Chapter 7. Statistical Mechanics

When one is faced with a system containing many molecules at or near thermal equilibrium, it is not necessary or even wise to try to describe it in terms of quantum wave functions or even classical trajectories following the positions and momenta of all of the constituent particles. Instead, the powerful tools of statistical mechanics allow one to focus on quantities that describe the many-molecule system in terms of the behavior it displays most of the time. In this Chapter, you will learn about these tools and see some important examples of their application.

7.1. Collections of Molecules at or Near Equilibrium

As introduced in Chapter 5, the approach one takes in studying a system composed of a very large number of molecules at or near thermal equilibrium can be quite different from how one studies systems containing a few isolated molecules. In principle, it is possible to conceive of computing the quantum energy levels and wave functions of a collection of many molecules (e.g., ten Na^+ ions, ten Cl^- ions and 550 H_2O molecules in a volume chosen to simulate a concentration of 1 molar NaCl (aq)), but doing so becomes impractical once the number of atoms in the system reaches a few thousand or if the molecules have significant intermolecular interactions as they do in condensed-phase systems. Also, as noted in Chapter 5, following the time evolution of such a large number of molecules can be confusing if one focuses on the short-time behavior of any single molecule (e.g., one sees jerky changes in its energy, momentum, and angular momentum). By examining, instead, the long-time average behavior of each molecule or, alternatively, the average properties of a significantly large number of molecules, one is often better able to understand, interpret, and simulate such condensed-media systems. Moreover, most experiments do not probe such short-time dynamical properties of single molecules; instead, their signals report on the behavior of many molecules lying within the range of their detection device (e.g., laser beam, STM tip, or

electrode). It is when one want to describe the behavior of collections of molecules under such conditions that the power of statistical mechanics comes into play.

7.1.1 The Distribution of Energy Among Levels

One of the most important concepts of statistical mechanics involves how a specified total amount of energy E can be shared among a collection of molecules and within the internal (rotational, vibrational, electronic) and intermolecular (translational) degrees of freedom of these molecules when the molecules have a means for sharing or redistributing this energy (e.g., by collisions). The primary outcome of asking what is the most probable distribution of energy among a large number N of molecules within a container of volume V that is maintained in equilibrium by such energy-sharing at a specified temperature T is the most important equation in statistical mechanics, the Boltzmann population formula:

$$P_j = \Omega_j \exp(- E_j /kT)/Q.$$

This equation expresses the probability P_j of finding the system (which, in the case introduced above, is the whole collection of N interacting molecules) in its jth quantum state, where E_j is the energy of this quantum state, T is the temperature in K, Ω_j is the degeneracy of the jth state, and the denominator Q is the so-called partition function:

$$Q = \Sigma_j \Omega_j \exp(- E_j /kT).$$

The classical mechanical equivalent of the above quantum Boltzmann population formula for a system with a total of M coordinates (collectively denoted q- they would be the internal and intermolecular coordinates of the N molecules in the system) and M momenta (denoted p) is:

$$P(q,p) = h^{-M} \exp (- H(q, p)/kT)/Q,$$

where H is the classical Hamiltonian, h is Planck's constant, and the classical partition function Q is

$$Q = h^{-M} \int \exp(-H(q, p)/kT) \, dq \, dp .$$

This probability density expression, which must integrate to unity, contains the factor of h^{-M} because, as we saw in Chapter 1 when we learned about classical action, the integral of a coordinate-momentum product has units of Planck's constant.

Notice that the Boltzmann formula does not say that only those states of one particular energy can be populated; it gives non-zero probabilities for populating all states from the lowest to the highest. However, it does say that states of higher energy E_j are disfavored by the $\exp(-E_j/kT)$ factor, but, if states of higher energy have larger degeneracies Ω_j (which they usually do), the overall population of such states may not be low. That is, there is a competition between state degeneracy Ω_j, which tends to grow as the state's energy grows, and $\exp(-E_j/kT)$ which decreases with increasing energy. If the number of particles N is huge, the degeneracy Ω grows as a high power (let's denote this power as K) of E because the degeneracy is related to the number of ways the energy can be distributed among the N molecules. In fact, K grows at least as fast as N. As a result of Ω growing as E^K, the product function $P(E) = E^K \exp(-E/kT)$ has the form shown in Fig. 7.1 (for K=10, for illustrative purposes).
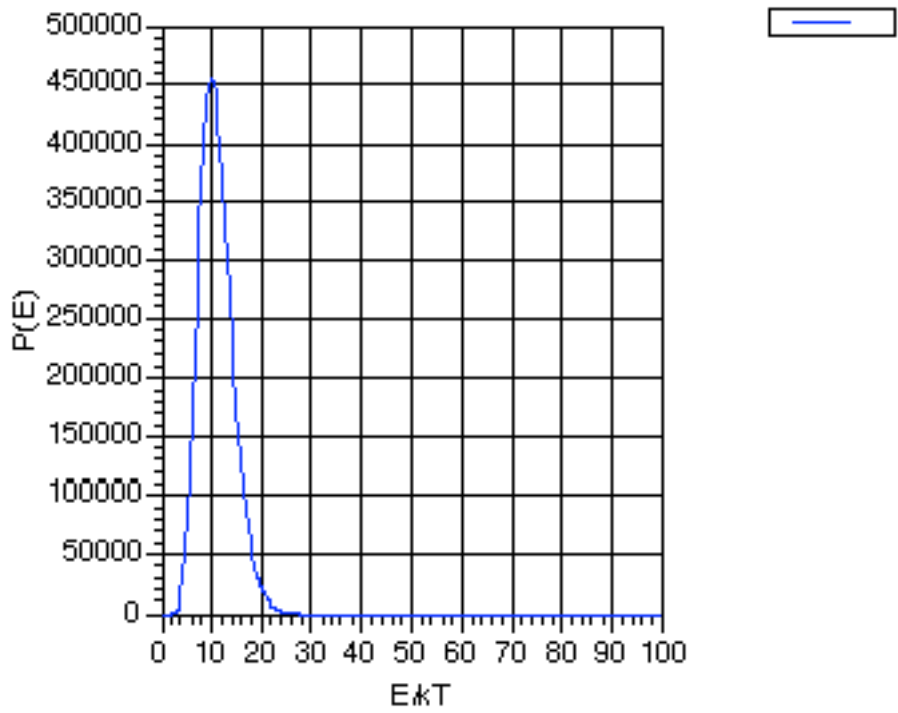
Figure 7.1 Probability Weighting Factor P(E) as a Function of E for K = 10.

By taking the derivative of this function P(E) with respect to E, and finding the energy at which this derivative vanishes, one can show that this probability function has a peak at $E^* = K\,kT$, and that at this energy value,

$$P(E^*) = (KkT)^K \exp(-K),$$

By then asking at what energy E' the function P(E) drops to exp(-1) of this maximum value P(E*):

$$P(E') = \exp(-1)\,P(E^*),$$

one finds

$$E' = K\,kT\,(1 + (2/K)^{1/2}).$$

So the width of the P(E) graph, measured as the change in energy needed to cause P(E) to drop to exp(-1) of its maximum value divided by the value of the energy at which P(E) assumes this maximum value, is

$$(E'-E^*)/E^* = (2/K)^{1/2}.$$

This width gets smaller and smaller as K increases.

The primary conclusion is that as the number N of molecules in the sample grows, which, as discussed earlier, causes K to grow, the energy probability function becomes more and more sharply peaked about the most probable energy E*. This, in turn, suggests that we may be able to model, aside from infrequent fluctuations which we may also find a way to take account of, the behavior of systems with many molecules by focusing on the most probable situation (i.e., those having the energy E*) and ignoring or making small corrections for deviations from this case.

It is for the reasons just shown that for macroscopic systems near equilibrium, in which N (and hence K) is extremely large (e.g., N ~ 10^{10} to 10^{24}), only the most probable distribution of the total energy among the N molecules need be considered. This is the situation in which the equations of statistical mechanics are so useful. Certainly, there are fluctuations (as evidenced by the finite width of the above graph) in the energy content of the N-molecule system about its most probable value. However, these fluctuations become less and less important as the system size (i.e., N) becomes larger and larger.

1. Basis of the Boltzmann Population Formula

To understand how this narrow Boltzmann distribution of energies arises when the number of molecules N in the sample is large, we consider a system composed of M identical containers, each having volume V, and each made out a material that allows for efficient heat transfer to its surroundings (e.g., through collisions of the molecules inside the volume with the walls of the container) but material that does not allow any of the N molecules in each container to escape. These containers are arranged into a regular lattice as shown in Fig. 7.2 in a manner that allows their thermally conducting walls to come into contact. Finally, the entire collection of M such containers is surrounded by a perfectly insulating material that assures that the total energy (of all NxM molecules) can not change. So, this collection of M identical containers each containing N molecules constitutes a closed (i.e., with no molecules coming or going) and isolated (i.e., so total energy is constant) system.
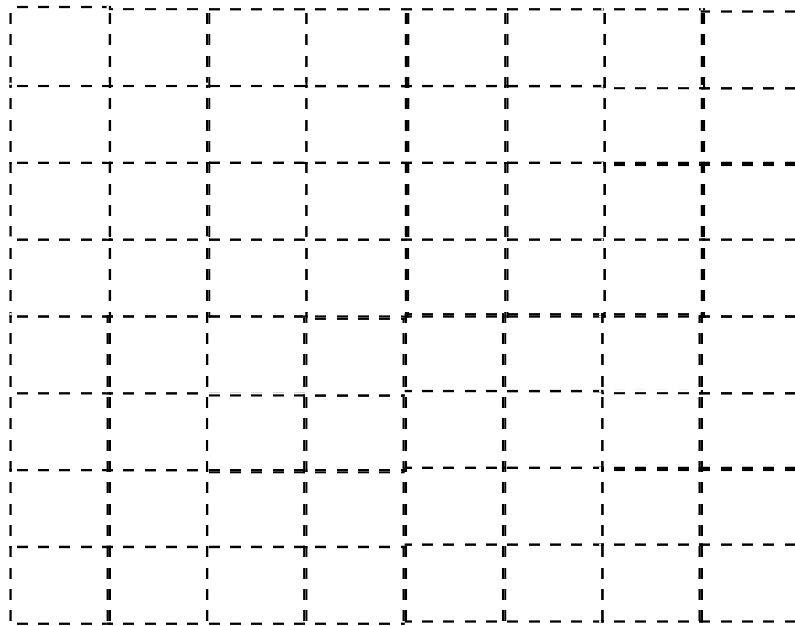
Each Cell Contains N molecules in Volume V. There are M such Cells and the Total Energy of These M Cells is E

Figure 7.2 Collection of M identical cells having energy-conducting walls that do not allow molecules to pass between cells.

2. Equal a priori Probability Assumption

One of the fundamental assumptions of statistical mechanics is that, for a closed isolated system at equilibrium, all quantum states of the system having energy equal to the energy E with which the system is prepared are equally likely to be occupied. This is called the assumption of equal a priori probability for such energy-allowed quantum states. The quantum states relevant to this case are not the states of individual molecules, nor are they the states of N of the molecules in one of the containers of volume V. They are the quantum states of the entire system comprised of NxM molecules. Because our system consists of M identical containers, each with N molecules in it, we can describe the quantum states of the entire system in terms of the quantum states of each such container. It may seem foolish to be discussing quantum states of the large system containing NxM molecules, given what I said earlier about the futility in trying to find such states. However, what I am doing at this stage is to carry out a derivation that is

based upon such quantum states but whose final form and final working equations will not actually require one to know or even be able to have these states in hand.

Let's pretend that we know the quantum states that pertain to N molecules in a container of volume V as shown in Fig. 7.2, and let's label these states by an index J. That is J=1 labels the lowest-energy state of N molecules in the container of volume V, J=2 labels the second such state, and so on. As I said above, I understand it may seem daunting to think of how one actually finds these N-molecule eigenstates. However, we are just deriving a general framework that gives the probabilities of being in each such state. In so doing, we are allowed to pretend that we know these states. In any actual application, we will, of course, have to use approximate expressions for such energies.

Assuming that the walls that divide the M containers play no role except to allow for collisional (i.e., thermal) energy transfer among the containers, an energy-labeling for states of the entire collection of M containers can be realized by giving the number of containers that exist in each single-container J-state. This is possible because, under the assumption about the role of the walls just stated, the energy of each M-container state is a sum of the energies of the M single-container states that comprise that M-container state. For example, if M= 9, the label 1, 1, 2, 2, 1, 3, 4, 1, 2 specifies the energy of this 9-container state in terms of the energies $\{\varepsilon_\varphi\}$ of the states of the 9 containers: $E = 4 \varepsilon_1 + 3 \varepsilon_2 + \varepsilon_3 + \varepsilon_4$. Notice that this 9-container state has the same energy as several other 9-container states; for example, 1, 2, 1, 2, 1, 3, 4, 1, 2 and 4, 1, 3, 1, 2, 2, 1, 1, 2 have the same energy although they are different individual states. What differs among these distinct states is which box occupies which single-box quantum state.

The above example illustrates that an energy level of the M-container system can have a high degree of degeneracy because its total energy can be achieved by having the various single-container states appear in various orders. That is, which container is in which state can be permuted without altering the total energy E. The formula for how many ways the M container states can be permuted such that:

i.   there are n_J containers appearing in single-container state J, with

ii.   a total of M containers, is

$$\Omega(\mathbf{n}) = M!/\{\Pi_J n_J!\}.$$

Here $\mathbf{n} = \{n_1, n_2, n_3, \ldots n_J, \ldots\}$ denote the number of containers existing in single-container states $1, 2, 3, \ldots J, \ldots$. This combinatorial formula reflects the permutational degeneracy arising from placing n_1 containers into state 1, n_2 containers into state 2, etc.

If we imagine an extremely large number of containers and we view M as well as the $\{n_J\}$ as being large numbers (n.b., we will soon see that this is the case at least for the most probable distribution that we will eventually focus on), we can ask- for what choices of the variables $\{n_1, n_2, n_3, \ldots n_J, \ldots\}$ is this degeneracy function $\Omega(\mathbf{n})$ a maximum? Moreover, we can examine $\Omega(\mathbf{n})$ at its maximum and compare its value at values of the $\{\mathbf{n}\}$ parameters changed only slightly from the values that maximized $\Omega(\mathbf{n})$. As we will see, Ω is very strongly peaked at its maximum and decreases extremely rapidly for values of $\{\mathbf{n}\}$ that differ only slightly from the optimal values. It is this property that gives rise to the very narrow energy distribution discussed earlier in this Chapter. So, let's take a closer look at how this energy distribution formula arises.

We want to know what values of the variables $\{n_1, n_2, n_3, \ldots n_J, \ldots\}$ make $\Omega = M!/\{\Pi_J n_J!\}$ a maximum. However, all of the $\{n_1, n_2, n_3, \ldots n_J, \ldots\}$ variables are not independent; they must add up to M, the total number of containers, so we have a constraint

$$\Sigma_J\, n_J = M$$

that the variables must obey. The $\{n_j\}$ variables are also constrained to give the total energy E of the M-container system when summed as

$$\Sigma_J\, n_J \varepsilon_J = E.$$

We have two problems: i. how to maximize Ω and ii. how to impose these constraints. Because Ω takes on values greater than unity for any choice of the $\{n_j\}$, Ω will experience its maximum where $\ln\Omega$ has its maximum, so we can maximize $\ln\Omega$ if doing so helps. Because the n_J variables are assumed to take on large numbers (when M is

large), we can use Sterling's approximation for the natural logarithm of the factorial of a large number:

$$\ln X! = X \ln X - X$$

to approximate $\ln \Omega$ as follows:

$$\ln \Omega = \ln M! - \Sigma_J \{n_J \ln n_J - n_J).$$

This expression will prove useful because we can take its derivative with respect to the n_J variables, which we need to do to search for the maximum of $\ln \Omega$.

To impose the constraints $\Sigma_J n_J = M$ and $\Sigma_J n_J \varepsilon_J = E$ we use the technique of Lagrange multipliers. That is, we seek to find values of $\{n_J\}$ that maximize the following function:

$$F = \ln M! - \Sigma_J \{n_J \ln n_J - n_J) - \alpha(\Sigma_J n_J - M) - \beta(\Sigma_J n_J \varepsilon_J - E).$$

Notice that this function F is exactly equal to the $\ln \Omega$ function we wish to maximize whenever the $\{n_J\}$ variables obey the two constraints. So, the maxima of F and of $\ln \Omega$ are identical if the $\{n_J\}$ have values that obey the constraints. The two Lagrange multipliers α and β are introduced to allow the values of $\{n_J\}$ that maximize F to ultimately obey the two constraints. That is, we first find values of the $\{n_J\}$ variables that make F maximum; these values will depend on α and β and will not necessarily obey the constraints. However, we will then choose α and β to assure that the two constraints are obeyed. This is how the Lagrange multiplier method works.

Taking the derivative of F with respect to each independent n_K variable and setting this derivative equal to zero gives:

$$- \ln n_K - \alpha - \beta \, \varepsilon_K = 0.$$

This equation can be solved to give $n_K = \exp(-\alpha) \exp(-\beta \, \varepsilon_K)$. Substituting this result into the first constraint equation gives $M = \exp(-\alpha) \Sigma_J \exp(-\beta \, \varepsilon_J)$, which allows us to solve for $\exp(-\alpha)$ in terms of M. Doing so, and substituting the result into the expression for n_K gives:

$$n_K = M \, \exp(-\beta \, \varepsilon_K)/Q$$

where

$$Q = \Sigma_J \exp(-\beta \, \varepsilon_J).$$

Notice that the n_K are, as we assumed earlier, large numbers if M is large because n_K is proportional to M. Notice also that we now see the appearance of the partition function Q and of exponential dependence on the energy of the state that gives the Boltzmann population of that state.

It is possible to relate the β Lagrange multiplier to the total energy E of the M containers by summing the number of containers in the Kth quantum state n_K multiplied by the energy of that quantum state ε_K

$$E = \Sigma_K n_K \varepsilon_K = M \Sigma_K \varepsilon_K \exp(-\beta \, \varepsilon_K)/Q$$

$$= -M \, (\partial \ln Q / \partial \beta)_{N,V}.$$

This shows that the average energy of a container, computed as the total energy E divided by the number M of such containers can be computed as a derivative of the logarithm of the partition function Q. As we show in the following Section of this Chapter, all thermodynamic properties of the N molecules in the container of volume V can be obtained as derivatives of the natural logarithm of this Q function. This is why the partition function plays such a central role in statistical mechanics.

To examine the range of energies over which each of the M single-container system varies with appreciable probability, let us consider not just the degeneracy $\Omega(\mathbf{n}^*)$ of that set of variables $\{\mathbf{n}^*\} = \{n^*_1, n^*_2, \ldots\}$ that makes Ω maximum, but also the degeneracy $\Omega(\mathbf{n})$ for values of $\{n_1, n_2, \ldots\}$ differing by small amounts $\{\delta n_1, \delta n_2, \ldots\}$ from the optimal values $\{\mathbf{n}^*\}$. Expanding $\ln \Omega$ as a Taylor series in the parameters $\{n_1, n_2, \ldots\}$ and evaluating the expansion in the neighborhood of the values $\{\mathbf{n}^*\}$, we find:

$$\ln \Omega = \ln \Omega(\{n^*_1, n^*_2, \ldots\}) + \Sigma_J \, (\partial \ln\Omega/\partial n_J) \, \delta n_J + 1/2 \, \Sigma_{J,K} \, (\partial^2 \ln\Omega/\partial n_J \partial n_K) \, \delta n_J \, \delta n_K + \ldots$$

We know that all of the first derivative terms $(\partial \ln\Omega/\partial n_J)$ vanish because $\ln\Omega$ has been made maximum at $\{\mathbf{n}^*\}$. To evaluate the second derivative terms, we first note that the first derivative of $\ln\Omega$ is

$$(\partial \ln\Omega/\partial n_J) \; = \partial(\ln M! - \Sigma_J \{n_J \ln n_J - n_J\})/\partial n_J = -\ln(n_J).$$

So the second derivatives needed to complete the Taylor series through second order are:

$$(\partial^2 \ln\Omega/\partial n_J \partial n_K) = - \, \delta_{J,K} \, n_j^{-1}.$$

Using this result, we can expand $\Omega(\mathbf{n})$ in the neighborhood of $\{\mathbf{n}^*\}$ in powers of $\delta n_J = n_J - n_J^*$ as follows:

$$\ln \Omega(\mathbf{n}) = \ln \Omega(\mathbf{n}^*) - 1/2 \, \Sigma_J \, (\delta n_J)^2/n_J^*,$$

or, equivalently,

$$\Omega(\mathbf{n}) = \Omega(\mathbf{n}^*) \exp[-1/2\Sigma_J \, (\delta n_J)^2/n_J^*]$$

This result clearly shows that the degeneracy, and hence, by the equal a priori probability hypothesis, the probability of the M-container system occupying a state having $\{n_1, n_2, ..\}$

falls off exponentially as the variables n_J move away from their most-probable values {**n***}.

3. The Thermodynamic Limit

As we noted earlier, the n_J^* are proportional to M (i.e., $n_J^* = M \exp(-\beta\varepsilon_J)/Q = f_J M$), so when considering deviations δn_J away from the optimal n_J^*, we should consider deviations that are also proportional to M: $\delta n_J = M \delta f_J$. In this way, we are treating deviations of specified percentage or fractional amount which we denote f_J. Thus, the ratio $(\delta n_J)^2/n_J^*$ that appears in the above exponential has an M-dependence that allows $\Omega(\mathbf{n})$ to be written as:

$$\Omega(\mathbf{n}) = \Omega(\mathbf{n}^*) \exp[-M/2\Sigma_J (\delta f_J)^2/f_J^*],$$

where f_J^* and δf_J are the fraction and fractional deviation of containers in state J: $f_J^* = n_J^*/M$ and $\delta f_J = \delta n_J/M$. The purpose of writing $\Omega(\mathbf{n})$ in this manner is to explicitly show that, in the so-called thermodynamic limit, when M approaches infinity, only the most probable distribution of energy {**n***} need to be considered because only {δf_J=0} is important as M approaches infinity.

