model has provided many useful results but is obviously not ideal for simulating atomic or molecular systems. In potentials such as the Lennard-Jones potential the force between two atoms or molecules changes continuously with their separation. By contrast, in the hard-sphere model there is no force between particles until they collide. The continuous nature of the more realistic potentials requires the equations of motion to be integrated by breaking the calculation into a series of very short time steps (typically between 1 femtosecond and 10 femtoseconds; $10^{-15}$ s to $10^{-14}$ s). At each step, the forces on the atoms are computed and combined with the current positions and velocities to generate new positions and velocities a short time ahead. The force acting on each atom is assumed to be constant during the time interval. The atoms are then moved to the new positions, an updated set of forces is computed, and so on. In this way a molecular dynamics simulation generates a
**Trajectory** that describes how the dynamic variables change with time. Molecular dynamics simulations are typically run for tens or hundreds of picoseconds (a 100 ps simulation using a 1 fs time step requires 100 000 steps). Thermodynamic averages are obtained from molecular dynamics as time averages using numerical integration of Equation (6.2):

\[
(A) = \frac{1}{M} \sum_{i=1}^{M} A(p_i^N, r_i^N)
\]

(6.5)

\(M\) is the number of time steps. Molecular dynamics is also extensively used to investigate the conformational properties of flexible molecules as will be discussed in Chapters 7 and 9.

### 6.1.3 The Basic Elements of the Monte Carlo Method

In a molecular dynamics simulation the successive configurations of the system are connected in time. This is not the case in a Monte Carlo simulation, where each configuration depends only upon its predecessor and not upon any other of the configurations previously visited. The Monte Carlo method generates configurations randomly and uses a special set of criteria to decide whether or not to accept each new configuration. These criteria ensure that the probability of obtaining a given configuration is equal to its Boltzmann factor, \(\exp[-\mathcal{V}(r^N)/k_B T]\), where \(\mathcal{V}(r^N)\) is calculated using the potential energy function. States with a low energy are thus generated with a higher probability than configurations with a higher energy. For each configuration that is accepted the values of the desired properties are calculated and at the end of the calculation the average of these properties is obtained by simply averaging over the number of values calculated, \(M\):

\[
\langle A \rangle = \frac{1}{M} \sum_{i=1}^{M} A(x_i^N)
\]

(6.6)

Most Monte Carlo simulations of molecular systems are more properly referred to as Metropolis Monte Carlo calculations after Metropolis and his colleagues, who reported the first such calculation. The distinction can be important because there are other ways in which an ensemble of configurations can be generated. As we shall see in Chapter 7, the Metropolis scheme is only one of a number of possibilities, though it is by far the most popular.

In a Monte Carlo simulation each new configuration of the system may be generated by randomly moving a single atom or molecule. In some cases new configurations may also be obtained by moving several atoms or molecules or by rotating about one or more bonds. The energy of the new configuration is then calculated using the potential energy function. If the energy of the new configuration is lower than the energy of its predecessor then the new configuration is accepted. If the energy of the new configuration is higher than the energy of its predecessor then the Boltzmann factor of the energy difference is calculated: \(\exp[-(\mathcal{V}_{\text{new}}(r^N) - \mathcal{V}_{\text{old}}(r^N))/k_B T]\). A random number is then generated between 0 and 1 and compared with this Boltzmann factor. If the random number is higher than the Boltzmann factor then the move is rejected and the original configuration is retained for the next iteration; if the random number is lower then the move is accepted and the new
configuration becomes the next state. This procedure has the effect of permitting moves to states of higher energy. The smaller the uphill move (i.e. the smaller the value of \( v_{\text{new}}(t^n) - v_{\text{old}}(t^n) \)) the greater is the probability that the move will be accepted.

