Matthew Schwartz
Statistical Mechanics, Spring 2019

Lecture 4: Temperature

1 Introduction

In the last lecture, we considered an ideal gas that is completely specified by the positions $\vec{q}_i$ and momenta $\vec{p}_i$ of the N molecules. We found the total number of states of the gas for large N was

$$\Omega(N, V, E) = 2e^{\frac{3}{2}N}\left(\frac{V}{(\Delta q \Delta p)^3}\right)^N \left(\frac{4\pi m E}{3N}\right)^{\frac{3N}{2}} \tag{1}$$

Then we asked how many fewer states were accessible if we know the velocity of single molecule. This told us the probability that the molecule had that velocity, leading to the Maxwell-Boltzmann velocity distribution

$$P(\vec{v}) = \left(\frac{3m}{4\pi\bar{\varepsilon}}\right)^{3/2} e^{-\frac{3}{4\bar{\varepsilon}}m\vec{v}^2} \tag{2}$$

where $\bar{\varepsilon} = \frac{E}{N}$.

The key to the computation of the Maxwell-Boltzmann distribution is that the number of states $\Omega(N, V, E)$ is an extremely rapidly varying function of energy, $\Omega \sim E^{3N/2}$. In this lecture we will see how to use the rapid variation of Ω to extract some general features of arbitrary systems. This will lead to the concept of temperature, as a constant among systems that can exchange energy.

2 Temperature

We defined $\Omega(E, V, N)$ for a system as the number of microstates compatible with some macroscopic parameters. What happens if we have two different types of systems that can interact? For example, nitrogen and oxygen gas in the air around you. The gas molecules can collide with each other and exchange energy, but the two types of gas are still distinct. What can we say about how the energy of the system is distributed among the two gases in equilibrium?

Say the total energy of both gases combined is E. By energy conservation, E does not change with time. So if there is energy E_1 in one gas, then the other gas has to have energy $E_2 = E - E_1$. Then the number of states with this partitioning is given by

$$\Omega(E, E_1) = \Omega_1(E_1)\Omega_2(E - E_1) \tag{3}$$

where $\Omega_1(E)$ and $\Omega_2(E)$ are the number of microstates of the two gases separately. The functions Ω_1 and Ω_2 do *not* have to be the same function, we just suppress their additional arguments for simplicity.

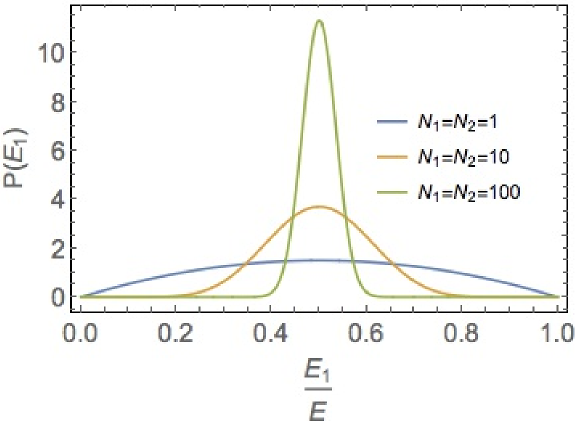
Now, the postulate of equal a priori probabilities implies that the probability of finding a system in a set of states is directly proportional to the number of states. Thus

$$P(E, E_1) = C \times \Omega_1(E_1)\Omega_2(E - E_1) \tag{4}$$

for some C, determined by normalizing the probabilities so that they integrate to 1.

For example, let's say these are ideal monatomic gases where $\Omega \sim E^{\frac{3}{2}N}$. Then

$$P(E,E_1) = C' \times E_1^{\frac{3N_1}{2}}(E-E_1)^{\frac{3N_2}{2}} = \tag{5}$$



where N_1 and N_2 are the numbers of the different types of gasses and the normalization C' is independent of E_1. We see that already for $N=100$ the central limit theorem is kicking in and the function is approaching a Gaussian with ever narrower width (plotting a Gaussian on top of the $N=100$ curve is indistinguishable). Note that the central limit applies here since we are looking at the probability of getting energy E_1 averaged over Ω microstates. The key observation is that because generically $\Omega(E)$ grows as some crazy power of E, $\Omega(E) \sim E^{10^{32}}$, this probability distribution is going to be some extremely sharply peaked function of energy. Adding or removing a little bit of energy changes Ω dramatically.

What is the expected value of E_1? For a Gaussian the mean is the same as the most probable value, and generally the most probable value is easier to compute (it's easier to differentiate than to integrate). The most probable value of E_1, is the one for which $\frac{\partial P}{\partial E_1} = 0$. For $P(E_1) = E_1^{\frac{3N_1}{2}}(E-E_1)^{\frac{3N_2}{2}}$ we have

$$\frac{\partial P}{\partial E_1} = \frac{3}{2} P(E_1)\left[\frac{N_1}{E_1} - \frac{N_2}{E-E_1}\right] \tag{6}$$

Setting this equal to zero implies that the most probable value (denoted by $\langle E_1 \rangle$ since it is also the mean) is $\langle E_1 \rangle = E\frac{N_1}{N_1+N_2}$ so that

$$\frac{\langle E_1 \rangle}{N_1} = \frac{\langle E_2 \rangle}{N_2} = \frac{E}{N} \tag{7}$$

Thus the average energies are equal. Since the function is so wildly varying, it is natural to expand its logarithm

$$\ln P(E_1) = \frac{3}{2}N_1 \ln E_1 + \frac{3}{2}N_2 \ln(E_2) = \frac{3}{2}N_1 \ln\langle E_1 \rangle + \frac{3}{2}N_2 \ln\langle E_2 \rangle - \frac{1}{2}\frac{(E_1 - \langle E_1 \rangle)^2}{\sigma^2} + \cdots \tag{8}$$

where

$$\sigma = \langle E_1 \rangle \sqrt{\frac{2N_2}{3N_1(N_1+N_2)}} \tag{9}$$

Thus we see that the width scales like $\sigma = \sqrt{\frac{1}{N}}$. Thus for 10^{24} particles, the chance of finding the configuration with anything other than the most probable energy allocation is not just exponentially small but exponentially exponentially small!