4. Fluctuations

Let's consider this very narrow distribution issue a bit further by examining fluctuations in the energy of a single container around its average energy $E_{ave} = E/M$. We already know that the number of containers in a given state K can be written as $n_K = M \exp(-\beta \varepsilon_K)/Q$. Alternatively, we can say that the probability of a container occupying the state J is:

$$P_J = \exp(-\beta \varepsilon_K)/Q.$$

Using this probability, we can compute the average energy E_{ave} as:

$$E_{ave} = \Sigma_J P_J \varepsilon_J = \Sigma_J \varepsilon_J \exp(-\beta \varepsilon_K)/Q = -(\partial lnQ/\partial\beta)_{N,V}.$$

To compute the fluctuation in energy, we first note that the fluctuation is defined as the average of the square of the deviation in energy from the average:

$$(E-E_{ave})^2_{ave.} = \Sigma_J (\varepsilon_J - E_{ave})^2 P_J = \Sigma_J P_J (\varepsilon_J^2 - 2\varepsilon_J E_{ave} + E_{ave}^2) = \Sigma_J P_J(\varepsilon_J^2 - E_{ave}^2).$$

The following identity is now useful for further re-expressing the fluctuations:

$$(\partial^2 \ln Q/\partial\beta^2)_{N,V} = \partial(-\Sigma_J\varepsilon_J \exp(-\beta\varepsilon_J)/Q)/\partial\beta$$

$$= \Sigma_J \varepsilon_J^2 \exp(-\beta\varepsilon_J)/Q - \{\Sigma_J \varepsilon_J\exp(-\beta\varepsilon_J)/Q\}\{\{\Sigma_L \varepsilon_L\exp(-\beta\varepsilon_L)/Q\}$$

Recognizing the first factor immediately above as $\Sigma_J \varepsilon_J^2 P_J$, and the second factor as $- E_{ave}^2$, and noting that $\Sigma_J P_J = 1$, allows the fluctuation formula to be rewritten as:

$$(E-E_{ave})^2_{ave.} = (\partial^2 \ln Q/\partial\beta^2)_{N,V} = - (\partial(E_{ave})/\partial\beta)_{N,V}).$$

Because the parameter β can be shown to be related to the Kelvin temperature T as $\beta = 1/(kT)$, the above expression can be re-written as:

$$(E-E_{ave})^2_{ave} = - (\partial(E_{ave})/\partial\beta)_{N,V}) = kT^2 (\partial(E_{ave})/\partial T)_{N,V}.$$

Recognizing the formula for the constant-volume heat capacity

$$C_V = (\partial(E_{ave})/\partial T)_{N,V}$$

allows the fractional fluctuation in the energy around the mean energy $E_{ave} = E/M$ to be expressed as:

$$(E-E_{ave})^2_{ave}/E_{ave}^2 = kT^2 C_V/E_{ave}^2.$$

What does this fractional fluctuation formula tell us? On its left-hand side it gives a measure of the fractional spread of energies over which each of the containers ranges about its mean energy E_{ave}. On the right side, it contains a ratio of two quantities that are extensive properties, the heat capacity and the mean energy. That is, both C_V and E_{ave} will be proportional to the number N of molecules in the container as long as N is reasonably large. However, because the right-hand side involves C_V/E_{ave}^2, it is proportional to N^{-1} and thus will be very small for large N as long as C_V does not become large. As a result, except near so-called critical points where the heat capacity does indeed become extremely large, the fractional fluctuation in the energy of a given container of N molecules will be very small (i.e., proportional to N^{-1}). This finding is related to the narrow distribution in energies that we discussed earlier in this section.

Let's look at the expression

$$(E-E_{ave})^2{}_{ave}/E_{ave}^2 = kT^2\, C_V/E_{ave}^2$$

in a bit more detail for a system that is small but still contains quite a few particles-a cluster of N Ar atoms at temperature T. If we assume that each of the Ar atoms in the cluster has 3/2 kT of kinetic energy and that the potential energy holding the cluster together is small and constant (so it cancels in $E-E_{ave}$), E_{ave} will be 3/2NkT and C_V will be 3/2 Nk. So,

$$(E-E_{ave})^2{}_{ave}/E_{ave}^2 = kT^2\, C_V/E_{ave}^2 = kT^2\, 3/2Nk\, /(3/2\ NkT)^2 = 2/3\ N^{-1}.$$

In a nano-droplet of diameter 100 Å, with each Ar atom occupying a volume of ca. 4/3 π $(3.8\text{Å})^3 = 232\ \text{Å}^3$, there will be ca.

$$N = 4/3\ \pi\ 100^3/[4/3\ \pi\ 3.8^3] = 1.8\ \text{x}10^4$$

Ar atoms. So, the average fractional spread in the energy

$$\sqrt{\frac{(E-E_{ave})^2}{E_{ave}^2}} = \sqrt{\frac{2}{3N}} = 0.006.$$

That is, even for a very small nano-droplet, the fluctuation in the energy of the system is only a fraction of a percent (assuming C_V is not large as near a critical point). This example shows why it is often possible to use thermodynamic concepts and equations even for very small systems, albeit realizing that fluctuations away from the most probable state are more important than in much larger systems.

7.1. 2 Partition Functions and Thermodynamic Properties

Let us now examine how this idea of the most probable energy distribution being dominant gives rise to equations that offer molecular-level expressions for other thermodynamic properties. The first equation is the fundamental Boltzmann population formula that we already examined:

$$P_j = \exp(- E_j /kT)/Q,$$

which expresses the probability for finding the N-molecule system in its Jth quantum state having energy E_j. Sometimes, this expression is written as

$$P_j = \Omega_j \exp(- E_j /kT)/Q$$

where now the index j is used to label an energy <u>level</u> of the system having energy E_j and degeneracy. It is important for the student to be used to either notation; a level is just a collection of those <u>states</u> having identical energy.

1. System Partition Functions

Using this result, it is possible to compute the average energy E_{ave}, sometimes written as <E>, of the system

$$<E> = \Sigma_j P_j E_j \, ,$$

and, as we saw earlier in this Chapter, to show that this quantity can be recast as

$$<E> = kT^2 \, \partial(\ln Q/\partial T)_{N,V} \, .$$

To review how this proof is carried out, we substitute the expressions for P_j and for Q into the expression for $<E>$ (I will use the notation labeling energy levels rather than energy states to allow the student to become used to this)

$$<E> = \{\Sigma_j E_j \, \Omega_j \exp(-E_j/kT)\}/\{\Sigma_l \, \Omega_l \exp(-E_l/kT)\}.$$

By noting that $\partial \, (\exp(-E_j/kT))/\partial T = (1/kT^2) \, E_j \exp(-E_j/kT)$, we can then rewrite $<E>$ as

$$<E> = kT^2 \, \{\Sigma_j \, \Omega_j \partial \, (\exp(-E_j/kT))/\partial T \, \}/\{\Sigma_l \, \Omega_l \exp(-E_l/kT)\}.$$

And then recalling that $\{\partial X/\partial T\}/X = \partial \ln X/\partial T$, we finally obtain

$$<E> = kT^2 \, (\partial \ln(Q)/\partial T)_{N,V}.$$

All other equilibrium properties can also be expressed in terms of the partition function Q. For example, if the average pressure $<p>$ is defined as the pressure of each quantum state (defined as how the energy of that state changes if we change the volume of the container by a small amount)

$$p_j = (\partial E_j /\partial V)_N$$

multiplied by the probability P_j for accessing that quantum state, summed over all such states, one can show, realizing that only E_j (not T or Ω) depend on the volume V, that

$$<p> = \Sigma_j \, (\partial E_j /\partial V)_N \, \Omega_j \exp(- E_j /kT)/Q$$

$$= kT(\partial \ln Q / \partial V)_{N,T} \ .$$

If you wonder why the energies E_J should depend on the volume V, think of the case of N gas-phase molecules occupying the container of volume V. You know that the translational energies of each of these N molecules depend on the volume through the particle-in-a-box formula

$$E_{n_x,n_y,n_z} = \frac{h^2}{8mL^2}(n_x^2 + n_y^2 + n_z^2) \ .$$

Changing V can be accomplished by changing the box length L. This makes it clear why the energies do indeed depend on the volume V. Of course, there are additional sources of the V-dependence of the energy levels. For example, as one shrinks V, the molecules become more crowded, so their intermolecular energies also change.

Without belaboring the point further, it is possible to express all of the usual thermodynamic quantities in terms of the partition function Q. The average energy and average pressure are given above, as is the heat capacity. The average entropy is given as

$$<S> = k \ln Q + kT(\partial \ln Q / \partial N)_{V,T}$$

the Helmholtz free energy A is

$$A = -kT \ln Q$$

and the chemical potential μ is expressed as follows:

$$\mu = -kT \ (\partial \ln Q / \partial N)_{T,V}.$$

As we saw earlier, it is also possible to express fluctuations in thermodynamic properties in terms of derivatives of partition functions and, thus, as derivatives of other properties. For example, the fluctuation in the energy $<(E-<E>)^2>$ was shown above to be given by

$$<(E-<E>)^2> = kT^2 C_V.$$

The text *Statistical Mechanics*, D. A. McQuarrie, Harper and Row, New York (1977) has an excellent treatment of these topics and shows how all of these expressions are derived.

So, if one were able to evaluate the partition function Q for N molecules in a volume V at a temperature T, either by summing the quantum-level degeneracy and $\exp(-E_j/kT)$ factors

$$Q = \Sigma_j \, \Omega_j \, \exp(- E_j /kT),$$

or by carrying out the phase-space integral over all M of the coordinates and momenta of the system

$$Q = h^{-M} \int \exp (- H(q, p)/kT) \, dq \, dp ,$$

one could then use the above formulas to evaluate any thermodynamic properties and their fluctuations as derivatives of lnQ.

The averages discussed above, derived using the probabilities $P_J = \Omega_J \exp(- E_J /kT)/Q$ associated with the most probable distribution, are called ensemble averages with the set of states associated with the specified values of N, V, and T constituting what is called a canonical ensemble. Averages derived using the probabilities P_J = constant for all states associated with specified values of N, V, and E are called ensemble averages for a microcanonical ensemble. There is another kind of ensemble that is often used in statistical mechanics; it is called the grand canonical ensemble and relates to systems with specified volume V, temperature T, and chemical potential μ (rather than particle number N). To obtain the partition function (from which

all thermodynamic properties are obtained) in this case, one considers maximizing the same function

$$\Omega(\mathbf{n}) = M!/\{\Pi_J n_J!\}$$

introduced earlier, but now considering each quantum (labeled J) as having an energy $E_J(N,V)$ that depends on the volume and on how may particles occupy this volume. The variables $n_J(N)$ are now used to specify how many of the containers introduced earlier contain N particles and are in the Jth quantum state. These variables have to obey the same two constraints as for the canonical ensemble

$$\Sigma_{J,N} n_J(N) = M$$

$$\Sigma_{J,N} n_J(N) \varepsilon_J(N,V) = E,$$

but they also are required to obey

$$\Sigma_{J,N} N n_J(N) = N_{total}$$

which means that the sum adds up to the total number of particles in the isolated system's large container that was divided into M smaller container. In this case, the walls separating each small container are assumed to allow for energy transfer (as in the canonical ensemble) and for molecules to move from one container to another (unlike the canonical ensemble). Using Lagrange multipliers as before to maximize $\ln\Omega(\mathbf{n})$ subject to the above three constraints involves maximizing

$$F = \ln M! - \Sigma_{J,N} \{n_{J,N} \ln n_{J,N} - n_{J,N}\} - \alpha(\Sigma_{J,N} n_{J,N} - M) - \beta(\Sigma_{J,N} n_{J,N} \varepsilon_J - E) - \gamma(\Sigma_{J,N} N n_{J,N}(N) - N_{total})$$

and gives

$$- \ln n_{K,N} - \alpha - \beta \varepsilon_K - \gamma N = 0$$

or

$$n_{K,N} = \exp[-\alpha - \beta\, \varepsilon_K - \gamma\, N].$$

Imposing the first constraint gives

$$M = \Sigma_{K,N} \exp[-\alpha - \beta\, \varepsilon_K - \gamma\, N], \text{ or}$$

$$\exp(-\alpha) = \frac{M}{\displaystyle\sum_{K,N} \exp(-\beta\varepsilon_K(N) - \gamma N)} = \frac{M}{Q(\gamma,V,T)}$$

where the partition function Q is defined by the sum in the denominator. So, now the probability of the system having N particles and being in the Kth quantum state is

$$P_K(N) = \frac{\exp(-\beta\varepsilon_K(N,V) - \gamma N)}{Q}.$$

Very much as was shown earlier for the canonical ensemble, one can then express thermodynamic properties (e.g., E, C_V, etc.) in terms of derivatives of lnQ. The text *Statistical Mechanics*, D. A. McQuarrie, Harper and Row, New York (1977) goes through these derivations in good detail, so I will not repeat them here because we showed how to do so when treating the canonical ensemble. To summarize them briefly, one again uses $\beta = 1/(kT)$, finds that γ is related to the chemical potential μ as

$$\gamma = -\mu\, \beta$$

and obtains

$$p = \sum_{N,K} P_K(N)\{\frac{-\partial\varepsilon_K(N,V)}{\partial V}\}_N = kT\left(\frac{\partial \ln Q}{\partial V}\right)_{\mu,T}$$

$$N_{ave} = \sum_{N,K} N P_K(N) = kT\left(\frac{\partial \ln Q}{\partial \mu}\right)_{V,T}$$

$$S = kT\left(\frac{\partial \ln Q}{\partial T}\right)_{\mu,V} = k \ln Q$$

$$E = \sum_{N,K} \varepsilon_K(N) P_K(N) = kT^2\left(\frac{\partial \ln Q}{\partial T}\right)_{\mu,V}$$

$$Q = \sum_{N,K} \exp(-\beta\varepsilon_K(N) + \mu\beta N).$$

The formulas look very much like those of the canonical ensemble, except for the result expressing the average number of molecules in the container N_{ave} in terms of the derivative of the partition function with respect to the chemical potential μ.

In addition to the equal a priori probability postulate stated earlier (i.e., that, in the thermodynamic limit (i.e., large N), every quantum state of an isolated system in equilibrium having fixed N, V, and E is equally probable), statistical mechanics makes another assumption. It assumes that, in the thermodynamic limit, the ensemble average (e.g., using equal probabilities P_J for all states of an isolated system having specified N, V, and E or using $P_j = \exp(- E_j /kT)/Q$ for states of a system having specified N, V, and T or using $P_K(N) = \dfrac{\exp(-\beta\varepsilon_K(N,V) + \mu\beta N)}{Q}$ for the grand canonical case) of any quantity is equal to the long-time average of this quantity (i.e., the value one would obtain by monitoring the dynamical evolution of this quantity over a very long time). This second postulate implies that the dynamics of an isolated system spends equal amounts of time in every quantum state that has the specified N, V, and E; this is known as the ergodic hypothesis.

Let's consider a bit more what the physical meaning or information content of partition functions is. Canonical ensemble partition functions represent the thermal-

averaged number of quantum states that are accessible to the system at specified values of N, V, and T. This can be seen best by again noting that, in the quantum expression,

$$Q = \Sigma_j \, \Omega_j \, \exp(-E_j \, /kT)$$

the partition function is equal to a sum of the number of quantum states in the jth energy level multiplied by the Boltzmann population factor $\exp(-E_j/kT)$ of that level. So, Q is dimensionless and is a measure of how many states the system can access at temperature T. Another way to think of Q is suggested by rewriting the Helmholtz free energy definition given above as $Q = \exp(-A/kT)$. This identity shows that Q can be viewed as the Boltzmann population, not of a given energy E, but of a specified amount of free energy A.

For the microcanonical ensemble, the probability of occupying each state that has the specified values of N, V, and E is equal

$$P_J = 1/\Omega(N,V,E)$$

where $\Omega(N,V,E)$ is the total number of such states. In the microcanonical ensemble case, $\Omega(N,V,E)$ plays the role that Q plays in the canonical ensemble case; it gives the number of quantum states accessible to the system.

2. Individual-Molecule Partition Functions

Keep in mind that the energy levels E_j and degeneracies Ω_j and $\Omega(N,V,E)$ discussed so far are those of the full N-molecule system. In the special case for which the interactions among the molecules can be neglected (i.e., in the dilute ideal-gas limit) at least as far as expressing the state energies, each of the energies E_j can be written as a sum of the energies of each individual molecule: $E_j = \Sigma_{k=1,N} \, \varepsilon_j(k)$. In such a case, the above partition function Q reduces to a product of individual-molecule partition functions:

$$Q = (N!)^{-1} q^N$$

where the N! factor arises as a degeneracy factor having to do with the permutational indistinguishability of the N molecules (e.g., one must not count both $\varepsilon_j(3) + \varepsilon_k(7)$ with molecule 3 in state j and molecule 7 in state k and $\varepsilon_j(7) + \varepsilon_k(3)$ with molecule 7 in state j and molecule 3 in state k; they are the same state), and q is the partition function of an individual molecule

$$q = \Sigma_l \omega_l \exp(-\varepsilon_l/kT).$$

Here, ε_l is the energy of the lth level of the molecule and ω_l is its degeneracy.

The molecular partition functions q, in turn, can be written as products of translational, rotational, vibrational, and electronic partition functions if the molecular energies ε_l can be approximated as sums of such energies. Of course, these approximations are most appropriate to gas-phase molecules whose vibration and rotation states are being described at the lowest level.

The following equations give explicit expressions for these individual contributions to q in the most usual case of a non-linear polyatomic molecule:

Translational:

$$q_t = (2\pi mkT/h^2)^{3/2}\, V,$$

where m is the mass of the molecule and V is the volume to which its motion is constrained. For molecules constrained to a surface of area A, the corresponding result is $q_t = (2\pi mkT/h^2)^{2/2}\, A$, and for molecules constrained to move along a single axis over a length L, the result is $q_t = (2\pi mkT/h^2)^{1/2}\, L$. The magnitudes these partition functions can be computed, using m in amu, T in Kelvin, and L, A, or V in cm, cm^2 or cm^3, as

$$q_t = (3.28 \times 10^{13}\, mT)^{1/2,2/2,3/2}\, L, A, V.$$

Clearly, the magnitude of q_t depends strongly on the number of dimensions the molecule and move around in. This is a result of the vast differences in translational state densities in 1, 2, and 3 dimensions; recall that we encountered these state-density issues in Chapter 2.

Rotational:

$$q_{rot} = \pi^{1/2}/\sigma \, (8\pi^2 I_A kT/h^2)^{1/2} \, (8\pi^2 I_B kT/h^2)^{1/2} \, (8\pi^2 I_C kT/h^2)^{1/2},$$

where I_A, I_B, and I_C are the three principal moments of inertia of the molecule (i.e., eigenvalues of the moment of inertia tensor). σ is the symmetry number of the molecule defined as the number of ways the molecule can be rotated into a configuration that is indistinguishable from its original configuration. For example, σ is 2 for H_2 or D_2, 1 for HD, 3 for NH_3, and 12 for CH_4. The magnitudes of these partition functions can be computed using bond lengths in Å and masses in amu and T in K, using

$$(8\pi^2 I_A kT/h^2)^{1/2} = 9.75 \times 10^6 (I\,T)^{1/2}$$

Vibrational:

$$q_{vib} = \Pi_{k=1,3N-6} \{\exp(-h\nu_j/2kT)/(1- \exp(-h\nu_j/kT))\},$$

where ν_j is the frequency of the jth harmonic vibration of the molecule, of which there are 3N-6. If one wants to treat the vibrations at a level higher than harmonic, this expression can be modified by replacing the harmonic energies $h\nu_j$ by higher-level expressions.

Electronic:

$$q_e = \Sigma_J \omega_J \exp(-\varepsilon_J/kT),$$

where ε_J and ω_J are the energies and degeneracies of the Jth electronic state; the sum is carried out for those states for which the product $\omega_J \exp(-\varepsilon_J/kT)$ is numerically significant (i.e., levels that any significant thermal population). It is conventional to define the energy of a molecule or ion with respect to that of its atoms. So, the first term in the electronic partition function is usually written as $\omega_e \exp(-D_e/kT)$, where ω_e is the degeneracy of the ground electronic state and D_e is the energy required to dissociate the molecule into its constituent atoms, all in their ground electronic states.

Notice that the magnitude of the translational partition function is much larger than that of the rotational partition function, which, in turn, is larger than that of the vibrational function. Moreover, note that the 3-dimensional translational partition function is larger than the 2-dimensional, which is larger than the 1-dimensional. These orderings are simply reflections of the average number of quantum states that are accessible to the respective degrees of freedom at the temperature T which, in turn, relates to the energy spacings and degeneracies of these states.

The above partition function and thermodynamic equations form the essence of how statistical mechanics provides the tools for connecting molecule-level properties such as energy levels and degeneracies, which ultimately determine the E_j and the Ω_j, to the macroscopic properties such as <E>, <S>, <p>, μ, etc.

If one has a system for which the quantum energy levels are not known, it may be possible to express all of the thermodynamic properties in terms of the classical partition function, if the system could be adequately described by classical dynamics. This partition function is computed by evaluating the following classical phase-space integral (phase space is the collection of coordinates q and conjugate momenta p as we discussed in Chapter 1)

$$Q = h^{-NM} (N!)^{-1} \int \exp(-H(q, p)/kT) \, dq \, dp.$$

In this integral, one integrates over the internal (e.g., bond lengths and angles), orientational, and translational coordinates and momenta of the N molecules. If each molecule has K internal coordinates, 3 translational coordinates, and 3 orientational

coordinates, the total number of such coordinates per molecule is $M = K + 6$. One can then compute all thermodynamic properties of the system using this Q in place of the quantum Q in the equations given above for <E>, <p>, etc.

