Chapter 7. Statistical Mechanics

When one is faced with a condensed-phase system, usually containing many molecules, that is 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 of all of the constituent molecules. Instead, the powerful tools of statistical mechanics allow one to focus on quantities that describe the most important features of the many-molecule system. In this Chapter, you will learn about these tools and see some important examples of their application.

I. Collections of Molecules at or Near Equilibrium

As noted 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, 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. 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. This is where the power of statistical mechanics comes into play.

A. The Distribution of Energy Among Levels

One of the most important concepts of statistical mechanics involves how a specified amount of total energy E can be shared among a collection of molecules and among the internal (translational, rotational, vibrational, electronic) degrees of freedom of these molecules. 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 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 j^{th} 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 M coordinates (collectively denoted q) 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

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

Notice that the Boltzmann formula does not say that only those states of a given 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).

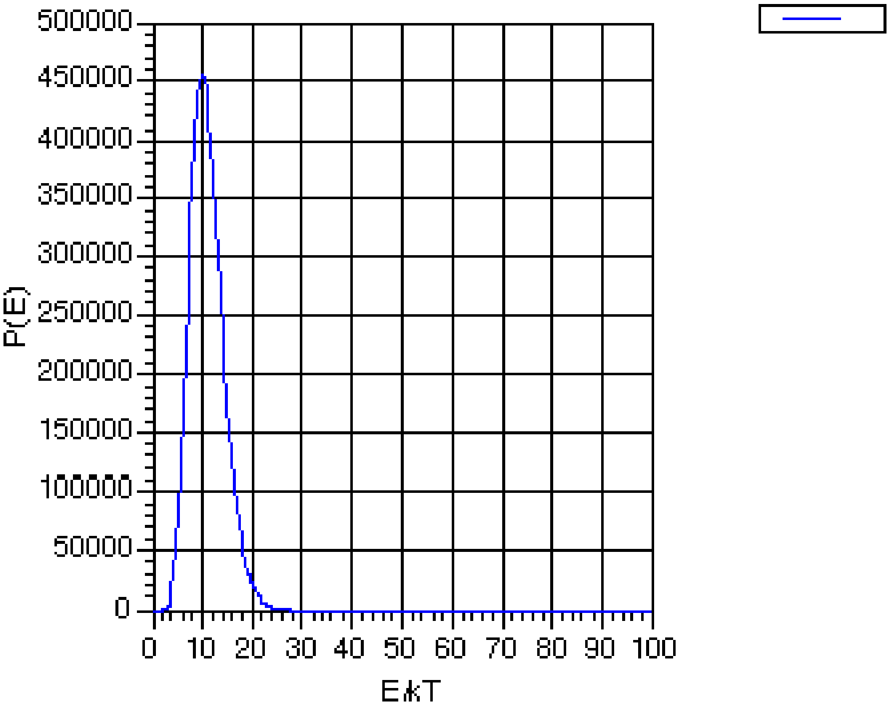


Figure 7.1 Probability Weighting Factor P(E) as a Function of E for K = 10.