Now, let's generalize this to situations where we do not know the explicit form of $\Omega(E)$. Starting only with Eq. (4),

$$\frac{\partial P}{\partial E_1} = C\left[\frac{\partial \Omega_1(E_1)}{\partial E_1}\Omega_2(E-E_1) + \Omega_1(E_1)\frac{\partial \Omega_2(E-E_1)}{\partial E_1}\right]_{E_1 = \langle E_1 \rangle} \tag{10}$$

$$= C\Omega_1(E_1)\Omega_2(E_2)\left[\frac{1}{\Omega_1(E_1)}\frac{\partial \Omega_1(E_1)}{\partial E_1} - \frac{1}{\Omega_2(E_2)}\frac{\partial \Omega_2(E_2)}{\partial E_2}\right]_{E_1 = \langle E_1 \rangle, E_2 = E - \langle E_1 \rangle} \tag{11}$$

Setting this equal to zero and writing $\frac{1}{f}\frac{df}{dx}=\frac{d\ln f}{dx}$ we then have,

$$\boxed{\left.\frac{\partial\ln\Omega_1(E)}{\partial E}\right|_{E=\langle E_1\rangle}=\left.\frac{\partial\ln\Omega_2(E)}{\partial E}\right|_{E=\langle E_2\rangle}} \tag{12}$$

This motivates us to define the quantity

$$\beta\equiv\frac{\partial\ln\Omega(E)}{\partial E} \tag{13}$$

Then Eq. (12) implies that $\beta_1=\beta_2$ in equilibrium. So, even without specifying Ω we can say quite generally that there is a quantity which *is* equal in equilibrium: β.

It is customary to write

$$\beta=\frac{1}{k_BT} \tag{14}$$

where T called the **temperature** and $k_B = 1.38 \times 10^{-23}\frac{J}{K}$ a constant called **Boltzmann's constant** that converts units from temperature to energy.

So we have found that **any two systems that can exchange energy will be at the same temperature in equilibrium**.

Of course, we have not yet shown that this temperature is the same thing as what we measure with a thermometer. To do that, all we have to do is show that the thing measured by one kind of thermometer is inversely proportional to β. We will do this for mercury bulb thermometers in the next lecture. Then since any two systems in equilibrium will be at the same temperature, we can identify temperature as the thing measured by any thermometer.

2.1 Entropy

We also define the **entropy** as

$$S(N,V,E)\equiv k_B\ln\Omega \tag{15}$$

Entropy is a critical element of statistical mechanics and we start to study it in Lecture 5 then study it in depth in Lecture 6. We just introduce it here as a symbol, related mathematically to the logarithm of Ω. We then find

$$\frac{1}{T}=\frac{\partial S(N,V,E)}{\partial E} \tag{16}$$

This is the first of many thermodynamic relations, generally called Maxwell relations, that we encounter as we go along.

3 Temperature of a monatomic ideal gas

We will spend a lot of time studying ideal gases. For an ideal gas, we make two assumptions

1. The molecules are pointlike, so they take up no volume.
2. The molecules only interact when they collide.

The second point means we ignore van der Walls forces, Coulombic attraction, dipole-dipole interactions, etc. Most gases act like ideal gases to an excellent approximation, and in any case, the ideal gas approximation makes a good starting point for the study of any gas. That is, we can add in effects of finite volume or molecular attraction as small perturbations to ideal gas behavior. The most ideal gases are the noble gases, helium, xenon, etc. These gases are monatomic. Diatomic gases, like H_2 or O_2 are very close to ideal as well. A big differences is that diatomic and polyatomic molecules can store energy in vibrational and rotational modes, while monatomic gases only store energy in the kinetic motion of the atoms. Bigger molecules like CH_4 tend to be less ideal (their volume is more relevant), but the ideal gas approximation still works for them quite well.

For a monatomic ideal gas, we already computed the number of states Ω in Eq. (1). Taking the logarithm and multiplying by k_B gives the entropy as defined in Eq. (15):

$$S = Nk_B\left[\ln V + \frac{3}{2}\ln\left(\frac{4\pi mE}{3N[\Delta p\Delta q]^2}\right) + \frac{3}{2}\right] \tag{17}$$

We will call this the classical Sackur-Tetrode equation.

The classical Sackur-Tetrode is not quite right, but it is close. You are not expected to understand this yet, but the correct formula for an ideal gas is the **Sackur-Tetrode equation**:

$$S = Nk_B\left[\ln\frac{V}{N} + \frac{3}{2}\ln\left(\frac{4\pi mE}{3Nh^2}\right) + \frac{5}{2}\right] \tag{18}$$

There are 3 differences between this and the classical one we derived. The first is that $\Delta p\Delta q$ is replaced by h. This follows from quantum mechanics by the uncertainty principle (see Lecture 10). With h instead of $\Delta p\Delta q$ we can talk about the absolute size of entropy, rather then just differences of entropy. The other 2 differences are that V gets replaced by V/N and the $\frac{3}{2}$ is replaced by $\frac{5}{2}$. Both of these changes come from replacing V by $\frac{V}{N!}$ in Eq. (1) and using Stirling's approximation. The $N!$ comes from saying that the particles are *indistinguishable*, so saying particle 1 is in position 1 and particle 2 in position 2 is an identical configuration to the particles being in opposite places. Thus we have overcounted by the number of independent microstates by $N!$. We'll talk about distinguishability in Lecture 6 and quantum mechanics in Lecture 10. For now, we'll stick with the classical Sackur-Tetrode equation, Eq. (17) since it's the one we actually derived.

We can now compute the temperature of a monatomic ideal gas from Eq. (16):

$$\frac{1}{T} = \frac{\partial S(N,V,E)}{\partial E} = \frac{3}{2}Nk_B\frac{1}{E} \tag{19}$$

Thus,

$$E = \frac{3}{2}Nk_BT \tag{20}$$

The average energy per molecule is

$$\bar{\varepsilon} = \frac{E}{N} = \frac{3}{2}k_BT \tag{21}$$

Next, recall the Maxwell-Boltzmann distribution for momenta

$$P(\vec{p}) = \left(\frac{3}{4\pi m\,\bar{\varepsilon}}\right)^{3/2} e^{-\frac{3}{4\bar{\varepsilon}}\frac{\vec{p}^2}{m}} \tag{22}$$

Using Eq. (21) this becomes

$$\boxed{P(\vec{p}) = \left(\frac{1}{2\pi mk_BT}\right)^{3/2} e^{-\frac{1}{k_BT}\frac{\vec{p}^2}{2m}}} \tag{23}$$

As we will see, this is a special case of a general result, that in thermal equilibrium, the chance of finding something energy ε is $P(\varepsilon) = e^{-\varepsilon/k_BT}$ (cf. Eq. (71) below).