The classical partition functions discussed above are especially useful when substantial intermolecular interactions are present (and, thus, where knowing the quantum energy levels of the N-molecule system is highly unlikely). In such cases, the classical Hamiltonian is often written in terms of H^0 which contains all of the kinetic energy factors as well as all of the potential energies other than the intermolecular potentials, and the intermolecular potential U, which depends only on a subset of the coordinates: $H = H^0 + U$. For example, let us assume that U depends only on the relative distances between molecules (i.e., on the 3N translational degrees of freedom which we denote r). Denoting all of the remaining coordinates as y, the classical partition function integral can be re-expressed as follows:

$$Q = \{h^{-NM} (N!)^{-1} \int \exp (- H^0(y, p)/kT) \, dy \, dp \; \{\int \exp (-U(r)/kT) \, dr\}.$$

The factor

$$Q_{ideal} = h^{-NM} (N!)^{-1} \int \exp (- H^0(y, p)/kT) \, dy \, dp \; V^N$$

would be the partition function if the Hamiltonian H contained no intermolecular interactions U. The V^N factor arises from the integration over all of the translational coordinates if U(r) is absent. The other factor

$$Q_{inter} = (1/V^N) \{\int \exp (-U(r)/kT) \, dr\}$$

contains all the effects of intermolecular interactions and reduces to unity if the potential U vanishes. If, as the example considered here assumes, U only depends on the positions of the centers of mass of the molecules (i.e., not on molecular orientations or internal geometries), the Q_{ideal} partition function can be written in terms of the molecular translational, rotational, and vibrational partition functions shown earlier:

$$Q_{ideal} = (N!)^{-1} \{(2\pi mkT/h^2)^{3/2} V \pi^{1/2}/\sigma (8\pi^2 I_A kT/h^2)^{1/2} (8\pi^2 I_B kT/h^2)^{1/2} (8\pi^2 I_C kT/h^2)^{1/2}$$

$$\Pi_{k=1,3N-6} \{\exp(-h\nu_j/2kT)/(1- \exp(-h\nu_j/kT))\} \Sigma_J \omega_J \exp(-\varepsilon_J/kT)\}^N .$$

Because all of the equations that relate thermodynamic properties to partition functions contain lnQ, all such properties will decompose into a sum of two parts, one coming from lnQ_{ideal} and one coming from lnQ_{inter}. The latter contains all the effects of the intermolecular interactions. This means that, in this classical mechanics case, all the thermodynamic equations can be written as an ideal component plus a part that arises from the intermolecular forces. Again, the *Statistical Mechanics* text by McQuarrie is a good source for reading more details on these topics.

7.1.3. Equilibrium Constants in Terms of Partition Functions

One of the most important and useful applications of statistical thermodynamics arises in the relation giving the equilibrium constant of a chemical reaction or for a physical transformation (e.g., adsorption of molecules onto a metal surface or sublimation of molecules from a crystal) in terms of molecular partition functions. Specifically, for any chemical or physical equilibrium (e.g., the former could be the HF $\Leftrightarrow$ H^+ + F^- equilibrium; the latter could be $H_2O(l)$ $\Leftrightarrow$ $H_2O(g)$), one can relate the equilibrium constant (expressed in terms of numbers of molecules per unit volume or per unit area, depending on whether species undergo translational motion in 3 or 2 dimensions) in terms of the partition functions of these molecules. For example, in the hypothetical chemical equilibrium A + B $\Leftrightarrow$ C, the equilibrium constant K can be written, if the species can be treated as having negligibly weak intermolecular potentials, as:

$$K = (N_C/V)/[(N_A/V) (N_B/V)] = (q_C/V)/[(q_A/V) (q_B/V)].$$

Here, q_J is the partition function for molecules of type J confined to volume V at temperature T. As another example consider the isomerization reaction involving the

normal (N) and zwitterionic (Z) forms of arginine that were discussed in Chapter 5. Here, the pertinent equilibrium constant would be:

$$K = (N_Z/V)/[(N_N/V)] = (q_Z/V)/[(q_N/V)].$$

So, if one can evaluate the partition functions q for reactant and product molecules in terms of the translational, electronic, vibrational, and rotational energy levels of these species, one can express the equilibrium constant in terms of these molecule-level properties.

Notice that the above equilibrium constant expressions equate ratios of species concentrations (in, numbers of molecules per unit volume) to ratios of corresponding partition functions per unit volume. Because partition functions are a count of the number of quantum states available to the system (i.e., the average density of quantum states), this means that we equate species number densities to quantum state densities when we use the above expressions for the equilibrium constant. In other words, statistical mechanics produces equilibrium constants related to numbers of molecules (i.e., number densities) not molar or molal concentrations.

7.2. Monte Carlo Evaluation of Properties

A tool that has proven extremely powerful in statistical mechanics since computers became fast enough to permit simulations of complex systems is the Monte Carlo (MC) method. This method allows one to evaluate the integrations appearing in the classical partition function described above by generating a sequence of configurations (i.e., locations of all of the molecules in the system as well as of all the internal coordinates of these molecules) and assigning a weighting factor to these configurations. By introducing an especially efficient way to generate configurations that have high weighting, the MC method allows us to simulate extremely complex systems that may contain millions of molecules.

To appreciate why it is useful to have a tool such as MC, let's consider how one might write a computer program to evaluate the classical partition function

$$Q = h^{-NM} (N!)^{-1} \int \exp(-H(q, p)/kT) \, dq \, dp$$

For a system consisting of N Ar atoms in a box of volume V at temperature T. The classical Hamiltonian H(q,p) consists of a sum of kinetic and inter-atomic potential energies

$$H(q,p) = \sum_{i=1}^{N} \frac{p_i^2}{2m} + V(q)$$

The integration over the 3N momentum variables can be carried out analytically and allows Q to be written as

$$Q = \frac{1}{N!} \left(\frac{2\pi mkT}{h^2} \right)^{3N/2} \int \exp\left(\frac{-V(q_1, q_2, \dots q_{3N})}{kT} \right) dq_1 dq_2 \dots dq_{3N} \, .$$

The contribution to Q provided by the integral over the coordinates is often called the configurational partition function

$$Q_{config} = \int \exp\left(\frac{-V(q_1, q_2, \dots q_{3N})}{kT} \right) dq_1 dq_2 \dots dq_{3N}$$

If the density of the N Ar atoms is high, as in a liquid or solid state, the potential V will depend on the 3N coordinates of the Ar atoms in a manner that would not allow substantial further approximations to be made. One would thus be faced with evaluating an integral over 3N spatial coordinates of a function that depends on all of these coordinates. If one were to discretize each of the 3N coordinate axes using say K points along each axis, the numerical evaluation of this integral as a sum over the 3N coordinates would require computational effort scaling as K^{3N}. Even for 10 Ar atoms with each axis having K = 10 points, this is of the order of 10^{30} computer operations. Clearly, such a straightforward evaluation of this classical integral would be foolish to undertake.

The MC procedure allows one to evaluate such high-dimensional integrals by

1. not dividing each of the 3N axes into K discrete points, but rather

2. selecting values of $q_1, q_2, \ldots q_{3N}$ for which the integrand exp(-V/kT) is non-negligible, while also

3. avoiding values of $q_1, q_2, \ldots q_{3N}$ for which the integrand exp(-V/kT) is small enough to neglect.

By then summing over only values of $q_1, q_2, \ldots q_{3N}$ that meet these criteria, the MC process can estimate the integral. Of course, the magic lies in how one designs a rigorous and computationally efficient algorithm for selecting those $q_1, q_2, \ldots q_{3N}$ that meet the criteria.

To illustrate how the MC process works, let us consider carrying out a MC simulation representative of liquid water at some density ρ and temperature T. One begins by placing N water molecules in a box of volume V chosen such that N/V reproduces the specified density. To effect the MC process, we must assume that the total (intramolecular and intermolecular) potential energy V of these N water molecules can be computed for any arrangement of the N molecules within the box and for any values of the internal bond lengths and angles of the water molecules. Notice that, as we showed above when considering the Ar example, V does not include the kinetic energy of the molecules; it is only the potential energy. Often, this energy V is expressed as a sum of intra-molecular bond-stretching and bending contributions, one for each molecule, plus a pair-wise additive intermolecular potential:

$$V = \Sigma_J V(\text{internal})_J + \Sigma_{J,K} V(\text{intermolecular})_{J,K},$$

although the MC process does not require that one employ such a decomposition; the energy V could be computed in other ways, if appropriate. For example, V might be evaluated as the Born-Oppenheimer energy if an ab initio electronic structure calculation on the full N-molecule system were feasible. The MC process does not depend on how V is computed, but, most commonly, it is evaluated as shown above.

7.2.1 Metropolis Monte Carlo

In each step of the MC process, this potential energy V is evaluated for the current positions of the N water molecules. In its most common and straightforward implementation known as the Metropolis Monte-Carlo process, a single water molecule is then chosen at random and one of its internal (bond lengths or angle) or external (position or orientation) coordinates is selected at random. This one coordinate (q) is then altered by a small amount (q → q +δq) and the potential energy V is evaluated at the new configuration (q+δq). The amount δq by which coordinates are varied is usually chosen to make the fraction of MC steps that are accepted (by following the procedure detailed below) approximately 50%. This has been shown to optimize the performance of the MC algorithm.

In implementing the MC process, it is usually important to consider carefully how one defines the coordinates q that will be used to generate the MC steps. For example, in the case of N Ar atoms discussed earlier, it might be acceptable to use the 3N Cartesian coordinates of the N atoms. However, for the water example, it would be very inefficient to employ the 9N Cartesian coordinates of the N water molecules. Displacement of, for example, one of the H atoms along the x-axis while keeping all other coordinates fixed would alter the intramolecular O-H bond energy and the H-O-H bending energy as well as the intermolecular hydrogen bonding energies to neighboring water molecules. The intramolecular energy changes would likely be far in excess of kT unless a very small coordinate change δq were employed. Because it is important to the efficiency of the MC process to make displacements δq that produce ca. 50% acceptance, it is better, for the water case, to make use of coordinates such as the center of mass and orientation coordinates of the water molecules (for which larger displacements produce energy changes within a few kT) and smaller displacements of the O-H stretching and H-O-H bending coordinates (to keep the energy change within a few kT).

Another point to make about how the MC process is often used is that, when the inter-molecular energy is pair wise additive, evaluation of the energy change V(q+δq) – V(q) = δV accompanying the change in q requires computational effort that is proportional to the number N of molecules in the system because only those factors V(intermolecular)$_{J,K}$, with J or K equal to the single molecule that is displaced need be computed. This is why pair wise additive forms for V are often employed.

Let us now return to how the MC process is implemented. If the energy change δV is negative (i.e., if the potential energy is lowered by the coordinate displacement), the change in coordinate δq is allowed to occur and the resulting new configuration is counted among the MC-accepted configurations. On the other hand, if δV is positive, the move from q to q + δq is not simply rejected (to do so would produce an algorithm directed toward finding a minimum on the energy landscape, which is not the goal). Instead, the quantity P = exp(-δV/kT) is used to compute the probability for accepting this energy-increasing move. In particular, a random number between, for example, 0.000 and 1.000 is selected. If the random number is greater than P (expressed in the same decimal format), then the move is rejected. If the random number is less than P, the move is accepted and the new location is included among the set of MC-accepted configurations. Then, new water molecule and its internal or external coordinate are chosen at random and the entire process is repeated.

In this manner, one generates a sequence of MC-accepted moves representing a series of configurations for the system of N water molecules. Sometimes this series of configurations is called a Monte Carlo trajectory, but it is important to realize that there is no dynamics or time information in this series. This set of configurations has been shown to be properly representative of the geometries that the system will experience as it moves around at equilibrium at the specified temperature T (n.b., T is the only way that information about the molecules' kinetic energy enters the MC process), but no time or dynamical attributes are contained in it.

As the series of accepted steps is generated, one can keep track of various geometrical and energetic data for each accepted configuration. For example, one can monitor the distances R among all pairs of oxygen atoms in the water system being discussed and then average this data over all of the accepted steps to generate an oxygen-oxygen radial distribution function g(R) as shown in Fig. 7.3. Alternatively, one might accumulate the intermolecular interaction energies between pairs of water molecules and average this over all accepted configurations to extract the cohesive energy of the liquid water.
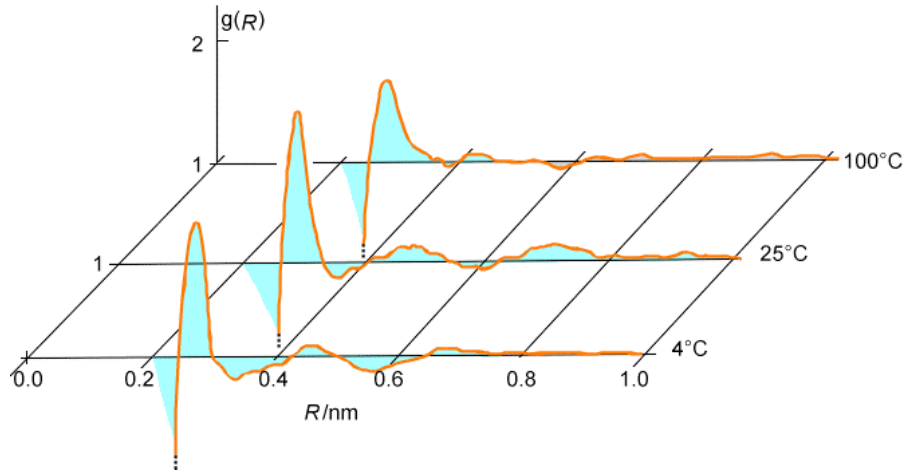
Figure 7.3. Radial distribution functions between pairs of Oxygen atoms in H_2O at three different temperatures.

The MC procedure also allows us to compute the equilibrium average of any property A(q) that depends on the coordinates of the N molecules. Such an average would be written in terms of the normalized coordinate probability distribution function P(q) as:

$$< A >= \int P(q)A(q)dq = \frac{\int \exp(-\beta V(q))A(q)dq}{\int \exp(-\beta V(q))dq}.$$

The denominator in the definition of P(q) is, of course, proportional to the coordinate-contribution to the partition function Q. In the MC process, this average is computed by forming the following sum over the M MC-accepted configurations q_J:

$$< A >= \frac{1}{M}\sum_{J=1}^{M}A(q_J).$$

In most MC simulations, millions of accepted steps contribute to the above averages. At first glance, it may seem that such a large number of steps represent an extreme computational burden. However, recall that straightforward discretization of the 3N axes

produced a result whose effort scaled as K^{3N}, which is unfeasible even for small numbers of molecules

So, why do MC simulations work when the straightforward way fails? That is, how can one handle thousands or millions of coordinates when the above analysis would suggest that performing an integral over so many coordinates would require K^{3N} computations? The main thing to understand is that the K-site discretization of the 3N coordinates is a stupid way to perform the above integral because there are many (in fact, most) coordinate values where the value of the quantity A whose average one wants multiplied by $\exp(-\beta V)$ is negligible. On the other hand, the MC algorithm is designed to select (as accepted steps) those coordinates for which $\exp(-\beta V)$ is non-negligible. So, it avoids configurations that are stupid and focuses on those for which the probability factor is largest. This is why the MC method works!

The standard Metropolis variant of the MC procedure was described above where its rules for accepting or rejecting trial coordinate displacements δq were given. There are several other ways of defining rules for accepting or rejecting trial MC coordinate displacements, some of which involve using information about the forces acting on the coordinates, all of which can be shown to generate a series of MC-accepted configurations consistent with an equilibrium system. The book *Computer Simulations of Liquids*, M. P. Allen and D. J. Tildesley, Oxford U. Press, New York (1997) provides good descriptions of these alternatives to the Metropolis MC method, so I will not go further into these approaches here.

7.2.2 Umbrella Sampling

It turns out that the MC procedure as outlined above is a highly efficient method for computing multidimensional integrals of the form

$$\int P(q)\, A(q)\, dq$$

where P(q) is a normalized (positive) probability distribution and A(q) is any property that depends on the multidimensional variable q.

There are, however, cases where this conventional MC approach needs to be modified by using so-called umbrella sampling. To illustrate how this is done and why it is needed, suppose that one wanted to use the MC process to compute an average, with $\exp(-\beta V(q))$ as the weighting factor, of a function $A(q)$ that is large whenever two or more molecules have high (i.e., repulsive) intermolecular potentials. For example, one could have

$$A(q) = \Sigma_{I<J} \, a/|\mathbf{R}_I - \mathbf{R}_J|^n.$$

Such a function could, for example, be used to monitor when pairs of molecules, with center-of-mass coordinates $\mathbf{R}_J$ and $\mathbf{R}_I$, approach closely enough to undergo a reaction that requires them to surmount a high inter-molecular barrier.

The problem with using conventional MC methods to compute

$$<A> = \int A(q) \, P(q) \, dq$$

in such cases is that

i. $P(q) = \exp(-\beta V(q))/ \int \exp(-\beta V) dq$ favors those coordinates for which the total potential energy V is low. So, coordinates with high $V(q)$ are very infrequently accepted.

ii. However, $A(q)$ is designed to identify events in which pairs of molecules approach closely and thus have high $V(q)$ values.

So, there is a competition between $P(q)$ and $A(q)$ that renders the MC procedure ineffective in such cases because the average one wants to compute involves the product $A(q) \, P(q)$ which is small for most values of q.

What is done to overcome this competition is to introduce a so-called umbrella weighting function $U(q)$ that

i. attains it largest values where $A(q)$ is large, and

ii. is positive and takes on values between 0 and 1 so it can be used as shown below to define a proper probability weighting function.

One then replaces P(q) in the MC algorithm by the product P(q) U(q) and uses this as a weighting function. To see how this replacement works, we re-write the average that needs to be computed as follows:

$$< A >= \int P(q)A(q)dq = \frac{\int \exp(-\beta V(q))A(q)dq}{\int \exp(-\beta V(q))dq}$$

$$= \frac{\dfrac{\int U(q)\exp(-\beta V(q))[A(q)/U(q)]dq}{\int U(q)\exp(-\beta V(q))dq}}{\dfrac{\int U(q)\exp(-\beta V(q))[1/U(q)]dq}{\int U(q)\exp(-\beta V(q))dq}} = \frac{<\dfrac{A}{U}>_{Ue^{-\beta V}}}{<\dfrac{1}{U}>_{Ue^{-\beta V}}}$$

The interpretation of the last identity is that <A> can be computed by

i. using the MC process to evaluate the average of (A(q)/U(q)) but with a probability weighting factor of U(q) exp(-βV(q)) to accept or reject coordinate changes, and

ii. also using the MC process to evaluate the average of (1/U(q)) again with U(q) exp(-βV(q)) as the weighting factor, and finally

iii. taking the average of (A/U) divided by the average of (1/U) to obtain the final result.

The secret to the success of umbrella sampling is that the product U(q) exp(-βV(q)) causes the MC process to emphasize in its acceptance and rejection procedure coordinates for which both exp(-βV) and U (and hence A) are significant. Of course, the tradeoff is that the quantities (A/U and 1/U) whose averages one computes using U(q) exp(-βV(q)) as the MC weighting function are themselves susceptible to being very small at coordinates q where the weighting function is large. Let's consider some examples of when and how one might want to use umbrella sampling techniques.

Suppose one has one system for which the evaluation of the partition function (and thus all thermodynamic properties) can be carried out with reasonable computational effort and another similar system (i.e., one whose potential does not differ much from the first) for which this task is very difficult. Let's call the potential function of the first

system V^0 and that of the second system $V^0 + \Delta V$. The latter system's partition function can be written as follows

$$Q = \sum_J \exp(-\beta(V^0 + \Delta V)) = Q^0 \sum_J \exp(-\beta(V^0 + \Delta V))/Q^0$$

$$= Q^0 < \exp(-\beta\Delta V) >^0$$

where Q^0 is the partition function of the first system and $< \exp(-\beta\Delta V) >^0$ is the ensemble average of the quantity $\exp(-\beta\Delta V)$ taken with respect to the ensemble appropriate to the first system. This result suggests that one can form the ratio of the partition functions (Q/Q^0) by computing the ensemble average of $\exp(-\beta\Delta V)$ using the first system's weighting function in the MC process. Likewise, to compute, for second system, the average value of any property A(q) that depends only on the coordinates of the particles, one can proceed as follows

$$< A >= \frac{\sum_J A_J \exp(-\beta(V^0 + \Delta V))}{Q} = \frac{Q^0}{Q} < A\exp(-\beta\Delta V) >^0$$

where $< A\exp(-\beta\Delta V) >^0$ is the ensemble average of the quantity $A\exp(-\beta\Delta V)$ taken with respect to the ensemble appropriate to the first system. Using the result derived earlier for the ratio (Q/Q^0), this expression for <A> can be rewritten as

$$< A >= \frac{Q^0}{Q} < A\exp(-\beta\Delta V) >^0 = \frac{< A\exp(-\beta\Delta V) >^0}{< \exp(-\beta\Delta V) >^0}.$$

In this form, we are instructed to form the average of A for the second system by
a. forming the ensemble average of $A\exp(-\beta\Delta V)$ using the weighting function for the first system,

b. forming the ensemble average of $\exp(-\beta \Delta V)$ using the weighting function for the first system, and

c. taking the ratio of these two averages.

This is exactly what the umbrella sampling device tells us to do if we were to choose as the umbrella function

$$U = \exp(\beta \Delta V).$$

In this example, the umbrella is related to the difference in the potential energies of the two systems whose relationship we wish to exploit.

