Statistical Mechanics

Physics 181

Matthew D. Schwartz

Department of Physics
Harvard University

Spring 2025

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 1: Probability

1 Basic probability

We are going to be dealing with systems with enormous degrees of freedom, typically governed by Avogadro's number $N_A = 6.02 \times 10^{23}$. This is the number of hydrogen atoms in a gram, or more intuitively, the number of molecules of water in a tablespoon. Even a tiny cell, with a diameter of only 100 microns ($10^{-4}\, m$), contains a trillion molecules. In most areas of physics, we work with small numbers (the fine structure constant $\alpha = \frac{1}{137}$ for example), and calculate things as a Taylor series in the coupling $f(\alpha) = \sum c_n \alpha^n$, often keeping only the leading term $f(\alpha) \approx c_1 \alpha$. In statistical mechanics, we work with a large number N and calculate things as a Taylor expansion in $\frac{1}{N}$, often keeping only the leading term ($N = \infty$). The key to doing this is not to ask what each particle is doing, which would be both impossible and impractical, but rather to ask what the *probability* is that a particle is doing something. It is imperative therefore to begin statistical mechanics with statistics.

In general, we will be interested in probabilities of states of a system which we write as P_a or $P(a)$. The parameter a represents the microstate – e.g. the positions $\{\vec{q}_i\}$ and momenta $\{\vec{p}_i\}$ of all the particles in a gas, or the square of the wavefunction $|\psi(\vec{q})|^2$ in quantum mechanics. We will sometimes think of a as a discrete index (e.g. if we flip a coin, it can land heads up with $P_H = \frac{1}{2}$ or tails up with $P_T = \frac{1}{2}$) and sometimes continuous. In the continuous case, we call $P(x)$ the probability density, so that $\int_{x_1}^{x_2} P(x)\, dx$ is the probility of finding x values between x_1 and x_2. Probability densities only become probabilities when integrated.

We will get to know a number of different probability distributions:

$$\text{Gaussian:} \quad P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - x_0)^2}{2\sigma^2} \right) \tag{1}$$

$$\text{Poisson:} \quad P_m(t) = \frac{(\lambda t)^m}{m!} e^{-\lambda t} \tag{2}$$

$$\text{Binomial:} \quad B_N(m) = a^m b^{N-m} \frac{N!}{m!(N-m)!} \tag{3}$$

$$\text{Lorentzian:} \quad P(x) = \frac{\Gamma}{2\pi} \frac{1}{(x - x_0)^2 + \left(\frac{\Gamma}{2}\right)^2} \tag{4}$$

$$\text{Flat:} \quad P(x) = \text{constant} \tag{5}$$

Probabilities distributions are always normalized so that they integrate/sum to 1:

$$\int dx\, P(x) = 1, \qquad \sum_a P_a = 1 \tag{6}$$

Given a probability distribution, we can calculate the expected value of any observable by integrating/summing against the probability. For example, the expected value of x (the **mean**) is

$$\bar{x} \equiv \langle x \rangle = \int dx\, x\, P(x) \tag{7}$$

or the mean-square is

$$\langle x^2 \rangle = \int dx\, x^2\, P(x) \tag{8}$$

The **variance** of a distribution is the difference between the mean of the square and the square of the mean:

$$\text{Var} \equiv \langle x^2 \rangle - \langle x \rangle^2 \tag{9}$$

The square root of the variance is called the **standard deviation**.

$$\sigma \equiv \sqrt{\langle x^2\rangle - \langle x\rangle^2} \tag{10}$$

While the mean has the intuitive interpretation as the expected outcome, variance is more subtle. Indeed, developing intuition for variance is a key to mastering statistics.

For example, a Gaussian has two parameters, x_0 and σ_0. The first parameter is the mean:

$$\langle x\rangle = \int_{-\infty}^{\infty} dx\, x \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{(x-x_0)^2}{2\sigma_0^2}\right) = x_0 \tag{11}$$

The mean of x^2 is

$$\langle x^2\rangle = \sigma^2 + x_0^2 \tag{12}$$

So that the standard deviation is $\sigma = \sqrt{\langle x^2\rangle - \langle x\rangle^2} = \sigma_0$. This is why we usually write a Gaussian in this way ($e^{-x^2/2\sigma}$ rather than, say, $e^{-\lambda x^2}$).

The standard deviation has an interpretation as the width of a distribution – how far you can go from the mean before the probability has decreased substantially. For example, in a Gaussian, the probability of finding x between $x_0 - \sigma$ and $x_0 + \sigma$ is

$$(\Delta x)_{1\sigma} = \int_{x_0-\sigma}^{x_0+\sigma} dx\, P(x) = 0.68 \tag{13}$$

So, *for a Gaussian*, there is a 68% that the values of x fall within 1 standard deviation of the mean. That's not true for all distributions. For example, with a constant probability distribution, there is a 58% chance of finding x within 1σ of the mean.

We will often be interested in situations where the mean is zero. Then the standard deviation is equivalent to the **root-mean-square**

$$x_{\text{RMS}} = \sqrt{\langle x^2\rangle} \tag{14}$$

For example, in a gas the velocities point in random directions, so $\langle \vec{v}\rangle = 0$. Thus the characteristic speed of a gas is characterized not by the mean but by the RMS velocity $v_{\text{RMS}} = \sqrt{\langle \vec{v}^2\rangle}$.

Another important concept is how probability distributions behave when they are combined. For example, say $P_A(x)$ and $P_B(y)$ are the probabilities of winning x dollars when betting on horse A and y dollars when betting on horse B. The probability of getting z total dollars is then

$$P_{AB}(z) = \int_{-\infty}^{\infty} dx\, P_A(x) P_B(z-x) \tag{15}$$

This is the definition of the mathematical operation of **convolution** between two functions. We say P_{AB} is the convolution of P_A and P_B and write it as

$$P_{AB} = P_A * P_B \tag{16}$$

Convolutions are extremely important in statistical mechanics, since we often measure only the sum of a great many independent processes. For example, the pressure on the wall of a container is due to the sum of the forces of all the little molecules hitting it, each with its own probability.

1.1 Examples

Consider the system of a gas molecule bouncing around in a 1D box of size L centered on $x = 0$. If there are no external forces and no position-dependent interactions, the molecule is equally likely to be anywhere in the box. So

$$P(x) = \frac{1}{L} \tag{17}$$

The mean value of the position of the molecule is

$$\langle x\rangle = \frac{1}{L}\int_{-\frac{L}{2}}^{\frac{L}{2}} dx\, x = 0 \tag{18}$$

Similarly, the mean value of x^2 is

$$\langle x^2\rangle = \frac{1}{L}\int_{-\frac{L}{2}}^{\frac{L}{2}} dx\, x^2 = \frac{L^2}{12} \tag{19}$$

So that the standard deviation is

$$\sigma = \sqrt{\langle x^2\rangle - \langle x\rangle^2} = \frac{1}{\sqrt{12}} L \approx 0.29 L \tag{20}$$

Note that the probability of finding x within $\langle x\rangle \pm \sigma$ is $\frac{2\sigma}{L} = 58\%$. It is not 68% because the probability distribution is not Gaussian. This illustrates that the interpretation of σ as a 68% confidence interval is not always accurate.

Suppose instead that there is some electric field so that the particles in the box are more likely to be on one side than the other. We might find some crazy function $P(x) = \frac{0.74}{L}\ln(1 + e^{2x/L})$ for these probabilities. Then, by numerical integration we find

$$\langle x\rangle = 0.59 L, \quad \langle x^2\rangle = 0.42 L^2, \quad \sigma = 0.28 L \tag{21}$$

Also, $\int_{\langle x\rangle - \sigma}^{\langle x\rangle + \sigma} P(x) dx = 0.6$ so 60% within $\langle x\rangle \pm \sigma$. This is just a contrived example. You should be able to compute $\langle x\rangle$ and σ with any function $P(x)$, at least numerically, and you will generally find that not exactly 68% are within $\langle x\rangle \pm \sigma$, but often you get something close.

2 Law of large numbers

An extremely important result from probability is that even if $P(x)$ is very complicated, when you average over many measurements, the result dramatically simplifies. More precisely, the **law of large numbers** states that

The average of the results from a set of independent trials
varies less and less the more trials are performed

More mathematically, we can state it this way

- If $P(x)$ has standard deviation σ, then the probability $P_N(x)$ of finding that the average over N draws from $P(x)$ is x will have standard deviation $\frac{\sigma}{\sqrt{N}}$.

Thus as $N \to \infty$, the standard deviation of the average $\frac{\sigma}{\sqrt{N}} \to 0$.

To derive the law of large numbers, lets consider the probability distribution for the center of mass of molecules in a box. Say there are N molecules in the box and the probability function of finding each is $P(x)$. Some examples for $P(x)$ are Section 1.1. We assume that the probabilities for each molecule are independent – having one at x does not tell us anything about where the others might be. In this case, what is the mean value of the center of mass of the system? We'll write $\langle x\rangle_N$, $\langle x^2\rangle_N$ and σ_N for quantities involving the N-body system and drop the subscript for the $N = 1$ case: $\langle x\rangle_1 = \langle x\rangle$ and $\sigma_1 = \sigma$.

For $N = 2$, the center of mass is $x = \frac{x_1 + x_2}{2}$, so the mean value of the center of mass is

$$\langle x\rangle_2 = \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_2 P(x_1) P(x_2) \frac{x_1 + x_2}{2} = \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 P(x_1) \frac{x_1}{2} + \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_2 P(x_2) \frac{x_2}{2} = \langle x\rangle \tag{22}$$

So the mean value for 2 molecules is the same as for 1 molecule. The expectation of x^2 with 2 molecules is

$$\langle x^2\rangle_2 = \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_2 P(x_1) P(x_2) \left(\frac{x_1 + x_2}{2}\right)^2 \tag{23}$$

$$= \frac{1}{4}\int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 P(x_1) x_1^2 + \frac{1}{2}\int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 P(x_1) x_1 \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_2 P(x_2) x_2 + \frac{1}{4}\int_{-\frac{L}{2}}^{\frac{L}{2}} dx_2 P(x_2) x_2^2 \tag{24}$$

$$= \frac{1}{2}\langle x^2\rangle + \frac{1}{2}\langle x\rangle^2 \tag{25}$$

So the standard deviation of the center-of-mass for 2 particles is:

$$\sigma_2 = \sqrt{\langle x^2\rangle_2 - (\langle x\rangle_2)^2} = \sqrt{\frac{1}{2}\langle x^2\rangle + \frac{1}{2}\langle x\rangle^2 - \langle x\rangle^2} = \frac{1}{\sqrt{2}}\sqrt{\langle x^2\rangle - \langle x\rangle^2} = \frac{\sigma}{\sqrt{2}} \tag{26}$$

That is, the standard deviation has shrunk by a factor of $\sqrt{2}$ from the one particle case *for any* $P(x)$.

Now say there are N particles. The mean value of the center of mass is

$$\langle x\rangle_N = \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 \cdots dx_N P(x_1) \cdots P(x_N) \left(\frac{x_1 + \cdots + x_N}{N}\right) = \frac{1}{N}\left[N \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 x_1 P(x_1)\right] = \langle x\rangle \tag{27}$$

independent of N. The expectation value of x^2 is

$$\langle x^2\rangle_N = \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 \cdots dx_N P(x_1) \cdots P(x_N) \left(\frac{x_1 + \cdots + x_N}{N}\right)^2 \tag{28}$$

When we expand $(x_1 + \cdots + x_N)^2$ there are N terms that give $\langle x^2\rangle$ and the remaining $(N^2 - N)$ terms are the same as $\langle x_1 x_2\rangle = \langle x\rangle^2$. So,

$$\langle x^2\rangle_N = \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 \cdots dx_N P(x_1) \cdots P(x_N) \frac{1}{N^2}[N x_1^2 + (N^2 - N) x_1 x_2] \tag{29}$$

$$= \frac{1}{N}\langle x^2\rangle + \left(1 - \frac{1}{N}\right)\langle x\rangle^2 \tag{30}$$

Therefore

$$\sigma_N = \sqrt{\langle x^2\rangle_N - \langle x\rangle^2} = \frac{1}{\sqrt{N}}\sqrt{\langle x^2\rangle - \langle x\rangle^2} = \frac{\sigma}{\sqrt{N}} \tag{31}$$

The appearance of $\sqrt{N}$ is called **the law of large numbers**. Note that Eq. (31), describing how the standard deviation scales as we average over many molecules, holds for any function $P(x)$. Different $P(x)$ will give different values of σ, but the relation between σ_N with N molecules and σ with one molecule is universal.

For the gas in the box with a flat $P(x) = \frac{1}{L}$, as in Section 1.1, the expected value of the center of mass is $\langle x\rangle_N = 0$, just like for any individual gas molecules, and the standard deviation is $\sigma_N = \frac{\sigma}{\sqrt{N}} \approx 10^{-11}\frac{L}{\sqrt{12}}$. Thus, even though we don't know very well where any of the molecules are, we know the center of mass to extraordinary precision.

The law of large numbers is the reason that statistical mechanics is possible: we can compute macroscopic properties of systems (like the center of mass, or pressure, or all kinds of other things) with great confidence even if we don't know exactly what is going on at the microscopic level.

3 Central Limit Theorem

We saw how for when we average over a large number N of draws from a probability distribution $P(x)$ the mean stays fixed and the standard deviation shrinks by $\sigma \to \frac{\sigma}{\sqrt{N}}$. What can we say about the shape of the probability distribution $P_N(x)$? It turns out we can say a lot. In fact, in the limit $N \to \infty$ we know $P_N(x)$ exactly: it is a Gaussian!

More precisely the **central limit theorem** states that

> When *any* probability distribution is sampled N times
> the average of the samples approaches a Gaussian distribution as $N \to \infty$
> with width scaling like $\sigma \sim \frac{1}{\sqrt{N}}$

There are a lot of ways to prove it. I find the "moment" approach the most accessible, as discussed next. Another proof using convolutions an Fourier transforms in is Appendix C.

3.1 CLT proof using moments

One way to prove the central limit theorem is by computing moments. If you specify the complete set of moments of a function, you know its shape completely. These moments are

$$\text{mean:} \quad \bar{x} = \langle x \rangle \tag{32}$$

$$\text{variance:} \quad \sigma^2 = \langle x - \bar{x} \rangle^2 = \langle x^2 \rangle - \bar{x}^2 \tag{33}$$

$$\text{skewness:} \quad S = \frac{\langle (x-\bar{x})^3 \rangle}{\sigma^3} = \frac{1}{\sigma^3}[\langle x^3 \rangle - 3\bar{x}\langle x^2 \rangle + 2\bar{x}^3] \tag{34}$$

$$\text{kurtosis:} \quad K = \frac{\langle (x-\bar{x})^4 \rangle}{\sigma^4} \tag{35}$$

$$n^{\text{th}}\,\text{moment:} \quad M_n = \frac{\langle (x-\bar{x})^n \rangle}{\sigma^n} \tag{36}$$

Skewness measures how asymmetric a distribution is around its mean. Kurtosis measures the 4th derivative, which is a measure of curvature. More intuitively, higher kurtosis means a probability distribution has a longer tail, i.e. more outliers from the mean. The higher moments do not have simple interpretations.

Notice that all the higher-order moments are normalized by dividing by powers of σ so that they are dimensionless. To understand this normalization imagine plotting $P_N(x)$, but shift it to center around $x = 0$ and rescale the x axis by σ so that the width is always 1. Then the curve will not get any smaller as $N \to 0$ because its width is fixed to be 1, but its shape may change. The shape is determined by the numbers M_n with $n > 2$. See Fig. 2 below for an example.

For the Gaussian probability distribution in Eq. (1) the moments are easy to calculate in Mathematica:

$$\bar{x} = 0, \quad \sigma = \sigma, \quad S = 0, \quad K = 3, \quad M_5 = 0, \quad M_6 = 15, \quad M_7 = 0, \quad M_8 = 105, \cdots (\textbf{Gaussian}) \tag{37}$$

Note that skewness is zero for a Gaussian because it is symmetric. For a Gaussian, in fact all the odd moments (M_n with n odd) vanish. The even moments, normalized to powers of σ, are dimensionless numbers given by the formula

$$M_n = \begin{cases} 0 & , n\,\text{odd} \\ 2^{-\frac{n}{2}} \dfrac{n!}{\left(\frac{n}{2}\right)!} & , n\,\text{even} \end{cases} \tag{38}$$

These M_n completely determine the shape of a Gaussian. If a function has all of these moments, it is a Gaussian.

Now let's compute the moments of the center of mass of our N molecules-in-a-box with probability $P(x)$. We'll do this for a general $P(x)$, but shift the domain so that $\langle x \rangle = \bar{x} = 0$ in order to simplify the formulas in Eqs. (33)-(36). For example, the 3rd moment of $P_N(x)$ is

$$\langle x^3 \rangle_N = \int_{-\frac{L}{2}}^{\frac{L}{2}} dx_1 \cdots dx_N P(x_1) \cdots P(x_N) \left(\frac{x_1 + \cdots + x_N}{N} \right)^3 \tag{39}$$

Since $\langle x \rangle = 0$ the only terms in this expression which don't vanish are the ones of the form x_j^3. So

$$\langle x^3 \rangle_N = \frac{1}{N^2} \langle x^3 \rangle \tag{40}$$

We conclude that the skewness S_N with N molecules is related to the skewness S_1 for 1 molecule by

$$S_N = \frac{\langle (x-\bar{x})^3 \rangle_N}{\sigma_N^3} = \frac{\langle (x-\bar{x})^3 \rangle / N^2}{(\sigma/\sqrt{N})^3} = \frac{S_1}{\sqrt{N}} \tag{41}$$

In particular, the skewness goes to zero as $N\to\infty$. That is, the distribution becomes more and more symmetric abound the mean as $N\to\infty$.

Now let's look at the 4th moment, kurtosis. Following the same method we need

$$\langle x^4\rangle_N=\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_1\cdots dx_N P(x_1)\cdots P(x_N)\Big(\frac{x_1+\cdots+x_N}{N}\Big)^4 \tag{42}$$

In this case, since $\langle x\rangle=0$, the terms that don't vanish are x_j^4 or $x_j^2x_i^2$ with $i\neq j$. Thinking about the combinatorics a little you can convince yourself that there N terms of the form x_i^4 and $3N(N-1)$ terms of the form $x_i^2x_j^2$.[1] So,

$$\langle x^4\rangle_N=\frac{1}{N^3}\langle x^4\rangle+\frac{3(N-1)}{N^3}\langle x^2\rangle\langle x^2\rangle \tag{43}$$

Then, calling $K_1=\frac{1}{\sigma^4}\langle x^4\rangle$ the kurtosis for $N=1$ we have

$$K_N=\frac{\langle(x-\bar{x})^4\rangle_N}{\sigma_N^4}=\frac{1}{\sigma^4/N^2}\Big[\frac{1}{N^3}\langle x^4\rangle+\frac{3(N-1)}{N^3}\langle x^2\rangle\langle x^2\rangle\Big]=\frac{K_1}{N}+3\Big(1-\frac{1}{N}\Big) \tag{44}$$

This is interesting – it says that as $N\to\infty$ the kurtosis $K_N\to3$ *independent* of the kurtosis of the one particle probability distribution! So the skewness goes to zero and the kurtosis goes to 3.

For the 6th moment the term which dominates at large N is the non-vanishing one with the largest combinatoric factor: $\langle x^2\rangle^3$. There are ${}_NC_3\times{}_6C_2\times{}_4C_2=\frac{1}{6}N(N-1)(N-2)\times15\times2\to15$ of these. So $(M_6)_N\to15$. Similarly, $(M_8)_N\to105$. In other words, for any $P(x)$ we find that as $N\to\infty$

$$S_N\to0,\quad K_N\to3,\quad (M_5)_N\to0,\quad (M_6)_N\to15,\quad (M_7)_N\to0,\quad (M_8)_N\to105,\quad \cdots \tag{45}$$

What we are seeing is that at large N all of the higher moments go to those of a Gaussian! If you work out the details, the general formula is

$$(M_r)_N\to\begin{cases}0 & ,n\,\text{odd}\\ 2^{-\frac{r}{2}}\frac{r!}{\left(\frac{r}{2}\right)!} & ,n\,\text{even}\end{cases} \tag{46}$$

In exact agreement with the moments of a Gaussian. Thus we always get a Gaussian and the central limit is proven. Another proof using convolutions is in Appendix C.

3.2 Combining flat distributions

Because the central limit theorem is so important, let's try to understand why it is true more physically. Again, say we have some probability distribution $P(x)$ for molecules in a box, with $-\frac{L}{2}<x<\frac{L}{2}$. We want to pick N molecules and compute their mean position (center of mass position) $x=\frac{1}{N}\sum_j x_j$. What is the probability distribution $P_N(x)$ that the mean value is x?

To be concrete, let's take the flat distribution $P(x)=\frac{1}{L}$. For $N=1$, we pick only molecule with position x_1. Then $x=x_1$ and so $P(x)=\frac{1}{L}$: any value for the center-of-mass position is equally likely.

Now say $N=2$, so we pick two molecules with positions x_1 and x_2. What is the probability that they will have mean x? For a given x we need $\frac{x_1+x_2}{2}=x$. For example if $x=0$, then for any x_1 there is an x_2 that works, namely $x_2=-x_1$. However, if the mean is all the way on the edge, $x=\frac{L}{2}$, then not all x_1 work; in fact, we need both x_1 and x_2 to be exactly $\frac{L}{2}$. Thus there are fewer possibilities when x is close to the boundaries of the box than if x is central. One way to see this is graphically

[1] There are $\left(\begin{array}{c}N\\1\end{array}\right)=N$ of the x_j^4 terms. There are $\left(\begin{array}{c}N\\2\end{array}\right)=\frac{N!}{2!(N-2)!}=\frac{N(N-1)}{2}$ possible pairs $i\neq j$ and there are $\left(\begin{array}{c}4\\2\end{array}\right)=6$ ways of picking which two of the 4 terms in the expansion are i. So the total number of these terms is $3N(N-1)$.

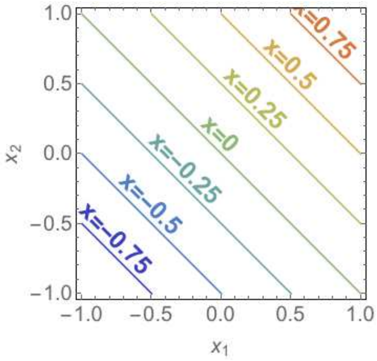


Figure 1. The regions in the x_1/x_2 plane with mean value x are diagonal lines for $L=2$. The length of the line is the probability $P_2(x)$. For $x=0$, the line is longest and probability greatest. For $x=1$, the line reduces to a point and the probability to zero.

To be quantitative, the easiest way to calculate the probability is with the Dirac δ function $\delta(x)$ (see Appendix A for a refresher on $\delta(x)$). Using the δ-function, we can write the probability for getting a mean value $x=\frac{x_1+x_2}{2}$ as

$$P_2(x)=\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_1P(x_1)\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_2\,P(x_2)\delta\Big(\frac{x_1+x_2}{2}-x\Big) \tag{47}$$

This is another way of writing a convolution, as in Eq. (15): $P_2=P*P$.

As a check, we can verify that this probability distribution is normalized correctly

$$\begin{aligned}\int_{-\frac{L}{2}}^{\frac{L}{2}}dx\,P_2(x)&=\int_{-\frac{L}{2}}^{\frac{L}{2}}dx\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_1P(x_1)\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_2\,P(x_2)\delta\Big(\frac{x_1+x_2}{2}-x\Big)\\&=\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_1P(x_1)\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_2\,P(x_2)=1\end{aligned} \tag{48}$$

where we have used the δ-function to integrate over x to get to the second line.

To evaluate $P_2(x)$ we first pull a factor of 2 out of the δ-function using Eq. (82), giving

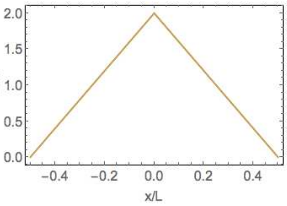
$$P_2(x)=2\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_1P(x_1)\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_2\,P(x_2)\delta(x_1+x_2-2x) \tag{49}$$

Now, the δ-function can only fire if its argument hits zero in the integration region. Since $\frac{x_1+x_2}{2}=x$ we can solve for $x_1=2x-x_2$. If $x<0$ then the most x_1 can be is $2x-\left(-\frac{L}{2}\right)=\frac{L}{2}+2x$. In other words, we have

$$P_2(x<0)=2\int_{-\frac{L}{2}}^{\frac{L}{2}+2x}dx_1P(x_1)P(2x-x_1) \tag{50}$$

Taking the flat distribution $P(x)=\frac{1}{L}$ this evaluates to $P_2(x<0)=2L+4x$. Similarly, for $x>0$ the limit is $x_1>2x-\frac{L}{2}$ and for a flat distribution $P_2(x>0)=2L-4x$. Thus we have

$$L^2P_2(x)=\begin{cases}2L+4x, & x<0\\ 2L-4x, & x>0\end{cases}= \tag{51}$$



You can also check this by evaluating Eq. (47) with Mathematica:

```
P=Integrate[DiracDelta[x1+x2-2x],{x1,-1,1},{x2,-1,1}];
```

Plot[P, {x, -1, 1}]

For $N=3$ we compute

$$P_3(x)=\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_1P(x_1)\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_2\,P(x_2)\int_{-\frac{L}{2}}^{\frac{L}{2}}dx_3\,P(x_3)\delta\Big(\frac{x_1+x_2+x_3}{3}-x\Big) \tag{52}$$

and so on. These successive approximations look like

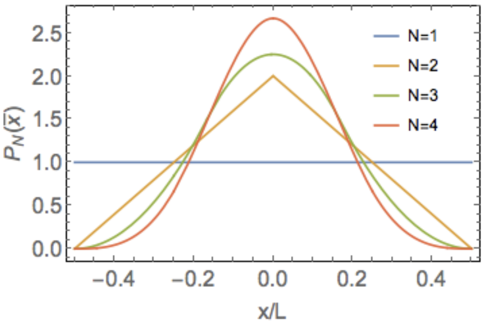


Figure 2. The average position of $N=1,2,3,4$ particles, each of which separately has a flat probability distribution.

We see that already at $N=4$ the flat probability distribution is becoming a Gaussian. Note also that the widths of the distributions are getting narrower.

The **central limit theorem** says that the distribution of the mean of N draws from a probability distribution approaches a Gaussian of width $\frac{\sigma}{\sqrt{N}}$ as $N\to\infty$ *independent* of the original probability distribution. That is,

$$P_N(x)\to\sqrt{\frac{N}{2\pi\sigma^2}}\exp\Big(-N\frac{(x-\bar{x})^2}{2\sigma^2}\Big) \tag{53}$$

Sometimes we sum the values of the draws from a distribution instead of averaging them. In this case, the mean grows as $\bar{x}\to N\bar{x}$ and the standard deviation grows like $\sigma\to\sqrt{N}\sigma$. Thus an equivalent phrasing of the central limit theorem is

- **Central Limit Theorem**: A function with mean $\bar{x}$ and standard deviation σ convolved with itself N times approaches a Gaussian with mean $N\bar{x}$ and standard deviation $\sqrt{N}\sigma$ as $N\to\infty$.

Summing the values is what happens when you convolve a function with itself. So for summing the values, the central limit theorem has the form

$$P_N^{\text{sum}}(x)=\underbrace{P*P*\cdots*P}_{N}\to\frac{1}{\sqrt{2\pi\sigma^2N}}\exp\Big(-\frac{(x-N\bar{x})^2}{2\sigma^2N}\Big) \tag{54}$$

A proof of the CLT using convolutions is in Appendix B.

We put the "sum" superscript to remind ourselves that we sum the values from each draw from $P(x)$ rather than average their values. The relation is simply

$$P_N^{\text{sum}}(x)=\frac{1}{N}P_N\Big(\frac{x}{N}\Big) \tag{55}$$

The $\frac{1}{N}$ comes from the fact that the probability distributions are differential, so we should technically write $P_N^{\text{sum}}(x)dx=P_N\big(\frac{x}{N}\big)\,d\frac{x}{N}$. Note when we average $\bar{x}\to\bar{x}$ and $\sigma\to\frac{\sigma}{\sqrt{N}}$ and when we sum $\bar{x}\to N\bar{x}$ and $\sigma\to\sqrt{N}\sigma$, so either way

$$\frac{\sigma}{\bar{x}}\to\frac{1}{\sqrt{N}}\frac{\sigma}{\bar{x}} \tag{56}$$

Thus a foolproof way to think of the scaling is that the dimensionless ratio $\frac{\sigma}{\bar{x}}$ should decrease as $\frac{1}{\sqrt{N}}$.

3.3 Why we take logarithms in statistical mechanics

In statistical mechanics, we will make great use out of the central limit theorem. Generally we have systems composed of enormously large numbers of particles $N \sim$ Avogadro's number $\sim 10^{24}$. The things we measure are macroscopic: the pressure a gas puts on a wall is the *average* pressure. Microscopically, the gas has a bunch of little molecules hitting and bouncing off the wall and the force these molecules impart is constantly varying. We don't care about these tiny fluctuations, just the average. So any time we try to measure something, like the pressure in a gas, or the concentration of a chemical, we will necessarily be averaging over an enormous number of fluctuations. Because of the central limit theorem, the distribution of any macroscopic quantity will be close to a Gaussian around its mean. This central limit theorem itself doesn't tell us what the mean is, or how various macroscopic quantities are related – we need physics for that. But it tells us that we don't need to worry about the precise details of the microscopic description.

Normally when a function $f(x)$ is rapidly falling away from $x \approx \bar{x}$ we Taylor expand $x = \bar{x}$ and keep the first few terms. We can do this for $P_N(x)$ too. However, the Taylor expansion of a Gaussian has an infinite number of terms

$$e^{-\frac{x^2}{2\sigma^2}} = \sum_{m=0}^{\infty} \frac{1}{m!}\left(-\frac{x^2}{2\sigma^2}\right)^m = 1 - \frac{x^2}{2\sigma^2} + \frac{1}{2}\left(\frac{x^2}{2\sigma^2}\right)^2 - \frac{1}{6}\left(\frac{x^2}{2\sigma^2}\right)^3 + \cdots \tag{57}$$

You need all the terms to reconstruct the original Gaussian. However, if we take the logarithm first, then Taylor expand, we find

$$\ln e^{-\frac{x^2}{2\sigma^2}} = -\frac{x^2}{2\sigma^2} \tag{58}$$

with only one term. So it will be extremely convenient to start taking the logarithms of our probabilities. By the central limit theorem, when we average the values,

$$\ln P_N(x) \to -N\frac{(x-\bar{x})^2}{2\sigma^2} + \ln\sqrt{\frac{N}{2\pi\sigma^2}} \tag{59}$$

As $N \to \infty$ there are no higher order terms.

In other words, a Gaussian is an unusual function. It is flat near the peak, but then quickly drops off and has a long tail. Since the function is smooth near the peak, it's hard to know what's going on at the tail from expanding near the peak. In particular, you have to work very hard to get information about points with $x \gtrsim \sigma$ from information at the peak. Taking the logarithm puts the peak and the tail on the same footing. Of course, we can't get something for nothing: taking logarithms alone won't solve any problems. But taking logarithms often makes it easier to solve problems. We will see many examples of this as the course progresses.

4 Poisson distribution

In many physical situations, there is a large number N of possible events each occurring with very small probability λ for a given time interval. For example, if you put a glass out in the rain, there are lots of possible drops of water that could fall into the glass, but each has a small probability. Or you have lots of friends on Instagram, each one has a small probability of posting something interesting. Or we have a gas of molecules and each one has a small chance of being in some tiny volume. Probabilities in situations like this, where each event is uncorrelated with the previous event, are described by the Poisson distribution.

Let's take a concrete example, radioactive decay. A block of ^{235}U has $N \sim 10^{24}$ atoms each of which can decay with a tiny probability

$$dP = \lambda dt \tag{60}$$

λ is called the **decay rate**. It has units of $\frac{1}{\text{time}}$. For a single atom of ^{235}U, this decay rate is $\lambda = 3 \times 10^{-17} s^{-1}$. In a mole of Uranium ($10^{24}$ atoms), 10^7 Uranium atoms decay, on average, each second. What is the chance of seeing m decays in a time t?

Let's start with $m=0$ and the time t very small (compared to $\frac{1}{\lambda}$), $t=\Delta t$. If the rate to decay is $dP=\lambda dt$ then the probability of not decaying in time $t=\Delta t$ is

$$P_{\text{no decay}}(\Delta t)=1-\lambda\Delta t \tag{61}$$

For the system to survive to a time $2\Delta t$ with no decays, it would have to not decay in Δt and then not decay again in the next Δt. Since the probability of two uncorrelated occurrences (or not-occurrences in this case) is the product of the probabilities, $P(a\&b)=P(a)P(b)$ we then have

$$P_{\text{no decay}}(2\Delta t)=(1-\lambda\Delta t)^2 \tag{62}$$

Now we can get all the way to time t by sewing together small times $\Delta t=\frac{t}{N}$ and taking $N\to\infty$. We thus have

$$P_{\text{no decay}}(t)=\lim_{N\to\infty}\left(1-\lambda\frac{t}{N}\right)^N=e^{-\lambda t} \tag{63}$$

So that's the $m=0$ case: no particles decay.

Using this formula, how long will it take for the probability of some decay to be $\frac{1}{2}$? That's the same as the probability of no decay being $1-\frac{1}{2}=\frac{1}{2}$. So we just solve

$$\frac{1}{2}=e^{-\lambda t_{1/2}} \qquad\Rightarrow\qquad t_{1/2}=\frac{1}{\lambda}\ln 2=\frac{0.7}{\lambda} \tag{64}$$

We often say $\frac{1}{\lambda}$ is the **lifetime** and $t_{1/2}$ is the **halflife**. The two numbers are related by a factor of ln2: $t_{1/2}=\frac{1}{\lambda}\ln 2$.

Now try $m=1$. We need the probability that there is exactly one decay in exactly one of the time intervals. There are N intervals we can pick. So,

$$P_{1\text{ decay}}(t)=\lim_{N\to\infty}N\underbrace{\left(1-\lambda\frac{t}{N}\right)^{N-1}}_{N-1\text{ no decays}}\underbrace{\left(\lambda\frac{t}{N}\right)}_{\text{one decay}}=\lim_{N\to\infty}-t\partial_t\left(1-\lambda\frac{t}{N}\right)^N \tag{65}$$

In the third term, we have simply rewritten the expression in a smart way with a derivative so we can reduce it to a previously solved problem – a powerful physicist trick. Now we switch the order of the limit and the ∂_t and use Eq. (63) to get

$$P_{1\text{ decay}}(t)=-t\partial_t P_{\text{nodecay}}(t)=\lambda t e^{-\lambda t} \tag{66}$$

For two decays there are $\left(\begin{smallmatrix}N\\2\end{smallmatrix}\right)=\frac{N!}{(N-2)!2!}=\frac{1}{2}N(N-1)$ ways and we have

$$P_{2\text{decays}}(t)=\lim_{N\to\infty}\underbrace{\frac{N(N-1)}{2}}_{\text{pick 2 to decay}}\underbrace{\left(1-\lambda\frac{t}{N}\right)^{N-2}}_{N-2\text{ no decays}}\underbrace{\left(\lambda\frac{t}{N}\right)^2}_{\text{two decays}}=\frac{1}{2}t^2\partial_t^2P_{\text{no decay}}(t)=\frac{(\lambda t)^2}{2}e^{-\lambda t} \tag{67}$$

For general m the result is

$$\boxed{P_m(t)=\frac{(\lambda t)^m}{m!}e^{-\lambda t}} \tag{68}$$

This is called the **Poisson distribution**. It gives the probability for exactly m events in time t when each event has a probability per unit time of λ and the events are uncorrelated.

In any time t there must have been some number of decays between 0 and ∞. Indeed,

$$\sum_m P_m(t)=\sum\frac{(\lambda t)^m}{m!}e^{-\lambda t}=1 \tag{69}$$

So that's consistent (as is the t-independence of this sum).

The way we derived the Poisson distribution was for a fixed m, as a function of t. But it can be more useful to think of it as a function of m at a fixed value of t: $P(m,t)=P_m(t)$. Keep in mind though that for fixed t, $P(m,t)$ as a function of m is a discrete probability distribution (meaning m is an integer). In contrast for fixed m, $P(m,t)$ is a continuous function of t. Moreover, while it is a normalized probability in m, it is a simply a function (not a probability distribution) of t. There is not a sense in which $\int dt P_m(t)=1$; this doesn't even have the right units.

For a given fixed t, how many particles do we expect to have decayed? In other words, what is the expected value $\langle m \rangle$ in a time t? We compute the mean value for m, by summing the value of m times the probability of getting m

$$\langle m \rangle = \sum_m m P_m(z) = \sum_m m \frac{(\lambda t)^m}{m!} e^{-\lambda t} = \lambda t \tag{70}$$

The last step is a little tricky – see if you can figure out how to do the sum yourself. (You can always run Mathematica if you get stuck on steps like this.) The result implies that the expected number of decays in a time t is λt. It makes sense that if you double the time, twice as many particles decay. How long will it take for half the particles to decay?

The standard deviation of the Poisson distribution is

$$\sigma = \sqrt{\langle m^2 \rangle - \langle m \rangle^2} = \sqrt{\lambda t} \tag{71}$$

Again, you can check this yourself as an exercise.

So the Poisson distribution as a function of m at fixed t has mean λt and width $\sqrt{\lambda t}$. Thus the width compared to the mean is

$$\frac{\sigma}{\langle m \rangle} = \frac{1}{\sqrt{\lambda t}} \tag{72}$$

This goes to 0 as $t \to \infty$. In other words, the Poisson distribution is narrower and narrower as t gets larger. What does this mean physically? It means if we wait one lifetime ($t = \frac{1}{\lambda}$) we should expect 1 ± 1 particle to decay. If we wait 2 lifetimes, we expect $2 \pm \sqrt{2}$ to decay ($t = \frac{2}{\lambda}$, $\langle m \rangle = 2$ and $\sigma = \frac{2}{\sqrt{2}} = \sqrt{2}$). If we wait 100 lifetimes, we expect 100 ± 10 to decay. So the longer we wait, not only are there more decays, but we know more precisely how many decays there will be. This is, of course, a consequence of the central limit theorem.