Still for the monatomic ideal gas, the average energy in kinetic energy in the x direction is

$$\langle\frac{p_x^2}{2m}\rangle = \int d^3p\,\frac{p_x^2}{2m}P(\vec{p}) = \frac{1}{2}k_BT \tag{24}$$

similarly, the energy in p_y and p_z are also both $\frac{1}{2}k_BT$, so

$$\langle\frac{\vec{p}^2}{2m}\rangle = \frac{3}{2}k_BT \tag{25}$$

Thus we can interpret the $\frac{3}{2}$ in Eq. (21) as saying that there are 3 degrees of freedom for the energy to be stored in for this monatomic gas: kinetic energy in p_x, p_y and p_z. Each kinetic energy degree of freedom gets $\frac{1}{2}k_BT$ of energy.

It may be worth building a little intuition for the size of $k_B T$. At room temperature $k_B T = 25\,\text{meV}$. Thus the kinetic energy of any given molecule at room temperature is 36 meV. You can compare this to the typical electronic excitation energy, or order 1 Rydberg = 13 eV. So typical kinetic energies are way too small to excite electronic excitations (vibrational excitations are lower energy than electronic ones, in the sub-eV range, while rotational excitations are even lower, be in the meV range, see below). Boltzmann's constant $k_B \sim 10^{-23} J/K$ is about as small as Avogadro's number is big. That's because it measures the typical energy of a molecule, so the energy of a whole mole of molecules is in the Joule range which is macroscopic (1 J is about the energy in a heartbeat). The ideal gas constant is a mole of Boltzmann's constants: $R = k_B N_A = 8.3 \frac{J}{\text{mol} \cdot K}$.

4 Equipartition theorem

Recall that for a monatomic ideal gas the energy is quadratic in all the momenta:

$$E = \frac{1}{2m}[p_{1x}^2 + \cdots + p_{Nx}^2 + p_{1y}^2 + \cdots + p_{Ny}^2 + p_{1z}^2 + \cdots + p_{Nz}^2] \tag{26}$$

There are $3N$ components in the sum and each gets $\frac{1}{2}k_B T$ of energy on average so the total energy is $E = \frac{3}{2} N k_B T$. If there weren't 3N components in the sum, but rather f components, each would still get $\frac{1}{2} k_B T$ of energy and the total energy would be $\frac{f}{2} k_B T$. This happens for example with rotational and vibrational modes of diatomic or polyatomic molecules (we'll get to these soon). The general calculation is summarized in

- The **equipartition theorem**: in equilibrium, the available energy is distributed equally among available quadratic modes of any system, each getting $\frac{1}{2} k_B T$.

A **mode** is the set of excitations of the same system (like momentum, or vibration, or rotation, or normal modes on a string). A quadratic mode is one for which the energy is quadratic in the variable.

4.1 Non-quadratic modes

What is special about quadratic modes? Nothing really. The equipartition theorem in this form is particularly useful in chemistry, since with non-relativistic gases and liquids all the degrees of freedom are quadratic. In physics, systems are more varied. For example, a system in which the energy is linear in the variable is the kinetic energy of an ultrarelativistic gas. Such gases are present in stars for example (as we'll discuss in Lecture 15). The relativistic formula for energy is

$$\varepsilon = \sqrt{m^2 c^4 + c^2 p^2} \approx \begin{cases} cp + \cdots & , p \gg mc \\ mc^2 + \dfrac{p^2}{2m} + \cdots & , p \ll mc \end{cases} \tag{27}$$

where $p = |\vec{p}|$. For $p \ll mc$ energy reduces to $mc^2 + \frac{1}{2}\frac{\vec{p}^2}{m}$ which is the rest mass energy plus a quadratic part, the non-relativistic kinetic energy. For $p \gg mc$ energy reduces to $\varepsilon = cp$ which is linear in the variable.

It is not hard to repeat the calculation we did for the non-relativistic momentum for a situation in which the energy is linear. Energy for N particles is

$$E = c(p_1 + p_2 + \cdots + p_N) \tag{28}$$

where $p_j = |\vec{p}_j|$. Thus

$$\Omega_N(E) = \left(\frac{1}{\Delta p}\right)^{3N} \int_0^{E/c} 4\pi p_1^2 dp_1 \cdots \int_0^{E/c} 4\pi p_N^2 dp_N \delta(cp_1 + \cdots + cp_N - E) \tag{29}$$

Computing the energy dependence of Ω_N (you should check this!), or more simply by dimensional analysis, we find

$$\Omega_N = C \times \left(\frac{E}{c\Delta p}\right)^{3N} \tag{30}$$

for some constant C. The temperature is then

$$\frac{1}{T} = \frac{\partial k_B \ln\Omega}{\partial E} = 3k_B \frac{N}{E} \tag{31}$$

and therefore

$$\bar{\varepsilon} = \frac{E}{N} = 3k_B T \tag{32}$$

So we find that there is $k_B T$ (not $\frac{1}{2}k_B T$) of energy for each linear mode of the system. Note that we did not need the constant C to compute the relationship between E and T.

We could imagine more exotic systems where the possible excitations are evenly spaced in a variable y (meaning the measure is dy) with $E(y) = y^a$. So $a = 2$ are kinetic, vibrational or rotational modes, while $a = 1$ is a non-relativistic gas. For general a, the calculation is as above giving $\Omega_N(E) \sim E^{N/a} \times \Omega_{\text{rest}}$ where Ω_{rest} are the degrees of freedom in other modes. More precisely, if there are f modes with each giving energy $E \sim y_i^{a_i}$, then

$$\Omega_N = C \times E^{\frac{N}{a_1}} \cdots E^{\frac{N}{a_f}} \tag{33}$$

Then,

$$\frac{1}{T} = \frac{\partial k_B \ln\Omega}{\partial E} = N k_B \left(\frac{1}{a_1} + \cdots + \frac{1}{a_f}\right) \frac{\partial}{\partial E} \ln E = \frac{N k_B}{E} \left(\frac{1}{a_1} + \cdots + \frac{1}{a_f}\right) \tag{34}$$

Thus,

$$\frac{E}{N} = \left(\frac{1}{a_1} + \cdots + \frac{1}{a_f}\right) k_B T. \tag{35}$$

So each mode contributes $\frac{1}{a_i} N k_B T$ to the total energy.