Under what circumstances would this kind of approach be useful? Suppose one were interested in performing a MC average of a property for a system whose energy landscape $V(q)$ has many local minima separated by large energy barriers, and suppose it was important to sample configurations characterizing the many local minima in the sampling. A straightforward MC calculation using $\exp(-\beta V)$ as the weighting function would likely fail because a sequence of coordinate displacements from near one local minimum to another local minimum would have very little chance of being accepted in the MC process because the barriers are very high. As a result, the MC average would likely generate configurations representative of only the system's equilibrium existence near one local minimum rather than representative of its exploration of the full energy landscape.

However, if one could identify those regions of coordinate space at which high barriers occur and construct a function ΔV that is large and positive only in those regions, one could then use

$$U = \exp(\beta \Delta V)$$

as the umbrella function and compute averages for the system having potential $V(q)$ in terms of ensemble averages for a modified system whose potential V^0 is

$$V^0 = V - \Delta V .$$

In Fig. 7. 3a, I illustrate how the original and modified potential landscapes differ in regions between two local minima.
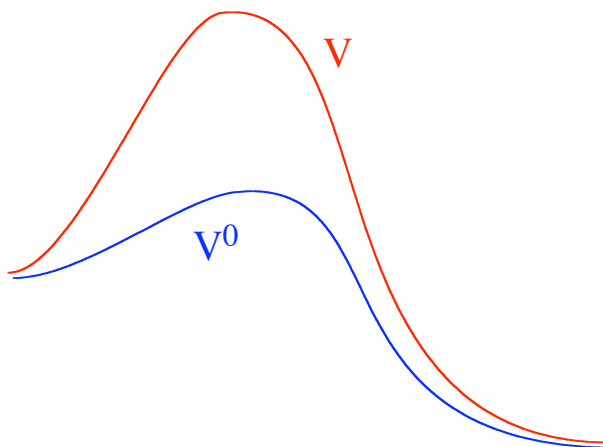

Figure 7. 3 a. Qualitative depiction of the potential V for a system having a large barrier and for the umbrella-modified system with potential $V^0 = V-\Delta V$.

The MC-accepted coordinates generated using the modified potential V^0 would sample the various local minima and thus the entire landscape in a much more efficient manner because they would not be trapped by the large energy barriers. By using these MC-accepted coordinates, one can then estimate the average value of a property A appropriate to the potential V having the large barriers by making use of the identity

$$< A >= \frac{Q^0}{Q} < A \exp(-\beta \Delta V) >^0 = \frac{< A \exp(-\beta \Delta V) >^0}{< \exp(-\beta \Delta V) >^0} .$$

The above umbrella strategy could be useful in generating a good sampling of configurations characteristic of the many local minima, which would be especially beneficial if the quantity A(q) emphasized those configurations. This would be the case, for example, if A(q) measured the intramolecular and nearest-neighbor oxygen-hydrogen interatomic distances in a MC simulation of liquid water. On the other hand, if one wanted to use as A(q) a measure of the energy needed for a Cl^- ion to undergo, in a 1 M

aqueous solution of NaCl, a change in coordination number from 6 to 5 as illustrated in Fig. 7.3 b, one would need a sampling that is accurate both near the local minima corresponding to the 5- and 6-coordinate and the transition-state structures.
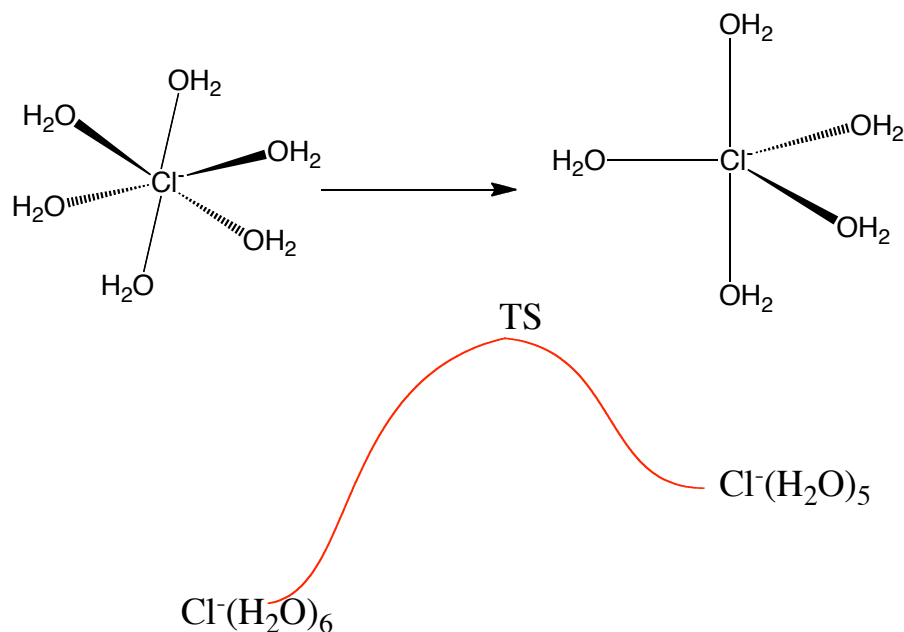

Figure 7.3 b Qualitative depiction of 5- and 6-coordinate Cl⁻ ion in water and of the energy profile connecting these two structures.

Using an umbrella function similar to that discussed earlier to simply lower the barrier connecting the two Cl⁻ ion structures may not be sufficient. Although this would allow one to sample both local minima, its sampling of structures near the transition state would be questionable if the quantity ΔV by which the barrier is lowered (to allow MC steps moving over the barrier to be accepted with non-negligible probability) is large. In such cases, it is wise to employ a series of umbrellas to connect the local minima to the transition states.

Assuming that one has knowledge of the energies and local solvation geometries characterizing the two local minima and the transition state as well as a reasonable guess or approximation of the intrinsic reaction path (refer back to Section 3.3 of Chapter 3)

connecting these structures, one proceeds as follows to generate a series of so-called windows within each of which the free energy A of the solvated Cl$^-$ ion is evaluated.

1. Using the full potential V of the system to constitute the unaltered weighting function $\exp(-\beta V(q))$, one multiplies this by an umbrella function

$$U(q) = \left(\frac{0; \{s_1 - \delta/2 \le s(q) \le s_1 + \delta/2\}}{\infty; otherwise} \right)$$

to form the umbrella-altered weighting function $U(q) \exp(-\beta V(q))$. In $U(q)$, $s(q)$ is the value of the value of the intrinsic reaction coordinate IRC evaluated for the current geometry of the system q, s_1 is the value of the IRC characterizing the first window, and δ is the width of this window. The first window could, for example, correspond to geometries near the 6-coordinate local minimum of the solvated Cl$^-$ ion structure. The width of each window δ should be chosen so that the energy variation within the window is no more than a 1-2 kT; in this way, the MC process will have a good (i.e., ca. 50%) acceptance fraction and the configurations generated will allow for energy fluctuations uphill toward the TS of about this amount.

2. As the MC process is performed using the above $U(q) \exp(-\beta V(q))$ weighting, one constructs a histogram $P_1(s)$ for how often the system reaches various values s along the IRC. Of course, the severe weighting caused by $U(q)$ will not allow the system to realize any value of s outside of the window $s_1 - \delta/2 \le s(q) \le s_1 + \delta/2$.

3. One then creates a second window $s_2 - \delta/2 \le s(q) \le s_2 + \delta/2$ that connects to the first window (i.e., with $s_1 + \delta/2 = s_2 - \delta/2$) and repeats the MC sampling using

$$U(q) = \left(\frac{0; \{s_2 - \delta/2 \le s(q) \le s_2 + \delta/2\}}{\infty; otherwise} \right)$$

to generate a second histogram $P_2(s)$ for how often the system reaches various values of s along the IRC within the second window.

4. This process is repeated at a series of connected windows $s_k - \delta/2 \le s(q) \le s_k + \delta/2$

whose centers s_k range from the 6-coordinate Cl^- ion ($k = 1$), through the transition state ($k = TS$), and to the 5-coordinate Cl^- ion ($k = N$).

After performing this series of N umbrella-altered samplings, one has in hand a series of N histograms $\{P_k(s); k = 1, 2, \ldots TS, \ldots N\}$. Within the kth window, $P_k(s)$ gives the <u>relative</u> probability of the system being at a point s along the IRC. To generate the normalized absolute probability function $P(s)$ expressing the probability of being at a point s, one can proceed as follows:

1. Because the first and second windows are connected at the point $s_1+\delta/2 = s_2 - \delta/2$, one can scale $P_2(s)$ (i.e., multiply it by a constant) to match $P_1(s)$ at this common point to produce a new $P'_2(s)$ function

$$P'_2(s) = P_2(s) \frac{P_1(s_1 + \delta/2)}{P_2(s_2 - \delta/2)}.$$

This new $P'_2(s)$ function describes exactly the same relative probability within the second window, but, unlike $P_2(s)$, it connects smoothly to $P_1(s)$.

2. Because the second and third windows are connected at the point $s_2+\delta/2 = s_3 - \delta/2$, one can scale $P_3(s)$ to match $P'_2(s)$ at this common point to produce a new $P'_3(s)$ function

$$P'_3(s) = P_3(s) \frac{P'_2(s_2 + \delta/2)}{P_3(s_3 - \delta/2)}.$$

3. This process of scaling P_k to match $P'_{k-1}(s)$ at $s_k - \delta/2 = s_{k-1} + \delta/2$ is repeated until the final window connecting $k = N-1$ to $k = N$. Upon completing this series of connections, one has in hand a continuous probability function $P(s)$, which can be normalized

$$P_{normalized} = \frac{P(s)}{\int_{s=0}^{s_{final}} P(s)ds}.$$

In this way, one can compute the probability of accessing the TS, $P_{normalized}(s = TS)$, and the free energy profile

$$A(s) = -kT \ln P_{normalized}(s)$$

at any point along the IRC. It is by using a series of connected windows, within each of which the MC process samples structures whose energies can fluctuate by 1-2 kT, that one generates a smooth connection from low-energy to high-energy (e.g., TS) geometries.

E. Molecular Dynamics Simulations

One thing that the MC process does not address directly is the time evolution of the system. That is, the steps one examines in the MC algorithm are not straightforward to associate with a time-duration, so it is not designed to compute the rates at which events take place. If one is interested in simulating such dynamical processes, even when the N-molecule system is at or near equilibrium, it is more appropriate to carry out a classical molecular dynamics (MD) simulation. In such an MD calculation, one has to assign initial values for each of the internal and external coordinates of each of the N molecules and an initial value of the kinetic energy or momentum for each coordinate, after which a time-propagation algorithm generates values for the coordinates and momenta at later times. For example, the initial coordinates could be chosen close to those of a local minimum on the energy surface and the initial momenta associated with each coordinate could be assigned values chosen from a Maxwell-Boltzmann distribution characteristic of a specified temperature T. In such cases, it is common to then allow the MD trajectory to be propagated for a length of time Δt long enough to allow further equilibration of the energy among all degrees of freedom before extracting any numerical data to use in evaluating average values or creating inter-particle distance histograms, for example.

One usually does not choose just one set of such initial coordinates and momenta to generate a single trajectory. Rather, one creates an ensemble of initial coordinates and momenta designed to represent the experimental conditions the MD calculation is to

simulate. The time evolution of the system for each set of initial conditions is then followed using MD and various outcomes (e.g., reactive events, barrier crossings, folding or unfolding events, chemisorption ocurrences, etc.) are monitored throughout each MD simulation. An average over the ensemble of trajectories is then used in computing averages and creating histograms for the MD simulation. It is the purpose of this Section to describe how MD is used to follow the time evolution for such simulations.

7.3.1 Trajectory Propagation

With each coordinate having its initial velocity $(dq/dt)_0$ and its initial value q_0 specified, one then uses Newton's equations written for a time step of duration δt to propagate q and dq/dt forward in time according, for example , to the following first-order propagation formula:

$$q(t+\delta t) = q_0 + (dq/dt)_0 \, \delta t$$

$$dq/dt \, (t+\delta t) = (dq/dt)_0 - \delta t \, [(\partial V/\partial q)_0/m_q].$$

Here m_q is the mass factor connecting the velocity dq/dt and the momentum p_q conjugate to the coordinate q:

$$p_q = m_q \, dq/dt,$$

and $-(\partial V/\partial q)_0$ is the force along the coordinate q at the earlier geometry q_0. In most modern MD simulations, more sophisticated numerical methods can be used to propagate the coordinates and momenta. For example, the widely used Verlet algorithm is derived as follows.

1. One expands the value of the coordinate q at the $n+1^{st}$ and $n-1^{st}$ time steps in Taylor series in terms of values at the n^{st} time step

$$q_{n+1} = q_n + (dq/dt)_n \delta t + \frac{-(\partial V/\partial q)_n}{2m} \delta t^2 + O(\delta t^3)$$

$$q_{n-1} = q_n - (dq/dt)_n \delta t + \frac{-(\partial V/\partial q)_n}{2m} \delta t^2 - O(\delta t^3)$$

2. One adds these two expansions to obtain

$$q_{n+1} = 2q_n - q_{n-1} + \frac{-(\partial V/\partial q)_n}{m} \delta t^2 + O(\delta t^4)$$

which allows one to compute q_{n+1} in terms of q_n and q_{n-1} and the force at the nth step, while not requiring knowledge of velocities.

3. If the two Taylor expansions are subtracted, one obtains

$$(dq/dt)_{n+1} - \frac{q_{n+1} - q_{n-1}}{2\delta t} + O(\delta t^2)$$

as the expression for the velocity at the $n+1^{st}$ time step in terms of the coordinates at the $n+1^{st}$ and $n-1^{st}$ steps.

There are many other such propagation schemes that can be used in MD; each has strengths and weaknesses. In the present Section, I will focus on describing the basic idea of how MD simulations are performed while leaving treatment of details about propagation schemes to more advanced sources such as *Computer Simulations of Liquids*, M. P. Allen and D. J. Tildesley, Oxford U. Press, New York (1997).

The forces $-(\partial V/\partial q)$ appearing in the MD propagation algorithms can be obtained as gradients of a Born-Oppenheimer electronic energy surface if this is computationally feasible. Following this path involves performing what is called direct-dynamics MD. Alternatively, the forces can be computed from derivatives of an empirical force field. In the latter case, the system's potential energy V is expressed in terms of analytical functions of

i. intramolecular bond lengths, bond angles, and torsional angles, as well as

ii. intermolecular distances and orientations.

The parameters appearing in such force fields have usually been determined from electronic structure calculations on molecular fragments, spectroscopic determination of vibrational force constants, and experimental measurements of intermolecular forces.

7.3.2 Force Fields

Let's interrupt our discussion of MD propagation of coordinates and velocities to examine the ingredients that usually appear in the force fields mentioned above. In Fig. 7.3 c, we see a molecule in which various intramolecular and intermolecular interactions are introduced.
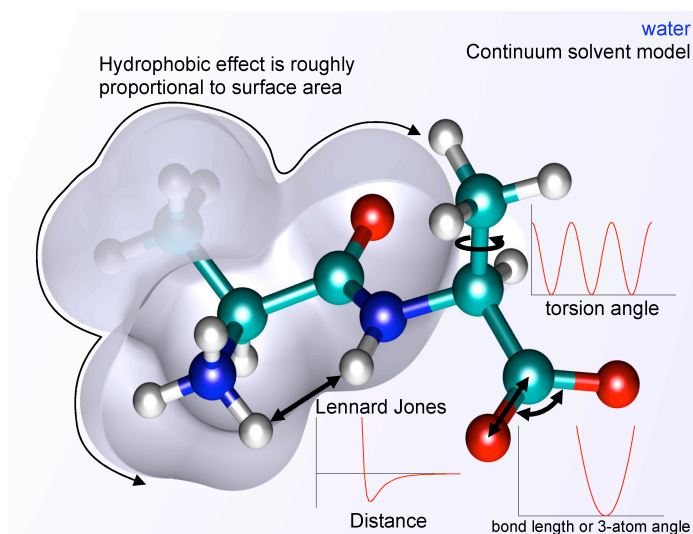
Figure 7. 3 c. Depiction of a molecule in which bond-stretching, bond-bending, intramolecular van der Waals, and intermolecular solvation potentials are illustrated.

The total potential of a system containing one or more such molecules in the presence of a solvent (e.g., water) it typically written as a sum of intramolecular potentials (one for each molecule in the system) and itermolecular potentials. The former are usually decomposed into a sum of covalent interactions describing how the energy varies with bond stretching, bond bending, and dihedral angle distortion as depicted in Fig. 7.3 d.

$$U = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2$$

periodicity
V2: 2 barriers *cis/trans*
V3: 3 barriers   -C-C-
Usually only 1, 2 and 3
(some force fields up to 6)

peak height

$$\sum_{dihedrals} \sum_{\eta} \frac{V_\eta}{2} [1 + \cos(\eta \phi - \gamma)]$$

phase
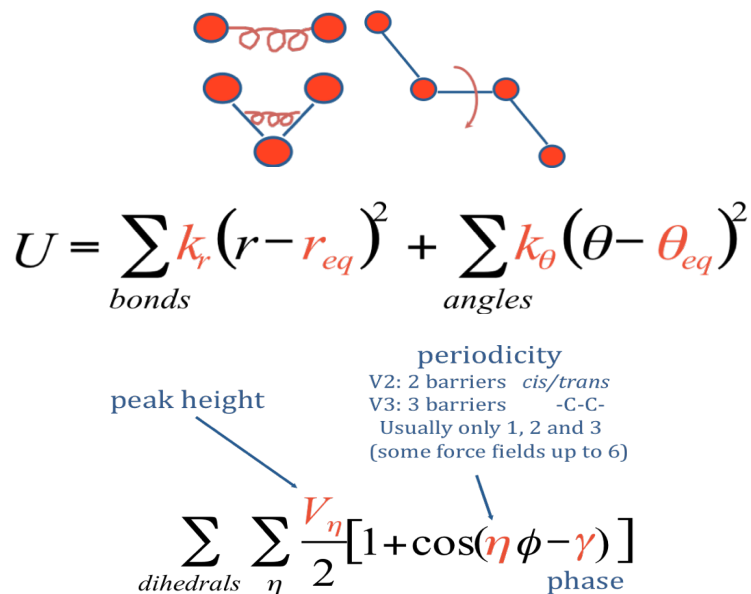
Figure 7. 3 d. Depiction of bond stretching and bending (top left) and dihedral angle distortion (top right) within a molecule and equations describing how the energy varies with these geometry changes.

and non-covalent interactions describing electrostatic and van der Waals interactions among the atoms in the molecule as

$$V_{noncovalent} = \sum_{i<j}^{atoms} \left\{ \frac{A_{i,j}}{r_{i,j}^{12}} - \frac{B_{i,j}}{r_{i,j}^{6}} + \frac{q_i q_j}{\varepsilon r_{i,j}} \right\} .$$

These functional forms would be used to describe how the energy V(q) changes with the bond lengths (r) and angles (θ, ϕ) within, for example, each of the molecules shown in Fig. 7. 3 c (let's call them solute molecules) as well as for any water molecules that may be present (if these molecules are explicitly included in the MD simulation).

The interactions among the solute and solvent moleulues are also often expressed in a form involving electrostatic and van der Waals interations between pairs of atoms-one on one molecule (solute or solvent) and the other on another molecule (solute or solvent).

$$V_{\text{intermolecular}} = \sum_{i<j}^{atoms} \{\frac{A_{i,j}}{r_{i,j}^{12}} - \frac{B_{i,j}}{r_{i,j}^{6}} + \frac{q_iq_j}{\varepsilon r_{i,j}}\}.$$

The Cartesian forces on any atom within a solute or solvent molecule are then computed for use in the MD simulation by using the chain rule to relate derivatives with respect to Cartesian coordinates to derivatives of the above intramolecular and intermolecular potentials with respect to the interatomic distances and the angles appearing in them.

Because water is such a ubiquitous component in condensed-phase chemistry, much effort has been devoted to generating highly accurate intermolecular potentials to describe the interactions among water molecules. In the popular TIP3P and TIP4P models, the water-water interaction is given by

$$V = \frac{A}{r_{OO}^{12}} - \frac{B}{r_{OO}^{6}} + \sum_{i,j} \frac{kq_iq_j}{r_{i,j}}$$

where r_{OO} is the distance between the oxygen atoms of the two water molecules in Å, and indices i and j run over 3 or 4 sites, respectively, for TIP3P or TIP4P, with i labeling sites on one water molecule and j labeling sites on the second water molecule. The parameter k is 332.1 Å kcal mol[-1]. A and B are conventional Lennard-Jones parameters for oxygen atoms and q_i is the magnitude of the partial charge on the ith site. In Fig. 7.3 d, we show how the 3 or 4 sites are defined for these two models.
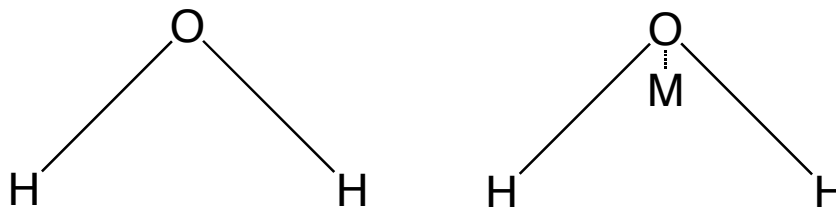

Figure 7.3 d Location of the 3 or 4 sites used in the TIP3P and TIP4P models.