6.1.4 Differences Between the Molecular Dynamics and Monte Carlo Methods

The molecular dynamics and Monte Carlo simulation methods differ in a variety of ways. The most obvious difference is that molecular dynamics provides information about the time dependence of the properties of the system whereas there is no temporal relationship between successive Monte Carlo configurations. In a Monte Carlo simulation the outcome of each trial move depends only upon its immediate predecessor, whereas in molecular dynamics it is possible to predict the configuration of the system at any time in the future – or indeed at any time in the past. Molecular dynamics has a kinetic energy contribution to the total energy whereas in a Monte Carlo simulation the total energy is determined directly from the potential energy function. The two simulation methods also sample from different ensembles. Molecular dynamics is traditionally performed under conditions of constant number of particles (\( N \)), volume (\( V \)) and energy (\( E \)) (the microcanonical or constant NVE ensemble) whereas a traditional Monte Carlo simulation samples from the canonical ensemble (constant \( N, V \) and temperature, \( T \)). Both the molecular dynamics and Monte Carlo techniques can be modified to sample from other ensembles; for example, molecular dynamics can be adapted to simulate from the canonical ensemble. Two other ensembles are common:

- isothermal-isobaric: fixed \( N, T, P \) (pressure)
- grand canonical: fixed \( \mu \) (chemical potential), \( V, T \)

In the canonical, microcanonical and isothermal-isobaric ensembles the number of particles is constant but in a grand canonical simulation the composition can change (i.e. the number of particles can increase or decrease). The equilibrium states of each of these ensembles are characterised as follows:

- canonical ensemble: minimum Helmholtz free energy (A)
- microcanonical ensemble: maximum entropy (S)
- isothermal-isobaric ensemble: minimum Gibbs function (G)
- grand canonical ensemble: maximum pressure x volume (PV)

6.2 Calculation of Simple Thermodynamic Properties

A wide variety of thermodynamic properties can be calculated from computer simulations; a comparison of experimental and calculated values for such properties is an important way in which the accuracy of the simulation and the underlying energy model can be quantified. Simulation methods also enable predictions to be made of the thermodynamic properties of systems for which there is no experimental data, or for which experimental data is difficult or impossible to obtain. Simulations can also provide structural information about the
conformational changes in molecules and the distributions of molecules in a system. The emphasis in our discussion will be on those properties that are routinely calculated in computer simulations and on the way in which they are obtained. It is important to recognise that the results we derive are for the canonical ensemble. Sometimes the equivalent expressions in other ensembles are provided. The result obtained from one ensemble may also be transformed to another ensemble, though this is strictly only possible in the limit of an infinitely large system. The expressions follow from standard statistical mechanical relationships, which are given in standard texts and summarised in Appendix 6.1.

6.2.1 Energy

The internal energy is easily obtained from a simulation as the ensemble average of the energies of the states that are examined during the course of the simulation:

$$U = \langle E \rangle = \frac{1}{M} \sum_{i=1}^{M} E_i$$ \hspace{1cm} (6.7)

6.2.2 Heat Capacity

At a phase transition the heat capacity will often show a characteristic dependence upon the temperature (a first-order phase transition is characterised by an infinite heat capacity at the transition but in a second-order phase transition the heat capacity changes discontinuously). Monitoring the heat capacity as a function of temperature may therefore enable phase transitions to be detected. Calculations of the heat capacity can also be compared with experimental results and so be used to check the energy model or the simulation protocol.

The heat capacity is formally defined as the partial derivative of the internal energy with respect to temperature:

$$C_V = \left( \frac{\partial U}{\partial T} \right)_V$$ \hspace{1cm} (6.8)

The heat capacity can therefore be calculated by performing a series of simulations at different temperatures, and then differentiating the energy with respect to the temperature. The differentiation can be done numerically or by fitting a polynomial to the data and then analytically differentiating the fitted function. The heat capacity may also be calculated from a single simulation by considering the instantaneous fluctuations in the energy as follows:

$$C_V = \frac{\langle (E^2) - \langle E \rangle^2 \rangle}{k_B T^2}$$ \hspace{1cm} (6.9)

