Week 3 Notes: Statistical Mechanics

Energy Levels of Many-Body Systems

We will look at the foundations of statistical mechanics from a quantum mechanical point of view. This avoids a number of logical problems which arise when attempting to treat statistical mechanics from a purely classical point of view. The classical limit does come out naturally for quantum systems at low density and high temperatures in any case.

Let us consider a simple example of a many body quantum system. For simplicity, let a single particle (which could be an atom, molecule, etc.) with Hamiltonian $\mathcal{H}_1$ have $t$ energy levels in a band from energy 0 to $u$. The average spacing between levels is therefore $u/t$. The inverse of this is called the density of states $g$, so $g = tu$, on average. Label the levels

$$\epsilon^{(1)}_1 = 0, \epsilon^{(2)}_1, \epsilon^{(3)}_1, \ldots, \epsilon^{(t)}_1 = u,$$

for particle number 1. If we have a second particle, also with $t$ levels from 0 to $u$ and no interactions between 1 and 2, the Hamiltonian is $\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2$, and there are $t^2$ levels:

$$\epsilon^{(1)}_1 + \epsilon^{(1)}_2, \epsilon^{(1)}_1 + \epsilon^{(2)}_2, \ldots, \epsilon^{(1)}_1 + \epsilon^{(t)}_2,$$

$$\epsilon^{(2)}_1 + \epsilon^{(1)}_2, \epsilon^{(2)}_1 + \epsilon^{(2)}_2, \ldots, \epsilon^{(2)}_1 + \epsilon^{(t)}_2,$$

$$\epsilon^{(t)}_1 + \epsilon^{(1)}_2, \epsilon^{(t)}_1 + \epsilon^{(2)}_2, \ldots, \epsilon^{(t)}_1 + \epsilon^{(t)}_2.$$

These are in the range from 0 to $2u$. Their average spacing is $2u/t^2$. For $N$ particles, there are $t^N$ levels between 0 and $Nu$. The average spacing is $Nu/t^N$. For a macroscopic system with $N \sim 10^{23}$ this is a fantastically small number. If the particles interact, there is no real change in this conclusion. Interactions can move the levels around, but they do not create or destroy levels. The point is that any thermodynamic system contains very many particles and has an incredibly dense energy level diagram.

Statistical mechanics is built on the idea that the macroscopic properties of large systems are obtained by doing averages over these very dense many-body energy levels. There are different ways to do this. Each different way is called an ensemble. This terminology comes from probability theory. The idea is that probabilities are experimentally determined by repeating an observation many times on identical systems (The members of an ensemble), and plotting a histogram to get the probability of a certain result.

Microcanonical ensemble

Let us now consider a system consisting of many particles which is isolated from its surroundings. It should then have a constant total energy. This brings us to the postulate as to how the properties of this system may be calculated. Let the total Hamiltonian be $\mathcal{H}$, and the eigenfunctions $\Phi_n(1, 2, 3, \ldots, N)$ satisfy $\mathcal{H}\Phi_n = E_n\Phi_n$, 1, 2, 3, $\ldots$ $N$ stands for all the coordinates of the $N$ particles, which may be spatial variables, spin variables, etc. If the system is in the state $\Phi_n$, has total energy $E_n$ and is truly isolated, then it should remain in this state, and the energy should be constant. However, given the very dense set of levels, it makes perhaps more sense to slightly relax the condition of constant energy - after all, the slightest perturbation (caused by, e.g., a measurement) will change the energy level. This leads us to make the postulate that the properties of the system are given by supposing that the system is equally likely to be in any of the states $\Phi_n$ whose energies satisfy $U < E_n < U + \Delta$, where $\Delta$ is an energy which is much smaller than any energy we can detect. This postulate is difficult, in fact probably impossible, to justify in complete generality. It does seem to apply to almost all the systems which will interest us in this course, so we will simply accept it as an experimental fact. Any measurement corresponds to some operator, say $\Theta$. 
The result of the measurement is given by

\[ \overline{\Theta} = \frac{\sum'_{n} \langle \Phi_{n} | \Theta | \Phi_{n} \rangle}{N_{\phi}}. \]

The prime on the summation indicates that \( n \) is such that \( U < E_{n} < U + \Delta \). \( N_{\phi} \) is the number of states which satisfy this condition. The average here is called an average over the microcanonical ensemble. One imagines a collection or ensemble of identical systems, one in each of the states \( \Phi_{n} \). The average corresponds to measuring \( \Theta \) in each member of the ensemble and taking the mean result.