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

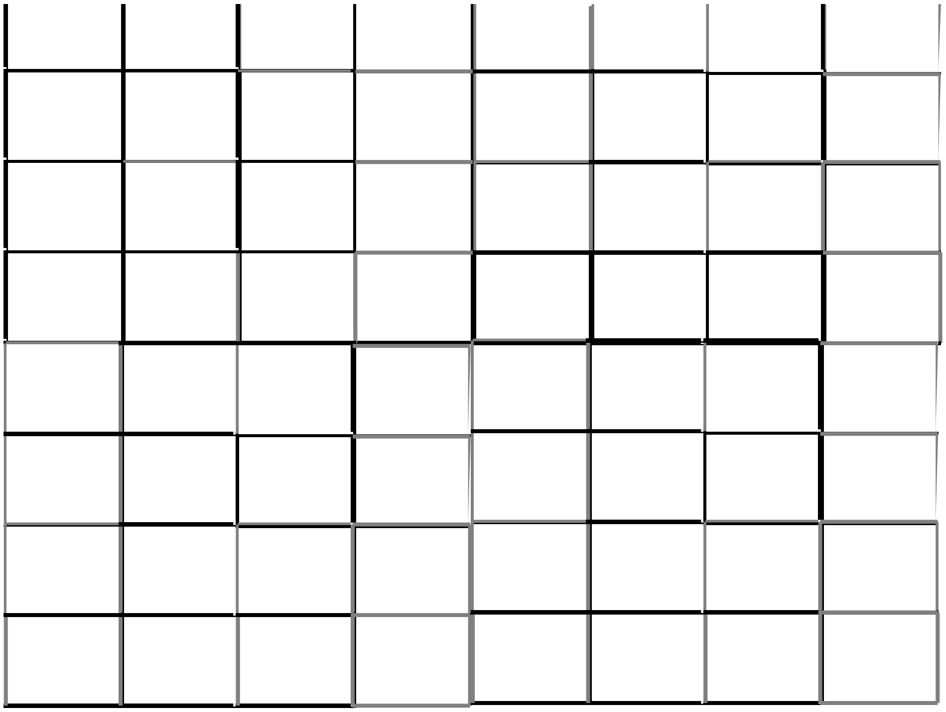
$$(E'-E^*)/E^* = (2/K)^{1/2}.$$

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, the behavior of systems with many molecules by focusing on the most probable situation (i.e., having the energy E*) and ignoring deviations from this case.

It is for the reasons just shown that for so-called macroscopic systems near equilibrium, in which N (and hence K) is extremely large (e.g., $N \sim 10^{10}$ 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.

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 but material that does not allow 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 Cell.

One of the fundamental assumptions of statistical mechanics is that, for a closed isolated system at equilibrium, all quantum states of the system having an 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.

In particular, 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 first energy state of N molecules in the container of volume V, J=2 labels the second such state, and so on. I understand that 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.

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 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 $\{\varepsilon_\phi\}$ 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, … J, …. 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), 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 Section. 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 lnΩ has its maximum, so we can maximize ln Ω 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 ln X! = X ln X – X to approximate ln Ω 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 Ω.

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 lnΩ function we wish to maximize whenever the $\{n_J\}$ variables obey the two constraints. So, the maxima of F and of lnΩ 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 will 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:

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

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 using

$$E = M \Sigma_J \varepsilon_J \exp(- \beta \varepsilon_K)/Q$$

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

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, all thermodynamic properties of the N molecules in the container of volume V can be obtained as derivatives of the 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 ranges 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 Ω as a Taylor series in the paramters $\{n_1, n_2, \ldots\}$ 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 lnΩ has been made maximum at $\{\mathbf{n}^*\}$. The first derivative of lnΩ as given above is $\partial\ln\Omega/\partial n_J = -\ln(n_J)$, so the second derivatives needed to complete the Taylor series through second order are:

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

We can thus express $\Omega(\mathbf{n})$ in the neighborhood of $\{\mathbf{n}^*\}$ 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 “optimal” values $\{\mathbf{n}^*\}$.

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 $\{\delta f_J=0\}$ is important as M approaches infinity.

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 nunmber 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:

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

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 β = 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:

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

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}). It is this fact that causes the narrow distribution in energies that we discussed earlier in this section.

B. 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 of thermodynamic properties. The first equation is the fundamental Boltzmann population formula that we already examined:

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

which expresses the probability for finding the N-molecule system in its J^{th} quantum state having energy E_j and degeneracy Ω_j.

Using this result, it is possible to compute the average energy <E> of the system

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

and, as we saw earlier in this Section, 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>:

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

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

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

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

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

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

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; the average entropy is given as

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

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 $\langle (E-\langle E\rangle)^2\rangle$ was shown above to be given by

$$\langle (E-\langle E\rangle)^2\rangle = kT^2\ C_V.$$

The *Statistical Mechanics* text by McQuarrie 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-state 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 as derivatives of lnQ.

What do these partition functions mean? They represent the thermal-average number of quantum states that are accessible to the system. 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 (i) the number of quantum states in the j^{th} energy level multiplied by (ii) 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.

Keep in mind that the energy levels E_j and degeneracies Ω_j 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), each of the energies E_j can be expressed 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, 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 l^{th} 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. 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

$$q = (3.28\ x10^{13}\ mT)^{1/2,2/2,3/2}\ L, A, V.$$

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 j^{th} harmonic vibration of the molecule, of which there are 3N-6.

Electronic:

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

where ε_J and ω_J are the energies and degeneracies of the J^{th} electronic state; the sum is carried out for those states for which the product $\omega_J \exp(-\varepsilon_J/kT)$ is numerically significant. It is conventional to define the energy of a molecule or ion with respect to that of its atoms. So, the first term above 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.