It is perhaps also informative to see how the Maxwell-Boltzmann distribution changes for non-quadratic degrees of freedom. Let's write the energy dependence of Ω as $\Omega = E^{Nb}$, and therefore get $\frac{E}{N} = b k_B T$. Suppose we want to know the probability of finding one mode excited with energy ε? For example, it might be a relativistic kinetic degree of freedom for which $\varepsilon \sim p$, or a vibrational degree of freedom for which $\varepsilon \sim x^2$ or $\varepsilon \sim \dot{x}^2$, or whatever. The probability we're after is given by the number of configurations with the other particles having $E - \varepsilon$ energy normalized to the total number of configurations. So,

$$P(\varepsilon) = \frac{\Omega_{N-1}(E-\varepsilon)}{\Omega_N(E)} = \frac{(E-\varepsilon)^{bN}}{E^{bN}} = \left(1 - \frac{\varepsilon}{E}\right)^{bN} = \left(1 - \frac{\varepsilon}{N b k_B T}\right)^{bN} \tag{36}$$

where $E = N k_B T$ was used. Now we take $N \to \infty$ recovering an exponential:

$$P(\varepsilon) \propto \exp\left(-\frac{\varepsilon}{k_B T}\right) \tag{37}$$

Note that the dependence on b has dropped out. The probability is also independent of the power a associated with the degree of freedom (i.e. whether it's a quadratic mode or a linear mode or whatever). Thus no matter how the energy is stored, the probability of finding energy ε in any mode is $e^{-\frac{\varepsilon}{k_B T}}$. This is very powerful and very general result in statistical mechanics.

4.2 Summary

The equipartition theory strictly speaking refers to quadratic modes only. Quadratic modes are not more fundamental, they were just historically the ones relavent when the equipartition theorem was discovered by chemists. For physics, the point of the equipartition theorem is that $k_B T$ characterizes the energy in each mode *independent of the other modes*. So $\frac{3}{2}k_B T$ is always the average kinetic energy of a non-relativistic gas, whether the gas molecules are monatomic, or complicated polymers.

As a simple application, consider a ball rolling down a hill. Why does it sit at the bottom of the hill? Why is energy minimized? There is nothing special about the minimum of energy from Newton's laws – a ball should roll down the hill, then roll back up the other side. But if it just oscillates around the minimum, it will have a lot of energy. Of course we know that there in real life the ball stops at the bottom because there is friction. The friction causes the energy of the ball to go into heating up (exciting kinetic modes) the molecules in the dirt and the air. There is one degree of freedom for the ball, but $N \sim 10^{24}$ for the air. Thus once equilibrium is reached, the ball only has 10^{-24} of the energy it started with. It is at rest. Thus equipartition implies that nature tends to a situation where no object contains more than its fair share of energy. We'll revisit this picture more quantitatively once we have introduced the concept of free energy in Lecture 8.

5 Heat capacity

Because of the equipartition theorem, the relationship between the temperature T and the total energy E of a system tells us about the degrees of freedom (modes) that can be excited. Generally, we don't care about the total energy of a system since we'll never actually measure it and certainly never use it for anything, so we define the relationship between E and T differentially:

$$C_V = \left(\frac{\partial E}{\partial T}\right)_V \tag{38}$$

C_V is called the **heat capacity**, or more precisely the **heat capacity at constant volume**. It tells you how much the energy goes up when you change the temperature (by adding heat, see next lecture) at constant volume. The notation $()_V$ means volume is held fixed when the system is heated. We could instead consider heating up the system and letting it expand so that the pressure is held constant, whence C_V is replaced by C_P. (Again, we'll talk about this more next lecture.)

For a non-relativistic ideal gas, we saw that $E = \frac{3}{2} N k_B T$ so $C_V = \frac{3}{2} N k_B$. For other systems, with more degrees of freedom, C_V will be greater. Here is a table of some gases and their measured values of C_V

Gas	Ar	He	O_2	CO	H_2	HCl	Br$_2$	N_2O	CO_2	CS_2
$\frac{1}{N k_B} C_V$	1.5	1.5	2.54	2.49	2.45	2.57	3.39	3.42	3.40	4.92

Heat capacities of various gases at $T = 15°C$ and $P = 1$ atm.

We see that the measured values of C_V for argon and helium are in excellent agreement with our prediction for monatomic molecules. To explain the others, we need to think about diatomic and polyatomic molecules.

5.1 Vibrational and rotational modes

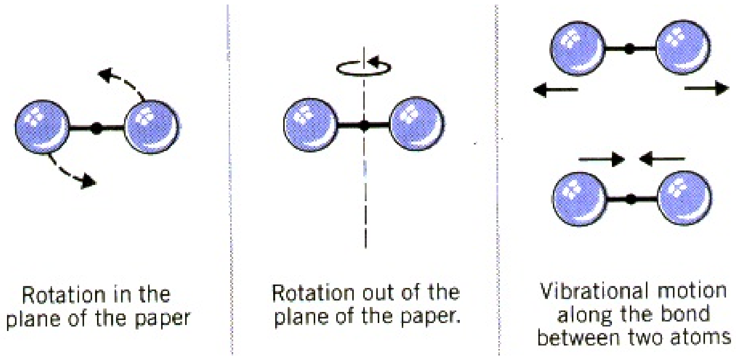
The energy of a monatomic molecule (atom) like Helium is all in the momentum. What happens with a diatomic molecule like H_2 or a more complicated polyatomic molecule like benzene or hydrodioxytetrohydramie? Or what happens with a solid?

Let's start with molecular hydrogen gas, H_2. We know from quantum mechanics that H_2 comprises two hydrogen atoms bound by a covalent bond – an electron orbital shared between the two atoms. The typical picture of a molecule is big heavy, slowly moving nuclei setting up a potential for the electrons on relatively long timescales, and then the electrons adjust and adapt their electron cloud based on the position of the nuclei. That is, we factorize the problem, first figuring out what is going on the nuclei, and then the electrons. This factorization approximation is called the Born-Oppenheimer approximation, and provides an excellent description of molecular dynamics.

