

Physics 3700

Statistical Thermodynamics – Basic concepts

Relevant sections in text: §6.1, 6.2, 6.5, 6.6, 6.7

The Boltzmann Factor

We now return to our analysis of microstates of thermodynamic systems, introducing some powerful new tools and getting some significant new insights. We are beginning a systematic study of *statistical thermodynamics*. We want to consider the statistical behavior of a system in thermal equilibrium with an environment at temperature T . Our system will not need to be macroscopic – we will be able to study individual microsystems, *e.g.*, atoms. Our system will not, in general, be closed — it can exchange energy with its environment. The energy is not going to be specified; instead, the temperature will be fixed. As we shall see, this will mean that many different microstates – including those with different energies – can correspond to the macrostate with the given temperature. We say that the energy “fluctuates”. “Fluctuation” is meant in a statistical sense: If you had many copies of the system all in the same macrostate — all at the same temperature in the current discussion — different microstates (characterized, *e.g.*, by their energy) would occur with various probabilities (to be discussed). This statistical thermodynamic set-up (with temperature T , volume V and number of particles N all fixed) is called the *canonical ensemble*. Our goal in what follows is to give a statistical characterization of microstates of a system at temperature T . As you will see, as a byproduct of this we will get a deeper, more statistical understanding of temperature. Indeed, we will give what is probably the ultimate interpretation of temperature using the methods of statistical thermodynamics, and this will be one of the most important ideas to come out of this course.

To begin, consider a thermodynamic system whose microscopic constituents can be in states labeled by some parameters denoted by s , and characterized by energy $E(s)$. For example, you can consider the Einstein solid, where the elementary constituents are 1-d harmonic oscillators, denoted by $A = 1, 2, \dots, N$, whose states are labeled by $n_A = 0, 1, 2, \dots$ (corresponding to the generic symbol s), with energy $E(n_A) = (n_A + \frac{1}{2})\hbar\omega$. The energy of the microstate (generically denoted by $E(s)$) is then

$$E(n_1, n_2, \dots, n_N) = \sum_{A=1}^N (n_A + \frac{1}{2})\hbar\omega,$$

Another example would be a system consisting of hydrogen atoms. The states of the microsystems – the atoms – are labeled by the energy and angular momentum values for the electron in each atom. For each microsystem there are two states with the lowest

energy (-13.6eV), there are 8 states with the next highest energy value (-3.4eV), differing in their angular momentum values, and so forth. (Note here I am including spin angular momentum states in the counting.) The microstate s is then a list of all energy and angular momentum values for all the atoms. The energy $E(s)$ of the microstate is obtained by adding up all the energies for all the atoms.

Our goal is to find a formula for the probability $P(s)$ that the system microstate s occurs given the system is at temperature T (and volume V and number of particles N). You may think we have already played this game some time ago when we computed multiplicities in various situations. We did indeed play a similar game — but there is a key difference now. Back then we assumed the total energy of the system was fixed — this specified the macrostate — and we computed multiplicities and probabilities accordingly. The statistical set up where the energy is fixed is known as the *micro-canonical ensemble* and is used to model closed systems. Now we are holding the temperature fixed; as mentioned above, we are considering the canonical ensemble. The canonical ensemble is used to treat open systems in thermal equilibrium with their environment.

Energy is still conserved, of course, and the system and environment form a closed system. Thus the total energy of the system *and environment* is fixed. Denote by E_t this total energy. With $E(s)$ the system energy as a function of microstate s and with E_e the environment energy we have

$$E_t = E(s) + E_e.$$

The energy of the system is not fixed, nor is that of the environment, just the total energy. Thus the system and/or environment can have a variety of energies, each with a corresponding multiplicity.

Now, let us suppose we have chosen a given microstate s of the system with corresponding energy $E(s)$. There will then be a variety of microstates s_e of the environment which go with this microstate of the system; these are any environment microstates whose energy is $E_t - E(s)$. All of these system-environment microstates correspond to a single given total energy of the system. Recall that the probability of getting a microstate of a *closed system* – here our system and environment – is proportional to the multiplicity of that microstate. Evidently, with a fixed system microstate s and with a fixed total energy E_t , the probability of finding a given system microstate s with energy $E(s)$ is proportional to the multiplicity $\Omega_e(s)$ of the environment macrostate with energy $E_e = E_t - E(s)$. The proportionality constant can be obtained by normalization and is just the reciprocal of the total number of microstates of the environment with energy as described. We can ignore our ignorance of this normalization constant if we just examine the *relative* probability, for the system to be in microstates s_1 and s_2 , which is:

$$\frac{P(s_2)}{P(s_1)} = \frac{\Omega_e(s_2)}{\Omega_e(s_1)}.$$

Our goal is to use this relative probability relation to find a formula for the probability $P(s)$ purely in terms of the system, making no reference to the environment (except via the temperature).

To begin, a more useful form of the relative probability relationship arises if we express it in terms of the entropy $S_e(s)$ of the environment. Using the relation between multiplicity and entropy

$$\Omega = e^{\frac{S}{k}},$$

we then have

$$\frac{P(s_2)}{P(s_1)} = e^{\frac{1}{k}(S_e(s_2) - S_e(s_1))}.$$

The exponent can be interpreted as $(1/k \text{ times})$ the change in the entropy of the environment corresponding to a change in the microstate of the system from s_1 to s_2 .

We now introduce into the analysis the ideas that the environment is (1) macroscopic and (2) “much bigger” than the system. We do this by assuming that any change of system microstate causes a very small change in the environment state. We can thus write

$$S_e(s_2) - S_e(s_1) \approx dS_e = \frac{1}{T}(dU_e + P_e dV_e - \mu_e dN_e),$$

where we used the thermodynamic identity for dS_e . (We used assumption (1) to allow us to use thermodynamics and we used (2) to justify the infinitesimals.) We will assume that the volume of the system – and hence the environment – are fixed, just as the temperature is fixed. We thus drop the PdV term.* For now we also suppose that the system cannot exchange particles with the environment, so we set $dN_e = 0$.† Thus we have

$$S_e(s_2) - S_e(s_1) \approx \frac{1}{T}dU_e.$$

Since the total energy is fixed – the change in environment energy is equal to *minus* the change in energy of the system:

$$dU_e = -[E(s_2) - E(s_1)].$$

So we have

$$S_e(s_2) - S_e(s_1) = -\frac{1}{T}[E(s_2) - E(s_1)],$$

* Typically the change in environment state does not involve a significant change of volume, and even when it does, the corresponding work is normally negligible compared to the change of energy dU_e . This is consistent with our assumption that the environment is suitably large. We idealize this idea by fixing the volume once and for all as part of the definition of the canonical ensemble