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 is possible to express all of the thermodynamic properties in terms of the classical partition function. 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)

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

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 usually 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 would arise from the integration over all of the translational coordinates if U(r) were absent (i.e., if U =0). The other factor

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

contains all of 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 $\ln Q_{ideal}$ and one coming from $\ln Q_{inter}$. The latter contains all of the effects of the intermolecular interactions. This means that all of the thermodynamic equations can, in this case, 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.

C. 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 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) 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, neglecting the effects of 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. Alternatively, for an isomerization reaction involving the normal (N) and zwitterionic (Z) forms of arginine that were discussed in Chapter 5, 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 thermal-average number of quantum states available to the system at temperature T (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.

D. 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 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 illustrate how this 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 with 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 E 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 E does not include the kinetic energy of the molecules; it is only the potential energy. Usually, this energy E is expressed as a sum of intra-molecular bond-stretching and bending contributions, one for each molecule, plus a pair-wise additive intermolecular potential:

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

However, the energy E could be computed in other ways, if appropriate. For example, E 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 E is computed, but, most commonly, it is evaluated as shown above.

In each "step" of the MC process, this potential energy E is evaluated for the current positions of the N water molecules. In its most common and straightforward implementation, 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 ($q \rightarrow q + \delta q$) and the potential energy E is evaluated at the "new" configuration ($q+\delta q$). The amount δq by which coordinates are varied is usually chosen to make the fraction of MC steps that are accepted (see below) approximately 50%. This has been shown to optimize the performance of the MC algorithm.

Note that, when the inter-molecular energy is pair-wise additive as suggested above, evaluation of the energy change $E(q+\delta q) - E(q) = \delta E$ accompanying the change in q requires computational effort that is proportional to the number N of molecules in the system because only those factors $E(\text{intermolecular})_{J,K}$, with J or K equal to the single molecule that “moved” need be computed. This is why pairwise additive forms for E are often employed.

If the energy change δE is negative (i.e., if the potential energy is lowered by the “move”), 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 δE is positive, the candidate 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 $P = \exp(-\delta E/kT)$ 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 number is greater than P (expressed in the same decimal format), then the move is accepted and included in the list of accepted MC configurations. If the random number is less than P, the move is not accepted. Instead, a new water molecule and its internal or external coordinate are chosen at random and the entire process is restarted.

In this manner, one generates a sequence of “accepted moves” that generate a series of configurations for the system of N water molecules. 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 the molecules' kinetic energy enters the MC process). 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 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.

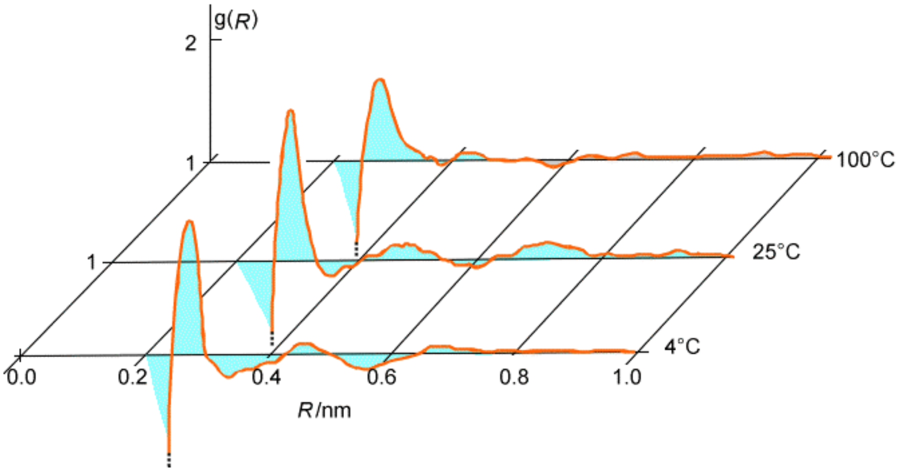


Figure 7.3. Radial Distribution Functions Between Pairs of Oxygen Atoms in H_2O at Three Different Temperatures.

The MC procedure thus 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:

$$\langle A \rangle = \int P(q)\, A(q)\, dq = \{\int \exp(-\beta E(q))\, A(q) dq\}/\int \exp(-\beta E(q)) dq.$$

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 same average is computed by forming the following sum over the M accepted MC configurations q_J:

$$<A> = (1/M)\ \Sigma_J\ A(q_J).$$

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, consider what might be viewed as an alternative procedure. Namely, suppose that the N molecules' 3 translational coordinates are the only variables to be treated (this certainly is a lower limit) and suppose one divides the range of each of these 3N coordinates into only 10 values. To compute an integral such as

$$\int \exp(-\beta E(q))\ A(q)\ dq$$

in terms of such a 10-site discretization of the 3N coordinates would require the evaluation of the following 3N-fold sum:

$$\Sigma_{j1,j2,\ldots j3N}\ A(q_1, q_2, \ldots q_{3N})\ \exp(-\beta E(q_1, \ldots q_{3N})).$$

This sum contains 10^{3N} terms! Clearly, even for N = 6 (i.e., six molecules), the sum would require as much computer effort as the one million MC steps mentioned above, and MC simulations are often performed on thousands and even millions of molecules.