To relate this to actual experiments where measurements are made on a single system, we need some form of the \textit{ergodic hypothesis}, which says in its original form that in a time short compared with the measurement, the system traverses all states available to it, spending on average, a time in each state given by the ensemble probability. In theory as well as in practice, this form of the hypothesis is too strong. Very few model systems are truly ergodic. However, this is usually not needed to justify statistical averages, which only requires that some reasonable sampling of the various states takes place.

There are other ways of writing the above average which are sometimes useful or interesting. For example, think of a single system which is in a state \( \Psi' = \sum_{m} b_{m} \Phi_{m} \), where \( |b_{m}| = N_{\phi}^{1/2} \), if \( U < E_{m} < U + \Delta \) and \( b_{m} = 0 \) otherwise. Then \( \langle \Psi' | \Psi' \rangle = \sum_{m} |b_{m}|^{2} = \sum' N_{\phi}^{-1} = 1 \), so that \( \Psi \) is normalized. Also

\[ \langle \Psi' | \Theta | \Psi' \rangle = \sum_{mn} b_{m}^{*} b_{n} \langle \Phi_{m} | \Theta | \Phi_{n} \rangle. \]

Now suppose that the phases of the \( b_{m} \) are random. Then all terms will cancel except for \( m = n \):

\[ \langle \Psi' | \Theta | \Psi' \rangle = \sum_{m} |b_{m}|^{2} \langle \Phi_{m} | \Theta | \Phi_{m} \rangle = \sum' N_{\phi}^{-1} \langle \Phi_{n} | \Theta | \Phi_{n} \rangle = \overline{\Theta} \]

Thus it is possible to think of the statistical average as like an incoherent quantum average. Another often seen device is the density matrix, defined as

\[ \rho_{mn} = \delta_{mn} |b_{n}|^{2}. \]

Then the average may be written as

\[ \overline{\Theta} = \frac{Tr(\rho \Theta)}{Tr(\rho)}. \]

This equation is exact and no statement is required about random phases. It is also independent of the representation (choice of basis) for \( \rho \) and \( \Theta \). - this is a well-known property of the trace. In the basis where \( \mathcal{H} \) is diagonal, we have
\[
\rho = \begin{pmatrix}
0 & 0 & \cdots & 1 \\
1 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 0 & \cdots & 0
\end{pmatrix}
\]

The diagonal elements change over from 0 to 1 where the energy of the state changes over to a value near \(U\). All off-diagonal elements are zero. In a different representation \(\rho\) might not be diagonal. However, we will always have \(Tr \rho = N_\phi\). (Note that many books normalize \(\rho\) so that \(Tr \rho = 1\), in contrast to what we have done here.)

The level density, \(\omega(E)\), the inverse separation between two adjacent levels, is called \(\omega(E)\). Let us define a dimensionless level density by \(\Gamma(E) = \omega(E) \Delta\). The entropy is now defined as \(S(U, V) = k \log[\Gamma(U)]\), where \(U\) is the internal energy. We take \(V\) to be fixed, which fixes the boundary conditions for the Schrödinger equation above. Only if we do that can we determine the energy levels and hence \(\Gamma(E)\) and \(S(E)\). There is arbitrariness in the definition of \(\Gamma\) because of the arbitrariness in \(\Delta\). This doesn’t really matter. For example, in our model system \(\omega(E) \sim t^N / N\), so that

\[
S(E) \sim k \ln(\Delta t^N / N) = kN \ln t + k \ln(\Delta/N).
\]

\(\Delta\) appears only in the second term and it is utterly negligible compared with the first. Once the second term is neglected, we also see that the entropy is extensive, \(i.e., S\) is proportional to \(N\). This is in accord with the thermodynamic definition of entropy.