Typical values for the parameters are given in the table below.

	r_{OH} (Å)	HOH angle degrees	r_{OM} (Å)	A ($Å^{12}$kcal/mol)	B ($Å^6$kcal/mol)	q_O or q_M	q_H
TIP3P	0.9572	104.52		582×10^3	595	-0.834	0.417
TIP4P	0.9672	104.52	0.15	600×10^3	610	-1.04	0.52

In the TIP3P model, the three sites reside on the oxygen and two hydrogen centers. For TIP4P, the fourth site is called the M-site and it resides off the oxygen center a distance of 0.15 along the bisector of the two O-H bonds as shown in Fig. 7.3 d. In using either the TIP3P or TIP4P model, the intramolecular bond lengths and angles are often constrained to remain fixed; when doing so, one is said to be using a rigid water model.

There are variants to these two 3-site and 4-site models that, for example, include van der Waals interactions between H atoms on different water molecules, and there are models including more than 4 sites, and models that allow for the polarization of each water molecule induced by the dipole fields (as represented by the partial charges) of the other water molecules and of solute molecules. The more detail and complexity one introduces, the more computational effort is needed to perform MD simulations. In particular, water molecules that allow for polarization are considerably more computationally demanding because they often involve solving self-consistently for the polarization of each molecule by the charge and dipole potentials of all the other molecules, with each dipole potential including both the permanent and induced dipoles of that molecule. Professor John Wampler has created a web page (http://www.bmb.uga.edu/wampler/399/lectures/mm1/index.htm) in which the details about molecular mechanics force fields introduced above are summarized. The web page (http://en.wikipedia.org/wiki/Molecular_mechanics) provides links to numerous software packages that use these kinds of force fields to carry out MD simulations. These links also offer more detailed information about the performance of various force fields as well as giving values for the parameters used in those force fields.

The parameter values are usually obtained by

a. fitting the intramolecular or intermolecular functional form (e.g., as shown above) to energies obtained in electronic structure calculations at a large number of geometries, or

b. adjusting them to cause MD or MC simulations employing the force field to reproduce certain thermodynamic properties (e.g., radial distribution functions, solvation energies, vaporization energies, diffusion constants), or some combination of both. It is important to observe that the kind of force fields discussed above have limitations beyond issues of accuracy. In particular, they are not designed to allow for bond breaking and bond forming, and they represent the Born-Oppenheimer energy of one (most often the ground) electronic state. There are force fields explicitly designed to include chemical bonding changes, but most MD packages do not include them. When one is interested in treating a problem that involves transitions from one electronic state to another (e.g., in spectroscopy or when the system undergoes a surface hop near a conical intersection), it is most common to use a combined QM-MM approach like we talked about in Section 6.1.3 of Chapter 6. A QM treatment of the portion of the system that undergoes the electronic transition is combined with a force-field (MM) treatment of the rest of the system to carry out the MD simulation. Let's now return to the issue of propagating trajectories given a force field and a set of initial conditions appropriate to describing the system to be simulated.

By applying one of the time-propagation algorithms to all of the coordinates and momenta of the N molecules at time t, one generates a set of new coordinates $q(t+\delta t)$ and new velocities $dq/dt(t+\delta t)$ appropriate to the system at time $t+\delta t$. Using these new coordinates and momenta as q_0 and $(dq/dt)_0$ and evaluating the forces $-(\partial V/\partial q)_0$ at these new coordinates, one can again use the propagation equations to generate another finite-time-step set of new coordinates and velocities. Through the sequential application of this process, one generates a sequence of coordinates and velocities that simulate the system's behavior. By following these coordinates and momenta, one can interrogate any dynamical properties that one is interested in. For example, one could monitor oxygen-oxygen distances throughout an MD simulation of liquid water with initial conditions chosen to represent water at a given temperature (T would determine the initial momenta) to generate a histogram of O-O distances. This would allow one to construct the kind of radial distribution function shown in Fig. 7. 3 using MD simulation rather than MC. The radial distribution function obtained in such an MD simulation should be identical to that obtained from MC because statistical mechanics assumes the ensemble average (MC) is

equal to the long-time average (MD) of any property for a system at equilibrium. Of course, one could also monitor quantities that depend on time, such as how often two oxygen atoms come within a certain distance, throughout the MD simulation. This kind of interrogation could not be achieved using MC because there is no sense of time in MC simulations.

In Chapter 8, I again discuss using classical molecular dynamics to follow the time evolution of a chemical system. However, there is a fundamental difference between the kind of simulations described above and the case I treat in Chapter 8. In the former, one allows the N-molecule system to reach equilibrium (i.e., either by carefully choosing initial coordinates and momenta or by waiting until the dynamics has randomized the energy) before monitoring the subsequent time evolution. In the problem discussed in Chapter 8, we use MD to follow the time progress of a system representing a single bimolecular collision in two crossed beams of molecules. Each such beam contains molecules whose initial translational velocities are narrowly defined rather than Maxwell-Boltzmann distributed. In this case, we do not allow the system to equilibrate because we are not trying to model an equilibrium system. Instead, we select an ensemble of initial conditions that represent the molecules in the two beams and we then follow the Newton dynamics to monitor the outcome (e.g., reaction or non-reactive collision).

Unlike the MC method, which is very amenable to parallel computation, MD simulations are more difficult to carry out in a parallel manner. One can certainly execute many different classical trajectories on many different computer nodes; however, to distribute one trajectory over many nodes is difficult. The primary difficulty is that, for each time step, all N of the molecules undergo moves to new coordinates and momenta. To compute the forces on all N molecules requires of the order of N^2 calculations (e.g., when pairwise additive potentials are used). In contrast, each MC step requires that one evaluate the potential energy change accompanying the displacement of only one molecule. This uses only of the order of N computational steps (again, for pair wise additive potentials).

Another factor that complicates MD simulations has to do with the wide range of times scales that may be involved. For example, for one to use a time step δt short enough to follow high-frequency motions (e.g., O-H stretching) in a simulation of an ion

or polymer in water solvent, δt must be of the order of 10^{-15} s. To then simulate the diffusion of an ion or the folding of a polymer in the liquid state, which might require 10^{-4} s or longer, one would have to carry out 10^{11} MD steps. This likely would render the simulation not feasible. In the table below we illustrate the wide range of time scales that characterize various events that one might want to simulate using some form of MD, and we give a sense of what is practical using MD simulations in the year 2010.

Examples of dynamical processes taking place over timescales ranging from 10^{-15} s through hundreds of seconds, each of which one may wish to simulate using MD.

10^{-15}-10^{-14} s	10^{-12} s	10^{-9} s	10^{-6} s	10^{-3} s	110 s
C-H, N-H, O-H bond vibration	Rotation of small molecule	Routinely accessible time duration for atomistic MD simulation	Time duration for heroic atomistic MD simulation	Time duration achievable using coarse-graining techniques[a]	Time needed for protein folding

a. These techniques are discussed in Section 7.3.3.

Because one can not afford to carry out simulations covering 10^{-3}-100 s using time steps needed to follow bond vibrations 10^{-15} s, it is necessary to devise strategies to focus on motions whose time frame is of primary interest while ignoring or approximating faster motions. For example, when carrying out long-time MD simulations, one can ignore the high-frequency intramolecular motions by simply not including these coordinates and momenta in the Netwonian dynamics (e.g., as one does when using a rigid-water model discussed earlier). In other words, one simply freezes certain bond lengths and angles. Of course, this is an approximation whose consequences must be tested and justified, and would certainly not be a wise step to take if those coordinates played a key role in the dynamical process being simulated. Another approach, called coarse graining involves replacing the fully atomistic description of

selected components of the system by a much-simplified description involving significantly fewer spatial coordinates and momenta.

7.3.3 Coarse Graining

The goal of coarse graining is to bring the computational cost of a simulation into the realm of reality. This is done by replacing the fully atomistic description of the system, in which coordinates sufficient to specify the positions (and, in MD, the velocities) of every atom, by a description in terms of fewer functional groups often referred to as "beads". The TIP4P and TIP3P models for the water-water interaction potential discussed above are not coarse-grained models because they contain as many (or more) centers as atoms. An example of a coarse-grained model for the water-water interaction is provided by the Stillinger-Weber model (that was originally introduced to treat tetrahedral Si) of water introduced in V. Molinero and E. B. Moore, J. Phys. Chem. B 2009, 113, 4008–4016. Here, each water molecule is described only by the location of its oxygen nucleus (labeled $\mathbf{r}_i$ for the ith water molecule), and the interaction potential is given as a sum of two-body and three-body terms

$$V = \sum_{i < j = 1}^{N} A\varepsilon\{B\left(\frac{\sigma}{r_{i,j}}\right)^p - \left(\frac{\sigma}{r_{i,j}}\right)^q\}\exp(\frac{\sigma}{r_{i,j} - a\sigma})$$

$$+ \sum_{i < j < k = 1}^{N} \lambda\varepsilon[\cos\theta_{i,j,k} - \cos\theta_0]^2 \exp(\frac{\gamma\sigma}{r_{i,j} - a\sigma})\exp(\frac{\gamma\sigma}{r_{i,k} - a\sigma})$$

where $r_{i,j}$ is the distance between the ith and j^{th} oxygen atom, $\theta_0 = 109.47$ deg, and $\theta_{i,j,k}$ is the angle between the ith (at the center), j^{th}, and k^{th} oxygen atom. The parameters A, B, ε, σ, and a are used to characterize various characteristics of the potential; different values are needed to describe the behavior of Si, Ge, diamond, or water even though they all can adopt tetrahedral coordination. The form of the three-body part of this potential is designed to guide the orientations among oxygen atoms to adopt tetrahedral character.

Although the above potential seems more complicated than, for example, the form used in the TIP3P or TIP4P potential, it has three important advantages when it comes to carrying out MD simulations:

1. Because the SW potential contains no terms varying with distance as r^{-1} (i.e., no Coulomb interactions among partial charges), it is of qualitatively shorter range than the other two potentials. This allows spatial cut-offs to be used (i.e., to ignore interactions beyond much shorter distances) efficiently.

2. For a system containing N water molecules, the TIP3P or TIP4P models require one to evaluate functions of the distances between $(3N)^2/2$ or $(4N)^2/2$ centers, whereas the SW's two-body component involves only $N^2/2$ interactions and the three-body component need only be evaluated for molecules j and k that are nearby molecule i.

3. If, for the atomistic models, one wishes to treat the O-H stretching and H-O-H bending motions, MD time steps of ca. 10^{-15} s must be employed. For the SW model, the fastest motions involve relative movements of the oxygen centers, which occur on time scales ca. 10 times longer. This means that one can use longer MD steps.

The net result is that this coarse-grained model of the water-water interaction allows MD simulations to be carried out for qualitatively longer time durations. Of course, this is only an advantage if the simulations provide accurate results. In the Table shown below (taken from the above reference), we see MD simulation results (as well as experimental results) obtained with the above (mW) model, with various TIPnP models, and with two other popular water-water potentials (SPC and SPCE) from which it is clear that the coarse-grained mW model is capable of yielding reliable results on a range of thermodynamic properties.

TABLE 1: Comparison of Water Models and Experiment[a]

	$T_{n\,HEX,\,ICE}$ (K)	ΔH_m (T_m) (kcal·mol⁻¹)	ρ_{liquid} (T_m) (g·cm⁻³)	ρ_{ice} (T_m) (g·cm⁻³)	ρ_{liquid}(298 K) (g·cm⁻³)	ΔH_{vap} (298 K) (kcal·mol⁻¹)	D (298 K) (10⁻⁵ cm²·s⁻¹)	γ_{LV} (300 K) mJ·m⁻²	TMD (K)	$\rho_{liquid,\,MAX}$ (TMD) (g·cm⁻³)
exp	273.15	1.436	0.999	0.917	0.997	10.52	2.3	71.6	277	0.99997
mW	**274.6**	**1.26**	**1.001**	0.978	**0.997**	10.65	6.5	**66.0**	250	**1.003**
SPC	(191)	0.62	0.991	**0.934**	0.977	**10.56**	4.0	53.4	228	1.008
SPCE	(215)	0.74	1.007	0.950	0.999	10.76	**2.4**	61.3	241	1.012
TIP3P	(146)	0.30	1.017	0.947	0.986	10.17	5.3	49.5	182	1.038
TIP4P	232	1.05	1.002	0.940	1.001	10.65	3.9	54.7	253	1.008
TIP5P	(274)	1.75	0.987	0.982	0.999	10.46	2.6	52.3	**285**	0.989

[a] Melting temperatures of hexagonal ice, densities of liquid, and crystal phase at coexistence and enthalpy of melting are from ref 71. Parentheses enclosing a T_n signal that the stable crystal is ice II, not hexagonal ice, for these models.[72] Diffusion coefficients D and density at 298 K are from refs 73 and 74. Liquid-vacuum surface tensions are from ref 75. TMD and its corresponding liquid density $\rho_{liquid,MAX}$ are from ref 76. Bold numbers signal the closest agreement with the experiment.

Another example of taking a coarse-grained approach to greatly simplify MD simulations is provided in the work T. A. Knotts, IV, N. Rathore, D. C. Schwartz, and J. J. de Pablo, J. Chem. Phys. 126, 084901 (2007). In Fig. 7.3 e, we see a coarse-grained representation of the DNA double helix (taken from this reference) as well as a depiction of how the beads are defined in terms of base, sugar, and phosphate units.
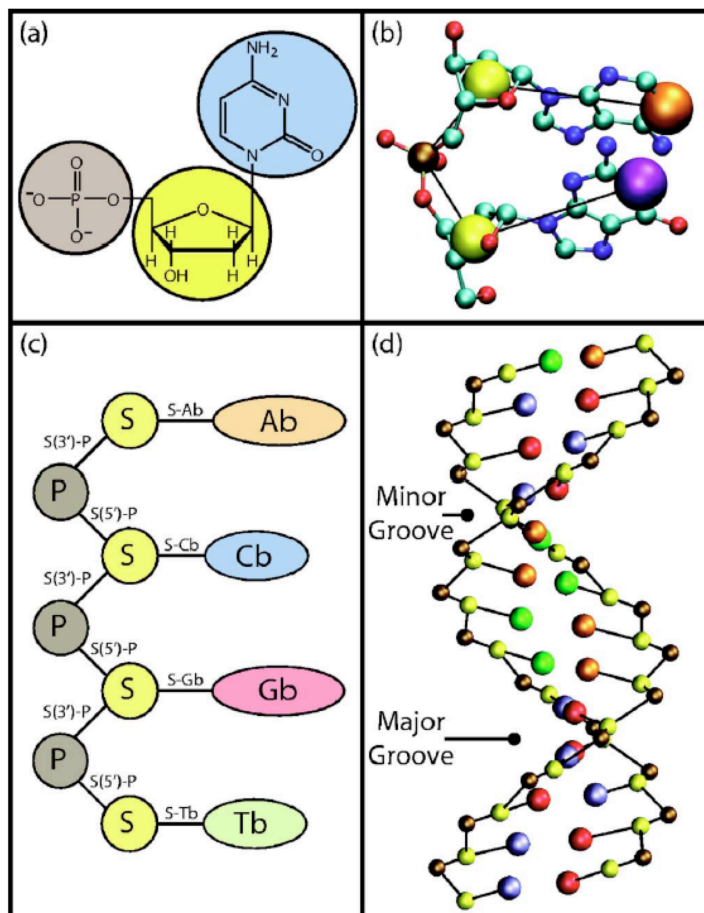
Figure 7.3 e Depiction of cytosine base, sugar, and phosphate units constituting blue, yellow, and brown beads, respectively (a); bead description of the double helix (d); locations of the beads relative to the atomic positions for the phosphate, sugar, and bases; and definition of various bead-bead interaction distances (c).

In the Table shown below, the reference cited above specifies the locations and masses of the phosphate, sugar, and base beads in the B form of the DNA helix. The masses need to be chosen so that the coarse-grained dynamical motions of these units replicate within reasonable tolerances the center of mass motions of the phosphate, sugar, and base moieties when atomistic MD simulations are carried out on smaller test systems containing these nucleotide units.

TABLE I. Cartesian and polar coordinates and masses for sites in the DNA model. For an atom at (x,y,z), the corresponding atom on the other strand is at $(x,-y,-z)$. Successive residues on each strand are placed using the screw symmetry for the B form of DNA (10 residues per turn and an axial rise per residue of 3.38 Å); thus, an atom at (r,ϕ,z) places the next at $(r,\phi+36°, z+3.38$ Å$)$. This scheme produces a molecule centered upon and aligned with the z axis. See Ref. 41 for more information.

| Site | x | y | z | r | ϕ | Mass |
(abbreviation)		(Å)			(degree)	(amu)
Phosphate (P)	−0.628	8.896	2.186	8.918	94.038	94.97
Sugar (S)	2.365	6.568	1.280	6.981	70.197	83.11
Adenine base (Ab)	0.575	0.516	0.051	0.773	41.905	134.1
Thymine base (Tb)	0.159	2.344	0.191	2.349	86.119	125.1
Cytosine base (Cb)	0.199	2.287	0.187	2.296	85.027	110.1
Guanine base (Gb)	0.628	0.540	0.053	0.828	40.691	150.1

The potential V used to carry out the coarse-grained MD simulations is given by the equations shown below taken from the above reference. In addition to the usual bond stretching, bending and dihedral terms (n.b., now the bonds relate to linkages between beads rather than between atoms) that are similar to what we saw earlier in our discussion of force fields, there are additional terms.

1. V_{stack} describes the interactions among π-stacked base pairs,

2. V_{bp} describes the hydrogen bonding interactions between bases, and

3. V_{ex} describes excluded-volume effects.

4. V_{qq} is the screened Coulombic interactions among phosphate units, with its exponential decay constant κ_D given in terms of a so-called Debye screening length as detailed in the above reference.

$$V_{\text{total}} = V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihedral}} + V_{\text{stack}} + V_{\text{bp}} + V_{\text{ex}} + V_{\text{qq}},$$

(1)

where

$$V_{\text{bond}} = \sum_{i}^{N_{\text{bond}}} [k_1(d_i - d_{0_i})^2 + k_2(d_i - d_{0_i})^4],$$

(2a)

$$V_{\text{angle}} = \sum_{i}^{N_{\text{angle}}} \frac{k_\theta}{2}(\theta_i - \theta_{0_i})^2,$$

(2b)

$$V_{\text{dihedral}} = \sum_{i}^{N_{\text{dihedral}}} k_\phi[1 - \cos(\phi_i - \phi_{0_i})],$$

(2c)

$$V_{\text{stack}} = \sum_{i<j}^{N_{\text{st}}} 4\varepsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right],$$

(2d)

$$V_{\text{bp}} = \sum_{\text{base pairs}}^{N_{\text{bp}}} 4\varepsilon_{\text{bp}_i} \left[5\left(\frac{\sigma_{\text{bp}_i}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{\text{bp}_i}}{r_{ij}}\right)^{10} \right],$$

(2e)

$$V_{\text{ex}} = \sum_{i<j}^{N_{\text{ex}}} \begin{cases} 4\varepsilon \left[\left(\frac{\sigma_0}{r_{ij}}\right)^{12} - \left(\frac{\sigma_0}{r_{ij}}\right)^{6} \right] + \varepsilon & \text{if } r_{ij} < d_{\text{cut}} \\ 0 & \text{if } r_{ij} \geq d_{\text{cut}} \end{cases},$$

(2f)

$$V_{\text{qq}} = \sum_{i<j}^{N} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_k r_{ij}} e^{-r_{ij}/\kappa_D}.$$

(2g)

The values of the parameters used in this force field potential given in the above reference are reproduced in the two Tables shown below.

TABLE II. Values for energy parameters found in the potential energy function.

Parameter	Value
k_1	ε
k_2	100ε
k_θ	$400\varepsilon/(\text{radian})^2$
k_φ	4ε
ε	0.26 kcal/mol
$\varepsilon_{b_{P_{GC}}}$	4ε
$\varepsilon_{b_{P_{AT}}}$	$\frac{2}{3}\varepsilon_{b_{P_{GC}}}$

TABLE III. Values for geometric parameters found in the potential energy function. A phosphate can bind to a sugar in either a 5′ or a 3′ sense. [See panel (c) of Fig. 1.] Thus, S(5′)-P represents a bond between a phosphate and a sugar belonging to the same nucleotide while S(3′)-P joins together neighboring residues. The bond angle P-(5′)S(3′)-P consists of both types of bonds. [Note: S(5′)−P=P−(5′)S.]

Bond	d_0 (Å)	Bond angle	θ_0 (degree)
S(5′)-P	3.899	S(5′)−P−(3′)S	94.49
S(3′)−P	3.559	P−(5′)S(3′)−P	120.15
S-Ab	6.430	P−(5′)S−Ab	113.13
S-Tb	4.880	P−(3′)S−Ab	108.38
S-Cb	4.921	P−(5′)S−Tb	102.79
S-Gb	6.392	P−(3′)S−Tb	112.72
		P−(5′)S−Cb	103.49
		P−(3′)S−Cb	112.39
		P−(5′)S−Gb	113.52
		P−(3′)S−Gb	108.12

Dihedral angle	ϕ_0 (degree)	Nonbonded	Length (Å)
P−(5′)S(3′)−P−(5′)S	−154.80	σ_{ij}	Interaction specific
S(3′)−P−(5′)S(3′)−P	−179.17	$\sigma_{b_{P_{AT}}}$	2.9002
Ab−S(3′)−P−(5′)S	−22.60	$\sigma_{b_{P_{GC}}}$	2.8694
S(3′)−P−(5′)S−Ab	50.69	σ_0 (mismatched bases)	$2^{-1/6}(1.0)$
Tb−S(3′)−P−(5′)S	−33.42	σ_0 (otherwise)	$2^{-1/6}d_{\text{cut}}$
S(3′)−P−(5′)S−Tb	54.69		
Cb−S(3′)−P−(5′)S	−32.72	d_{cut}	$\langle\sigma_{ij}\rangle \approx 6.86$
S(3′)−P−(5′)S−Cb	54.50		
Gb−S(3′)−P−(5′)S	−22.30		

Although there are numerous parameters in this potential, the key to the success of this coarse graining is that there are only six kinds of sites whose positions and velocities must be propagated in the MD simulation- phosphate sites, sugar sites, and four kinds of base sites. This is far fewer coordinates that would arise in a fully atomistic MD simulation. I will refer the reader to the reference cited above for details about how successful coarse graining was in this case, but I will not go further into it at this time. I think the two examples we discussed in this Section suffice for introducing the subject of coarse graining to the readers of this text.