† We will relax this assumption later when we define the *Grand Canonical Ensemble*.

and

$$\frac{P(s_2)}{P(s_1)} = e^{-\frac{1}{kT}(E(s_2)-E(s_1))} = \frac{e^{-\frac{E(s_2)}{kT}}}{e^{-\frac{E(s_1)}{kT}}}.$$

With a little rearrangement we have

$$\frac{P(s_2)}{e^{-\frac{E(s_2)}{kT}}} = \frac{P(s_1)}{e^{-\frac{E(s_1)}{kT}}}.$$

The left hand side is a function of s_2 only, and the right hand side is a function of s_1 only. Since s_1 and s_2 can be independently chosen, each side of the equation above must in fact be independent of s_1 and s_2 . (This is the same argument that we use when solving linear partial differential equations by separation of variables.)

$$\frac{P(s)}{e^{-\frac{E(s)}{kT}}} = \text{constant},$$

where constant means “independent of s ”.

We now almost have a complete formula for the probability:

$$P(s) = (\text{const.})e^{-\frac{E(s)}{kT}}$$

The constant is a normalization factor. We have

$$P(s) = \frac{1}{Z}e^{-\frac{E(s)}{kT}},$$

where, schematically,

$$Z = \sum_s e^{-\frac{E(s)}{kT}}.$$

The exponential is called the *Boltzmann factor*, the normalization constant Z is called the *partition function*,[†] and we have now derived the extremely important *Boltzmann probability distribution*.

To summarize, the probability for finding a microstate s in a system with given (fixed) values of temperature T , volume V and number of particles N is

$$P(s) = \frac{e^{-\frac{E(s)}{kT}}}{\sum_s e^{-\frac{E(s)}{kT}}}.$$

The Boltzmann probability distribution is very important, and remarkably simple. Let us take note of a few of its salient properties.

[†] As we shall see, there is a lot of physics in this normalization factor. The partition function will depend upon temperature (clearly), volume and number of particles (because the energies will). It is in this sense that it is a function.

First, for a given temperature it is a decreasing function of energy. This can be understood by recalling that the probability is proportional to the multiplicity for the *environment*; increasing the system's energy decreases the environment's energy and multiplicity for a macrostate defined by energy generally decreases with that energy. So, smaller energies are more likely than larger energies at any given temperature. But note that *all* energies will have a non-vanishing probability, if they are permitted at all. It is worth a moment of your time to reflect on this. Recall that the temperature of the system being the same as its environment means that there is no spontaneous flow of energy via heat to or from the system. What the Boltzmann probability distribution tell us that the only way this can happen is if *all* energies permitted by the system occur with some non-zero probability.

s Next, we note that the Boltzmann factor is an exponential. Where did that come from? From the exponential relationship between multiplicity and entropy. We also see a new way to think about temperature in this exponential behavior: kT sets the energy scale at which a given state has appreciable probability. As we shall see, states with energies less than kT above the ground state have significant probability, states with energies above the ground state greater than kT have a less significant probability. For this reason kT is sometimes called the thermal energy associated with temperature T . As an example, you can check that the thermal energy kT at room temperature is about $\frac{1}{40}$ eV.

Next, note that the Boltzmann distribution $P(s)$ is a probability distribution for the occurrence of a state s , but the state only enters via its energy $E(s)$. If there is only one state for a given energy then the Boltzmann formula will give the probability for that energy go occur. Usually, however, a given energy will correspond to more than one state – just think of the kinetic energy of a particle.* Such energies are called *degenerate*. States with the same energy have the same probability of occurring. If there are n states with energy E then the probability of E is bigger by a factor of n :

$$P(E) = \frac{n}{Z} e^{-\frac{E}{kT}}$$

Finally, the Boltzmann distribution gives us a deep, statistical way to think about temperature. The temperature is that quantity which, via the Boltzmann formula, gives the probability for the presence of microstates in thermal equilibrium.

Example: thermal excitation of atoms

With virtually no effort at all, the Boltzmann distribution can help us understand how atoms behave when in thermal equilibrium at a given temperature. Let our system

* A free particle moving to the right at speed v has the same energy as a free particle moving to the left with speed v .

consist of a single atom in thermal equilibrium at temperature T with its environment. (Notice that nothing in our derivation of the Boltzmann distribution requires a macroscopic system — only a macroscopic environment.) Let us label the ground state energy by E_1 and the first excited state energy by E_2 . Then the probability for finding E_2 *relative* to the probability for finding E_1 is given by the ratio (the normalization Z cancels):

$$\frac{P(E_2)}{P(E_1)} = \frac{d_2}{d_1} e^{-(E_2-E_1)/kT},$$

where d_2 is the degeneracy of E_2 and d_1 is the degeneracy of E_1 . Note that this relative probability only depends upon the difference between the two energies. This is as it should be; the actual value of each energy depends upon a choice of reference energy (here the energy of an ionized electron at rest), which is arbitrarily chosen. Only the energy differences have any physical significance. For hydrogen, modeled as just a spin 1/2 electron in a Coulomb field, $d_2 = 8$ and $d_1 = 2$, and the energy difference which appears above is given by

$$(E_2 - E_1) = 10.2 \text{ eV}.$$

Let's try some temperatures. First, consider room temperature, $T = 300\text{K}$. We get $kT = .026 \text{ eV}$ and

$$\frac{P(E_2)}{P(E_1)} \sim 10^{-170}.$$

The message here is that, for all practical purposes, you will never find hydrogen in its excited state at room temperature.* By the way, since the states with even higher energy are exponentially lower in probability, you won't find them either. You could check all this by shining light through a dilute hydrogen gas and examining the absorption spectrum. You will mostly just see spectral lines corresponding to transitions from the ground state to excited states.

Now suppose that the hydrogen atom is in a star. For example, in the atmosphere of the sun the temperature is about 5800 K . Now you will find $kT = 0.5 \text{ eV}$ and

$$\frac{P(E_2)}{P(E_1)} = 5 \times 10^{-9}.$$

So about 5 atoms in a billion will be in the first excited state due to thermal excitation – not a lot, but not negligible either. Despite the very high temperature in a star (by terrestrial standards) most of the atoms are in their ground state – just a little “thermal excitation”.