The fact that \(S\) is extensive means that if we put a system with entropy \(S\) together with a system of entropy \(S_2\), we expect that the total entropy will be \(S_{tot} = S_1 + S_2\). We can say more than this in fact. Let the two systems exchange energy but in such a way that \(E_{tot} = E_1 + E_2\) is fixed. The systems will exchange energy until they find the most probable state, \(i.e., they will maximize \(\Gamma_1(E_1) \Gamma_2(E_2)\) which is the total number of states between \(E_{tot}\) and \(E_{tot} + 2\Delta\) for the whole combined system. Hence \(\partial[(\Gamma_1(E_1) \Gamma_2(E_{tot} - E_1))] / \partial E_1 = 0\). Or, since the logarithm is a monotonic function

\[
\partial \ln[(\Gamma_1(E_1) \Gamma_2(E_{tot} - E_1))] / \partial E_1 = 0,
\]

so

\[
\frac{\partial \ln[(\Gamma_1(E_1))]}{\partial E_1} - \frac{\partial \ln[(\Gamma_2(E_2))]}{\partial E_2} = 0
\]

or

\[
\frac{1}{k} \left[ \frac{\partial S_1(E)}{\partial E} - \frac{\partial S_2(E)}{\partial E} \right] = 0.
\]

Since \(1/T = (\partial S / \partial E)_V\), we finally have \(T_1 = T_2\). The most probable state is the state where the temperatures are equal. The concept of equilibrium between two systems is now seen to have a microscopic basis. Equilibrium is just the macroscopic state in which the system has the most microscopic states available to it. I have used \(E\) rather than \(U\) here to emphasize that we are considering all states, not just the equilibrium state.
Canonical Ensemble

The microcanonical ensemble, which is the ensemble at fixed energy, is conceptually fairly simple, but usually not very useful. A more common experimental condition is that of complete isolation, but rather than immersion in a heat bath. This is the case of fixed temperature, and the resulting averaging is called the canonical ensemble. In this case, we divide the total system into a large reservoir (the bath) and the experimental system we wish to understand. We assume that the reservoir is much larger than the system, and the reservoir energy $E_r$ is much larger than the system energy $|E_r| >> |E_s|$. The total energy $E_{tot} = E_r + E_s$. The total system may still be thought of as isolated, and will maximize its probability $\Gamma_{tot}(E) = \Gamma_r(E_r)\Gamma_s(E_s)$. Because $\Gamma_r >> \Gamma_s$ and even $\log \Gamma_r >> \log \Gamma_s$, this will amount to maximizing $\Gamma_r(E_r) = \Gamma_r(E - E_s)$. Now

$$\log \Gamma_r(E_r) = \log \Gamma_r(E) - E_s \frac{\partial \log \Gamma_r(E)}{\partial E} = S/k - E_s/kT.$$ #

Hence $\Gamma_r(E_r) = \exp(S/k) \exp(-E_s/kT)$ is the probability distribution. The probability distribution for the system is then $\exp(-E_s/kT)$, and we can then forget about the reservoir.

From the standpoint of the ergodic hypothesis, it may be thought of in this way. The system is traversing all of its eigenstates $\Phi_n$ under the influence of its contact with the reservoir at temperature $T$. It does so at a rate fast enough that it traverses many, many states during the course of any measurement we might want to do. It spends a fraction of time in the state $\Phi_n$ which is proportional to $\exp(-E_s/kT)$. This finally determines the canonical ensemble. The averaging procedure for an operator $\Theta$ is

$$\overline{\Theta} = \sum_n <\Phi_n|\Theta|\Phi_n> \exp(-\beta E_n) \overline{\sum_n \exp(-\beta E_n)} \tag{1}$$

where $\beta = 1/kT$. This gives the expected value of a measurement of $\Theta$. Note that, unlike the microcanonical ensemble, it contains no arbitrary parameters. The density matrix in the energy representation is $\rho_{mn} = \delta_{mn} \exp(-\beta E_n)$. This gives $\overline{\Theta} = \text{Tr}(\rho \Theta)/\text{Tr}(\rho)$, which is valid in any representation (meaning any choice of basis) of $\rho$ and $\Theta$.

