N. Bodor *et al.* [31] decided to extend and enhance Klopman and Iroff's work by including the molecular volume, weight and surface area. In addition, they included the molecular electric dipole moment in the correlation, all possible products of pairs of charges and a further 57 compounds in the analysis. They used a more up-to-date quantum mechanical model, AM1 (again to be discussed in Chapter 13). The molecular volume and surface area were calculated by a grid method. In their equation

$$\begin{aligned}\log_{10}(P) = {}& -1.167 \times 10^{-4}S^2 - 6.106 \times 10^{-2}S + 14.87O^2 - 43.67O + 0.9986I_{\text{alkane}} \\ & + 9.57 \times 10^{-3}M_{\text{w}} - 0.1300D - 4.929Q_{\text{ON}} - 12.17Q_{\text{N}}^4 + 26.81Q_{\text{N}}^2 \\ & - 7.416Q_{\text{N}} - 4.551Q_{\text{O}}^4 + 17.92Q_{\text{O}}^2 - 4.03Q_{\text{O}} + 27.273\end{aligned} \tag{7.9}$$

where S (cm^2) is the molecular surface area, O the molecular ovality, I_{alkane} is the indicator variable for alkanes (it is 1 if the molecule is an alkane, 0 otherwise), M_{w} the relative molar mass, D the calculated electric dipole moment, and Q_{ON} the sum of absolute values of the atomic charges on nitrogen and oxygen. The remaining Qs differ from those used by Klopman and Iroff in that they are the square root of the sum of charges on the nitrogen or oxygen atoms. The most significant parameters are the volume and the surface area; the authors claim that this demonstrates that the most important contribution to log(P) is the creation of a hole in the structure of water.

8 Quick Guide to Statistical Thermodynamics

Molecular structure theory tends to deal with the details of individual atoms and molecules, and the way in which a small number of them interact and react. Chemical thermodynamics on the other hand deals with the bulk properties of matter, typically 10^{23} particles. There clearly ought to be a link between the two sets of theories, even though chemical thermodynamics came to maturity long before there was even a satisfactory atomic theory and does not at first sight draw on the concept of a molecule.

Suppose then that we have a macroscopic pure liquid sample, which might consist of 10^{23} particles, and we want to try to model some simple thermodynamic properties such as the pressure, the internal energy or the Gibbs energy. At room temperature, the individual particles making up the sample will be in motion, so at first sight we ought to try to solve the equations of motion for these particles. In view of the large number of particles present, such an approach would be foolhardy. Just to try to specify the initial positions and momenta of so many particles would not be possible, and in any case such a calculation would give too much information.

Even if we could do this impossible task, the next step would be to find a way in which we could relate the individual molecular information to the bulk properties.

For the sake of argument, suppose that the container is a cube. I have shown a two-dimensional slice through the cube as the left-hand side of Figure 8.1, and I have exaggerated the size of the particles by a factor of approximately 10^{10}.

The pressure exerted by a gas on a container wall depends on the rate at which particles collide with the wall. It is not necessary, or even helpful, to know which particle underwent a particular collision. What we need to know are the root mean square speed of the particles, their standard deviation about the mean, the temperature and so on. In chemical thermodynamics, we don't enquire about the behaviour of the individual particles that make up a macroscopic sample; we just enquire about their average properties.

Ludwig Boltzmann and Josiah Willard Gibbs understood all these problems, and invented the subject of *statistical thermodynamics* to get around them.

If we were to measure the pressure exerted on the walls at time intervals $t_1, t_2, \ldots, t_n$ then we might record results $p(t_1), p(t_2), \ldots, p(t_n)$. We could calculate a

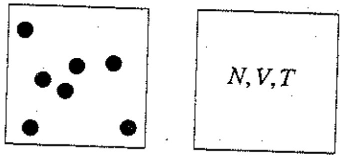

Figure 8.1 Box of particles

sample mean $\langle p \rangle$ and a sample standard deviation using these results

$$\begin{aligned} \langle p \rangle &= \frac{1}{n}\sum_{i=1}^{n} p(t_i) \\ \sigma_p &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(p(t_i) - \langle p \rangle\right)^2} \end{aligned} \tag{8.1}$$

We might expect that the greater the number of measurements, the closer the sample mean would be to the true mean, and the smaller the sample deviation would become.

8.1 The Ensemble

When we considered Figure 8.1, I was careful to draw your attention to the difference between particle properties and bulk properties. I also mentioned that classical thermodynamics is essentially particle-free; all that really matters to such a thermodynamicist are bulk properties such as the number of particles N, the temperature T and the volume of the container V. I have represented this information in the right hand box in Figure 8.1.

Rather than worry about the time development of the particles in the left-hand box in Figure 8.1, what we do is to make a very large number of copies of the system on the right-hand side. We then calculate average values over this large number of replications and according to the *ergodic theorem*, the average value we calculate is exactly the same as the time average we would calculate by studying the time evolution of the original system. The two are the same.

I am not suggesting that all the *cells* in the ensemble are exact replicas at the molecular level; all we do is to ensure that each cell has a certain number of thermodynamic properties that are the same. There is no mention of molecular properties at this stage of the game.

So Figure 8.2 is an ensemble of cells all with the same values of N, V and T. This array of cells is said to form a *canonical ensemble*. There are three other important ensembles in the theory of statistical thermodynamics, and they are named according to what is kept constant in each cell. Apart from the canonical ensemble, where N, V and T are kept constant, statistical thermodynamicists concern themselves also with three others.

N, V, T	N, V, T	N, V, T	N, V, T
N, V, T	N, V, T	N, V, T	N, V, T
N, V, T	N, V, T	N, V, T	N, V, T

Figure 8.2 Canonical ensemble

In a *microcanonical ensemble*, N, the total energy E and V are kept constant in each cell. In fact, this is a very simple ensemble because energy cannot flow from one cell to another. In an *isothermal–isobaric ensemble*, N, T and the pressure p are kept constant. Finally, we have the *grand canonical ensemble*, where V, T and the chemical potential are kept constant. The grand canonical ensemble is a fascinating one because the number of particles is allowed to fluctuate.

Suppose then that we consider a canonical ensemble of N^* cells, comprising the original cell together with $N^* - 1$ replications. Energy may flow between the cells, but the total energy of the ensemble is constant. Suppose that the possible total energies of the N particles contained in each cell are E_1^*, E_2^*, and so on. We take an energy snapshot, and find a distribution of energies amongst the cells as follows:

N_1^* cells have energy E_1^*,

N_2^* cells have energy E_2^*, etc.

According to Boltzmann, the E^* and the N^* are related by

$$\frac{N_i^*}{N^*} = \frac{\exp\left(-\frac{E_i^*}{k_B T}\right)}{\sum_i \exp\left(-\frac{E_i^*}{k_B T}\right)} \tag{8.2}$$

Be sure to understand that the energies E_i^* are not molecular energies; they are the total energies of the collection of the N particles contained in each cell. Also note that N^* is the number of cells in the ensemble, and that the energies are taken as relative to a common arbitrary zero.

The denominator in the expression above plays an important role in our theory, and so it is given a special name and symbol

$$Q = \sum_i \exp\left(-\frac{E_i^*}{k_B T}\right) \tag{8.3}$$

Q is (in this case) the *canonical partition function*, and it can be used to calculate the usual chemical thermodynamic functions as follows.

8.2 The Internal Energy U_{th}

The IUPAC recommended symbol for thermodynamic internal energy is U, but I have added a subscript 'th' for 'thermodynamic' so that there is no confusion with the total potential energy of a system U (often written Φ). Internal energy is obviously related to the ensemble energy average, but we have to exercise caution. Chemical measurements only give changes in the internal energy, not absolute values. I will therefore write the internal energy as $U_{th} - U_0$, where U_0 is an arbitrary constant. For most purposes we can take U_0 to be zero.

We have, for the ensemble of N^* members

$$U_{th} - U_0 = \frac{\sum_i N_i^* E_i^*}{N^*} \tag{8.4}$$

and according to the ergodic theorem, this is equal to the time average of $U_{th} - U_0$ for any one cell. Using the Boltzmann expression we have

$$U_{th} - U_0 = \frac{\sum_i E_i^* \exp\left(-\frac{E_i^*}{k_B T}\right)}{Q} \tag{8.5}$$

I can tidy up Equation (8.5) by noting

$$\left(\frac{\partial Q}{\partial T}\right)_{V,N} = \frac{1}{k_B T^2} \sum_i E_i^* \exp\left(-\frac{E_i^*}{k_B T}\right) \tag{8.6}$$

and so on substitution

$$U_{th} - U_0 = \frac{k_B T^2}{Q} \left(\frac{\partial Q}{\partial T}\right)_V \tag{8.7}$$

8.3 The Helmholtz Energy A

From the definition of A we have the following

$$\begin{aligned} A &= U - TS \\ S &= -\left(\frac{\partial A}{\partial T}\right)_{V,N} \\ A &= U + T\left(\frac{\partial A}{\partial T}\right)_{V,N} \end{aligned} \tag{8.8}$$

A little manipulation gives

$$\begin{aligned} \left(\frac{\partial}{\partial T}\left(\frac{A}{T}\right)\right)_{V,N} &= -k_B\left(\frac{\partial \ln Q}{\partial T}\right)_{V,N} \\ A - A_0 &= -k_B T \ln Q \end{aligned} \tag{8.9}$$

Again, the arbitrary constant A_0 can be taken as zero, since only changes in A are ever measured.

8.4 The Entropy S

Finally, since

$$S = \frac{U - A}{T}$$

we have

$$S = k_B T\left(\frac{\partial \ln Q}{\partial T}\right)_{V,N} + k_B \ln Q \tag{8.10}$$

8.5 Equation of State and Pressure

The pressure is related to the Helmholtz energy by

$$p = -\left(\frac{\partial A}{\partial V}\right)_{T,N}$$

and so we find

$$p = k_B T \left(\frac{\partial \ln Q}{\partial V} \right)_{T,N} \tag{8.11}$$

This equation is sometimes called the *equation of state*. The enthalpy and the Gibbs energy can be derived using similar arguments. They turn out to be

$$\begin{aligned} H &= k_B T^2 \left(\frac{\partial \ln Q}{\partial T} \right)_{V,N} + k_B T V \left(\frac{\partial \ln Q}{\partial V} \right)_{T,N} \\ G &= -k_B T \ln Q + k_B T V \left(\frac{\partial \ln Q}{\partial V} \right)_{T,N} \end{aligned} \tag{8.12}$$

8.6 Phase Space

Sophisticated methods such as those due to Hamilton and to Lagrange exist for the systematic treatment of problems in particle dynamics. Such techniques make use of *generalized coordinates* (written $q_1, q_2, \ldots, q_n$) and the *generalized momenta* (written $p_1, p_2, \ldots, p_n$); in Hamilton's method we write the total energy as the *Hamiltonian H*. H is the sum of the kinetic energy and the potential energy, and it is a constant provided that the potentials are time independent. H has to be written in terms of the ps and the qs in a certain way, and systematic application of Hamilton's equations gives a set of differential equations for the system.

