

**What is Entropy? A handout by Dr Colton for Physics 123 Section 2
a.k.a. “Derivation of the Boltzmann factor and the Maxwell-Boltzmann speed distribution”**

The “Boltzmann factor” is a vitally important expression which tells you how likely states are to be occupied due to thermal energy. It results from the Second Law of Thermodynamics. One formulation of the Second Law states:

Isolated physical systems will tend to be in the macrostate that has the most number of microstates.

For example, if you roll two six-sided dice, you will most likely get a 7 as the sum of the two, because there are more ways of getting a 7 than any other number (1+6, 2+5, 3+4, 4+3, 5+2, and 6+1). So, if you jumble two dice in a box and want to know (without peeking) what state they are in, you can say “They probably add up to a 7”. In thermodynamics terms, picture 10^{23} molecules instead of two dice; the microstates are the positions/velocities of the molecules, the macrostates are the macroscopic variables (P, V, T) that a given microscopic configuration produces.

When you combine two systems, the number of microstates of the combined system is the product of the individual microstates. For example, when you roll one die, there are 6 microstates. When you roll two dice, there are 36 microstates (six of which result in a sum of 7, as enumerated above). When you roll three dice, there are $6 \times 6 \times 6$ microstates. And so forth.

For large systems, the phrase “will tend to be in” becomes “will be extremely close to”. For example, if you roll 10^{23} dice their total will be VERY close to 3.5×10^{23} . There are a HUGE number of ways (# microstates) that that total could be achieved, though. So, when we have large systems, let’s deal with the logarithm of the number of microstates instead of the number of microstates itself. Log functions are very efficient at reducing huge numbers to much more manageable ones. ($\text{Log}_{10}(10^{23}) = 23$, for example.) By convention we will use log base e. Also, for reasons you will see shortly, let’s multiply the log by a constant which has units of Joules/Kelvin. I will call the multiplicative constant simply “constant” for now. We’ll define this new quantity as S, called “entropy”:

$$S = \text{constant} \times \ln(\#\text{microstates}) \quad [\text{units of J/K}]$$

Since S increases and decreases as the # microstates increases and decreases, we can rephrase Second Law as such:

Large isolated physical systems will be extremely close to the state which has the largest S.*

Using the logarithm instead of the # of microstates has this added benefit: when you combine two systems, the entropies of each system ADD. This is because the # microstates multiply, so

$$\begin{aligned} S_{\text{tot}} &= \text{constant} \times \ln(\#\text{microstates1} \times \#\text{microstates2}) \\ &= \text{constant} \times \ln(\#\text{microstates1}) + \text{constant} \times \ln(\#\text{microstates2}) \\ &= S_1 + S_2 \end{aligned}$$

Now, let’s think about two systems which can exchange thermal energy. We’ll suppose that we have a small system which is really our system of interest, which comes to thermal equilibrium with a much larger system (a “thermal reservoir”). For example, the small system could be a bit of material that I’m studying, with the reservoir being the flask of liquid nitrogen I put it in to cool it down. Or the small system could be the gas molecules in my room, with the reservoir being the walls and rest of the building. Let’s also suppose that the systems and the reservoir are isolated from the rest of the world, so that their total energy E is fixed.** Let’s call the small system we’re really interested in “system 1” (with energy E_1), and the reservoir “system 2” (with energy $E_2 = E - E_1$). Since system 1 is much smaller than system 2, E_1 is much less than both E_2 and E.

We want to know the properties of system 1 in equilibrium. Remember that it will be in (or extremely close to) the state that has the largest entropy. For S to be a maximum, dS/dE_1 must be zero.† Let’s write dS/dE_1 a different way:

* When the system reaches equilibrium, that is. One could deliberately set up a system in a state with a smaller entropy, but it would approach the state with the largest entropy as it comes to equilibrium.

** We’ll also assume that no work is done on/by either system, so we don’t need to account for integral of PdV-type factors.

† Just about all of the derivatives in this paper should really be “partial derivatives” which use the symbol ∂ . But since those are likely unfamiliar to most Physics 123 students, I’ll use the regular derivative symbol instead.

$$\begin{aligned}\frac{dS}{dE_1} &= \frac{d}{dE_1}(S_1 + S_2) \\ &= \frac{dS_1}{dE_1} + \frac{dS_2}{dE_1} \\ &= \frac{dS_1}{dE_1} - \frac{dS_2}{dE_2}\end{aligned}$$

The last step happens because $dE_2 = -dE_1$ (which is obvious; just take the differential of the equation $E_2 = E - E_1$, with $E = \text{constant}$).