So what do you expect the distribution to look like as $t \to \infty$ or $m \to \infty$? Let's first look numerically. We can plot $P_m(t)$ as a function m, which is a discrete index, or as a function of t, which is continuous:

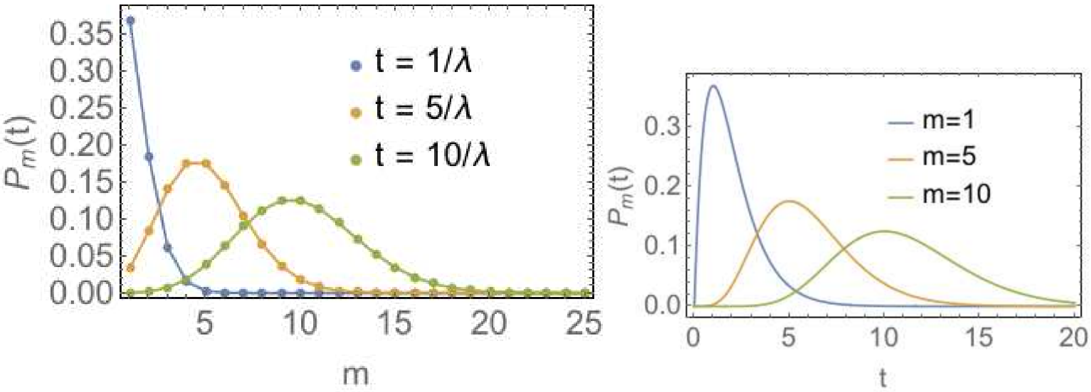


Figure 3. The Poisson distribution as a function of the discrete index m for various times (left) and time, for various values of m (right)

We clearly see the Gaussian shape emerging at large t (left) and at large m (right).

Now let's try to see how the Gaussian form arises analytically. First of all, we want the high statistics limit, which means large t in units of $\frac{1}{\lambda}$ which also means large m. When you see a factor of $m!$ and want to expand at large m, you should immediately think **Stirling's approximation**:

$$x! \approx e^{-x} x^x \times (\cdots) \tag{73}$$

or equivalently

$$\ln x! \approx x \ln x - x + \cdots \tag{74}$$

For a simple derivation, see Appendix B. We will use this expansion a lot.

The log of the Poisson distribution is

$$\ln P_m(t) = \ln\left[\frac{(\lambda t)^m}{m!}e^{-\lambda t}\right] = m\ln(\lambda t) - \lambda t - \ln m! \tag{75}$$

Then we use Stirling's approximation for $m!$

$$\ln P_m(t) \xrightarrow[m\gg 1]{} m\ln(\lambda t) - \lambda t - m\ln m + m + \cdots = m\ln\frac{\lambda t}{m} + (m - \lambda t) + \cdots \tag{76}$$

This is still a mess. But we expect $P_m(t)$ to be peaked around its mean $\langle m\rangle = \lambda t$. So let's Taylor expand $\ln P_m(t)$ around $m = \lambda t$. The leading term, from setting $m = \lambda t$ makes Eq. (76) vanish. The next term is

$$\frac{\partial}{\partial m}\ln P_m(t)\bigg|_{m=\lambda t} = \lim_{m\to\lambda t}[\ln(\lambda t) - \ln m] = 0 \tag{77}$$

which also vanishes. We have to go one more order in the Taylor expansion to get a nonzero answer:

$$\frac{\partial^2}{\partial m^2}\ln P_m(t)\bigg|_{m=\lambda t} = \lim_{m\to\lambda t}\left[-\frac{1}{m}\right] = -\frac{1}{\lambda t} \tag{78}$$

Thus,

$$\ln P_m(t) = -\frac{1}{2\lambda t}(m - \lambda t)^2 + \cdots \tag{79}$$

and therefore

$$P_m(t) \xrightarrow[m\gg 1]{} \frac{1}{\sqrt{2\pi\lambda t}}e^{-\frac{(m-\lambda t)^2}{2\lambda t}} \tag{80}$$

This is a Gaussian with mean $\langle m\rangle = \lambda t$ and width $\sigma = \sqrt{\lambda t}$ exactly as expected by the central limit theorem.

You might not be terribly impressed with this derivation as a check of the central limit theorem. After all, we expanded $\ln P_m$ to second order around $m = \langle m\rangle$. Doing that, for any function P_m is guaranteed to give a Gaussian. But that's really the whole point of the central limit theorem – any function *does* give a Gaussian. So in the end you should be impressed after all.

5 Summary

In this lecture, we introduced the basic concepts from probability that will be useful for statistical mechanics. The key concepts are

- Normalized **probability distributions** $P(x)$ with $\int dx P(x) = 1$
- **Mean**: $\bar{x} = \langle x\rangle = \int dx x P(x)$
- **Variance** $\text{var} = \int dx (x - \bar{x})^2 P(x)$,
- **Standard deviation** or **width** $\sigma = \sqrt{\text{var}}$
- **Gaussian** distribution $P(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right)$ has mean $\bar{x}$ and width σ.
- If you draw x from Gaussian it is 68% likely to between $\bar{x} - \sigma$ and $\bar{x} + \sigma$.
- The **convolution** of two distributions is defined as $(P_A * P_B)(z) = \int_{-\infty}^{\infty} dx P_A(z - x) P_B(x)$. It describes the probabilty of getting z as the sum of a number draw from P_A and another number drawn from P_B.
- Given a probabilty distribition $P(x)$ with mean $\bar{x}$ and width σ, you can construct a new probabilty distribution $P_N(x)$ by averaging over N draws from $P(x)$. The **central limit thoerem (CLT)** says that as $N \to \infty$ this new distribution will approach a Gaussian with the same mean as $P(x)$ ($\bar{x}_N = \bar{x}$) and a smaller standard deviation $\sigma_N \approx \frac{\sigma}{\sqrt{N}}$. All other properties of $P(x)$ are lost after this averaging at large N.
- The CLT also implies that if we *sum* (rather than average) the values from draws, the mean grows like $\bar{x}_N \approx N\bar{x}$ and the standard deviation like $\sigma_N \approx \sqrt{N}\sigma$. If we

- Because of the CLT, Gaussians are very common. Their exponential decay encourages us to study logarithms of distributions, which turns fast-varying exponentials into slow-varying polynomials: $\ln e^{-\frac{x^2}{2\sigma^2}} = -\frac{x^2}{2\sigma^2}$.
- When we have a rate $dP = \lambda dt$ for an event happening that is independent of time, then the probability of having m events after a time t is described by the **Poisson distribution** $P_m(t) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}$.
- **Stirling's approximation** is that $N! \approx \sqrt{2\pi N} N^N e^{-N}$ at large N. This works very well, even at $N = 1$.

Appendix A Dirac δ-function

The Dirac δ-function is very useful in physics, from quantum mechanics to statistical mechanics. The δ-function is not really a function but rather a distribution. $\delta(x)$ is zero everywhere except at $x = 0$. When you integrate a function against $\delta(x)$ you pick up the value of that function at 0. That is

$$\int dx \delta(x) f(x) = f(0) \tag{81}$$

This is the defining property of $\delta(x)$. The integration region has to include $x = 0$ but is otherwise arbitrary since $\delta(x) = 0$ if $x \neq 0$.

Another useful property of δ-functions is that if we rescale the argument of $\delta(x)$ by a number a then the δ-function rescales by $\frac{1}{a}$. That is,

$$\delta(ax) = \frac{1}{a}\delta(x) \tag{82}$$

To check this, we can change variables from $x \to \frac{x}{a}$ in the integral

$$\int dx \delta(ax) f(x) = \int d\frac{x}{a} \delta(x) f\left(\frac{x}{a}\right) = \frac{1}{a} f(0) = \int dx \left[\frac{1}{a}\delta(x)\right] f(x) \tag{83}$$

It's sometimes helpful to think of the δ function as the limit of a regular function. There are lots functions whose limits are δ functions. For example, Gaussians:

$$\delta(x) = \lim_{\sigma \to 0} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \tag{84}$$

As a check, note that the integral over the Gaussian is 1 regardless of σ, so the δ function also integrates to 1. As $\sigma \to 0$, the width of the Gaussian goes to zero, so it has zero support away from mean, that is it vanishes except at $x = 0$, just like the δ function.

Appendix B Stirling's approximation

There are many ways to derive Stirling's approximation. Here's a relatively easy one. We start by taking the logarithm

$$\ln N! = \ln N + \ln(N-1) + \ln(N-2) + \cdots + \ln 1 = \sum_{j=1}^{N} \ln j \tag{85}$$

For large N we then write the sum as an integral

$$\ln N! = \sum_{j=1}^{N} \ln j \approx \int_1^N dj \ln j = N \ln N - N - 1 \approx N \ln N - N \tag{86}$$

That's the answer.

One can include more terms in the expansion by using the Euler-McLauren formula for the difference between a sum and an integral. For example, the next term is

$$N! \approx \sqrt{2\pi N} N^N e^{-N} \tag{87}$$

An alternative derivation is to given an integral representation of the factorial as a Γ function: $n! = \Gamma(n+1) = \int_0^\infty x^n e^{-x} dx$. For example, Mathematica can simply series expand this around $n = \infty$ to reproduce Eq. (87). Try it!

The next order correction to this is down by $\frac{1}{12N}$, which gets small fast. You can check that Stirling's approximation is off by less than 8% already at $N = 1$ and by less than 2% by $N = 3$. For Avogadro's number $N = 6 \times 10^{23}$ it is off by one part in 10^{25}.

Appendix C Central limit theorem from convolutions

Here's another slick proof of the central limit theorem. We start with the definition

$$P_N(x) = \int dx_1 ... dx_n P(x_1) ... P(x_n) \delta\left(\frac{x_1 + \cdots + x_n}{N} - x\right) \tag{88}$$

Now we write the δ function in Fourier space as

$$\delta(x_1 + \cdots + x_n - x) = \int \frac{dk}{2\pi} e^{ik(x_1 + \cdots + x_n - x)} \tag{89}$$

So that

$$P_N(x) = \int dk \int dx_1 ... dx_n P(x_1) ... P(x_n) e^{ik\left(\frac{x_1 + \cdots + x_n}{N} - x\right)} \tag{90}$$

Defining the Fourier transform of P as

$$\tilde{P}(k) = \int dx e^{ikx} P(x) \tag{91}$$

we then have

$$P_N(x) = \int \frac{dk}{2\pi} \left[\tilde{P}\left(\frac{k}{N}\right)\right]^N e^{ikx} \tag{92}$$

Eq. (92) is just the statement that Fourier transforms turns convolutions into products.Then,

$$\tilde{P}\left(\frac{k}{N}\right) = \int dx e^{i\frac{kx}{N}} P(x) = \int dx \left(1 + \frac{ikx}{N} - \frac{1}{2}\left(\frac{kx}{N}\right)^2 + \frac{1}{3!}\left(\frac{ikx}{N}\right)^3 + \cdots\right) P(x) \tag{93}$$

$$= 1 + i\frac{k}{N}\langle x\rangle - \frac{k^2}{2}\frac{\langle x^2\rangle}{N^2} - i\frac{k^3\langle x^3\rangle}{6N^3} + \cdots \tag{94}$$

Now if we didn't do anything else, then as $N \to \infty$ we see immediately that $\tilde{P}\left(\frac{k}{N}\right) \to 1$ and so $P_N(x) \to \delta(x)$. This is because the whole distribution is shrinking down to be around $x = 0$. That result is not wrong, but it's happening because the width of the Guassian is going to zero. We want to work to first subleading order in $\frac{1}{N}$, i.e. keep the factor of N in the width.

To proceed, let's first set $\langle x\rangle = 0$ by shifting the offset of x. Then $\langle x^2\rangle = \sigma^2$ and we have

$$\left[\tilde{P}\left(\frac{k}{N}\right)\right]^N = \left[1 - \frac{k^2}{2}\frac{\sigma^2}{N^2} - i\frac{k^3\langle x^3\rangle}{6N^3} + \cdots\right]^N \tag{95}$$

Now, the biggest terms in this product will come from taking as many factors of 1 as we can. If you take one of the other terms you pay at least $\frac{1}{N^2}$. So taking either all $1's$ or only one non-1 term, we get

$$\left[\tilde{P}\left(\frac{k}{N}\right)\right]^N = 1 - N\left(\frac{k^2}{2}\frac{\sigma^2}{N^2} - i\frac{k^3\langle x^3\rangle}{6N^3} + \cdots\right) + \cdots \tag{96}$$

From here we see that the first term subleading in $\frac{1}{N}$ is the σ^2 term and the rest is order $\frac{1}{N^2}$ or lower. Thus to get the answer write to order $\frac{1}{N}$ we write

$$\left[\tilde{P}\left(\frac{k}{N}\right)\right]^N = \left[1-\frac{1}{N}\left(\frac{k^2}{2}\frac{\sigma^2}{N}\right)\right]^N + \mathcal{O}\left(\frac{1}{N^2}\right) \xrightarrow[N\gg 1]{} e^{-\frac{k^2}{2}\left(\frac{\sigma^2}{N}\right)} + \mathcal{O}\left(\frac{1}{N^2}\right) \tag{97}$$

where $e^x = \lim_{N\to\infty}\left(1+\frac{x}{N}\right)^N$ was used. We can then compute the inverse Fourier transform to get

$$P_N(x) = \int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{-\frac{k^2}{2}\left(\frac{\sigma^2}{N}\right)}e^{ikx} = \sqrt{\frac{2N}{\sigma^2}}\int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{-\frac{k^2}{2}}e^{ik\left(\frac{\sqrt{2N}x}{\sigma}\right)} = \sqrt{\frac{N}{2\pi\sigma^2}}e^{-\frac{Nx^2}{2\sigma^2}} \tag{98}$$

which is the desired result, the central limit theorem.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 2: Diffusion

1 Introduction

If you we put a drop of red dye in water, it will slowly diffuse throughout the water. Why does this happen? How fast does it happen? What is going on microscopically?

The microscopic mechanism of diffusion is very simple: the dye molecules start densely concentrated near one point. Then they get bumped by neighboring molecules until they are spread out all over. To model this process, we can suppose that the dye molecule moves a distance ℓ between collisions and after each collision its direction is completely randomized. This approximation is called a **random walk**. Although the distance ℓ between collisions has some variation and the direction of scattering is somewhat correlated with the initial direction, because molecules collide billions of times per second, the law of large numbers applies to their net displacement and random walks provide an excellent approximation to real diffusion.

Random walks are actually quite common. They can be used to model any stochastic process. For another example, say you're playing blackjack with a friend. You are both expert players and evenly matched. Sometimes you win, sometimes she wins. Each time you play, you bet 1 dollar. This is a 1-D random walk. Say you play N games. Although we can't say who will be winning after N games, we can predict how by much they would be winning.

The 2-dimensional random walk is sometimes called the **drunkard's walk**. The idea is that a drunkard leaves a party late at night, takes a step in one direction, then gets totally disoriented and takes a step in another direction. How far will she get after N steps?

2 1D random walk

Let's work out the blackjack problem. We'll make it a little more interesting. Say you have a probability a of winning and your opponent has a probability $b=1-a$ of winning. If you play N times, the chance of you winning m of them is

$$B_N(m)=a^m b^{N-m}\binom{N}{m} \tag{1}$$

This is known as the **binomial distribution**. The factor "N choose m"

$$\binom{N}{m}={}_N C_m=\frac{N!}{m!(N-m)!} \tag{2}$$

is known as the **binomial coefficient**. It is the number of ways of picking m of the games for you to have won out of the N total games.[1]

Binomial coefficients come up in the expansion of powers of sums. Namely

$$(x+y)^N=x^N+Nx^{N-1}y+\cdots+y^N=\sum_{m=0}^{N} x^m y^{N-m}\binom{N}{m} \tag{3}$$

So the binomial distribution is simply coefficient of the m^{th} term in this sum. In fact, this relationship makes it easy to see that the probabilities in the blackjack game sum to 1:

$$\sum_{m=0}^{N} B_N(m)=\sum_{m=0}^{N} a^m b^{N-m}\binom{N}{m}=(a+b)^N=1 \tag{4}$$

[1] To see this, first note that if we want to sort N objects, N of them can go first, then $N-1$ second, and so on, so there are $N!$ permutations. If we are instead splitting into a set of size m and a set of size $N-m$, we don't care about the order within each set. So we have to divide by the permutations of each set, giving $\frac{N!}{m!(N-m)!}$.

since $b=1-a$.

How much can you expect to be winning after N games? This is determined by the number of games you win, namely the expected value of m:

$$\langle m\rangle = \sum_{m=0}^{N} m B_N(m) = \sum_{m=0}^{N} m a^m b^{N-m}\left(\begin{array}{c} N \\ m \end{array}\right) \tag{5}$$

Although $b = 1 - a$ we can compute the sum on the right most easily if we allow a and b to be unrelated. Then, we note that this sum is the same as the sum in Eq. (4) if we mulitiply by the power of a. We can pull down this power if we differentiate with respect to a, then multiply by a. So we have

$$\sum_{m=0}^{N} m a^m b^{N-m}\left(\begin{array}{c} N \\ m \end{array}\right) = a\partial_a\left[\sum_{m=0}^{N} a^m b^{N-m}\left(\begin{array}{c} N \\ m \end{array}\right)\right] = a\partial_a(a+b)^N = Na(a+b)^{N-1} \tag{6}$$

Now that we can compute the sum for any a and b we can take the case of interest where $b=1-a$ and get

$$\langle m\rangle = Na \tag{7}$$

Similarly, the standard deviation is (check this, using the same trick!)

$$\sigma_m = \sqrt{\sum_{m=0}^{N} (m^2 - \langle m\rangle^2) B_N(m)} = \sqrt{Nab} \tag{8}$$

As $N\rightarrow\infty$ the binomial distribution approaches a Gaussian, by the central limit theorem. Thus knowing the mean and standard deviation, we know the whole answer at large N:

$$B_N(m) \rightarrow \frac{1}{\sqrt{2\pi Nab}}\exp\left[-\frac{(m-Na)^2}{2Nab}\right] \tag{9}$$

We can derive this by studying the falloff of ln $B_N(m)$ at large N, but using the central limit theorem is easier. Try checking the agreement of Eqs. (1) and (9) for some numerical values.

For the blackjack game, $\langle m\rangle = Na$ is the expected number of times you win. The expected number of times you lose is $\langle m\rangle = Nb$. If you gain $\$1$ when you win, or lose $\$1$ when you lose, then

$$\langle \text{winnings}\rangle = N(a-b) \times \$1 \tag{10}$$

and the standard deviation is (check this!):

$$\sigma_{\text{winnings}} = 2\sqrt{Nab} \times \$1 \tag{11}$$

For a fair match $a=b=\frac{1}{2}$ and so the expected winnings are $\langle\text{winnings}\rangle = 0$ with standard deviation

$$\sigma = \sqrt{N} \tag{12}$$

That σ grows as $\sqrt{N}$ is exactly what we expect for the sum of random values by the central limit theorem. The $a=b=\frac{1}{2}$ case is sometimes called an **unbiased 1D random walk**.

For example, if you play 100 games for \$1 each and are evenly matched, then $\langle\text{winnings}\rangle = 0$ and $\sigma_{\text{winnings}} = \sqrt{100} \times \$1 = \$10$. This means that after 100 games, we don't know who's winning but there is a 32% chance someone is up by at least \$10.

For an unbiased 1D random walk, the mean displacement is 0. In this case, the typical scale for displacement is better described by the root-mean-square (RMS) fluctuation, which reduces to the standard deviation when the mean is zero. That is, the RMS fluctuation is $x_{\text{rms}} = \sigma = \sqrt{N}$. Typically, RMS fluctuations are used for quantities that average to zero (as in a unbiased random walk), but there is no hard and fast rule about when to use the mean displacement and when to use the RMS fluctuation. Just like there are many measures of uncertainty (σ, fraction within 50% of mean, width at half maximum, etc), there are many measures of discplacement ($\bar{x}$, x_{rms}, etc.). Typically these are all similar (when nonzer) and determined by dimensional analysis, so you can just substitute one for another, often depending on which leads to the simplest looking final expression.

Let us compare the binomial distribution to the Poisson distribution. The binomial distribution $B_N(m)$ is defined for discrete N and m, in contrast to the Poisson distribution $P_m(t) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}$ which has discrete m but continuous t. For a binomial distribution, the smallest interval is one discrete step, with probability of occurrence a. For Poisson, we can take an arbitrarily small timestep Δt with probability $dP = \lambda \Delta t$. If we identify the interval for the binomial distribution with that of the Poisson distribution by setting $a = \lambda \Delta t = \lambda \frac{t}{N}$, and then take the limit $N \to \infty$ we find

$$\lim_{N\to\infty} B_N(m) = \lim_{N\to\infty} \underbrace{\begin{pmatrix} N \\ m \end{pmatrix}}_{\text{pick } m \text{ to decay}} \underbrace{\left(1 - \lambda \frac{t}{N}\right)^{N-m}}_{N-m \text{ do not decay}} \underbrace{\left(\lambda \frac{t}{N}\right)^m}_{m \text{ dceays}} = \frac{(\lambda t)^m}{m!} e^{-\lambda t} \tag{13}$$

which formally recovers the Poisson distribution from the binomial distribution. Note that this requires taking $N \to \infty$ holding $\lambda t = Na$ fixed, and therefore $a \to 0$ and $b \to 1$. In other words, a Poisson process is like a random walk that always goes in one direction but you don't know when the step will be taken.

Although they can be related, as we have seen, you should really think of binomial and Poisson distributions as being relevant in different contexts: binomial is used when the steps are discrete and incoherent (random directions) and Poisson is used when time is continuous but the steps are coherent (counts always increase). If we are flipping coins, then $a = \frac{1}{2}$ is fixed, and so the Poisson distribution is not relevant since it needs $a = 0$. For a decay process, a decay can happen at any time t and so the binomial distribution is not appropriate. After a given time t, a Poisson process can have potentially an infinite number of events. With a binomial process the time is the number of steps N, so the number of possible events is always bounded.

2.1 Random walks in 2D and 3D

For the 2D case, a popular picture of the random walk is a drunkard stumbling around. In each time step she moves a distance L in some random direction. In 3D you can imagine a dye molecule diffusing in water and in each time step it bumps into something, and then gets buffeted into a different direction. For simplicity, we'll assume in the 2D and 3D cases that the distance is the same each step and the angle totally random. Where will the drunkard or molecule be after N steps?

Let us say that in the j^{th} step she moves by a displacement $\vec{\ell}_j$. The vectors $\vec{\ell}_j$ all have length ℓ. The dot product of two vectors is

$$\vec{\ell}_j \cdot \vec{\ell}_k = \ell^2 \cos\theta_{jk} \tag{14}$$

where θ_{jk} is the angle between the two steps. Since we are assuming the angle is random, then the expectation value of this dot product is zero:

$$\langle \vec{\ell}_j \cdot \vec{\ell}_k \rangle = \ell^2 \frac{1}{\pi} \int_0^\pi d\theta \cos\theta = 0 \tag{15}$$

Now let $\vec{s}_N$ be the total displacement from the origin after N timesteps.

$$\vec{s}_N = \sum_{j=1}^N \vec{\ell}_j \tag{16}$$

Then

$$\langle \vec{s}_N^2 \rangle = \langle (\vec{s}_{N-1} + \vec{\ell}_N)^2 \rangle = \langle \vec{s}_{N-1}^2 \rangle + 2 \langle \vec{s}_{N-1} \cdot \vec{\ell}_N \rangle + \langle \vec{\ell}_N^2 \rangle \tag{17}$$

Now, the angle between the last step $\vec{\ell}_N$ and the sum $\vec{s}_{N-1}$ of the ones before is totally random, so $\langle \vec{s}_{N-1} \cdot \vec{\ell}_N \rangle = 0$. Also $\langle \vec{\ell}_N^2 \rangle = \ell^2$. So we find

$$\langle \vec{s}_N^2 \rangle = \langle \vec{s}_{N-1}^2 \rangle + \ell^2 \tag{18}$$

Since this is true for any N we can compute that

$$\langle \vec{s}_N^2 \rangle = \langle \vec{s}_{N-1}^2 \rangle + \ell^2 = \langle \vec{s}_{N-2}^2 \rangle + 2\ell^2 = \cdots = N\ell^2 \tag{19}$$

and the RMS distance moved is $\sqrt{N}\ell$, just like in the 1D case.

3 Diffusion from random walks

Diffusion refers to the net spreading of the distribution of molecules due to random molecular motion. Think about an individual molecule in a gas, say some CO molecule coming out of a car's exhaust. It leaves the exhaust and moves in a straight line until it hits another molecule, in which case it is buffeted essentially randomly in a different direction. As all the CO molecules are doing the same thing, on average, the net effect is a diffusion of the CO gas. We want to compute the probability distribution $P_t(x)$ for where a CO molecule is after a time t and then use this to determine the equation of motion of the density of the gas.

3.1 Collisions in a gas

It's helpful to discuss random walks for gases in terms of a set of convenient physical quantities. An important one is

- τ = the **collision time** is the average time a molecule goes before colliding with another molecule

The number of collisions in a time t is then

$$N = \frac{t}{\tau} \tag{20}$$

A related quantity is

- ℓ = the **mean free path** is the average distance a molecule goes between collisions

The mean free path is related to the collision time by

$$\ell = \bar{v}\,\tau \tag{21}$$

where

- $\bar{v}$ = the **average molecular speed**, $\bar{v} = \langle |\vec{v}| \rangle$.

Sometimes a more useful quantity is the root-mean-square velocity $v_{\text{rms}} = \sqrt{\langle \vec{v}^2 \rangle}$. We can also use the speed of sound c_s in a gas, which is of course limited by the speed by which the molecules move. All three of these, $\bar{v}, v_{\text{rms}}$ and c_s are related by coefficients of order one, as we will see once we understand gases in more detail in future lectures. For example, in air at room temperature, $\bar{v} = 467\frac{m}{s}$, $v_{\text{rms}} = 507\frac{m}{s}$ and $c_s = 346\frac{m}{s}$.

The mean free path is related to the density and size of the molecules. Treating molecules as spheres of radius R, two molecules will hit if their centers are within $2R$ of each other. Thus you can think of a moving molecule as having an effective cross sectional area of $\sigma = \pi(2R)^2$. This effective cross sectional area is also called the **collisional cross section**. After N collisions a molecule will have swept out a volume $V = N\ell\sigma$. The number of molecules it hits during this sweeping is $N = Vn$ with

- n = the **number density** = number of molecules per unit volume

We will use number density a lot in statistical mechanics. It is interchangeable with the

- ρ = the **mass density**.

as $\rho = mn$ where m is the mass of a molecule (or the average mass of a molecule if the gas is mixed).

Thus,

$$\ell = \frac{1}{n\sigma} \tag{22}$$

Bigger molecules have bigger cross sectional areas so they will have smaller mean free paths. Since liquids are more dense than gases, generally they will have smaller mean free paths.

For example, the radius of a typical atom is around the Bohr radius $1\ a_0 = 0.05\,\text{nm}$. So an air molecule, such as N_2 or O_2, has a radius of around $R \approx 2a_0 \sim 0.1\text{nm}$. Thus $\sigma \approx \pi(2R)^2 = 0.14\text{nm}^2$ in air. Air has a density of $\rho = 1.3\frac{\text{kg}}{m^3}$ and an average mass of $m = 4.81 \times 10^{-26}\frac{\text{kg}}{\text{molecule}}$, so its number density is $n = \frac{\rho}{m} = 2.6 \times 10^{25}\frac{1}{m^3}$. Note that $n^{-1/3} = 3.3\text{nm}$ so air molecules are around 3nm apart on average. The mean free path is $\ell = \frac{1}{n\sigma} = 0.26\mu m$. The collision time is then $\tau = \frac{\ell}{\bar{v}} = 0.57\text{ns}$.

These are useful numbers to have in your head: in air at room temperature, molecules have velocities around $v \sim 500\,\frac{m}{s} \approx 1100\,\frac{\text{miles}}{\text{hour}}$, are around $R \sim 0.1\,\text{nm}$ big and $n^{-1/3} \sim 1\,\text{nm}$ apart. They collide around once every nanosecond (one billion times per second) after having moved around $\ell \sim 100\,\text{nm}$ (one thousand molecule lengths).

3.2 Diffusion from random walks

Let's now consider the probability distribution $P_t(x)$ for where a CO molecule is after a time t. We'll start in one dimension. Treating molecular interactions as a random walk, we take $a = b = \frac{1}{2}$ since the molecule should be equally likely to be knocked left as right. Such a random walk is unbiased. For an unbiased random walk, the mean displacement is $\bar{x} = 0$ and therefore does not tell us much about how fast the molecules are diffusing. Instead, the RMS displacement is more useful. The RMS displacement after a time t is, from Eq. (12)

$$x_{\text{rms}} = \sqrt{N}\ell = \sqrt{\frac{t}{\tau}}\ell = \sqrt{\ell\bar{v}\,t} \tag{23}$$

This $x \sim \sqrt{t}$ behavior is the key characteristic of a random walk. Note that this is going to be a much smaller distance than an unhindered molecule would move on average, $\Delta x \sim vt$.

Knowing the mean ($\bar{x} = 0$) and the standard deviation ($\sigma = \sqrt{\ell\bar{v}t}$) we can immediately write down the full probability distribution for large times ($t \gg \tau$) using the central limit theorem:

$$P_t(x) = \sqrt{\frac{1}{2\pi\ell\bar{v}t}}\exp\left[-\frac{x^2}{2t\ell\bar{v}}\right] \tag{24}$$

Note that probability distribution satisfies the differential equation

$$\frac{\partial P_t(x)}{\partial t} = D\frac{\partial^2 P_t(x)}{\partial x^2} \tag{25}$$

where $D = \frac{1}{2}\ell\bar{v}$. This is the **1D diffusion equation**. You can easily check by plugging Eq. (24) into Eq. (25).

We defined $P_t(x)$ as a probability distribution for one particular CO molecule in a gas. But the same probability distribution holds for any molecule. Since there are usually an enormous number $N \sim 10^{24}$ of gas molecules, if each one has a probability $P_t(x)$ of being at the the point x then number density will be simply

$$n(x,t) = NP_t(x) \tag{26}$$

To be precise, the number density is not exactly the same as the probability distribution since, classically, a particle is either at a particular position or not. So we should think of $n(x,t) = NP_t(x)$ as referring to the number density averaged over time. (We'll return to this averaging in the next lecture, in the context of ergodicity.)

Thus we find

$$\frac{\partial n(x,t)}{\partial t} = D\frac{\partial^2 n(x,t)}{\partial x^2} \tag{27}$$

In 2 or 3 dimensions, the resulting equation is the rotationally symmetric version of this:

$$\boxed{\frac{\partial n(\vec{x},t)}{\partial t} = D\vec{\nabla}^2 n(\vec{x},t)} \tag{28}$$

This is known as the **diffusion equation**. It describes how substances move due to random motion. The coefficient is

$$D = \frac{1}{2}\ell\bar{v} = \frac{1}{2}\frac{\ell^2}{\tau} \tag{29}$$

This coefficient D is called the **diffusion constant** and the relation $D = \frac{1}{2}\frac{\ell^2}{\tau}$ is known as the **Einstein-Smoluchowski equation**.

For example, in air $\ell \approx 0.26\mu m$ and $\tau \approx 0.57\text{ns}$ so $D = 5.9 \times 10^{-5} m^2 / s$. This means that it takes 1 day for an individual gas molecule to diffuse 1 meter. Then it takes 100 days for it to diffuse 10 meters. These numbers are characteristic of diffusion processes: diffusion in air over macroscopic distances is generally very slow, and almost always dominated by convection (see Section 4.1) and other forms of energy transport. On the other hand, on smaller length scales where convention is irrelevant like a cell, diffusion can be dominant. For example, the diffusion constant for proteins in water is around $D \approx 10^{-11}\frac{m^2}{s}$. This is tiny compared to D for gases, but the typical distances proteins diffuse are also tiny, and time scales like distance squared. To diffuse across a cell of size $10^{-4}m$ it takes 15 minutes. To diffuse across a cell nucleus of size $6\,\mu m$ it takes only 3.6 seconds.

Just because the diffusion equation looks simple does not mean it has trivial consequences! For example, it is mathematically identical to the Schrödinger equation, which accounts for a great variety of interesting physics. It also also mathematically identical to the heat equation: heat conduction is a diffusive process.

The diffusion equation is linear, so that if $n_1(\vec{x}, t)$ and $n_2(\vec{x}, t)$ are solutions, then so is their sum. In particular if we start with a bunch of particles at some positions x_i, then they will diffuse independently of each other. This gives us a way to solve the diffusion equation in general. For one particle, starting at $\vec{x} = 0$, the solution is given by Eq. (24). Note that at $t = 0$, this solution really does represent a localized source. In fact, the limit as $t \to 0$ of this solution is one of the possible definitions of a δ-function:

$$\lim_{t \to 0} \left(\frac{1}{4\pi D t} \right)^{3/2} \exp\left[-\frac{\vec{x}^2}{4 D t} \right] = \delta^3(\vec{x}) \tag{30}$$

where $\delta^3(\vec{x}) = \delta(x)\delta(y)\delta(z)$. Thus Eq. (24) is a solution to the diffusion equation with boundary condition $n(\vec{x}, t) = \delta^3(\vec{x})$ at time $t = 0$: it describes the diffusion away from a point source. Since any function can be described as a set of points, we can construct any solution to the diffusion equation by combining the solutions as in Eq. (24). More precisely, if $n_0(\vec{x}, 0)$ is the number density at time 0, then the solution for all times is

$$n(\vec{x}, t) = \int d^3 y \left(\frac{1}{2\pi \ell \bar{v} t} \right)^{3/2} \exp\left[-\frac{(\vec{x} - \vec{y})^2}{2 t \ell \bar{v}} \right] n_0(\vec{y}, 0) \tag{31}$$

To check this, we note that the right-hand side satisfies the diffusion equation and Eq. (30) verifies the boundary condition at $t = 0$. Solving differential equations in this way is known as the **Green's function method**.[2] It converts solving a difficult differential equation to doing an integral.

Eq. (31) has a simple physical interpretation: the number of molecules at a point $\vec{x}$ are those that have walked there randomly from any other point $\vec{y}$ over the time t.

4 Fick's laws of diffusion

The approach to diffusion we discussed was based on a microscopic picture of random walks of individual molecules. We can also approach diffusion from the continuum perspective. Let us continue to denote the number density by $n(\vec{x}, t)$ and let us also denote the local velocity as the vector field $\vec{v}(\vec{x}, t)$. It can be helpful to think of $n(\vec{x}, t)$ as the density of a fluid, like water in a stream, and $\vec{v}(x, t)$ as its velocity at the point $\vec{x}$ at time t. For simplicity, lets assume that n and $\vec{v}$ are constant in the y and z directions, so they only depend on x, $n = n(x, t)$, $\vec{v} = (v_x(x, t), 0, 0)$. Now, the total number of molecules between x_1 and x_2 can only change if particles flow in or out of that region. So

$$\frac{\partial}{\partial t} \int_{x_1}^{x_2} dx\, n(x, t) = \underbrace{n(x_1, t) v_x(x_1, t)}_{\text{number coming in past } x_1 \text{ per unit time}} - \underbrace{n(x_2, t) v_x(x_2, t)}_{\text{number going out past } x_2 \text{ per unit time}} \tag{32}$$

$$= -\int_{x_1}^{x_2} dx\, \partial_x [n(x, t) v_x(x, t)] \tag{33}$$

[2] In general, a Green's function satisfies $\mathcal{O}G(x, t) = \delta(x)\delta(t)$ for some differential operator $\mathcal{O}$. In our case, the Green's function is $G(\vec{x}, t) = P(\vec{x}, t)\theta(t) = \sqrt{\frac{1}{4\pi D t}} \exp\left[-\frac{\vec{x}^2}{4 D t} \right] \theta(t)$ which satisfies $[\partial_t - D\vec{\nabla}^2] G(\vec{x}, t) = \delta^3(\vec{x})\delta(t)$.

Pulling the $\frac{d}{dt}$ on the left into the integral, and using that x_1 and x_2 are arbitrary, we get

$$\frac{\partial}{\partial t}n(x,t) = -\frac{\partial}{\partial x}J_x(x,t) \tag{34}$$

where

$$J_x(x,t) = n(x,t)v_x(x,t) \tag{35}$$

The 3D version of this equation is called the **continuity equation**

$$\frac{\partial n}{\partial t} + \vec{\nabla}\cdot\vec{J} = 0 \tag{36}$$

and $\vec{J}(\vec{x},t)$ is called the **flux**. The flux is the number density times velocity $\vec{J}(x,t) = n(x,t)\times\vec{v}(x, t)$. It gives the number of particles passing by a given point per unit area per unit time. Note that the velocity field and the density are in-principle independent, like position and velocity are independent in classical mechanics.

How does the flux related to the density during diffusion? Well, if the density is constant in position, then the net diffusion should be zero and the flux should vanish. Conversely, if the density has some spatial gradient, then there should be net flux from high density to low density. Thus, the leading order thing we could imagine is that $\vec{J}$ is proportional to the concentration gradient:

$$\vec{J} = -D\vec{\nabla}n \tag{37}$$

with D a proportionality constant. We put in the minus sign so that D would be a positive number (if $\partial_x n > 0$, so the gradient increases to the right, then particles flow to the *left*). This is known as **Fick's first law**. It's a law because we didn't derive it (neither did Fick), it just seems reasonable. Note that Fick's first law is not some general property of fluxes – you can certainly have a nonzero flux at constant density, like current in a river, if there's some potential driving the flow. Ficks first law it is a statement about fluxes in *diffusive systems*, where there is no other source of motion other than random motion.