To fix our ideas, consider the particle of mass m undergoing simple harmonic motion as discussed in Chapter 4. In this one-dimensional problem I wrote the potential as

$$U = \tfrac{1}{2} k_s (R - R_e)^2$$

so that the total energy is

$$\varepsilon = \tfrac{1}{2} m \left(\frac{dR}{dt} \right)^2 + \tfrac{1}{2} k_s (R - R_e)^2$$

If I put $q = R - R_e$, then the momentum p is $m\, dR/dt$ and I can write the Hamiltonian

$$H = \frac{p^2}{2m} + \frac{k_s q^2}{2} \tag{8.13}$$

We say that the particle moves through *phase space* and in this example the trajectory through phase space is an ellipse (see Figure 8.3), which can be easily seen by rewriting Equation (8.13) as

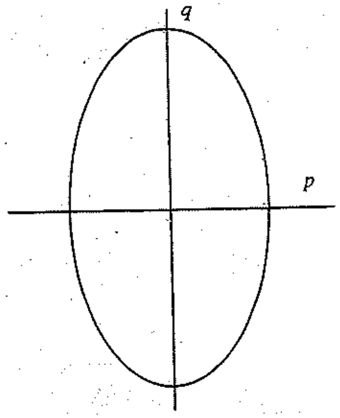

Figure 8.3 Phase space

$$\frac{p^2}{a^2} + \frac{q^2}{b^2} = 1$$

Hamilton's equations of motion are

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}; \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \tag{8.14}$$

so for a general problem with N atoms, we have to solve $6N$ first-order differential equations rather than the $3N$ second-order differential equations we would get from straightforward application of Newton's Law.

In the case of a one-particle, three-dimensional system, the Hamiltonian will be a function of the three coordinates $\mathbf{q}$ and the three momenta $\mathbf{p}$, and for a more general problem involving N particles the Hamiltonian will be a function of the $3N$ $\mathbf{q}$s and the $3N$ $\mathbf{p}$s. We say that the $\mathbf{p}$s and the $\mathbf{q}$s together determine a point in $6N$-dimensional phase space, and this point is often denoted Γ.

8.7 The Configurational Integral

Returning now to the canonical partition function, Equation (8.3)

$$Q = \sum_i \exp\left(-\frac{E_i^*}{k_B T} \right)$$

the first thing to note is that all points in phase space contribute to the sum, and the summation has to be replaced by an integral. For an ideal monatomic gas the expression becomes

$$Q = \frac{1}{N!}\frac{1}{h^{3N}}\iint \exp\left(-\frac{E}{k_\mathrm{B}T}\right)\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{q} \qquad (8.15)$$

The equation is often written with the Hamiltonian H replacing E, for the reasons discussed above.

The $N!$ term is needed in situations where the particles are completely indistinguishable from one another; for particles that can be distinguished there is no $N!$ term. The integrals have to be done over the spatial variables of all the N particles, and also the momentum variables of the N particles. The integral is therefore a $6N$-dimensional one.

The energy (the Hamiltonian) is always expressible as a sum of kinetic and potential energies, and I have written the mass of each particle m

$$E = \sum_{i=1}^{N}\frac{p_i^2}{2m} + \Phi(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N) \qquad (8.16)$$

Kinetic energies depend on the momentum coordinates $\mathbf{p}$. All the potential energies we will meet depend on the spatial coordinates $\mathbf{q}$ but not on the momenta and so the partition function can be factorized into a product of a kinetic part and a potential part

$$Q = \frac{1}{N!}\frac{1}{h^{3N}}\int \exp\left(-\frac{1}{k_\mathrm{B}T}\sum_{i=1}^{N}\frac{p_i^2}{2m}\right)\mathrm{d}\mathbf{p}\int \exp\left(-\frac{\Phi}{k_\mathrm{B}T}\right)\mathrm{d}\mathbf{q} \qquad (8.17)$$

The kinetic integral has to be done over the momentum coordinates of all N particles, and it can be seen to be a product of N identical three-dimensional integrals of the type

$$\int \exp\left(-\frac{1}{k_\mathrm{B}T}\frac{p_1^2}{2m}\right)\mathrm{d}\mathbf{p}_1$$

Each of these is a product of three identical standard integrals of the type

$$\int \exp\left(-\frac{1}{k_\mathrm{B}T}\frac{p_x^2}{2m}\right)\mathrm{d}p_x$$

and the final result is

$$Q = \frac{1}{N!}\left(\frac{2\pi m k_\mathrm{B}T}{h^2}\right)^{3N/2}\int \exp\left(-\frac{\Phi}{k_\mathrm{B}T}\right)\mathrm{d}\mathbf{q} \qquad (8.18)$$

The $3N$-dimensional integral over the position variables is often referred to as the *configurational integral*. For an ideal gas $\Phi = 0$ and so the configurational integral is V^N, where V is the volume of the container. Some authors include the $N!$ in the definition of the configurational integral.

The canonical partition function for an ideal gas is therefore

$$Q = \frac{V^N}{N!}\left(\frac{2\pi m k_\mathrm{B}T}{h^2}\right)^{3N/2} \qquad (8.19)$$

The partition function for a real system is often written as the product of an ideal part and an excess part due to non-ideal behaviour

$$Q = Q^{\mathrm{ideal}}Q^{\mathrm{excess}}$$

where

$$Q^{\mathrm{excess}} = \frac{1}{V^N}\int \exp\left(-\frac{\Phi}{k_\mathrm{B}T}\right)\mathrm{d}\mathbf{q} \qquad (8.20)$$

The point of doing this is that thermodynamic properties such as A are often measured experimentally as an ideal and an excess part

$$A = A^{\mathrm{ideal}} + A^{\mathrm{excess}}$$

The ideal part can be related to Q^{ideal} and the excess part to Q^{excess}.

8.8 The Virial of Clausius

Let me focus attention on one particular particle i moving in the box, Figure 8.1. As this particle moves it will be subject to some varying force $\mathbf{F}_i$ and

$$\mathbf{F}_i = m\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} \qquad (8.21)$$

Taking the scalar product of both sides of this equation with $\mathbf{r}_i$ I get

$$\mathbf{r}_i \cdot \mathbf{F}_i = m\mathbf{r}_i \cdot \left(\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t}\right) \qquad (8.22)$$

Consider now the vector identity

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r}_i \cdot \mathbf{v}_i) = \mathbf{r}_i \cdot \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} + \frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} \cdot \mathbf{v}_i \qquad (8.23)$$

which can also be written

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r}_i \cdot \mathbf{v}_i) = \mathbf{r}_i \cdot \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} + v_i^2 \tag{8.24}$$

On comparison of Equations (8.22) and (8.24), I have

$$\mathbf{r}_i \cdot \mathbf{F}_i = m\left(\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r}_i \cdot \mathbf{v}_i) - v_i^2\right) \tag{8.25}$$

or

$$-\tfrac{1}{2}\mathbf{r}_i \cdot \mathbf{F}_i = -\tfrac{1}{2}m\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r}_i \cdot \mathbf{v}_i + \tfrac{1}{2}mv_i^2 \tag{8.26}$$

The next step is to sum corresponding terms on both sides of the equation for each particle in the box. For N particles each of mass m, this gives

$$-\frac{1}{2}\sum_{i=1}^{N}\mathbf{r}_i \cdot \mathbf{F}_i = -\tfrac{1}{2}m\frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^{N}\mathbf{r}_i \cdot \mathbf{v}_i + \tfrac{1}{2}m\sum_{i=1}^{N}v_i^2 \tag{8.27}$$

Finally, we take a time average over all the particles in the box, which is assumed to be in an equilibrium state

$$-\frac{1}{2}\left\langle\sum_{i=1}^{N}\mathbf{r}_i \cdot \mathbf{F}_i\right\rangle = -\frac{m}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\langle\sum_{i=1}^{N}\mathbf{r}_i \cdot \mathbf{v}_i\right\rangle + \tfrac{1}{2}m\left\langle\sum_{i=1}^{N}v_i^2\right\rangle \tag{8.28}$$

The second term on the right-hand side is obviously the mean kinetic energy of all the particles in the box. This must be $\frac{3}{2}Nk_{\mathrm{B}}T$, according to the equipartition of energy principle.

Whatever the value of the first average quantity in brackets on the right-hand side it cannot vary with time because we are dealing with an equilibrium state and so the first time derivative must vanish

$$-\frac{m}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\langle\sum_{i=1}^{N}\mathbf{r}_i \cdot \mathbf{v}_i\right\rangle = 0$$

Thus, we have

$$-\frac{1}{2}\left\langle\sum_{i=1}^{N}\mathbf{r}_i \cdot \mathbf{F}_i\right\rangle = \tfrac{1}{2}m\left\langle\sum_{i=1}^{N}v_i^2\right\rangle \tag{8.29}$$

The summation term on the left hand side $-\frac{1}{2}\langle\sum_{i=1}^{N}\mathbf{r}_i \cdot \mathbf{F}_i\rangle$ involving the forces and coordinates is often referred to as the *virial of Clausius*.

9 Molecular Dynamics

Molecular mechanics (MM) these days tends to be concerned only with prediction of local minima on molecular potential energy surfaces. QSAR properties are often calculated in order to assist high-volume screening studies in pharmaceuticals applications. Should we want to study the motions of the molecule, all that would be needed would be to investigate the normal modes of vibration (which can be obtained from the hessian). MM does not take account of zero-point vibrations and the calculations refer to a molecule at 0 K, when it is completely at rest. Workers in the modelling field often refer to MM as *energy minimization*.