* Of course, this statement means you'll never find an excited state “caused” by “thermal excitation”. Other processes (*e.g.*, collisions, absorption of photons, *etc.*) can put the atoms in an excited state.

Example: Spin $\frac{1}{2}$, *i.e.*, the 2 state system

As you may know, electrons come equipped with intrinsic angular momentum – they have “spin” – of magnitude $\frac{\sqrt{3}}{2}\hbar$. They also have the (remarkable) property that the projection of their angular momentum along any given axis can only take two values: $\pm\frac{\hbar}{2}$. The intrinsic spin angular momentum corresponds to an intrinsic magnetic moment for the electron whose projection along any axis can only take two values $\pm\mu$, where $\mu = e\hbar/2m$. This means that the energy[†] of an electron due to its spin in a given magnetic field with magnitude B will only take two values,

$$E_{\pm} = \pm\mu B,$$

depending upon the (spin) state of the electron. The electron in a magnetic field (from the point of view of just the spin or magnetic moment behavior) is thus an example of a “2-state system”. In certain (para)magnetic materials this energy controls the magnetization of the material.

Let us consider an electron in equilibrium with an environment at temperature T . The electron is the system – and we just focus on its spin behavior. Typically, the electron is part of a solid; the rest of the solid is the environment. As before the system has just one constituent. So, for this system, the microstates are the same as the macrostates; they are determined by the energy E_{\pm} . We now want to find the probability for each of the two states, labeled by E_{\pm} . The partition function is

$$Z = e^{-E_-/kT} + e^{-E_+/kT} = 2 \cosh\left(\frac{\mu B}{kT}\right).$$

The probability distribution is then

$$P(E_{\pm}) = \frac{e^{-E_{\pm}/kT}}{e^{-E_-/kT} + e^{-E_+/kT}}.$$

If we divide top and bottom by $e^{-E_-/kT}$, the Boltzmann factor for the ground state energy, we get

$$P(E_-) = \frac{1}{1 + e^{-(E_+ - E_-)/kT}}, \quad P(E_+) = \frac{e^{-(E_+ - E_-)/kT}}{1 + e^{-(E_+ - E_-)/kT}}.$$

Thus the probabilities can be expressed purely in terms of differences of energies, which you might have expected since this is all that matters in non-gravitational physics. If desired, we could just measure energies relative to the ground state, which amounts to setting the ground state energy to zero. We see from the above formulas that the first excited state has exponentially smaller probability than the ground state. Also note that

[†] The energy of a magnetic moment $\vec{\mu}$ in a magnetic field \vec{B} is $-\vec{\mu} \cdot \vec{B}$.

as $T \rightarrow 0$ the probability for the excited state approaches zero, while that of the ground state approaches unity. As $T \rightarrow \infty$ the probabilities for the ground and excited states both approach $\frac{1}{2}$.

Finally, let us compute the average energy E_{ave} for a spin. To do this we multiply each energy value by its probability and then add them all up.* We have

$$E_{ave} = E_+P(E_+) + E_-P(E_-) = -\mu B \tanh\left(\frac{\mu B}{kT}\right).$$

If you have a collection of spins (*e.g.*, in a solid) at temperature T and you take a “census” of energies for all the spins the average result is what is displayed above. Note that as $T \rightarrow 0$ the average energy is just that of the ground state. As $T \rightarrow \infty$ the average energy approaches zero. Both of these results should make physical sense to you. As $T \rightarrow 0$ the probability for finding an excited state vanishes, so all spins are in their ground state and the average energy is just the ground state energy. As $T \rightarrow \infty$ the two states have equal probabilities to be found and the average of these two is just zero.

Of course, neither $T = 0$ nor $T = \infty$ are physically valid quantities. But notice what sets the energy scales here: the thermal energy is characterized by kT and the magnetic energy is characterized by μB . When the thermal energy is dominating, *i.e.*, $kT \gg \mu B$, we are in the large T regime and the magnetic alignment is suppressed by “thermal fluctuations”. When the magnetic energy is dominating the thermal fluctuations, *i.e.*, $\mu B \gg kT$, we are in the small T regime and magnetic alignment tends to occur. Note that the terminology “fluctuations” here is meant in the statistical sense.

Statistical averages and thermodynamic observables

The Boltzmann distribution applies to any system – microscopic or macroscopic – at a fixed temperature (and volume, and number of particles/constituents). For macroscopic systems, the Boltzmann distribution is the way in which we relate our microscopic model of a system to its thermodynamic properties. The key to connecting the microscopic, statistical description of a system to its macroscopic, thermodynamic description is to identify the bulk thermodynamic observables with corresponding statistical averages. For example, suppose you want to compute the internal energy U of a system consisting of N microsystems (in equilibrium at temperature T with volume V). We *define* this internal energy as the average energy in the canonical ensemble. What this means is we set

$$U \equiv \bar{E} = \frac{1}{Z} \sum_s E(s) e^{-E(s)/kT}.$$

* If it isn’t already obvious to you that this is the right thing to do, it is important that you prove this result to yourself.

Note that the s which appears here is a microstate built from the N constituents. That the right hand side is in fact the average energy can be understood by interpreting the $\frac{1}{Z}e^{-E(s)/kT}$ as the fractional number of systems in the ensemble with microstate s and energy $E(s)$.

More generally, if X is some observable of a macroscopic system we can try to compute it microscopically by first identifying the formula $X = X(s)$ which computes X given the microstate s . Then we define the thermodynamic value of X (in the state specified by V , T , N) via

$$X = \frac{1}{Z} \sum_s X(s) e^{-E(s)/kT}$$

Now, if N is not very large, the internal energy (or whatever observable we are interested in) is not too precisely defined in the sense that it is just an average and there will be large fluctuations from the average in the ensemble. The idea is, though, if $N \gg 1$, *i.e.*, the system is macroscopic, the fluctuations become negligible and the average value for the observable is, in effect, *the* macroscopic value of the observable. We will see an illustration of this below. First let us note a convenient fact.