An alternative way to write this expression uses the relationship:

$$\langle (E - \langle E \rangle)^2 \rangle = \langle E^2 \rangle - \langle E \rangle^2$$ \hspace{1cm} (6.10)

giving:

$$C_V = \frac{\langle (E - \langle E \rangle)^2 \rangle}{k_B T^2}$$ \hspace{1cm} (6.11)

A derivation of this result is provided in Appendix 6.2.
molecules in a system. The heat capacities are routinely calculated in this way. It is important to recognize that sometimes the equivalent heat capacity from one ensemble may be more realistic than that calculated from another ensemble. The heat capacities calculated from one ensemble may be only possible in the limit of the experimental system. The standard statistical mechanical calculations are used in Appendix 6.1.

The ensemble average of the internal energy with respect to the temperature is

\[ \langle E \rangle = k_B T. \]  

The first derivative of the internal energy with respect to the temperature is

\[ \gamma = \left. \frac{\partial E}{\partial T} \right|_{V} = -\frac{1}{T} \left( \frac{\partial W}{\partial V} \right). \]  

The second derivative of the internal energy is

\[ \chi = \left. \frac{\partial^2 E}{\partial T^2} \right|_{V} = -\frac{1}{T^2} \left( \frac{\partial W}{\partial V} \right) + \frac{1}{T} \left( \frac{\partial^2 W}{\partial V^2} \right). \]  

The third derivative of the internal energy is

\[ \eta = \left. \frac{\partial^3 E}{\partial T^3} \right|_{V} = -\frac{1}{T^3} \left( \frac{\partial W}{\partial V} \right) + \frac{1}{T^2} \left( \frac{\partial^2 W}{\partial V^2} \right) - \frac{1}{T} \left( \frac{\partial^3 W}{\partial V^3} \right). \]  

The heat capacity can therefore be obtained by keeping a running count of \( E^2 \) and \( E \) during the simulation, from which their expectation values \( \langle E^2 \rangle \) and \( \langle E \rangle \) can be calculated at the end of the calculation. Alternatively, if the energies are stored during the simulation then the value of \( \langle (E - \langle E \rangle)^2 \rangle \) can be calculated once the simulation has finished. This second approach may be more accurate due to round-off errors; \( \langle E^2 \rangle \) and \( \langle E \rangle^2 \) are usually very large numbers so there may be a large uncertainty in their difference.

### 6.2.3 Pressure

The pressure is usually calculated in a computer simulation via the virial theorem of Clausius. The virial is defined as the expectation value of the sum of the products of the coordinates of the particles and the forces acting on them. This is usually written

\[ W = \sum_{i=1}^{N} x_i \hat{p}_i, \]

where \( x_i \) is a coordinate (e.g., the \( x \) or \( y \) coordinate of an atom) and \( \hat{p}_i \) is the first derivative of the momentum along that coordinate (\( \hat{p}_i \) is the force, by Newton’s second law). The virial theorem states that the virial is equal to \(-3Nk_B T\).

In an ideal gas, the only forces are those due to interactions between the gas and the container and it can be shown that the virial in this case equals \(-3PV\). This result can also be obtained directly from \( PV = Nk_B T \).

Forces between the particles in a real gas or liquid affect the virial, and hence the pressure. The total virial for a real system equals the sum of an ideal gas part \((-3PV)\) and a contribution due to interactions between the particles. The result obtained is:

\[ W = -3PV + \sum_{i=1}^{N} r_{ij} \frac{d\nu(r_{ij})}{dr_{ij}} = -3Nk_B T \]  

The real gas part is derived in Appendix 6.3. If \( d\nu(r_{ij})/dr_{ij} \) is written as \( f_{ij} \), the force acting between atoms \( i \) and \( j \), then we have the following expression for the pressure:

\[ P = \frac{1}{V} \left[ \frac{3Nk_B T}{3} \sum_{i=1}^{N} \sum_{j=1}^{N} r_{ij} f_{ij} \right] \]

The forces are calculated as part of a molecular dynamics simulation, and so little additional effort is required to calculate the virial and thus the pressure. The forces are not routinely calculated during a Monte Carlo simulation, and so some additional effort is required to determine the pressure by this route. When calculating the pressure it is also important to check that the components of the pressure in all three directions are equal.