We now turn our attention to the time development of collections of atoms and molecules, for which the techniques of Molecular Dynamics and Monte Carlo are widely used.

I have stressed in previous chapters the intermolecular potential energy U_{mol} (often written Φ). Assuming pairwise additivity, Φ can be found by summing over all distinct pairs of particles

$$\Phi = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}U_{ij} \tag{9.1}$$

If the assumption of pairwise additivity is not valid, then we have to include all possible triples, and so on

$$\Phi = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}U_{ij} + \sum_{i=1}^{N-2}\sum_{j=i+1}^{N-1}\sum_{k=j+1}^{N}U_{ijk} + \cdots \tag{9.2}$$

In this book we will generally be concerned with situations where the potentials (and the forces) are pairwise additive. If we focus on particle A, then the mutual potential energy of A with all the other particles U_{A} is found from

$$U_{\mathrm{A}} = \sum_{j \neq \mathrm{A}} U_{\mathrm{A}j}$$

and we can find the force on particle A, $\mathbf{F}_\mathrm{A}$, by differentiating with respect to the coordinates of particle A

$$\mathbf{F}_\mathrm{A} = -\mathrm{grad}\, U_\mathrm{A}$$

For example, if we consider a pair of Lennard-Jones particles A and B where

$$U_\mathrm{AB} = 4\varepsilon\left(\left(\frac{\sigma}{R_\mathrm{AB}}\right)^{12} - \left(\frac{\sigma}{R_\mathrm{AB}}\right)^{6}\right)$$

we note that the potential only depends on the distance between the particles. The expression for grad U is particularly simple so that

$$\mathbf{F}_\mathrm{A} = -\frac{\partial U}{\partial R_\mathrm{A}}\frac{\mathbf{R}_\mathrm{BA}}{R_\mathrm{BA}}$$

which gives

$$\mathbf{F}_\mathrm{A} = 24\varepsilon\left(2\left(\frac{\sigma}{R_\mathrm{AB}}\right)^{12} - \left(\frac{\sigma}{R_\mathrm{AB}}\right)^{6}\right)\frac{\mathbf{R}_\mathrm{AB}}{R_\mathrm{AB}^2}$$

Newton's second law connects force and acceleration by

$$\mathbf{F}_\mathrm{A} = m_\mathrm{A}\frac{\mathrm{d}^2\mathbf{R}_\mathrm{A}}{\mathrm{d}t^2}$$

and in principle we could study the time development of a system by solving this second-order differential equation, one such equation for each of the particles in our system. Calculating the trajectories of N particles therefore appears to involve the solution of a set of $3N$ second-order differential equations. Alternatively, we could use an advanced method such as Hamilton's to solve $6N$ first-order differential equations. For any set of N particles it is always possible to find three coordinates that correspond to translation of the centre of mass of the system, and, if the particles have 'shape', three coordinates that correspond to rotations about three axes that pass through the centre of mass.

Most of the early molecular dynamics studies were directed at the problem of liquid structure, so that is where we will begin our discussion.

9.1 The Radial Distribution Function

Of the three states of matter, gases are the easiest to model because the constituent particles are so far apart on average that we can ignore intermolecular interactions, apart from during their brief collisions. This is why researchers were able to bring the kinetic theory of gases to such an advanced stage by the end of the nineteenth century (and before the existence of a satisfactory theory of molecular structure).

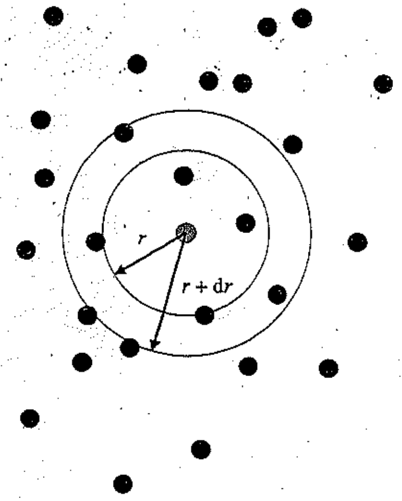

Figure 9.1 Radial distribution function

The atoms or molecules in a crystalline solid are arranged in a regular order, and for this reason we usually start a discussion of the solid state from the properties of regular solids. Once such patterns were truly understood at the beginning of the twentieth century, the theory of the solid state made rapid progress.

Liquids are much harder to model and to study experimentally than solids and gases; elementary textbooks usually state that liquids show neither complete order nor complete disorder. The basis of this remark concerns a property called the *radial distribution function* $g(r)$. Consider Figure 9.1, which is a snapshot of the particles in a simple atomic liquid.

We take a typical atom (the grey one, designated i) and draw two spheres of radii r and $r + \mathrm{d}r$. We then count the number of atoms whose centres lie between these two spheres, and repeat the process for a large number N of atoms. If the result for atom i is $g_i(r)\,\mathrm{d}r$, then the radial distribution function is defined as

$$g(r)\,\mathrm{d}r = \frac{1}{N}\sum_{i=1}^{N} g_i(r)\,\mathrm{d}r \tag{9.3}$$

This process then has to be repeated for many complete shells over the range of values of r thought to be significant.

In the case of an ideal gas, we would expect to find the number of particles to be proportional to the volume enclosed by the two spheres, which is $4\pi r^2\,\mathrm{d}r$. This gives $g(r) = 4\pi r^2$, a simple quadratic curve.

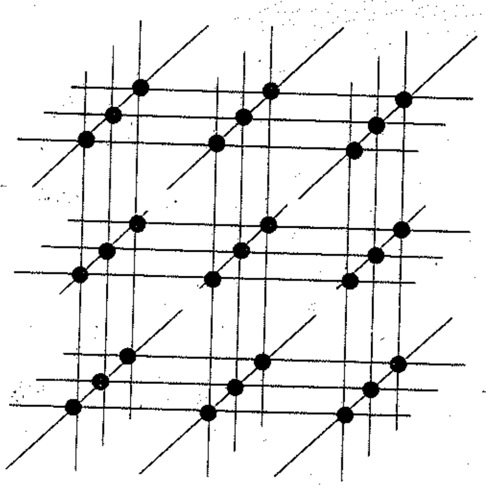

Figure 9.2 Simple cubic lattice

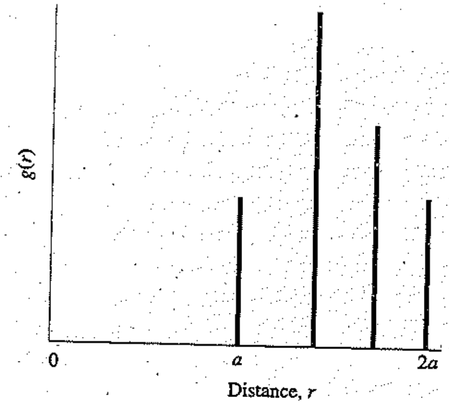

Figure 9.3 First part of the radial distribution function for a simple solid

Consider now the simple cubic solid shown in Figure 9.2 whose nearest neighbour distance is a. Each atom is surrounded by 6 nearest neighbours at a distance a, 12 at a distance $\sqrt{2}\ a$, 8 next-next nearest neighbours at a distance $\sqrt{3}\ a$, 6 at a further distance $2a$ and so on. We would therefore expect to find a radial distribution function similar to the one shown in Figure 9.3. The height of each peak is proportional to the number of atoms a distance r from any given atom.

Radial distribution functions can be deduced experimentally from diffraction studies. In the case of a liquid, Figure 9.4, the curve resembles that expected for a solid at low temperatures, and at high temperatures it resembles the quadratic expected for an ideal gas. At intermediate temperatures, the two features can be clearly seen; essentially a solid pattern is superimposed on the gas pattern. This gives the experimental basis for the well-known remark about liquid structure quoted above.

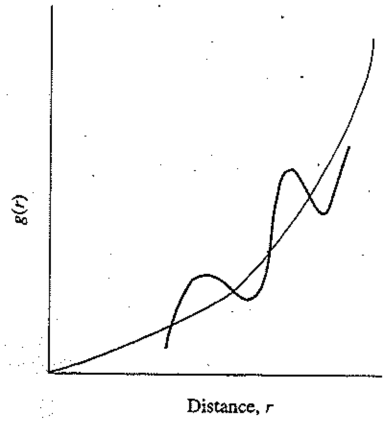

Figure 9.4 Radial distribution function for liquid superimposed on an ideal gas

9.2 Pair Correlation Functions

The radial distribution function for a gas varies as $4\pi r^2$ and so tends to infinity as r tends to infinity. It is usual to remove the $4\pi r^2$ dependence by defining a related quantity called the *pair correlation function* $g_{AB}(r)$, which gives information about the probability of finding two particles A and B separated by a distance r. If the volume of a system is V and it contains N_A species of type A and N_B species of type B, then the number densities are N_A/V and N_B/V. The fraction of time that the differential volume elements $d\tau_1$ and $d\tau_2$, which are separated by a distance r, simultaneously contain species of type A and B is given by

$$\frac{N_A}{V}\frac{N_B}{V}g_{AB}(r)\,d\tau_1 d\tau_2$$

In a mixture of A and B we would be interested in the three distinct pair correlation functions $g_{AA}(r)$, $g_{BB}(r)$ and $g_{AB}(r)$. These pair correlation functions have a limiting value of 1 for a fluid.

9.3 Molecular Dynamics Methodology

In an ideal gas, the particles do not interact with each other and so the potential Φ is zero. Deviations from ideality are due to the interparticle potential, and most of the early studies were made on just three types of particle: the hard sphere model, the finite square well and the Lennard-Jones model.

9.3.1 The hard sphere potential

The hard sphere potential of Figure 9.5 is the simplest one imaginable; the system consists of spheres of radii σ and $U(r)$ is zero everywhere except when two spheres touch, when it becomes infinite.

The hard sphere potential is of great theoretical interest not because it represents the intermolecular potential of any known substance, rather because any calculations based on the potential are simple. B. J. Alder and T. E. Wainwright introduced the modelling technique now known as Molecular Dynamics to the world in a short *Journal of Chemical Physics* 'Letters to the Editor' article in 1957 [32]. They reported a study of hard disks, the two-dimensional equivalent of hard spheres.

9.3.2 The finite square well

B. J. Alder and T. E. Wainwright's 1959 paper [33] is usually regarded the keynote paper in the field, and you might like to read the Abstract.