Once we have Eq. (37) we can plug into the continuity equation, Eq. (36) to get

$$\frac{\partial n(\vec{x},t)}{\partial t} = D\vec{\nabla}^2 n(\vec{x},t) \tag{38}$$

This is also known as **Fick's second law**. It is none other than the diffusion equation. Since Fick's second law follows from Fick's first law, in fact, we have justified Fick's first law through our analysis of random walks. Moreover, through our analysis of random walks, we have related D to properties of the gas, $D = \frac{1}{2}\ell\bar{v}$ as in Eq. (29).

For example, say we have some lemmings that come out of a hole, walk left or right randomly, and maybe fall off a cliff a distance a away. Let their density be $n(z)$, with $z = 0$ the hole and $z = a$ the cliff. At a given z, over a short time δt, the number of lemmings that leave to the right is proportional to $n(z)$ and those that come in from the right is propotional to $n(z+\delta z)$. Similarly, the number that leave to the left is $n(z)$ and those coming in from the left proportional to $n(z-\delta z)$. So the net change in the number at z is going to be proportional to $n(t+\delta t) - n(t) \propto [n(z+\delta z) - n(z)] + [-n(z) + n(z-\delta z)]$. That is, $\frac{\partial n}{\partial t} \propto \frac{\partial^2 n}{\partial z^2}$. This is a physical way to see why diffusion only occurs if the second derivative is nonzero. In a steady state situation, $\frac{\partial n}{\partial t} = 0$, the solution to the diffusion equation is $n(z) = n_0 + \frac{z}{a}(n_1 - n_0)$. Then $J_z = -\frac{D}{a}(n_1 - n_0)$. This is a constant flux of lemmings coming out of the hole and falling of the cliff. Note that since J_z is constant, $\partial_z J_z = 0$ and the continuity equation is satisfied.

To get a feel for how fast diffusion is, the diffusion constant in water for nitrogen molecules is $D = 2\times 10^{-9}\frac{m^2}{s}$. Recalling from Eq. (22) that $\ell = \frac{1}{n\sigma}$, so $D = \frac{1}{2}\frac{\bar{v}}{n\sigma}$, bigger molecules should have smaller diffusion rates. Indeed, benzene molecules C_6H_6 in water have $D = 1\times 10^{-9}\,\frac{m^2}{s}$. For large molecules like proteins in water the diffusion constant is even smaller $D \approx 10^{-11}\frac{m^2}{s}$. In gases, densities n are smaller so ℓ is larger and the diffusion constants are generally larger. For example, CO molecules in air at room temperature and pressure have $D = 2\times 10^{-5}\frac{m^2}{s}$.

To use the diffusion contant, we can either plug in the exponential solution, Eq. (24), or more simply use Eq. (23):

$$x_{\text{rms}} = \sqrt{2Dt} \tag{39}$$

For example, taking a dye molecule in water with $D \approx 10^{-9}\frac{m^2}{s}$, to move $\Delta x = 1m$ would take $\frac{(\Delta x)^2}{2D} = 31$ years. So clearly diffusion is not the main mechanism by which dyes move around in water.

By the way, thermal conduction is very much like diffusion. Instead of the diffusion equation, temperature satisfies the **heat equation**:

$$\frac{\partial T(x, y, z, t)}{\partial t} = \alpha \vec{\nabla}^2 T(x, y, z, t) \tag{40}$$

where α is called the coefficient of thermal diffusivity. This equation describes diffusion of temperature, rather than diffusion of particle number. The derivation of the heat equation is identical to the derivation of Fick's second law, with conservation of energy replacing conservation of particle number. The analog of Fick's first law for thermal conduction is called **Fourier's law**. Fourier's law is an empirical observation that the rate of heat flow is proportional to the temperature difference. We'll return to thermal conduction when we talk about temperature and heat in future lectures.

4.1 Convection (optional)

Diffusion refers to the motion of a molecule through random collisions. Think of a liquid in equilibrium and just try to follow one molecule. **Convection** occurs when the system is not in equilibrium to begin with. In such situations, there can be coherent convective currents, like a hot or cold wind, that move the dye much faster than through a random walk. Or if you dropped the dye into the water with a dropper it hits with some force and has some inertia; then it takes a while for the system to equilibrate and the dye molecules are for a while moving much faster than they would if there were only diffusion.

If there is some external effect causing the medium to flow with velocity $\vec{v}_{\text{conv}}(\vec{x}, t)$, then there will be flux even if there is no concentration gradient. We can introdude the **convective flux**

$$\vec{J}_{\text{conv}}(\vec{x}, t) = \vec{v}_{\text{conv}}\, n(\vec{x}, t) \tag{41}$$

to describe this situation. Adding this convective flux to the diffusive flux, we get a new term:

$$\frac{\partial n(\vec{x}, t)}{\partial t} = D\vec{\nabla}^2 n(\vec{x}, t) + \vec{v}_{\text{conv}} \cdot \vec{\nabla} n(\vec{x}, t) \tag{42}$$

This is called the **generalized diffusion equation** and describes situations where diffusion and convection are both important. Unfortunately, it is usually impossible to determine $\vec{J}_{\text{conv}}(\vec{x},\, t)$, since when there is convection usually molecules are all moving around in different directions and it is a horribly non-linear process. Think about this next time you pour milk into your coffee – all those little eddy currents and funny shapes are convective. Good luck describing them analytically! Convection is almost always studied with numerical simulations.

So for diffusion to actually be visible, a system has to be very calm – no chemical, temperature, or density gradients. A place where diffusion is more important than convection is in biology. In biological systems, temperature is often very constant, convection is small, and molecules do not have to move very far. Diffusion of heat (thermal conduction) is the dominant mechanism of heat transfer in solids, for example as you heat up a pan on the stove. However, when you heat a room, convection dominates and the heat equation, Eq. (40) is not relevant.

5 Brownian motion

An important application of the diffusion equation is to study Brownian motion. In 1827, a botanist named Robert Brown collected some pollen one Spring afternoon and put it in some water in a Petri dish in his lab, then went to bed. When he woke up, he found that the pollen grain had moved a significant distance. "It's alive!" he concluded. In fact, the pollen moved not because it is alive, but rather because it underwent a random walk due to the water molecules surrounding it constantly giving it little kicks. This movement is called **Brownian motion**, after Robert Brown.

Brownian motion refers to the random walk of a large particle due to stochastic collisions with smaller particles. Although each hit from a small particle does very little, the hits add up to a macroscopically observable displacement. You can see Brownian motion easily with a microscope, where a dust particle or a bacterium will move a finite distance in a reasonable time. What is fascinating about this migration is that you cannot resolve the small molecules, like water molecules, in the microscope, so it looks like the big particle is moving by magic. Of course, it is not magic, and indeed we can deduce the existence of "invisible" molecules from Brownian motion of something visible. Einstein used this insight to measure Avogadro's number, as we will now see.

The molecular collisions have another effect too – they slow down a moving particle, through a drag force. Indeed, drag, that you experience running your hand through the air or in water is a collective effect of many small molecules impeding the motion. Drag forces are macroscopic and can be measured without ever talking about molecules. A drag force, by definition, slows down a particle, so if $\vec{v}=0$ it should vanish. Thus, the leading effect in an expansion around $\vec{v}=0$ of drag is that it is linear in the velocity is that the equations of motion $F=ma$ get modified to

$$m\frac{d^2\vec{x}}{dt^2}+\mu\frac{d\vec{x}}{dt}=\vec{F}_{\text{ext}} \tag{43}$$

where m is the mass of a particle $\vec{F}_{\text{ext}}$ is some external force. μ is the **drag coefficient** (also called **mobility**). To measure μ we could, for example, could rub the particle to make it electrically charged, then pull it with the electric force and measure the resistance. Or we could tie a tether to it, add a weight, and pull it with gravity.

Once μ is measured and $\vec{F}_{\text{ext}}$ is turned off we can look at the distance the dust particle moves due to Brownian motion alone. Of course, the expected value is $\langle\vec{x}\rangle=0$, by symmetry, so we want to look at the RMS displacement $x_{\text{rms}}=\sqrt{\langle\vec{x}^2\rangle}$. Why should the particle move at all? With $\vec{F}_{\text{ext}}=0$ there is a solution $\vec{x}(t)=0$ to the equations of mition. The key piece of additional information is that the mass m is not going to stay at rest since it gets buffeted by the molecules. A key result that we derive soon is that everything in equilibrium has the same average kinetic energy determined by temperature. More preciely $\langle\frac{1}{2}m\vec{v}^2\rangle=\frac{3RT}{2N_A}$ for anything in equilibrium: the mass, the gas, anything. It is the random buffeting of the particles into each other due to random walks than causes equilibrium. The bottom line is that $\vec{x}=0$ cannot be correct since then $\langle\vec{v}^2\rangle=0$ which is wrong. What Einstein did was found a way to express $\langle\vec{x}^2\rangle$ in terms of $\langle\vec{v}^2\rangle$.

To begin, we note

$$\frac{d}{dt}(\vec{x}\cdot\vec{v})=\left(\frac{d}{dt}\vec{x}\right)\cdot\vec{v}+\vec{x}\cdot\left(\frac{d}{dt}\vec{v}\right)=\vec{v}\cdot\vec{v}+\vec{x}\cdot\left(\frac{d}{dt}\vec{v}\right) \tag{44}$$

and that Eq. (43) with $\vec{F}_{\text{ext}}$ turned off imples implies

$$\frac{d}{dt}\vec{v}=-\frac{\mu}{m}\vec{v} \tag{45}$$

So then we have

$$\frac{d}{dt}\langle\vec{x}\cdot\vec{v}\rangle=\langle\vec{v}^2\rangle-\frac{\mu}{m}\langle\vec{x}\cdot\vec{v}\rangle \tag{46}$$

The general solution of this equation starting at $\vec{x}=0$ is

$$\langle\vec{x}\cdot\vec{v}\rangle=\frac{m}{\mu}\langle\vec{v}^2\rangle\left(1-e^{-\frac{\mu t}{m}}\right) \tag{47}$$

For late times $t\gg\frac{m}{\mu}$, the exponential is very small and we see that $\langle\vec{x}\cdot\vec{v}\rangle=\frac{m}{\mu}\langle\vec{v}^2\rangle$ is constant in time. Next we use that

$$\frac{d}{dt}\langle\vec{x}^2\rangle=2\,\langle\vec{x}\cdot\frac{d\vec{x}}{dt}\rangle=2\,\langle\vec{x}\cdot\vec{v}\rangle=\frac{m}{\mu}\langle\vec{v}^2\rangle \tag{48}$$

This says that the displacement-squared grows linearly with time, characteristic of a random walk. that is

$$\langle\vec{x}^2\rangle=\frac{m}{\mu}\langle\vec{v}^2\rangle\,t \tag{49}$$

The RMS displacement is therefore

$$x_{\rm rms} = \sqrt{\langle \vec{x}^2 \rangle} = v_{\rm rms}\sqrt{\frac{2m}{\mu}}\sqrt{t} \tag{50}$$

Comparing to Eq. (39) we have

$$\boxed{D = \frac{1}{\mu}\langle m\vec{v}^2 \rangle} \tag{51}$$

This is known as the Einstein relation.

As mentioned, to proceed, we need to borrow a result here from Lecture 4, that the average kinetic energy per particle is $\langle \frac{1}{2}m\,\vec{v}^2 \rangle = \frac{3R}{2N_A}T$ with R the ideal gas constant and T the temperature. (It should not be obvious to you that the average kinetic energy is determined by the temperature, but it's true as we'll show in Lecture 4.) Then Eq. (51) implies

$$N_A = \frac{3RT}{\mu D} \tag{52}$$

So by measuring the temperature (with a thermometer), the drag coefficient (with an external force) and the rate of diffusion (from Brownian motion), the number of water molecules N_A can be determined. This is how Albert Einstein proposed to measure Avogadro's number N_A using Brownian motion in 1905.

5.1 Viscosity

Consider a ball falling through a fluid. It is accelerated from a downward force due to gravity and encounters resistance, or drag, from the fluid. There are two sources of drag. The first, called **inertial drag**, is due to the ball bonking molecules in front of it and speeding them up. In a time Δt the ball will sweep out a volume $\Delta V = \pi R^2 |\vec{v}| \Delta t$, thereby accelerating a mass $\Delta m = \rho \Delta V$ of fluid to velocity roughly $\vec{v}$. The average acceleration is then $\vec{a} = \frac{\vec{v}}{\Delta t}$, so that the force is $\vec{F}_{\rm inertial} = \Delta m\,\vec{a} = \pi R^2 \rho |\vec{v}| \vec{v}$. This goes like the *square* of the velocity. This inertial drag is relevant at large velocities, but for slow objects it is always going to be subdominant to a drag force that is linear in v. For slow objects (slow compared to typical molecular speeds), the dominant drag force is due to viscosity, called **viscous drag**. It is viscous drag that is important for Brownian motion, and related to random walks.

Viscosity is another physical effect whose microscopic origin is in the stochastic collision of molecules. The more precise name for viscosity is dynamic shear viscosity. It measures how a fluid responds to shear forces: you push the top layer and ask how much the bottom layer moves. Intuitively, viscosity is a measure of how well a fluid flows.

A shear force is applied to an area. Think of floating a block of wood on water and applying a external force $\vec{F}_{\rm ext}$ to move it parallel to the surface. The bigger the area of the block, the more force it puts on the water. The water then responds by picking up some velocity $\vec{v}$, in the same direction as the force. The deeper you go into the water, the slower it will go. So we might expect

$$\eta \frac{\vec{v}}{z} \overset{?}{=} \frac{\vec{F}_{\rm ext}}{A} \tag{53}$$

with z the depth. This isn't quite right, because we don't know that the z dependence is exactly $\frac{1}{z}$.

The right way to think about this shear force is that we apply it only to the top of the water. When the top of the water moves, it pulls along the layer below that (by layer we mean layer of molecules, or just some abstract infinitesimal thickness of the fluid), and so on. So we write

$$\eta \frac{\partial \vec{v}}{\partial z} = \frac{\vec{F}_{\rm ext}}{A} \tag{54}$$

The parameter η is this equation is called the dynamic shear **viscosity**.

Now lets return to the sphere falling through the fluid. It has a downward force due to gravity. As it moves down, it displaces the molecules, but also imparts velocity to the fluid in the direction transverse to its motion

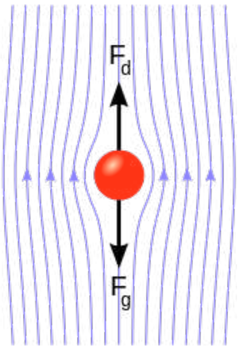


Figure 1. A ball falling through a viscous fluid has a downward force due to gravity and a drag force due to viscosity.

The definition of viscosity tells us that the sphere induces a velocity of the fluid

$$\eta\frac{\partial\vec{v}}{\partial z} = \frac{\vec{F}_g}{4\pi R^2} \tag{55}$$

where $4\pi R^2$ is the surface area of the sphere. At the surface of the sphere, the fluid velocity is the same as the sphere velocity. Thus we can solve this equation to see how the fluid velocity changes with distance. The faster the sphere is falling, the faster the fluid will go. Eventually, all the energy used by gravity to accelerate the sphere will be taken up by the work done to move the water, and the sphere will stop accelerating. Thus, there will be some effective drag force. Working out all the factors (an annoyingly tedious calculation), the result is

$$\vec{F}_{\text{drag}} = 6\pi\eta R\,\vec{v} \tag{56}$$

This is known as the **Stokes drag force**. This equation relates a property of the fluid (the viscosity), to the resistance experienced by an external object as it is being forced through the fluid (the viscous drag force). It is linear in the velocity of the object, compared to the inertial force which was quadratic in velocity, so at small velocities, the viscous drag force dominates. Keep in mind that viscosity is a property of the fluid itself, while drag depends on what is being dragged through the fluid. The Stokes force applies when that thing is a sphere of radius R. For a different object, like a cube, the precise form of the force would be different, but the scaling with η and $\vec{v}$ and some measure of size R is universal. Indeed, it is fixed by dimensional analysis.

Recall that we defined mobility as the drag coefficient in Eq. (43). So

$$\mu = 6\pi\eta R \tag{57}$$

This is known as the **Stokes relation**. Plugging it into the Einstein relation in Eq. (51) gives

$$D = \frac{1}{6\pi\eta R}\langle m\vec{v}^2\rangle \tag{58}$$

This is known as the **Einstein-Stokes relation.** It relates the diffusion constant D and the viscosity, telling us that viscosity also has its origin in the microscopic random walks of the molecules in the fluid.

6 Summary

In this lecture we have studied diffusion. The main concepts to understand are

- **Random walks**: A particle/system has a fixed probability of moving in each direction. We are interested in the net motion after N steps.
- If we move left with probability a and right with probability $b = 1 - a$, the chance of taking m steps right is given by the **binomial distribution**: $B_N(m) = a^m b^{N-m}\binom{N}{m}$ where $\binom{N}{m} = \frac{N!}{m!(N-m)!}$.

- The average distance moved in a random walk after N steps scales as $\sigma \sim \sqrt{N}\ell$ where ℓ is the step size. This is true in any number of dimensions.
- A bunch of molecules random-walking is called **diffusion**. Then the number of steps N is proportional to the time t so the distance moved is $\Delta x \sim \sqrt{t}\ell$. This scaling of **distance $\propto$ square-root of time** is the characteristic feature of diffusion.
- In the continuum limit, molecules are described by a **number-density** $n(\vec{x}, t)$.
- The number density satisfies the **diffusion equation** $\frac{\partial n(\vec{x}, t)}{\partial t} = D\vec{\nabla}^2 n(\vec{x}, t)$ when diffusion is dominant.
- The exact solution to the diffusion equation with boundary conditions $n(\vec{x}, t) = \delta^3(\vec{x})$ is $n(\vec{x}, t) = \sqrt{\frac{1}{4\pi D t}} \exp\left[-\frac{\vec{x}^2}{4Dt}\right]$. This confirms the scaling $\Delta x \sim \sqrt{t}$.
- The diffusion equation comes from a modeling of the microscopic system as undergoing random walks. An alternative classical-field approach (no particles) uses **Fick's laws**. These assume the conservation of the amount of stuff (the continuity equation $\frac{\partial n}{\partial t} + \vec{\nabla} \cdot \vec{J} = 0$, where $\vec{J} = n \times \vec{v}$ is the **flux**) and that flux is proportional to the concentration gradient $\vec{J} = -D\vec{\nabla} n$, leads to the same diffusion equation.
- The conduction of heat is described by the **heat equation**: $\frac{\partial T(\vec{x}, t)}{\partial t} = \alpha \vec{\nabla}^2 T(\vec{x}, t)$. This equation has identical form as the diffusion equation, and so also has $\Delta x \sim \sqrt{t}$.
- In liquids and gases, diffusion is rarely dominant unless the system is very calm. **Convection**, coming from non-equilibrium initial conditions, is usually is more important. In solids, diffusion and the heat equation work well.
- A large molecule in a reasonably calm liquid or gas can often be modeled well by a random walk/diffusion. When this applies, we say it undergoes **Brownian motion**. For such molecules, the diffusion consent is given by the **Einstein relation**: $D = \frac{1}{\mu}\langle m\vec{v}^2\rangle$ with μ the drag coefficient. Measuring D and μ is one way to measure Avogadro's number: $N_A = \frac{3RT}{\mu D}$.
- The drag coefficient is also related to **viscosity** η by $\mu = 6\pi\eta R$. Then $D = \frac{1}{6\pi\eta R}\langle m\vec{v}^2\rangle$. This is called the **Einstein-Stokes relation**,

The main lesson from the last section is simply that the drag force, mobility, diffusion, viscosity and random walks are all related. I don't expect you to remember all these formulas, and I certainly don't want you to memorize them. Just try to have the basic ideas straight. Viscosity is a macroscopically measurable property of a material. When *a large particle* moves in a viscous material, it undergoes Brownian motion. The bigger the particle, the smaller the diffusion constant, and the slower it moves. The higher the viscosity, the larger the drag force, and the slower the particle moves.

Small molecules moving in a fluid also undergo random walks. For small molecules, of around the same size as the molecules in the fluid, the diffusion constant is $D = \frac{1}{2}\ell\bar{v}$ with ℓ the mean free path. In this case, it is not useful to think in terms of drag forces and viscosity since the diffusing particle has essentially no inertia.

A tricky point from this lecture is that **diffusion is an equilibrium phenomena**. Equilibrium does not mean there is no time-dependence at all, but that the macroscopic properties of the system are static. We'll discuss equilibrium more next lecture. You should think of $\frac{\partial n(\vec{x}, t)}{\partial t} = D\vec{\nabla}^2 n(\vec{x}, t)$ as describing the change in the number density of a small subset of the molecules in a bath, for example, dye in water, or of a macromolecule undergoing Brownian motion in a bath. The water in the bath doesn't change, even though individual water molecules might move. When the whole system is out of equilibrium, for example when the boundary conditions are open (gas released into a vacuum) or if you violently drop the dye in, then the diffusion equation doesn't apply.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 3: Equilibrium

1 Introduction

Now we have a little bit of sense of how things simplify when large numbers of events are sampled from a probability distribution. The next thing to do is apply this simplification to general physical systems. The key observation, which allows statistical mechanics to be useful at all, is that systems equilibrate. This means that at some point the macroscopic properties of the system stop changing. Note that equilibrium does not mean the system is static: a gas at fixed temperature still has moving molecules, but its macroscopic properties, pressure, temperature, etc. are not changing. From a microscopic perspective, the probability distribution of the states does not change. In this lecture, we will show that probabilities eventually stop changing.

It is not hard to build intuition for equilibration of probabilities. For example, take a deck of cards and pick two consecutive cards from somewhere in the middle. What is the probability that the cards have the same suit after the deck has been shuffled t times? When you first open a new deck, it is all in order, so the chance that the two cards have the same suit pretty high, $P(0)\approx 1$. Then you shuffle it. After shuffling once, the probability is of two suited cards in a row is lower than before, $P(t=1)\lesssim 1$ but probably still pretty high since one shuffle doesn't mix them much. Eventually, after shuffling a bunch of times, the probability of finding two consecutive suited cards is going to stop changing ($\lim_{t\to\infty}P(t)=\frac{12}{52}$). Note that in equilibrium the cards change with each shuffle, but the probabilities don't: after each shuffle the identity of the top card is not constant, but the probability that the top card is the ace of spades *is* constant, $P=\frac{1}{52}$.

One of the most important properties of equilibrium is that in equilibrium, all possible states are equally likely. This is known as the *postulate of equal a priori probabilities*. For example, with a shuffled deck, the chance of the top card being any of the 52 cards is the same $P=\frac{1}{52}$.

In this lecture we introduce a number of important concepts related to equilibration

- Chaos: the state of a system is uncontrollably sensitive to small perturbations.
- Molecular chaos: correlations among states are lost over time.
- Coarse-graining: averaging over nearby position/momenta.
- Ergodicity: the time average of a system is the same as the average over possible states.

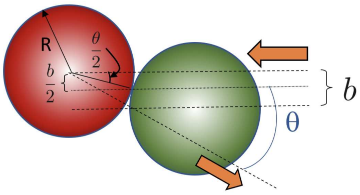
The example we focus on most in this lecture, and for much of the course, is that of an **ideal gas**. An ideal gas is one for which all the interactions are short-ranged and collisions are elastic. In the simplest ideal gas, the gas molecules have no internal structure. The state of the ideal gas is specified classically by giving the positions $\vec{q}_i(t)$ and velocities $\vec{v}_i(t)$ (or momenta $\vec{p}_i(t)=m\vec{v}_i(t)$) of the particles $i=1...N$ at any time t. The set of allowed $\vec{q}_i$ and $\vec{p}_i$ for the N particles is called the **phase space** of the gas. The full phase space is $6N$ dimensional. The coordinates of a point in 6N dimensions tells you the state of the system at any fixed time. The equations of motion of the then determine the trajectory through phase space. Thinking of time-evolution as a trajectory in phase space provides a useful language for discussing (and proving) results about classical systems.

2 Chaos

A key to understanding equilibration (and statistical mechanics!) is to appreciate why we go from knowing exactly what is going on in a system to accepting that we can only discuss the probability of what is going on. This is the transition from mechanics (classical, where you know positions and velocities or quantum where you know the wavefunction) to statistical mechanics.

The easiest way to see why we *must* go to a probabilistic treatment is that, it is literally *impossible* to keep track of every particle in the system. In fact, systems with large numbers of degrees of freedom are always **chaotic**: they are uncontrollably sensitive to infinitesimal inaccuracies of the specification of the system. Chaos is sometimes called the "butterfly effect", since a butterfly flapping its wings in Australia can affect the weather in Boston. We can understand the basic observation about chaos with a simple example.

Let's treat the molecules in a gas as hard sphere particles of radius R. If you follow one sphere, it will bounce off another sphere after travelling, on average, a distance ℓ (the mean free path). The angle θ it deflects will depend on the impact parameter b, defined as the distance between the sphere's centers perpendicular to the direction of motion



(1)

By working out the geometry as in the figure, we see that b, R and θ are related by:

$$\frac{b}{2} = R\sin\frac{\theta}{2} \tag{2}$$

You can see this relation from a little triangle in the red sphere.

Let us follow one ball. It first collides with an impact parameter b_1 deflecting at an angle θ_1, then hits another ball, with impact parameter b_2 at angle θ_2 and so on. We are interested in how the trajectory of this molecule changes upon a really weak force. If the force changes the impact parameter of the first collision by some small Δb_1, then the after the first collision, the scattering angle will change by $\Delta\theta_1$ where, by Taylor expanding around θ_1 we get

$$\frac{b + \Delta b_1}{2} = R\sin\left(\frac{\theta_1 + \Delta\theta_1}{2}\right) \approx R\sin\left(\frac{\theta_1}{2}\right) + R\frac{\Delta\theta_1}{2}\cos\frac{\theta_1}{2} + \cdots \tag{3}$$

The situation is shown in Fig. 1:

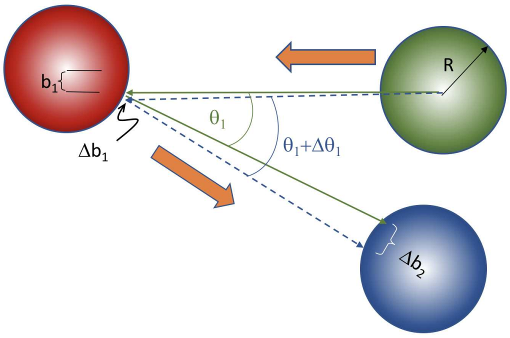


Figure 1. The green ball hits the red ball at an impact parameter b_1, scatters at an angle θ_1, then hits the blue ball at impact parameter b_2. A small change Δb_1 in b_1 leads to a change $\Delta\theta_1$ in θ_1 and a change Δb_2 in b_2.

Assuming $\frac{\Delta b_1}{R} \ll 1$ we can then solve Eq. (3) for $\Delta\theta_1$ using Eq. (2), giving $\Delta\theta_1 \approx \frac{2\Delta b_1}{R\cos\theta_1}$. Let us not assume anything special about θ_1, so that $\cos\theta_1$ is not unusually big or small, and therefore $\Delta\theta_1 \approx \frac{\Delta b_1}{R}$ up to some number of order 1. This makes sense: if Δb_1 is small, then $\Delta\theta_1$ is small too. However, after the first collision, the sphere moves a distance ℓ to the next collision. Then the impact parameter for the second collision changes by $\Delta b_2 = \ell\Delta\theta_1$. This implies $\Delta\theta_2 \approx \frac{\Delta b_2}{R} = \frac{\ell}{R}\Delta\theta_1$.

In this way, the angle change has grown by a factor of $\frac{\ell}{R}$. After the next collision, we would similarly find $\Delta\theta_3 = \frac{\ell}{R}\Delta\theta_2 = \left(\frac{\ell}{R}\right)^2 \Delta\theta_1$ and so on. Thus after N collisions,

$$\Delta\theta_N \approx \left(\frac{\ell}{R}\right)^N \Delta\theta_1 \tag{4}$$

Even if $\Delta\theta_1$ is very very small this factor of $\left(\frac{\ell}{R}\right)^N$ quickly becomes very large. For a gas at room temperature, $\ell \approx 10^{-7}$ m and $R \approx 10^{-10}$ m. Then $\frac{\ell}{R} \approx 10^3$. Thus after just a few collisions, this $\left(\frac{\ell}{R}\right)^N$ factor can make small effects very very big.

For a concrete example, let's estimate the effect on the trajectory of a gas molecule from waving your arm from far away. Displacing your arm by the distance Δr will change the gravitational force by $\Delta F = \frac{\partial F}{\partial r}\Delta r$. So,

$$\Delta F = \frac{d}{dr}\left(G_N \frac{m_1 m_2}{r^2}\right)\Delta r = -2G_N \frac{m_{\text{atom}} m_{\text{arm}}}{r^3}\Delta r \tag{5}$$

Say $m_{\text{atom}} = 10^{-27}$kg, $m_{\text{arm}} = 1$kg and $\Delta r = 1$m and you are standing $r = 10$km away. Then

$$\Delta F \approx \left(10^{-11}\frac{Nm^2}{\text{kg}^2}\right)\frac{(10^{-27}\text{kg})(1\text{kg})}{(10^4 m)^3}(1m) = 10^{-50}N \tag{6}$$

This will cause an acceleration of $a = \frac{\Delta F}{m_{\text{atom}}} = 10^{-23}\frac{m}{s^2}$ on the atom. Over a time $\tau \sim 10^{-9}s$ between two collisions, this hand-waving has moved the atom by around $\Delta b_1 = a\tau^2 \approx 10^{-32}m$, much much less than the size of an atom, and so $\Delta\theta_1 \sim \frac{\Delta b_1}{R} \approx 10^{-22}$. This is a tiny tiny angle. However, after N collisions

$$\Delta\theta_N = 10^{-22}\left(\frac{\ell}{R}\right)^N = 10^{-22}(10^3)^N \tag{7}$$

Thus after only 8 collisions over $10^{-8}s$, the change in angle is of order 1! This is chaos.

Suppose instead of a hand waving on earth, we consider the wing-waving of a flea across the universe, at a distance of $r \approx 10^{27}m$ with $\Delta r = 10^{-6}\ m$. Once the gravitational waves from the flea arrive to our gas, the would change the initial deflection to $\Delta\theta_1 \approx 10^{-97}$. Even for such a tiny tiny angle, it would take only 33 collisions (10^{-7} seconds) for the trajectory of the atom to change completely. After 1 second, every molecule has a vastly different position and velocity from what it would have had if the flea across the universe had not waved its hand.

A chaotic system is one in which the late time behavior is exponentially sensitive to initial conditions: changing the initial condition by a little bit has an enormous effect on the outcome. A flea across the universe changing the trajectory of an atom by order-one after one microsecond illustrates this point. Because of chaos, we can never hope to know the state of a gas exactly. There is simply no physical limit in which a state can be isolated and well defined. We *must* resort to probabilities if we are to make any physical predictions for gases.

In addition, even if we pretend a system is completely isolated – turn off gravity and fleas, etc – the system is still chaotic due to (classical) uncertainty. With an even exponentially small uncertainty on the initial condition, the final state after long enough time will be completely unknown. Using the same numbers as above, if we specify the initial condition to one part in 10^{97} after $1\,\mu s$ the state has order 1 uncertainty. In quantum mechanics, the relevant uncertainty on the initial state is not the Heisenberg uncertainty (on knowing position given momentum), but on how well we can actually know the initial state wavefunction. Again, there is no hope of solving the time-evolution exactly.

3 Maxwell and Molecular Chaos

Next, we want to show that systems tend towards flat probability distributions. This was first understood by Maxwell, who developed his understanding through the kinetic theory of gases.

3.1 Equilibration of molecular velocities

Say we have a gas with different types of molecules in it, of different masses, m_1, m_2, etc. If the gas is in equilibrium, then there will be well-defined probabilities for the velocities of the different molecule types. What can we say about these probabilities?

Pick randomly one molecule of mass m_1 and one of mass m_2. Say these two molecules come in to hit each other with incoming velocities $\vec{v}_1$ and $\vec{v}_2$, respectively. If there is no preferred direction in the system, then the velocities are equally likely to point in any direction. Thus the dot product $\vec{v}_1 \cdot \vec{v}_2 = |\vec{v}_1||\vec{v}_2|\cos\theta$ is equally like to be positive or negative and therefore the expectation value of the dot product of the incoming velocities must be zero:

$$\langle \vec{v}_1 \cdot \vec{v}_2 \rangle = \langle |\vec{v}_1| \rangle \langle |\vec{v}_2| \rangle \langle \cos\theta \rangle = \langle |\vec{v}_1| \rangle \langle |\vec{v}_2| \rangle \frac{1}{\pi} \int_0^\pi d\theta \cos = 0 \tag{8}$$

That is, two random initial velocities are **uncorrelated**.

For collisions, it is often helpful to work in the center of mass frame. The velocity of the center of mass is

$$\vec{v}_{\text{cm}} = \frac{1}{m_1 + m_2}(m_1 \vec{v}_1 + m_2 \vec{v}_2) \tag{9}$$

Note that $\vec{v}_{\text{cm}}$ does not change as a result of the collision, by momentum conservation. We can shift to the center of mass frame by $\vec{v}_1 \to \vec{v}_1 - \vec{v}_{\text{cm}}$ and $\vec{v}_2 \to \vec{v}_2 - \vec{v}_{\text{cm}}$. The relative velocity:

$$\overrightarrow{\Delta v} = \vec{v}_1 - \vec{v}_2 \tag{10}$$

is unchanged by this shift. Since $\vec{v}_1$ and $\vec{v}_2$ pointed in random directions, $\overrightarrow{\Delta v}$ also points in a random direction as does $\vec{v}_{\text{cm}}$. Moreover: $\overrightarrow{\Delta v}$ and $\vec{v}_{\text{cm}}$ are totally uncorrelated: one can have any relative velocity with any $\vec{v}_{\text{cm}}$. Thus,

$$\langle \overrightarrow{\Delta v} \cdot \vec{v}_{\text{cm}} \rangle = 0 \tag{11}$$

Next, write

$$\overrightarrow{\Delta v} \cdot \vec{v}_{\text{cm}} = \frac{1}{m_1 + m_2}(\vec{v}_1 - \vec{v}_2) \cdot (m_1 \vec{v}_1 + m_2 \vec{v}_2) \tag{12}$$

$$= \frac{m_1 v_1^2 - m_2 v_2^2 + (m_2 - m_1)\vec{v}_1 \cdot \vec{v}_2}{m_1 + m_2} \tag{13}$$

Now take the average value of the terms in this equation over all possible choices for molecules 1 and 2 with masses m_1 and m_2. Since we already established that $\langle \vec{v}_1 \cdot \vec{v}_2 \rangle = 0$ and $\langle \overrightarrow{\Delta v} \cdot \vec{v}_{\text{cm}} \rangle = 0$ we must therefore have

$$\langle m_1 \vec{v}_1^2 \rangle = \langle m_2 \vec{v}_2^2 \rangle \tag{14}$$

Dividing by two, we conclude that the **average kinetic energy** of any molecular species in the gas is the same. This calculation of Maxwell's was one of the first theoretical results demonstrating progress towards equilibrium.

I think Maxwell's calculation is remarkable because it is so simple, yet so profound. All we used was that the velocities point in random directions, and we learned something non-trivial about the magnitudes of the velocities. Not only do heavier molecules move slower, on average, but the average kinetic energy for any molecules is a universal quantity determined by the state of the gas, independent of the mass.

3.2 Molecular chaos

Note that Maxwell's argument depended on $\langle \vec{v}_1 \cdot \vec{v}_2 \rangle = 0$, which we justified with the logic that there is no preferred direction in the system, so the angle between two randomly chosen velocities should be evenly distributed. This justification is very reasonable, but it is still an assumption:

- The **assumption of molecular chaos**: velocities of colliding particles are independent of each other, and independent of the position of the collision.

This assumption is an excellent approximation for most physical systems. It is however, never exactly true.

To turn the assumption into an equation, it is helpful to refer to the state of the entire gas at time t as a point $(\vec{q}_i, \vec{p}_i)$ in phase space. Note that phase space is enormous; it is $6N \approx 10^{24}$ dimensional. We are interested in the probability $P(\vec{q}_i, \vec{p}_i, t)$ of finding all the particles with given positions and momenta at the same time t; i.e. if there are N molecules in the gas then P is a $6N+1$ dimensional function: $P = P(\vec{q}_1, \vec{p}_1, \cdots, \vec{q}_N, \vec{p}_N, t)$. The assumption of molecular chaos lets us write

$$P(\vec{q}_1, \vec{p}_1, \cdots, \vec{q}_N, \vec{p}_N, t) = P_1(\vec{q}_1, \vec{p}_1, t) \times \cdots \times P_1(\vec{q}_N, \vec{p}_N, t) = \prod_j P_1(\vec{q}_j, \vec{p}_j, t) \tag{15}$$

for a simpler function $P_1(\vec{q}, \vec{p}, t)$ of just 7 variables. Now, the expectation value of observables, like the average velocity-squared, can be computed just with 7-dimensional function P_1 rather than the 10^{23} dimensional function P:

$$\langle \vec{v}_k^2 \rangle(t) = \int d^{3N}q d^{3N}q \frac{\vec{p}_k^2}{m^2} P(\vec{q}_1, \vec{p}_1, \cdots, \vec{q}_N, \vec{p}_N, t) = \int d^3q_k d^3p_k \frac{\vec{p}_k^2}{m^2} P_1(\vec{q}_k, \vec{p}_k, t) \tag{16}$$

In the last step, we have used $\int d^3q d^3p P_1(p, q) = 1$ for all the $p's$ and q's other than p_k, q_k. So the velocity distribution is the same for any particle, and determined by a single 7-dimensional function.

To see the subtlety in the assumption of molecular chaos, return to the example of two colliding particles of masses m_1 and m_2 with uncorrelated initial velocities $\vec{v}_1$ and $\vec{v}_2$. Are the outgoing velocities $\vec{v}_1'$ and $\vec{v}_2'$ uncorrelated as well? If you think about it for a moment, it is easy to see that the answer must be no. For example, say $m_1 \gg m_2$, like a truck hitting a bicycle. After the collision, the truck and the bicycle will be going close to the truck's initial direction. So the angle θ measured with respect to the incoming truck's direction will likely be close to zero and therefore $\langle \vec{v}_1' \cdot \vec{v}_2' \rangle \neq 0$. In other words, after a collision, two uncorrelated velocities become correlated.