### 6.2.4 Temperature

In a canonical ensemble the total temperature is constant. In the microcanonical ensemble, however, the temperature will fluctuate. The temperature is directly related to the kinetic energy of the system as follows:

\[ \mathcal{E} = \sum_{i=1}^{N} \frac{|p_i|^2}{2m_i} = \frac{k_B T}{2} (3N - N_c) \]

\[ N_c \]
In this equation, $p_i$ is the total momentum of particle $i$ and $m_i$ is its mass. According to the theorem of the equipartition of energy each degree of freedom contributes $k_B T / 2$. If there are $N$ particles, each with three degrees of freedom, then the kinetic energy should equal $3Nk_B T / 2$. $N_c$ in Equation (6.14) is the number of constraints on the system. In a molecular dynamics simulation the total linear momentum of the system is often constrained to a value of zero, which has the effect of removing three degrees of freedom from the system and so $N_c$ would be equal to 3. Other types of constraint are also possible as we shall discuss in Section 7.5.

### 6.2.5 Radial Distribution Functions

Radial distribution functions are a useful way to describe the structure of a system, particularly of liquids. Consider a spherical shell of thickness $\delta r$ at a distance $r$ from a chosen atom (Figure 6.1). The volume of the shell is given by:

$$
V = \frac{4}{3} \pi (r + \delta r)^3 - \frac{4}{3} \pi r^3 \\
= 4\pi r^2 \delta r + 4\pi r \delta r^2 + \frac{4}{3} \pi \delta r^3 \\
\approx 4\pi r^2 \delta r
$$

(6.15)

If the number of particles per unit volume is $\rho$, then the total number in the shell is $4\pi \rho r^2 \delta r$ and so the number of atoms in the volume element varies as $r^2$.

The pair distribution function, $g(r)$, gives the probability of finding an atom (or molecule, if simulating a molecular fluid) a distance $r$ from another atom (or molecule) compared to the ideal gas distribution. $g(r)$ is thus dimensionless. Higher radial distribution functions (e.g. the triplet radial distribution function) can also be defined but are rarely calculated and so references to the ‘radial distribution function’ are usually taken to mean the pairwise version. In a crystal, the radial distribution function has an infinite number of sharp peaks whose separations and heights are characteristic of the lattice structure.

The radial distribution function of a liquid is intermediate between the solid and the gas, with a small number of peaks as short distances, superimposed on a steady decay to a constant value at longer distances. The radial distribution function calculated from a molecular dynamics simulation of liquid argon (shown in Figure 6.2) is typical. For short distances (less
is its mass. According to the equipartition theorem, the kinetic energy should equal \( k_B T / 2 \). If there are \( n \) degrees of freedom in the system, in a molecular model, the system is often constrained to a finite number of degrees of freedom from the system so that in a liquid or gas natural state as possible as we shall discuss later.

![Radial distribution function](image)

Fig. 6.2: Radial distribution function determined from a 100 ps molecular dynamics simulation of liquid argon at a temperature of 100 K and a density of 1.396 g cm\(^{-3}\).

than the atomic diameter) \( g(r) \) is zero. This is due to the strong repulsive forces. The first (and largest) peak occurs at \( r \approx 3.7 \text{ Å} \), with \( g(r) \) having a value of about 3. This means that it is three times more likely that two molecules would have this separation than in the ideal gas. The radial distribution function then falls and passes through a minimum value around \( r \approx 5.4 \text{ Å} \). The chances of finding two atoms with this separation are less than for the ideal gas. At long distances, \( g(r) \) tends to the ideal gas value, indicating that there is no long-range order.