> *A method is outlined by which it is possible to calculate exactly the behaviour of several hundred interacting classical particles. The study of this many-body problem is carried out by an electronic computer that solves numerically the simultaneous equations of motion. The limitations of this numerical scheme are enumerated and the important steps in making the program efficient on computers are indicated. The applicability of this method to the solution of many problems in both equilibrium and nonequilibrium statistical thermodynamics is discussed.*

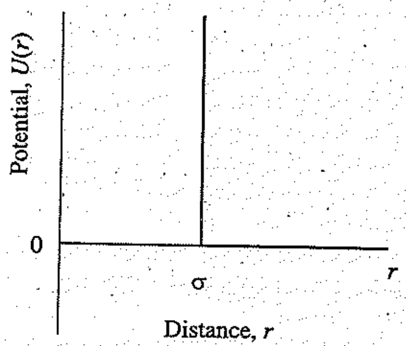

Figure 9.5 Hard sphere potential

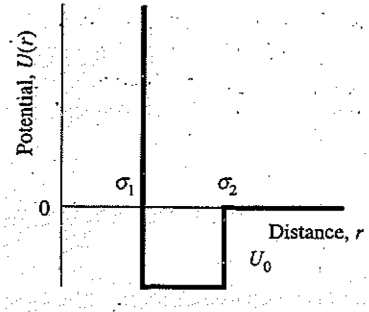

Figure 9.6 Finite square well

In this second paper they chose a three-dimensional system of particles and the finite square well potential shown in Figure 9.6. This potential is especially simple because a given particle does not experience any change in velocity except when it is separated from another particle by σ_1 (when it undergoes an attractive collision) or σ_2 (when it undergoes a repulsive collision). On collision, the velocities are adjusted and the calculation restarts. Statistical data are collected every collision.

In their dynamic calculation all the particles were given initial velocities and positions. In one example, the particles were given equal kinetic energies with the three direction cosines of the velocity vector chosen at random, and initial positions corresponding to a face-centred cubic lattice. Once the initial configuration was set up, they calculated exactly the time at which the first collision occurs. The collision time can be found by evaluating, for every pair in the system, the time taken for the projected paths to reach a separation of σ_1 or σ_2.

If two particles A and B have initial positions $\mathbf{r}_{A,0}$ and $\mathbf{r}_{B,0}$ and velocities $\mathbf{v}_A$ and $\mathbf{v}_B$, then the instantaneous positions at time t will be

$$\mathbf{r}_A = \mathbf{r}_{A,0} + \mathbf{u}_A t$$
$$\mathbf{r}_B = \mathbf{r}_{B,0} + \mathbf{u}_B t$$

giving

$$\mathbf{r}_A - \mathbf{r}_B = \mathbf{r}_{A,0} - \mathbf{r}_{B,0} + (\mathbf{u}_A - \mathbf{u}_B)t$$

and so

$$(\mathbf{r}_A - \mathbf{r}_B)^2 = (\mathbf{r}_{A,0} - \mathbf{r}_{B,0})^2 + 2t(\mathbf{r}_{A,0} - \mathbf{r}_{B,0}) \cdot (\mathbf{u}_A - \mathbf{u}_B) + t^2(\mathbf{u}_A - \mathbf{u}_B)^2$$

If we rewrite the last equation as a quadratic in t as

$$u_{AB}^2 t^2 + 2b_{AB} t + r_{AB,0}^2 = \sigma_\alpha^2$$

where α takes values 1 or 2, then we see that the time required for a repulsive or attractive collision is

$$t_{AB}^{(\alpha)} = \frac{-b_{AB} \pm (b_{AB}^2 - u_{AB}^2(r_{AB}^2 - \sigma_\alpha^2))^{1/2}}{u_{AB}^2} \tag{9.4}$$

In order to find the first collision time, all pairs have to be analysed. All the particles are then allowed to move for such time, and the velocities of the colliding pair are adjusted according to the equations of motion.

The finite square well occupies an important place in the history of molecular modelling. Real atomic and molecular systems have much more complicated mutual potential energy functions, but the finite square well does at least show a minimum. On the other hand, because of the finite square well potential, the equations of motion are particularly simple and no complicated numerical techniques are needed. There are no accelerations until two particles collide.

9.3.3 Lennardjonesium

The first simulation of a 'real' chemical system was A. Rahman's 1964 study of liquid argon [34]. He studied a system comprising 864 Lennard-Jones particles under conditions appropriate to liquid argon at 84.4 K and a density of 1.374 g cm^{-3}. Once again, there is much to be gained by studying the Abstract, so here is the first part of it.

> *A system of 864 particles interacting with a Lennard-Jones potential and obeying classical equations of motion has been studied on a digital computer (CDC 3600) to simulate molecular dynamics in liquid argon at 94.4 K and a density of 1.374 g cm^{-3}. The pair correlation function and the constant of self-diffusion are found to agree well with experiment; the latter is 15% lower than the experimental value. The spectrum of the velocity autocorrelation function shows a broad maximum in the frequency range $\omega = 0.25\ (2\pi k_B T/h)$. The shape of the Van Hove function $G_s(r, t)$ attains a maximum departure from a Gaussian at about $t = 0.3 \times 10^{-2}$ s and becomes a Gaussian again at about 10^{-11} s.*

There are several interrelated problems. A sample size has to be chosen; this is usually determined by the available computer resource and the complexity of the potential function, because the potential function has to be calculated very many times during the simulation. The number of particles and the density determine the size of the container. At the same time we need to decide on a potential function; the natural choice for the inert gases is the Lennard-Jones potential, and we should note that the L-J potential is essentially short range.

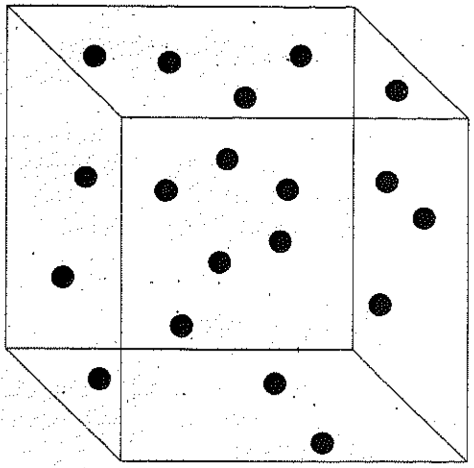

Figure 9.7 Box of argon atoms

So many early papers used the L-J potential that the noun *Lennardjonesium* was coined to describe a non-existent element whose atoms interacted via the L–J potential.

9.4 The Periodic Box

Figure 9.7 shows a suitable virtual box of argon atoms. Examination of the figure reveals two problems. Atoms near the edges of the box will experience quite different resultant forces from the atoms near the centre of the box. Secondly, the atoms will be in motion if the temperature is non-zero. As the system evolves in time, it is quite likely that one of the atoms will pass through the container walls and so disappear from the calculation. This has the undesirable effect of reducing the density.

There is a third subtle point: if the atoms are sufficiently light (He rather than Ar), we would need to take the quantum mechanical zero point energy into effect; even at 0 K, quantum mechanical particles have a residual motion.

The periodic box concept, illustrated in Figure 9.8, gives a solution to the first two problems. We appeal to the ensemble concept of statistical thermodynamics, and surround our system with a large number of *identical* copies. In this case the boxes are truly identical at the atomic level rather than in the usual thermodynamic sense of having N, V and T in common.

Figure 9.8 shows a two-dimensional slice through a small portion of the system (the central box where the atoms are shown grey) and the copies (where the atoms are

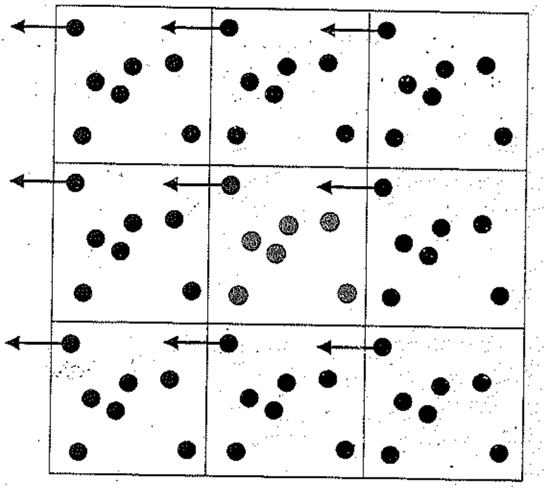

Figure 9.8 Small part of the molecular ensemble

shown black). Each copy is identical at the atomic level, and each atom undergoes the same time development as its image in every other copy. As the dark grey atom (top left, central cell) leaves the central cell, its image enters from an adjoining copy, shown by the vector displacements in the figure. This keeps the density constant. There are no effects due to the walls because each atom in the central cell is under the influence of every other atom in the central cell and in all other cells.

Consider now the dark grey atom (top left in the central cell). We need to calculate the force on this atom in order to understand its time development. To do this we should in principle sum the pair potential of the atom with every other atom. Differentiation of the potential with respect to the coordinates of the dark grey atom gives the force on the particle. This would give an infinite sum.

In this particular case, there is no great problem because the L-J potential is short range. We decide on a cut-off distance beyond which the pair potential will be negligible; this defines a sphere. In order to treat the dark grey atom, we have to include contributions from all other atoms in the sphere. This is illustrated as a two-dimensional slice in Figure 9.9.