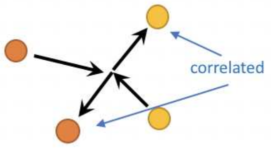


Figure 2. When two molecules collide, the outgoing velocities are correlated.

Since outgoing particles eventually become incoming particles when they collide again, the assumption of molecular chaos must not strictly hold. Why then can we use it?

The key to tracking what happens to the correlations is that the dynamics of multibody systems are chaotic, as discussed in Section 2. Since we never know exactly what the initial state is of any physical system – there is some measurement uncertainty or uncertainty due to motion of fleas across the universe – we should properly specify the state not as a point in phase space but as a region R in phase space around the point (p_i, q_i) of volume $\Delta V = (\Delta q)^{3N}(\Delta p)^{3N}$ with Δq and Δp our (classical) uncertainty on the position and momentum. To be concrete, say we have a gas of hard spheres, and let us track two of them that collide head on at $t = 0$. Right after the collision, their outgoing momenta will be highly correlated (still back-to-back). If we shift the initial momenta p_i or positions q_i by a little bit, the outgoing momenta will be slightly different, but still essentially back-to-back. So molecular chaos, Eq. (15) does not seem to apply after one collision: given one outgoing momentum, we have a pretty good idea of what the other one is.

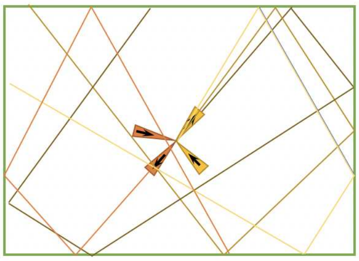


Figure 3. After a short time, nearby points in phase space follow highly-correlated trajectories through phase space.

But now let's wait a bit longer. Suppose for the exact point $(\vec{q}_i, \vec{p}_i)$ there are 9 collisions in $10^{-8}\, s$. Now consider a point $10^{-32}m < \Delta q$ away still within the region R. As we saw in Eq. (7) the trajectory of such a point will be off by an angle of order 1 after the 8 collisions, so will miss the 9th collision and move on to very different region in phase space. Thus, as time moves on, the original region R of volume ΔV gets fragmented and split up into an enormous number of disconnected regions:

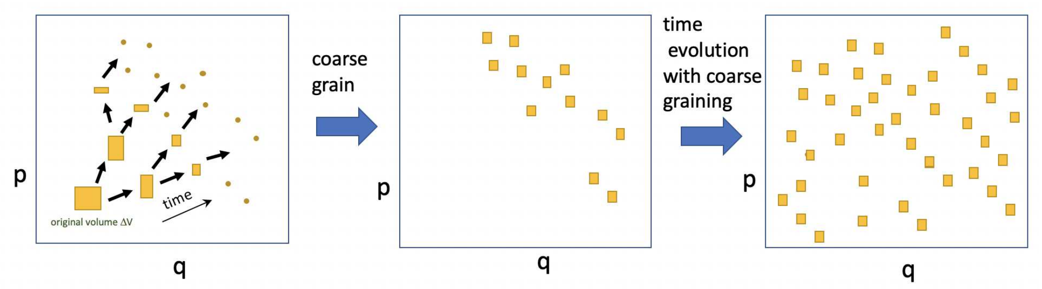


Figure 4. Over time, a phase space region R of size ΔV fragments into an enormous number of small regions with the same total volume (left). When we coarse grain, the phase space volume increases (middle). Further time evolution with coarse graining fills up more and more of phase space (right).

This doesn't yet explain molecular chaos – the small disconnected regions are still highly correlated with each other.

Now we invoke a result from classical mechanics called **Liouville's theorem**, which says that the sum of the phase-space volumes of all the fragments is the same as the original volume ΔV of the region R (it is easy to prove Louiville's theorem using Hamilton's equations of motion[1]). This means the volumes of the fragments are getting smaller and smaller after each collision. This is illustrated in the first panel of Fig. 4. However, we already said that we cannot possibly know what point in phase space our system is in with a precision better than ΔV. So we must **coarse-grain** these small phase space volumes, treating them all the same way. That is, we must accept that we cannot distinguish nearby points. Although the original small fragmented regions were highly correlated, nearby fragments that we absorb through coarse graining are not. In other words, the correlations are still there, but we cannot ever measure anything sensitive to them: it would be beyond our experimental resolution. So the effective phase space volume, from the point of view of things we can physically distinguish, is increasing (middle panel). When we coarse-grain, the correlations are completely washed out.

In summary, if we had perfect knowledge of a system, the trajectories of all the particles would be highly correlated. However, with any arbitrarily small amount of uncertainty, as we *must* have do to the lack of knowledge of fleas across the universe, those correlations get dispersed into small phase space fragments due to the chaotic nature of multibody systems. When these fragments are coarse-grained, again due to our classical uncertainty, the correlations are lost forever.

This simple fact, that chaos forces us to discard correlations, is our first indication of the arrow of time. Microscopic laws of physics are time-reversal invariant (a video of two point masses orbiting each other looks realistic in reverse), but average properties of the system made up of a large number of microscopic particles change in a fixed time direction (a video of a gas expanding does not look realistic in reverse). We saw that irreversibility arises because the correlations, which would be there if we had the exact solution to the equations of motion, become dilute in phase

[1] To prove Liouville's theorem, think about an infinitesimal time evolution as a change of variables, from q, p to $q' = q + \dot{q}\, dt$ and $p' = p + \dot{p}\, dt$. Then, to leading order in dt, the phase space volume $dq dp$ changes to

$$dq'\, dp' = \frac{\partial q'}{\partial q}\frac{\partial q'}{\partial p} dq dp = \left(1 + \frac{\partial \dot{q}}{\partial q} dt + \frac{\partial \dot{p}}{\partial p} dt\right) dq dp = \left(1 + \frac{\partial}{\partial q}\frac{\partial H}{\partial p} dt - \frac{\partial}{\partial q}\frac{\partial H}{\partial p} dt + \cdots\right) dq dp = dq dp \tag{17}$$

where Hamilton's equations of motion $\dot{q} = \frac{\partial H}{\partial p}$, $\dot{p} = -\frac{\partial H}{\partial q}$ were used. Thus the phase space volume does not change upon time evolution. The generalization to many q_i, p_i just involves adding indices.

space and inaccessible. In other words, the information in the correlations becomes lost. This loss of information is a key feature of the progression of time. The arrow of time is the direction in which our ignorance grows.

4 Boltzmann's *H* theorem

Not long after Maxwell's work, Ludwig Boltzmann attempted to make the arguments for approaching equilibrium more robust and general.

Say we have some state a. For concreteness, think of a as a specification of all the momenta and positions of all the particles in a gas. Let's denote by $P_a(t)$ the probability of finding the gas in state a. Let's denote the rate for a to turn into some other state b by T_{ab}, and the rate for b to turn into a by T_{ba}. Then the rate of change of $P_a(t)$ is given by summing over possible states b by the equation

$$\frac{d}{dt}P_a(t) = \underbrace{\sum_b P_b(t)T_{ba}}_{\text{transitions } b\to a} - \underbrace{P_a(t)\sum_b T_{ab}}_{\text{transitions } a\to b} \tag{18}$$

If a state b in the sum can never turn into a (so $T_{ab}=0$) or if a can never turn into b (so $T_{ab}=0$), there is no point in including b in this sum together. So we can separate the problem into exclusive sets of states that can turn into each other. So without loss of generality, we assume $T_{ab}\neq 0$.

The key property of physical systems that allows equilibrium to be approached is **the principle of detailed balance**: the transition rate from one state a to b is the same as the rate for b going to a: $T_{ab}=T_{ba}$. In classical mechanics, this follows from time-reversal invariance of the equations of motion; in quantum mechanics, it follows from the fact that the interaction Hamiltonian H_{int} is Hermitian, i.e. the transition rate, as appearing in the Born approximation for example satisfies

$$T_{ab} = |\langle a|H_{\text{int}}|b\rangle|^2 = \langle a|H_{\text{int}}|b\rangle\langle b|H_{\text{int}}^\dagger|a\rangle = \langle b|H_{\text{int}}|a\rangle\langle a|H_{\text{int}}^\dagger|b\rangle = |\langle b|H_{\text{int}}|a\rangle|^2 = T_{ba} \tag{19}$$

In quantum field theory, detailed balance follows from unitarity (probability conservation). The principle of detailed balance is often used in chemistry: in equilibrium the rate for a reaction $A\to B$ must be the same as the rate for the reverse reaction $B\to A$.

Once we know that $T_{ab}=T_{ba}$ it follows from Eq. (18) that

$$\frac{d}{dt}P_a(t) = \sum_b T_{ab}[P_b(t) - P_a(t)] \tag{20}$$

This is a powerful equation. For example, say there are only two states. Then this equation says that if $P_a(t) > P_b(t)$ then $P_a(t)$ will go down, and if $P_b(t) > P_a(t)$ then $P_a(t)$ will go up. Thus, over time, the probabilities will become the same: $\lim_{t\to\infty}P_a(t) = \lim_{t\to\infty}P_b(t)$.

To see what happens when there are N states, we consider the quantity[2]

$$H(t) = -\sum_a P_a(t)\ln P_a(t) \tag{21}$$

Then

$$\frac{d}{dt}H(t) = -\sum_a \left[\frac{d}{dt}P_a(t)\right]\ln P_a(t) - \frac{d}{dt}\sum_a P_a(t) \tag{22}$$

Since $\sum_a P_a(t) = 1$ the second term is zero. Thus

$$\frac{d}{dt}H(t) = \sum_a \sum_b T_{ab}\,[P_a(t) - P_b(t)]\ln P_a(t) \tag{23}$$

Switching the labels a and b we also have

$$\frac{d}{dt}H(t) = \sum_a \sum_b T_{ab}[P_b(t) - P_a(t)]\,\ln P_b(t) \tag{24}$$

[2] This definition of H may seem like it was pulled out of thin air. We will see in the next lecture that it is in fact related to the number of configurations and to entropy.

Averaging these two equations gives

$$\frac{d}{dt}H(t) = \frac{1}{2}\sum_a \sum_b T_{ab}[\ln P_a(t) - \ln P_b(t)][P_a(t) - P_b(t)] \tag{25}$$

Now, $\ln x$ is a monotonic function of x, so if $P_a > P_b$ then $\ln P_a > \ln P_b$. This means every term in the sum is non-negative and therefore

$$\boxed{\frac{d}{dt}H(t) \geqslant 0} \tag{26}$$

This is known as the **Boltzmann H theorem**.

If $H(t)$ is changing, then the probabilities must also be changing and we cannot be in equilibrium. Thus, equilibrium is only possible if $\frac{d}{dt}H(t) = 0$ which only happens if $P_a(t) = P_b(t)$ for all states a and b. That is,

- In equilibrium, the probabilities of finding the system in any two states a and b for which transitions can possibly occur ($T_{ab} \neq 0$) are the same

This is the postulate of equal a priori probabilities.

For a simple example, imagine you have 5 coins in a box and they all start heads up. Then you start throwing golf balls into the box, one by one. Each golf ball could hit a coin and flip it so that it may then land heads or tails. So initially, $P(\text{HHHHH}) = 1$. But after a long enough time, the probability of any configuration will be the same, $\frac{1}{2^5} = \frac{1}{32}$ and $P(\text{HHHHH}) = \frac{1}{32}$.

Note that the Boltzmann H theorem is not time-reversal invariant: H increases as we move forward in time, not backwards in time. How did this happen? The microscopic equations of motion are time-reversal invariant, so where did this arrow of time come from? In other words, for each sequence of events which takes $a \to b \to c$ there is exactly one sequence of events which goes $c \to b \to a$. For colliding molecules, we simply reverse the velocities of the outgoing molecules and then we get the initial velocities back. This mystery is known as **Loschmidt's paradox**.

To understand Loschmidt's paradox, we have to decode the implicit assumptions in Boltzmann's H theorem. First note that if we knew exactly what the state was and evolved it with time perfectly, we would always know the state, so $P_a(t)$ would either be zero or 1 for all time, just for different a at different times. For example, with the coins, if they start as HHHHH and after one hit go to HHTTH then HTHTT and so on, there is only ever one configuration possible. So $P_a(t) = 1$ for that configuration and $P_a(t) = 0$ for the others and thus $H = 0$ for all time. Thus, $\frac{dH}{dt} = 0$, which is consistent with Boltzmann's H theorem, but does not lead to the postulate of equal a priori probabilities.

So why do we say the probabilities change with the coins example? In that example, what we mean by the probability is that if threw the ball in a bunch of times, but didn't keep count, and didn't try to calculate everything, then the chance of finding any given configuration at some random later time is the same. One way to isolate our assumption is that we are implicitly talking about the time-averaged probabilities: over an interval T, what fraction of the time is a given configuration present? For large enough T these time-averaged probabilities will average out. Alternatively, we can try the experiment over and over again. Because it's hard to control the ball, we would get a different answer each time; there is inherent chaos in the system that makes it impossible to actually predict what happens. So when we repeat the experiment, what we mean by the probability is an average over the unknown parts of the initial conditions: it's not just HHHHH, but HHHHH with a flea in Australia flapping its wings, etc. Either way, the probabilities are changing because we do not have perfect information, either by choice (the time averaging) or necessity (chaos).

The key assumption in Boltzmann's H theorem is that a state a can transition to multiple states b. This is generally not possible in a unitary causal theory, since the time evolution should uniquely determine $a(t)$. So even though the theory is causal at the microscopic level, the apparent violation of causality, allowing $H(t)$ to grow, comes about because we discard information that is in principle available but practically inaccessible. We do this by time-averaging, by averaging over unknown parts of the initial conditions, and by coarse graining. These, often implicit, operations are the key to the Boltzmann H theorem, the postulate of equal a priori probabilities, and the increase of entropy (Lecture 6).

5 The ergodic hypothesis

Another concept related to chaos and coarse graining that we will need is **ergodicity**.

- An **ergodic system** is one for which the average over all possible states (the ensemble average) is the same as the average over the states that a given state will evolve into over time (the time average).

In other words, we can find the probabilities of a system being in a state at a given time t by looking at the possible states a system passes through over time.

We used ergodicity already in describing diffusion when we equated the probability of finding a particle at position x at time t, $P(x, t)$ with the number density: $n(x, t) = NP(x, t)$. Strictly speaking, $P(x,t)$ is a smooth function but $n(x,t)$ is not. At any time $n_{\text{true}}(x,t) = \frac{1}{V}\sum_i \delta(x - x_i(t))$ since the molecules are only ever at some precise points. When we write $n(x,t) = NP(x,t)$ what we mean is the time-averaged number density $n(x,t) = \frac{1}{T}\int_0^T dt' n_{\text{true}}(x, t + t')$ for some time T greater than the typical collision time τ will agree with $NP(x, t)$. Similarly, $P(x, t)$ is an average: we average over the possible random-walk paths any molecules could have taken. Each walk for each molecule constitutes a microstate. So the ensemble average over microstates, $P(x,t)$, is the time-average for a particular microstate, $\langle n(x,t)\rangle$. We usually write $n(x,t)$ instead of $\langle n(x,t)\rangle$ with the use of the ergodic hypothesis implicit.

The idea behind ergodicity that a classical trajectory through phase space $\vec{q}(t), \vec{p}(t)$ will eventually pass close to any other accessible phase space point $\vec{q}_0$, $\vec{p}_0$. For example, a gas molecule bouncing around a room, will eventually go everywhere and eventually have any momentum. Unfortunately, most systems are not ergodic, in the strict mathematical sense. Hence ergodic "hypothesis". An example of a non-ergodic system is one with closed orbits in phase space. It is not hard to find such systems. For example, imagine a circular pool table. A billiard ball bouncing around this table would never reach points closer than a certain distance from the center. It is non-ergodic. A cardioid pool table is ergodic:

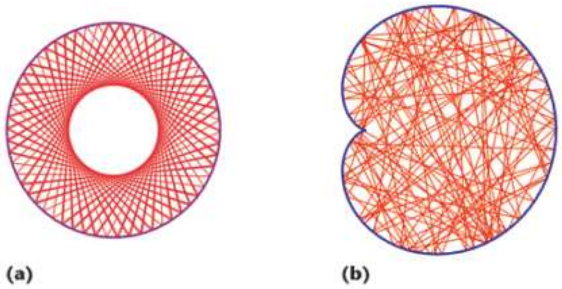


Figure 5. A circular container (left) is non-ergodic while a cardioid container (right) is. If let to continue indefinitely, the left trajectory would never reach the center, while the right trajectory would eventually fill the entire volume.

Most systems are believed to have regions of phase space which do not mix, so generally ergodicity is not exact. Moreover, even in systems where it does hold, the time for all the points in phase space to be passed through is astronomical. This is simply because the volume phase space is astronomical: $6N \sim 10^{24}$ dimensional (in constrast to the pictures in Fig. 5) which are 2 dimensional. For the trajectory of molecules in a gas to fill out a 10^{24} dimensional space will take a very very long time.

The fact that systems are not strictly ergodic and take a long time to be approximately ergodic is largely irrelevant. The reason is that we coarse grain the phase space. So instead of a $6N \sim 10^{24}$ dimensional space, we treat phase space as essentially 6 dimensional as in Eq. (15). Although it takes forever for the entire collection of 10^{24} molecules to pass near any point in the 10^{24} dimensional phase space, it does not take long at all for *one* of the 10^{24} molecules to get close to any given position and velocity.

In any case, the main reason that people care about ergodicity it is the simple fact that experiments measure time averages, but the things we compute in statistical mechanics are ensemble averages. Without ergodicity our calculations would not allow us to make any physical predictions.

6 Counting states

Boltzmann's H theorem immediately implies the

- **Postulate of equal a priori probabilities**: all accessible microstates are equally likely.

This postulate is really a theorem, to the extent that Boltzmann's H theorem is a theorem. It is rigorously true, as long as we coarse-grain.

The word "accessible" is required because we only proved Boltzmann's H theorem for the sets of states for which $T_{ab}\neq 0$, as in Eq. (18). For example, if we have a box of gas, a state with all the molecules on the left side and a state with them all on the right side are both accessible to each other. However, if our box had a partition in the middle then the two states would not be accessible. Accessibility can be limited by physical barriers, or by conservation laws (number of particles, energy, charge, etc.).

To make the postulate precise, we need to know how to compare probabilities. There is always some measure for the probabilities. For example, if the states are the phase space points of a gas $(\vec{q}_i,\, \vec{p}_i)$ there is some intrinsic resolution $(\Delta q)^{3N}(\Delta p)^{3N}$ to how well we can determine the points. Ultimately, the phase space resolution is limited by quantum uncertainty: $\Delta q \Delta p \geqslant \hbar$. For classical statistical mechanics, one does not need to invoke Planck's constant,[3] but one does need some notion of Δq and Δp – the probability of finding a system at an *exact* phase space point is necessarily zero. Thus, we will stick to the general notation of Δq and Δp for our phase space resolution, and plan to take $\Delta p \Delta q \rightarrow 0$ at the end to recover the infinite precision by which a classical system can in principle be specified. It is only in situations that are very dense, so that more than one particle might be in the same phase space point, that the actual value of $\Delta p \Delta q$ is relevant. For these situations, quantum statistical mechanics is necessary, as we will see starting in Lecture 10.

Let's take an example: an ideal gas in a box with energy E. An **ideal gas** is one where all the collisions are perfectly elastic. We treat it classically, so that positions and momenta are continuous. We assume that there are no external forces or external potential, so the energy of a particle is independent of position $\vec{q}_i$. Then the number of states is the product of the number of choices for momenta and number of states of position

$$\Omega = \Omega_q \Omega_p \tag{27}$$

Then for every state $(\vec{q}_i, \vec{p}_i)$ there is another state with the same $\vec{q}_i$ but different $\vec{p}_i$.

To count the $\vec{q}_i$, let us assume that each gas molecule can be some box of size Δq, so that there are $\Omega_1 = \frac{L}{\Delta q}$ choices for one molecule in 1 dimension and $\frac{V}{(\Delta q)^3}$ choices in 3 dimensions. We assume that any of the molecules can be in any position. Thus, for the whole ensemble,

$$\Omega_q = \left[\frac{V}{(\Delta q)^3}\right]^N \tag{28}$$

where N is the number of particles. Although Ω_q depends on the arbitrary scale Δq, we will be able to take $\Delta q \rightarrow 0$ once we have used Ω_q to compute a physical quantity.

For momenta, the calculation is harder since energy is involved. Say we have a classical monatomic gas where all the energy is in kinetic energy. Then if we know the total energy E we have a constraint:

$$E = \sum_j \frac{\vec{p}_j^{\,2}}{2m} \tag{29}$$

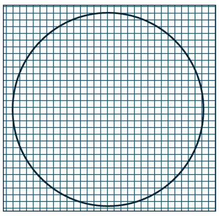
[3] In classical mechanics, the properties of phase space and probability distributions on it have been very thoroughly studied. We will try to keep our discussion intuitive and non-technical as much as possible. See Grandy's book in pdf form on the Canvas site for more discussion.

There are going to be many choices of $\vec{p}_j$ for this constraint equation to hold. How do we count them?

Let's start with 2 particles in 1 dimension. Then

$$2mE = p_1^2 + p_2^2 \tag{30}$$

We want to count the number of small boxes of size Δp^2 that a circle with radius $R = \sqrt{2mE}$ passes through. To do this, we note that the length of circle passing through each relevant box is Δp:



(31)

So the total number of boxes around the circumference is $\Omega_p = \frac{2\pi R}{\Delta p} = \frac{2\pi\sqrt{2mE}}{\Delta p}$. If there are 3 particles in 1 dimension, then

$$2mE = p_1^2 + p_2^2 + p_3^2 \tag{32}$$

the number of states is determined by the surface area of a sphere: $\sigma_3 = 4\pi$. Then we get that $\Omega_p = \frac{4\pi R^2}{\Delta p^2} = \frac{4\pi\,(2mE)}{\Delta p^2}$.

If there are N particles in 3 dimensions, then

$$2mE = p_{1x}^2 + p_{1y}^2 + p_{1z}^2 + \cdots + p_{Nx}^2 + p_{Ny}^2 + p_{Nz}^2 \tag{33}$$

this is the sum of the squares of $3N$ independent momenta. Then

$$\Omega_p = \sigma_{3N}\left(\frac{\sqrt{2mE}}{\Delta p}\right)^{3N} \tag{34}$$

where σ_d is the surface area of a d-dimensional sphere with radius $r = 1$. For $d = 2$, $\sigma_2 = 2\pi$, for $d = 3$, $\sigma_3 = 4\pi$. For general d the result is[4]

$$\sigma_d = \frac{2(\sqrt{\pi})^d}{\left(\frac{d}{2} - 1\right)!} \tag{37}$$

Therefore,

$$\Omega_p = \frac{2(\sqrt{\pi})^{3N}}{\left(\frac{3}{2}N - 1\right)!}\left(\frac{\sqrt{2mE}}{\Delta p}\right)^{3N} \tag{38}$$

For large N, we can write $3N - 1 \approx 3N$ and also Stirling's approximation $N! \approx e^{-N}N^N$ so that $\left(\frac{3}{2}N\right)! \approx e^{-\frac{3}{2}N}\left(\frac{3}{2}N\right)^{\frac{3N}{2}}$ giving

$$\Omega_p = e^{\frac{3}{2}N}\left(\frac{4\pi mE}{3N(\Delta p)^2}\right)^{\frac{3N}{2}} \tag{39}$$

Combining with the phase space for position in Eq. (28), we get

$$\boxed{\Omega(N, V, E) = e^{\frac{3}{2}N}\left(\frac{V}{(\Delta q\Delta p)^3}\right)^N\left(\frac{4\pi mE}{3N}\right)^{\frac{3N}{2}}} \tag{40}$$

[4] To derive Eq. (37) we first compute the 2D integral

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, e^{-x^2 - y^2} = 2\pi\int_0^{\infty} r\, dr\, e^{-r^2} = \pi \tag{35}$$

Then we can generalize to d dimensions:

$$(\sqrt{\pi})^d = \left[\int_{-\infty}^{\infty} dp\, e^{-p^2}\right]^d = \int_{-\infty}^{\infty} dp_1 \cdots dp_n e^{-p_1^2 - \cdots - p_d^2} = \sigma_d \int dr\, r^{d-1}\, e^{-r^2} \tag{36}$$

The last integral is a 1D integral that we can do with mathematica. It gives $\frac{\sigma_d}{2}\Gamma\left(\frac{d}{2}\right)$ where $\Gamma(x)$ is the Gamma function. The Gamma function is a generalization of the factorial function. For positive integers, $\Gamma(n) = (n-1)!$.

A key feature of the number of states is that it is an extremely rapidly growing function of energy – it grows like energy to the power 10^{24}. That is,

$$\Omega(E) \sim E^{10^{24}} \tag{41}$$

So when you add energy to a system, the number of states grows exponentially. For example, say we increased the energy by 0.00001% ($E \rightarrow E + 10^{-7}E$). Then the number of states grows by

$$\#\text{new states} = \Omega(E + 10^{-7}E) - \Omega(E) = 10^{-7}\partial_E E^{10^{24}} = 10^{18}\,\Omega(E) \tag{42}$$

This is an enormous number, 10^{18}, times a ridiculously enormous number $E^{10^{24}}$.

In the next lecture, we will consider situations that have more contributions to the energy in Eq. (33), such as from vibrational modes of a molecule. In such situations, the calculation is similar, resulting in $\Omega \sim E^{\frac{f}{2}N}$ rather of $\Omega \sim E^{\frac{3}{2}N}$ with f the number of degrees of freedom in which the molecules can store energy.

7 Maxwell-Boltzmann distribution

After all this rather formal introduction, we are finally ready to compute something observable: the velocity distribution of molecules in a gas.

What is the probability of finding the p_x component of the momentum of one molecule in a small region of size Δp around p_x? According to the postulate of equal a priori probabilities, this probability is proportional to the number of states compatible with this restriction, divided by the total number of states: $P(p_x) = \frac{\Omega_{p_x \text{fixed}}}{\Omega_{\text{total}}}$.

Once we fix p_x the remaining energy is $E' = E - \frac{p_x^2}{2m}$. So the remaining phase space volume, with energy E' is

$$\Omega(E') = e^{\frac{3N-1}{2}}\left(\frac{V}{(\Delta q \Delta p)^3}\right)^N \left(\frac{4\pi m E'}{3N-1}\right)^{\frac{3N-1}{2}} \Delta p \approx e^{\frac{3N}{2}}\left(\frac{V}{(\Delta q \Delta p)^3}\right)^N \left(\frac{4\pi m E'}{3N}\right)^{\frac{3N}{2}} \Delta p \tag{43}$$

The probability of finding the x component of momentum between p_x and $p_x + \Delta p$ is therefore

$$\frac{\Delta P}{\Delta p} \equiv \frac{P(p_x \,\text{to}\, p_x + \Delta p)}{\Delta p} = \frac{\Omega(E')}{\Omega(E)\Delta p} = \frac{(E')^{3N/2}}{(E)^{3N/2}} = \left(1 - \frac{p_x^2}{2mE}\right)^{3N/2} \tag{44}$$

This function ΔP looks a lot like the limit definition of the exponential function

$$e^{-x} = \lim_{N \rightarrow \infty}\left(1 - \frac{x}{N}\right)^N \tag{45}$$

To make it match exactly, let us introduce the average energy

$$\bar{\varepsilon} = \frac{E}{N} \tag{46}$$

Writing Eq. (44) in terms of $\bar{\varepsilon}$ we can then take $N \rightarrow \infty$ using Eq. (45):

$$\frac{\Delta P}{\Delta p} = \left[\left(1 - \frac{1}{N}\frac{p_x^2}{2m\,\bar{\varepsilon}}\right)^N\right]^{3/2} \xrightarrow[N \rightarrow \infty]{} \left[\exp\left(-\frac{p_x^2}{2m\,\bar{\varepsilon}}\right)\right]^{3/2} = \exp\left(-\frac{3\,p_x^2}{4m\,\bar{\varepsilon}}\right) \tag{47}$$

In taking the limit, we have messed up the normalization. We would like the probabilities to be normalized so that $\sum_{p_x} \Delta P(p_x) = 1$. Rather than working with the discrete sum, it is easier to go straight to the continuum limit. Defining $\frac{dP}{dp_x} = C\frac{\Delta P}{\Delta p}$ and choosing C so that $\int \frac{dP}{dp_x} dp_x = 1$ we get

$$\frac{dP(p_x)}{dp_x} = \sqrt{\frac{3}{4\pi m\,\bar{\varepsilon}}}\exp\left(-\frac{3\,p_x^2}{4m\,\bar{\varepsilon}}\right) \tag{48}$$

Note that Δq and Δp have dropped out of this expression so we take $\Delta p \rightarrow 0$ and $\Delta q \rightarrow 0$

We could repeat the calculation for p_y and p_z. Since the theory is rotationally symmetric, we get the same answer. Therefore,

$$\frac{d^3P(\vec{p})}{dp_xdp_ydp_z}=\left(\frac{3}{4\pi m\,\bar{\varepsilon}}\right)^{3/2}\exp\left(-\frac{3p_x^2}{4m\,\bar{\varepsilon}}\right)\exp\left(-\frac{3p_y^2}{4m\,\bar{\varepsilon}}\right)\exp\left(-\frac{3p_z^2}{4m\,\bar{\varepsilon}}\right) \tag{49}$$

In other words

$$\boxed{\frac{d^3P(\vec{p})}{dp_xdp_ydp_z}=\left(\frac{3}{4\pi m\,\bar{\varepsilon}}\right)^{3/2}e^{-\frac{1}{\bar{\varepsilon}}\frac{3\vec{p}^2}{4m}}} \tag{50}$$

This is known as the **Maxwell-Boltzmann distribution**.

As a check, we compute the average value of kinetic energy:

$$\langle\frac{\vec{p}^2}{2m}\rangle=\int d^3p\left(\frac{\vec{p}^2}{2m}\right)\frac{d^3P(\vec{p})}{d^3p}=\int d^3p\frac{\vec{p}^2}{2m}\left(\frac{3}{4\pi m\,\bar{\varepsilon}}\right)^{3/2}e^{-\frac{1}{\bar{\varepsilon}}\frac{3\vec{p}^2}{4m}}=\bar{\varepsilon}=\frac{E}{N} \tag{51}$$

This is as expected, since all energy is kinetic. Note that this result is consistent with Eq. (14) which we derived using kinetic theory.

Now wait a minute ... didn't we say that each state is equally likely, but now we say that each state has probability $e^{-\frac{1}{\bar{\varepsilon}}\frac{3\vec{p}^2}{4m}}$? How are these statements consistent? Let's be careful. What we said is that each microstate with total energy E is equally likely. This is still true. But if we now start grouping the microstates the value of $\vec{p}^2$ for a given molecule, then we find fewer and fewer such microstates with larger values of $\vec{p}^2$. This is because the bigger $\vec{p}$ is, the fewer ways there are to split up the remaining energy among the other molecules, and thus the probability of finding $\vec{p}$ goes down as $\vec{p}^2$ goes up.

If we change variables from $\vec{p}^2$ to speed $v=\frac{\sqrt{\vec{p}^2}}{m}$ (using $d^3p=4\pi|\vec{p}|^2d|\vec{p}|=4\pi m^3v^2dv$) and substitute $\bar{\varepsilon}=\frac{3}{2}k_BT$ (a result we'll derive in the next lecture), we get the **Maxwell-Boltzmann** distribution:

$$\boxed{\frac{dP(v)}{dv}=4\pi v^2\left(\frac{m}{2\pi k_BT}\right)^{3/2}e^{-\frac{mv^2}{2k_BT}}} \tag{52}$$

This looks like

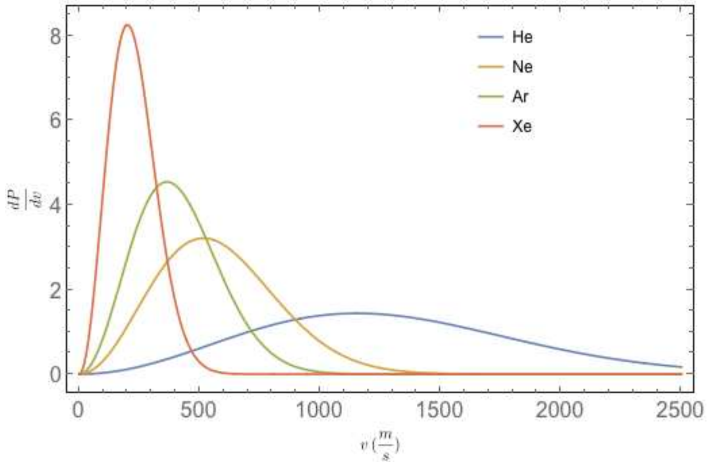


Figure 6. Maxwell-Boltzmann velocity distributions for some gases with the same value of the average energy $\bar{\varepsilon}=\frac{E}{N}=\frac{3}{2}k_BT$.

8 Summary

There were a lot of new concepts introduced in this lecture and some very important results. So let's recap:

- **Chaos**: the trajectory of molecules in a gas are strongly sensitive to even the smallest perturbations, like a flea flapping its wings across the universe.

- **Molecular chaos**: when molecules scatter, their outgoing velocities are correlated. Due to chaos, these correlations are rapidly disperse throughout phase space, dissolving into every smaller, separated regions, like a kind of phase space dust.
- Because the correlations are dilute, they get lost when we **coarse grain** (average over nearby regions in phase space). Coarse graining lets us treat the probability of each molecule occupying a point in its phase space as independent of what the other molecules are doing, as in Eq. (15). It breaks time-reversal invariance.
- The **Boltzmann *H*-Theorem**: with correlations discarded, probabilities tend towards uniformity over phase space.
- The H-theorem follows from the **principle of detailed balance**: the rate for a process and the reverse process is the same.
- **Ergodicity**: the probability of finding a molecule at point $(\vec{p}, \vec{q})$ at a given time t is the same as the probability of finding it at $(\vec{p}, \vec{q})$ averaged over t.
- **Postulate of equal a priori probabilities**: in equilibrium, a system is equally likely to be found in any accessible state.
- **Maxwell-Boltzmann distribution**: the distribution of velocities of gas molecules is computed by counting the number of ways the the total energy of the gas can be distributed among the molecules.

Much of the material in this lecture was abstract and foundational. It built up to the postulate of equal a priori probabilities. We then applied this postulate to derive the Maxwell-Boltzmann distribution in Sections 6 and 7. Going forward, we will do many more applications, referring back to this foundational material when appropriate. If you are confused, don't despair: this is probably the most conceptually difficult lecture in the entire course.

Matthew Schwartz

Statistical Mechanics, Spring 2025

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

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

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

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

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.

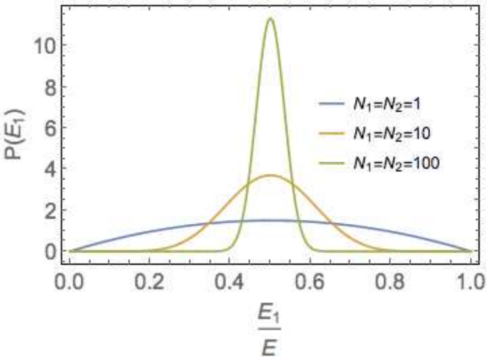
Now, the postulate of equal a priori probabilities implies that the probability of finding a system in a set of states is directly proportional to the number of states. Thus

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

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

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



where N_1 and $N_2=N-N_1$ are the numbers of the different types of gasses and the normalization $C'=E^{-1-\frac{3N}{2}}\frac{\left(1+\frac{3N}{2}\right)!}{\frac{3N_1}{2}!\frac{3N_2}{2}!}$ depends only on the total energy $E=E_1+E_2$, not on how the energy is distributed. We see from the plot 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 because the energy E_1 is the sum over possible values of the energies of the all the particles; we're summing over independent draws from a flat distribution.[1]

What is the expected value of E_1? For a Gaussian (which $P(E_1)$ approaches at large N) 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 $\frac{\partial P(E_1)}{\partial E_1}=0$. For our function $P(E_1)=C'\times E_1^{\frac{3N_1}{2}}(E-E_1)^{\frac{3N_2}{2}}$ 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. When we Taylor expand $\ln P(E,E_1)$ around $E_1=\langle E_1\rangle=E\frac{N_1}{N}$ we find

$$\ln P(E_1)=-\frac{3}{4}(N_1+N_2)\frac{(E_1-\langle E_1\rangle)^2}{\langle E_1\rangle\langle E_2\rangle}+\cdots \tag{8}$$

Writing this quadrtic term as $-\frac{1}{2}\frac{(E_1-\langle E_1\rangle)^2}{\sigma^2}$ as for a Gaussian we can identify

$$\sigma=\sqrt{\frac{2\langle E_1\rangle\langle E_2\rangle}{3(N_1+N_2)}} \tag{9}$$

Thus we see that the width scales like $\sigma\sim\sqrt{\frac{1}{N}}$ at large N_1 or N_2. For 10^{24} particles, the chance of finding the configuration with anything other than the most probable energy allocation scales like $e^{-N}\sim e^{-10^{23}}$: this is 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),

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

[1] To verify the convergence to a Gaussian analytically, you can check that the skewness $S=\frac{\langle(E_1-\langle E_1\rangle)^3\rangle}{\sigma^3}$ and higher moments go to their Gaussian values as $N\to\infty$ as discussed in Lecture 1.