There is a slick formula for \overline{E} in terms of the partition function. First, recall that

$$Z = \sum_s e^{-E(s)/kT}.$$

We view Z as a function of

$$\beta \equiv \frac{1}{kT},$$

via

$$Z(\beta) = \sum_s e^{-\beta E(s)}.$$

Then it is easy to see that

$$\overline{E} = \frac{1}{Z} \sum_s E(s) e^{-E(s)/kT} = -\frac{\partial \ln(Z)}{\partial \beta}.$$

So, to compute \overline{E} (and hence U in the macroscopic setting) it is sufficient to compute the partition function and take a derivative. This is an illustration of a fact we shall see again and again: everything you need to know is stored in the partition function.

To illustrate this last point, recall that for the single spin in a magnetic field we found

$$Z = 2 \cosh\left(\frac{\mu B}{kT}\right).$$

Since

$$\frac{d}{dx} \cosh(x) = \sinh(x),$$

we have

$$-\frac{d}{d\beta} \ln(Z(\beta)) = -\frac{\mu B}{\cosh(\mu B\beta)} \sinh(\mu B\beta) = -\mu B \tanh(\mu B\beta) = \overline{E},$$

as we should.

Example: paramagnet

As an application of this stuff, we can consider a paramagnet modeled as a collection of N non-interacting magnetic moments $\vec{\mu}_i$, $i = 1, 2, \dots, N$. Assume the magnetic moments come from an electron spin for each microsystem, as described previously. We have two energies possible for each magnetic moment

$$E_{i\pm} = \pm\mu B,$$

where μ is the magnitude of the magnetic moment of the electron. An assignment of a plus or minus to each of the spins determines the microstate of the system. In particular, since the total number of magnetic moments N is fixed, and because $N = N_+ + N_-$, once we know N_+ — how many magnetic moments are “up” — we know the microstate. The energy of the paramagnet in an applied magnetic field *for a given microstate* is then

$$E(N_+) = -\sum_{i=1}^N \vec{\mu}_i \cdot \vec{B} = -\mu B(N_+ - N_-),$$

where N_{\pm} is the number of electrons with spin along/against \vec{B} . (Note: $N_+ + N_- = N$.)

Let's compute the partition function. We are assuming all the spins are distinguishable, *e.g.*, by their location, and for a given number of spins N_+ there are

$$\binom{N}{N_+} = \frac{N!}{N_+!N_-!} = \frac{N!}{N_+!(N - N_+)!}$$

possible microstates.* Summing over all possible microstates, we have

$$Z = e^{-N\mu\beta B} + N e^{-(N-2)\mu\beta B} + \dots + \frac{N!}{N_+!N_-!} e^{(N_+-N_-)\mu\beta B} + \dots + N e^{(N-2)\mu\beta B} + e^{N\mu\beta B}$$

The first term is the microstate with $N_- = N$. The second term is the microstate with $N_- = N - 1$, and so forth. Recall that the fundamental meaning of the binomial coefficient is to define the binomial formula:

$$(a+b)^N = \binom{N}{0} a^N + \binom{N}{1} a^{N-1} b + \dots + \binom{N}{M} a^{N-M} b^M + \dots + \binom{N}{N-1} a b^{N-1} + \binom{N}{N} b^N$$

* This is just the binomial coefficient expression we studied some time ago in the context of coin tosses.

If we identify

$$a \leftrightarrow e^{-\mu\beta B}, \quad b \leftrightarrow e^{\mu\beta B},$$

then you can see that

$$Z = (e^{-\mu\beta B} + e^{\mu\beta B})^N = [2 \cosh(\mu\beta B)]^N.$$

Note that this is just the partition function for a single spin (computed earlier) raised to the N^{th} power! This can be understood *a priori*, as will be explained below.

It is now not hard to compute the average energy

$$\overline{E} = -\frac{\partial \ln(Z)}{\partial \beta} = -N\mu B \tanh(\beta\mu B).$$

Note that this is just N times the average energy for a single spin. This kind of result will hold more generally, as will be explained below.

As mentioned before, in an ensemble of systems all at temperature T , the energy will have a statistical fluctuation about this average value. We expect this fluctuation to become negligible for a macroscopic system. Let's see how this happens for the paramagnet. The simplest measure of the statistical spread is given by the variance ΔE^2 which is the difference between the average of the square $\overline{(E^2)}$ and the square of the average \overline{E}^2 . We have just computed

$$\overline{E} = -N\mu B \tanh(\beta\mu B).$$

As you can easily check we have that

$$\Delta E^2 = \overline{(E^2)} - \overline{E}^2 = \frac{1}{Z} \frac{\partial^2 Z}{\partial \beta^2} - \frac{1}{Z^2} \left(\frac{\partial Z}{\partial \beta} \right)^2.$$

A simple computation then shows:

$$\Delta E^2 = \frac{N\mu^2 B^2}{\cosh^2(\mu\beta B)}.$$

Is this large or small? It depends – note that the variance has units! To see how big it is, we compare it to the average value:

$$\frac{\Delta E^2}{\overline{E}^2} = \frac{1}{N \sinh^2(\mu\beta B)}.$$

Clearly, for large enough N the fluctuations from the mean value become negligible. Thus, in the macroscopic limit the average energy in the canonical ensemble is, in effect, *the* energy and it makes sense to identify $U = \overline{E}$.

We can compute other observables of interest using a similar analysis. For example, denote by μ_i the i^{th} magnetic moment $\vec{\mu}_i$ projected along the direction of \vec{B} . This value

is $\mu_i = \pm\mu$ depending upon the state, of course. For the i^{th} energy we have $E_i = -\mu_i B$. The total magnetization of the material in a given microstate — specified by fixing values for each μ_i — is then computed via

$$M_i = \sum_{i=1}^N \mu_i = \mu(N_+ - N_-).$$

Note the energy of a microstate is just $(-B)$ times the magnetization of the microstate,

$$E(s) = -\mu B(N_+ - N_-).$$

We have computed U via

$$U = \bar{E} = \sum_s E(s)P(s).$$

We can compute the macroscopic magnetization M — the statistical average of the magnetization — by simply dividing the internal energy by $(-B)$:

$$M = -\frac{1}{B} \sum_s E(s)P(s) = -\frac{U}{B}.$$

We thus have

$$M = N\mu \tanh(\beta\mu B).$$

This is a really nice formula because it can be easily compared with experiment. For materials which can be understood as getting their magnetization from a single electron per atom in a zero orbital angular momentum state, the agreement is quite good except at low temperatures where other effects come into play (to be discussed later). You can see that as T gets very small the magnetization approaches the value $N\mu$. As T gets large the magnetization approaches zero. Notice that “large” temperatures are those where $kT \gg \mu B$. In this case we have

$$\tanh(\beta\mu B) \approx \beta\mu B,$$

so that

$$M \approx \frac{N\mu^2 B}{kT}.$$

The fact that M is proportional to $1/T$ (for large enough temperatures) was discovered empirically a long time ago and is known as “Curie’s Law”.