Let's start with a classical description of the nuclei. To specify the the state of the nuclei classically we need to give the positions and velocities of two atoms in the molecule, so we need 6 positions and 6 velocities. We want to know the energy cost for moving any of these 12 classical coordinates. Since there is no energy cost to translate the whole system, it makes sense to separate the motion of the center-of-mass from the motion relative to the center of mass. The center-of-mass we have already discussed: the energy does not depend on the position of the center-of-mass of any molecule, and the motion of the center of mass contributes the usual kinetic energy $\frac{\vec{p}^2}{2m}$. That is, the center-of-mass acts just like a monatomic molecules, so all the new features for more complicated molecules will come from internal motion, relative to the center of mass.

For motion relative to the center of mass of H_2 we need to specify 3 positions and 3 velocities. To understand the positions, lets hold momentum fixed, say at zero. Then we can rotate the two atoms around the center of mass without stretching the covalent bond. There is no energy cost to doing so, and so the energy doesn't depend on 2 of the 3 positions. It makes sense therefore to work in spherical coordinates (r, θ, ϕ) where we see the energy can depend on r but not on θ or ϕ. The important coordinate is the distance between the two hydrogen nuclei, r. There is an equilibrium value r_0 and the atoms can oscillate about this value due to the restoring potential induced by the covalent bond. Near equilibrium the force is linear (just Taylor expand – every force is linear near equilibrium), so the atoms will vibrate like simple harmonic oscillators, i.e. like springs. Call the spring constant k, so $F = -kx$ with $x = r - r_0$ the distance from equilibrium. The energy of a spring is $E = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2$: it is part kinetic energy and part potential energy from stretching the spring. Both of the energies are quadratic in the variable, so we expect $\frac{1}{2}k_BT$ of energy for each by the equipartition theorem; equivalently, we sometimes say that the vibrational mode has k_BT of energy. This takes care of the remaining position and one of the velocities.

The two remaining velocities to specify the 2 nuclei completely are angular velocities $\dot{\theta}$ and $\dot{\phi}$. The molecule can rotate in two independent ways



Two rotational modes and one vibrational mode of a diatomic molecule.

Classically, rotational kinetic energy is $E = \frac{1}{2}I\omega^2$, with I the moment of inertia and ω the angular velocity. So for two rotational modes we would have $E = \frac{1}{2}I\omega_1^2 + \frac{1}{2}I\omega_2^2$. Note that the moments of inertia are the same for the two rotations since they are the same motion in a different plane (see Fig. 1). Each of these energies is quadratic in the variable, so we have 2 more $\frac{1}{2}k_BT$ excitations. You might wonder about the rotation around the axis connecting the molecules. Indeed, there is some rotational energy associated with a spinning nucleus. However, the moment of inertia in this direction is so small that this energy is absolutely tiny and can totally be neglected compared to the other energies. (Roughly, $I \sim R^2$ so this third 3rd moment of inertia is down by a factor of $\left(\frac{R_{\text{nucleus}}}{R_{\text{molecule}}}\right)^2 \sim 10^{-10}$ compared to the others).

Putting everything together, we find the total heat capacity for H_2 is

$$\frac{1}{Nk_B}C_V = \underbrace{\frac{3}{2}}_{\text{kinetic}} + \underbrace{2 \times \frac{1}{2}}_{\text{rotations}} + \underbrace{2 \times \frac{1}{2}}_{\text{vibrations}} = \frac{7}{2} = 3.5 \quad \text{(classical diatomic molecule)} \tag{39}$$

Another way to see this is that we found of the 6 position coordinates needed to describe the molecule, only the relative distance between the atoms costs energy. Of the 6 momenta, all 6 of them cost energy. So we have $(1+6)\times\frac{1}{2}k_BT$ total. So we conclude that $C_V\approx 3.5Nk_B$ for H_2. Looking at table 1 we see the data says $C_V=2.45k_BT$ for H_2. So something is not right.

As you might have guessed, the reason our prediction is off is quantum mechanics. We assumed that we could have arbitrarily little energy in any vibrational or rotational mode. Instead, there is a lower limit. In the quantum system, we know that the energies of a harmonic oscillator are $\varepsilon_{\text{sho}}=\hbar\sqrt{\frac{k}{m}}\left(n+\frac{1}{2}\right)$ with k the spring constant and m the mass. For a diatomic atom we have

$$\varepsilon_n=\hbar\sqrt{\frac{k}{\mu}}\left(n+\frac{1}{2}\right)=\left(n+\frac{1}{2}\right)\varepsilon_{\text{vib}} \tag{40}$$

where $\mu=\frac{m_1m_2}{m_1+m_2}$ is the reduced mass and $\varepsilon_{\text{vib}}=\hbar\sqrt{\frac{k}{\mu}}$ is the energy to excite the first vibration, from $n=0$ to $n=1$.

Rotational modes have a quantized spectrum as well

$$\varepsilon_j=j(j+1)\frac{\hbar}{2\mu c r_0^2}=\frac{j(j+1)}{2}\varepsilon_{\text{rot}} \tag{41}$$

So there is also a characteristic energy ε_{rot} scale for rotations. Generally rotational energies are lower than vibrational ones.

It's not easy, but in principle possible, to compute the energies for vibrational and rotational modes using quantum mechanics, at least numerically. It is easier to measure the energies experimentally, since there's a resonance absorption of energy of photons corresponding to the vibrational or rotational transitions. For example, hydrogen has a resonance at wavenumber $\tilde{\nu}_{\text{vib}}=4342\,\text{cm}^{-1}$ corresponding to the vibrational mode and at $\tilde{\nu}_{\text{rot}}=60\,\text{cm}^{-1}$ corresponding to a rotational mode. Here $\tilde{\nu}$ is spectroscopy notation for wavenumber, defined as the the inverse of the wavelength $\tilde{\nu}=\frac{1}{\lambda}=\frac{\nu}{c}$ (physicists use $k=\frac{2\pi}{\lambda}=2\pi\frac{\nu}{c}$ for wavenumber which differs by 2π). To convert between wavenumbers and Kelvin we use

$$k_B=1.38\times10^{-23}\frac{J}{K}=8.617\times10^{-5}\frac{\text{eV}}{K} \tag{42}$$

$$h=4.14\times10^{-15}\,\text{eV}\cdot s \tag{43}$$

$$c=2.99\times10^{8}\frac{m}{s} \tag{44}$$

So room temperature, $298K$ corresponds to 25 meV or $4\times10^{-21}J$. Also,

$$1K\times\frac{k_B}{hc}=0.52\,\text{cm}^{-1} \tag{45}$$

Thus for H_2

$$\varepsilon_{\text{vib}}=hc\,\tilde{\nu}_{\text{vib}}=0.54\text{ eV},\qquad \frac{\varepsilon_{\text{vib}}}{k_B}=\frac{ch}{k_B}\tilde{\nu}_{\text{vib}}=6300K \tag{46}$$