Using that $E = E_1 + E_2$ is fixed, then $\frac{\partial}{\partial E_1} = -\frac{\partial}{\partial E_2}$ and so we can rewrite the second term to get

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

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

$$\beta = \frac{1}{k_B T} \tag{14}$$

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 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 Waals 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 difference 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):

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

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**:

$$S = N k_B \left[\ln \frac{V}{N} + \frac{3}{2} \ln \left(\frac{4\pi m E}{3N h^2} \right) + \frac{5}{2} \right] \tag{18}$$

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 (i.e. how big is S?), rather than just differences of entropy (i.e. $S(E_1) - S(E_2)$ where h and $\Delta p \Delta q$ drop out). The other two 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):

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

Thus,

$$E = \frac{3}{2} N k_B T \tag{20}$$

The average energy per molecule is

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

Next, recall the Maxwell-Boltzmann distribution for momenta

$$\frac{d^3 P(\vec{p})}{dp^3} = \left(\frac{3}{4\pi m \bar{\varepsilon}} \right)^{3/2} e^{-\frac{1}{\bar{\varepsilon}} \frac{3\vec{p}^2}{4m}} \tag{22}$$

Using Eq. (21) this becomes

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

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 $P(\varepsilon) = e^{-\varepsilon / k_B T}$ (cf. Eq. (72) below).

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

$$\langle \frac{p_x^2}{2m} \rangle = \int d^3 p \frac{p_x^2}{2m} \frac{d^3 P(\vec{p})}{dp^3} = \int dp_x dp_y dp_z \frac{p_x^2}{2m} \left(\frac{1}{2\pi m k_B T} \right)^{3/2} e^{-\frac{1}{k_B T} \frac{p_x^2 + p_y^2 + p_z^2}{2m}} = \frac{1}{2} k_B T \tag{24}$$

The same integral gives that the average energy in p_y and p_z are also both $\frac{1}{2} k_B T$. Adding these up,

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

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 $\frac{1}{2} k_B T$ of energy.

It may be worth building a little intuition for the size of $k_B T$. At room temperature $k_B T =$ 25 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, 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 components:

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

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 an independent excitation of the system, like momentum, or vibration, or rotation, or normal modes on a string. Technically speaking, mode means "normal mode" as in 15c, which is an eigenvalue of the Hamiltonian for small displacements from equilibrium. A **quadratic mode** is one for which the energy depends on the square of the phase space coordinate. The word *available* in this thoerem is important too. Classically, all modes are available. Due to quantum mecahnics, however, modes always have a lowest energy ε that can possibly be in that mode. So if $k_B T < \varepsilon$ for a given mode, then the mode is not available and is not included in the equipartition thoerem. We'll see how this works in Section 5.1 below.

By the way, this theorem was proposed in 1859 by Maxwell, but without the words "quadratic" and "available". Without these words, it's sometimes called the "classical equipartition theorem". Most of the time the classical version is correct: almost all excitations are quadratic (momentum, vibrations, rotations), and almost all modes are usually available. However, to understand the theorem from the modern perspective, it is important to understand what happens with non-quadratic modes (next section), and how modes become not available (Section 5.1).

4.1 Non-quadratic modes

What is special about quadratic modes? In chemistry, where the equipartition theory was first understood, the possible excitations of molecules are either kinetic, vibrational, or rotational, all of which lead to quadartic dependence of the energy on the displacement (see Eq. (38) below, for example). 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

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

To determine the scaling of of Ω_N with E, we can rescale all the integration variables by $p_j \to Ep_j$ use $\delta(Ex) = \frac{1}{E}\delta(x)$ to pull all the E dependence of the integral, giving $E^{3N-1} \approx E^{3N}$. The remaining dimensionless integral just gives some number C. So we get

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

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.

So for example, if we have some quadaratic and some linear modes

$$E = \underbrace{\frac{1}{2m}p_x^2 + \frac{1}{2}mx^2 + \frac{1}{2}I\omega^2 + \cdots +}_{Nf_q \text{quadraticmodes}} \underbrace{cp_y + \cdots}_{Nf_\ell \text{linearmodes}} \tag{33}$$

Then

$$E = N\left[f_q\left(\frac{1}{2}k_B T\right) + f_\ell(k_B T)\right] \tag{34}$$

More generally, you can show (try it!) that if the energy scales like the coordinate to some power $\varepsilon(y) \sim y^{\frac{2}{f}}$ then $\Omega \sim E^{\frac{f}{2}}$ and we would get a contribution $E = \frac{f}{2}k_B T$ to the total energy.

Suppose we want to know the probability of finding one mode excited with energy ε. Does this probability depend on whether the mode is quadratic or linear? The probability is given by the number of configurations with the other particles having $E - \varepsilon$ energy normalized to the total number of configurations. We have found that or quadratic modes $\Omega \sim E^{\frac{Nf_q}{2}}$ and $E = \frac{f_q}{2}Nk_B T$ and for linear modes $\Omega \sim E^{Nf_\ell}$ and $E = f_\ell N k_B T$. So, let us write the general case as $\Omega \sim E^{\frac{1}{2}fN}$ with $E = \frac{f}{2}Nk_B T$, with $f = 1$ for quadratic modes and $f = 2$ for linear modes. Then, the probability we are after is

$$P(\varepsilon) = \frac{\Omega_{N-1}(E-\varepsilon)}{\Omega_N(E)} = \frac{(E-\varepsilon)^{\frac{fN}{2}}}{E^{\frac{fN}{2}}} = \left(1 - \frac{\varepsilon}{E}\right)^{\frac{fN}{2}} = \left(1 - \frac{\varepsilon}{\frac{fN}{2}Nk_B T}\right)^{\frac{fN}{2}} \tag{35}$$

where $E = \frac{fN}{2}Nk_B T$ was used. Now we take $N \to \infty$ recovering an exponential:

$$P(\varepsilon) \propto \exp\left(-\frac{\varepsilon}{k_B T}\right) \tag{36}$$

Note that the dependence on f has dropped out: the probability of finding a mode with energy ε is independent of what type of mode it is (i.e. whether it's a quadratic mode or a linear mode, vibrational or rotation, or any microscopic properties). 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 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.

More physically, equipartition implies that nature tends to a situation where no object contains more than its fair share of energy. 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 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 (so it is at rest) and the remaining energy is in the air. 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:

$$C_V = \left(\frac{\partial E}{\partial T}\right)_V \tag{37}$$

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 $E = \frac{3}{2}Nk_B T$ so $C_V = \frac{3}{2}Nk_B$. 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
$\frac{1}{Nk_B}C_V$	1.5	1.5	2.54	2.49	2.45	2.57	3.39	3.42	3.40	4.92

Table 1. 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 Molecules

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?

The typical picture of a molecule is of big heavy, slowly moving nuclei setting up a potential for the electrons on relatively long timescales, and then the electrons adjusting and adapting their electron cloud based on the position of the nuclei. That is, we factorize the problem, first figuring out what is going on with the nuclei, and then the electrons. This factorization approximation is called the Born-Oppenheimer approximation, and provides an excellent description of molecular dynamics. Additionally, we treat the nuclei like points, with no internal structure. This is justified because the energy required to excite nuclei is typically $E_{\text{nuc}} \sim$ MeV. This is one million times higher than typical molecular excitation energy scales $E_{\text{rot/vib}} \sim \text{eV}$ as will now see.

Once we have approximated the moleucle as pointlike atoms, we can then consider how these atoms can be excited. The molecule can either be excited to rotate along one of 3 possible axes, or it can be excited to vibrate in various ways. During rotations all the distances betwen atoms are kept fixed, while during vibrations, at least one interatomic distance changes. Any possible small displacement of the atoms in the molcule can be decomposed into a basis of translations, rotations and vibrations. Both rotations and vibrations are periodic oscillations. To an excellent approximation, we can figure out the energies of these various excitations modes by first determining the classical oscillation frequences, ν_i and then quantizing each mode as a rotor or simple harmonic oscillator. Either way, the energies become quantized in units of $E_i = h\nu_i$.

To be concrete, let's begin 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. To specify the state of the molecule classically we need to give the positions and velocities of two H atoms, 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 $F = -kd$ with $d = |\vec{x} - \vec{x}_0|$ the distance from equilibrium. The energy of a spring is $E = \frac{1}{2}\mu\dot{d}^2 + \frac{1}{2}kd^2$, where $\mu = \frac{m_1 m_2}{m_1 + m_2}$ is the reduced mass. The energy is part kinetic energy and part potential from stretching the spring. Both of the energies are quadratic in the variable, so we expect $\frac{1}{2}k_B T$ of energy for each by the equipartition theorem; equivalently, we sometimes say that the vibrational mode has $k_B T$ 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

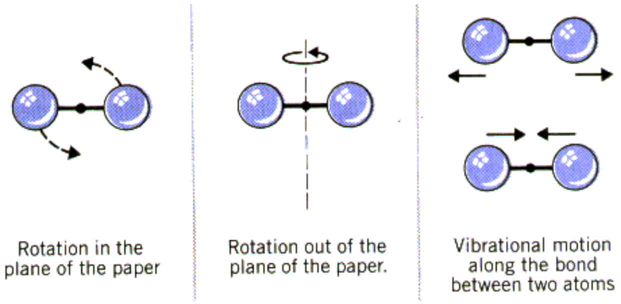

Rotation in the plane of the paper
Rotation out of the plane of the paper.
Vibrational motion along the bond between two atoms

Figure 1. Two rotational modes and one vibrational mode of a diatomic molecule.

Note that treating the atoms like points there is no rotation along the axis between the nuclei; Such a rotation leaves the state of the molecule unchanged. To have a thrid rotation axis, the molecule must be non-linear (like H_2O).

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. In summary, the energy for a diatomic molecule is

$$E=\frac{1}{2m}p_x^2+\frac{1}{2m}p_y^2+\frac{1}{2m}p_z^2+\frac{1}{2}\mu\dot{d}^2+\frac{1}{2}kd^2+\frac{1}{2}I\omega_1^2+\frac{1}{2}I\omega_2^2 \tag{38}$$

This form should clarify what we mean by "mode". In general the energy depends on the various phase-space coordinates. Writing the energy in a basis of orthogonal directions of displacement from equilibrium (the normal modes), makes it a sum of terms like in Eq. (38). Each term represents a mode.

From Eq. (38) and the equipartition theorem, we conclude that the total heat capacity for a diatomic molecule is

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

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 $(1+6)\times\frac{1}{2}Nk_B$ total.

We conclude that $C_V\approx 3.5Nk_B$ for a diatomic molecule. Looking at table 1 we see the data shows $C_V=3.39k_BT$ for Br_2, in pretty good agreement with this prediction. However, the data also lists that $C_V=2.45k_BT$ for H_2. So something is not right. The conflict between the heat capacity of H_2 and that predicted by the (classical) equipartition theorem was first appreciated by Maxwell in 1875.

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. After a straightforward quantum mecahnics calculation, one finds

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

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

It's in principle possible, to compute the energies for vibrational and rotational modes using quantum mechanics, at least numerically. But it is easy 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 $\tilde{\nu}_{\text{vib}}=4342\,\mathrm{cm}^{-1}$ corresponding to the vibrational mode and at $\tilde{\nu}_{\text{rot}}=60\,\mathrm{cm}^{-1}$ 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

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

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

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

So room temperature, $298K$ corresponds to 25 meV or $4\times 10^{-21}J$. Also,

$$\frac{k_B}{hc} = \frac{0.69\,\mathrm{cm}^{-1}}{K} \tag{45}$$

Thus for H_2

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

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

Thus at room temperature, rotational modes of H_2 are excited but not vibrational ones. In other words, the vibrational modes are not availble at room temperature so the equipartition thoerem does not apply to them. Using this, we can refine our prediction. With the the 3 center-of-mass momenta degrees of freedom, plus 2 rotational ones we get

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

now in great agreement with measured value of 2.45 in Table 1. This result is one of the most accessible tests of quantum mechanics: classical mechancis predicts $C_V = \frac{7}{2}Nk_B$ while quantum mechanics predicts $C_V = \frac{5}{2}Nk_B$.

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

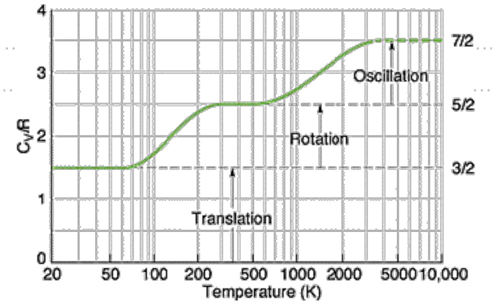


Figure 2. Cartoon of the heat capacity of H_2 as a function of temperature

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

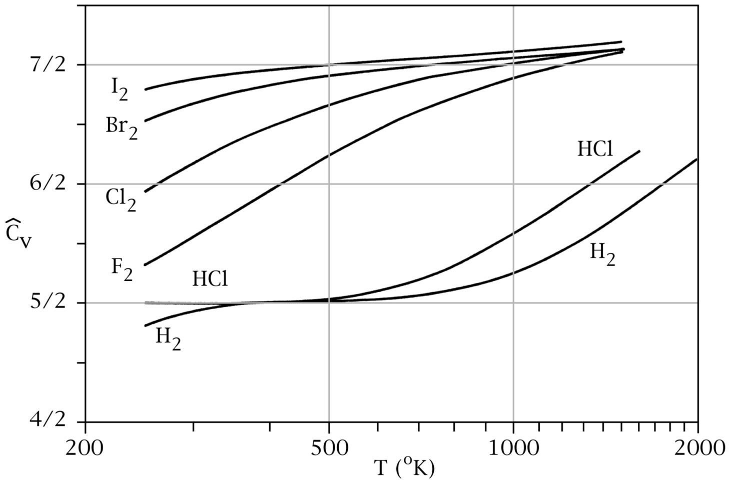


Figure 3. Heat capacities for various molecules

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 $1000\,K$ while the lighter ones, hydrogen and hydrogen-chloride, only excite 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?

At room temperature, a good approximation is that for most molecules vibrational modes cannot be excited, but rotational ones can. Monatomic molecules (like He) have no rotational modes, linear molecules (like H_2) have two, and non-linear molecules (like $\mathrm{H}_2\mathrm{O}$) have 3 rotational modes. Thus, including the three kinetic degrees of freedom, the rule of thumb is

$$\frac{C_V}{Nk_B} = \frac{3}{2} \quad \text{(monatomic)}, \qquad \frac{C_V}{Nk_B} = \frac{5}{2} \quad \text{(diatomic/linear)}, \qquad \frac{C_V}{Nk_B} = 3 \quad \text{(nonlinear)} \tag{49}$$

For air, for example, which is mostly N_2 and O_2, the experimental value is 2.63, in pretty good agreement with this prediction. More accurate predictions require details of the vibrational spectrum. We'll come back to the vibrational and rotational spectra, computing the heat capacity precisely and quantitatively comparing to data, in Lecture 7 in the context of the canonical ensemble, and again Lecture 10 with quantum statistical mechanics.

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 $c = 3N_Ak_B$ = constant. When using moles we also use

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

"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

Table 2. 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

$$S \equiv \frac{\Delta E}{m\Delta T} = \frac{C_V}{m} \tag{50}$$

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

6 Principle of maximum entropy

In Section 2 we defined the entropy as

$$S = k_B \ln\Omega \tag{51}$$

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 $i = 1...m$. 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 pick 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

$$\sum_{i=1}^{m} n_i = N \tag{52}$$

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 ${}_N C_n = \binom{N}{n} = \frac{N!}{n!(N-n)!}$ 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 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

$$\Omega = \frac{N!}{n_1! \cdots n_m!} \tag{53}$$

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:

$$\ln\Omega \sim N\ln N - N - \sum_{i=1}^{m}(n_i\ln n_i - n_i) = N\ln N - \sum_{i=1}^{m} n_i\ln n_i = -N\sum_{i=1}^{m}\frac{n_i}{N}\ln\frac{n_i}{N} \tag{54}$$

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

$$\boxed{\ln\Omega = -N\sum_i f_i\ln f_i} \tag{55}$$

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. (55) looks nearly identical to Boltzmann's quantity H from his H theorem

$$H = -\sum P_i\ln P_i \tag{56}$$

we just have to identify $H = \frac{\ln\Omega}{N}$. Since $S = k_B\ln\Omega$ we can also identify Boltzmann'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 detail 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 powerful mathematical tool. 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

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

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. (57) through by N gives

$$\frac{\ln\Omega}{N} = -\sum_{i=1}^{m} f_i\ln f_i - \alpha\Big(\sum f_i - 1\Big) \tag{58}$$

Taking the derivative with respect to f_i then gives

$$\frac{\partial}{\partial f_i}\frac{\ln\Omega}{N} = -(1+\ln f_i) - \alpha \tag{59}$$

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

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

Varying Eq. (58) with respect to α gives

$$1 = \sum_{i=1}^{m} f_i = m e^{-\alpha-1} \tag{61}$$

So that

$$\alpha = \ln m - 1 \tag{62}$$

and therefore

$$f_i = \frac{1}{m} \tag{63}$$

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. (63) is a 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

$$\sum f_i \varepsilon_i = \bar{\varepsilon} = \frac{E}{N} \tag{64}$$

Now we want to maximize lnΩ subject to the constraints in Eqs. (52) and (64). We introduce two Lagrange multipliers for this case, giving

$$\frac{1}{N} \ln \Omega = -\sum_{i=1}^{m} f_i \ln f_i - \alpha \left(\sum_i f_i - 1 \right) - \beta \left(\sum_i f_i \varepsilon_i - \bar{\varepsilon} \right) \tag{65}$$

Differentiating with respect to f_i gives

$$-\ln f_i - 1 - \alpha - \beta \epsilon_i = 0 \tag{66}$$

So that

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

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. Differentiating Eq. (65) with respect to α gives the total-number constraint

$$1 = \sum_i f_i = e^{-1-\alpha} \sum_i e^{-\beta \varepsilon_i} \tag{68}$$

It is handy at this point to define

$$Z \equiv \sum_i e^{-\beta \varepsilon_i} \tag{69}$$

Thus the α constraint implies

$$Z = e^{1+\alpha} \tag{70}$$

In terms of Z, Eq. (67) becomes

$$f_i = \frac{1}{Z} e^{-\beta \varepsilon_i} \tag{71}$$

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 particle with energy ε is

$$\boxed{P(\varepsilon) = \frac{1}{Z} e^{-\beta \varepsilon}} \tag{72}$$

The variations around this probability are very small, scaling like $\frac{1}{\sqrt{N}}$ by the law of large numbers. This is 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. (65) with respect to β gives

$$\bar{\varepsilon} = \sum_i f_i \varepsilon_i = \frac{1}{Z} \sum_i \varepsilon_i e^{-\beta\varepsilon_i} \tag{73}$$

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. (65) is that (after putting back $\bar{\varepsilon} = NE$)

$$\beta = \frac{\partial \ln\Omega}{\partial E} \tag{74}$$

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.

6.3 Preview

While on the subject of maxent, it is convenient to preview some connections that will be clearer a little later on. You can just skim this discussion but may want to come back to it later.

Since β has the interpretation as (inverse) temperature, you might naturally ask what is the interpretation of α? Just like Eq. (74) we find

$$\alpha = \frac{\partial \ln\Omega}{\partial N} \tag{75}$$

This quantity α is related to the **chemical potential $\boldsymbol{\mu}$** as $\alpha = -\frac{\mu}{k_B T}$. The chemical potential is a quantity, like temperature, that is constant among systems in equilibrium. A third is pressure, $P \propto \frac{\partial \ln\Omega}{\partial V}$. 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

$$H[P] = -\int_{-\infty}^{\infty} dx\, P(x) \ln P(x) \tag{76}$$

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: $P(x) = Ze^{-\beta x}$ 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?[2] Similarly, one finds the Poisson distribution and the binomial distribution as those that maximize $H[P]$ subject to appropriate constraints.[3] Thus, maxent gives a satisfying way to understand the universality of various statistical distributions.

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

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

You might recognize $H[P]$ as Boltzmann's H function from Lecture 3. Boltzmann's H theorem says that H always increases with time. We'll come back to H, and see that it directly proportional to entropy, in Lecture 6.

7 Summary

This lecture defined temperature using statistical mechanics. The main points are

- If two systems with numbers of microstates $\Omega_1(N_1,V_1,E_1)$ and $\Omega_2(N_2,V_2,E_2)$ are allowed to exchange energy keeping to total energy $E=E_1+E_2$ fixed, then in equilibrium

$$\left.\frac{\partial \ln \Omega_1(N_1,V_1,E)}{\partial E}\right|_{E=\langle E_1\rangle}=\left.\frac{\partial \ln \Omega_2(N_2,V_2,E)}{\partial E}\right|_{E=\langle E_2\rangle} \tag{77}$$

- Because of these derivatives are equal, we define $\beta=\frac{\partial \ln \Omega(E)}{\partial E}$. β is the same for any two systems in equilibrium and **temperature** is defined as $T=\frac{1}{k_B\beta}$ with k_B Boltzmann's constant.
- At temperature T, the probability of finding energy ε in *anything* is $P(\varepsilon)=\frac{1}{Z}e^{-\frac{\varepsilon}{k_BT}}$, where Z is defined so the sum of probabilities is 1. This very general rule is the most important result of statistical mechanics.
- A special case is the Maxwell-Boltzmann distribution: $P(\vec{p})=\left(\frac{1}{2\pi m k_BT}\right)^{3/2}e^{-\frac{1}{k_BT}\frac{\vec{p}^2}{2m}}$. This gives the distribution of momenta/velocities of any system in equilibrium.
- **Entropy** is defined in terms of Ω as $S=k_B\ln\Omega$, so then $\frac{1}{T}=\frac{\partial S(N,V,E)}{\partial E}$.
- Entropy of a monatomic ideal gas is given by the Sakur-Tetrode equation

$$S=Nk_B\left[\ln\frac{V}{N}+\frac{3}{2}\ln\left(\frac{4\pi mE}{3Nh^2}\right)+\frac{5}{2}\right] \tag{78}$$

- In equilibrium, energy is distributed evenly among all available modes. Each mode gets $\frac{1}{2}k_BT$ of energy if the energy depends quadratically on the degree of freedom for the mode, or k_BT of energy if the dependence is linear. The quadratic case is the one relevant for chemistry. In that case, the **equipartition theorem** says that each mode gets $\frac{1}{2}k_BT$ of energy.
- Heat capacity at constant volume is defined as $C_V=\left(\frac{\partial E}{\partial T}\right)_V$.
- In gases of molecules, there are translational, rotational, and vibrational modes. Typically, at room temperature, only rotational and translational modes can be excited due to the $e^{-\varepsilon/k_BT}$ factor, since typically $\varepsilon_{\text{rot}}\lesssim k_BT\lesssim\varepsilon_{\text{vib}}$.
- The law of Dulong and Petite says that the heat capacity per unit mass of most metals is constant. It follows from the equipartition theorem.
- The **principle of maximum entropy** is a general statistical trick for determining probability distributions using constraints: you maximize the entropy function subject to the constraints. Constraining only the total number, maxent predicts P is constant. Constraining the total number and the mean predicts P is an exponential: $P=ae^{-xa}$. Knowing the temperature means knowing the mean energy, thus, maxent predicts the Boltzmann factor $P\sim e^{-\beta\varepsilon}$. The simplicity of maxent, and how little it assumes, helps explain why the Boltzmann factor is universal.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 5: Thermodynamics

1 Introduction

Thermodynamics is the study of heat and temperature. One thing that makes thermodynamics hard (and generally unpopular) is all the damn variables. Everything is related and it's often tough to keep straight what is an independent and what is a dependent variable. We will do our best to write the dependent variables explicitly whenever possible. Another thing that makes thermodynamics hard is that we give new definitions to common words. Words like system, energy, work, heat, temperature, etc. have precise meanings in physics that do not always agree with their everyday meanings. For example, we defined temperature in terms of the number of states, not what you measure with a thermometer. So learning thermodynamics means unlearning some of your associations.

Let's start by defining some concepts we will use a lot. We define a **system** as the thing we are observing or calculating properties of. The system is separated from the **surroundings** by a barrier. The barrier could be **open**, so matter and energy can pass through, or **closed**.

For example, when we talk about doing work on a gas by moving a piston, the piston is usually part of the surroundings and the gas is the system. We could alternatively treat the whole gas/piston complex as the system, and our arm that pushes the piston as the surroundings. Or we could attach the piston to a weight, put a box around the whole thing and call the everything together the system. In such a case, where the surroundings have no influence on the system, we say the system is **isolated**.

Energy we take to mean the **internal energy** of the system, stored in chemical bonds, or kinetic motion of the atoms. Generally, we won't include in energy the gravitational energy from the earth or nuclear energy within the atoms. We use the symbol E for energy, although sometimes U is used for internal energy. Quantum mechanically, energy is the expectation value of the Hamiltonian; classically it is the value of the (classical) Hamiltonian. I like to use ε for energies for individual molecules (microscopic energy) and E for macroscopic energy. An important constraint on energy is

- **The first law of thermodynamics**: the total energy is conserved in time; energy is neither created nor destroyed.

This law follows from Noether's theorem: energy is conserved in any system whose Hamiltonian is independent of time (you would prove it in an advanced classical mechanics or field theory class).

One kind of energy is in the motion of molecules. This is called kinetic energy, $\frac{1}{2}m\vec{v}^2$. Another kind of energy is vibrational and rotational energy. When we have a large group of molecules all moving/vibrating/rotating incoherently we call the energy **thermal energy**. A kind of non-thermal energy is potential energy, such as from lifting a rock over your head (gravitational potential energy), or from separating two opposite charges, like in a battery (chemical potential energy). These types of potential energy are coherent and can be easily converted into other forms of energy. One goal of this lecture is to understand limits on how to convert thermal energy into other energy forms.

Work is an amount of energy we put in to change a system by organized motion. For example, lifting a weight up a height h against gravity requires work. The work required is $W = F h = m g h$. In fact, we can define work as follows: a process that does work is one that could be used to lift a weight by some height. Of course we will rarely be using work for lifting weights, but we *could*. For example, the energy in your phone battery *could* have been used to lift a weight. The work done to spin a wheel underwater *could* have been used to lift a weight. One theme for this Lecture is how thermal energy can be turned into work, since it is not immediately clear how incoherent random motion can be coerced into coherent organized motion.

Heat is the transfer of energy from a temperature difference between two systems. Heat is a transfer of thermal energy. Energy transfer by heat involves incoherent chaotic molecular motion. Thus, one question we will try to answer in this lecture is how heat can be used to do work. A machine that does so is called a **heat engine**.

Importantly, the distinction between work and heat is about the surroundings, not the system itself. If we attach a weight to a pinwheel in a pot of water, as the weight falls, the pinwheel turns. This does work on the water, increasing its molecular motion and increasing its temperature. We did work on the water. We could have induced the same change on the system (the pot of water) by heating it over a fire. Thus these two processes leave the system the same. The distinction between work and heat is in the surroundings (the weight or the fire).

Another useful concept is a **quasistatic** process. This is one where things are changed slowly enough that equilibrium holds at all times. For example, if we allow a gas to expand by easing the pressure on the walls, that would be quasistatic. It doesn't even have to be that slow, just slower than the rate by which the molecules in the gas are hitting the walls (around the speed of sound), so that equilibrium is re-established continually as the gas is expanded. If we abruptly remove the wall of a gas, like popping a balloon, that would not be quasistatic. Or if we take two gases at different temperatures and put them in contact, as heat flows from one to another equilibrium does not hold, so this heat transfer is not quasistatic. If there is a time in the process when the system is clearly not in equilibrium, it's not quasistatic, otherwise it is.

Finally, we can talk about **reversibility**. A process is reversible if it can be brought back to a previous state exactly, with *no net change in the system or surroundings*. Compressing a gas quasi-statically with a piston is reversible: allow the gas to push back on the piston and it will end up exactly where it started (we have to check this of course, and we will). Letting a gas expand abruptly is not reversible – it takes work to compress it back to where it was, and the work takes energy from the surroundings. You can have situations that are quasistatic but not reversible. For example, slowly heating up a system by adding infinitesimal amounts of heat. Or, more simply, a ball rolling on the ground with friction is quasistatic but not reversible. Generally, any reversible thermodynamic process must be quasistatic.

The **thermodynamic limit** means taking N large and V large holding the number density $n = \frac{N}{V}$ fixed. More precisely, we Taylor expand in $\frac{1}{N}$ after setting $\frac{1}{V} = n\frac{1}{N}$ and keep only the leading terms. The thermodynamic limit lets us use the central limit theorem: systems in equilibrium are characterized by their mean values, which are the same as the most probable values, and fluctuations around these means scale like $\frac{1}{\sqrt{N}}$.

For completeness, we list the 4 laws of thermodynamics here:

0. If two systems are in equilibrium with a third system, they are in equilibrium with each other.

1. Energy is conserved in time. Energy is neither created nor destroyed.

2. The total entropy (system+surroundings) never decreases: $\Delta S_{\text{total}} \geqslant 0$.

3. Entropy is finite as $T \to 0$.

The first and second laws are sometimes summarized as "you can't win" and "you can't break even."

The 0th law we already proved by showing that $\beta = \frac{1}{k_B T} = \frac{\partial \ln \Omega}{\partial E}$ is the same for any two systems in equilibrium. The 1st law follows from classical mechanics (Noether's theorem, as mentioned above). The 2$^{\text{nd}}$ law is enormously important and will be discussed in this lecture and the next.

The 3rd law was orignally formulated (by Nerst around 1910) as that entropy of a crystal goes to zero as $T \to 0$. Having a crystal, rather than a gas, is necessary to fix the positions of the molecules beacuse otherwise there is the configurational entropy $\ln \Omega_q \sim N \ln V$, even at $T = 0$. However, even crystals can have nonzero entropy at $T = 0$ in quantum mechanics if they have degenerate ground states, since then at $T = 0$, $S = k_B \ln N_{\text{ground}}$. Thus the modern version of the third law is that entropy goes to a constant as $T \to 0$, which holds for crystals, gases, classical and quantum systems. The constant is not important physically, and in practice we are free to add an offset to the entropy (as we do with energy), using that only changes in entropy matter. There is still content in the 3rd law when doing this, because the entropy is nonzero at finite temperature, so it should scale as a positive power of T as $T \to 0$. We'll come back to the 3rd law in Lectures 10 and onward, after we discuss quantum statistical mechanics, since at low temperature quantum effects become important.

2 Heat

Say we have two systems with different temperatures T_1 and T_2, energies E_1 and E_2 and numbers of particles N_1 and N_2. The number of microstates of each is $\Omega_1(N_1, E_1)$ and $\Omega_2(N_2, E_2)$ and their entropies are $S_1 = k_B \ln \Omega_1$ and $S_2 = k_B \ln \Omega_2$. When the systems are isolated, they are each in their own equilibrium. So each one separately has $\frac{1}{T_1} = \frac{\partial S_1}{\partial E_1}$ and $\frac{1}{T_2} = \frac{\partial S_1}{\partial E_2}$.

Let us then put these systems into thermal contact. Thermal contact means that the two can exchange energy, for example, by vibrating the molecules in a thermally-conducting barrier between them. We assume they cannot exchange particles. Thus E_1 and E_2 can change, as can T_1 and T_2 but N_1 and N_2 are fixed. At the moment when we put the two systems in contact, the total number of microstates is $\Omega_{\text{initial}} = \Omega_1(E_1)\Omega_2(E_2)$. The two systems will then exchange energy until equilibrium is reached, where $T_1 = T_2$. As equilibrium has the maximum value of Ω or equivalently maximal entropy $S = k_B \ln \Omega$ we must have $\Omega_{\text{final}} \geqslant \Omega_{\text{initial}}$. So entropy of the whole system increases. This is the second law of thermodynamics. Note that we only know the total entropy goes up

$$\Delta S_1 + \Delta S_2 \geqslant 0 \tag{1}$$

It is certainly possible for the entropy of one side to go down.

The heat Q is defined as the amount by which the energy changes through thermal contact

$$Q = -\Delta E_1 = \Delta E_2 \tag{2}$$

When a positive amount of heat $Q > 0$ leaves system 1, its energy goes down. When a negative amount of heat $Q < 0$ leaves system 1, a positive amount of heat $-Q > 0$ enters. So Q can take either sign and we just have to keep track of which sign we want by thinking about the physical situation.

Note that heat is *not* a property of a system in equilibrium. Heat is a non-equilibrium quantity related to how energy is transferred in the approach to equilibrium.

When energy enters a system, through heat or work, usually the temperature goes up too. But this is not guaranteed. Whether the temperature goes up is indicated by the heat capacity $C_V = \frac{\partial E}{\partial T}\Big|_V$. Most systems have positive heat capacities, but some have negative heat capacities. For the total energy to go down when the temperature goes up, so $C_V < 0$, we generally need some kind of long-range interactions, like gravity. For example when stars get hotter, they expand and their gravitational potential energy goes down by more than their kinetic energy goes up. Thus their total energy goes down and $C_V < 0$. Stars, black holes, galaxies, etc, all have $C_V < 0$. We'll explore gravitational systems in more detail later on.

Now, suppose we have an infinitesimal heat transfer between two systems. We write dQ for such a small heat transfer (note that dQ just means a small amount of heat; it is not the difference between two heats). Then the change in entropy for a small dQ is

$$dS_{\text{heat transfer}} = \left(\frac{\partial S}{\partial E}\right) dE = \left(\frac{1}{T}\right) dQ = \frac{dQ}{T} \tag{3}$$

So to find the net change in entropy due to a heat transfer we can integrate over dQ, presuming we know the functional dependence of the temperature on the amount of heat transferred. Note that the entropy can increase in ways other than heat transfer, for example in the free expansion of a gas (see Eq (30) below).

2.1 Heat reservoirs

A heat reservoir is a system that does not change temperature or volume when heat is put in. For example, a bowl of soup acts like a heat reservoir for a fly that falls into it. Air in a room acts like a heat reservoir for a bowl of soup. Any system acts like a reservoir if the heat transferred is much less than the energy of the system $Q \ll E_{\text{res}}$.

Another way to define a reservoir is that $|\Delta T_{\text{res}}| \ll T_{\text{res}}$. Since ΔT_{res} is assumed small, we can write $\Delta T_{\text{res}} = \frac{\partial T}{\partial E} \Delta E = \frac{\partial T}{\partial E} Q$ where Q is the heat transferred as equilibrium is approached. So the condition for a heat reservoir can be written as

$$\left| \frac{\partial T}{\partial E} Q \right| \ll T \tag{4}$$

For a heat reservoir, the temperature and volume doesn't change throughout the heat transfer. So we can trivially integrate Eq. (3):

$$\Delta S_{\text{heat reservoir}} = \frac{Q}{T} \tag{5}$$

Because $S = S(E, V, N)$ and V and N are unchanging, this is the *total* change in entropy of the reservoir.

2.2 Heat + work

In general, we can transfer energy through heat and work at the same time. The net energy change is

$$\Delta E = Q + W \tag{6}$$

or, differentially,

$$dE = dQ + dW \tag{7}$$

Note that there is an implicit sign convention here: Q is the heat put *into* the system and W is the work done *on* the system. If we defined W as the work done *by* the system, then we would have $dE = dQ - dW$.

You can write the equation however you like as long as you are consistent with your definitions for W and Q. For example, say we compress a gas that is in thermal contact with a heat reservoir. As we push down on the gas, we do work on it, so its energy goes up. This will heat up the gas. As it heats up, it will go out of equilibrium with the reservoir so heat will start flowing out, lowering the internal energy of the gas. So we can write

$$\Delta E = W - Q \tag{8}$$

In this case we have defined W as the work done *on* the system, so $W > 0$, and Q as the heat flowing *out* of the system. It's always a good habit to write down explicitly what your sign convention is for W and Q. The energy change of the reservoir is $\Delta E_{\text{res}} = Q$, by energy conservation (1st law of thermodynamics).

2.3 Clausius entropy

Historically, the first definition of entropy was due to Rudolf Clausius in 1865. Clausius did not have a microscopic definition of entropy, and only specified a way to calculate entropy differences. His formula was that to compute the entropy change from a state A to a state B, find a reversible path connecting the two and then compute

$$\Delta S_{\text{system}} = S_A - S_B = \int_{\text{rev.}} \frac{dQ_{\text{in}}}{T} \tag{9}$$

Here, dQ_{in} is the infinitesimal heat absorbed along the path from A to B. Clausius was implicitly using that there are two sources of entropy increase: irreversibility and heat transfer. If you make the path reversible, then all you need to know is how much heat comes in as a function of temperature and you can compute the entropy change.

Note that entropy is a function of state: it doesn't matter how you get from A to B, the entropy change will be the same. For example, a gas could expand freely, doubling its volume. This is irreversible. However, we can find a reversible path to double the volume, by letting the gas expand slowly, doing work against a piston. To keep it from cooling as it does work, we can connect it to a heat bath. Then the entropy change in doubling the volume is given by $\Delta S_{\text{system}} = \frac{Q}{T}$, with Q the heat drawn in from the bath to do the expansion work. Although the surroundings change differently in the irreversible free expansion or the expansion with the heat bath, the *system* changes the same way, and so its entropy change is the same. More details of this example, and other examples of reversible and irreversible processes, are in Section 3.3.

The connection between reversibility and entropy was understood first by Clausius, as was the second law of thermodynamics, $\Delta S_{\text{tot}} \geqslant 0$. Clausius showed that the second law implies that $\Delta S_{\text{tot}} = 0$ for any reversible process. To see that, note that if states A and B (including system and surroundings) can be connected reversibly, then $S_B^{\text{tot}} - S_A^{\text{tot}} \geqslant 0$. For the reverse process $S_A^{\text{tot}} - S_B^{\text{tot}} \geqslant 0$. Thus $S_A^{\text{tot}} = S_B^{\text{tot}}$ and so $\Delta S_{\text{tot}} = 0$ for any reversible process.

Since $dS = \frac{1}{T} dE$ (at fixed V and N), it might seem like any change in energy, from heat *or* work would result in a change in entropy. Clausius argued that this is not true: only heat flow matters. To see this, consider doing work on a system, for example, by compressing it. This work can be done through a non-thermal mechanism, for example, a weight. Then $\Delta S_{\text{surroundings}} = 0$ in the compression. If the work is done reversibly, then since $\Delta S_{\text{tot}} = 0$ for any reversible process we must also have $\Delta S_{\text{system}} = 0$. Therefore, along a reversible path, work does not cause an entropy increase. This confirms Eq. (9) – entropy change is due to heat flow.