So, how do MC simulations work? 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 10^{1000} or $10^{1,000,000}$ computations? The main thing to understand is that the 10-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 A exp(-βE) is negligible. On the other hand, the MC algorithm is designed to select (as accepted steps) those coordinates for which exp(-βE) 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!

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, suppose that one wanted to use the MC process to compute an average, with the above

$\exp(-\beta E(q))$ 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 reaction.

The problem with using conventional MC methods to compute

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

in such cases is that

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

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

So, there is a competition between P(q) and A(q) that renders the MC procedure ineffective in such cases.

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.

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

$$\langle A\rangle = \int P(q)\, A(q)\, dq = \{\int A(q) \exp(-\beta E(q))\, dq\}/\int \exp(-\beta E(q)\, dq$$

$$= \int (A(q)/U(q))\, (U(q) \exp(-\beta E(q)))\, dq\}/\int (U(q) \exp(-\beta E(q)))\, dq$$

$$/\{\{\int (1/U(q))\, (U(q) \exp(-\beta E(q)))\, dq\}/\int (U(q) \exp(-\beta E(q)))\, dq\}.$$

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(-βE(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(-βE(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 Uexp(-βE) causes the MC process to focus on those coordinates for which both exp(-βE) and U (and hence A) are significant.

E. Molecular Dynamics Simulations of Properties

One thing that the MC process does not address directly is information about 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 a MD calculation, one usually assigns to each of the internal and external coordinates of each of the N molecules an initial amount of kinetic energy (proportional to T). However, whether one assigns this initial kinetic energy equally to each coordinate or not does not matter much because, as time evolves and the molecules interact, the energy becomes more or less randomly shared in any event and eventually properly simulates the dynamics of the equilibrium system. Moreover, one usually waits until such energy randomization has occurred before beginning to use data extracted from the simulation to compute properties. Hence, any effects caused by improper specifications of the initial conditions can be removed.

With each coordinate having its initial velocity $(dq/dt)_0$ and its initial value q_0 specified as above, 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:

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

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

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 $-(\partial E/\partial q)_0$ is the force along the coordianate q at the "initial" geometry q_0. In most modern MD simulations, more sophisticated numerical methods can be used to propagate the coordinates and momenta. However, what I am outlining here provides you with the basic idea of how MD simulations are performed. The forces can be obtained from gradients of a Born-Oppenheimer electronic energy surface if this is computationally feasible. Alternatively, it can be computed from derivatives of an empirical force field. In the latter case, the system's potential energy E is expressed in terms of analytical functions of

i. intramolecular bond lengths, 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.

By applying this time-propagation to all of the coordinates and momenta of the N molecules, one generates a set of "new" coordinates q(t+δt) and new velocities dq/dt(t+δt) appropriate to the system at time t+δt. Using these new coordinates and momenta as q_0 and $(dq/dt)_0$ and evaluating the forces $-(\partial E/\partial q)_0$ at these new coordinates, one can again use the Newton 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.

In Chapter 8, I again discuss using Newtonian dynamics to follow the time evolution of a chemical system. There is a fundamental difference between the situation just described and the case treated in Chapter 8. In the former, one allows the N-molecule system to reach equilibrium (i.e., by waiting until the dynamics has randomized the energy) before monitoring the subsequent time evolution. In the problem of 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 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 initial conditions that represent 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 pairwise 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, one would have to carry out 10^{11} MD steps. This likely would render the simulation not feasible. For such reasons, when carrying out long-time MD simulations, it is common to ignore the high-frequency intramolecular motions by, for example, simply not including these coordinates and momenta in the Netwonian dynamics. Of course, this is an approximation whose consequences must be tested and justified.

In summary, MD simulations are not difficult to implement if one has available a proper representation of the intramolecular and intermolecular potential energy E. 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.