A very useful quantity is the partition function $Q = \text{Tr} \rho = \sum_n \exp(-\beta E_n)$. For example, the internal energy

$$U = \mathcal{H} = \frac{1}{Q} \sum_n <\Phi_n|\mathcal{H}|\Phi_n> \exp(-\beta E_n) = \frac{1}{Q} \sum_n E_n \exp(-\beta E_n). \tag{2}$$

Comparing $Q$ and $U$ shows that

$$U = -\left( \frac{\partial \ln Q}{\partial \beta} \right)_V. \tag{3}$$

Always keep in mind that $E_n$ depends on $V$, but not on $T$. A simple relation between $Q$ and $A$ is given by noting that

$$\left( \frac{\partial A}{\partial \beta} \right)_V = A + \beta \left( \frac{\partial A}{\partial \beta} \right)_T = A + TS = U = -\left( \frac{\partial \ln Q}{\partial \beta} \right)_V. \tag{4}$$

Thus we must have $\beta A = -\ln Q$. The Helmholtz free energy is therefore $A = kT \ln Q$. Inverting this, we get $1/Q = \exp(\beta A)$, so the general formula for the average of an operator may be written as

$$\overline{\Theta} = \sum_n \exp[-\beta(E_n - A)] <\Phi_n|\Theta|\Phi_n>.$$ #

These averages appear at first sight to be rather puzzling. We seem to be taking into account enormously high energies for the system (though with low probability) and very low energies (with high probability). However, we know that a real macroscopic system at a reasonable temperature
always has pretty much the same energy. The resolution is that, according to the averaging procedure we have defined, the actual fluctuations for a large system are small. A measure is the root-mean-square deviation:

$$\overline{(\Delta E^2)^{1/2}} = \left(\overline{\mathcal{H}^2} - \overline{\mathcal{H}}^2\right)^{1/2}.$$ #

Now

$$U = \overline{\mathcal{H}} = \sum_n E_n \exp[-\beta(E_n - A)],$$ #

whereas

$$\overline{\mathcal{H}^2} = \sum_n E_n^2 \exp[-\beta(E_n - A)].$$ #

We can use a trick to get \((\Delta E^2)\):

$$\left(\frac{\partial U}{\partial \beta}\right)_V = \sum_n E_n (A - E_n) \exp[-\beta(E_n - A)]$$ #

$$+ \beta \left(\frac{\partial A}{\partial \beta}\right)_V \sum_n E_n \exp[-\beta(E_n - A)]$$ #

$$= \sum_n E_n \left[-E_n + A + \beta \left(\frac{\partial A}{\partial \beta}\right)_V \right] \exp[-\beta(E_n - A)]$$ #

Comparing with the relation between \(A\) and \(U\), we find

$$\left(\frac{\partial U}{\partial \beta}\right)_V = \sum_n (UE_n - E_n^2) \exp[-\beta(E_n - A)]$$ #

$$= U\mathcal{H} - \overline{\mathcal{H}^2} = \mathcal{H}^2 - \overline{\mathcal{H}^2}$$ #

Hence

$$\overline{(\Delta E^2)^{1/2}} = -\left(\frac{\partial U}{\partial T}\right)_V = kT^2 \left(\frac{\partial U}{\partial T}\right)_V = kT^2 C_V.$$ #

Finally, the rms deviation is

$$\overline{(\Delta E^2)^{1/2}} = T(kC_V)^{1/2}.$$ #

The only extensive quantity involved is \(C_V\), proportional to \(N\), the number of particles. Hence

$$\overline{(\Delta E)^2}^{1/2} \sim N^{1/2},$$ #

whereas \(U \sim N\), and the fractional deviation

$$\overline{\frac{(\Delta E^2)^{1/2}}{U}} \sim N^{-1/2}.$$ #

If we were to concentrate on a small piece of the system, just a few particles, we would indeed find that the fractional deviations were large. A system of 100 particles would have deviations of the order of 10%. For real macroscopic systems with \(N \sim 10^{24}\), these are completely negligible.