Eq. (9) is important because it lets us compute the entropy change without knowing anything about the microscopic composition of our system. We just need to find a reversible path between two states, and determine how much heat flows in. There is no counting of states involved. For example, we can apply Eq. (9) to a steam engine or an internal combustion engine without knowing anything about water or chemistry. In other words, Eq. (9) is useful in thermodynamics, as in the 19th century, regardless of whether there is a microscopic statistical mechanics description.a

3 Ideal gases

Much of thermodynamics will be about relations such as $\frac{1}{T} = \left.\frac{\partial S}{\partial E}\right|_{N,V}$ that hold for any substance in equilibrium in the thermodynamic limit. Ideal gases, and monatomic ideal gases in particular, provide very useful systems for which we will be able to check many thermodynamic relations explicitly. Recall that the molecules of ideal gases are assumed to occupy no volume and have no long-range interactions.

3.1 Pressure

Pressure is the force per unit area on the walls of a container: $P = \frac{|\vec{F}_{\text{wall}}|}{A}$. This force depends on the average velocities of the molecules. Since the average velocity is the same in every direction, by the equipartition theorem, the force is the same in every direction. That's why pressure is a scalar, not a vector quantity: it is a number characterizing the force per unit area in any direction on any wall of any container.

For an ideal gas, we can work out a relation between pressure, volume and temperature using Maxwell's kinetic theory of gases. In a time Δt, a certain number N_{hit} of molecules will hit the wall. Then the net force is

$$\vec{F}_{\text{wall}} = N_{\text{hit}} m\, \vec{a} = N_{\text{hit}} \frac{\Delta \vec{p}}{\Delta t} \tag{10}$$

where $\Delta \vec{p}$ is the impulse (change in momentum). If a particle hits the wall in the x direction then only the x component of its velocity changes, by $\Delta v_x = -2 v_x$, so the impulse is $|\Delta \vec{p}| = 2 m v_x$. Then the number of particles which hit the wall in time Δt are the ones within a distance $\Delta x = v_x \Delta t$ of the wall, within the area A, going *towards* the wall:

$$N_{\text{hit}} = \frac{1}{2} n A \Delta x = \frac{1}{2} n A v_x \Delta t \tag{11}$$

where $n = \frac{N}{V}$ is the number density. The factor of $\frac{1}{2}$ is because the ones going away from the wall not hit in the time Δt. Thus,

$$|\vec{F}_{\text{wall}}| = \left(\frac{1}{2} n A v_x \Delta t\right) \frac{2 m v_x}{\Delta t} = m n v_x^2 A \tag{12}$$

Taking the average, we then get

$$P = \frac{|\vec{F}_{\text{wall}}|}{A} = m n \langle v_x^2 \rangle \tag{13}$$

Recalling that $\langle\frac{1}{2}mv_x^2\rangle = \frac{1}{2}k_BT$, as for any quadratic degree of freedom, we can then write

$$P = nm\langle v_x^2\rangle = \frac{N}{V}k_BT \tag{14}$$

so that

$$\boxed{PV = Nk_BT} \tag{15}$$

This is the **ideal gas law**. It is very general and holds for any ideal gas, not just a monatomic one.

We also sometimes write the ideal gas law in terms of the gas constant $R = k_BN_A = 8.3\frac{J}{K\cdot\text{mol}}$ in the form

$$PV = nRT \tag{16}$$

where in this context n is the number of moles (not the number density; it's unfortunate that we use the same letter for n for number density and number of moles, but the notation is standard).

Note that for the force in Eq. (10) to be constant, we must average over time. If we don't time average, then as $\Delta t\to 0$ all the quantities N_{hit}, Δp, F_{wall} and P would be wildly fluctuating functions of time. On the other hand, $\langle v_x^2\rangle$ refers to the average of v_x^2 over all the possible microstates (the ensemble average). Thus this derivation of the ideal gas law makes an implicit assumption of ergodicity, that the time average and the ensemble averarge are the same.

3.2 Heat capacity

Heat capacity characterizes how temperature changes when heat is put in. There are different types of heat capacity, depending on whether the volume is held fixed, in which case all the heat goes to raise the temperature, or whether the pressure is held fixed, in which case some of the heat goes into work in expanding the system. We give different symbols for the heat capacity in these different circumstances

$$C_V = \left(\frac{\partial Q}{\partial T}\right)_V, \quad C_P = \left(\frac{\partial Q}{\partial T}\right)_P \tag{17}$$

where the $|_V$ means volume is held fixed and $|_P$ means pressure is held fixed.

When the volume is held fixed, no work is done and all of the heat goes into internal energy of the gas. Then

$$C_V = \left(\frac{\partial E}{\partial T}\right)_V \tag{18}$$

This was the heat capacity we computed in the previous lecture for ideal gases. Recall that for most comon gases, vibrational modes cannot be excited at room temperature. Then $C_V = \frac{3}{2}Nk_B$ for a monatomic gas, $C_V = \frac{5}{2}Nk_B$ for a diatomic gas like H_2 or N_2 (or air, which is mostly N_2), and $C_V = 3Nk_B$ for water or other non-linear molecules. If there are f excitable degrees of freedom then $C_V = \frac{f}{2}Nk_B$ for a general molecule.

The total internal energy of a gas can be determined by integrating the heat capacity:

$$E = \int_0^T C_V(T')dT' \tag{19}$$

If C_V is temperature-independent (as it is for a classical ideal gas), then we have the simple relation between the total internal energy of a gas and its temperature:

$$E = C_VT = \frac{C_V}{Nk_B}PV \qquad \text{(ideal gas)} \tag{20}$$

where the ideal gas law has been used in the second step. Note that Eq. (20) implies

- The internal energy of a classical ideal gas is linearly proportional to the temperature
- The internal energy of a classical ideal gas only depends on P and V through the combination PV.

To be clear, for a monatomic ideal gas, we know exactly the internal energy $E = \frac{3}{2}PV$ but for a general ideal gas, we know only that $E = \frac{C_V}{Nk_B}PV$ for some C_V.

What is the difference between heating a gas at constant volume and constant pressure? If we heat it at constant volume all of the heat goes into increasing the internal energy of the gas. For heating at constant pressure, imagine a syringe where you seal the end and put a 5 kg weight on top. The weight compresses it until equilibrium is reached. The pressure due to the weight is $P_{\text{weight}} = \frac{mg}{A}$, with A the area of the pistion in the syringe. The total pressure is $P = P_{\text{weight}} + P_0$ with P_0 atmostpheric pressure. Then you heat it up. As the gas expands, it does work on its weight, so some of the energy goes into this work. The work the gas does to psuh the weight by a small amount dx is

$$dW = Fdx = F\frac{dV}{A} = PdV \tag{21}$$

Note that the work done involves the total pressure, not just the pressure from the weight.

To separate out the weight, we often use the term **gauge pressure**: $P_{\text{gauge}} = P - P_0$, which just means pressure with atmopsheric pressure subtracted. For example, when you pump up a tire or a ball, the pressure listed on the pump is the gauge pressure: when a ball totally deflated the pressure inside is atmospheric pressure P_0, not $P = 0$. The gauge pressure when deflated is $P_{\text{gauge}} = 0$.

If pressure is constant, then the change in the energy of the gas is the heat we put in minus the work it does:

$$dE = dQ - PdV \tag{22}$$

To compute C_P we need to express the quantities in this equation entirely in terms of T and P, so that we can take $\frac{\partial Q}{\partial T}\Big|_P$. To do so, we will use that $E = C_V T$ for an ideal gas[1], so that $dE = C_V dT$ By the ideal gas law, $V = \frac{N k_B T}{P}$ so at constant pressure $dV = \frac{N k_B}{P} dT$. Putting these two relations into Eq. (22) gives

$$C_V dT = dQ - N k_B dT \tag{23}$$

or equivalently,

$$(C_V + N k_B) dT = dQ \tag{24}$$

and therefore

$$C_P = \left(\frac{\partial Q}{\partial T}\right)_P = C_V + N k_B \tag{25}$$

C_P is always larger than C_V since it takes more heat to increase the temperature at constant pressure than at constant volume: the extra heat goes into work on expanding the gas against the pressure.

Recall that for monatomic gas the number of microstates was $\Omega \sim V^N \left(\frac{mE}{N}\right)^{\frac{3}{2}N}$ and the entropy $S = k_B \ln \Omega$ was given by the classical Sackur-Tetrode equation

$$S = N k_B \ln V + \frac{3}{2} N k_B \ln \frac{E}{N} + \text{const} \tag{26}$$

Then the temperature is

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_V = \frac{3}{2} N k_B \frac{1}{E} \tag{27}$$

so that $E = \frac{3}{2} N k_B T$. For a general ideal gas, $\Omega \sim V^N \left(\frac{mE}{N}\right)^{\frac{f}{2}N}$ with f the number of degrees of freedom. Then $E = C_V T$ with $C_V = \frac{Nf}{2} k_B$ (this follows from $\frac{1}{T} = k_B \frac{\partial \ln \Omega}{\partial E}$) and entropy has the form

$$S = N k_B \ln V + C_V \ln \frac{E}{N} + \text{const} \tag{28}$$

Using Eq. (25) we can equivalently write

$$S = (C_P - C_V) \ln V + C_V \ln \frac{E}{N} + \text{const} \tag{29}$$

Eqs. (28) and (29) hold for any ideal gas (not just a monatomic one).

[1] "But wait!" you say, that calculation was for constant volume, and now we are discussing constant pressure. In fact the relation between energy and temperature, $E = C_V T$, is an exact statement about any ideal gas in equilibrium (constant V, P, T, N), independently of how it got there or what we're going to do to the gas.

3.3 Compressing/expanding an ideal gas

We want to know how V, P, T, E and S change as we compress or expand a gas. There are in fact many ways to compress or expand a gas: we can pressurize it and increase its volume so the temperature is fixed, we can pressurize it and heat it at fixed volume, or we can heat it and let it expand with fixed pressure. There are other ways too. The trick to understanding how things change as a gas is compressed or expanded is to always express the quantities in terms of dependent variables that involve the thing being held fixed.

Free expansion means letting a gas expand on its own. If a gas expands freely from V_1 to V_2, its energy doesn't change (energy is conserved). So its temperature doesn't change either ($E = C_V T$). Remember, temperature is the kinetic energy of the molecules. Just because they can move in a larger volume, it doesn't mean they are slowing down. Since temperature doesn't change, using the ideal gas law $PV = Nk_B T$, the pressure goes down by $P_2 = P_1 \frac{V_1}{V_2}$. Also, from the Sackur-Tetrode equation,

$$\Delta S = Nk_B \ln \frac{V_2}{V_1} \tag{30}$$

Since $V_2 > V_1$ this says that entropy increases upon free expansion.

Quasistatic isothermal expansion means letting a gas expand at fixed temperature. To use the ideal gas law, or any of the thermodynamic relations we've derived so far using statistical mechanics, the system has to be in equilibrium at all times. So the expansion should be done quasistatically. This means that the expansion has to be done more slowly than the molecules are moving. As mentioned in the introduction, quasistatic processes are pretty generic. Most natural processes, motion of clouds for example, are quasistatic. An example of a non-quasistatic process is abruptly moving a piston that had been compressing a gas. When the expansion is quasistatic, all the work done on moving the piston can be employed, say for lifting a weight. If the piston were abruptly moved, the gas would expand and its pressure would drop. We would then have to do work to compress it again.

A helpful visualization for quasistatic expansion is to picture the expansion being done against a pile of sand sitting on top of a piston. We slowly let sand slide off the piston. As each grain slides off, the pressure decreases and the piston rises. Equilibrium is maintained at all times:

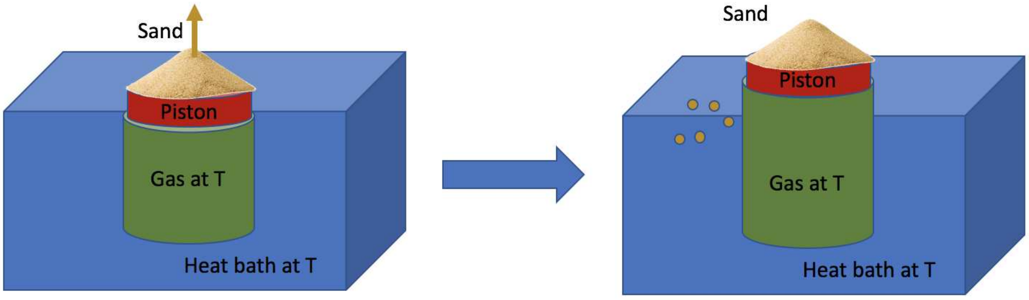


Figure 1. Isothermal expansion against a piston covered with sand

I like the sand picture too because there you see where the work is going – it's lifting the sand. We only take off a little bit to get the piston to go up and it doesn't cost any energy to let the sand fall off the piston. After the expansion, the rest of the sand is higher so work has clearly been done. If you have the piston without the sand it's a little confusing sometimes as to where the energy in the work has gone.

We already established in Eq. (21) that $dW = PdV$. In isothermal expansion $P = \frac{Nk_B T}{V}$ so

$$W = \int_{V_1}^{V_2} PdV = Nk_B T \int_{V_1}^{V_2} \frac{dV}{V} = Nk_B T \ln \frac{V_2}{V_1} \tag{31}$$

Where did the energy come from? Not from the gas, as we already established that its energy is unchanged if its temperature is unchanged. If we assume no friction, and there is no external source of energy, all the energy must have come from heat withdrawn from the reservoir. Thus

$$Q = N k_B T \ln \frac{V_2}{V_1} \tag{32}$$

Using Eq. (5) the entropy change in the reservoir is therefore

$$\Delta S_{\text{res}} = -\frac{Q}{T} = -N k_B \ln \frac{V_2}{V_1} \tag{33}$$

The entropy of the reservoir goes down, since heat is extracted.

Since T is fixed and the energy of the gas is fixed, the entropy change of the gas is the same as if it expanded freely

$$\Delta S_{\text{gas}} = N k_B \ln \frac{V_2}{V_1} \tag{34}$$

So we see that the entropy of the gas goes up by exactly the same amount that the entropy of the reservoir goes down by. The net entropy change of the whole system and surroundings is zero. The process is reversible.

Quasistatic adiabatic expansion. Adiabatic means without transferring heat. To achieve adiabaticity, we can simply insulate a system. Let us consider the quasistatic adiabatic expansion of a gas. Then because $\Delta S = \int \frac{1}{T} dQ$, the system's entropy doesn't change.

We often colloquially say "adiabatic" to mean "slow". Usually, physicists mean a quasistatic adiabatic process when they say "adiabatic". Indeed, while a quistatic process must be slow (slower than the average molecular speed), an adiabatic process does not have to be slow, it just has to be insulated. Free expansion of a gas is an adiabatic process that is not quasistatic. The expansion of a gas against an insulated piston is also adiabatic, and if done slow enough, can be quasistatic as well.

By adiabatic expansion, we usually mean an adiabatic quasistatic expansion, as the expansion against a piston:

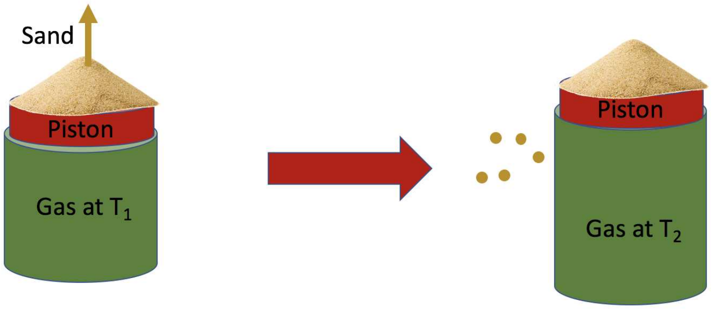


Figure 2. In adiabatic expansion there is no heat bath, so no heat transfer.

In adiabatic expansion, even though no heat flows, the temperature may still change. Consider quaistatic adiabatic expansion against a piston: the gas will do work on the piston as it expands, losing energy, so its temperature will go down. The work done by the gas is $dW = P dV$. To figure out by how much the energy of the gas changes, we use Eq. (20):

$$dE = \frac{C_V}{N k_B} d(PV) = \frac{C_V}{N k_B}(P dV + V dP) \tag{35}$$

By energy conservation $dE + dW = 0$ (since there is no heat involved), we then have

$$\frac{C_V}{N k_B}(P dV + V dP) + P dV = 0 \tag{36}$$

Recalling Eq. (25), $C_P = C_V + N k_B$, this simplifies to

$$C_P \frac{dV}{V} = -C_V \frac{dP}{P} \tag{37}$$

which integrates to $C_P \ln V = -C_V \ln P$ + const. It is conventional to define the **adiabatic index** γ as

$$\gamma = \frac{C_P}{C_V} \tag{38}$$

so that $\ln PV^\gamma =$ constant and therefore

$$PV^\gamma = \text{constant} \tag{39}$$

So that

$$P_1 V_1^\gamma = P_2 V_2^\gamma \tag{40}$$

This equation tells us how pressure and volume are related as we compress a gas adiabatically.

Equivalently, since $P = \frac{nRT}{V}$ for an ideal gas, $TV^{\gamma-1}$ is constant, as is $VT^{\frac{1}{\gamma-1}}$, $TP^{\frac{1-\gamma}{\gamma}}$ and $PT^{\frac{\gamma}{1-\gamma}}$ in an adiabatic compression/expansion process. These alternative formulas are sometimes more useful for solving problems, depending on what you are trying to solve for.

As a check, we can conform Eq. (40) by using Eq. (29) and $E \propto PV$:

$$\Delta S = S_2 - S_1 = (C_P - C_V)\ln\frac{V_2}{V_1} + C_V \ln\frac{P_2 V_2}{P_1 V_1} = C_V \ln\frac{P_2}{P_1} + C_P \ln\frac{V_2}{V_1} = C_V \ln\left(\frac{P_2 V_2^{C_P/C_V}}{P_1 V_1^{C_P/C_V}}\right) \tag{41}$$

Setting $\Delta S = 0$ for adiabatic expansion gives Eq. (40).

By the way, you may have seen the adiabatic index before in 15c where you learned that the speed of sound in a gas is $c_{\text{sound}} = \sqrt{\gamma \frac{k_B T}{m}}$. A sound wave is adiabatic compression and rarefaction of a gas. We know it's adiabatic since there is no heat transfer involved.

Isobaric expansion. Isobaric means constant pressure. For a gas to expand and keep the same pressure it must heat up. Thus some heat must be absorbed into the system. We define the the **coefficient of thermal expansion** as

$$\alpha_V = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P \tag{42}$$

We usually treat α_V as approximately constant over a small range of temperature, in which case

$$\Delta V = V \alpha_V \Delta T \tag{43}$$

Thus the bigger α_V is the more the volume changes for the same temperature change. You can get a more accurate formula for the change in volume by integrating Eq. (42), but most of the time Eq. (43) is sufficient.

For an ideal gas,

$$\alpha_V = \frac{1}{V}\frac{\partial}{\partial T}\left(\frac{N k_B T}{P}\right) = \frac{N k_B}{VP} = \frac{1}{T} \tag{44}$$

For example, at $298.2K$, the measured coefficient of thermal expansion for N_2 (i.e air) is $\alpha_{N_2} = 0.003365K^{-1} = \frac{1}{297.2K}$ in excellent agreement with this prediction. Comparing α_V to $\frac{1}{T}$ is an indication of how close to an ideal gas a substance is.

Liquids and solids generally do not act as ideal gases. They expand much less when heated. For example, liquid mercury at $T = 298.2K$ has $\alpha_{\text{Hg}} = 0.000181K^{-1} = \frac{1}{5524K}$, an expansion coefficient about 20 times smaller than for air.

Figure 3. Mercury thermometer.

A mercury thermometer has liquid mercury in a tiny capillary with vacuum above it. An important feature of mercury that makes it good for a thermometer is that essentially none of the liquid evaporates into the air. This prevents pressure from forming which would limit the expansion and make the volume/temperature relation nonlinear as the volume of the capillary fills up. Mercury thermometers were historically so important that a unit of pressure is defined with them: $1\,\text{mm}\,\text{Hg} = 1\,\text{torr}$. Torr is related to bar as $1\,\text{bar} = 750.06$ torr.

Now we can understand why the temperature measured by a thermometer is the same as the "T" we defined by $\frac{1}{k_B T} = \frac{\partial S}{\partial E}$. The explanation is almost tautological. All we need about the statistical-mechanics T is that it takes the same value for all systems in equilibrium with each other. So, let us put our mercury thermometer in equilibrium with a heat bath. As we make the bath hotter T changes and the mercury rises. We put little lines on the mercury and give them labels as "temperature". There is therefore a 1-to-1 correspondence with those labels and T. We just then need one system where we know both T and "temperature", such as the mercury thermometer, to determine the correspondence. Then since any two systems in equilibrium are at the same T, this correspondance between temperature and "T" holds for any system.

4 Carnot's engine

In the 1820's Sadi Carnot became interested in understanding how efficient steam engines could be. A steam engine is a type of heat engine that uses a temperature difference to do work. An example is the Watt engine, which looks like this

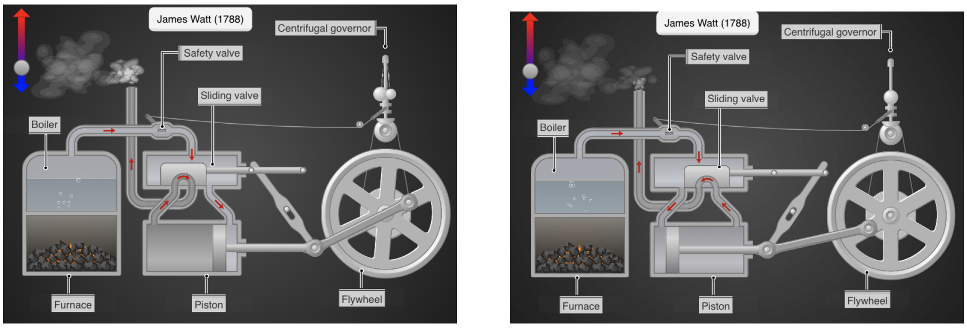


Figure 4. James Watt steam engine. Animation here https://www.youtube.com/watch?v=U-f9UqFiXXk

In the Watt engine, a furnace heats water which produces steam. The steam has high pressure which is used to move a piston, and then is vented out to the ambient air. Carnot wanted to know how much work W can be done *in principle* with the heat Q_1 drawn from the furnace. His important observation was that there is a maximal efficiency for such an engine, depending only on the temperatures of the steam and the air.

An idealized heat engine has two heat baths at temperatures T_H (furnace) and T_C (air) with $T_H > T_C$. Exploiting the temperature difference, we should be able to do some work W. How much work can we do? If we draw an amount of heat Q_{in} from the hot bath, then the efficiency by which the heat is turned into work is

$$\varepsilon = \frac{W}{Q_{\text{in}}} \tag{45}$$

By energy conservation, the energy not used for work must be exhausted as heat $Q_{\text{out}} = Q_{\text{in}} - W$ into the cold bath.

One idea is that we could just put the two baths in contact with each other. Then the heat transfers are equal $Q_{\text{in}} = Q_{\text{out}}$ and no work is done. This has $\varepsilon = 0$.

Carnot realized that the most efficient conversion should involve only reversible processes. First we will describe Carnot's cycle and calculate its efficiency, then in the next section we will explore why reversibility is key to maximizing efficiency. We will use the convention that all the heats and work are positive.

Carnot's 4-stage process starts with a gas in volume V_1 at temperature T_H. The 4 stages are

1. Connect the gas to the T_H heat bath. Perform isothermal expansion at T_H to a volume V_2.
2. Remove the heat bath and insulate the gas. Cool the gas through adiabatic expansion to T_C and a volume V_3.
3. Put the gas in heat bath at T_C. Perform isothermal compression at T_C to some V_4.
4. Remove the heat bath and insulate the gas. Adiabatic compression to back up to T_H and back to V_1.

At the end of this cycle, we are back to where we started with a gas at T_H and V_1 and heat has been transferred from the bath at T_H to the bath at T_C.

Note that these volumes V_1, V_2, V_3 and V_4 are not independently selectable. In the adiabatic stages, the temperature change fixes the final volume. In these stages, PV^γ constant. Using the ideal gas law $PV \propto T$ we also have $TV^{\gamma-1}$ is constant which is more useful for current purposes. So V_2 is a free parameter, but V_4 is not free – it is fixed by what it needs to be for stage 4 to go back to V_1. So there is one free parameter V_2 which determines how big of a cycle we want.

Now let's calculate the efficiency.

Stage 1. Here heat is taken from the hot bath and some work is done. The heat removed is

$$Q_{\text{in}} = N k_B T_H \ln\frac{V_2}{V_1} > 0 \tag{46}$$

As the gas does not heat up, its internal energy is unchanged, so all the heat goes into work. The work we get out is

$$W_{\text{out}}^1 = N k_B T_H \ln\frac{V_2}{V_1} > 0 \tag{47}$$

Stage 2. In adiabatic expansion PV^γ is constant. Since no heat comes in, the energy change from doing work comes out of the internal energy of the gas, $E = C_V T$. Thus the work done this stage is

$$W_{\text{out}}^2 = C_V(T_H - T_C) > 0 \tag{48}$$

Using $TV^{\gamma-1}$ is constant we also find that

$$V_3 = V_2 \left(\frac{T_H}{T_C}\right)^{\frac{1}{\gamma-1}} > V_2 \tag{49}$$

Stage 3. Like stage 1, but compressing to $V_4 < V_3$. Heat goes out and work needs to be done

$$Q_{\text{out}} = W_{\text{in}}^3 = N k_B T_C \ln\frac{V_3}{V_4} > 0 \tag{50}$$

Stage 4 is like stage 2 but compressing

$$W_{\text{in}}^4 = C_V(T_H - T_C) > 0 \tag{51}$$

and

$$V_1 = V_4 \left(\frac{T_C}{T_H}\right)^{\frac{1}{\gamma-1}} < V_4 \tag{52}$$

Note that

$$\frac{V_3}{V_4} = \frac{V_2}{V_1} \tag{53}$$

The net work is then

$$W = W_{\text{out}}^1 + W_{\text{out}}^2 - W_{\text{in}}^3 - W_{\text{in}}^4 = N k_B T_H \ln\frac{V_2}{V_1} - N k_B T_C \ln\frac{V_3}{V_4} = N k_B (T_H - T_C) \ln\frac{V_2}{V_1} \tag{54}$$

So the efficiency is the net work divided by the heat extracted:

$$\boxed{\varepsilon = \frac{W}{Q_{\text{in}}} = \frac{T_H - T_C}{T_H}} \tag{55}$$

This is the Carnot efficiency.

It's sometimes helpful to think about the stages in the Carnot cycle as paths in the PV or ST plane:

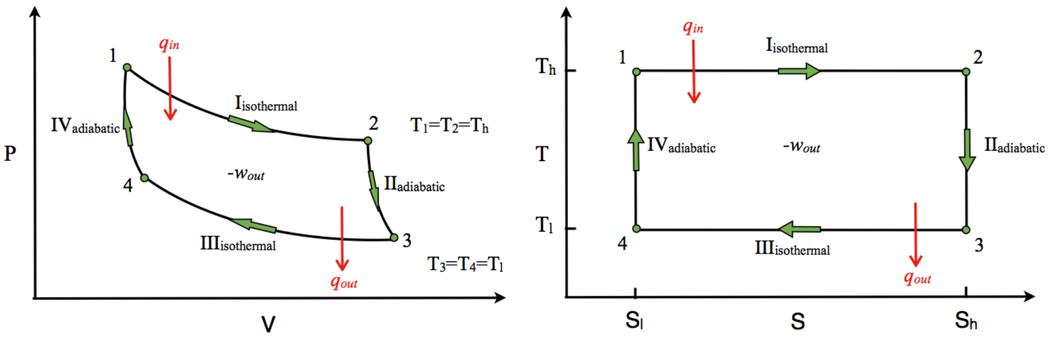

P
V
1
2
3
4
q_{in}
q_{out}
$\mathrm{I}_{isothermal}$
$\mathrm{II}_{adiabatic}$
$\mathrm{III}_{isothermal}$
$\mathrm{IV}_{adiabatic}$
$-W_{out}$
$T_1=T_2=T_h$
$T_3=T_4=T_l$
T_h
T
T_l
S_l
S
S_h

Figure 5. PV diagram and ST diagram for Carnot cycle. In the PV diagram, the isothermal curves have PV = const i.e $P\sim\frac{1}{V}$ and the adiabatic curves have $PV^{\gamma} = \text{const}$ i.e. $P\sim V^{-\gamma}$.

The area under a curve in the PV diagram is $\int PdV$, which is the work done. So the area of the closed region of a reversible cycle is the net work done over a cycle, W. The area under a *reversible* curve in a TS diagram is the heat, $Q=\int TdS$, by Eq. (9), so the area in a closed region in a TS diagram with reversible paths is the net heat transferred, $Q_{\text{out}}-Q_{\text{in}}$. By energy conservation, we must have $Q_{\text{out}}-Q_{\text{in}}=W_{\text{net}}$ by conservation of energy. The efficiency is just $\frac{W}{Q_{\text{in}}}$ which is less than 1 if $Q_{\text{out}}\neq 0$.

5 Second law of thermodynamics

For the Carnot cycle

$$\frac{Q_{\text{in}}}{T_H}=Nk_B\ln\frac{V_2}{V_1}=Nk_B\ln\frac{V_3}{V_4}=\frac{Q_{\text{out}}}{T_C} \tag{56}$$

Recall that the entropy change of a heat bath is $\Delta S=\frac{Q}{T}$. So we see that the entropy decrease of the hot bath $\Delta S_{\text{hot}}=-\frac{Q_{\text{in}}}{T_H}$ is exactly compensated by an entropy increase in the cold bath $\Delta S_{\text{cold}}=\frac{Q_{\text{out}}}{T_C}$. There is no entropy change in the gas, $\Delta S_{\text{gas}}=0$, since it comes back to its initial state after the cycle. Thus the net $\Delta S_{\text{total}}=\Delta S_{\text{hot}}+\Delta S_{\text{cold}}+\Delta S_{\text{gas}}=0$.

In fact, if we were to use that $\Delta S_{\text{total}}=0$ the efficiency would have been easy to calculate. $\Delta S_{\text{total}}=0$ along with $\Delta S_{\text{gas}}=0$ after a complete cycle, we then immediately have

$$\frac{Q_{\text{in}}}{T_H}=\frac{Q_{\text{out}}}{T_C} \tag{57}$$

So the efficiency is

$$\varepsilon=\frac{W}{Q_{\text{in}}}=\frac{Q_{\text{in}}-Q_{\text{out}}}{Q_{\text{in}}}=\frac{T_H-T_C}{T_H} \tag{58}$$

Any efficiency higher than this would *decrease* the entropy overall. (This calculation does not tell whether this efficiency can be achieved, which is why we needed the cycle.) Note that for $T_C=0$ (which is not possible in any real system) the efficiency can be 1. The point is not that $\varepsilon<1$ always, but rather that for a given T_H and T_C there *is* a maximum efficiency.

Another important observation about the Carnot cycle is that it's **reversible**. In fact, each step – adiabatic compression/expansion or isothermal compression/expansion – can be done in reverse, and each has $\Delta S=0$. The result of reversing the whole cycle would be that we put work in and move heat from the cold bath to the hot bath. Thus the efficiency also tells us how efficient a refrigerator can be.

Can we make an engine more efficient than the Carnot engine? If we could, then this new energy would be able to do more work $W' = Q'_{\text{in}} - Q_{\text{out}} > W$ when depositing the same heat Q_{out} into the bath at T_C. We could then run the original Carnot engine in reverse, applying work W to pulling Q_{out} heat out of the T_C bath and putting Q_{in} heat into the hot bath. The net effect is that the difference in heat extracted $Q'_{\text{in}} - Q_{\text{in}}$ is turned *directly* into work $W' - W$. We could then use the work to power something, collect all the heat it produces, and put it back into the heat bath. Such a process is a type of **perpetual motion machine**: it converts heat directly into work (see Section 6).

Note also that the more efficient engine would decrease the entropy of the first bath by $\Delta S_1 = \frac{Q'_{\text{in}}}{T_H} > \frac{Q_{\text{in}}}{T_H} = \frac{Q_{\text{out}}}{T_C} = \Delta S_2$. Thus the net effect would be the overall total entropy decreases.

This leads us to the **second law of thermodynamics**. It has many forms

- Total entropy of system+surroundings never decreases: $\Delta S_{\text{total}} \geqslant 0$.
- The maximum amount of work extractible from heat Q_{in} removed from a heat bath at temperature T_H is $W = Q_{\text{in}} \frac{T_H - T_C}{T_H}$ where T_C is the temperature of the bath to which the energy not used as work is dumped.
- Maximally efficient heat engines are reversible, have $\Delta S_{\text{total}} = 0$, and have $\frac{Q_H}{T_H} = \frac{Q_C}{T_C}$.
- All reversible heat engines have the same efficiency.
- No work can be extracted from heat using a system at a uniform temperature.
- Perpetual motion machines that convert thermal energy into mechanical work with no other consequence (no dumping of heat into a cold bath) are impossible.
- No machine can have the sole effect of transferring heat from a cold bath to a hot bath.

Any of these formulations imply the others.

Note that we used the ideal gas law in our calculation of the Carnot cycle efficiency. We could do this because all reversible engines have the same efficiency by the second law, so we can use any engine we can conceive of, such as an ideal gas one, to calculate that efficiency.

The second law does not say that we cannot convert thermal energy into work. There is no limitation on doing this, other than that imposed by the first law of thermodynamics (energy conservation). For example, we could have a hot gas expand against a piston, and expand and expand until the gas is infinitely large and at zero degrees. Then all the thermal energy would have been turned into work. This doesn't violate the second law, since a gas pushing against a piston really has a vacuum on the other side of the piston, which is at zero temperature, so this is like a heat engine with $T_C = 0$. The kind of perpetual motion machine that is forbidden by the second law is one that takes a single heat source and converts all the thermal energy into work, without dumping any heat anywhere, i.e. without heating anything up.

The second law is stronger than the first law. For example, if you have a hot bath and a cold bath, you can put them in contact and heat will spontaneously flow from hot to cold. You cannot however take heat out of the cold bath and dump it in the hot bath without doing work. Such a refrigerator would be consistent with the first law, since energy is conserved, but violate the second law. Two example constructions that attempt to violate the second law are the Brownian Ratchet, discussed below, and Maxwell's demon, discussed in Lecture 6.

6 Brownian Ratchet

The second law of thermodynamics is one of the most amazing results in all of physics. It seems to be true, has enormously important consequences (like not producing perpetual motion machines) but no one can prove it. All the "proofs" for it (like the Boltzmann H theorem) involve assumptions that are mostly true, but not true in every situation (such as the assumption of molecular chaos or ergodicity).

There are many great examples illustrating how a perpetual motion machine is foiled by a conspiracy of thermodynamics. One of the most famous is called the Brownian Ratchet. It was conceived of in 1912 by Marian Smoluchowski and popularized by Richard Feynman. There is a fantastic discussion of it in the Feynman Lectures on Physics.

A ratchet and pawl is a tool designed to turn only in one direction. The pawl is a little strip of metal held against a set of angular gears (the ratchet) by a spring. When a small force is applied in one direction, the ratchet will turn slowly lifting the pawl until it slips over a gear. But if the force is applied in the other direction, the ratchet will go nowhere - it pushes against the pawl in the wrong direction.

The set up for the Brownian ratchet is shown in this Figure:

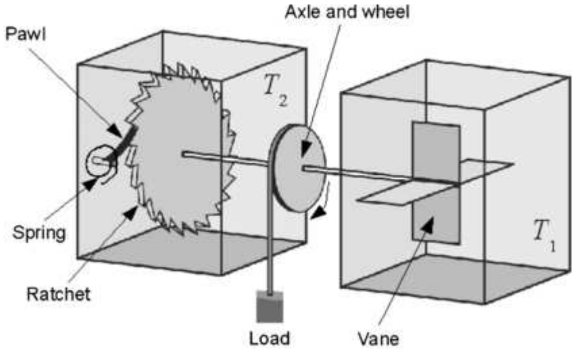


Figure 6. Brownian ratchet

A ratchet and pawl are placed in a box with a gas at temperature T_2. The ratchet is connected via an axle to a vane in another box, at a temperature T_1. The vane is symmetric and can rotate in either direction. Finally, a small weight (the load) is attached to a string tied to the axle.

The Brownian ratchet is meant to work as follows. There is thermal motion of the gas molecules on the vane side at temperature T_1. This motion jiggles the vane back and forth. Sometimes, perhaps rarely, a large enough displacement accumulates counterclockwise to push the pawl over the ratchet. This turns the axle lifting the load. The key point is that the ratchet can only turn in one direction because of the pawl. Thus over time, the wheel will spin in one direction and thermal energy is converted into work. This is called a Brownian ratchet because the ratchet undergoes a biased random walk and therefore Brownian motion.

Note that nothing in this explanation seems to require $T_1 \neq T_2$. If the Brownian ratchet works when $T_1 = T_2$ then it generates work from a system at constant temperature, producing perpetual motion and violating the second law of thermodynamics. What's the catch?

The catch is that thermal fluctuations can affect the pawl, just as they affect the vane. You can see this from the ratchet/pawl/vane system alone, without the weight. Say it takes energy ε for the pawl to be pushed far enough to slip over a gear. This energy is the same whether it accumulates by molecules pushing the vane that pushes the ratchet that pushes the pawl or by molecules that directly push the pawl. The probability for forward motion of the ratchet is then the probability of the energy ε accumulating through Brownian motion on the vane side $P_{\text{forward}} = \exp\left(-\frac{\varepsilon}{k_B T_1}\right)$. The probability of backwards motion, when energy ε accumulates through Brownian motion on the ratchet and pawl side, is $P_{\text{backward}} = \exp\left(-\frac{\varepsilon}{k_B T_2}\right)$. If $T_1 = T_2$ then these probabilities are the same, so it is no more likely for the ratchet to turn one way than the other, and no work gets done.