Truncation of the intermolecular potential at a cut-off distance introduces two technical difficulties. First, the pair potential has a discontinuity at the cut-off distance r_c, and secondly, whenever a pair of particles A and B have separation greater than r_c the total energy is not conserved. The first problem is solved by shifting the potential function by an amount $U(r_c)$, that is we take

$$U^{s}(r_{AB}) = \begin{cases} U(r_{AB}) - U(r_c) & \text{if } r_{AB} \le r_c \\ 0 & \text{if } r_{AB} > r_c \end{cases} \qquad (9.5)$$

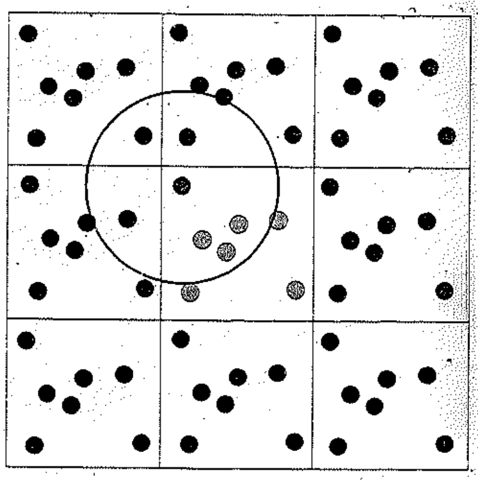

Figure 9.9 The cut-off distance

The second problem can be solved by adding a small linear term to the potential, chosen so that its derivative is zero at the cut-off distance

$$U^{\text{lin}}(r_{AB}) = \begin{cases} U(r_{AB}) - U(r_c) - \left(\dfrac{dU}{dr_{AB}}\right)_{R_c}(r_{AB} - r_c) & \text{if } r_{AB} \le r_c \\ 0 & \text{if } r_{AB} > r_c \end{cases} \qquad (9.6)$$

9.5 Algorithms for Time Dependence

Once we have calculated the potential and hence the force by differentiation, we have to solve Newton's equation of motion. If $\mathbf{F}_A$ is the force on particle A, whose position vector is $\mathbf{r}_A$ and whose mass is m_A, then

$$\begin{aligned} \mathbf{F}_A &= m_A \frac{d^2\mathbf{r}_A}{dt^2} \\ &= m_A \mathbf{a}_A \end{aligned}$$

This is a second-order differential equation that I can write equivalently as two first-order differential equations for the particle position $\mathbf{r}_A$ and the velocity $\mathbf{v}_A$

$$\begin{aligned} \mathbf{F}_A &= m_A \frac{d\mathbf{v}_A}{dt} \\ \mathbf{v}_A &= \frac{d\mathbf{r}_A}{dt} \end{aligned}$$

9.5.1 The leapfrog algorithm

A simple algorithm for integration of these two equations numerically in small time steps Δt can be found by considering the Taylor expansion for $\mathbf{v}(t)$

$$\begin{aligned}\mathbf{v}_\mathrm{A}\left(t+\frac{\Delta t}{2}\right) &= \mathbf{v}_\mathrm{A}(t) + \left(\frac{\mathrm{d}\mathbf{v}_\mathrm{A}}{\mathrm{d}t}\right)_t \frac{\Delta t}{2} + \frac{1}{2}\left(\frac{\mathrm{d}^2\mathbf{v}_\mathrm{A}}{\mathrm{d}t^2}\right)_t \left(\frac{\Delta t}{2}\right)^2 + \cdots \\ \mathbf{v}_\mathrm{A}\left(t-\frac{\Delta t}{2}\right) &= \mathbf{v}_\mathrm{A}(t) - \left(\frac{\mathrm{d}\mathbf{v}_\mathrm{A}}{\mathrm{d}t}\right)_t \frac{\Delta t}{2} + \frac{1}{2}\left(\frac{\mathrm{d}^2\mathbf{v}_\mathrm{A}}{\mathrm{d}t^2}\right)_t \left(\frac{\Delta t}{2}\right)^2 + \cdots \end{aligned} \tag{9.7}$$

Subtracting and rearranging we get

$$\mathbf{v}_\mathrm{A}\left(t+\frac{\Delta t}{2}\right) = \mathbf{v}_\mathrm{A}\left(t-\frac{\Delta t}{2}\right) + \mathbf{a}_\mathrm{A}(t)\Delta t + \cdots \tag{9.8}$$

I will switch back and forth between, for example, $\mathbf{v}$ and $\mathrm{d}\mathbf{r}/\mathrm{d}t$ in order to try to improve the readability of the equations. Also, I could have written $\mathbf{v}_\mathrm{A}(t)$ or $(\mathbf{v}_\mathrm{A})_t$ to mean the instantaneous velocity of particle A at time t. The acceleration $\mathbf{a}$ is calculated from the force.

Using the same procedure for the Taylor expansion of $\mathbf{r}_\mathrm{A}$ at the time point $t + 1/2\ \Delta t$ we get

$$\mathbf{r}_\mathrm{A}(t+\Delta t) = \mathbf{r}_\mathrm{A}(t) + \mathbf{v}_\mathrm{A}\left(t+\frac{\Delta t}{2}\right)\Delta t + \cdots \tag{9.9}$$

Equations (9.8) and (9.9) form the so-called *leapfrog algorithm*, which is reputed to be one of the most accurate and stable techniques for use in molecular dynamics. A suitable time increment Δt for molecular dynamics is a femtosecond (10^{-15} s). In the leapfrog scheme the velocities are first calculated at time $t + 1/2\ \Delta t$. These are used to calculate the positions of the particles at time $t + \Delta t$ and so on. In this way the velocities leap over the positions and then the positions leap over the velocities.

9.5.2 The Verlet algorithm

If instead we start from the Taylor expansion of $\mathbf{r}_\mathrm{A}(t)$ we have

$$\begin{aligned}\mathbf{r}_\mathrm{A}(t+\Delta t) &= \mathbf{r}_\mathrm{A}(t) + \left(\frac{\mathrm{d}\mathbf{r}_\mathrm{A}}{\mathrm{d}t}\right)_t \Delta t + \frac{1}{2}\left(\frac{\mathrm{d}^2\mathbf{r}_\mathrm{A}}{\mathrm{d}t^2}\right)_t (\Delta t)^2 + \cdots \\ \mathbf{r}_\mathrm{A}(t-\Delta t) &= \mathbf{r}_\mathrm{A}(t) - \left(\frac{\mathrm{d}\mathbf{r}_\mathrm{A}}{\mathrm{d}t}\right)_t \Delta t + \frac{1}{2}\left(\frac{\mathrm{d}^2\mathbf{r}_\mathrm{A}}{\mathrm{d}t^2}\right)_t (\Delta t)^2 + \cdots \end{aligned} \tag{9.10}$$

which gives (assuming that third-order and higher terms are negligible)

$$\mathbf{r}_\mathrm{A}(t+\Delta t) = 2\mathbf{r}_\mathrm{A}(t) - \mathbf{r}_\mathrm{A}(t-\Delta t) + \left(\frac{\mathrm{d}^2\mathbf{r}_\mathrm{A}}{\mathrm{d}t^2}\right)_t (\Delta t)^2 \tag{9.11}$$

This is known as the *Verlet algorithm*. The acceleration is obtained from the force experienced by atom A at time t. The velocity does not appear in the expression, but it can be obtained from the finite difference formula

$$\mathbf{v}_\mathrm{A}(t) = \frac{\mathbf{r}_\mathrm{A}(t+\Delta t) - \mathbf{r}_\mathrm{A}(t-\Delta t)}{2\Delta t} \tag{9.12}$$

The Verlet algorithm uses positions and accelerations at time t and the position at time $t - \Delta t$ to calculate a new position at time $t + \Delta t$. All these have to be stored at every iteration.

A variant is the *velocity Verlet algorithm*, which requires only the storage of positions, velocities and accelerations that all correspond to the same time step. It takes the form

$$\begin{aligned}\mathbf{r}_\mathrm{A}(t+\Delta t) &= \mathbf{r}_\mathrm{A}(t) + \left(\frac{\mathrm{d}\mathbf{r}_\mathrm{A}}{\mathrm{d}t}\right)_t \Delta t + \frac{1}{2}\left(\frac{\mathrm{d}^2\mathbf{r}_\mathrm{A}}{\mathrm{d}t^2}\right)_t (\Delta t)^2 \\ \mathbf{v}_\mathrm{A}(t+\Delta t) &= \left(\frac{\mathrm{d}\mathbf{r}_\mathrm{A}}{\mathrm{d}t}\right)_t + \frac{1}{2}\left(\left(\frac{\mathrm{d}^2\mathbf{r}_\mathrm{A}}{\mathrm{d}t^2}\right)_t + \left(\frac{\mathrm{d}^2\mathbf{r}_\mathrm{A}}{\mathrm{d}t^2}\right)_{t+\Delta t}\right)\Delta t \end{aligned} \tag{9.13}$$

There are many other algorithms in the literature, each with their own strengths and weaknesses.

9.6 Molten Salts

Molten salts (such as the alkali halides) are of great technological interest in the field of metal extraction. The first simulations were done by L. V. Woodcock in 1971 [35]. Molten salts introduce a new problem because the potential energy terms are long range. Consider a (hypothetical) one-dimensional infinite crystal, part of which is shown in Figure 9.10. The unshaded ions have charge $-Q$, the shaded ions have charge $+Q$ and the spacing between the ions is a. Suppose we start with the central (grey) ion at infinity, and all the other ions at their lattice positions as shown. The work done, W, in bringing the grey ion from infinity to its place in the lattice is

$$W = -\frac{2Q^2}{4\pi\epsilon_0 a}\left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots\right)$$

and the term in brackets converges very slowly to its limiting value of ln (2). Such series have to be summed when calculating the force on a given ion in a periodic box

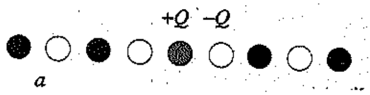

Figure 9.10 Part of a one-dimensional crystal, where the separation between ions is *a*

Table 9.1 Summation of series for *W*

No of terms	Sum
1	1.0000
2	0.5000
3	0.8333
4	0.5833
100	0.6882

such as Figure 9.8; in the case of neutral species, the sum is very quickly convergent because of the short-range nature of the forces (see Table 9.1). For neutral systems, a cut-off radius usually is taken beyond which the interactions are set to zero.

9.7 Liquid Water

Water plays a prominent place amongst solvents for obvious reasons. Two-thirds of our planet is covered in water, chemical reactions tend to be done in aqueous solution, and so on. The relatively strong hydrogen bonding in liquid water causes many of its physical properties to be 'anomalous', and the structure of ice has long interested both theoreticians and experimentalists.