Some insight into why this is true comes from thinking in general about operators \(\Theta(E)\) which depend only on the energy. (\(\mathcal{H}\) and \(\mathcal{H}^2\) above are examples.) We have \(\overline{\Theta} = \sum_n \Theta(E_n) \exp(-\beta E_n) / \mathcal{Q}\). Now let us define the state density \(\omega(E) = (\sum_n \delta(E - E_n)) / \mathcal{Q}\), and normally \(\omega \sim E^N\) for large \(E\), by analogy to our two-particle model system. Now consider the function

$$p(E) = \omega(E) e^{-\beta E} \sim E^N e^{-\beta E}.$$ #
that weights the states in an average.

$$\ln p(E) \sim N \ln E - \beta E.$$ #

This function has a peak at an energy $E_{\text{peak}}$ satisfying $d \ln p(E) = 0$, which gives $E_{\text{peak}} \sim N/\beta = NkT$. If $\ln p(E)$ has a peak, then $p(E)$ has a very high and sharp peak. So

$$\Theta = \int_0^\infty dE \omega(E) \Theta(E) e^{-\beta E} = \int_0^\infty dE p(E) \Theta(E) \approx \Theta(E_{\text{peak}}) \int_0^\infty dE p(E).$$ #

For practical purposes, only a relatively small range of energies, very close to $E_{\text{peak}}$, actually comes into play in our averages. This means that, again for practical purposes, (meaning when fluctuations can be ignored) the microcanonical (fixed energy) and canonical (fixed temperature) ensembles are equivalent, and we can identify $E_{\text{peak}} = U$, the thermodynamic energy.

## Grand Canonical Ensemble

There is one last ensemble which is very useful, in fact almost essential for our discussion of Fermi and Bose systems. This is the grand canonical ensemble. As in the canonical ensemble, we imagine the system to be in a large heat bath. In the canonical ensemble, however, we always implicitly assumed that the number of particles in the system was a fixed number $N$. The system exchanged energy, but not particles, with the bath. In the grand canonical ensemble, we will allow $N$ to vary - the system exchanges both energy and particles with the bath. This may not sound very physical. Indeed, in many cases, we will just use the grand canonical ensemble as a calculational convenience, when, from a physical point of view, we could just as well use the canonical ensemble. On the other hand, there are certain cases, such as a solid in equilibrium with its vapor, when the system and the bath really do exchange particles. There are also cases where we wish to focus on a small subvolume of the system. We can think of the subvolume as a system in its own right with perfectly permeable walls, if we are willing to allow $N$ to vary. Our thermodynamic quantities should now have $N$ explicitly shown as a variable, i.e., $U = U(S, V, N)$, $P = P(V, T, N)$, etc. The quantum states of the system are now $\Phi_{nN}$ and satisfy

$$\mathcal{H}(N) \Phi_{nN} = E_{nN} \Phi_{nN}.$$ #

The Hamiltonian also depends on the total number of particles. For fixed $N$, $n$ runs over all eigenstates of $\mathcal{H}(N)$. The probability of a state of the combined system and bath is

$$\Gamma_r(E_r, N_r) = \Gamma_r(E_{\text{tot}} - E_{nN}, N_\text{tot} - N)$$

$$= \exp[S_r(E_{\text{tot}} - E_{nN}, N_\text{tot} - N)]$$

$$= \exp\left[ S(E_{\text{tot}}, N_\text{tot}) - E_{nN}\left( \frac{\partial S_r}{\partial E} \right)_{E,V} - N\left( \frac{\partial S_r}{\partial N} \right)_{E,V} \right]$$