Now let's take $T_1 \neq T_2$ and see how much work we do. First of all, note that the pawl has to dissipate heat. If the pawl were elastic and completely frictionless, then after one click of the ratchet, it would just bounce up and down conserving energy, like an undamped harmonic oscillator. Such a pawl wouldn't do a very good job holding the ratchet in place – a clockwise force on the ratchet would let it easily slip under the pawl at the top of its bounce. So we already see the most important part of this whole thought experiment: doing work requires some dissipative element (this dissipation plays an important role in the second law of thermodynamics, as we'll see in Lecture 6).

We conclude that for this setup to work in any circumstance, the pawl must have some damping. Then, instead of bouncing forever, it dissipates the energy ε it took to bend back the pawl entirely into heat until the pawl is back to its rest position. Through this heat dissipation, the pawl, the ratchet, and the gas in volume T_2 will heat up. Thus, over time as the system is working T_2 will increase.

Say each time the pawl slips over a gear (one “click”), the ratchet turns by an angle θ. If the weight is exerting a torque L on the ratchet, then the work done in lifting the weight by one click is $\Delta = L\theta = mg\Delta h$, where m is the mass to be lifted by a height Δh. So it takes a total energy $\varepsilon + \Delta$ to lift the weight one click after which ε of the energy is dissipated as heat into the T_2 bath. The chance of this happening due to Brownian motion of the vane is

$$P_{\text{lift}} = \exp\left(-\frac{\varepsilon + \Delta}{k_B T_1}\right) \tag{59}$$

For Brownian motion on the ratchet side to cause the pawl to slip over the ratchet and drop the weight one click, only energy ε is required, so that

$$P_{\text{drop}} = \exp\left(-\frac{\varepsilon}{k_B T_2}\right) \tag{60}$$

If these probabilities are equal, then the weight doesn't move on average and no work is done. In such an equilibrium configuration $P_{\text{lift}} = P_{\text{drop}}$ so that

$$\frac{\varepsilon + \Delta}{T_1} = \frac{\varepsilon}{T_2} \tag{61}$$

Once we are in this equilibrium situation, say we then take a little bit of weight off, a grain-of-sand's worth. This will lower Δ and increase P_{lift} and the weight will start to slowly rise. Alternatively, by adding a grain of sand, the ratchet would turn the other way. So for a very small change in Δ we can make infinitesimal, quasistatic changes. Or we could make small changes in T_1 or T_2 with the same effect.

Now say we make a change so that the weight starts to rise, very very slowly. Each time it moves up it's because energy $\varepsilon + \Delta$ accumulated on the vane side of which Δ went to lifting the weight and ε is dissipated as heat into the T_2 bath. Thus this is an engine from which a heat $Q_1 = \varepsilon + \Delta$ is withdrawn from the T_1 bath, work $W = \Delta$ is done, and heat $Q_2 = \varepsilon$ is deposited into the T_2 bath. Thus

$$\frac{Q_1}{Q_2} = \frac{\varepsilon + \Delta}{\varepsilon} = \frac{T_1}{T_2} \tag{62}$$

So that $\frac{Q_1}{T_1} = \frac{Q_2}{T_2}$. This is exactly the condition satisfied by Carnot's engine! If we are farther from equilibrium, so T_2 is much hotter, then the efficiency goes down.

In other words, the Brownian Ratchet is maximum efficiency heat engine. As Feynman says, an advantage of this system over Carnot's is that you can see what is happening physically. Indeed, it gives you clues as to how to show perpetual motion machines cannot work.

7 Summary

This lecture introduced some of the basic concepts of thermodynamics. Thermodynamics concerns macroscopic quantities like heat, work, energy, temperature, volume etc. These quantities can be studied and measured and relations derived among them without recourse to a microscopic description (statistical mechanics). The main concepts are

- **Work**: energy transferred in an organized manner. Denoted by W
- **Heat**: energy transferred through thermal motion. Denoted by Q.
- **Reversibility**: A process $A \to B$ is reversible if the final state B can be brought back to the original state A exactly, i.e. both system and surroundings must be identical.
- Clausius defined entropy as a state variable (independent of history). To find the entropy change between systems A_{sys} and B_{sys} he said find suitable surroundings A_{surr} and B_{surr} such that the path $A \to B$ is reversible. Then $S_{A_{\text{sys}}} - S_{B_{\text{sys}}} = \int_{\text{rev.}} \frac{dQ_{\text{in}}}{T}$, where the integral is a path integral along the reversible path. That is, the source of entropy change is *heat flow*.
- Ideal gases have pointlike particles with pointlike interactions. These satisfy the ideal gas law $\text{PV} = Nk_B T$.

- **Adiabatic** means no heat flow. Adiabatic process do not have to be slow or reversible. The free expansion of a gas is adiabatic.
- **Quaistatic** means slow enough so equilibrium can be assumed to hold at all times. Reversible processes must be quasistatic, but quasistatic processes may not be reversible.
- A **heat engine** is a way to convert heat transfer into work.
- The maximum possible efficiency $\varepsilon = \frac{W}{Q_{\text{in}}}$ of a heat engine is $\varepsilon \leqslant \varepsilon_{\max} = \frac{T_H - T_C}{T_H}$. **Carnot's** engine achieves this efficiency. Many other idealized engines do too. Any maximally efficient engine is reversible.
- The second law of thermodynamics says that $\Delta S_{\text{sys}} + \Delta_{\text{surr}} \geqslant 0$ in *any process*. If you violated this, you could make an engine with efficiency greater than $\varepsilon_{\max}$. There are many other equivalent versions of the second law.
- The Brownian Ratchet is a "perpetual motion machine" that seems like it converts heat directly into work ($\varepsilon = 1$). A careful quantitative analysis shows that in fact, it does not convert heat to work, but has $\varepsilon \leqslant \varepsilon_{\max}$.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 6: Entropy

1 Introduction

In this lecture, we discuss many ways to think about entropy. The most important and most famous property of entropy is that it never decreases

$$\Delta S_{\text{tot}} \geqslant 0 \tag{1}$$

Here, ΔS_{tot} means the change in entropy of a system plus the change in entropy of the surroundings. This is the second law of thermodynamics that we met in the previous lecture.

There's a great quote from Sir Arthur Eddington from 1927 summarizing the importance of the second law:

> If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations—then so much the worse for Maxwell's equations. If it is found to be contradicted by observation—well these experimentalists do bungle things sometimes. But if your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation.

Another possibly relevant quote, from the introduction to the statistical mechanics book by David Goodstein refers to a curious fact about some physicists who struggled with entropy:

> Ludwig Boltzmann who spent much of his life studying statistical mechanics, died in 1906, by his own hand. Paul Ehrenfest, carrying on the work, died similarly in 1933. Now it is our turn to study statistical mechanics.

There are many ways to define entropy. All of them are equivalent, although it can be hard to see. In this lecture we will compare and contrast different definitions, building up intuition for how to think about entropy in different contexts.

The original definition of entropy, due to Clausius, was thermodynamic. As we saw in the last lecture, Clausius noted that entropy is a function of state, we can calculate the entropy difference between two states by connecting them however we like. If we find a reversible path to connect them, then the entropy change is determined simply by the heat absorbed:

$$\Delta S_{\text{system}} = \int_{\text{rev.}} \frac{dQ_{\text{in}}}{T} \qquad (\text{Clausius entropy}) \tag{2}$$

This definition of entropy (change) called **Clausius entropy**. It is a thermodynamic, rather than statistical-mechanic, definition. It says that entropy is generated (or removed) from heating (or cooling) a system. Clausius entropy is important in that it directly connects to physics. As we saw in the last lecture, if the total Clausius entropy change were negative, it would be possible to create a system whose sole function is to turn heat into work.

In statistical mechanics, we can define the entropy as

$$S = k_B \ln \Omega \qquad (\text{Boltzmann entropy}) \tag{3}$$

where Ω is the number of microstates compatible with some macroscopic parameters (E,V,N). This form is usually attributed to Boltzmann, although it was Planck who wrote it down in this form for the first time. We'll call this the **Boltzmann entropy** since it's on Boltzmann's gravestone. Note that there is no arbitrariness in deciding which states to count in Ω or how to weight them – we count all states compatible with the macroscopic parameters equally, with even weight. That is part of the definition of S. From S, we extract the temperature as $\frac{1}{T}=\frac{\partial S}{\partial E}$ and then integrating over $dE = dQ$ we recover Eq. (2). Thus the Boltzmann entropy definition implies Clausius formula for how to compute entropy change.

Defining S in terms of microstates is useful in that it lets us compute S from a microscopic description of a system. For example, we saw that for a monatomic ideal gas, the Boltzmann entropy is given by

$$S = Nk_B\left[\ln V + \frac{3}{2}\ln\left(\frac{mE}{N}\right) + c\right] \tag{4}$$

for some constant c. This is an example showing that entropy is a state variable: it depends only on the current state of a system, not how it got there (heat and work are not state variables). Clausius formula assumed entropy was a state variable. With the Boltzmann formula, we can check.

Another way to compute entropy came from considering the number of ways N particles could be split into m groups of sizes n_i. This number is $\Omega = \frac{N!}{n_1!\cdots n_m!}$. Expanding for large n_i gives

$$S = -k_B N \sum_{i=1}^{m} f_i \ln f_i \tag{5}$$

where $f_i = \frac{n_i}{N}$. Since $\Sigma n_i = N$ then $\sum f_i = 1$ and so f_i has the interpretation of a probability: f_i are the probabilities of finding a particle picked at random in the group labeled i.

The entropy written in terms of probabilities, but without the factor of N, is called the **Gibbs entropy**:

$$S = -k_B \sum_{i=1}^{m} P_i \ln P_i \qquad \text{(Gibbs entropy)} \tag{6}$$

If all we do is maximize S at fixed N, the prefactor doesn't matter. However, sometimes we care about how S depends on N, in which case we need to get the prefactor right. We'll return to this in Section 3. Taking $P_i = \frac{1}{\Omega} = \text{const.}$ the Gibbs entropy reduces to the Boltzmann entropy. So the Gibbs entropy formula is a generalization of the Botlzmann entropy formula: it applies when probabilities are all equal and also when they are not.

All of these ways of thinking about entropy are useful. They are all ways of understanding **entropy as disorder**: the more microstates there are, the less organized are the particles. A solid has lower entropy than a gas because the molecules are more ordered: the constraints on the positions of the atoms in the solid and limitations on their velocities drastically reduce the number of possible configurations. Entropy as disorder is certainly intuitive, and conforms with common usage: a child playing with blocks will most certainly leave them more disordered than how she found them, so we say she has increased the entropy. There are fewer ways for something to be ordered than disordered.

In the second half of the 20th century, it was realized that a more general and useful way of thinking about entropy than entropy as disorder is **entropy as uncertainty**. That is, we associate entropy with our ignorance of the system, or our **lack of information** about what microstate it's in. Entropy as uncertainty makes a lot of unintuitive aspects of the second law of thermodynamics easier to understand. We have already come across this the connection between entropy and information when discussing the principle of molecular chaos – the velocities of particles become correlated when they scatter, but over time the information of their correlations disperses over phase space and is lost. A solid has lower entropy than a gas because we have more information about the location of the atoms.

Thus, a single definition of entropy that unifies all the other is

- Entropy measures the amount of information you would need to completely specify a microstate for given macroscopic properties.

In this lecture, we'll see how thinking of entropy as information can resolve a number of classical paradoxes, such as Gibbs paradox and Maxwell's demon, and how entropy connects to quantum mechanical measurement as well.

2 Free expansion

An example that helps elucidate the different definitions of entropy is the free expansion of a gas from a volume V_1 to a volume V_2.

First, consider the Boltzmann entropy, defined as $S = k_B \ln\Omega$ with Ω the number of accessible microstates. Using Eq. (4) which follows from the Boltzmann entropy definition, in going from a volume V_1 to a volume V_2, the gas gains an amount of entropy equal to $\Delta S = N k_B \ln\frac{V_2}{V_1}$. That the Boltzmann entropy increases makes sense because there are more accessible microstates in the larger volume, $\Omega_2 > \Omega_1$.

What about the Gibbs entropy? If P_i is the probability of finding the system in microstate i, then $P_i = \frac{1}{\Omega_1}$ when the gas is at a volume V_1. When it expands to V_2, each microstate of the gas in V_1 corresponds to exactly one microstate of the gas in V_2, so we should have $P_i = \frac{1}{\Omega_1}$ also at volume V_2, and therefore the Gibbs entropy appears to be unchanged! Although there are more possible microstates in V_2, we know that, since the gas came from V_1, that only a small fraction of these could possibly be populated. That is, the state after expansion is in a subset $\mathcal{M}_{\text{sub}} \subset \mathcal{M}_2$ of the full set $\mathcal{M}_2$ of microstates in V_2. The microstates in $\mathcal{M}_{\text{sub}}$ are exactly those for which if we reverse time, they would go back to V_1. The size Ω_1 of the set $\mathcal{M}_1$ of microstates in V_1 is the same as the size Ω_{sub} of $\mathcal{M}_{\text{sub}}$.

So in the free expansion of a gas, Boltzmann entropy increases but Gibbs entropy does not appear to. How do we reconcile these two concepts?

The origin to this inconsistency is that Boltzmann entropy is defined in terms of the number of states Ω consistent with some macroscopic parameters, V, E, N, etc.. In contrast "the set of states that when time reversed to back to V_1" used for Gibbs entropy depends on more than just these parameters. So we are computing the different entropies using different criteria. If we define the probabilities P_i for the Gibbs entropy the same way as we define Ω for the Boltzmann entropy, that is, as the probability for finding a state with given values of V, E, N, the two definitions will agree. Indeed, the number of new states is $\frac{\Omega_2}{\Omega_1} = \exp\left(\frac{\Delta S}{k_B}\right) = \left(\frac{V_2}{V_1}\right)^N$. Including these new states makes the Gibbs entropy go up by $\Delta S = N k_B \ln\frac{V_2}{V_1}$; removing them makes the Boltzmann entropy go down by the same amount. So if we are consistent with our criterion for computing entropy, the different definitions agree.

Now, you may be wondering why we choose to define entropy using V, E, N and not using the information about where the particles originated from. As an extreme example, we could even say that the gas starts in exactly one phase space point, $(\vec{q}_i, \vec{p}_i)$. So $\Omega = 1$ and the probability being in this state is $P = 1$ with $P = 0$ for other states. We can then evolve the gas forward in time using Newton's laws, which are deterministic and reversible, so that in the future there is still only one state $P = 1$ or $P = 0$. If we do so, $S = 0$ for all time. While we could choose to define entropy this way, it would clearly not be a useful concept. Entropy, and statistical mechanics, more broadly, is useful only if we coarse grain. Entropy is *defined* in terms of the number of states or probabilities compatible with macroscopic parameters. Coarse graining is part of the definition. Remember, in statistical mechanics, we are not in the business of predicting what will happen, but what is overwhelmingly likely to happen. Defining entropy in terms of coarse grained probabilities is the trick to making statistical mechanics a powerful tool for physics.

To justify why we must coarse grain from another perspective, recall the arguments from Lecture 3. Say we have a minimum experimental or theoretical phase space volume $\Delta q\Delta p$ that can be distinguished. Due to molecular chaos, the trajectory of a $\Delta q\Delta p$ phase space region quickly fragments into multiple disconnected regions in phase space that are smaller than $\Delta q\Delta p$ (since phase space volume is preserved under time evolution by Louiville's theorem). Then we coarse grain to increase the phase space volume of each disconnected region to a size $\Delta q\Delta p$. In this way, the phase-space volume of $\mathcal{M}_{\text{sub}}$ grows with time. By ergodicity, every point in $\mathcal{M}_2$ will eventually get within $\Delta q\Delta p$ of a point in $\mathcal{M}_{\text{sub}}$, so if we wait long enough, $\mathcal{M}_{\text{sub}}$, through coarse graining, will agree with $\mathcal{M}_2$.

It may be helpful to see that if we use Gibbs entropy definition that entropy does in fact increase during diffusion. For simplicity consider a diffusing gas in 1 dimension, with number density $n(x, t)$. By the postulate of equal-a-priori probabilities, the probability of finding a gas molecule at x, t is proportional to the number density $P(x,t) \propto n(x,t)$. Then the Gibbs entropy is $S(t) = -c\int dx n(x,t) \ln n(x,t)$ for some normalization constant c. Now, the diffusing gas satisfies the diffusion equation $\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial x^2}$. Using this we find

$$\frac{dS}{dt} = -c\int dx \frac{\partial}{\partial t}[n \ln n] = -c\int dx\, [1 + \ln n]\frac{dn}{dt} = -cD\int dx\, [1 + \ln n]\frac{\partial}{\partial x}\frac{\partial}{\partial x}n \tag{7}$$

$$= \underbrace{-cD(1 + \ln n)\frac{\partial}{\partial x}n\Big]_{-\infty}^{\infty}}_{=0} + cD\int dx \frac{1}{n}\left(\frac{\partial n}{\partial x}\right)^2 > 0 \tag{8}$$

where we integrated by parts in the last step. The boundary terms vanish, since the density should go to zero at spacial infinity as we are at finite volume. We conclude that during diffusion the entropy strictly grows with time. It stops growing when $\frac{\partial n}{\partial x} = 0$, i.e. when the density is constant, which is the state of maximum entropy.

This calculation illustrates that the problem from the beginning of this section was not that the Gibbs entropy wouldn't increase, but rather that imposing constraints on the probabilities based on inaccessible information about past configurations is inconsistent with the postulate of equal a priori probabilities.

3 Entropy of mixing

Although entropy is a theoretical construction – it cannot be directly measured – it is nevertheless extremely useful. In fact, entropy can be used to do work. One way to do so is though the entropy of mixing.

Say we have a volume V of helium with N molecules and another volume V of xenon also with N molecules (both monatomic gases) at the same temperature. If we let the gases mix, then each expands from V to $2V$. The energy of each is constant so the entropy change of each is (from integrating the discussion of free expansion before or directly from Eq. (4)):

$$\Delta S = Nk_B \ln\frac{2V}{V} = Nk_B \ln 2 \tag{9}$$

So we get a total entropy change of

$$\Delta S = 2Nk_B \ln 2 \tag{10}$$

This increase is called the **entropy of mixing**.

The entropy of mixing is a real thing, and can be used to do work. For example, say we had a vessel with xenon on one side and helium on the other, separated by a semi-permeable membrane that lets helium pass through and not xenon. Say the sides start at the same temperature and pressure. As the helium inevitably diffuses through the membrane, it dilutes the helium side lowering its pressure and adds to the pressure on the xenon side. The net effect is that there is pressure on the membrane. This pressure can be used to do work.

The pressure when there is a mixed solution (e.g. salt water) in which the solute (e.g. salt) cannot penetrate a semi-permeable barrier (e.g. skin) is is called **osmotic pressure**. For example, when you eat a lot of salt, your veins increase salt concentration and pull more water in from your body to compensate, giving you high blood pressure.

In chemistry and biology, **concentration gradients** are very important. A concentration gradient means the concentration of some ion, like Na^+ or K^+ is not constant in space. When there is a concentration gradient, the system is not completely homogeneous, so entropy can be increased by entropy of mixing. Only when the concentrations are constant is there no way to increase the entropy more. Concentration gradients are critical for life. Neurons fire when the concentration of certain chemicals or ions passes a threshold. If systems always work to eliminate concentration gradients, by maximizing entropy, how do the concentration gradients develop? The answer is work! Cells use energy to produce concentration gradients. There are proteins in the cell wall that work to pump sodium and potassium (or other ions) from areas of low concentration to areas of high concentration.

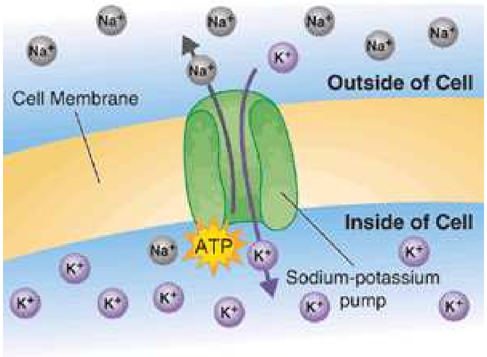


Figure 1. A cellular pump, present in practically every cell of every living thing. It uses energy in the form of ATP to establish concentration gradients

Another familiar effect due the entropy of mixing is how salt is able to melt ice. The salt draws water out of the ice (i.e. melts it) because saltwater has higher entropy than salt and ice separately.

In order to study these physical processes in quantitative detail, we need first to understand entropy better (this lecture), as well as free energy, chemical potentials, and phase transitions, which are topics for the next few lectures. We'll quantify osmotic pressure and temperature changes due to mixing (like in saltwater) in Lectures 8 and 9.

3.1 Gibbs paradox

So entropy of mixing is a real thing, and is very important in physics, chemistry and biology. But it is still a little puzzling. Say instead of two different ideal gases, we just have helium. Again we start with two volumes V, N and E each, and remove a partition between them to let them mix. The calculation is identical to the calculation for helium and xenon and we still find $\Delta S = 2Nk_B \ln 2$.

What happens if we put the partition back in. We start with helium gas with volume $2V$ number $2N$ and energy $2E$. Its initial entropy, by Eq. (4) is

$$S_{\text{init}} = 2Nk_B\left[\ln 2V + \frac{3}{2}\ln\left(\frac{mE}{N}\right) + c\right] \tag{11}$$

Now put a partition right down the middle of the gas, splitting it into two equal halves, each with V, N and E. Then the entropy is the sum of the entropies of the two halves:

$$S_{\text{final}} = 2\left\{Nk_B\left[\ln V + \frac{3}{2}\ln\left(\frac{mE}{N}\right) + c\right]\right\} = S_{\text{init}} - 2Nk_B\ln 2 \tag{12}$$

Thus the entropy has gone down, by exactly the entropy of mixing. If we mix the gases entropy goes, up, if we split them entropy goes down. That entropy could go down by simply sticking a partition in a gas seems very strange, and apparently violates the second law of thermodynamics. This is called the **Gibbs paradox**.

There are two parts to resolving the Gibbs paradox. First, we have to discuss indistinguishability. So far we assumed a microstate with helium molecules at positions A and B was different from one where the molecules swapped places. Is that correct? But let's say it is correct as if each molecule had a little label on it. Then the counting that we have been using, which is indeed counting *something*, would have entropy apparently go down. We'll address the first issue first and return to the second in Section 6.

3.2 Distinguishable microstates

To resolve the Gibbs paradox, let's think about why the entropy is changing from the viewpoint of Boltzmann entropy and ergodicity. Again, we start with xenon on the left and helium on the right each in a volume V. At a time t_0 we let them mix. The number of microstates for each gas increases by $\Delta S = Nk_B\ln\frac{2V}{V}$, so the net entropy of mixing is $\Delta S = 2Nk_B\ln 2$ (as we have seen). The entropy increases because there are now new microstates with xenon on the right or helium on the left that weren't there before. By ergodicity, we coarse grain and add these the microstates to Ω. The new microstates we include are not the ones actually populated by the expanding gas, but rather ones exponentially close to those microstates that we can't tell apart (recall Fig. 4 from Lecture 3). The new states don't trace back to being separated at $t = t_0$ but the original ones do.

Now let's do the same thing for two volumes of helium that are allowed to mix starting at time t_0. As they mix, the entropy change is $\Delta S = Nk_B\ln 2$ just as for the xenon/helium mixture. This entropy increase comes because we are adding to the original microstates new microstates that, when traced back to t_0 have the molecules from the two original volumes still distributed throughout the whole system. But for the helium/helium mixing, these states *do* correspond to states we started with: half helium and the other half helium. So we already had these states and we are including them again and overcounting.

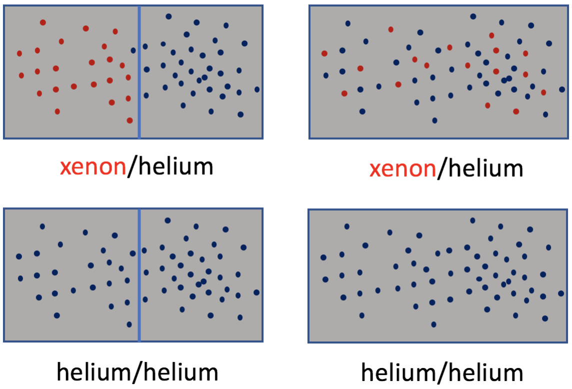


Figure 2. When xenon and helium mix, the new microstates we add don't look like the old ones. When helium mixes with helium, the new microstates are indistinguishable from the old ones.

In our original counting of states, we said that in a length L, with some minimal size Δq for discretizing the box, the number of states was $\frac{L}{\Delta q}$. Each particle could be anywhere, and for $\Delta q \ll L$ the chance of finding two particles in the same place is negligible. This leads to $\Omega = \left(\frac{L}{\Delta q}\right)^N$.

Let's take L to be very small, twice the minimal size ($L = 2\Delta q$). Then there are only two possible states. If there are N helium molecules, then there are $\Omega = 2^N$ configurations and $S = Nk_B \ln 2$. Now say we partition the system into two halves, with $\frac{N}{2}$ particles in each half. For each half, there is only one possible place for each particle, so the system is fixed, $\Omega = 1$ and $S = 0$. Thus entropy has decreased! What went wrong? The catch is that there were $\Omega_{\text{split}} = \begin{pmatrix} N \\ N/2 \end{pmatrix} = \frac{N!}{\frac{N}{2}!\frac{N}{2}!} \sim 2^N$ different ways to split the molecules, and we only considered one of them. If each molecule were different, say we had N different colored balls, then we could tell which of the states we ended up with. In that case, entropy of the system would go down $\Delta S_{\text{sys}} < 0$, however it would take a lot of work to count all the balls and entropy of the surroundings would increase from doing all this counting. It's easier just to say we do not know which colored ball went where and include the Ω_{split} combinatoric factor. For helium, maybe we can tell them apart, or maybe not, but if we're not planning to try, then we need to account for the Ω_{split} possibilities. We can do this by saying that all the Ω_{split} partitionings of the particles are the same microstate. See section 3.3 for a detailed calculation of these cases.

It's actually easier to account for the combinatoric factor in a large volume than with $L = 2\Delta q$. In a large volume, there are many more positions than there are particles so we can assume no two particles are in the same state. In fact we have to make this assumption in the classical limit, since if there is a reasonable chance that two particles are in the same state we need to know if they are fermions or bosons and must use the appropriate quantum statistics (Lecture 10). In a large volume with every particle in a different position, we can simply divide by the $N!$ for permuting those positions. This leads to

$$\Omega(q, p) = 2e^{\frac{3}{2}N} \frac{1}{N!} \left(\frac{V}{(\Delta q \Delta p)^3}\right)^N \left(\frac{4\pi m E}{3N}\right)^{\frac{3N}{2}} \tag{13}$$

This is the same as our old formula but has an extra $\frac{1}{N!}$ in front. The extra factor of $N!$ was introduced by Gibbs. After using Stirling's approximation the entropy $S = k_B \ln \Omega$ is

$$\boxed{S = Nk_B \left[\ln \frac{V}{N} + \frac{3}{2} \ln\left(\frac{4\pi m E}{3Nh^2}\right) + \frac{5}{2}\right]} \tag{14}$$

We have used $\Delta q \Delta p = h$ (as will be explained in Lecture 10). This is the **Sackur-Tetrode** equation.

Does the extra $N!$ solve Gibbs paradox? For the gas with $2V, 2N$ and $2E$, Eq. (11) becomes

$$S_{\text{init}} = 2Nk_B \left[\ln \frac{V}{N} + \frac{3}{2} \ln\left(\frac{mE}{N}\right) + c\right] \tag{15}$$

After splitting into two halves, Eq. (12) becomes

$$S_{\text{final}} = 2\left\{Nk_B \left[\ln \frac{V}{N} + \frac{3}{2} \ln\left(\frac{mE}{N}\right) + c\right]\right\} = S_{\text{init}} \tag{16}$$

So the entropy is unchanged by adding, or removing the partition.

What about the xenon/helium mixture? The gases are independent and do not interact, so each one separately acts just like helium alone. Thus inserting a partition in a helium/xenon mixture has a net effect of $\Delta S = 0$ on each separately and therefore $\Delta S = 0$ total as well.

What about the entropy of mixing? Let's start with two separate gases. Using our new formula, the initial entropy is the sum of the two gases' entropies. Each one has volume V, energy E and N. So,

$$S_{\text{init}} = 2\left\{Nk_B \left[\ln \frac{V}{N} + \frac{3}{2} \ln\left(\frac{mE}{N}\right) + c\right]\right\} \tag{17}$$

After letting the gasses mix, each gas goes from $V \to 2V$ but N and E are the same, so we have

$$S_{\text{final}} = 2\left\{ N k_B \left[\ln\frac{2V}{N} + \frac{3}{2}\ln\left(\frac{mE}{N}\right) + c \right] \right\} = S_{\text{init}} + 2 N k_B \ln 2 \tag{18}$$

So now, with the $N!$ factor added, we get a result that makes sense: there is only entropy of mixing if the two gases are different. Inserting a partition never changes the entropy.

3.3 Mixing example details

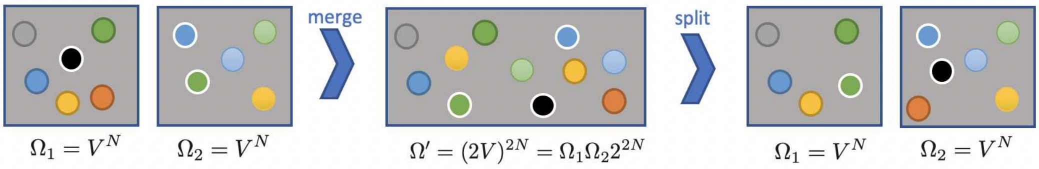
Since distinguishability is an important and confusing topic, let's do an example of mixing and unmixing in full detail, with three different assumptions

1. N colored balls + N more colored balls, all *distinguishable*
2. N molecules helium + N more molecules helium, all *indistinguishable*
3. N molecules helium + N molecules xenon, each *separately indistinguishable*

In all cases, we start with a volume V of each, then let them mix, then put a partition in to separate them.

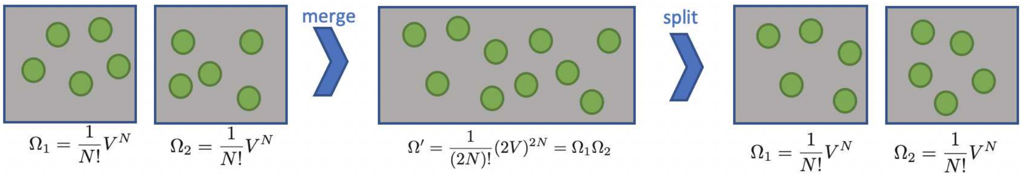
Note that we are using distinguishable here to mean classical indistinguishability, not quantum. That is, it's just about whether we use identifying information about individual molecules in the definition of the microstates. Due to quantum statistics, it's impossible to actually give color labels to separate helium molecules since they are fundamentally indistinguishable. But here we just care about what information we have about the system, not where that information came from. Just to be safe, we talk about colored balls instead of molecules to avoid any possible conflicts with quantum mechanics. But quantum mechanics is not required to resolve the Gibbs paradox.

For the first case we have something like this



So it looks like the entropy goes up by $\Delta S = 2N \ln 2$ when the two are mixed and then *down* by $\Delta S = -2N \ln 2$ when they are split. However, note that there are $\Omega_{\text{split}} = \binom{2N}{N} \approx 2^{2N}$ ways to split them. So in enumerating the final states, we should include this factor, writing $\Omega'' = \Omega_1 \Omega_2 \Omega_{\text{split}} = V^N V^N 2^{2N}$ so that $\Delta S = 0$ upon the split. If we actually look in the box and enumerate which balls are where, we lose Ω_{split}, but the entropy of the surroundings must go up to compensate, as we explain in Section 5.

We can contrast this to the pure helium (indistinguisable) case when


merge
split
$\Omega_1 = \frac{1}{N!} V^N$
$\Omega_2 = \frac{1}{N!} V^N$
$\Omega' = \frac{1}{(2N)!} (2V)^{2N} = \Omega_1 \Omega_2$
$\Omega_1 = \frac{1}{N!} V^N$
$\Omega_2 = \frac{1}{N!} V^N$

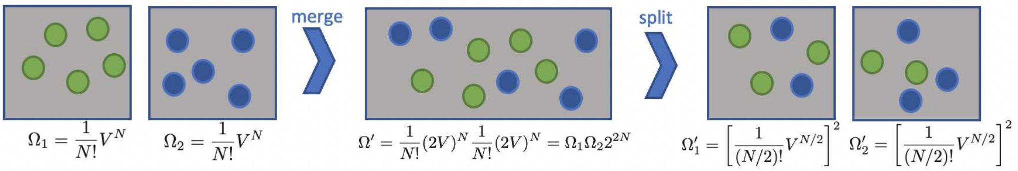
Here, we add the $\frac{1}{N!}$ factors for identical particles. When the two sets are merged, then the number of states is

$$\Omega' = \frac{1}{(2N)!}(2V)^{2N} = \Omega_1 \Omega_2 \frac{N!N!}{(2N)!} 2^{2N} \approx \Omega_1 \Omega_2 \frac{N^N N^N}{(2N)^{2N}} 2^{2N} = \Omega_1 \Omega_2 \tag{19}$$

So $\Delta S = 0$ upon merging. Similarly, $\Omega'' = \Omega_1 \Omega_2 = \Omega'$, so $\Delta S = 0$ upon splitting.[1]

[1] You might be bothered the fact that we had to take the thermodynamic limit $N \to \infty$, which lets us use that the most probable configuration is the only configuration, i.e. that there are always N particles in each side after the splitting. At finite N then indeed $\Omega' > \Omega_1 \Omega_2$, so entropy goes up on merging. After splitting, we must allow for m particles on one side and $N - m$ on the other. Then $\Omega'' = \sum_{m=0}^{2N} \frac{1}{(2N-m)!} V^{(2N-m)} \frac{1}{m!} V^m = \frac{(2V)^{2N}}{2N!} = \Omega'$ exactly, so entropy still does not go down upon splitting.

When we mix helium and xenon we have


merge
split
$\Omega_1 = \frac{1}{N!}V^N$
$\Omega_2 = \frac{1}{N!}V^N$
$\Omega' = \frac{1}{N!}(2V)^N \frac{1}{N!}(2V)^N = \Omega_1\Omega_2 2^{2N}$
$\Omega_1' = \left[\frac{1}{(N/2)!}V^{N/2}\right]^2$
$\Omega_2' = \left[\frac{1}{(N/2)!}V^{N/2}\right]^2$

In this case, after mixing, each set of N molecules occupy the volume $2V$, so the entropy of mixing is $\Delta S = 2\text{Nln}\,2$, just as in the colored balls case. When we split them, since the particles are identical, there is no way to tell apart one splitting from the other. Each half has $\frac{N}{2}$ of each species in a volume V. So the total number of states in this case is

$$\Omega'' = \left[\frac{1}{\left(\frac{N}{2}\right)!}V^{N/2}\right]^4 = \Omega_1\Omega_2\frac{N!N!}{\left(\frac{N}{2}!\right)^4} \approx \Omega_1\Omega_2\frac{N^NN^N}{\left(\frac{N}{2}\right)^{2N}} = \Omega_1\Omega_2 2^{2N} = \Omega' \tag{20}$$

And therefore $\Delta S = 0$ for the splitting case.

So we see that in the distinguishable case (colored balls) or the helium/xenon mixture case, there is a $2N\ln 2$ entropy of mixing – each of the 2N molecules now has an extra binary choice of where to be, so we get $2N\ln 2$. In no situation does entropy go down when we split the volume back into two halves.

3.4 Entropy is extensive

Note that with the extra factor $N!$ in Eq. (13) entropy has become an **extensive quantity**. Extensive quantities are defined as those that double when you have twice as much of the same thing. For example, energy E is extensive, as are N and V. The ratio of two extensive, quantities is an **intensive property**: one that characterizes the stuff itself, independent of how much you have. Temperature and pressure and the heat capacity C_V are intensive.

To see that entropy is extensive, note from the Sackur-Tetrode formula that doubling V, N and E makes S double. This makes sense from the original definition – if you have two isolated systems with Ω_1 microstates in one (entropy $S_1 = k_B\ln\Omega_1$) and Ω_2 microstates in the other (entropy $S_2 = k_B\ln\Omega_2$) then the total number of microstates is $\Omega = \Omega_1\Omega_2$. So

$$S_{12} = S_1 + S_2 \tag{21}$$

This is true if the systems are truly isolated, whether or not we have the extra factor of N in the formula. If the systems are not isolated, Eq. (21) only works if we specify whether the particles are distinguishable – so that we know if we need to add new states (by coarse graining) – or if they are indistinguishable – so that coarse graining would not add anything new. Getting entropy to be extensive both for distinguishable and indistinguishable particles was what motivated Gibbs to add the $N!$ to Ω. The $N!$ is associated with indistinguishable particles.

We can also see the extensive property from the definition of Gibbs entropy in terms of probabilities, in Eq. (6). Say we have two systems A and B with probabilities P_i^A and P_j^B. Then the Gibbs entropy of the combined system is

$$S_{AB} = -k_B\sum_{i,j} P_i^A P_j^B \ln(P_i^A P_j^B) \tag{22}$$

$$= -k_B\left(\sum_i P_i^A\right)\sum_j P_j^B\ln(P_j^B) - \left(\sum_j P_j^B\right)\sum_i P_i^A\ln(P_i^A) \tag{23}$$

$$= S_A + S_B \tag{24}$$

So Gibbs entropy is an extensive quantity. If we had used the formula with the extra factor of N, Eq. (5), we would have found $S_{\text{AB}} \neq S_A + S_B$.

The extensivity often simplifies properties of entropy. For example, say we have m possible states a single particle can be in, and the probability of being in state i is P_i. Then the entropy for a single particle is the Gibbs entropy $S = -k_B \sum P_i \ln P_i$. Now if we have N *indistinguishable* particles, we get simply N times this

$$S = -k_B N \sum_{i=1}^{m} P_i \ln P_i \tag{25}$$

This is identical to the Boltzmann entropy formula in Eq. (5). Indeed, the Boltzmann formula was based on a state counting $\Omega = \frac{N!}{n_1! \ldots n_m!}$. The factors of $n_i!$ in the denominator precisely compensate for the fact that we do not attempt to distinguish the $n_i!$ particles in state i. That is, the factor of N difference between Eq. (5) and Eq. (6) is due to Boltzmann's formula applying for indistinguishable particles, and Gibbs' formula for distinguishable ones.