To calculate the pair distribution function from a simulation, the neighbours around each atom or molecule are sorted into distance 'bins', or histograms. The number of neighbours in each bin is then averaged over the entire simulation. For example, a count is made of the number of neighbours between (say) 2.5 Å and 2.75 Å, 2.75 Å and 3.0 Å and so on for every atom or molecule in the simulation. This count can be performed during the simulation itself or by analysing the configurations that are generated.

Radial distribution functions can be measured experimentally using X-ray diffraction. The regular arrangement of the atoms in a crystal gives the characteristic X-ray diffraction pattern with bright, sharp spots. For liquids, the diffraction pattern has regions of high and low intensity but no sharp spots. The X-ray diffraction pattern can be analysed to calculate an experimental distribution function, which can then be compared with that obtained from the simulation.

Thermodynamic properties can be calculated using the radial distribution function, if pairwise additivity of the forces is assumed. These properties are usually given as an ideal gas part plus a real gas part. For example, to calculate the energy of a real gas, we consider the spherical shell of volume \( 4\pi r^2 \Delta r \) that contains \( 4\pi r^2 \rho g(r) \Delta r \) particles. If the pair potential at a distance \( r \) has a value \( \phi(r) \) then the energy of interaction between the particles in the shell and the central particle is \( 4\pi r^2 \rho g(r) \phi'(r) \Delta r \). The total potential energy of the real gas is obtained by integrating this between 0 and \( \infty \) and multiplying the result...
by \( N/2 \) (the factor \( 1/2 \) ensures that we only count each interaction once). The total energy is then given by:

\[
E = \frac{3}{2} N k_B T + 2\pi N \rho \int_0^\infty r^2 \phi(r) g(r) \, dr
\]  
(6.16)

In a similar way the following expression for the pressure can be derived:

\[
PV = N k_B T - \frac{2\pi N \rho}{3k_B T} \int_0^\infty r^2 \frac{d\phi(r)}{dr} g(r) \, dr
\]  
(6.17)

It is usually more accurate to calculate such properties directly, partly because the radial distribution function is not obtained as a continuous function but is derived by dividing the space into small but discrete bins.

For molecules, the orientation must be taken into account if the true nature of the distribution is to be determined. The radial distribution function for molecules is usually measured between two fixed points, such as between the centres of mass. This may then be supplemented by an orientational distribution function. For linear molecules, the orientational distribution function may be calculated as the angle between the axes of the molecule, with values ranging from \(-180^\circ\) to \(+180^\circ\). For more complex molecules it is usual to calculate a number of site-site distribution functions. For example, for a three-site model of water, three functions can be defined \( (g(O-O), g(O-H) \text{ and } g(H-H)) \). An advantage of the site-site models is that they can be directly related to information obtained from the X-ray scattering experiments. The O-O, O-H and H-H radial distribution functions have been particularly useful for refining the various potential models for simulating liquid water.

### 6.3 Phase Space

An important concept in computer simulation is that of the phase space. For a system containing \( N \) atoms, \( 6N \) values are required to define the state of the system (three coordinates per atom and three components of the momentum). Each combination of \( 3N \) positions and \( 3N \) momenta (usually denoted by \( \Gamma_N \)) defines a point in the \( 6N \)-dimensional phase space; an ensemble can thus be considered to be a collection of points in phase space. The way in which the system moves through phase space is governed by Hamiltonian's equations:

\[
\frac{d\mathbf{r}_i}{dt} = \frac{\partial H}{\partial \mathbf{p}_i}
\]  
(6.18)

\[
\frac{d\mathbf{p}_i}{dt} = -\frac{\partial H}{\partial \mathbf{r}_i}
\]  
(6.19)