In summary for this Section, MD classical simulations are not difficult to implement if one has available a proper representation of the intramolecular and intermolecular potential energy V. Such calculations are routinely carried out on large bio-molecules or condensed-media systems containing thousands to millions of atomic centers. There are, however, difficulties primarily connected to the time scales over which molecular motions and over which the process being simulated change that limit the success of this method and which often require one to employ reduced representations of the system such as in coarse graining. In contrast, quantum MD simulations such as we describe in the following Section are considerably more difficult to carry out.

7.4 Time Correlation Functions

One of the most active research areas in statistical mechanics involves the evaluation of so-called equilibrium time correlation functions such as we encountered in Chapter 6. The correlation function C(t) is defined in terms of two physical operators A and B, a time dependence that is carried by a Hamiltonian H via $\exp(-iHt/\hbar)$, and an equilibrium average over a Boltzmann population $\exp(-\beta H)/Q$.

The quantum mechanical expression for C(t) is

$$C(t) = \Sigma_j <\Phi_j | A \exp(iHt/\hbar) B \exp(-iHt/\hbar) |\Phi_j > \exp(-\beta E_j)/Q,$$

while the classical mechanical expression (here, we allow the h^{-M} factor that occurs in the partition function shown in Section 7.1.2 to be cancelled out in the numerator and denominator for simplicity) is

$$C(t) = \int dq(0) \int dp(0) \ A(q(0),p(0)) \ B(q(t),p(t)) \ \exp(-\beta H(q(0),p(0)))/Q,$$

where q(0) and p(0) are the values of all the coordinates and momenta of the system at t=0 and q(t) and p(t) are their values, according to Newtonian mechanics, at time t. As shown above, an example of a time correlation function that relates to molecular spectroscopy is the dipole-dipole correlation function that we discussed in Chapter 6:

$$C(t) = \Sigma_j \ <\Phi_j \ | \ \mathbf{e} \bullet \boldsymbol{\mu} \ \exp(iHt/\hbar) \ \mathbf{e} \bullet \boldsymbol{\mu} \ \exp(-iHt/\hbar) \ |\Phi_j > \exp(-\beta E_j)/Q,$$

for which A and B are both the electric dipole interaction $\mathbf{e} \bullet \boldsymbol{\mu}$ between the photon's electric field whose direction is characterized by the vector $\mathbf{e}$ and the molecule's dipole operator $\boldsymbol{\mu}$. The Fourier transform of this particular C(t) relates to the absorption intensity for light of frequency ω:

$$I(\omega) = \int dt \ C(t) \ \exp(i\omega t).$$

It turns out that many physical properties (e.g., absorption line shapes, Raman scattering intensities) and transport coefficients (e.g., diffusion coefficients, viscosity) can be expressed in terms of time-correlation functions. It is beyond the scope of this text to go much further in this direction, so I will limit my discussion to the optical spectroscopy case at hand, which requires that we now discuss how the time-evolution aspect of this problem is dealt with. The text *Statistical Mechanics*, D. A. McQuarrie, Harper and Row, New York (1977) has a nice treatment of such other correlation functions, so the reader is directed to that text for further details.

The computation of correlation functions involves propagating either wave functions or classical trajectories which produce the q(t), p(t) values entering into the expression for C(t). In the classical case, one carries out a large number of Newtonian trajectories with initial coordinates q(0) and momenta p(0) chosen to represent the equilibrium condition of the N-molecule system. For example, one could use the MC method to select these variables employing $\exp(-\beta H(p(0),q(0)))$ as the probability

function for accepting or rejecting initial q(0) and p(0) values. In this case, the weighting function contains not just the potential energy but also the kinetic energy (and thus the total Hamiltonian H) because now we need to also select proper initial values for the momenta. So, with many (e.g., M) selections of the initial q and p variables of the N-molecules being made, one would allow the Newton dynamics of each set of initial conditions to proceed. During each such trajectory, one would monitor the initial value of the A(q(0), p(0)) property and the time progress of the B(q(t),p(t)) property. One would then compute the MC average to obtain the correlation function:

$$C(t) = (1/M) \Sigma_{J=1,M} A(q_J(0),p_J(0)) B(q_J(t),p_J(t)) \exp(-\beta H(q_J(0),p_J(0))).$$

Where the index J labels the M accepted configurations and momenta of the MC sampling.

In the quantum case, the time propagation is especially challenging and is somewhat beyond the scope of this text. However, I want to give you some idea of the steps that are involved, realizing that this remains an area of very active research development. As noted in Section 1.3.6, it is possible to time-propagate a wave function Φ that is known at $t = 0$ if one is able to expand Φ in terms of the eigenfunctions of the Hamiltonian H. However, for systems comprised of many molecules, which are most common in statistical mechanics studies, it is impossible to compute (or realistically approximate) these eigenfunctions. Thus, it is not productive to try to express C(t) in terms of these eigenfunctions. Therefore, an entirely new set of tools has been introduced to handle time-propagation in the quantum case, and it is these new devices that I now attempt to describe in a manner much like we saw in Section 1.3.6's discussion of time propagation of wave functions.

To illustrate, consider the time propagation issue contained in the quantum definition of C(t) shown above. One is faced with

1. propagating $|\Phi_j>$ from t=0 up to time t, using $\exp(-iHt/\hbar) |\Phi_j>$ and then acting with the operator B

2. acting with the operator A^+ on $|\Phi_j>$ and then propagating $A^+ |\Phi_j>$ from t=0 up to time t, using $\exp(-iHt/\hbar)A^+ |\Phi_j>$;

3. C(t) then requires that these two time-propagated functions be multiplied together and integrated over the coordinates that Φ depends on.

The exp(-βH) operator that also appears in the definition of C(t) can be combined, for example, with the first time propagation step and actually handled as part of the time propagation as follows:

$$\exp(-iHt/\hbar)\ |\Phi_j> \exp(-\beta E_j) = \exp(-iHt/\hbar)\ \exp(-\beta H)\ |\Phi_j>$$

$$=\exp(-i[t+\beta\ \hbar\ /i]H/\hbar)\ |\Phi_j>.$$

The latter expression can be viewed as involving a propagation in complex time from t = 0 to t = t + β ℏ /i. Although having a complex time may seem unusual, as I will soon point out, it turns out that it can have a stabilizing influence on the success of these tools for computing quantum correlation functions.

Much like we saw earlier in Section 1.3.6, so-called Feynman path integral techniques can be used to carry out the above time propagations. One begins by dividing the time interval into P discrete steps (this can be the real time interval or the complex interval)

$$\exp[-i\ Ht/\hbar] = \{\exp[-i\ H\delta t/\hbar\ ]\}^{P}\ .$$

The number P will eventually be taken to be large, so each time step $\delta t = t/P$ has a small magnitude. This fact allows us to use approximations to the exponential operator appearing in the propagator that are valid only for short time steps. For each of these short time steps one then approximates the propagator in the most commonly used so-called split symmetric form:

$$\exp[-i\ H\delta t/\hbar] = \exp[-i\ V\delta t/2\hbar]\ \exp[-i\ T\delta t/\hbar]\ \exp[-i\ V\delta t/2\hbar].$$

Here, V and T are the potential and kinetic energy operators that appear in H = T + V. It is possible to show that the above approximation is valid up to terms of order $(\delta t)^4$. So,

for short times (i.e., small δt), these symmetric split operator approximation to the propagator should be accurate.

The time evolved wave function $\Phi(t)$ can then be expressed as

$$\Phi(t) = \{ \exp[-i\, V\delta t/2\hbar] \exp[-i\, T\delta t/\hbar] \exp[-i\, V\delta t/2\hbar]\}^P\, \Phi(t=0).$$

The potential V is (except when external magnetic fields are present) a function only of the coordinates $\{q_i\}$ of the system, while the kinetic term T is a function of the momenta $\{p_j\}$ (assuming Cartesian coordinates are used). By making use of the completeness relations for eigenstates of the coordinate operator

$$1 = \int dq \mid q_j> < q_j|$$

and inserting this identity P times (once between each combination of $\exp[-i\, V\delta t/2h] \exp[-i\, T\delta t/h] \exp[-i\, V\delta t/2h]$ factors), the expression given above for $\Phi(t)$ can be rewritten as follows:

$$\Phi(q_P\,,t)= \int dq_{P-1}\ dq_{P-2} \ldots dq_1\ dq_0\ \Pi_{j=1,P}\ \exp\{(-i\delta t/2\hbar)[V(q_j) + V(q_{j-1})]\}$$

$$< q_j| \exp(-i\delta tT/\hbar)\ |q_{j-1}>\Phi(q_0,0).$$

Then, by using the analogous completeness identity for the momentum operator

$$1 = (1/\hbar) \int dp_j|\ p_j>< p_j\ |$$

one can write

$$< q_j| \exp(-i\delta tT/h)\ |q_{j-1}> = (1/\hbar) \int dp < q_j|p > \exp(-ip^2\delta t/2m\hbar)\ < p|q_{j-1}>.$$

Finally, by using the fact (recall this from Section 1.3.6) that the momentum eigenfunctions |p>, when expressed as functions of coordinates q are given by

$$< q_j|p > = (1/2\pi)^{1/2} \exp(ipq/\hbar),$$

the above integral becomes

$$< q_j \mid \exp(-i\delta t T /\hbar) \mid q_{j-1}> = (1/2\pi\hbar) \int dp \exp(-ip^2 \delta t /2m\hbar) \exp[ip(q_j - q_{j-1})/\hbar].$$

This integral over p can be carried out analytically to give

$$< q_j \mid \exp(-i\delta t T /\hbar) \mid q_{j-1}> = (m/2\pi i\hbar \delta t)^{1/2} \exp[im(q_j - q_{j-1})^2 /2\hbar \delta t].$$

When substituted back into the multidimensional integral for $\Phi(q_P ,t)$, we obtain

$$\Phi(q_P ,t)= (m/2\pi i\hbar \delta t)^{P/2} \int dq_{P-1} dq_{P-2} \ldots dq_1 dq_0 \; \Pi_{j=1,P} \exp\{(-i\delta t/2\hbar)[V(q_j) + V(q_{j-1})]\}$$

$$\exp[im(q_j - q_{j-1})^2 /2\hbar \delta t] \; \Phi (q_0,0)$$

or

$$\Phi(q_P ,t)= (m/2\pi i\hbar \delta t)^{P/2} \int dq_{P-1} dq_{P-2} \ldots dq_1 dq_0 \; \exp\{\Sigma_{j=1,P}[(-i\delta t/2\hbar)[V(q_j) + V(q_{j-1})]$$

$$+ (i \, m(q_j - q_{j-1})^2 /2\hbar \delta t)]\} \; \Phi (q_0,0).$$

Recall what we said earlier that the time correlation function was to be computed by:

1. propagating $|\Phi_j >$ from t=0 up to time t, using $\exp(-iHt/\hbar) |\Phi_j >$ and then acting with the operator B

2. acting with the operator A^+ on $|\Phi_j>$ and then propagating $A^+ |\Phi_j >$ from t=0 up to time t, using $\exp(-iHt/\hbar)A^+ |\Phi_j >$;

3. multiplying together these two functions and integrating over the coordinates that Φ depends on.

So all of the effort described above would have to be expended for $\Phi(q_0,0)$ taken to be $|\Phi_j>$ after which the result would be multiplied by the operator B, as well as for $\Phi(q_0,0)$ taken to be $A^+|\Phi_j>$ to allow the quantum time correlation function C(t) to be evaluated. These steps can be performed, but they are very difficult to implement, so I will refer the student to *Computer Simulations of Liquids*, M. P. Allen and D. J. Tildesley, Oxford U. Press, New York (1997) for further discussion on this topic.

Why are the multidimensional integrals of the form shown above called path integrals? Because the sequence of positions q_1, ... q_{P-1} describes a path connecting q_0 to q_P. By integrating over all of the intermediate positions q_1, q_2,... q_{P-1} for any given q_0 and q_P one is integrating over all paths that connect q_0 to q_P. Further insight into the meaning of the above is gained by first realizing that

$$(m/2\delta t)\ (q_j - q_{j-1})^2 = (m/2(\delta t)^2)\ (q_j - q_{j-1})^2\ \delta t = \int T\ dt$$

is the finite-difference representation, within the P discrete time steps of length δt, of the integral of Tdt over the jth time step, and that

$$(\delta t/2)\ [V(q_j) + V(q_{j-1})] = \int V(q)dt$$

is the representation of the integral of Vdt over the jth time step. So, for any particular path (i.e., any specific set of q_0, q_1, , ... q_{P-1}, q_P values), the sum over all such terms $\Sigma_{j=1,P-1}\ [m(q_j - q_{j-1})^2 / 2\delta t - \delta t(V(q_j) + V(q_{j-1}))/2]$ represents the integral over all time from t=0 until t = t of the so-called Lagrangian L = T - V:

$$\Sigma_{j=1,P-1}\ [m(q_j - q_{j-1})^2 / 2\delta t - \delta t(V(q_j) + V(q_{j-1}))/2] = \int L dt.$$

This time integral of the Lagrangian is called the action S in classical mechanics (recall that in Chapter 1, we used quantization of the action in the particle-in-a-box problem). Hence, the N-dimensional integral in terms of which $\Phi(q_P,t)$ is expressed can be written as

$$\Phi\,(q_P\,,t) = (m/2\pi i\hbar\,\delta t)^{P/2}\;\Sigma_{\text{all paths}}\;\exp\{i\,/\hbar\,\textstyle\int dt\,L\,\}\;\Phi\,(q_0\,,t{=}0).$$

Here, the notation "all paths" is realized in the earlier version of this equation by dividing the time axis from $t = 0$ to $t = t$ into P equal divisions, and denoting the coordinates of the system at the jth time step by q_j. By then allowing each q_j to assume all possible values (i.e., integrating over all possible values of q_j using, for example, the Monte-Carlo method discussed earlier), one visits all possible paths that begin at q_0 at $t = 0$ and end at q_P at $t = t$. By forming the classical action S

$$S = \textstyle\int dt L$$

for each path and then summing $\exp(iS/\hbar)\,\Phi(q_0\,,t{=}0)$ over all paths and multiplying by $(m/2\pi\hbar\,\delta t)^{P/2}$, one is able to form $\Phi(q_P\,,t)$.

The difficult step in implementing this Feynman path integral method in practice involves how one identifies all paths connecting q_0, $t = 0$ to q_P, t. Each path contributes an additive term involving the complex exponential of the quantity

$$\Sigma_{j=1,P-1}\;[m(q_j - q_{j-1})^2\,/\,2\delta t - \delta t(V(q_j) + V(q_{j-1}))/2]$$

Because the time variable $\delta t = t/P$ appearing in each action component can be complex (recall that, in one of the time evolutions, t is really $t + \beta\hbar\,/i$), the exponentials of these action components can have both real and imaginary parts. The real parts, which arise from the $\exp(-\beta H)$, cause the exponential terms to be damped (i.e., to undergo exponential decay), but the imaginary parts give rise (in $\exp(iS/\hbar)$) to oscillations. The sum of many, many (actually, an infinite number of) oscillatory $\exp(iS/\hbar) = \cos(S/\hbar) + i\,\sin(S/\hbar)$ terms is extremely difficult to evaluate because of the tendency of contributions from one path to cancel those of another path. The practical evaluation of such sums remains a very active research subject.

The most commonly employed approximation to this sum involves finding the path(s) for which the action

$$S = \Sigma_{j=1,P-1} \, [m(q_j - q_{j-1})^2 / 2\delta t - \delta t(V(q_j) + V(q_{j-1}))/2]$$

is smallest because such paths produce the lowest-frequency oscillations in $\exp(iS/\hbar)$, and thus may be less subject to cancellation by contributions from other paths.

The path(s) that minimize the action S are, in fact, the classical paths. That is, they are the paths that the system whose quantum wave function is being propagated would follow if the system were undergoing classical Newtonian mechanics subject to the conditions that the system be at q_0 at t=0 and at q_P at t=t. In this so-called semi-classical approximation to the propagation of the initial wave function using Feynman path integrals, one finds all classical paths that connect q_0 at t = 0 and at q_P at t = t, and one evaluates the action S for each such path. One then applies the formula

$$\Phi(q_P, t) = (m/2\pi i\hbar \, \delta t)^{P/2} \, \Sigma_{\text{all paths}} \, \exp\{i /\hbar \int dt \, L\} \, \Phi(q_0, t=0)$$

but includes in the sum only the contribution from the classical path(s). In this way, one obtains an approximate quantum propagated wave function via a procedure that requires knowledge of only classical propagation paths.

Clearly, the quantum propagation of wave functions, even within the semi-classical approximation discussed above, is a rather complicated affair. However, keep in mind the alternative that one would face in evaluating, for example, spectroscopic line shapes if one adopted a time-independent approach. One would have to know the energies and wave functions of a system comprised of many interacting molecules. This knowledge is simply not accessible for any but the simplest molecules. For this reason, the time-dependent framework in which one propagates classical trajectories or uses path-integral techniques to propagate initial wave functions offers the most feasible way to evaluate the correlation functions that ultimately produce spectral line shapes and other time correlation functions for complex molecules in condensed media.

Before finishing this Section, it might help if I showed how one obtains the result that classical paths are those that make the action integral $S = \int L dt$ minimum. This provides the student with an introduction to the subject called calculus of variations or functional analysis, which most students reading this text have probably not studied in a

class. First, let's clarify what a functional is. A function f(x) depends on one or more variables x that take on scalar values; that is, given a scalar number x, f(x) produces the value of the function f at this value of x. A functional F[f] is a function of the function f if, given the function f, F acts on it to produce a value. In more general functionals, F[f] might depend not only of f, but on various derivatives of f. Let's consider an example. Suppose one has a functional of the form

$$F[f] = \int_{t_0}^{t_f} F(t, f(t), \frac{df(t)}{dt}) dt$$

meaning that the functional involves an integral from t_0 through t_f of an integrand that may contain (i) the variable t explicitly, (ii) the function f(t), and (iii) the derivative of this function with respect to the variable t. This is the kind of integral one encounters when evaluating the action integral

$$S = \int_{t_0}^{t_f} [T - V] dt = \int_{t_0}^{t_f} [\frac{m}{2}(\frac{dx(t)}{dt})^2 - V(x(t))] dt$$

where the function f(t) is the coordinate x(t) that evolves from $x(t_0)$ to $x(t_f)$. The task at hand is to determine that function x(t) for which this integral is a minimum.

We solve this problem proceeding much as one would do if one had to minimize a function of a variable; we differentiate with respect to the variable and set the derivative to zero. However, in our case, we have a function of a function, not a function of a variable; so how do we carry out the derivative? We assume that the function x(t) that minimizes S is known, and we express any function that differs a little bit from the correct x(t) as

$$x(t) + \varepsilon \eta(t)$$

where ε is a scalar quantity used to suggest that x(t) and $x(t) + \varepsilon \eta(t)$ differ by only a small amount and $\eta(t)$ is a function that obeys

$$\eta(t) = 0 \text{ at } t=t_0 \text{ and at } t = t_f;$$

this is how we guarantee that we are only considering paths that connect to the proper x_0 at t_0 and x_f at t_f. By considering all possible functions $\eta(t)$ that obey these conditions, we have in $x(t) + \varepsilon\eta(t)$ a parameterization of all paths that begin (at t_0) and end (at t_f) where the exact path x(t) does but differ by a small amount from x(t). Substituting $x(t) + \varepsilon\eta(t)$ into

$$S = \int_{t_0}^{t_f} [\frac{m}{2}(\frac{dx(t)}{dt})^2 - V(x(t))]dt$$

gives

$$S = \int_{t_0}^{t_f} [\frac{m}{2}\{\frac{dx(t)}{dt} + \varepsilon\frac{d\eta(t)}{dt}\}^2 - V\{x(t)) + \varepsilon\eta(t)\}]dt .$$

The terms in the integrand are then expanded in powers of the ε parameter

$$\{\frac{dx(t)}{dt} + \varepsilon\frac{d\eta(t)}{dt}\}^2 = \frac{dx(t)}{dt} + 2\varepsilon\frac{dx(t)}{dt}\frac{d\eta(t)}{dt} + \varepsilon^2[\frac{d\eta}{dt}]^2$$

$$-V(x(t) + \varepsilon\eta(t)) = -V(x(t)) - \varepsilon\frac{\partial V(x(t))}{\partial x(t)}\eta(t) - 1/2\varepsilon^2\frac{\partial^2 V(x(t))}{\partial x(t)^2}\eta^2(t) - ...$$

and substituted into the integral for S. Collecting terms of each power of ε allows this integral to be written as

$$S(\varepsilon) = \int_{t_0}^{t_f} [\frac{m}{2}\{(\frac{dx(t)}{dt})^2 + 2\varepsilon\frac{dx(t)}{dt}\frac{d\eta(t)}{dt} + O(\varepsilon^2)\} - V(x(t)) - \varepsilon\frac{\partial V(x(t))}{\partial x(t)}\eta(t) - O(\varepsilon^2)]dt .$$

The condition that S(ε) be stable with respect to variations in ε can be expressed as

$$\frac{dS(\varepsilon)}{d\varepsilon} = 0 = \lim_{\varepsilon \to 0} \frac{S(\varepsilon) - S(0)}{\varepsilon}$$

which is equivalent to requiring that the terms linear in ε in the above expansion for S(ε) vanish

$$0 = \int_{t_0}^{t_f} [m \frac{dx(t)}{dt} \frac{d\eta(t)}{dt} - \frac{\partial V(x(t))}{\partial x(t)} \eta(t)]dt$$

Next, we use integration by parts to rewrite the first term involving $\frac{d\eta(t)}{dt}$ as a term involving $\eta(t)$ instead

$$\int_{t_0}^{t_f} m \frac{dx(t)}{dt} \frac{d\eta(t)}{dt} dt = m\left[\frac{dx(t)}{dt} \eta(t)\right]_{t_0}^{t_f} - \int_{t_0}^{t_f} m \frac{d^2 x(t)}{dt^2} \eta(t)dt$$

Because the function $\eta(t)$ vanishes at t_0 and t_f, the first term vanishes, so this identity can be used to rewrite the condition that the terms in S(ε) that are linear in ε vanish as

$$0 = \int_{t_0}^{t_f} [-m \frac{d^2 x(t)}{dt^2} - \frac{\partial V(x(t))}{\partial x(t)}]\eta(t)dt.$$

Because this result is supposed to be valid for any function $\eta(t)$ that vanishes at t_0 and t_f, the factor multiplying $\eta(t)$ in the above integral must itself vanish

$$-m \frac{d^2 x(t)}{dt^2} - \frac{\partial V(x(t))}{\partial x(t)} = 0.$$

This shows that the path x(t) that makes S stationary is the path that obeys Newton's equations- the classical path. I urge the student reader to study this example of the use of functional analysis because this mathematical device is an important tool too master.