Therefore, since $dS/dE_1 = 0$ when S is a maximum,

$$\frac{dS_1}{dE_1} = \frac{dS_2}{dE_2}$$

This is interesting! Remember the Joules/Kelvin units we chose for S ? That means that each side of this equation has units of 1/Kelvin. So here we have a quantity which is the same for two systems in thermal contact, which behaves very similarly to the inverse of the temperature! Well, the inverse of the temperature IS the same for two systems in thermal contact (as is the temperature itself, or the temperature squared, etc.), so let's just posit that this quantity is in fact the inverse of the temperature:*

$$\boxed{\frac{dS}{dE} = \frac{1}{T}} \quad \text{Relationship between temperature and entropy}^{**} \quad (1)$$

If we pick our "S-constant" appropriately, this turns out to be exactly true. The proper constant to make this definition of temperature correspond to our usual temperature in Kelvin is called "Boltzmann's constant" and is given the symbol k_B :

$$\boxed{S = k_B \ln(\#\text{microstates})} \quad \text{Definition of entropy} \quad (2)$$

Boltzmann's constant has the experimentally measured value of 1.38×10^{-23} J/K.

Back to the derivation. E_1 could be any small amount of energy that corresponds to an allowed state of system 1. But what if system 1 has multiple energy states it could be in? Let's call them E_{1A} , E_{1B} , etc. How likely is it that the system is in E_{1A} compared to E_{1B} ? If E_{1A} is lower than E_{1B} , it seems logical that system 1 will more likely be in state E_{1A} ... but how much more likely? Can we quantify that? Yes, we can. Let's continue.

In order for system 1 to be in state E_{1A} , it must be using energy E_{1A} that would otherwise be in system 2. This *decreases* the number of states available to system 2; that is, system 2, the thermal reservoir, would have more allowed microstates if it had energy E than it does with energy $E_{2A} = E - E_{1A}$. Since the #microstates of individual systems multiply when you put them together, the probability of being in state E_{1A} is proportional to the number of microstates of system 1 having energy E_{1A} times the number of microstates of system 2 having energy E_{2A} . To simplify things for now, let's assume that E_{1A} has just one microstate; each if not, when calculating probabilities in a problem we'll have to multiply by the number of microstates it does have. (Repeat paragraph for state 1B.) The probability ratio we want is then just the ratio of available microstates of system 2 for the two cases, like this:

$$\frac{P_{1A}}{P_{1B}} = \frac{\#\text{microstates of 2 for case A}}{\#\text{microstates of 2 for case B}} = \frac{e^{S_{2A}/k_B}}{e^{S_{2B}/k_B}}$$

* Disclaimer: That step may seem like it involves a little "hand-waving". For example, if we had defined the constant in front of S in terms of different units, then we wouldn't have been able to make the supposition that dS/dE is exactly equal to $1/T$. However, in more advanced thermodynamics, instead of this being a supposition, this is actually the DEFINITION of temperature, $T = (dS/dE)^{-1}$. If you define temperature that way, all of the properties do indeed fit perfectly with "everyday temperature".

** This is only true if the volume is constant; see the footnote about PdV factors on the previous page. And if there's no volume change there's no work, so $dQ = dE$ by the First Law. Thus Eqn 1 fits exactly with the definition of S in the book: $dS = dQ/T$.

We're nearly there. We just have to figure out how S_2 changes when E_{1A} and E_{1B} are extracted, S_{2A} and S_{2B} respectively. Since E_{1A} is much smaller than E , we can use the definition of a derivative to do an expansion:

$$\frac{df}{dx} = \frac{f(x) - f(x - \Delta x)}{\Delta x} \quad (\text{for small } \Delta x) \quad (\text{The parentheses indicate "function of", not "times".})$$

$$\frac{dS_2}{dE} = \frac{S_2(E) - S_2(E - E_{1A})}{E_{1A}}$$

$$S_2(E - E_{1A}) = S_2(E) - \frac{dS_2}{dE} \cdot E_{1A}$$

But $\frac{dS_2}{dE} = \frac{1}{T_2}$ by Eqn 1 above ($= 1/T_1 = 1/T$, since both systems are at the same temperature T), so

$$S_2(E - E_{1A}) = S_2(E) - E_{1A}/T$$

Similarly, for state E_{1B} ,

$$S_2(E - E_{1B}) = S_2(E) - E_{1B}/T$$

Plugging those two expressions in and doing a little bit of algebra,

$$\begin{aligned} \frac{P_{1A}}{P_{1B}} &= \frac{e^{S_{2A}/k_B}}{e^{S_{2B}/k_B}} \\ &= \frac{e^{S_2(E) - E_{1A}/k_B T}}{e^{S_2(E) - E_{1B}/k_B T}} \\ &= \frac{e^{S_2(E)} e^{-E_{1A}/k_B T}}{e^{S_2(E)} e^{-E_{1B}/k_B T}} \\ &= \frac{e^{-E_{1A}/k_B T}}{e^{-E_{1B}/k_B T}} \end{aligned}$$

Thus the probability of system 1 being in any state is proportional to $e^{-E/k_B T}$ where E is the energy of that state. That deserves boxing:

$$\boxed{P \sim e^{-E/k_B T}} \quad \text{The Boltzmann factor} \quad (3)$$

(This E is not the same as the total energy E above. Rather, it is a general representation of the energy of system 1: E_{1A} , E_{1B} , etc.)