Here the derivatives are taken at constant volume and evaluated at $E_{\text{tot}}$. Recall that

$$\left( \frac{\partial S_r}{\partial E} \right)_{N,V} = \frac{1}{T}$$

and define

$$\mu = -\left( \frac{\partial S_r}{\partial N} \right)_{E,V} T.$$ #

$\mu$ is known as the chemical potential. We then have

$$\Gamma = \text{const.} \times \exp(-\beta E_{nN} + \beta \mu N).$$ #

This is not normalized. The normalized probability of a state $\Phi_{nN}$ of the system is

$$\frac{\exp[-\beta(E_{nN} - \mu N)]}{\sum_{nN} \exp[-\beta(E_{nN} - \mu N)]} = \frac{1}{Q} \exp[-\beta(E_{nN} - \mu N)].$$ #
This defines the grand partition function
\[ \mathcal{Q} = \sum_{nN} \exp[-\beta(E_{nN} - \mu N)]. \]

The average measured value of the thermodynamic quantity corresponding to \( \Theta \) is
\[ \Theta = \frac{1}{\mathcal{Q}} \sum_{nN} \exp[-\beta(E_{nN} - \mu N)] < \Phi_{nN} | \Theta | \Phi_{nN} > \]

It is necessary to go back to thermodynamics for a moment at this point. From the definition of \( \mu \), we can write a generalized form of the entropy for variable \( N \):
\[ dS(U, V, N) = \frac{1}{T} dU + \frac{P}{T} dV - \frac{\mu}{T} dN. \]

Similarly,
\[ dU = T dS - P dV + \mu dN \]

From this equation, we may get some insight into the physical meaning of \( \mu \), the chemical potential. A high temperature means that a system will wish to give off heat (or entropy). A high chemical potential means that a system will tend to give off particles. More precisely, if a system at high temperature is placed in thermal contact with a system at lower temperature, heat will flow from high to low temperature. If a system at high chemical potential is placed in diffusive contact with a system at lower chemical potential, particles will flow from high to low chemical potential, until \( n \) is the same on the two sides. Since \( A = U - TS \), and \( G = U - TS + PV \) we have also
\[ dA = -SdT - PdV + \mu dN \] and \( dG = -SdT + VdP + \mu dN. \]

Once we have variable \( N \), then we can have a complete discussion of intensive and extensive quantities. We know that, for example,
\[ U(\alpha S, \alpha V, \alpha N) = \alpha U, \]

since \( U, S, V, N \) are all extensive. (It is a good idea to remember that this is actually an approximation that holds good only as the volume goes to infinity. We saw an exception in the problems when we considered liquid drops that had a contribution to the energy that was proportional to surface area, i.e., \( N^{2/3} \).) This relation puts severe restraints on the possible functional forms of \( U \). Similarly
\[ A(T, \alpha V, \alpha N) = \alpha A \]

puts restrictions on the possible forms of \( A \). But \( G \) is the most interesting:
\[ G(T, P, \alpha N) = \alpha G(T, P, N) \]

The only function that satisfies this is \( G(P, T, N) = Ng(P, T) \), where \( g(P, T) \) is the Gibbs free energy per particle. But \( \mu = (\partial G/\partial N)_{P,T} \), so \( G = N\mu(P,T). \) Hence \( g = \mu. \) The chemical potential is just the same as the Gibbs free energy per particle.

We may also define a further useful potential \( \Omega = A - \mu N \). The natural variables for \( \Omega \) are \( (T, V, \mu) \) because \( d\Omega = -SdT - PdV - Nd\mu \). \( \Omega \) may also be rewritten as \( \Omega = A - \mu N = A - G = -PV. \) In the canonical ensemble, the connection between statistical mechanics and thermodynamics was \( A = -kT \ln Q. \) By a similar deviation, it may easily be shown that \( \Omega = -kT \ln Q. \) The equation of state is therefore given directly by \( PV = kT \ln Q. \)

We are considering a system with fluctuating \( N \), just as in the canonical ensemble we considered a system with fluctuating \( U \). However, the statements we made about the energy fluctuations being small in the canonical ensemble apply to the fluctuations of \( N \) in the grand canonical ensemble. The fractional fluctuations satisfy \( (\Delta N)^2/N \sim N^{-1/2}, \) so \( N \) never deviates substantially from its average value \( \bar{N}. \)