7.5 Some Important Chemical Applications of Statistical Mechanics

In this Section, I introduce several applications of statistical mechanics that are important for students to be aware of because they arise frequently when chemists make use of the tools of statistical mechanics. These examples include

1. The basic equations connecting the translational, rotational, vibrational, and electronic properties of isolated (i.e., gas-phase) molecules to their thermodynamics.

2. The most basic descriptions of the vibrations of ions, atoms, or molecules within crystals.

3. The most elementary models for describing cooperative behavior and phase transitions in gas-surface and liquid-liquid systems.

4. The contributions of intermolecular forces to the thermodynamics of gases.

7.5.1 Gas-Molecule Thermodynamics

The equations relating the thermodynamic variables to the molecular partition functions can be employed to obtain the following expressions for the energy E, heat capacity C_V, Helmholz free energy A, entropy S, and chemical potential μ in the case of a gas (i.e., in the absence of intermolecular interactions) of polyatomic molecules:

$$E/NkT = 3/2 + 3/2 + \Sigma_{J=1,3N-6} [h\nu_J/2kT + h\nu_J/kT (\exp(h\nu_J/kT)-1)^{-1}] - D_e/kT,$$

$$C_V/Nk = 3/2 + 3/2 + \Sigma_{J=1,3N-6} (h\nu_J/kT)^2 \exp(h\nu_J/kT) (\exp(h\nu_J/kT)-1)^{-2},$$

$$-A/NkT = \ln \{[2\pi mkT/h^2]^{3/2} (Ve/N)\} + \ln[(\pi^{1/2}/\sigma) (8\pi^2 I_A kT/h^2)^{1/2} (8\pi^2 I_B kT/h^2)^{1/2}$$

$$(8\pi^2 I_C kT/h^2)^{1/2}] - \Sigma_{J=1,3N-6} [h\nu_J/2kT + \ln(1-\exp(-h\nu_J/kT))] + D_e/kT + \ln\omega_e$$

$$S/Nk = \ln \{[2\pi mkT/h^2]^{3/2} (Ve^{5/2}/N)\} + \ln [(\pi^{1/2}/\sigma) (8\pi^2 I_A kT/h^2)^{1/2} (8\pi^2 I_B kT/h^2)^{1/2}$$

$(8\pi^2 I_C kT/h^2)^{1/2}] + \Sigma_{J=1,3N-6} [h\nu_J/kT (exp(h\nu_J/kT)-1)^{-1} - ln(1-exp(-h\nu_\vartheta/kT))] + ln\omega_e$

$$\mu/kT = - ln \{[2\pi mkT/h^2]^{3/2} (kT/p)\} - ln[(\pi^{1/2}/\sigma) (8\pi^2 I_A kT/h^2)^{1/2} (8\pi^2 I_B kT/h^2)^{1/2}$$

$(8\pi^2 I_C kT/h^2)^{1/2}] + \Sigma_{J=1,3N-6} [h\nu_J/2kT + ln(1-exp(-h\nu_J/kT))] - D_e/kT - ln\omega_e.$

Earlier in this Chapter in Section 7.1.2, we showed how these equations are derived, so I refer the reader back to that treatment for further details.

Notice that, except for the chemical potential μ, all of these quantities are extensive properties that depend linearly on the number of molecules in the system N. Except for the chemical potential μ and the pressure p, all of the variables appearing in these expressions have been defined earlier when we showed the explicit expressions for the translational, vibrational, rotational, and electronic partition functions. These are the working equations that allow one to compute thermodynamic properties of stable molecules, ions, and even reactive species such as radicals in terms of molecular properties such as geometries, vibrational frequencies, electronic state energies and degeneracies, and the temperature, pressure, and volume.

7.5.2 Einstein and Debye Models of Solids

These two models deal with the vibrations of crystals that involve motions among the neighboring atoms, ions, or molecules that comprise the crystal. These inter-fragment vibrations are called phonons. In the Einstein model of a crystal, one assumes that:
1. Each atom, ion, or molecule from which the crystal is constituted is trapped in a potential well formed by its interactions with neighboring species. This potential is denoted $\phi(V/N)$ with the volume-to-number V/N ratio written to keep in mind that it likely depends on the packing density (i.e., the distances among neighbors) within the crystal. Keep in mind that ϕ represents the interaction of any specific atom, ion, or molecule with the N-1 other such species. So, N $\phi/2$, not N ϕ is the total interaction energy among all of the species; the factor of 1/2 is necessary to avoid double counting.

2. Each such species is assumed to undergo local harmonic vibrational motions about its equilibrium position ($\mathbf{q}_J^0$) within the local well that traps it. If the crystal is isotropic, the force constants k_J that characterize the harmonic potential $1/2 \, k_J \, (q_J - q_J^0)^2$ along the x, y, and z directions are equal; if not, these k_J parameters may be unequal. It is these force constants, along with the masses m of the atoms, ions, or molecules, that determine the harmonic frequencies $\nu_J = 1/2\pi \, (k_J/m)^{1/2}$ of the crystal.

3. The inter-species phonon vibrational partition function of the crystal is then assumed to be a product of N partition functions, one for each atom, ion, or molecule in the crystal, with each partition function taken to be of the harmonic vibrational form:

$$Q = \exp(-N \, \phi/2kT) \, \{\Pi_{J=1,3} \exp(-h\nu_J/2kT) \, (1-\exp(-h\nu_J/kT))^{-1}\}^N.$$

There is no factor of N! in the denominator because, unlike a gas of N species, each of these N species (atoms, ions, or molecules) are constrained to stay put (i.e., not free to roam independently) in the trap induced by their neighbors. In this sense, the N species are distinguishable rather than indistinguishable as they are in the gas case. The $N\phi/2kT$ factor arises when one asks what the total energy of the crystal is, aside from its vibrational energy, relative to N separated species; in other words, what is the total cohesive energy of the crystal. This energy is N times the energy of any single species ϕ, but, as noted above, divided by 2 to avoid double counting the inter-species interaction energies.

This partition function can be subjected to the thermodynamic equations discussed earlier to compute various thermodynamic properties. One of the most useful to discuss for crystals is the heat capacity C_V, which is given by (see the vibrational contribution to C_V expressed in Section 7.5.1) :

$$C_V = Nk \, \Sigma_{J=1,3} \, (h\nu_J/kT)^2 \exp(h\nu_J/kT) \, (\exp(h\nu_J/kT) -1)^{-2}.$$

At very high temperatures, this function can be shown to approach 3Nk, which agrees with the experimental observation know as the law of Dulong and Petit. However, at very low temperatures, this expression approaches:

$$C_V \rightarrow \Sigma_{J=1,3} Nk \, (h\nu_J/kT)^2 \exp(-h\nu_J/kT),$$

which goes to zero as T approaches zero, but not in a way that is consistent with experimental observation. That is, careful experimental data shows that all crystal heat capacities approach zero proportional to T^3 at low temperature; the Einstein model's C_V approaches zero but not in the T^3 form found in experiments.

So, although the Einstein model offers a very useful model of how a crystal's stability relates to $N\phi$ and how its C_V depends on vibrational frequencies of the phonon modes, it does not work well at low temperatures. Nevertheless, it remains a widely used model in which to understand the phonons' contributions to thermodynamic properties as long as one does not attempt to extrapolate its predictions to low T.

In the Debye model of phonons in crystals, one abandons the view in which each atom, ion, or molecule vibrates independently about it own equilibrium position and replaces this with a view in which the constituent species vibrate collectively in wave-like motions. Each such wave has a wave length λ and a frequency ν that are related to the speed c of propagation of such waves in the crystal by

$$c = \lambda \, \nu.$$

The speed c is a characteristic of the crystal's inter-species forces; it is large for stiff crystals and small for soft crystals.

In a manner much like we used to determine the density of quantum states $\Omega(E)$ within a three-dimensional box, one can determine how many waves can fit within a cubic crystalline box having frequencies between ν and $\nu + d\nu$. The approach to this problem is to express the allowed wave lengths and frequencies as:

$$\lambda_n = 2L/n,$$

$$\nu_n = n \, c/2L,$$

where L is the length of the box on each of its sides and n is an integer $1, 2, 3, ...$ This prescription forces all wave lengths to match the boundary condition for vanishing at the box boundaries.

Then carrying out a count of how many ($\Omega(\nu)$) waves have frequencies between ν and $\nu + d\nu$ for a box whose sides are all equal gives the following expression:

$$\Omega(\nu) = 12\pi\ V\ \nu^2/c^3.$$

The primary observation to be made is that the density of waves is proportional to ν^2:

$$\Omega(\nu) = a\ \nu^2.$$

It is conventional to define the parameter a in terms of the maximum frequency ν_m that one obtains by requiring that the integral of $\Omega(\nu)$ over all allowed ν add up to 3N, the total number of inter-species vibrations that can occur:

$$3N = \int \Omega(\nu)\ d\nu = a\ \nu_m^3/3.$$

This then gives the constant a in terms of ν_m and N and allows $\Omega(\nu)$ to be written as

$$\Omega(\nu) = 9N\nu^2/\nu_m^3.$$

The Debye model uses this wave picture and computes the total energy E of the crystal much as done in the Einstein model, but with the sum over 3N vibrational modes replaced by a continuous integral over the frequencies ν weighted by the density of such states $\Omega(\nu)$ ((see the vibrational contribution to E expressed in Section 7.5.1):

$$E = N\phi/2 + (9NkT/\nu_m^3) \int [h\nu/2kT + (h\nu/kT)\ (\exp(h\nu/kT) -1)^{-1}\ ]\nu^2\ d\nu,$$

where the integral over ν ranges from 0 to ν_m. It turns out that the C_V heat capacity obtained by taking the temperature derivative of this expression for E can be written as follows:

$$C_V = 3Nk\ [\ 4\ D(h\nu_\mu/kT) - 3(h\nu_\mu/kT)\ (\exp(h\nu_\mu/kT) - 1)^{-1}\ ]$$

where the so-called Debye function $D(u)$ is defined by

$$D(u) = 3\ u^{-3} \int x^3\ (\exp(x) - 1)^{-1}\ dx,$$

and the integral is taken from $x = 0$ to $x = u$.

The important thing to be noted about the Debye model is that the heat capacity, as defined above, extrapolates to 3Nk at high temperatures, thus agreeing with the law of Dulong and Petit, and varies at low temperature as

$$C_V \rightarrow (12/5)\ Nk\pi^4\ (kT/h\nu_m)^3.$$

So, the Debye heat capacity does indeed vary as T^3 at low T as careful experiments indicate. For this reason, it is appropriate to use the Debye model whenever one is interested in properly treating the energy, heat capacity, and other thermodynamic properties of crystals at temperatures for which $kT/h\nu_m$ is small. At higher temperatures, it is appropriate to use either the Debye or Einstein models. The major difference between the two lies in how they treat the spectrum of vibrational frequencies that occur in a crystal. The Einstein model says that only one (or at most three, if three different k_J values are used) frequency occurs $\nu_J = 1/2\pi\ (k_J/\mu)^{1/2}$; each species in the crystal is assumed to vibrate at this frequency. In contrast, the Debye model says that the species vibrate collectively and with frequencies ranging from $\nu = 0$ up to $\nu = \nu_m$, the so-called Debye frequency, which is proportional to the speed c at which phonons propagate in the crystal. In turn, this speed depends on the stiffness (i.e., the inter-species potentials) within the crystal.

7.5.3 Lattice Theories of Surfaces and Liquids

This kind of theory can be applied to a wide variety of chemical and physical problems, so it is a very useful model to be aware of. The starting point of the model is to consider a lattice containing M sites, each of which has c nearest neighbor sites (n.b., clearly, c will depend on the structure of the lattice) and to imagine that each of these sites can exist in either of two states that we label A and B. Before deriving the basic equations of this model, let me explain how the concepts of sites and A and B states are used to apply the model to various problems.

1. The sites can represent binding sites on the surface of a solid and the two states A and B can represent situations in which the site is either occupied (A) or unoccupied (B) by a molecule that is chemisorbed or physisorbed to the site. This point of view is taken when one applies lattice models to adsorption of gases or liquids to solid surfaces.

2. The sites can represent individual spin = 1/2 molecules or ions within a lattice, and the states A and B can denote the α and β spin states of these species. This point of view allows the lattice models to be applied to magnetic materials.

3. The sites can represent positions that either of two kinds of molecules A and B might occupy in a liquid or solid in which case A and B are used to label whether each site contains an A or a B molecule. This is how we apply the lattice theories to liquid mixtures.

4. The sites can represent cis- and trans- conformations in linkages within a polymer, and A and B can be used to label each such linkage as being either cis- or trans-. This is how we use these models to study polymer conformations.

In Fig. 7.4, I show a two-dimensional lattice having 25 sites of which 16 are occupied by dark (A) species and 9 are occupied by lighter (B) species.
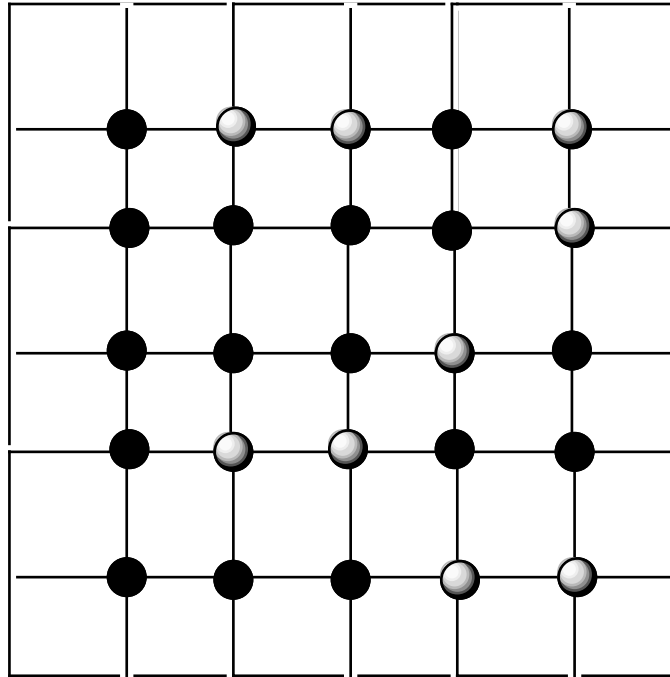
Figure 7.4 Two-dimensional lattice having 25 sites with 16 A and 9 B species

The partition function for such a lattice is written in terms of a degeneracy Ω and an energy E, as usual. The degeneracy is computed by considering the number of ways a total of $N_A + N_B$ species can be arranged on the lattice:

$$\Omega = (N_A+N_B)!/[N_A!\ N_B!].$$

The interaction energy among the A and B species for any arrangement of the A and B on the lattice is assumed to be expressed in terms of pair wise interaction energies. In particular, if only nearest neighbor interaction energies are considered, one can write the total interaction energy E_{int} of any arrangement as

$$E_{int} = N_{AA}\ E_{AA} + N_{BB}\ E_{BB} + N_{AB}\ E_{AB}$$

where N_{IJ} is the number of nearest neighbor pairs of type I-J and E_{IJ} is the interaction energy of an I-J pair. The example shown in Fig. 7.4 has $N_{AA} = 16$, $N_{BB} = 4$ and $N_{AB} = 19$.

The three parameters N_{IJ} that characterize any such arrangement can be re-expressed in terms of the numbers N_A and N_B of A and B species and the number of nearest neighbors per site c as follows:

$$2N_{AA} + N_{AB} = cN_A$$

$$2N_{BB} + N_{AB} = cN_B.$$

Note that the sum of these two equations states the obvious fact that twice the sum of AA, BB, and AB pairs must equal the number of A and B species multiplied by the number of neighbors per species, c.

Using the above relationships among N_{AA}, N_{BB}, and N_{AB}, we can rewrite the interaction energy as

$$E_{int} = E_{AA} (c\, N_A - N_{AB})/2 + E_{BB} (c\, N_B - N_{AB})/2 + E_{AB} N_{AB}$$

$$= (N_A E_{AA} + N_B E_{BB})\, c/2 + (2\, E_{AB} - E_{AA} - E_{BB})\, N_{AB}/2$$

The reason it is helpful to write E_{int} in this manner is that it allows us to express things in terms of two variables over which one has direct experimental control, N_A and N_B, and one variable N_{AB} that characterizes the degree of disorder among the A and B species. That is, if N_{AB} is small, the A and B species are arranged on the lattice in a phase-separated manner; whereas, if N_{AB} is large, the A and B are well mixed.

The total partition function of the A and B species arranged on the lattice is written as follows:

$$Q = q_A^{\,NA}\, q_B^{\,NB} \Sigma_{NAB}\, \Omega(N_A, N_B, N_{AB})\, \exp(-E_{int}/kT).$$

Here, q_A and q_B are the partition functions (electronic, vibrational, etc.) of the A and B species as they sit bound to a lattice site and $\Omega(N_A, N_B, N_{AB})$ is the number of ways that N_A species of type A and N_B of type B can be arranged on the lattice such that there are

N_{AB} A-B type nearest neighbors. Of course, E_{int} is the interaction energy discussed earlier. The sum occurs because a partition function is a sum over all possible states of the system. There are no $(1/N_J!)$ factors because, as in the Einstein and Debye crystal models, the A and B species are not free to roam but are tied to lattice sites and thus are distinguishable.

This expression for Q can be rewritten in a manner that is more useful by employing the earlier relationships for N_{AA} and N_{BB}:

$$Q = (q_A \exp(-cE_{AA}/2kT))^{NA} (q_B \exp(-cE_{BB}/2kT))^{NB} \Sigma_{NAB} \ \Omega(N_A, N_B, N_{AB}) \ \exp(N_{AB}X/2kT),$$

where

$$X = (-2 E_{AB} + E_{AA} + E_{BB}).$$

The quantity X plays a central role in all lattice theories because it provides a measure of how different the A-B interaction energy is from the average of the A-A and B-B interaction energies. As we will soon see, if X is large and negative (i.e., if the A-A and B-B interactions are highly attractive), phase separation can occur; if X is positive, phase separation will not occur.

The problem with the above expression for the partition function is that no one has yet determined an analytical expression for the degeneracy $\Omega(N_A, N_B, N_{AB})$ factor. Therefore, in the most elementary lattice theory, known as the Bragg-Williams approximation, one approximates the sum over N_{AB} by taking the following average value of N_{AB}:

$$N_{AB}^* = N_A (cN_B)/(N_A+N_B)$$

in the expression for Ω. This average is formed by taking the number of A sites and multiplying by the number of neighbor sites (c) and by the fraction of these neighbor sites

that would be occupied by a B species if mixing were random. This aproximation produces

$$Q = (q_A \exp(-cE_{AA}/2kT))^{N_A} (q_B \exp(-cE_{BB}/2kT))^{N_B} \exp(N_{AB}*X/2kT) \Sigma_{NAB} \Omega(N_A, N_B, N_{AB}).$$

Finally, we realize that the sum $\Sigma_{NAB} \Omega(N_A, N_B, N_{AB})$ is equal to the number of ways of arranging N_A A species and N_B B species on the lattice regardless of how many A-B neighbor pairs there are. This number is, of course, $(N_A+N_B)!/[(N_A!)(N_B!)]$.

So, the Bragg-Williams lattice model partition function reduces to:

$$Q = (q_A \exp(-cE_{AA}/2kT))^{N_A} (q_B \exp(-cE_{BB}/2kT))^{N_B} (N_A+N_B)!/[(N_A!)(N_B!)] \exp(N_{AB}*X/2kT).$$

The most common connection one makes to experimental measurements using this partition function arises by computing the chemical potentials of the A and B species on the lattice and equating these to the chemical potentials of the A and B as they exist in the gas phase. In this way, one uses the equilibrium conditions (equal chemical potentials in two phases) to relate the vapor pressures of A and B, which arise through the gas-phase chemical potentials, to the interaction energy X.