This is an amazingly helpful expression, and to my mind is the single most important thing you will learn if/when you take Physics 360. The factor $e^{-E/k_B T}$ is called the "Boltzmann factor" of the state, and as I stated at the outset, it tells you how likely states are to be occupied due to thermal energy. The "states" could be discrete states, like electrons being in quantum-mechanical atomic energy levels, or continuous states, like molecules having different kinetic energies.

Warning: Although equation (3) gives the *relative* probabilities that two (or more) states are occupied, to find the *absolute* probability, you have to normalize things—make it so that the total probability adds up to one. Typically you do that by dividing the Boltzmann factor for each state by the sum of all the Boltzmann factors.

Example 1: A made-up two level system. Suppose an atom has only two available energy levels, which are separated by $2e-23$ J. If the temperature is 1.5 K, what is the probability the atom is in the lower state?

Solution: Taking $E = 0$ to be the lower state, the two Boltzmann factors are:

$$\text{Upper state: BF} = e^{-2 \cdot 10^{-23} / (1.38 \cdot 10^{-23} \cdot 1.5)} = 0.38$$

$$\text{Lower state: BF} = e^{-0 / (1.38 \cdot 10^{-23} \cdot 1.5)} = 1$$

Dividing each one by the sum, 1.38, in order to normalize the probabilities, we get:

Upper state: 27.6% probable

Lower state: 72.4% probable

Example 2: The Maxwell-Boltzmann velocity distribution for molecules. This is an example of a continuously varying set of energy levels, because the energy of the molecules just depends on their velocities as $\text{KE} = \frac{1}{2} mv^2$. Since they can have any velocity, they can have any kinetic energy. That makes the BF more tricky to write out compared to the last example, because we can't just list all the states in a table. Instead, we have to write:

$$\text{State with } v: \text{BF} = e^{-\frac{1}{2}mv^2 / k_B T} \quad \text{incorrect expression}$$

That's not quite right, though. Recall a couple of pages back I said:

To simplify things for now, let's assume that E_{1A} has just one microstate; if not, when calculating probabilities in a problem we'll have to multiply by the number of microstates it does have.

In this situation, the state corresponding to speed v has more than one microstate, so as promised we're going to have to multiply by the number of microstates it does have. Actually, there are a ton of possible microstates, so we'll have to be clever.

For example, suppose the state corresponds to speed $v_1 = 5$ m/s. In terms of the x , y , and z velocities in m/s, some possible microstates could be $(5,0,0)$, $(3,4,0)$, $(0,5,0)$, $(-3,0,4)$, and so forth. If you picture all of the vectors starting at the origin and having length 5, the microstates can be represented as the surface of the sphere made up of all of the tips of those vectors (a sphere with radius = 5).

Although that's an infinite number of states, it's clearly "less", in some sense, than the number of microstates for the state corresponding to speed $v_2 = 6$ m/s—because the microstates of that state would be represented by the surface of a sphere having radius = 6. Thus the state corresponding to v_2 has more microstates than the v_1 state. Or, to use the usual term, we say it has a higher "multiplicity".

To account for the difference in multiplicity, we have to multiply each of the probabilities by the surface areas of the relevant spheres; that is, we must multiply by v^2 .^{*} In short:

$$\text{State with } v: \text{multiplicity} \times \text{BF} = v^2 e^{-\frac{1}{2}mv^2 / k_B T} \quad \text{correct expression}$$

In order to normalize the probabilities, we must divide each BF by the sum of all the BFs; in other words, we must

divide $v^2 e^{-\frac{1}{2}mv^2 / k_B T}$ by the integral $\int_0^\infty v^2 e^{-\frac{1}{2}mv^2 / k_B T} dv$. When we do that and evaluate the integral, we get:

$$\boxed{\text{State with } v: \text{Probability} = \frac{v^2 e^{-\frac{1}{2}mv^2 / k_B T}}{\sqrt{\frac{\pi}{2} \left(\frac{k_B T}{m}\right)^3}}} \quad \text{The Maxwell-Boltzmann distribution function} \quad (4)$$

Acknowledgement: The Boltzmann factor derivation presented here was modeled after the derivations given on these two websites: <http://world.std.com/~mmcirvin/boltzmann.html> and <http://mysite.du.edu/~jcalvert/phys/boltz.htm>

^{*} More properly, we should multiply by $4\pi v^2$, the surface area of a sphere with radius v , but since we have to normalize things later anyway, let's not worry about the 4π .