II. 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

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

while the classical mechanical expression is

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

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:

$$C(t) = \Sigma_j <\Phi_j \mid e\bullet\mu \exp(iHt/\hbar)\ e\bullet\mu \exp(-iHt/\hbar) \mid\Phi_j > \exp(-\beta E_j)/Q,$$

for which A and B are both the electric dipole interaction $e\bullet\mu$ between the photon's electric field and the molecule's dipole operator. 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 now requires that we discuss how the time-evolution aspect of this problem is dealt with. The *Statistical Mechanics* text by McQuarrie 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 $\exp(-\beta H(p,q))$ as the probability function for accepting or rejecting initial q and p 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:

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

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 the Background Material, 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 the Background Material'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 $\exp(-\beta H)$ 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 $t = t + \beta \hbar /i$. 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 the Background Material, 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 very large, so each time step $\delta t = t/P$ has a very 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 $(\delta t)^4$, whereas the form used in the Background Material is valid only to order δt^2 . 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 $\{q_j\}$ 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

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

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 the Background Material) 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

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

This integral over p can be carried out analytically to give

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

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

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

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

Why are such multidimensional integrals 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 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 P such terms $\Sigma_{j=1,P}\ [m(q_j - q_{j-1})^2 / 2\delta t - \delta t(V(q_j) + V(q_{j-1}))/2]$ represents the integral over all time from t=0 until t = t of the so-called Lagrangian L = T - V:

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

This time integral of the Lagrangian is called the "action" S in classical mechanics (recall that in the Background Material 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

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

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 j^{th} 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

$$S = \int dtL$$

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

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

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 real 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

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

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.

III. Some Important Chemical Applications of Statistical Mechanics

A. 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:

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

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

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

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

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

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

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

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

Notice that, except for 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.

B. 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 $\varphi(V/N)$ with the 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, $N \varphi/2$, not $N \varphi$ is the total interaction energies 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 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 species in the crystal, with each partition function taken to be of the harmonic vibrational form:

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

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. The $N\varphi/2kT$ 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 give by:

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

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 does not.

So, although the Einstein model offers a very useful model of how a crystal's stability relates to $N\varphi$ 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 $\Omega(\nu)$:

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

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.

C. 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. For example,

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 chemi-sorbed or physi-sorbed 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 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.

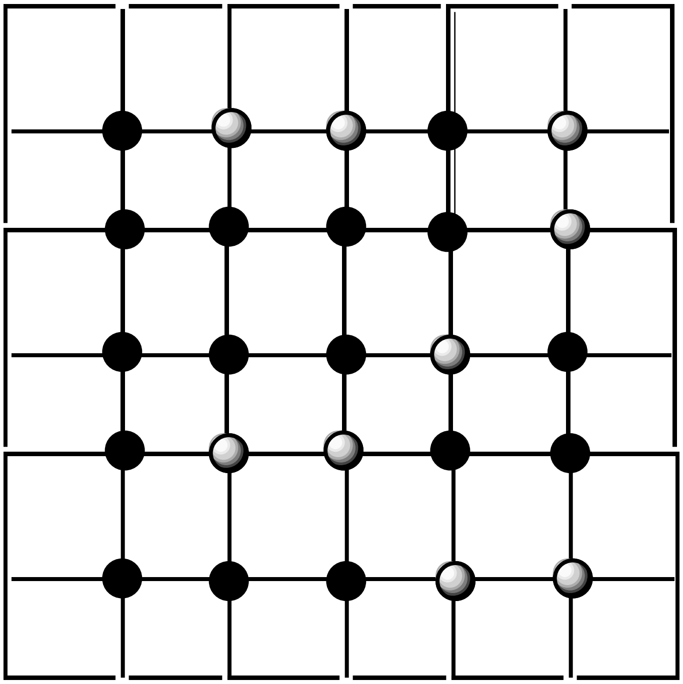


Figure 7.4 Two-dimensional Lattice Having 25 sites With 16 A and 9 B Species

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 pairwise 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 $N_{AB} = 22$.

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:

$$N_{AA} + 1/2\; N_{AB} = cN_A$$

$$N_{BB} + 1/2\; N_{AB} = cN_B.$$

The factor of 1/2 is needed to make sure that one does not double count the AB pairs. Note that the sum of these two equations states the obvious fact that 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 then produces

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

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:

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

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 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 μ_A^0 denotes all of the other factors in μ_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.