Let me now show you how this is done. First, we use

$$\mu_J = -kT \, (\partial \ln Q/\partial N_J)_{T,V}$$

to compute the A and B chemical potentials on the lattice. This gives

$$\mu_A = -kT\{ \ln(q_A \exp(-cE_{AA}/2kT)) - \ln(N_A/(N_A+N_B)) + (1-[N_A/(N_A+N_B)])^2 \, cX/2kT \}$$

and an analogous expression for μ_B with N_B replacing N_A. The expression for the gas-phase chemical potentials $\mu_A{}^g$ and $\mu_B{}^g$ given earlier in this Chapter has the form:

$$\mu = -kT \ln \{[2\pi mkT/h^2]^{3/2} (kT/p)\} - kT \ln[(\pi^{1/2}/\sigma) (8\pi^2 I_A kT/h^2)^{1/2} (8\pi^2 I_B kT/h^2)^{1/2}$$

$$(8\pi^2 I_C kT/h^2)^{1/2}] +kT \ \Sigma_{J=1,3N-6} [h\nu_J/2kT + \ln(1-\exp(-h\nu_J/kT))] - D_e - kT \ \ln\omega_e,$$

within which the vapor pressure appears. The pressure dependence of this gas-phase expression can be factored out to write each μ as:

$$\mu_A{}^g = \mu_A{}^0 + kT \ \ln(p_A),$$

where p_A is the vapor pressure of A (in atmosphere units) and $\mu_A{}^0$ denotes all of the other factors in $\mu_A{}^g$. Likewise, the lattice-phase chemical potentials can be written as a term that contains the N_A and N_B dependence and a term that does not:

$$\mu_A = -kT\{ \ln(q_A\exp(-cE_{AA}/2kT)) - \ln X_A + (1-X_A)^2 \ cX/2kT \},$$

where X_A is the mole fraction of A ($N_A/(N_A+N_B)$). Of course, an analogous expression holds for μ_B.

   We now perform two steps:
1. We equate the gas-phase and lattice-phase chemical potentials of species A in a case where the mole fraction of A is unity. This gives

$$\mu_A{}^0 + kT \ \ln(p_A{}^0) = -kT\{ \ln(q_A\exp(-cE_{AA}/2kT))\}$$

where $p_A{}^0$ is the vapor pressure of A that exists over the lattice in which only A species are present.
2. We equate the gas- and lattice-phase chemical potentials of A for an arbitrary chemical potential X_A and obtain:

$$\mu_A{}^0 + kT \ \ln(p_A) = -kT\{ \ln(q_A\exp(-cE_{AA}/2kT)) - \ln X_A + (1-X_A)^2 \ cX/2kT \},$$

which contains the vapor pressure p_A of A over the lattice covered by A and B with X_A being the mole fraction of A.

Subtracting these two equations and rearranging, we obtain an expression for how the vapor pressure of A depends on X_A:

$$p_A = p_A^0 X_A \exp(-cX(1-X_A)^2/2kT).$$

Recall that the quantity X is related to the interaction energies among various species as

$$X = (-2\, E_{AB} + E_{AA} + E_{BB}).$$

Let us examine that physical meaning of the above result for the vapor pressure. First, if one were to totally ignore the interaction energies (i.e., by taking $X = 0$), one would obtain the well known Raoult's Law expression for the vapor pressure of a mixture:

$$p_A = p_A^0 X_A$$

$$p_B = p_B^0 X_B.$$

In Fig. 7.5, I plot the A and B vapor pressures vs. X_A. The two straight lines are, of course, just the Raoult's Law findings. I also plot the p_A vapor pressure for three values of the X interaction energy parameter. When X is positive, meaning that the A-B interactions are more energetically favorable than the average of the A-A and B-B interactions, the vapor pressure of A is found to deviate negatively from the Raoult's Law prediction. This means that the observed vapor pressure is lower than is that expected based solely on Raoult's Law. On the other hand, when X is negative, the vapor pressure deviates positively from Raoult's Law.
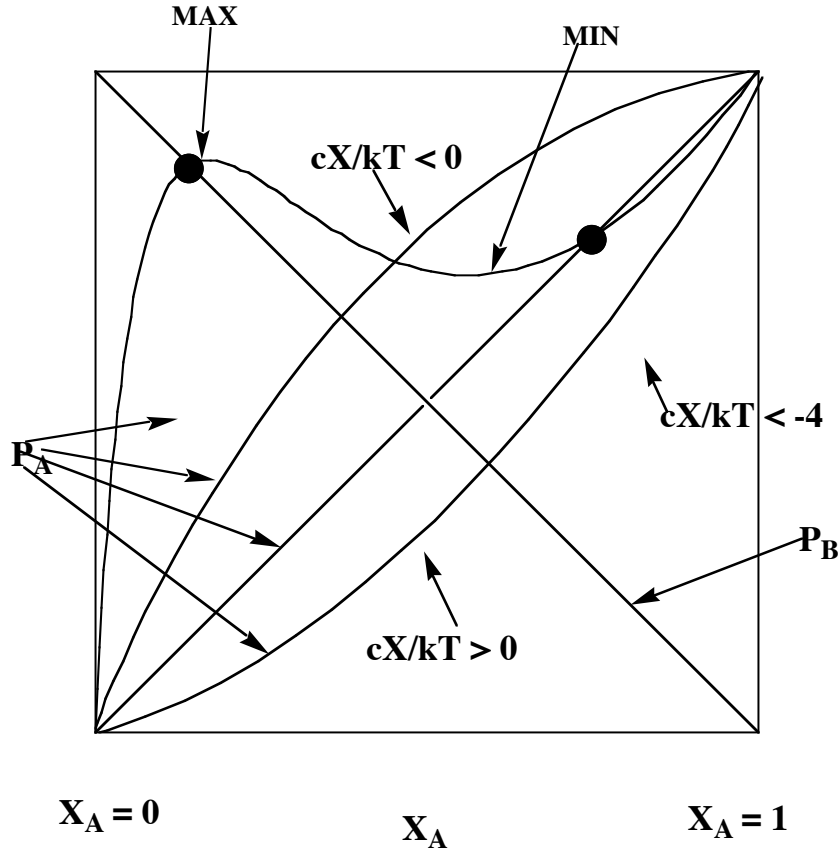
Figure 7.5. Plots of vapor pressures in an A, B mixture as predicted in the lattice model with the Bragg-Williams approximation.

An especially important and interesting case arises when the X parameter is negative and has a value that makes $cX/2kT$ be more negative than -4. It turns out that in such cases, the function p_A suggested in this Bragg-Williams model displays a behavior that suggests a phase transition may occur. Hints of this behavior are clear in Fig. 7.5 where one of the plots displays both a maximum and a minimum, but the plots for $X > 0$ and for $cX/2kT > -4$ do not. Let me now explain this further by examining the derivative of p_A with respect to X_A:

$$dp_A/dX_A = p_A^0 \{1 + X_A(1-X_A) \, 2cX/2kT\} \exp(-cX(1-X_A)^2/2kT).$$

Setting this derivative to zero (in search of a maximum or minimum), and solving for the values of X_A that make this possible, one obtains:

$$X_A = 1/2 \{1 \pm (1+4kT/cX)^{1/2} \}.$$

Because X_A is a mole fraction, it must be less than unity and greater than zero. The above result giving the mole fraction at which $dp_A/dX_A = 0$ will not produce a realistic value of X_A unless

$$cX/kT < -4.$$

If $cX/kT = -4$, there is only one value of X_A (i.e., $X_A = 1/2$) that produces a zero slope; for $cX/kT < -4$, there will be two such values given by $X_A = 1/2 \{1 \pm (1+4kT/cX)^{1/2}\}$, which is what we see in Fig. 7.5 where the plot displays both a maximum and a minimum.

What does it mean for cX/kT to be less than –4 and why is this important? For X to be negative, it means that the average of the A-A and B-B interactions are more energetically favorable than is the A-B interactions. It is for this reason that a phase separation is may be favored in such cases (i.e., the A species prefer to be near other A species more than to be near B species, and similarly for the B species). However, thermal motion can overcome a slight preference for such separation. That is, if X is not large enough, kT can overcome this slight preference. This is why cX must be less than -4kT, not just less than zero.

So, the bottom line is that if the A-A and B-B interactions are more attractive, on average, than are the A-B interactions, one can experience a phase separation in which the A and B species do not remain mixed on the lattice but instead gather into two distinct kinds of domains. One domain will be rich in the A species, having an X_A value equal to that shown in the right dot in Fig. 7.5. The other domains will be rich in B and have an X_A value of that shown by the left dot.

As I noted in the introduction to this Section, lattice models can be applied to a variety of problems. We just analyzed how it is applied, within the Bragg-Williams approximation, to mixtures of two species. In this way, we obtain expressions for how the vapor pressures of the two species in the liquid or solid mixture display behavior that

reflects their interaction energies. Let me now briefly show you how the lattice model is applied in some other areas.

In studying adsorption of gases to sites on a solid surface, one imagines a surface containing M sites per unit area A with N_{ad} molecules (that have been adsorbed from the gas phase) bound to these sites. In this case, the interaction energy E_{int} introduced earlier involves only interactions among neighboring adsorbed molecules; there are no lateral interactions among empty surface sites or between empty surface sites and adsorbed molecules. So, we can make the following replacements in our earlier equations:

$$N_A \rightarrow N_{ad}$$

$$N_B \rightarrow M - N_{ad}$$

$$E_{int} = E_{ad,ad} N_{ad,ad},$$

where $N_{ad,ad}$ is the number of nearest neighbor pairs of adsorbed species and $E_{ad,ad}$ is the pairwise interaction energy between such a pair. The primary result obtained by equating the chemical potentials of the gas-phase and adsorbed molecules is:

$$p = kT (q_{gas}/V) (1/q_{ad}) [\theta/(1-\theta)] \exp(E_{ad}c\theta/kT).$$

Here q_{gas}/V is the partition function of the gas-phase molecules per unit volume, q_{ad} is the partition function of the adsorbed molecules (which contains the adsorption energy as $\exp(-\phi/kT)$) and θ is called the coverage (i.e., the fraction of surface sites to which molecules have adsorbed). Clearly, θ plays the role that the mole fraction X_A played earlier. This so-called adsorption isotherm equation allows one to connect the pressure of the gas above the solid surface to the coverage.

As in our earlier example, something unusual occurs when the quantity $E_{ad}c\theta/kT$ is negative and beyond a critical value. In particular, differentiating the expression for p with respect to θ and finding for what θ value(s) $dp/d\theta$ vanishes, one finds:

$$\theta = 1/2 \, [\, 1 \pm (1 + 4kT/cE_{ad})^{1/2} \,].$$

Since θ is a positive fraction, this equation can only produce useful values if

$$cE_{ad}/kT < -4.$$

In this case, this means that if the attractions between neighboring adsorbed molecules is strong enough, it can overcome thermal factors to cause phase-separation to occur. The kind of phase separation on observes is the formation of islands of adsorbed molecules separated by regions where the surface has little or no adsorbed molecules.

There is another area where this kind of lattice model is widely used. When studying magnetic materials one often uses the lattice model to describe the interactions among pairs of neighboring spins (e.g., unpaired electrons on neighboring molecules or nuclear spins on neighboring molecules). In this application, one assumes that up or down spin states are distributed among the lattice sites, which represent where the molecules are located. N_α and N_β are the total number such spins, so $(N_\alpha - N_\beta)$ is a measure of what is called the net magnetization of the sample. The result of applying the Bragg-Williams approximation in this case is that one again observes a critical condition under which strong spin pairings occur. In particular, because the interactions between α and α spins, denoted $-J$, and between α and β spins, denoted $+J$, are equal and opposite, the X variable characteristic of all lattice models reduces to:

$$X = -2E_{\alpha,\beta} + E_{\alpha,\alpha} + E_{\beta,\beta} = -4 \, J.$$

The critical condition under which one expects like spins to pair up and thus to form islands of α-rich centers and other islands of β-rich centers is

$$-4 \, cJ/kT < -4$$

or

$$cJ/kT > 1.$$

7.5.4 Virial Corrections to Ideal-Gas Behavior

Recall from our earlier treatment of classical partition function that one can decompose the total partition function into a product of two factors:

$$Q = \{h^{-NM} (N!)^{-1} \int \exp (- H^0(y, p)/kT) \, dy \, dp \; \{\int \exp (-U(r)/kT) \, dr\}$$

one of which

$$Q_{ideal} = h^{-NM} (N!)^{-1} \int \exp (- H^0(y, p)/kT) \, dy \, dp \; V^N$$

is the result if no intermolecular potentials are operative. The second factor

$$Q_{inter} = (1/V^N) \, \{\int \exp (-U(r)/kT) \, dr\}$$

thus contains all of the effects of intermolecular interactions. Recall also that all of the equations relating partition functions to thermodynamic properties involve taking $\ln Q$ and derivatives of $\ln Q$. So, all such equations can be cast into sums of two parts; that arising from $\ln Q_{ideal}$ and that arising from $\ln Q_{inter}$. In this Section, we will be discussing the contributions of Q_{inter} to such equations.

The first thing that is done to develop the so-called cluster expansion of Q_{inter} is to assume that the total intermolecular potential energy can be expressed as a sum of pair wise additive terms:

$$U = \Sigma_{I<J} \, U(r_{IJ})$$

where r_{IJ} labels the distance between molecule I and molecule J. This allows the exponential appearing in Q_{inter} to be written as a product of terms, one for each pair of molecules:

$$\exp(-U/kT) = \exp(- \Sigma_{I<J} U(r_{IJ})/kT) = \Pi_{I<J} \exp(- U(r_{IJ})/kT).$$

Each of the exponentials $\exp(- U(r_{IJ})/kT)$ is then expressed as follows:

$$\exp(- U(r_{IJ})/kT) = 1 + (\exp(- U(r_{IJ})/kT) -1) = 1 + f_{IJ},$$

the last equality being what defines f_{IJ}. These f_{IJ} functions are introduced because, whenever the molecules I and J are distant from one another and thus not interacting, $U(r_{IJ})$ vanishes, so $\exp(- U(r_{IJ})/kT)$ approaches unity, and thus f_{IJ} vanishes. In contrast, whenever molecules I and J are close enough to experience strong repulsive interactions, $U(r_{IJ})$ is large and positive, so f_{IJ} approaches -1. These properties make f_{IJ} a useful measure of how molecules are interacting; if they are not, $f = 0$, if they are repelling strongly, $f = -1$, and if they are strongly attracting, f is large and positive.

Inserting the f_{IJ} functions into the product expansion of the exponential, one obtains:

$$\exp(-U/kT) = \Pi_{I<J} (1 + f_{IJ}) = 1 + \Sigma_{I<J} f_{IJ} + \Sigma_{I<J} \Sigma_{K<L} f_{IJ} f_{KL} + \ldots$$

which is called the cluster expansion in terms of the f_{IJ} pair functions. When this expansion is substituted into the expression for Q_{inter}, we find:

$$Q_{inter} = V^{-N} \int (1 + \Sigma_{I<J} f_{IJ} + \Sigma_{I<J} \Sigma_{K<L} f_{IJ} f_{KL} + \ldots) \, dr$$

where the integral is over all 3N of the N molecule's center of mass coordinates.

The integrals involving only one f_{IJ} function are all equal (i.e., for any pair I, J, the molecules are identical in their interaction potentials) and reduce to:

$$N(N-1)/2 \; V^{-2} \int f(r_{1,2}) \; dr_1 \; dr_2.$$

The integrals over $dr_3 \ldots dr_N$ produce V^{N-2}, which combines with V^{-N} to produce the V^{-2} seen. Finally, because $f(r_{1,2})$ depends only on the relative positions of molecules 1 and 2, the six dimensional integral over $dr_1 \; dr_2$ can be replaced by integrals over the relative location of the two molecules r, and the position of their center of mass R. The integral over R gives one more factor of V, and the above cluster integral reduces to

$$4\pi \; N(N-1)/2 \; V^{-1} \int f(r) \; r^2 \; dr.$$

with the 4π coming from the angular integral over the relative coordinate r. Because the total number of molecules N is very large, it is common to write the $N(N-1)/2$ factor as $N^2/2$.

The cluster integrals containing two $f_{IJ} \; f_{KL}$ factors can also be reduced. However, it is important to keep track of different kinds of such factors (depending on whether the indices I, J, K, L are all different or not). For example, terms of the form

$$V^{-N} \int f_{IJ} \; f_{KL} \; dr_1 \; dr_2 \ldots dr_N \text{ with I, J, K, and L all unique}$$

reduce (again using the equivalence of the molecules and the fact that f_{IJ} depends only on the relative positions of I and J) to:

$$1/4 \; N^4 \; (4\pi)^2 \; V^{-2} \int f_{12} \; r_{12}{}^2 \; dr_{12} \int f_{34} \; r_{34}{}^2 \; dr_{34},$$

where, again I used the fact that N is very large to replace $N(N-1)/2 \; (N-2)(N-3)/2$ by $N^4/4$.

On the other hand, cluster integrals with, for example, I=K but J and L different reduce as follows:

$$V^{-N} \int f_{12} \; f_{13} \; dr_1 \; dr_2 \ldots dr_N = 1/2 \; V^{-3} N^3 \int f_{12} \; f_{13} \; dr_1 \; dr_2 \; dr_3.$$

Because f_{12} depends only on the relative positions of molecules 1 and 2 and f_{13} depends on the relative positions of 1 and 3, the nine-dimensional integral over $dr_1 \, dr_2 \, dr_3$ can be changed to a six-dimensional integral over $dr_{12} \, dr_{13}$ and an integral over the location of molecule 1; the latter integral produces a factor of V when carried out. Thus, the above cluster integral reduces to:

$$(4\pi)^2 \, 1/2 \, V^{-2} N^3 \int f_{12} \, f_{13} \, r_{12}^2 \, r_{13}^2 \, dr_{12} \, dr_{13} \, .$$

There is a fundamental difference between cluster integrals of the type $f_{12} \, f_{34}$ and those involving $f_{12} \, f_{13}$. The former are called unlinked clusters because they involve the interaction of molecules 1 and 2 and a separate interaction of molecules 3 and 4. The latter are called linked because they involve molecule 1 interacting simultaneously with molecules 2 and 3 (although 2 and 3 need not be close enough to cause f_{23} to be non-zero). The primary differences between unlinked and linked cluster contributions are:
1. The total number of unlinked terms is proportional to N^4, while the number of linked terms is proportional to N^3. This causes the former to be more important than the latter because they are more numerous.
2. The linked terms only become important at densities where there is a significant probability that three molecules occupy nearby regions of space. The unlinked terms, on the other hand, do not require that molecules 1 and 2 be anywhere near molecules 3 and 4. This also causes the unlinked terms to dominate especially at low and moderate densities.
I should note that a similar observation was made in Chapter 6 when we discussed the configuration interaction and coupled-cluster expansion of electronic wave functions. That is, we noted that doubly excited configurations (analogous to f_{IJ}) are the most important contributions beyond the single determinant, and that quadruple excitations in the form of unlinked products of double excitations were next most important, not triple excitations. The unlinked nature in this case was related to the amplitudes of the quadruple excitations being products of the amplitudes of two double excitations. So, both in electronic structures and in liquid structure, one finds that pair correlations followed by unlinked pair correlations are the most important to consider.

Clearly, the cluster expansion approach to Q_{inter} can be carried to higher and higher-level clusters (e.g., involving $f_{12}f_{34}f_{56}$ or $f_{12}f_{13}f_{34}$, etc.). Generally, one finds that the unlinked terms (e.g., $f_{12}f_{34}f_{56}$ in this example) are most important (because they are proportional to higher powers of N and because they do not require more than binary collisions). It is most common, however, to employ a severely truncated expansion and to retain only the linked terms. Doing so for Q_{inter} produces at the lower levels:

$$Q_{inter} = 1 + 1/2 \ (N/V)^2 \ 4\pi \ V \int f \ r^2 \ dr + 1/4 \ (N/V)^4 \ [4\pi \ V \int f \ r^2 \ dr \ ]^2$$

$$+ 1/2 \ (N/V)^3 \ V \ (4\pi)^2 \int f_{12} f_{13} \ r_{12}^2 r_{13}^2 \ dr_{12} \ dr_{13}.$$

One of the most common properties to compute using a partition function that includes molecular interactions in the cluster manner is the pressure, which is calculated as:

$$p = kT \ (\partial \ln Q / \partial V)_{N,T}.$$

Using $Q = Q_{ideal} Q_{inter}$ and inserting the above expression for Q_{inter} produces the following result for the pressure:

$$pV/NkT = 1 + B_2 \ (N/V) + B_3 \ (N/V)^2 + \dots$$

where the so-called virial coefficients B_2 and B_3 are defined as the factors proportional to (N/V) and $(N/V)^2$, respectively. The second virial coefficient's expression in terms of the cluster integrals is:

$$B_2 = - \ 2\pi \int f \ r^2 \ dr = - \ 2\pi \int [\exp(-U(r)/kT) -1] \ r^2 \ dr.$$

The third virial coefficient involves higher order cluster integrals.

The importance of such cluster analyses is that they allow various thermodynamic properties (e.g., the pressure above) to be expressed as one contribution that would occur if the system consisted of non-interacting molecules and a second contribution that arises from the intermolecular forces. It thus allows experimental measurements of the deviation from ideal (i.e., non-interacting) behavior to provide a direct way to determine intermolecular potentials. For example, by measuring pressures at various N/V values and various temperatures, one can determine B_2 and thus gain valuable information about the intermolecular potential U.

7.6 Chapter Summary

In this Chapter, you were introduced to many of the main concepts and methods of statistical mechanics. You should be familiar with the following topics by now

a. Microcanonical, canonical, and grandcanonical ensembles and their partition funcitons.

b. Ensemble averages being equal to long-time averages; the equal a priori postulate.

c. Fluctuations

d. Expressions for thermodynamic properties in terms of partition functions.

e. Monte Carlo methods including Metropolis sampling and umbrella sampling.

f. Molecular dynamics simulations, including molecular mechanics force fields.

g. Coarse graining methods.

h. Time correlation functions.

i. Einstein and Debye models for solids' phonons.

j. Lattice theories of adsorption, liquids, and phase transitions.

k. Virial expansions of thermodynamic properties.