Neutron diffraction studies on heavy ice D_2O have shown that water molecules retain their identity in condensed phases with very little distortion of their molecular geometry. This means that water molecules may be treated as rigid asymmetric rotors (with six degrees of freedom) rather than explicitly treating the three nuclei separately (nine degrees of freedom). The classical energy for a collection of *N* rigid rotor molecules consists of the kinetic energy for translation and rotation, together with the intermolecular potential. Each water molecule is described by six coordinates; three specify the centre of mass and three angles that fix the spatial orientation about the centre of mass. In this section I will denote these coordinates by the six-dimensional vector **X**. In terms of the linear velocities $\mathbf{v}_i$, the angular velocity vectors ω_i, the moments of inertia $\mathbf{I}_i$ and the coordinates $\mathbf{X}_i$ the energy turns out to be

$$\varepsilon = \frac{1}{2}\sum_{i=1}^{N}(mv_i^2 + \omega_i^{\mathrm{T}}\mathbf{I}_i\omega_i) + U(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)$$

The question then is the extent to which the intermolecular potential is pairwise additive; such functions may always be resolved into contributions from pairs, triples and higher contributions

$$U(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N) = \sum_{i<j}^{N} U^{(2)}(\mathbf{X}_i, \mathbf{X}_j) + \sum_{i<j<k}^{N} U^{(3)}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k) + \cdots + U^{(N)}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N) \tag{9.14}$$

In the case of simple fluids such as liquid argon it is widely believed that *U* is practically pairwise additive. In other words, the contributions $U^{(p)}$ for $p > 2$ are negligible. However, the local structure in liquid water is thought to depend on just these higher order contributions and so it is unrealistic in principle to terminate the expansion with the pair contributions. It is legitimate to write the potential as a sum of pair potentials, provided one understands that they are effective pair potentials that somehow take higher terms into account.

The classic molecular dynamics study of liquid water is that of A. Rahman and F. H. Stillinger [36]. They wrote the effective pair potential as a sum of two contributions: a Lennard-Jones 12-6 potential U_{LJ} for the two oxygen atoms

$$U_{\mathrm{LJ}}(R) = 4\varepsilon\left(\left(\frac{\sigma}{R}\right)^{12} - \left(\frac{\sigma}{R}\right)^{6}\right)$$

and a function U_{el}, modulated by a function $S(R_{ij})$ that depends sensitively on the molecular orientations about the oxygen atoms

$$U^{(2)}(\mathbf{X}_i, \mathbf{X}_j) = U_{\mathrm{LJ}}(R_{ij}) + S(R_{ij})U_{\mathrm{el}}(\mathbf{X}_i, \mathbf{X}_j) \tag{9.15}$$

The Lennard-Jones parameters were taken to be those appropriate to neon, $\sigma = 282\,\mathrm{pm}$ and $\varepsilon = 5.01 \times 10^{-22}\,\mathrm{J}$, on the grounds that neon is isoelectronic with water.

Four point charges *Q*, each of magnitude 0.19 e and each 100 pm from the oxygen nucleus, are embedded in the water molecule in order to give U_{el}. Two charges are positive, to simulate the hydrogen atoms, whilst the other two are negative and simulate the lone pairs. The four charges are arranged tetrahedrally about the oxygen. The set of 16 Coulomb interactions between two water molecules gives U_{el}. Figure 9.11 shows the

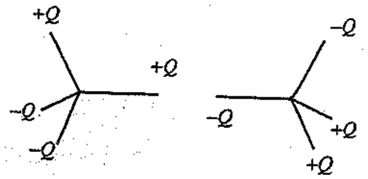

Figure 9.11 Two water molecules

minimum energy conformation, with the two 'molecules' 76 pm apart (i.e. an oxygen–oxygen distance of 276 pm).

The so-called switching function prevents the charges overlapping, and takes values between 0 and 1 depending on the distance between the oxygen atoms. The authors took $N = 216$ water molecules in a cube of side 1862 pm, which corresponds to a density of $1\,\mathrm{g\,cm^{-3}}$. They took periodic boundary conditions in order to eliminate undesirable wall effects, and a cut-off of distance of 916.5 pm beyond which the intermolecular interactions were set to zero.

The first step in a molecular dynamics calculation is to generate a suitable initial state. In general, random structures created using density values consistent with experiment may give atoms that are very close together. This causes undesirable large repulsive forces in small localized regions of space. One approach is to start from a low-density system with the atoms or molecules placed on simple cell lattices. The density is then increased gradually.

Starting velocities then have to be assigned. One technique is to calculate a value of the speed from equipartition of energy, and then assign random directions to each particle, with the constraint that the system has zero overall momentum. The next step is to let the system progress for a sufficient number of time steps in order to eliminate any undesirable effects due to the choice of initial conditions. After that, the collection of statistics begins. The temperature is inferred from the average values of the translational and rotational kinetic energies, both of which should equal $\frac{3}{2}Nk_{\mathrm{B}}T$.

It is possible in principle to follow a molecular dynamics simulation by displaying the atom coordinates at a series of time points. It proves more convenient to calculate certain statistical quantities. I mentioned the pair correlation function $g^{(2)}(r)$ earlier; in water the three distinct types of nuclear pairs lead to three corresponding functions $g^{(2)}_{\mathrm{OO}}(r)$, $g^{(2)}_{\mathrm{OH}}(r)$ and $g^{(2)}_{\mathrm{HH}}(r)$, which give information about the fraction of time that differential volume elements separated by a distance r simultaneously contain pairs of the nuclei given by the subscripts. These functions all have a limiting value of 1 as $r \to \infty$. The $g^{(2)}_{\mathrm{OO}}(r)$ curve is shown schematically in Figure 9.12. The horizontal axis is actually a reduced distance r/σ.

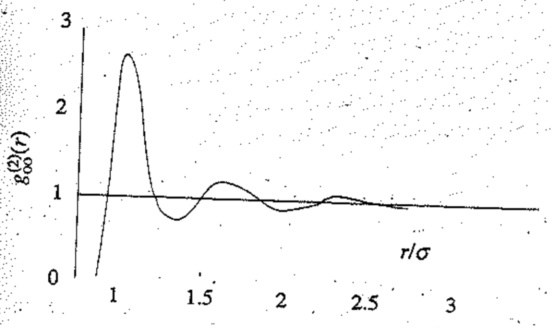

Figure 9.12 O–O pair distribution function

Two important features of the curve are as follows

1. The narrow first peak at 275 pm corresponds to 5.5 neighbouring oxygen atoms.

2. The second peak is low and broad, with a maximum at about 465 pm. The ratio of the second peak to the first, $465\,\mathrm{pm}/275\,\mathrm{pm} = 1.69$, is close to that observed for an ideal ice structure ($2\sqrt{2}/\sqrt{3} = 1.633$).

The *mean square displacement* is often calculated. This indicates the average displacement of an atom during a fixed time period.

9.7.1 Other water potentials

Over the years, a number of authors have tackled the problem of finding a suitable pair potential for liquid water. All involve a rigid water monomer with a number of interaction sites. The original TIPS 3-site model proposed by W. L. Jorgensen [37] has positive charges on the hydrogens and a negative charge ($q_{\mathrm{O}} = -2q_{\mathrm{H}}$) on the oxygen. The interaction potential is taken as the sum of all intermolecular Coulomb terms together with a single Lennard-Jones 12-6 term between the oxygens

$$U_{\mathrm{AB}} = \frac{1}{4\pi\epsilon_0} \sum_{i\text{ on A}} \sum_{j\text{ on B}} \frac{q_i q_j}{R_{ij}} + \frac{A}{R_{\mathrm{OO}}^{12}} - \frac{B}{R_{\mathrm{OO}}^{6}} \tag{9.16}$$

The parameters q_i, A and B were chosen to give reasonable energetic results for gas-phase complexes of water and alcohols. The author subsequently reoptimized the parameters to give TIP3P, an improved three-site model.

Four-site models have also been used; the oldest is that due to J. D. Bernal and R. H. Fowler [38], with a more modern version referred to as TIP4P by W. L. Jorgensen *et al.* [39]. In this case the negative charge is moved off the oxygen and towards the hydrogens at a point along the bisector of the HOH angle. The pair potential is still calculated according to the equation above, but more distances have to be evaluated.

Stillinger and Rahman's ST2 potential is widely used; this is an improved version of the one discussed above, but adds a Lennard-Jones 12-6 term between the oxygen atoms. In all, 17 distances have to be evaluated.

9.8 Different Types of Molecular Dynamics

When Newton's equations of motion are integrated, the energy is conserved. Since the energy and temperature are related by the equipartition of energy principle, the temperature should also be constant. Slow temperature drifts do occur as a result of the numerical integration and also because of the truncation of the forces. Several methods for performing molecular dynamics at constant temperature have been proposed, of which the simplest is known as *rescaling*. If the current temperature at time t is $T(t)$, then for a system of N particles and $3N$ degrees of freedom we have

$$\langle \varepsilon_{\text{kin}} \rangle = \left\langle \sum_{i=1}^{N} \tfrac{1}{2} m_i v_i^2 \right\rangle = \tfrac{3}{2} N k_B T(t)$$

To adjust the temperature to exactly a reference temperature T_{ref} we simply rescale the velocities by a factor

$$\left(\frac{T_{\text{ref}}}{T(t)} \right)^{1/2}$$

Scaling the velocities at appropriate intervals can therefore control the temperature of the system. We speak about *constant-temperature molecular dynamics*.

S. Nose [40] proposed an alternative method that adds an extra degree of freedom, referred to as a heat bath, to the atomic degrees of freedom. Extra kinetic energy and potential energy terms are added to the total energy, and the molecular dynamics is carried out with this one extra degree of freedom. Energy can flow back and forth between the heat bath and the system, and an equation of motion for the extra degree of freedom is solved.

Similar considerations apply to the pressure; in order to control the pressure in a system, any change in volume must be monitored and adjusted. Methods are available, and such calculations are referred to as *constant-pressure molecular dynamics*.

9.9 Uses in Conformational Studies

Molecular dynamics generally is used to study the structure and behaviour of materials, where one is concerned with the intermolecular potential. It is of course also perfectly possible to study single molecules and it is sometimes used as an aid in conformational analysis. As noted previously, there are very many local minima on the potential energy surface and it may be that small variations in torsional angles do not facilitate searching of all regions.

The idea is to start with a molecular mechanics local minimum, which corresponds to a temperature of 0 K. We use molecular dynamics to raise the temperature, let the molecule vibrate for a while at an elevated temperature and cool it back to 0 K. Hopefully the molecule will then have crossed over any potential barriers and a re-optimization will give a new structure.