where \( i \) varies from 1 to \( N \). Molecular dynamics generates a sequence of points in phase space that are connected in time. These points correspond to the successive configurations of the system generated by the simulation. A molecular dynamics simulation performed in the microcanonical (constant \( NVE \)) ensemble will sample phase space along a contour of constant energy. There is no momentum component in a Monte Carlo simulation and such simulations sample from the \( 3N \)-dimensional space corresponding to the positions of
the atoms. It might seem odd that thermodynamic properties can be obtained from Monte Carlo simulations, given that there is no momentum contribution and so 3N degrees of freedom are not explored. In fact, all of the deviations from ideal gas behaviour are a consequence of interactions between the atoms and are encapsulated in the potential function, \( \Phi(r^N) \), which only depends upon the positions of the atoms. A Monte Carlo simulation does sample from the positional degrees of freedom and so can be used to provide the deviations of thermodynamic properties from ideal gas behaviour, which is what we want to calculate. We shall return to this point in Chapter 8.

If it were possible to visit all the points in phase space then the partition function could be calculated by summing the values of \( \exp(-E/k_BT) \). The phase-space trajectory in such a case would be termed ergodic and the results would be independent of the initial configuration. For the systems that are typical of those studied using simulation methods the phase space is immense, and an ergodic trajectory is not achievable (indeed, even for relatively small systems with only a few tens of atoms the time that would be required to cycle round all of the points in phase space is longer than the age of the universe). A simulation can thus only ever provide an estimate of the ‘true’ energies and other thermodynamic properties and so a sequence of simulations using different starting conditions would be expected to give similar, but different, results.

The thermodynamic properties that we have considered so far, such as the internal energy, the pressure and the heat capacity are collectively known as the mechanical properties and can be routinely obtained from a Monte Carlo or molecular dynamics simulation. Other thermodynamic properties are difficult to determine accurately without resorting to special techniques. These are the so-called entropic or thermal properties: the free energy, the chemical potential and the entropy itself. The difference between the mechanical and thermal properties is that the mechanical properties are related to the derivative of the partition function whereas the thermal properties are directly related to the partition function itself. To illustrate the difference between these two classes of properties, let us consider the internal energy, \( U \), and the Helmholtz free energy, \( A \). These are related to the partition function by:

\[
U = k_BT^2 \frac{\partial Q}{\partial T} \quad (6.20)
\]

\[
A = -k_B T \ln Q \quad (6.21)
\]

\( Q \) is given by Equation (6.4) for a system of identical particles. We shall ignore any normalisation constants in our treatment here to enable us to concentrate on the basics, and so it does not matter whether the system consists of identical or distinguishable particles. We also replace the Hamiltonian by the energy, \( E \). The internal energy is obtained via Equation (6.20):

\[
U = k_BT^2 \frac{1}{Q} \int \int dp^N \, dr^N \frac{E(p^N, r^N)}{k_BT^2} \exp\left(-\frac{E(p^N, r^N)}{k_BT}\right)
\]

\[
= \int \int dp^N \, dr^N \frac{E(p^N, r^N)}{Q} \exp\left(-\frac{E(p^N, r^N)}{k_BT}\right) \quad (6.22)
\]
Now consider the probability of the state with energy $E(p^N, r^N)$:

$$\frac{\exp(-E(p^N, r^N)/k_B T)}{Q}$$

(6.23)

This probability is written $\rho(p^N, r^N)$; the internal energy is thus given by

$$U = \iint d^Np d^N r E(p^N, r^N) \rho(p^N, r^N)$$

(6.24)

The crucial point about Equation (6.24) is that high values of $E(p^N, r^N)$ have a very low probability and make an insignificant contribution to the integral. The Monte Carlo and molecular dynamics methods preferentially generate states of low energy, which are the states that make a significant contribution to the integral in Equation (6.24). These methods sample from phase space in a way that is representative of the equilibrium state and are able to generate accurate estimates of properties such as the internal energy, heat capacity, and so on.