Partition function for composite systems

In our last example we noted a close relationship between the partition function and average energy of a single microsystem and the partition function and average energy of a

system composed of N (distinguishable) such microsystems. This relationship is not too hard to understand in general. First we consider the partition function.

Let a system consist of N microsystems which are identical but distinguishable. Each microsystem has a state, say, s_i , $i = 1, \dots, N$ and a specification of these states, (s_1, s_2, \dots, s_N) , defines a microstate of the composite system. *If the microsystems are not interacting*, we can write the energy of the total system $E(s)$ in terms of the energies $\epsilon(s_i)$ of the microsystems via

$$E(s) = \sum_i \epsilon(s_i).$$

Thus the partition function can be written as

$$\begin{aligned} Z &= \sum_{s_1, s_2, \dots, s_N} e^{-\beta E(s)} \\ &= \sum_{s_1, s_2, \dots, s_N} e^{-\beta \sum_i \epsilon(s_i)} \\ &= \left(\sum_{s_1} e^{-\beta \sum_i \epsilon(s_1)} \right) \left(\sum_{s_2} e^{-\beta \sum_i \epsilon(s_2)} \right) \dots \left(\sum_{s_N} e^{-\beta \sum_i \epsilon(s_N)} \right) \\ &= \left(\sum_{s_1} e^{-\beta \sum_i \epsilon(s_1)} \right)^N. \end{aligned}$$

Thus for non-interacting microsystems the partition function for the composite system is just the product of the individual microsystem partition functions.

From this it follows that the average energy for the whole system is just N times the average energy of a microsystem:

$$U = -\frac{\partial \ln(Z)}{\partial \beta} = -N \frac{\partial}{\partial \beta} \ln \left(\sum_{s_1} e^{-\beta \sum_i \epsilon(s_1)} \right) = N \frac{\sum_{s_1} \epsilon(s_1) e^{-\beta \sum_i \epsilon(s_1)}}{e^{-\beta \sum_i \epsilon(s_1)}}$$

This kind of result generalizes to other thermodynamics observables: the value of the observable can be computed as N times the average value of that observable for a single microsystem. Keep in mind though that this kind of result only works insofar as you can approximate the individual microsystems as being independent, *i.e.*, not interacting, so that the energy of the system can be viewed as the sum of energies of the individual microsystems, with no potential energies of interaction occurring.

Distinguishability

We have seen how the partition function for N distinguishable, non-interacting particles is just the N^{th} power the partition function for a single particle.

$$Z = \sum_s e^{-\beta E(s)} = \sum_{s_1} \sum_{s_2} \dots \sum_{s_N} e^{-\beta [E(s_1) + E(s_2) + \dots + E(s_N)]} = Z_1 Z_2 \dots Z_N.$$

Often times the particles are to be viewed as *indistinguishable* and this changes the partition function. To see this, first consider the situation where the system is just two non-interacting, distinguishable particles. The indistinguishability means that the same microstate occurs in the two scenarios: particle 1 is in state i and particle 2 is in state j , versus particle 2 is in state i and particle 1 is in state j . In this case, this same state is erroneously counted twice in the naive sum over states in the partition function. *Assuming that the number of microstates with particle 1 and particle 2 in the same state is negligible* we thus have

$$Z = \frac{1}{2} Z_1 Z_2.$$

Now, we should not divide by two on the terms in the sum over states in which two (or more) particles are in the same 1-particle state. We shall see how to handle the situation soon using quantum theory. For now, we can simply suppose that the number of microstates which have two (or more) particles in the same state is negligible. This is a pretty good approximation for systems at reasonably high temperatures and low densities. At lower temperatures and/or higher densities these kinds of microstates are important and we must use quantum theory to deal with them.

Let us generalize our discussion to N particles. For identical particles, assuming that the number of microstates with 2 or more particles in the same state is negligible (low density and/or high temperature), we have

$$Z = \frac{1}{N!} Z_1 Z_2 \dots Z_N.$$

This is a classical (non-quantum) approximation to the partition function to be used when one is working with identical microsystems.

Equipartition of energy

We are now in a position to understand – indeed, derive – the famous equipartition of energy theorem, which we have been using throughout this course. Let us prove it in a particularly simple case – the 1-d harmonic oscillator – and then indicate the generalization. Our example is, of course, directly relevant to the Einstein model of a solid as a large collection of uncoupled 1-d oscillators.

For a 1-d oscillator the state of the system can be specified via the position and velocity (x, v) . The idea here is that if you know the position and velocity then any other observable can be computed.* Thus the generic symbol “ s ” is, for a 1-d particle, an ordered pair of real numbers, (x, v) . The oscillator energy as a function of the state is

$$E(s) \longrightarrow E(x, v) = \frac{1}{2}mv^2 + \frac{1}{2}kx^2.$$

* This is what it *means* to model a system as a particle moving in 1-d.

Our goal is to compute the average energy,

$$\overline{E} = \frac{1}{Z} \sum_s E(s) e^{-E(s)/kT}, \quad Z = \sum_s e^{-E(s)/kT},$$

of an oscillator in the canonical ensemble. The generic “sum over states” is now a 2-d integral in x - v space. In particular, the partition function is

$$Z = \eta \int_{-\infty}^{\infty} dv \int_{-\infty}^{\infty} dx \exp \left\{ -\beta \left(\frac{1}{2} m v^2 + \frac{1}{2} k x^2 \right) \right\}, \quad \beta = \frac{1}{kT},$$

where η is a normalization constant for phase space integration,** and the average energy is

$$\overline{E} = \frac{1}{Z} \eta \int_{-\infty}^{\infty} dv \int_{-\infty}^{\infty} dx \left(\frac{1}{2} m v^2 + \frac{1}{2} k x^2 \right) \exp \left\{ -\beta \left(\frac{1}{2} m v^2 + \frac{1}{2} k x^2 \right) \right\}.$$