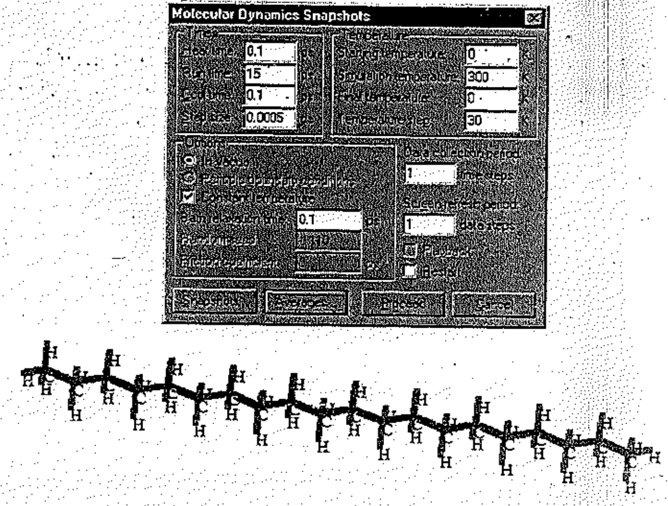

Figure 9.13 Molecular dynamics calculation on hydrocarbon

To give a simple example, consider the straight chain hydrocarbon $C_{20}H_{42}$, shown in Figure 9.13. A MM+ optimization starting from the fully stretched structure gave the structure shown. There is restricted rotation about each C—C bond and in a later chapter I will show how theories of polymeric molecules focus attention on the end-to-end distance; the simplest theories predict relationships between this and the number of C—C links in such a molecule. $C_{20}H_{42}$ is barely a polymer, but the principles are the same. There are very many conformers, all similar in energy.

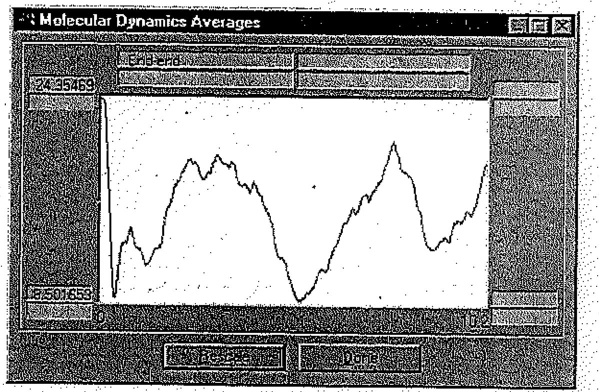

Figure 9.14 End-to-end distance

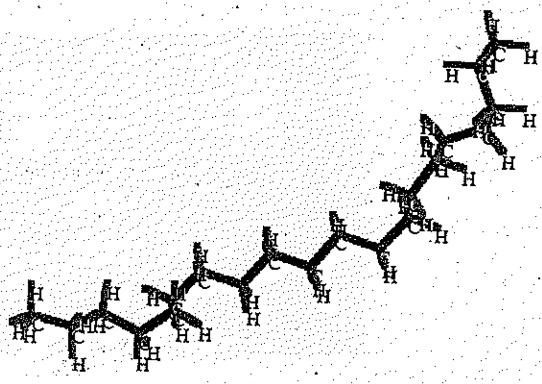

Figure 9.15 C20 hydrocarbon at end of MD experiment

I therefore set up a 15 ps molecular dynamics calculation, as shown in Figure 9.13. The options are shown, and statistics were collected in an MS/EXCEL file. The end-to-end distance varied dramatically over the experiment, from its initial value of 2435 pm through 850 pm. The average end-to-end distance was 1546 pm, as shown in Figures 9.14 and 9.15. The resulting structure was then optimized. One strategy in such calculations is to take snapshots of the structure at regular time intervals and then optimize from each snapshot.

10 Monte Carlo

10.1 Introduction

When discussing QSAR in Chapter 7, I mentioned that the molecular volume was thought to be a useful index and explained how it could be calculated within the framework of a molecular mechanics (MM) study.

Figure 10.1 shows a diatomic molecule, where the atomic radii are R_A and R_B. The problem is to find the volume of the molecule AB. The volume of atom A is $\frac{4}{3}\pi R_A^3$ and the total for AB is given by this plus the part of B that is not shared with atom A. Thus, we have to be careful not to count volume X twice in any calculation.

I suggested that the total could be calculated by surrounding atom B with a regular grid of side $2R_B$, and examining which grid points lay inside B but excluding those lying in region X, and those that lay outside atom B. A simple ratio then gives the contribution to the volume from atom B. The process can be easily extended to any polyatomic system.

Rather than take a regularly spaced cubic array of points around atom B, we can surround atom B by an (imaginary) cube of side $2R_B$ and choose points at random inside this cube. For each point, we examine whether it lies outside atom B, in which case we reject it. If the point lies inside atom B but also inside atom A (i.e. it lies in the X region), we also reject it. Otherwise we increment the number of successful points. A simple proportionality between the number of successful points and the total number tried gives the required volume.

The following BASIC program accomplishes the calculation.

```
REM   MONTE CARLO CALCULATION OF DIATOMIC A–B VOLUME
REM   ATOMIC RADII RA AND RB, SEPARATION RAB.
REM
PI = 3.14159265359#
NPOINTS = 10 000
RA = 2
RB = 1
RAB = 2!
XA = 0!
YA = 0!
```

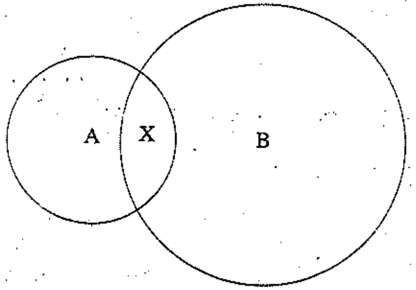

Figure 10.1 A MM diatomic

```
ZA = 0!
XB = 0!
YB = 0!
ZB = RAB
VA = 4/3 * PI * RA^3
VB = 0
FOR I = 1 TO NPOINTS
XI = XB + RB * (2 * RND - 1)
YI = YB + RB * (2 * RND - 1)
ZI = ZB + RB * (2 * RND - 1)
DISTB = SQR ((XI - XB)^2 + (YI - YB)^2 + (ZI - ZB)^2)
IF DISTB > RB THEN 2000
DISTA = SQR ((XI - XA)^2 + (YI - YA)^2 + (ZI - ZA)^2)
IF DISTA < RA THEN 2000
VB = VB + 1
2000 NEXT I
VB = 8 * RB^3 * VB/NPOINTS
VTOTAL = VA + VB
PRINT
PRINT "VOLUME AFTER"; NPOINTS; "POINTS IS"; VTOTAL
END
```

You might have to modify the statements involving RND to make it run with your version of BASIC. I have taken two atoms with radii 2 units and 1 unit, and the atoms are separated by R_{AB} that I have set to 2 units. I have calculated the volume of atom A as $\frac{4}{3}\pi R_A^3$. I then chose 10 000 points at random within a cube of side $2R_B$ surrounding atom B.

If the point lies outside atom B, then we do nothing apart from increment the counter and progress to the next point. If the point lies inside atom B, then we test whether it also lies inside atom A (in which case we have already counted it). I hope the code is transparent; this little program can be easily extended to deal with a polyatomic molecule.

This calculation is an example of the *Monte Carlo* technique (denoted MC). MC is a generic term applied to calculations that involve the use of random numbers for sampling; it became widely used towards the end of the Second World War by physicists trying to study the diffusion of neutrons in fissionable material. In fact, the MC method was first discussed by the French eighteenth-century naturalist Buffon, who discovered that if a needle of length l were dropped at random onto a set of parallel lines with spacing $d > l$, then the probability of the needle crossing a line is $2l/\pi d$.

Straightforward thermodynamic quantities such as the pressure, the internal energy and the Gibbs energy turn out to be impossible to calculate directly for a macroscopic system, simply because of the large number of particles involved. In fact, it is not even sensible to contemplate recording an initial starting point for 10^{23} particles, let alone devising methods for solving the equations of motion. Boltzmann and Gibbs recognized this problem, and invented the subject of *statistical thermodynamics*.

Suppose for the sake of argument that we have a system of N simple atomic particles (such as argon atoms) and we are interested in the total potential energy Φ. If the pair potential is $U_{ij}(R)$ and the potentials are pairwise additive, then we have

$$\Phi = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} U_{ij}(R_{ij})$$

If I denote the position vectors of the N particles measured at time t by $\mathbf{R}_A(t), \mathbf{R}_B(t), \ldots, \mathbf{R}_N(t)$, then the position vectors will depend on time and the instantaneous value of Φ will depend on the values of the variables at that time. If we make enough measurements $\Phi(t_1), \Phi(t_2), \ldots, \Phi(t_n)$ at times $t_1, t_2, \ldots, t_n$, then the average of these measurements will approach the mean value of Φ

$$\langle\Phi\rangle = \frac{1}{n}\sum_{i=1}^{n} \Phi(t_i)$$

Other properties such as the self-diffusion depend on fluctuations about mean values. In statistical thermodynamics, rather than calculating a time average we consider the average over a large number of replications of the system (an ensemble). The *ergodic theorem* tells us that the two are the same.

In Chapter 8 I considered the case of a canonical ensemble, which consists of a large number of replications (or cells), each of which is identical in the sense that the number of particles, the volume and the temperature are the same. The cells are not identical at the atomic level, all that matters is N, V and T. Energy can flow from one cell to another, but the total energy of the ensemble is constant.

Under these conditions, the chance of finding a cell with energy E is proportional to the Boltzmann factor

$$\exp\left(-\frac{E}{k_B T}\right)$$

The mean value $\langle E \rangle$ can be found by averaging over the replications

$$\langle E \rangle = \frac{\sum E_i \exp\left(-\frac{E_i}{k_B T}\right)}{\sum \exp\left(-\frac{E_i}{k_B T}\right)} \tag{10.1}$$

I explained the importance of the configurational integral, which depends only on exponential terms involving the total mutual potential energy Φ

$$\int \exp\left(-\frac{\Phi(\mathbf{q})}{k_B T}\right) d\mathbf{q} \tag{10.2}$$

and showed how it could be related to the so-called excess thermodynamic functions.

I have written the configurational integral in a simplified way; if there are N particles, then it is actually a $3N$ dimensional integral over the positions of the N particles. We might want to take $N = 10\,000$ particles, but a little thought shows that it is impracticable to carry out such a multidimensional integral by the usual techniques of numerical analysis. Instead we resort to the MC method and generate a representative number of points in conformation space. For each point we calculate a value of Φ. The integral is approximated by a finite sum.