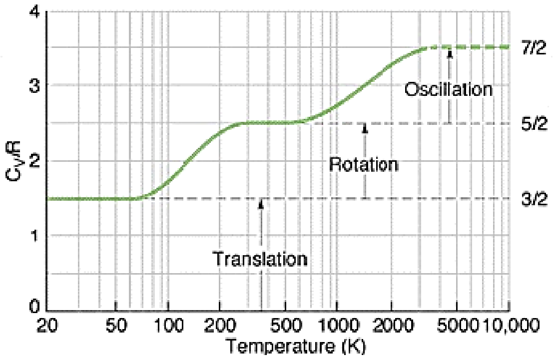
$$\varepsilon_{\text{rot}}=hc\,\tilde{\nu}_{\text{rot}}=7.4\,\text{meV},\qquad \frac{\varepsilon_{\text{rot}}}{k_B}=\frac{ch}{k_B}\tilde{\nu}_{\text{rot}}=86K \tag{47}$$

Thus at room temperature, rotational modes are excited but not vibrational ones. Using this, we can refine our prediction. With the the 3 center-of-mass momenta degrees of freedom, plus 2 rotational ones we get

$$\frac{1}{Nk_B}C_V=\frac{5}{2}\quad(\text{hydrogen, rotation only}) \tag{48}$$

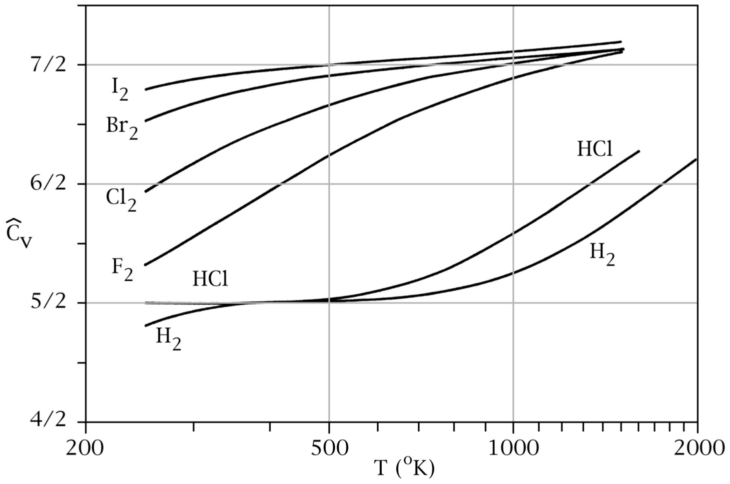
now in great agreement with measured value of 2.45 in Table 1.

If we were to increase the temperature, the heat capacity would look something like this



Cartoon of the heat capacity of H_2 as a function of temperature

Fig 2 is a cartoon, and not very realistic. Actual data looks more like this :



Heat capacities for various molecules

To understand this data, we can think about the scaling of the rotational and vibrational energies with mass. From Eqs. (40) and (41) we see that a larger reduced mass decreases both the vibrational and rotational energies. Thus what is going on in this figure is that the heavier molecules, iodine and bromine are exciting all 7 degrees of freedom by $1000K$ while the lighter ones, hydrogen and hydrogen-chloride still only exciting rotational modes at that temperature. That the heavier ones get above $\frac{7}{2}$ says that something else is getting excited too. Can you guess what?

We'll come back to the vibrational and rotational spectra, computing the heat capacity precisely and quantitatively comparing to data, once we've introduced quantum statistical mechanics in Lecture 10.

5.2 Solids

What is the heat capacity of a solid? A simple model of a solid is a lattice of atoms connected by little springs (covalent bonds). If there are N atoms, then we specify the state of the system with $3N$ degrees of freedom. The contribution to the heat capacity due to the momentum of any atom is the usual $\frac{3}{2}k_BT$. The contribution from the position of each atom is also $\frac{3}{2}k_BT$. So the total is $C_V = 3Nk_B$. Another way to see it is that there are vibrational modes in each direction. Each vibrational mode gives k_BT (half potential and half kinetic), so the total is again $C_V = 3Nk_B$.

This leads to

- **The law of Dulong and Petit**: the molar heat capacity of many metals is roughly constant.

where

- The **molar heat capacity** is the heat capacity per mole of a substance. Molar heat capacity is also called **molar specific heat** and often denoted by a lowercase c.

and

- A **mole** of a substance is Avogadro's number $N_A = 6.02 \times 10^{23}$ of that thing

So the number of moles n in N particles of something is $n = \frac{N}{N_A}$. Thus the molar heat capacity is $c = 3N_A k_B$ = constant. When using moles we also use

- **The ideal gas constant:** $R \equiv N_A k_B = 8.314 \frac{J}{\text{mol} \cdot K}$

"mol" in this expression is a very strange object for a physicist, it's a unit without dimensions. In terms of R, the molar heat capacity is $c = 3R$, as observed experimentally by Dulong and Petit in 1819. Here is a table of various molar specific heats similar to what Dulong and Petit might have measured.

Element	Al	Bi	Cu	Au	Ag	Pb		H_2O	C_2H_5OH
Molar specific heat $\left(\frac{J}{\text{mol} \cdot K}\right)$	24.3	25.7	24.5	25.6	24.9	26.4		75.2	111

Molar specific heats for various metals, contrasted with water and ethanol which do not satisfy the law of Dulong and Petit.

To understand deviations from the law of Dulong and Petit, we need a better model of a solid. We will construct such models, such as the Debye model and the Einstein model, in Lecture 13.

A common related quantity is the

- The **specific heat**: $S =$ heat capacity per unit mass.

That is

$$S \equiv \frac{\Delta E}{m \Delta T} = \frac{C_V}{m} \tag{49}$$

Specific heat has units of $\frac{J}{\text{kg} \cdot K}$.