Note that the normalization η drops out for this computation. Both of the integrals (for Z and \overline{E}) are of the Gaussian type, although the partition function is a simpler integral. In fact, to compute \overline{E} we only need to compute Z , since (as noted in general earlier)

$$\overline{E} = -\frac{\partial}{\partial \beta} \ln(Z).$$

We can evaluate the partition function as follows. First we write it as

$$Z = \eta \int_{-\infty}^{\infty} dv \exp \left\{ -\frac{\beta m}{2} v^2 \right\} \int_{-\infty}^{\infty} dx \exp \left\{ -\frac{\beta k}{2} x^2 \right\}.$$

This is the product of two Gaussian integrals. Fortunately, such integrals are well understood. We have (see your text, or any decent calculus book, for a proof):

$$\int_{-\infty}^{\infty} dy e^{-\alpha y^2} = \sqrt{\frac{\pi}{\alpha}}, \quad \alpha > 0.$$

Thus the partition function is

$$Z = \sqrt{\frac{2\pi}{m\beta}} \sqrt{\frac{2\pi}{k\beta}} = \frac{2\pi}{\beta} \frac{\eta}{\sqrt{km}}.$$

The average value of E can now be computed using our derivative formula:

$$\overline{E} = -\frac{\partial \ln(Z)}{\partial \beta} = kT.$$

** Just on dimensional (units) grounds, such a constant is necessary — the sum over states should be a dimensionless process. It can be obtained using the mass of the particle and Planck’s constant, but we won’t need it for what we do here.

Notice – and this is important – each of the integrals (for x and for v) contributed $\frac{1}{2}kT$ to the total.

For a macroscopic system consisting of a large number of non-interacting oscillators we now recover the simple expression for the internal energy:

$$U = N\overline{E} = NkT.$$

As you know, we have been using this formula extensively when analyzing the Einstein model of a solid.

It is now easy (I hope) to see how to generalize these results. Suppose the state of the microsystem is specified by n variables, denoted $w_i = w_1, w_2, w_3, \dots, w_n$. Some of these could be positions, velocities, or other types of observables. Suppose that the energy is a sum of squares of f of these variables, say, the first f of them.

$$E = \sum_{i=1}^f a_i w_i^2 = a_1 w_1^2 + a_2 w_2^2 + \dots + a_f w_f^2.$$

Then the partition function will be of the form*

$$Z = \sqrt{\frac{\pi}{\beta a_1}} \sqrt{\frac{\pi}{\beta a_2}} \cdots \sqrt{\frac{\pi}{\beta a_f}}.$$

The logarithm of Z takes the form

$$\ln(Z) = -\frac{f}{2} \ln \beta + X,$$

where X is all the stuff independent of β . We then have

$$\overline{E} = -\frac{\partial \ln Z}{\partial \beta} = \frac{f}{2\beta} = \frac{f}{2}kT.$$

This is the equipartition result. If the the energy is a sum of squares of f degrees of freedom, then the average energy of the canonical ensemble is $\frac{f}{2}kT$. We know that, for a macroscopic system, $U = N\overline{E}$. So, now we have *derived* the familiar formula

$$U = \frac{f}{2}NkT.$$

You might wonder how the equipartition result can be so useful when almost all physical systems have energies which are not quadratic functions of their basic observables. The answer is that, for many purposes, one can model the constituents of a system as

* As we discussed earlier, if the microsystems are indistinguishable we should divide this result by $\frac{1}{n!}$ (in the low density/high temperature regime).

moving in the vicinity of stable equilibrium. Such motion is always described by an energy function which is quadratic in coordinates and momenta. Systems which cannot be well-approximated by a quadratic energy will not have a useful equipartition description.

I have mentioned in the past that the equipartition result, *e.g.*, for an ideal gas, does not hold if the temperature is lowered sufficiently. The reason for this (to be elaborated later) is that at low temperatures the system becomes quantum mechanical and our Newtonian description is invalid. In the quantum mechanics regime, the microstates are determined in a rather different way than specifying the values for (x, v) for each particle, so the calculation of the partition function changes and the energy formula then changes as well.

Maxwell's speed distribution for an ideal gas

A gas in equilibrium has an internal energy (understood as the average energy of the molecules making up the gas multiplied by the large number of molecules). But this doesn't mean all molecules have the same energy - there is a distribution of energies. In particular, there is a distribution of speeds for molecules in an ideal gas which depends upon the temperature. It is of great interest to understand this distribution in more detail. Here we use the Boltzmann distribution to compute the probability distribution for speeds of particles in an ideal gas. Roughly speaking, we are computing the fraction of particles moving at this or that speed given the temperature. We will ignore all degrees of freedom except translation degrees of freedom of the particles (molecules) making up the gas.

Because the speed is a continuous variable, it is actually not that useful to talk about the probability of finding a particular speed* (except maybe among friends), rather it is useful to talk about the probability for the speed to be in some range of possible values (although that range can be as small as you like). Thus, our goal is to find a function $\mathcal{D}(v)$ such that the probability $P(v_1, v_2)$ for finding speeds between v_1 and v_2 is given by

$$P(v_1, v_2) = \int_{v_1}^{v_2} dv \mathcal{D}(v).$$

$\mathcal{D}(v)$ is called the *probability density*. As with the probabilistic description of any continuous variable (or variables) we can interpret $\mathcal{D}(v)$ as giving the probability $P(v, v + dv)$ for finding the speed in the range $[v, v + dv]$ via

$$P(v, v + dv) = \mathcal{D}(v)dv.$$

We can obtain a formula for $P(v_1, v_2)$ from the Boltzmann distribution as follows (see the text for an alternate approach). The state of a particle in an ideal gas is determined

* The probability for finding a single real number out of a continuous range of real numbers is always zero!

by specifying its position and velocity,

$$s \longleftrightarrow (\vec{x}, \vec{v}).$$

Sums over states are now integrals:

$$\sum_s \longleftrightarrow \eta \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dv,$$

where we have again inserted a normalization factor for phase space integration. The energy of the state (\vec{x}, \vec{v}) is the kinetic energy:

$$E(s) \longleftrightarrow E(x, v) = \frac{1}{2}mv^2.$$

The probability $P(\mathcal{R})$ for finding a particle in a cell \mathcal{R} of size $dx dy dz dv_x dv_y dv_z$ centered about the point (\vec{x}, \vec{v}) in state space is

$$P(\mathcal{R}) = \eta \exp \left\{ -\frac{mv^2}{2kT} \right\} dx dy dz dv_x dv_y dv_z,$$

where η is determined by normalization.