The first serious MC calculation was that of N. Metropolis *et al.* [41]. In fact, this paper marked the birth of computer simulation as a statistical mechanical technique, and it preceded the molecular dynamics studies mentioned in Chapter 9. I can't do better than repeat the words of the authors (i.e. quote the Abstract):

> *A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC, and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.*

They took a two-dimensional square array of 224 hard disks with finite radius. Each disk was moved in turn according to the formulae

$$X \rightarrow X + \alpha\xi_1; \qquad Y \rightarrow Y + \alpha\xi_2$$

where α is some maximum allowed distance and the ξ_i are random numbers between 0 and 1. In the case that the move would put one disk on top of another disk, the move is not allowed and we progress to the next disk. This is a primitive form of the technique known as *importance sampling* that is a key feature of modern MC simulations.

In the case of a more complicated problem such as the Lennard-Jones potential, a more sophisticated treatment of importance sampling is needed. The average potential energy can be obtained from the configuration integral

$$\langle \Phi \rangle = \frac{\int \Phi(\mathbf{q}) \exp\left(-\frac{\Phi(\mathbf{q})}{k_B T}\right) d\mathbf{q}}{\int \exp\left(-\frac{\Phi(\mathbf{q})}{k_B T}\right) d\mathbf{q}} \tag{10.3}$$

and each integral is approximated by a finite sum of M terms in MC

$$\langle \Phi \rangle \approx \frac{\sum_{i=1}^{M} \Phi_i(\mathbf{q}) \exp\left(-\frac{\Phi_i(\mathbf{q})}{k_B T}\right)}{\sum_{i=1}^{M} \exp\left(-\frac{\Phi_i(\mathbf{q})}{k_B T}\right)} \tag{10.4}$$

If we calculate the mutual potential energy $\Phi_i(\mathbf{q})$ for a random array of atoms, then the array will have a Boltzmann weight of

$$\exp\left(-\frac{\Phi(\mathbf{q})}{k_B T}\right)$$

and if some atoms are close together, then Φ_i will be large and the Boltzmann factor small.

Most of phase space corresponds to non-physical configurations with very high energies and only for a very small proportion of phase space does the Boltzmann factor have an appreciable value. The authors proposed a modified MC scheme where, instead of choosing configurations randomly and then weighting them with $\exp(-\Phi/k_B T)$, configurations were chosen with a probability distribution of $\exp(-\Phi/k_B T)$ and then weighted equally. This adds a step to the algorithm as follows.

Before making a move, we calculate the energy change of the system caused by the move. If the energy change is negative, then the move is allowed. If the energy change is positive, then we allow the move but with a probability of $\exp(-\Phi/k_B T)$; to do this, we generate a random number ξ_3 between 0 and 1, and if

$$\xi_3 < \exp\left(-\frac{\Phi}{k_B T}\right)$$

then we allow the move to take place and increment the numerator and denominator in Equation (10.1). Otherwise we leave the particle in its old position and move on to the next particle. This summarizes the *Metropolis Monte Carlo* technique.

The box is taken to be periodic, so if a move places the particle outside the box, then an image particle enters the box from the opposite side. The procedure is repeated for very many trials, and the configurational integral is estimated as a sum.

Hard disk systems were the first ones investigated, and they occupy an important place in the history of modelling. The results of the first successful 'realistic' MC

Simulation of an atomic fluid was published in 1957 by W. W. Wood and F. R. Parker [42]. Again, let the authors tell the story in their own words.

> *Values obtained by Monte Carlo calculations are reported for the compressibility factor, excess internal energy, excess constant-volume heat capacity, and the radial distribution function of Lennard-Jones (12-6) molecules at the reduced temperature $k_BT/\varepsilon^* = 2.74$, and at thirteen volumes between $v/v^* = 0.75$ and 7.5 (v is the molar volume; $v^* = 2^{-1/2}\ N_0r^{*3}$; N_0 is Avogadro's number; ε^* is the depth and r^* the radius of the Lennard-Jones potential well). The results are compared with the experimental observations of Michels (≈150–2000 atm) and Bridgman (≈2000–15 000 atm) on argon at 55 °C, using Michels' second virial coefficient values for the potential parameters. Close agreement with Michels is found, but significant disagreement with Bridgman. The Monte Carlo calculations display the fluid–solid transition; the transition pressure and the volume and enthalpy increments are not precisely determined. The Lennard-Jones–Devonshire cell theory gives results which disagree throughout the fluid phase, but agree on the solid branch of the isotherm. Limited comparisons with the Kirkwood–Born–Green results indicate that the superposition approximation yields useful results at least up to $v/v^* = 2.5$.*

They therefore studied a three-dimensional Lennard-Jones fluid (i.e. argon) and compared their results with experimental equation of state data. One of the most important technical questions discussed in the paper is the procedure used to truncate a potential in a system with periodic boundary conditions.

10.2 MC Simulation of Rigid Molecules

For rigid non-spherical molecules, it is necessary to vary their orientation as well as the position in space; usually each molecule is translated and rotated by random amounts once per cycle. The simplest approach is to rotate about each Cartesian axis in turn by an angle chosen at random but subject to a suitable maximum allowed variation. For example, if the position vector of an atom is the column vector **R**, then the position vector after rotation by α about the x-axis is given by

$$\mathbf{R}' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} \mathbf{R} \qquad (10.5)$$

The three *Euler angles* ϕ, θ and ψ are often used when discussing rotational motion. First of all we rotate by ϕ about the Cartesian z-axis. This changes the x and y-axes to x' and y' as shown in Figure 10.2. Then we rotate by θ about the new x-axis which changes the y and z-axes. Finally, we rotate by ψ about the new z-axis.

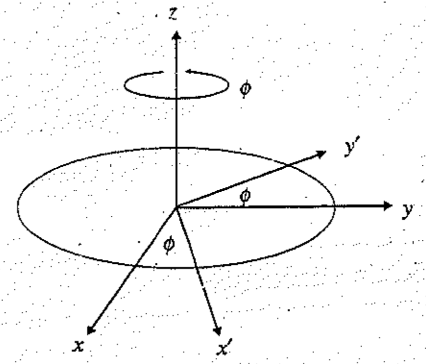

Figure 10.2 First Euler angle

The relationship between a position vector before and after the three rotations is

$$\mathbf{R}' = \begin{pmatrix} \cos\phi\cos\psi - \sin\phi\cos\theta\sin\psi & \sin\phi\cos\psi + \cos\phi\cos\theta\sin\psi & \sin\theta\sin\psi \\ -\cos\phi\sin\psi - \sin\phi\cos\theta\cos\psi & -\sin\phi\sin\psi + \cos\phi\cos\theta\cos\psi & \sin\theta\cos\psi \\ \sin\phi\sin\theta & -\cos\phi\sin\theta & \cos\theta \end{pmatrix} \mathbf{R} \qquad (10.6)$$

There are two technical problems with the use of Euler angles. First, sampling the angles at random does not give a uniform distribution; it is necessary to sample ψ, $\cos\theta$ and ψ. Second, there are a total of six trigonometric function evaluations per rotation.

A preferred alternative makes use of *quaternions*, which are four-dimensional unit vectors. A quaternion **q** is usually written in terms of the scalar quantities q_0, q_1, q_2 and q_3 as

$$\mathbf{q} = (q_0 \quad q_1 \quad q_2 \quad q_3)^{\mathrm{T}}$$

and the components satisfy

$$q_0^2 + q_1^2 + q_2^2 + q_4^2 = 1$$

They can be related to the Euler angles as follows

$$\begin{aligned} q_0 &= \cos\tfrac{1}{2}\theta\cos\tfrac{1}{2}(\phi + \psi) \\ q_1 &= \sin\tfrac{1}{2}\theta\cos\tfrac{1}{2}(\phi - \psi) \\ q_2 &= \sin\tfrac{1}{2}\theta\sin\tfrac{1}{2}(\phi - \psi) \\ q_3 &= \cos\tfrac{1}{2}\theta\sin\tfrac{1}{2}(\phi + \psi) \end{aligned} \qquad (10.7)$$

The rotation matrix can be written as

$$\mathbf{R}' = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 + q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 - q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 + q_0q_1) \\ 2(q_1q_3 + q_0q_2) & 2(q_2q_3 - q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} \mathbf{R} \quad (10.8)$$

and all that is necessary is to generate four suitable random numbers.

10.3 Flexible Molecules

Monte Carlo simulations of flexible molecules are difficult to perform unless the system is small or some of the internal degrees of freedom are kept fixed. The simplest way to generate a new configuration is to perform random changes to the Cartesian coordinates of individual atoms in the molecule but it is a common experience that very small changes are needed in order to produce an acceptable Boltzmann factor (in the Metropolis MC sense).

11 Introduction to Quantum Modelling

By the early days of the twentieth century, scientists had successfully developed three of the four great cornerstones of physics; Sir Isaac Newton's mechanics, James Clerk Maxwell's electromagnetic theory and Albert Einstein's theory of special relativity. They had a picture of the physical world where matter was made from point particles and radiation consisted of electromagnetic waves, and this picture seemed to explain all known physical phenomena with just a few untidy exceptions.

These untidy exceptions comprised phenomena such as *the theory of black body radiation, the photoelectric effect, Compton scattering, atomic structure and spectra* and a few other apparently unrelated experimental findings. I don't have space in this book to go into the historical detail; I will simply say that a thorough study of such phenomena led to the fourth cornerstone, quantum theory. Every textbook that has to deal with quantum theory has its own particular treatment and starting point, and this one is no exception; I am going to assume that you have heard of Erwin Schrödinger and his famous equations, and start the discussion at that point. Perhaps I can reassure you by saying that most professional scientists perceive quantum theory as a hard subject (along with electromagnetism). Even Schrödinger didn't fully understand the physical meaning that we now attach to his wavefunctions when he first wrote down his famous equation and solved it for the hydrogen atom.

11.1 The Schrödinger Equation

Consider then the simple case of a particle of mass m constrained to the x-axis by a potential $U(x, t)$ that depends on x and time t. I have allowed the potential to be time dependent in order to cover the possibility of (for example) an electron being influenced by an external electromagnetic wave. Schrödinger's time-dependent equation states that

$$\left(-\frac{h^2}{8\pi^2 m}\frac{\partial^2}{\partial x^2} + U(x,t)\right)\Psi(x,t) = \mathrm{j}\frac{h}{2\pi}\frac{\partial \Psi(x,t)}{\partial t} \qquad (11.1)$$