6 Principle of maximum entropy

In Section 2 we defined the entropy as

$$S = k_B \ln \Omega \tag{50}$$

We showed that $\frac{\partial S}{\partial E} = \frac{1}{T}$ was the same among systems that can exchange energy. To show this, we used that the number of configurations $\Omega(E)$ was a very rapidly growing function of energy. Then the system was exponentially likely to be close to the value of E to where $\Omega(E)$ is maximized. Another way to say this is that the system is exponentially likely to be the state of maximum entropy.

The idea of maximizing entropy is very powerful. In this section we will see a very general way to use it, called **the principle of maximum entropy** or **maxent**. This principle says to find the probability distributions that maximize $\ln \Omega$ using only what is known about the system (total number of particles, total energy, total volume, etc). It is a very general method, proposed in 1957 by E.T Jaynes. It turns out to be very powerful, not just for physics, but for statistics, information theory, artificial intelligence, finance and many other areas of intellectual enquiry. We will use it in physics for a general derivation of the Boltzmann factor $e^{-\varepsilon / k_B T}$ and to provide a new powerful formula for entropy, related to the H in the Boltzmann H theorem.

6.1 Fixed particle number

For our first application of the principle of maximum entropy, consider the question: suppose we have N particles and have m categories in which to split them. For example, we might have m energy levels in a quantum system, or m regions in a box. Let's label the m groups $i=1...n$. If you pick a particle at random, what is the probability it would come from group i? This question is so simple that you can probably guess the answer. The next question we will ask is if the particles in box i have energies ε_i and the total energy is E, what is the probability that if you a particle at random it will have energy ε_i? The answer to the second question is not so obvious. But by solving the first question the right way, the solution to the second will be easy.

For the first question, with no mention of energy, we consider the ways to divide the N particles into the m groups. Each group can have some number n_i of particles in it. Since we know the total number of particles, we have

$$\sum_{i=1}^{m} n_i = N \tag{51}$$

Note that we are not fixing n_i, so you should imagine there are many possible values that the n_i can take for a given m and N.

Now, how many ways are there of splitting the N particles into m groups of size n_i? For example with 8 particles ($N=8$) and 1 group ($m=1$) then $n_1=8$ and there is $1=\frac{8!}{8!}$ way. With two groups ($m=2$) there are ${}_NC_n = \binom{N}{n} = \frac{N!}{n!(N-n)!}$ ways of picking $n_1=n$ particles for the first group, with the other $n_2=N-n$ particles in the second group. For an arbitrary number of groups, we can work out the formula by putting all the particles all in a row. There are $N!$ ways of doing this. Then we take the first n_1 into group 1, the second n_2 into group 2 and so on. There are $n_1!$ of the original orderings which put the same n_1 particles in group 1, and $n_2!$ which put the particles in group 2, and so on. Thus, the total number of ways of divvying up the particles is

$$\Omega = \frac{N!}{n_1!\cdots n_m!} \tag{52}$$

This formula is due to Boltzmann. It is a generalization of the binomial distribution.

Now, if N and the n_i are all very large, then we can use Stirling's approximation:

$$\ln\Omega \sim N\ln N - N - \sum_{i=1}^{m}(n_i \ln n_i - n_i) = N\ln N - \sum_{i=1}^{m} n_i \ln n_i = -N\sum_{i=1}^{m}\frac{n_i}{N}\ln\frac{n_i}{N} \tag{53}$$

Defining $f_i = \frac{n_i}{N}$ this gives

$$\boxed{\ln\Omega = -N\sum_i f_i \ln f_i} \tag{54}$$

This is a very important result, originally due to Boltzmann. It is a form of writing entropy in terms of fractions of particles, rather than Ω which is just the total number of microstates.

Since $\sum_i n_i = N$ so that $\sum f_i = 1$ the fractions have interpretation of probabilities: f_i is the probability that if you pick a particle at random it will be from group i. Now wait, you say, we already know that $f_i = \frac{n_i}{N}$, so we know these probabilities. Yes, that's true. The probability f_i is just the number of particles in that group divided by N. Picking any particle is equally likely, as with the postulate of equal a priori probabilities. But we're allowing n_i to vary. The principle of maximum entropy will tell us what the most probable values for the f_i (and hence n_i) are.

Before computing f_i, it is worth noting that $\ln\Omega$ in Eq. (54) looks nearly identical to Boltzmann's quantity H from his H theorem

$$H = -\sum P_i \ln P_i \tag{55}$$

we just have to identify $H = \frac{\ln\Omega}{N}$. Since $S = k_B \ln\Omega$ we can also identify Boltmann's H with entropy. Recall that the H theorem says that (assuming molecular chaos) H always increases. A general consequence of this is that entropy always increases as well (the 2$^{\text{nd}}$ law of thermodynamics). We'll discuss entropy in great deal more over the next few lectures.

Now let's apply the principle of maximum entropy. What is the most probable configuration? We want to maximize $\ln\Omega$ over f_i subject to the constraint that $\sum_i n_i = N$ or equivalently $\sum f_i = 1$. To maximize a function with constraints, we use Lagrange multipliers. Langrange multipliers are a fairly magical mathematical trick. The idea behind them is to turn the problem of maximizing a function of n variables with constraints into a problem of maximizing a function of more than n variables with no constraints. Explicitly, we want to find values of f_i and α that maximize

$$\ln\Omega = -N\sum_{i=1}^{m} f_i \ln f_i - \alpha\left(\sum n_i - N\right) \tag{56}$$

Here α is the Lagrange multiplier. Variations of $\ln\Omega$ with respect to α would enforce the constraint. Importantly however, we don't want to impose the constraints directly in $\ln\Omega$. Instead, we compute partial derivatives with respect to f_i first, then we put the constraints in afterwards.

Since N is constant we can equally well maximize $\frac{\ln\Omega}{N}$ as $\ln\Omega$. Dividing Eq. (56) through by N gives

$$\frac{\ln\Omega}{N} = -\sum_{i=1}^{m} f_i \ln f_i - \alpha\left(\sum f_i - 1\right) \tag{57}$$

Taking the derivative with respect to f_i then gives

$$\frac{\partial}{\partial f_i}\frac{\ln\Omega}{N} = -(1+\ln f_i) - \alpha \tag{58}$$

This is zero (and $\ln\Omega$ is maximized) when

$$f_i = e^{-\alpha-1} \tag{59}$$

Varying Eq. (57) with respect to α gives

$$1 = \sum_{i=1}^{m} f_i = m e^{-\alpha-1} \tag{60}$$

So that

$$\alpha = \ln m - 1 \tag{61}$$

and therefore

$$f_i = \frac{1}{m} \tag{62}$$

We have found that, if we don't know anything about the system except for the total number of particles and the number of groupings m, then the probability of finding a particle in group i is $f_i = \frac{1}{m}$.