We determine the normalization constant by integrating $P(\mathcal{R})$ over a spatial region of volume V and a velocity region of infinite size and setting the result to unity:

$$1 = \eta \int_V d^3x \int_{-\infty}^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z \exp \left\{ -\frac{mv^2}{2kT} \right\}.$$

This means simply that the gas is in a box of volume V and the probability of finding a molecule somewhere in V with some/any velocity \vec{v} is unity. Because

$$v^2 = v_x^2 + v_y^2 + v_z^2,$$

the velocity integral is just a product of 3 Gaussians. We then get

$$1/\eta = V \left(\frac{2\pi kT}{m} \right)^{3/2}.$$

All together we have

$$P(\mathcal{R}) = \frac{1}{V} \left(\frac{m}{2\pi kT} \right)^{3/2} \exp \left\{ -\frac{mv^2}{2kT} \right\} dx dy dz dv_x dv_y dv_z.$$

Evidently, the probability density for states (\vec{x}, \vec{v}) is given by

$$\mathcal{D}(\vec{x}, \vec{v}) = \frac{1}{V} \left(\frac{m}{2\pi kT} \right)^{3/2} \exp \left\{ -\frac{mv^2}{2kT} \right\}.$$

Note that this density does not depend upon position, nor does it depend upon the direction of velocity. This should make physical sense to you, given our model of the ideal gas. Notice also that the probability density has precisely the right units so that $P(\mathcal{R})$ is dimensionless.

We can get a probability density function for velocity only, $\mathcal{D}(\vec{v})$, by computing the probability that \vec{x} is anywhere in the volume V , but \vec{v} is still in the cell of size $d\vec{v}$ around \vec{v} . This is obtained by integrating over all \vec{x} , which just cancels out the volume V in the formula. We thus get (trivially)

$$\mathcal{D}(\vec{v}) = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left\{-\frac{mv^2}{2kT}\right\}.$$

This probability density function defines a probability for velocities, although it only depends upon the speed, not the direction of the velocity. Thermal equilibrium affects speeds, but not directions of velocity. We can obtain a probability density for speed by computing the probability that the speed is in the range $[v, v + dv]$ but the direction of the velocity is anything. To do this we introduce spherical polar coordinates (v, θ, ϕ) in 3-d velocity space. Here the speed is playing the role of the radius, the polar and azimuthal angles (θ, ϕ) determine the direction of the velocity. Note that the Boltzmann factor depends only on the “radius” in velocity space, not the angles. We have

$$\mathcal{D}(\vec{v}) dv_x dv_y dv_z = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left\{-\frac{mv^2}{2kT}\right\} v^2 \sin\theta dv d\theta d\phi.$$

We want to integrate over all angles to get the probability density function $\mathcal{D}(v)$ for speed:

$$\mathcal{D}(v) = \int_0^{2\pi} d\phi \int_0^\pi d\theta \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left\{-\frac{mv^2}{2kT}\right\} v^2 \sin\theta d\theta d\phi.$$

The angular integrals just provide a factor of 4π . We thus get

$$\mathcal{D}(v) = \left(\frac{m}{2\pi kT}\right)^{3/2} 4\pi v^2 \exp\left\{-\frac{mv^2}{2kT}\right\}.$$

This is the famous Maxwell probability density for speed.

The Maxwell probability density vanishes exponentially at large speeds and vanishes like v^2 as $v \rightarrow 0$. It is, of course, of interest to see where is the maximum, v_{max} of \mathcal{D} . To get this we simply see where the first derivative vanishes:

$$0 = \frac{d\mathcal{D}(v)}{dv} = \left(\frac{m}{2\pi kT}\right)^{3/2} 8\pi v \exp\left\{-\frac{mv^2}{2kT}\right\} \left[1 - \frac{mv^2}{kT}\right] \implies v_{max} = \sqrt{\frac{2kT}{m}}.$$

This peak grows like the square root of temperature. For nitrogen at 300K we have $v_{max} = 422 \text{ m/s}$. There are a couple of other relevant speeds here. The average speed, \bar{v} is given by

$$\bar{v} = \int_0^\infty dv v \mathcal{D}(v) = \int_0^\infty dv \left(\frac{m}{2\pi kT}\right)^{3/2} 4\pi v^3 \exp\left\{-\frac{mv^2}{2kT}\right\} = \sqrt{\frac{8kT}{\pi m}}.$$

It is worth mentioning how you can do such an integral. We have the Gaussian integral result:

$$\int_0^\infty dx e^{-\alpha x^2} = \frac{1}{2} \sqrt{\frac{\pi}{\alpha}}, \quad \alpha > 0.$$

From this you can easily obtain all integrals of the form

$$\int_0^\infty dx x^{2n} e^{-\alpha x^2} = (-1)^n \frac{d}{d\alpha} \left(\frac{1}{2} \sqrt{\frac{\pi}{\alpha}} \right).$$

But we have an odd power of v under the integral sign in our application above. What do we do? Let us first compute

$$\int_0^\infty dx x e^{-\alpha x^2} = \int_0^\infty dx \left(-\frac{1}{2\alpha} \right) \frac{d}{dx} e^{-\alpha x^2} = \frac{1}{2\alpha}.$$

Now we can use our trick:

$$\int_0^\infty dx x^3 e^{-\alpha x^2} = -\frac{d}{d\alpha} \int_0^\infty dx x e^{-\alpha x^2} = \frac{1}{2\alpha^2}.$$

From this formula you can check our result for \bar{v} .

The final characteristic speed we can mention is the root-mean-square speed, v_{rms} . This is obtained by computing the average of v^2 then taking the square root. Using our trick above we have

$$v_{rms}^2 = \int_0^\infty dv v^2 \mathcal{D}(v) = \int_0^\infty dv \left(\frac{m}{2\pi kT} \right)^{3/2} 4\pi v^4 \exp \left\{ -\frac{mv^2}{2kT} \right\} = \frac{3kT}{m},$$

so that

$$v_{rms} = \sqrt{\frac{3kT}{m}}.$$

Note that the root-mean-square speed is that corresponding to the average kinetic energy, \bar{K} :

$$\bar{K} = \frac{1}{2} m v_{rms}^2,$$

which we got via the equipartition theorem.