Let us now consider the problem of calculating the Helmholtz free energy of a molecular liquid. Our aim is to express the free energy in the same functional form as the internal energy, that is as an integral which incorporates the probability of a given state. First, we substitute for the partition function in Equation (6.21):

$$A = -k_B T \ln Q = k_B T \ln \left( \frac{1}{\iint d^Np d^N r \exp(-E(p^N, r^N)/k_B T)} \right)$$

(6.25)

Next we recognise that the following integral is equal to 1:

$$1 = \frac{1}{(8\pi^2 V)^N} \iint d^Np d^N r \exp\left(-\frac{E(p^N, r^N)}{k_B T}\right) \exp\left(\frac{E(p^N, r^N)}{k_B T}\right)$$

(6.26)

Inserting this into the expression for the free energy and ignoring the constants (which act to change the zero point from which the free energy is calculated) gives:

$$A = k_B T \ln \left( \frac{1}{\iint d^Np d^N r \exp(-E(p^N, r^N)/k_B T)} \right)$$

(6.27)

We can now substitute for the probability density, $\rho(p^N, r^N)$ in this equation, leading to the final result (in which we have again ignored the normalisation factors):

$$A = k_B T \ln \left( \iint d^Np d^N r \exp\left(\frac{E(p^N, r^N)}{k_B T}\right) \rho(p^N, r^N) \right)$$

(6.28)

The important feature of this result is that the configurations with a high energy make a significant contribution to the integral due to the presence of the exponential term $\exp(\pm E(p^N, r^N)/k_B T)$. A Monte Carlo or molecular dynamics simulation preferentially samples the lower-energy regions of phase space. An ergodic trajectory would, of course, visit all of these high-energy regions, but in practice these will never be adequately sampled
by a real simulation. The results for the free energy and other entropic properties will as a consequence be poorly converged and inaccurate.

To reiterate a point that we made earlier, these problems of accurately calculating the free energy and entropy do not arise for isolated molecules that have a small number of well-characterised minima which can all be enumerated. The partition function for such systems can be obtained by standard statistical mechanical methods involving a summation over the minimum energy states, taking care to include contributions from internal vibrational motion.

6.4 Practical Aspects of Computer Simulation

6.4.1 Setting Up and Running a Simulation

There are significant differences between the molecular dynamics and Monte Carlo simulation methods, but the same general strategies are used to set up and run either type of simulation. The first task is to decide which energy model is to be used to describe the interactions within the system. Simulations are usually performed with relatively large numbers of atoms over many iterations or time steps. The intra- and intermolecular interactions are therefore almost always described using an empirical (i.e. molecular mechanics) energy model. Faster computers and new theoretical techniques do now enable simulations to be performed using models based only on quantum mechanics or mixed models based on molecular mechanics/quantum mechanics as discussed in Section 11.13. Having chosen an energy model, the simulation itself can be broken into four distinct stages. First, an initial configuration for the system must be established. An equilibration phase is then performed, during which the system evolves from the initial configuration. Thermodynamic and structural properties are monitored during the equilibration until stability is achieved. Several distinct steps may be required during the equilibration, particularly for inhomogeneous systems. At the end of the equilibration the production phase commences. It is during the production phase that simple properties of the system are calculated. At regular intervals the configuration of the system (i.e. the atomic coordinates) is output to a disk file. Finally, the simulation is analysed; properties not calculated during the simulation are determined and the configurations are examined, not only to discover how the structure of the system changed but also to check for any unusual behaviour that might indicate a problem with the simulation.

6.4.2 Choosing the Initial Configuration

Before a simulation can be performed it is obviously necessary to select an initial configuration of the system. This should be done with some care, as the initial arrangement can often determine the success or failure of a simulation. For simulations of systems at equilibrium (the most common sort) it is wise to choose an initial configuration that is close to the state which it is desired to simulate. For example, it would be unwise to initiate a simulation of a face-centred cubic solid from a body-centred cubic starting point. It is also good practice to ensure that the initial configuration does not contain any high-energy