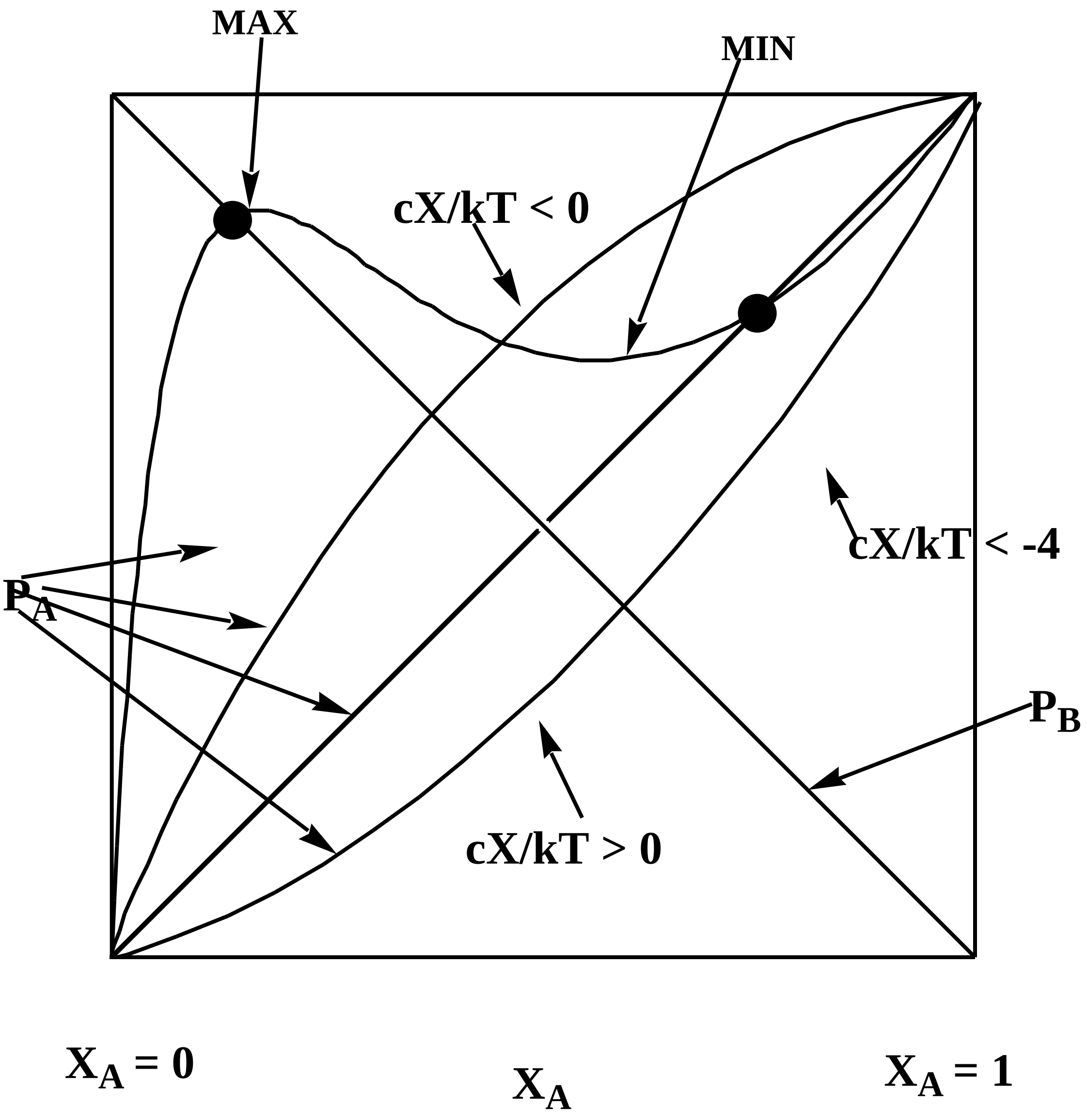


Figure 7.5. Plots of Vapor Pressures in an A, B mixture as Predicted in the Lattice Model With the Bragg-Williams Approximation.

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:

$$X_A = 1/2\ \{1 \pm (1+4kT/cX)^{12}\}.$$

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 $X_A = 1/2\ \{1 \pm (1+4kT/cX)^{12}\}$, which is what we see in Fig. 7.4 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 eth 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 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 $\exp(-\varphi/kT)$) 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_α - N_β) 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

$$cJ/kT > 1.$$

D. 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 lnQ and derivatives of lnQ. 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 pairwise 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:

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

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.

2. The linked terms only become important at densities where there is a significant probability that three molecules occupy nearby regions of space. The linked 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:

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

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.