To summarize:

$$v_{max} = \sqrt{\frac{2kT}{m}}, \quad \bar{v} = \sqrt{\frac{8kT}{\pi m}}, \quad v_{rms} = \sqrt{\frac{3kT}{m}}.$$

Note that

$$v_{rms} > \bar{v} > v_{max}.$$

The most fundamental application of $\mathcal{D}(v)$ is to calculate the probability $P(v_1, v_2)$ that a molecule has its speed in some range $[v_1, v_2]$,

$$P(v_1, v_2) = \int_{v_1}^{v_2} dv \mathcal{D}(v).$$

Such integrals do not have a useful closed form expression; you should evaluate them numerically. For example, the probability $P(v > 1000 \text{ m/s})$ that a nitrogen molecule has a speed higher than, say, 1000 m/s at 300K. We have

$$P(v > 1000 \text{ m/s}) = \int_{1000}^{\infty} dv \left(\frac{m}{2\pi kT} \right)^{3/2} 4\pi v^2 \exp \left\{ -\frac{mv^2}{2kT} \right\},$$

where we still need to substitute in for m , k and T . But I have already told you that for nitrogen at 300 K

$$v_{max} = \sqrt{\frac{2kT}{m}} = 422 \text{ m/s}.$$

Set

$$x = \frac{v}{v_{max}} = \sqrt{\frac{m}{2kT}} v,$$

and change variables in the usual way you learned in calculus. Then the integral is*

$$P(v > 1000 \text{ m/s}) = \int_{2.37}^{\infty} dx x^2 e^{-x^2} = .011.$$

So, about 1% of nitrogen molecules exceed 1000 m/s.

Partition function and free energy

We have already seen how the partition function for a macroscopic system Z is related — indeed can be used to compute — the internal energy:

$$U = -\frac{\partial \ln(Z)}{\partial \beta}, \quad \beta = \frac{1}{kT}.$$

An even more fundamental link between the microscopic, statistical model of a macroscopic system and its thermodynamic observables comes via the relation between the partition function and the Helmholtz free energy F . Indeed, we will now demonstrate the fundamental result:

$$Z = e^{-\beta F}.$$

Recall we are in the canonical ensemble, characterized by fixed values for number of particles N , volume V and temperature T (or β). The partition function depends upon these variables and we can consider its mathematical response to an infinitesimal change in these variables (for simplicity we hold N fixed):

$$d \ln Z = \frac{\partial \ln Z}{\partial \beta} d\beta + \frac{\partial \ln Z}{\partial V} dV = -U d\beta + \frac{\partial \ln Z}{\partial V} dV.$$

* As a general rule, numerical evaluations are always facilitated by working with dimensionless variables, *e.g.*, x in this case.

We can write

$$Ud\beta = d(\beta U) - \beta dU$$

to get

$$d \ln Z = \beta dU - d(\beta U) + \frac{\partial \ln Z}{\partial V} dV.$$

This can be rewritten as

$$dU = \frac{1}{\beta} d(\ln Z + \beta U) - \frac{1}{\beta} \frac{\partial \ln Z}{\partial V} dV,$$

or

$$dU = T d(k \ln Z + \frac{1}{T} U) - T \frac{\partial(k \ln Z)}{\partial V} dV.$$

From the thermodynamic identity we see that

$$k \ln Z + \frac{1}{T} U = S + \text{const.},$$

and

$$\frac{\partial(kT \ln Z)}{\partial V} = P.$$

The first relation will give us the desired result if we can prove the additive constant (which may depend upon N) vanishes. Since the constant is, well, constant, it does not depend upon T and so we can compute the constant by evaluating the expression at a single value of T . Let us consider the limit $T \rightarrow 0$, where we invoke the third law: As $T \rightarrow 0$ we assume there is a single microstate*, and we adjust our zero point of energy so that its energy vanishes. This means that the partition function is unity as $T \rightarrow 0$, that the entropy is zero as $T \rightarrow 0$, and that $U \rightarrow 0$ as $T \rightarrow 0$. This means the constant is zero.†

With this $T \rightarrow 0$ behavior understood, we now have

$$kT \ln Z = -U + TS = -F,$$

i.e.,

$$F = -\frac{1}{\beta} \ln Z$$

or

$$Z = e^{-\beta F}.$$

Note that the partition function gives us $F = F(T, V, N)$. The thermodynamic identity for F ,

$$dF = -SdT - PdV + \mu dN,$$

* This assumption can be generalized.

† Of course, if we don't want to play this game, we can just leave the constant in the relation between Z and F ; its presence just reflects the behavior of the entropy as $T \rightarrow 0$.

says that

$$P = - \left(\frac{\partial F}{\partial V} \right)_{T,N}, \quad S = - \left(\frac{\partial F}{\partial T} \right)_{V,N}, \quad \mu = - \left(\frac{\partial F}{\partial N} \right)_{T,V}.$$

So, once we've computed the partition function — and hence F — we can compute P , S , μ .

As a simple example, let us compute the partition function for a macroscopic Einstein solid. From it we will compute the free energy and the entropy.

The energy quantum for a single oscillator is (as usual) denoted by ϵ . The energies relative to the ground state energy are then $E_n = n\epsilon$, $n = 0, 1, 2, \dots$. The partition function for a single oscillator is

$$Z_1 = \sum_{n=0}^{\infty} e^{-\beta n\epsilon} = \sum_{n=0}^{\infty} \left(e^{-\beta\epsilon} \right)^n = \frac{1}{1 - e^{-\beta\epsilon}}.$$

Here I have used the Taylor series representation

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots, \quad x < 1.$$

The partition function for N independent oscillators is then

$$Z = Z_1^N = \frac{1}{(1 - e^{-\beta\epsilon})^N}$$

The free energy is then (in the macroscopic/thermodynamic limit)

$$F = -\frac{1}{\beta} \ln Z = \frac{N}{\beta} \ln(1 - e^{-\beta\epsilon}).$$

Let us compute the entropy in the high temperature approximation and compare it with what we found from the multiplicity formula. The entropy *in the limit* $kT \gg \epsilon$ (our familiar high temperature limit) is (try it!)

$$S = -\frac{\partial}{\partial T} \left[NkT \ln(1 - e^{-\frac{\epsilon}{kT}}) \right] \approx Nk \ln\left(\frac{kT}{\epsilon} + 1\right), \quad kT \gg \epsilon.$$

Recall that $U = q\epsilon$, and (*e.g.*, from equipartition) we have

$$U = NkT.$$

This gives

$$S = Nk \ln\left(\frac{q}{N} + 1\right),$$

which we computed (from multiplicity) a long time ago.