As a special case, consider taking $m = N$ so that each microstate is a single particle configuration. Then maxent reproduces the postulate of equal a priori probabilities. That should not be surprising, since we derived the postulate of equal a priori probabilities from the Boltzmann H theorem. But Eq. (62) is more general result than the postulate.

6.2 Fixed average energy

Now for the harder question. Suppose that the groups labelled i have energies ε_i and that we know the total energy $E = \sum n_i \varepsilon_i$ or equivalently the average energy $\bar{\varepsilon} = \frac{E}{N}$. What are the most probable values of the probabilities $f_i = \frac{n_i}{N}$ given the ε_i and $\bar{\varepsilon}$? The constraint on the average energy is that

$$\sum f_i \varepsilon_i = \bar{\varepsilon} = \frac{E}{N} \tag{63}$$

Now we want to maximize $\ln\Omega$ subject to the constraints in Eqs. (51) and (63). We introduce two Lagrange multipliers for this case, giving

$$\frac{1}{N}\ln\Omega = -\sum_{i=1}^{m} f_i \ln f_i - \alpha\left(\sum_i f_i - 1\right) - \beta\left(\sum_i f_i \varepsilon_i - \bar{\varepsilon}\right) \tag{64}$$

Varying with respect to f_i gives

$$-\ln f_i - 1 - \alpha - \beta\epsilon_i = 0 \tag{65}$$

So that

$$f_i = e^{-1-\alpha} e^{-\beta\varepsilon_i} \tag{66}$$

Already this is a very powerful result. It says that if all we know is the average value of some quantity, through $\sum \varepsilon f(\varepsilon) = \bar{\varepsilon}$, then our best guess at the probabilities should be exponential functions $f(\varepsilon) \propto e^{-\beta\varepsilon}$.

Next we impose the constraints. Varying Eq. (64) with respect to α gives the total-number constraint

$$1 = \sum_i f_i = e^{-1-\alpha} \sum_i e^{-\beta\varepsilon_i} \tag{67}$$

It is handy at this point to define

$$Z \equiv \sum_i e^{-\beta\varepsilon_i} \tag{68}$$

Thus the α constraint implies

$$Z = e^{1+\alpha} \tag{69}$$

In terms of Z, Eq. (66) becomes

$$f_i = \frac{1}{Z} e^{-\beta\varepsilon_i} \tag{70}$$

This is our answer. The probability of picking a particle at random and finding it to have energy ε_i decreases exponentially with ε_i. Starting from basically nothing we have found a very general formula, that the probability of finding a with energy ε is

$$\boxed{P(\varepsilon) = \frac{1}{Z} e^{-\beta\varepsilon}} \tag{71}$$

The variations around this probability are very small, scaling like $\frac{1}{\sqrt{N}}$ by the law of large numbers. This is the one of the most important equations in all of statistical mechanics, perhaps all of physics. It says that, in equilibrium, the probability of finding something with energy ε is proportional to a Boltzmann factor $e^{-\beta\varepsilon}$.

Varying Eq. (64) with respect to β gives

$$\bar{\varepsilon} = \sum_i f_i \varepsilon_i = \frac{1}{Z} \sum_i \varepsilon_i e^{-\beta\varepsilon_i} \tag{72}$$

This says that the average energy is given by the sum over the possible energies times the probability of finding those energies.

Another consequence of Eq. (64) is that

$$\beta = \frac{\partial \ln\Omega}{\partial E} \tag{73}$$

Using $S = k_B \ln\Omega$ we see that the Lagrange multiplier β is the same as the $\beta = \frac{1}{k_B T}$ in Eq. (13) (which is why we chose the same letter). Note that we are not maximizing $\ln\Omega$ with respect to E (only to f_i, α and β), so we do not set $\frac{\partial \ln \Omega}{\partial E}$ equal to zero.

Since β has the interpretation as (inverse) temperature, you might naturally ask what is the interpretation of α? Just like Eq. (73) we find

$$\alpha = \frac{\partial \ln\Omega}{\partial N} \tag{74}$$

This quantity α is related to the **chemical potential $\boldsymbol{\mu}$** as $\alpha = -\frac{\mu}{k_B T}$. The chemical potential is another quantity, like temperature, that is constant among systems in equilibrium. A third is pressure, $p = \frac{\partial \ln\Omega}{\partial V}$. While you are probably somewhat familiar with pressure already, chemical potential may take some time to understand. We will tackle chemical potential starting in Lecture 7. I only introduced the partition function and the chemical potential here because they appear naturally when using the principle of maximum entropy. It will take some time to understand both concepts, so don't worry about them now.

We have treated the possible energies as discrete values ε_i. It is often easier to think of the energies as being continuous, as in a classical system. So rather than a discrete index i of the groups of particles, we can use a continuous label x. In the continuum limit, the results from this section can be phrased set of mathematical results about functions $P(x)$ that maximize a functional $H[P(x)]$ defined as

$$H[P] = -\int_{-\infty}^{\infty} dx\, P(x) \ln P(x) \tag{75}$$

For example, if we only constrain the probabilities to be less than 1, $0 \leqslant P(x) \leqslant 1$, and properly normalized, $\int dx\, P(x) = 1$ then $H[P]$ is maximized when $P(x) = \text{constant}$. This is what we found in Section 6.1. If we also constrain the mean, $\bar{x} = \int dx\, x\, P(x)$, then H is maximized by an exponential: $P(x) = Z e^{-\beta x}$ with Z and β fixed by imposing the mean and normalization constraints, as we found in Section 6.2. Can you guess what distribution maximizes $H[P]$ if we fix the mean $\bar{x}$ *and* the standard deviation σ?[1] Similarly, one finds the Poisson distribution and the binomial distribution as those that maximize $H[P]$ subject to appropriate constraints.[2] Thus, maxent gives a satisfying way to understand the universality of various statistical distributions.

---

[1] Answer: a Gaussian. You should check this yourself!

[2] A summary of some probability distributions and associated constraints can be found on wikipedia: https://en.wikipedia.org/wiki/Maximum_entropy_probability_distribution