To be clear, **indistinguishability** just means we are not interested in trying to tell them apart. There is a quantum definition of indistinguishability for identical particles. We're not talking about that. We're just talking about whether we want to use a device that can distinguish all of the $N \sim 10^{24}$ helium molecules from each other. If we're not planning to exploit their differences, we should treat the particles as indistinguishbale. Recall that to do work using the entropy of mixing, we need something like a semipermeable membrane that lets one type of thing through and not the other. Perhaps we could tell a He^4 isotope from a He^3 isotope. An entropy definition that can tell them apart would differ from one that says they are distinguishable by 2!, which hardly matters. You are never going to be able to pick out each of the 10^{24} individual He atoms in a gas, so the $N!$ should be included. In a metal, on the other hand, you can tell where each atom is, so the atoms in a solid should be treated as distinguishable, even if they are identical elements (like in solid gold). We'll talk about metals in Lecture 13.

It's worth adding that extensivity is a convenient property for entropy to have, but it is not guaranteed by any of the definitions of entropy. Indeed, in some systems, such as stars, entropy is not extensive due to long-range forces (gravity). With long-range interactions when you double the amount of stuff, the system can be qualitatively different (a star not much smaller than the sun would not be hot enough to burn hydrogen, see Lecture 15). With that caveat in mind, for the vast majority of systems we consider, where interactions are local (nearest neighbor or contact interactions), entropy will be extensive and we can exploit that property to simplify many formulae and calculations.

4 Information entropy

Next, we introduce the concept of information entropy, as proposed by Claude Shannon in 1948. We'll start by discussing information entropy in the context of computation, as it was originally introduced, and then connect it back to physics once we understand what it is.

Consider the problem of data compression: we have a certain type of data and want to compress it into as small a file as possible. How good is a compression algorithm? Of course if we have two algorithms, say .jpg and .gif, and some data, say a picture of a cat, then we can just compress the data with the algorithms and see which is smaller. But it may turn out that for one picture of a cat the jpeg comes out smaller, and for another, the gif is smaller. Then which is better? Is it possible to make an algorithm better than either? What is the absolute best an algorithm can do?

If you want *lossless compression*, so that the data can always be restored exactly from the compressed file, then it is impossible for any algorithm to compress all data. This follows from the "pigeonhole principle": you cannot put m pigeons in n holes if $n < m$ without some hole having more than one pigeon.

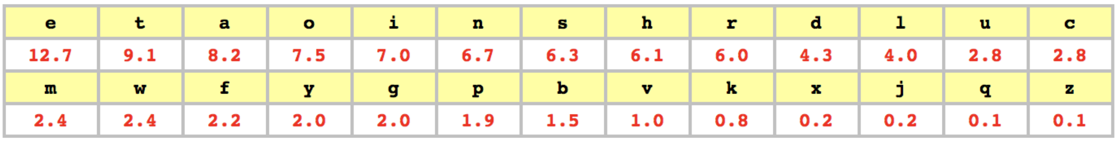
So non algorithm can losslessly compress every possible file. But that's ok. Most data files have some structure. For example, images often have similar colors next to each other. This leads to the idea of run-length-encoding: instead of giving all the colors as separate bytes, encode the information in pairs of bytes: the first byte in the pair gives the color and the second byte gives the number of pixels of that color. Run-length-encoding was used in early versions of .gif compression. It will compress almost all pictures. But if you give it white noise, where neighboring pixels are uncorrelated, then the compressed file will be bigger than the original.

Another feature in image data is that it often has smooth features separated by relatively sharp edges. Thus taking a discrete Fourier transform becomes efficient, since high frequency modes are often absent in images. jpeg compression is based on discrete Fourier transforms. Again, white noise images will not compress because their Fourier transforms are not simpler than the original images.

4.1 Text

For data that is text, the information is a sequence of letters. Different letters appear in a typical text with different frequencies. The standard way to write uncompressed text as numbers is with the ASCII code (American Standard Code for Information Interchange). A **bit** (short for binary digit) is the smallest unit of information. It can take the value 0 or 1. In ASCII every letter is assigned a number from 0 to 127 which takes up 7 bits. For example, the ASCII code for "e" is 101 and the ASCII code for "&" is 38. There is an extended ASCII code as well, with 8 bits, allowing for letters such as "ä" which is 228. Since "e" is much more common than "&" it should be possible to efficiently compress text, allowing for random sequences of symbols not to compress well.

Here is a table of the probability of getting a given letter in some English text:



e	t	a	o	i	n	s	h	r	d	l	u	c
12.7	9.1	8.2	7.5	7.0	6.7	6.3	6.1	6.0	4.3	4.0	2.8	2.8
m	w	f	y	g	p	b	v	k	x	j	q	z
2.4	2.4	2.2	2.0	2.0	1.9	1.5	1.0	0.8	0.2	0.2	0.1	0.1

Figure 3. Probabilities of finding different letters in English text.

For example, if you open a book and pick a letter at random, 12.7% of the time the letter you pick will be "e" and only 0.1% of the time it will be "q". Exploiting these frequencies, a good compression algorithm would find a way to use fewer bits for "e" and more bits for "q".

What is the minimal number of bits you need to encode a given type of data? Shannon's answer was

$$H = \text{minimal}\,\#\text{bits} = -\sum_i P_i \log_2 P_i \tag{26}$$

This quantity H is called the **Shannon entropy** or **information entropy**. (You may recognize this H as the same as Boltzmann's H with $\log_e$ replaced by $\log_2$.) Shannon proved that you can't ever encode data with fewer than H bits, on average. This is known as the **source coding theorem**.

For example, if you have a single coin, with $P_{\text{head}} = \frac{1}{2}$ and $P_{\text{tails}} = \frac{1}{2}$. Then

$$H = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = 1 \tag{27}$$

So a coin has 1 bit. Note we are taking logarithms base 2 here. So one bit has $H = 1$ not $H = \ln 2$

With the letter sequences above we find

$$H = -[0.127\log_2 0.127 + 0.091\log_2 0.091 + \cdots + 0.001\log_2 0.001] = 4.17 \tag{28}$$

This means the best we can possibly do is to represent each letter with 4.17 bits, on average. Having a non-integer number of bits it is not a problem; it just means that you could encode 100 characters with 417 bits and so on (see coin example below).

Note that 4.17 is better than the 7 bits in ASCII. Of course, we don't need 7 bits to represent 26 letters, but a naive encoding would use 5 bits (so $2^5 = 32$ characters), so since $H = 4.17 < 5$ this says you can do than better than 5 bits. Since we are considering non-integer bits, you might say we should use $\log_2 26 = 4.7$ bits. Indeed, the 4.7 bits is exactly what Shannon entropy would say is the best encoding if the probabilities of finding each letter were equal: $P_i = \frac{1}{26}$. That is

$$H_{\text{equal}} = -\sum_{i=1}^{26} \frac{1}{26} \log_2 \frac{1}{26} = \log_2 26 = 4.7 \tag{29}$$

That reason that we can use only 4.17 bits on average instead of 4.7 is because the probabilities are not equal. A better algorithm *uses this extra information.*

Shannon also noted that in text, letters aren't randomly distributed with probabilities but form words. Using words rather than letters, in his 1948 paper Shannon estimated that $H \approx 2.62/\text{letter}$ for the entropy of English text. That is, it only takes 2.62 bits/letter to encode words. If a letter is given one byte (8 bits) in extended ASCII, this says that the maximal compression you could get is a compression factor of $\frac{8}{2.62} = 3.05$.

The prize is a 50,000€ competition to compress a 100MB snapshot of Wikipedia. The current record is 16 MB. For each 1% improvement you get 1,000€. Note that already the compression factor is $\frac{100}{16} = 6.25$ so that each character is represented by $\frac{8\,\text{bits}}{6.25} = 1.28$ bits. This is already much better than Shannon's original estimate. The improvement implies that Shannon's estimate of H is off, probably because he did not use all the information about the regularity of English text (for example, sentence structure); perhaps also Wikipedia articles are not typical text.

4.2 Algorithms

The source coding theorem doesn't tell you how to maximally compress the data, just what the maximal compression rate is. Let's think a little about what an optimal compression algorithm might look like. This should give us a better feel for Shannon entropy.

First, we'll take some limits. If all the probabilities are equal with $P_i = \frac{1}{N}$ then

$$H = -\sum_i \frac{1}{N} \log_2 \frac{1}{N} = \log_2 N \tag{30}$$

This is just the number of bits to encode N letters. I.e. if $N = 32$, it takes 5 bits, if $N = 64$ it takes 6 bits and so on.

If one of the probabilities is zero, then the effect on H is $\lim_{P \to 0} P \ln P = 0$. So adding something else to the data that never occurs does not affect the entropy. Conversely, if the data is completely uniform, so $P = 1$ for some character, then $H = -1 \log_2 1 = 0$. So it takes zero bits to encode a completely determined sequence.

Say we have a fair coin with 50% heads and 50% tails probabilities. Then

$$H = -\left(\frac{1}{2}\log_2 \frac{1}{2} + \frac{1}{2}\log_2 \frac{1}{2}\right) = 1 \tag{31}$$

So it takes one bit to encode each flip. This is the best we can do. If the coin has 2 heads and no tails, then $H = 0$: we know it will always be heads.

Now let's look at a more complicated example. Say the coin is weighted so it is 90% likely to get heads and 10% likely to get tails. Then we find that

$$H = -(0.9 \log_2 0.9 + 0.1 \log_2 0.1) = 0.468 \tag{32}$$

So it is inefficient to encode each coin flip with just 0 or 1. What we really want is a code that uses less than a bit for heads and more than a bit for tails. How could we do this? One way is to say 0 means two heads in a row, 10 means heads and 11 means tails:

HH	H	T
0	10	11

(33)

This lets us represent a sequence of two flips with one bit if it's , and with 4 bits otherwise. Thus for example, the sequence HHHHHHHHTHHH, with 12 digits, becomes 000011100, with 9 digits. For all possible 2 flip sequences, we find:

Sequence	HH	HT	TH	TT
Probability	81%	9%	9%	1%
Code	0	1011	1110	1111
#bits	1	4	4	4

(34)

The expected number of bits needed to encode a 2 flip sequence is the number of bits in the code (1 or 4) times the probabilities, namely $\#\text{bits} = 1\times 0.81 + 4\times 0.09 + 4\times 0.09 + 4\times 0.01 = 1.57$. So instead of 2 bits, we are using 1.57 on average, corresponding to an entropy per flip of $\frac{1.57}{2} = 0.785$. This is not as good as 0.468, but it is better than 1. In other words, our algorithm compresses the data, but not optimally.

Check your understanding. What is the entropy per flip in the following encoding?

HH	HT	TH	TT
0	10	110	111

(35)

Can you think of a better compression algorithm?

4.3 Uniqueness

You might like to know that Shannon's formula for information entropy is not as arbitrary as it might seem. This formula is the unique function of the probabilities satisfying three criteria

1. It does not change if something with $P_i = 0$ is added.
2. It is maximized when P_i are all the same.
3. It is additive on uncorrelated probabilities.

You can find the proof of the uniqueness of Shannon's formula in various textbooks on information theory, but I'm not going present it here since it's not particularly illuminating.

This last criteria needs a little explanation. First, let's check it. Say we have two sets of probabilities P_i and Q_j for different things. For example, P_i could be the probability of having a certain color hair and Q_j the probability for wearing a certain size shoe. If these are uncorrelated, then the probability of measuring i and j is $P_i Q_j$. So the total entropy is

$$H_{\text{PQ}} = -\sum_{i,j} P_i Q_j \log_2(P_i Q_j) = -\sum_i \sum_j Q_j P_i \log_2(P_i) - \sum_j \sum_i P_i Q_j \log_2(Q_j) \tag{36}$$

Doing the sum over j in the first term or i in the second term gives 1 since Q_j and P_i are normalized probabilities. Thus

$$H_{\text{PQ}} = -\sum_i P_i \log_2(P_i) - \sum_j Q_j \log_2(Q_j) = H_P + H_Q \tag{37}$$

In other words, entropy is extensive. This is the same criterion Gibbs insisted on. So Gibbs entropy is also unique according to these criteria.

By the way, there are other measures of information entropy other than Shannon entropy, such as the Renyi entropy

$$H_{\text{Renyi}}^{\alpha} = \frac{1}{1-\alpha} \log_2 \left(\sum P_i^{\alpha} \right) \tag{38}$$

The Shannon entropy is the limit as $\alpha \to 1$ of the Renyi entropy (check it!). Other limits give the Hartley entropy or the collision entropy. These other entropies do not satisfy the conditions 1-3 above. Instead, they satisfy some other conditions and consequently they have different applications and interpretations. For example, the second order Renyi entropy ($\alpha = 2$) is $H_{\text{Renyi}}^{\alpha} = -\ln \text{Tr} \rho^2$ with ρ the density matrix, is related to the purity of a quantum system. See Section 7 for a longer discussion of quantum information. In this course, when we discuss information entropy, we will mean Shannon entropy unless otherwise stated.

5 Information entropy to thermodynamic entropy

One way to connect information entropy to thermodynamics is to say that **entropy measures uncertainty**. For example, suppose you have gas in a box. In reality, all of the molecules have some velocities and positions (classically). If you knew all of these, there would be only one microstate compatible with it and the entropy would be zero. But the entropy of the gas is not zero. It is nonzero because we *don't know* the positions and velocities of the molecules, even though we could in principle. So entropy is not a property of the gas itself but of our knowledge of the gas.

In information theory, if we flip some biased coins that always come out heads, then the Shannon entropy is 0; because we know exactly what will come next – a head – so our uncertainty is zero. If the coin is fair and half the time gives heads and half the time tails, then $H=1$: we are maximally ignorant. We know nothing about what happens next. If the coin is unfair, 90% chance of heads, then we have a pretty good sense of what will happen next, but are still a little uncertain. If we know the data ahead of time (or the sequence of flips), we can write a simple code to compress it: 1 = the data. So there is no ignorance in that case, as with knowing the position of the gas molecules.

With the uncertainty idea in mind, we can make more direct connections between information theory and thermodynamics.

5.1 Gibbs = Shannon entropy

The easiest way to connect information entropy to thermodynamic entropy is simply by interpreting microstates, and their associated probabilities, as the data. In general, suppose that the probability of finding a system in a given microstate is P_i. Then we can compute a thermodynamic entropy by multiplying the information entropy by a constant (recalling the relation $\frac{\ln x}{\ln 2}=\log_2 x$)

$$k_B(\ln 2)H=-k_B\ln 2\sum_i P_i\log_2 P_i=-k_B\sum_i P_i\ln P_i=S \tag{39}$$

This is the Gibbs entropy from Eq. (6). Note that from the information theory point of view, the bits are necessarily indistinguishable (if a bit had a label, it would take more than one bit!), so it makes sense that the information entropy leads to Gibbs entropy.

What values of P_i will maximize S? Given no other information of constraints on P_i, the postulate of equal a priori probabilities (or the principle of maximum entropy) gives that the probability of a microstate i is $P_i=\frac{1}{\Omega}$ with Ω the total number of microstates. In terms of information theory, if the P_i are equal, it means that the data is totally random: there is equal probability of finding any symbol. Thus there should be no way to compress the data at all. The data can be compressed only if there is some more information in the probabilities. So the minimal information leads to the maximum entropy. With $P_i=\frac{1}{\Omega}$ the entropy is

$$S=-\sum_{j=1}^{\Omega}k_B\left[\frac{1}{\Omega}\ln\frac{1}{\Omega}\right]=k_B\ln\Omega \tag{40}$$

which is of course the original Boltzmann entropy formula in Eq. (3).

For another connection, consider the free expansion of a gas from a volume V to a volume $2V$. The change in entropy is $\Delta S=k_B N\ln 2$ or equivalently, $\Delta H=\frac{1}{k_B\ln 2}\Delta S=N$. So the number of bits we need to specify the system has gone up by N. But that's exactly what we should have expected: each bit says which of the 2V volumes each particle is in, so we need N more bits to specify the system.

5.2 Irreversibility of information storage

We made an abstract connection between Gibbs entropy and information entropy. Actually, the connection is not just formal, they are actually the same thing. To see this, we need a little bit of the physics of computation.

A key observation was made by Rolf Landauer in 1961. Landauer was interested in making powerful computers that used as little energy as possible. How little energy could they use? He considered a model of a bit with a (classical) double well. A ball on the left was 0 and a ball on the right was 1.

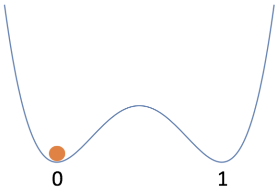


Figure 4. A model of information storage where a ball on the left is 0 and a ball on the right is 1.

The first question is whether we can change the bit from 0 to 1 without using any energy. It seems that the answer is yes. For example, we could hook a line to the ball and tie it to a counterweight with some pulleys on a frictionless rod:

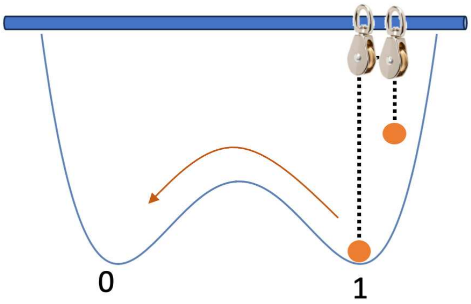


Figure 5. Hooking a pulley to the bit, we can move it from 1 to 0 without doing any work.

We just have to give the ball an infinitesimal nudge and it will roll across, then another infinitesimal nudge will stop it. So no energy is needed to flip this bit. This action is adiabatic and reversible, and can be also used to set 0 to 1 by running in reverse. Even if we need a little energy to actually do it, we get the energy back, so can flip 1000 bits with the same net energy cost as for one bit.

So if the bit is 0 it takes no energy to set it to 1 and if the bit is 1 it takes no energy to set it to 0. But what if we don't know what state the bit is in? Here's where it gets interesting. Suppose you had some automated mechanism for "SetToZero" to take the bit from 1 *or* 0 to 0. This is the kind of operation computers need to do all the time. Can we use our pulley gizmo to do it? The answer is no. In fact, the answer is no for any gizmo and any way of representing a bit. The reason as that we are just using Newton's laws, which are time-reversible. So if whatever action we do must be some some kind of 1-to-1 invertible function F acting on the position and momenta of the stuff in the bit. If phase space point $(\vec{q}_i^0, \vec{p}_i^0)$ represents 0 and point $(\vec{q}_i^1, \vec{p}_i^1)$ represents 1, then we want $F(\vec{q}_i^0, \vec{p}_i^0) = (\vec{q}_i^0, \vec{p}_i^0)$ and $F(\vec{q}_i^1, \vec{p}_i^1) = (\vec{q}_i^0, \vec{p}_i^0)$. But this is impossible if F is invertible. This argument is very rigorous and holds even in quantum mechanics, since the Schrödinger equation can also be run backwards in time.

Now, computers are very good at SetToZero. How do they do it? If we are allowed to dissipate energy, it's easy. For example, if there is friction in our double well system, then SetToZero could be "swing a mallet on the 1 side with enough energy to knock a ball over the hill." If there is no ball on the 1 side, the this does nothing $0 \to 0$. If there is a ball on the 1 side, it will go over to 0 and then settle down to the minimum due to friction, $1 \to 0$. Note that without friction this wouldn't work, since the ball would come back to 1. So heat *must* be generated. If we don't know the initial states then to flip 1000 bits would take, at minimum, 1000 times the minimum energy needed to flip 1 bit. In a real computer, the information might be stored in the spin of a magnet on a magnetic tape. Applying a field to flip the bit would release energy if it flips which would then dissipate as heat. No matter how you cut it, we find

- **Landauer's principle**: erasing information requires energy be dissipated as heat.

Erasing information is an essential step in computation. Every time we store information, we erase the information that was previously there. But it not the storing of information but the erasing, the throwing out of information, that dissipates heat. That was Landauer's critical insight.

The key element to showing that SetToZero on an unknown bit is impossible without dissipation was the reversibility of the laws of physics. Erasing information cannot be done with a reversible process. Thus thermodynamic entropy increases when information is thrown out.

To be absolutely clear, strictly speaking the information is not really lost. The laws of physics are still reversible, even with friction, so the final state could be run backwards to get the initial state. The final state however requires not just knowing the bit we are interested in, but all the positions and momenta of all the particles carrying off the heat. If we only record the bit, we are averaging over all possible states of the other stuff. It is in that averaging, that purposeful forgetting, where the information is actually lost. Dissipation into thermal energy implies this purposeful forgetting. Coarse graining erases information.

5.3 Energy from information

The connection between information entropy and thermodynamics was pushed further by Charles Bennett in the 1980s. Bennett was very interested in how much energy computers require, in principle. That is, what are the fundamental limits on computing determined by thermodynamics?

The first relevant observation is that information itself can be used to do work. The setup Bennett considered was a digital tape where the information is stored in the position of gas molecules. We say the tape is made up of little cells with gas molecules either in the bottom of the cell, which we call 0, or in the top of the cell which we call 1. Imagine there is a thin dividing membrane between the top and bottom keeping the molecule from moving from top to bottom.

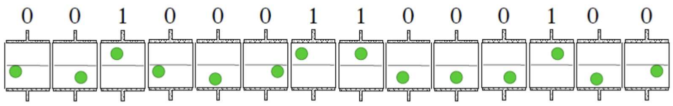


Figure 6. Information is stored in the position of gas molecules.

Let us keep this molecular tape in thermal contact with a heat bath so that the molecules are always at constant temperature. By fixing the temperature, we fix the molecule's momentum. Conversely, if we had allowed the temperature to vary, then the momentum would be variable too, and there would be more degrees of freedom that just the single bit represented by position.

Now, for an individual cell, if we know whether it's 0 or 1, we can use that information to do work. Bennett proposed putting pistons on both the top and bottom of each cell. Say the molecule is on the bottom. If so, we lower the top piston to isolate the molecule on the bottom half. This doesn't cost any energy. Then we remove the dividing membrane and let the thermal motion of the molecule slowly push the piston up, drawing heat from the heat bath and doing useful work:

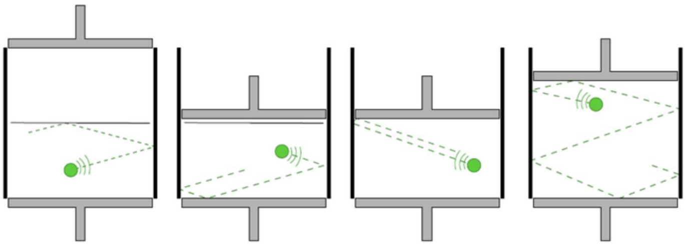


Figure 7. Extracting work from information.

If the molecule were in the top, we would move the bottom piston to the middle, remove the membrane, and let the system do work pushing it back down. Either way, the final state of the system has the gas molecule bouncing around the whole cell so we have lost the information – we don't know if it is 0 or 1 – but we have done work. This way of getting work out of information is known as **Szilard's engine**.

Once the piston is open and the molecule is free to bounce around the whole container, we have lost the information. We can then set the bit to 0 (or 1) by pushing down (up) on the gas with the piston. The work we do during this compression goes off into the thermal bath as heat. This is the SetToZero operation again that we discussed in Section 5.2. Just like there, acting on an unknown bit SetToZero dissipates energy. Once the bit is set, we can then use it to do work. But the work we get out is the same as the work we put in to set the bit. So we cannot do useful work if we do not know the state.

Let's make this a little more precise. Suppose the molecule is an ideal gas, so $PV = k_B T$. Let's also put the engine in a heat bath and extract the work isothermally. Then the work done is

$$W = \int P dV = k_B T \int \frac{dV}{V} = k_B T \ln \frac{V_2}{V_1} = k_B T \ln 2 \tag{41}$$

The heat absorbed from the bath is then $Q_{\text{in}} = k_B T \ln 2$. This is the maximum amount of work you can get from a bit of information. If we run the system in reverse when we don't know the state of the system, we need to use this amount of work (on average) to set the bit, so the amount of heat dissipated into the bath is $Q = k_B T \ln 2$. This is therefore minimal amount of heat dissipation from SetToZero. In summary

- Setting a bit erases information and dissipates at least $Q = k_B T \ln 2$ in heat
- The minimum amount of work it takes to set a bit is $W = k_B T \ln 2$

The entropy cost of losing the information, and of not being able to do work, is the entropy increase in doubling the volume available to the molecule

$$\Delta S_{\text{sys}} = k_B \ln \frac{V_2}{V_1} = k_B \ln 2 \tag{42}$$

or equivalently

$$\Delta H = 1 \tag{43}$$

The information entropy goes up by one bit because we have lost one bit of information – the position of the molecule. In this isothermal quasistatic expansion, the entropy change of the bath is $\Delta S_{\text{bath}} - \frac{Q_{\text{in}}}{T} = -k_B \ln 2$, so overall $\Delta S_{\text{tot}} = 0$ since we have extracted the information reversibly.

To be clear, when we say information can be used to do work, we don't mean the first law of thermodynamics is violated. Really, it is the thermal motion in heat that is actually doing the work. Szilard's engine can convert heat directly into work without violating the second law because entropy decrease of the gas is compensated for by an entropy increase from erasing information of the initial state of the system.

6 Maxwell's demon

We're now ready to tackle the most famous paradox about entropy, invented by Maxwell in 1867. Suppose we have a gas of helium and xenon, all mixed together in a box. Now say a little demon is sitting by a little shutter between the two sides of the box. When he sees a helium molecule come in from the right, he opens a little door and lets it go left. But when it's a xenon molecule coming in from the right, he doesn't open the door.

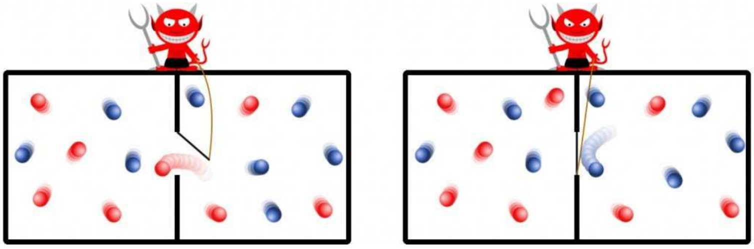


Figure 8. Fig 2: Maxwell's demon lets helium through, but not xenon

After a while, enough helium will be on the left that it will increase the pressure and this can be used to do work. The demon has made the system more ordered and the entropy has gone down.

Your first thought might be that the resolution to this puzzle has to do with identical particles. But this is not true. The paradox holds for pure helium. If the demon lets helium molecules go left and not right, the entropy would go down. Moving the door up and down doesn't take any work (it can be moved with an arbitrarily small push, and moreover the work can be completely recovered by stopping it), yet when the helium gets concentrated on one side it will exert a pressure on the barrier that can be used to do work. So this little demon is converting heat directly into work at constant temperature. In Maxwell's original formulation, the demon would only let the fastest molecules through one way and the slow ones the other way, so that the temperature difference of the two sides would increase. These are all different ways of saying the second law of thermodynamics is violated.

The demon doesn't have to be alive either. A robot could do his job, governed by the laws of physics. You just have to program him to tell xenon from helium or fast from slow. So consciousness doesn't have anything to do with it (although people sometimes like to say it does). For a mechanical example, say you had a little door with a very weak spring on it that only opens one way. If a molecule hits it from the left it will open and let the molecule through, but will not open when hit form the right (do you think this would really work?).

Maxwell's demon has exasperated generations of physicists, for over 100 years. In the 1920s the great physicists Szilard and Brillouin argued that it must take the robot some energy to find out which way the gas is going. The robot must shine light at least one photon of light on the molecule or something equivalent. The energy of this photon will then dissipate as heat increasing the entropy, so the total entropy of the demon/gas system would not go down. While it is true that doing the measurement with light does use energy and increase the entropy, it is possible to make a measurement using an arbitrarily small amount of energy, for example with an arbitrarily low frequency photon.

The correct resolution to Maxwell's demon is that somewhere in the process of letting the molecules pass through from left to right, the robot has to ask: is the molecule on the left? He must store the answer to this question in some variable in his program, somewhere. So he must set a bit, using the "SetToZero" operation. This operation takes work, at least as much work as we get out from moving the molecule to the right side. In terms of information, we gain one bit of information by identifying the particle as left/right, xenon/helium or fast/slow. But we also erase one bit of information by using SetToZero, sending heat into the surrounds. So the net effect is $\Delta S = 0$. Entropy does not go down and there is no contradiction with the second law.

You might instead suppose that we knew the bit on our tape was 0 to begin with. Then recording the position of the molecule with $0 \to 0$ or $0 \to 1$ does not require heat dissipation. In this case, it seems that Maxwell's demon does violate the second law. Note, however, that as we write the (random) locations of the molecules to our tape, our tape randomizes. So we are just moving the disorder from our gas into the disorder of the tape. In other words, $\Delta S_{\text{gas}} = -\Delta S_{\text{tape}}$ so the net entropy increase is still zero.

In this way Maxwell's demon was resolved by Bennett in 1982. after 115 years of confusion. As Feynman says in his Lectures on Computation (p. 150)

> This realization that it is the erasure of information, and not measurement, that is the source of entropy generation in the computational process, was a major breakthrough in the study of reversible computation.

Maxwell's demon clarifies why information entropy really is entropy. If we had a tape of 0's, then the demon could indeed make a finite amount of heat flow from a hot bath to cold bath. As Bennett puts it, a tape of 0's has "fuel value" that can be used to do work. So we must include the specification of this tape as part of the definition of the system. Once we do so, then the entropy never decreases, it just moves from the disorder of the molecules in the box to the disorder of the information on the tape. Thus the entropy for which $\Delta S \geqslant 0$ is strictly true should include both thermodynamic and information-theoretic entropy.

Finally, it is worth returning briefly to Gibbs paradox, from the information entropy point of view. Recall that Gibbs paradox was that when you separate a box of distinguishable particles in half, the entropy seems to go down. The resolution, as discussed in depth in Section 3.3, was that there are $\Omega_{\text{split}} = \binom{2N}{N}$ ways of separating the distinguishable particles. If we include Ω_{split} in the counting of microstates, then entropy does not go down. Now suppose we look at the balls, so we know which balls go where. Then there is no more Ω_{split}. Has entropy gone down? No! To find out which ball went where, we had to look at the color of each ball. Equivalently, we have to measure for each ball which side it's on. Each such measurement must SetToZero some bit in whatever we're using to do the measurement. The entropy consumed by this measurement is exactly the entropy lost by the system. The unifying viewpoint that resolves both Gibbs paradox and Maxwell's demon is that you must always include all forms information in the assessment of entropy changes.

6.1 Proof of the second law of thermodynamics

The information theory definition of entropy allows us to finally prove the second law of thermodynamics. We start with some system, specified with a given set of information (P, V, E or which colored balls are where or the initial condition for our tape etc). The initial entropy is $S = k_B \ln 2H$. In order for entropy to go down, we need to acquire more information about the system, so that fewer states are consistent with what we know. So we have to make some sort of measurement on the system. Since the laws of physics are deterministic, for each possible state of the system, the measurement will have a different outcome. Thus the unknown elements of the system translate directly into unknown elements of our measuring device. You might think that after the measurement we know the result, so the state of our measuring device is known. However, the state that we measured is *only* stored on our measuring device. Thus the device becomes just as unknown as the original system and we've just transported uncertainty from one place to another.

7 Quantum mechanical entropy (optional)

This section requires some advanced appreciation of quantum mechanics. It's not a required part of the course, but some students might find this discussion interesting.

In quantum mechanics, distinguishability takes a more fundamental role, as does measurement. Thus, naturally, there are additional ways to quantify entropy in quantum mechanics. These all involve the density matrix ρ. Recall that in quantum mechanics, the states of the system are linear combinations of elements $|\psi\rangle$ of a Hilbert space. You may know exactly what state a system is in, in which case we say that the system is in a pure state $|\psi\rangle$. Alternatively, you may only know the probabilities P_i that the system is in the state $|\psi_i\rangle$. In such situations, we say that the system is in a mixed ensemble of states. The density matrix is defined as

$$\rho = \sum P_j |\psi_j\rangle\langle\psi_j| \tag{44}$$

The **von Neumann entropy** is defined as

$$S = -k_B \mathrm{Tr}[\rho \ln\rho] \tag{45}$$

Because S is defined from a trace, it is basis independent. Of course, we can always work in the basis for which ρ is diagonal, $\rho = \sum P_i |\psi_i\rangle\langle\psi_i|$, then $\rho \ln\rho$ is diagonal too and

$$S = -k_B \sum_j \langle\psi_j|\rho \ln\rho|\psi_j\rangle = -k_B \sum_j P_j \ln P_j \tag{46}$$

In agreement with the Gibbs entropy. Thus, in a pure state, where $P_j = 1$ for some j and $P_j = 0$ for everything else, $S = 0$. That is, in a pure state we have no ignorance. The von Neumann entropy therefore gives a basis-independent way of determining how pure a state is.

For example, say we start with a pure state $|\psi\rangle = |\rightarrow\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$. This has $P_1 = 1$ and so $S = 0$. Now say we measure the spin along the z axis, but don't record the result. Then the system is either in the state $|\psi\rangle = |\uparrow\rangle$ with probability $P_1 = \frac{1}{2}$ or the state $|\psi\rangle = |\downarrow\rangle$ with $P_2 = \frac{1}{2}$. The density matrix is therefore

$$\rho = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \tag{47}$$

and the entropy is $S = k_B \ln 2$. The entropy has gone up since the measurement has collapsed the wavefunction from a pure state to a mixed state. We no longer know what the state is exactly, so our ignorance has gone up.

The von Neumann entropy also gives a useful way to quantify correlations. In quantum mechanics correlations among different particles are encoded through entanglement. For example, if there are two electrons, possible states have their spins aligned $|\psi\rangle = |\uparrow\uparrow\rangle$, anti-aligned $|\psi\rangle = |\uparrow\downarrow\rangle$, or entangled, $|\psi\rangle = |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle$. To quantify entanglement in general, let us suppose our Hilbert space has two subspaces A and B, so $H_{AB} = H_A \otimes H_B$. Then we can compute a reduced density matrix for a subspace A by tracing over B, and for B by tracing over A

$$\rho_A = \mathrm{Tr}_B[\rho], \quad \rho_B = \mathrm{Tr}_A[\rho] \tag{48}$$

The von Neumann entropies of the reduced density matrices

$$S_A = -k_B \mathrm{Tr}[\rho_A \ln\rho_A], \quad S_B = -k_B \mathrm{Tr}[\rho_B \ln\rho_B] \tag{49}$$

are called the **entanglement entropies** of the subspaces.

For example, consider the system in pure state

$$|\psi\rangle = \frac{1}{2}(|\uparrow\rangle_A + |\downarrow\rangle_A) \otimes (|\uparrow\rangle_B + |\downarrow\rangle_B) = \frac{1}{2}\Big[\, |\uparrow\uparrow\rangle + |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle + |\downarrow\downarrow\rangle \Big] \tag{50}$$

Because the state is pure, the density matrix $\rho = |\psi\rangle\langle\psi|$ has zero von Neumann entropy. The density matrix for A is

$$\rho_A = \mathrm{Tr}_B(\rho) = \langle\uparrow|_B\, \rho |\uparrow\rangle_B + \langle\downarrow|_B\, \rho |\downarrow\rangle_B \tag{51}$$

$$= \frac{1}{2}(|\uparrow\rangle_A + |\downarrow\rangle_A)(\langle\uparrow|_A + \langle\downarrow|_A) \tag{52}$$

$$= \frac{1}{2}(|\uparrow\rangle\langle\uparrow| + |\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|) \tag{53}$$

This is the density matrix for a pure state, $|\psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle_A + |\downarrow\rangle_A)$, so $S_A = 0$. Similarly, $S_B = 0$. There is no entanglement entropy.

Now consider the system in an entangled pure state

$$\psi = \frac{1}{\sqrt{2}}\Big[\, |\uparrow\rangle_A \otimes |\downarrow\rangle_B + |\downarrow\rangle_A \otimes |\uparrow\rangle_B \Big] = \frac{1}{\sqrt{2}}\Big[\, |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle \Big] \tag{54}$$

Then, $S=0$ since $\rho=|\psi\rangle\langle\psi|$ is still based on a pure state. Now the reduced density matrix is

$$\rho_A=\mathrm{Tr}_B(\rho)=\frac{1}{2}\Big[\ |\uparrow\rangle_A\langle\uparrow|_A+|\downarrow\rangle_A\langle\downarrow|_A\Big]=\left(\begin{array}{cc}\frac{1}{2} & 0\\ 0 & \frac{1}{2}\end{array}\right) \tag{55}$$

This is is now a mixed state $P_1=\frac{1}{2}$ and $P_2=\frac{1}{2}$. Thus $S_A=k_B\ln 2$. So when we start with an entangled state, there is entanglement entropy. Tracing over B amounts to throwing out any chance of measuring B. By doing so, we cannot exploit the entanglement anymore, so the information is lost and entropy goes up.

We can think of the whole universe as being described by a single wavefunction evolving in time. It's a pure state with entropy zero. Everything is entangled with everything else. As it becomes practically impossible to exploit that entanglement, exactly like it was impossible to exploit the correlations among scattered molecules classically, we coarse grain. Coarse graining in quantum mechanics means tracing over unmeasurable components. This increases the entropy and moreover turns a pure state into a mixed state. In this way, classical probabilities emerge from a completely deterministic quantum system.

Another aspect of quantum mechanics that entropy sheds light on is the no-cloning theorem. This theorem states that no unitary operator can evolve the state $|\phi\rangle|\psi\rangle\rightarrow|\phi\rangle|\phi\rangle$ for all possible states $|\phi\rangle$ and $|\psi\rangle$. The theorem follows imply from the fact that unitary operators are invertible, so they cannot map multiple different states to the same state. In other words, the evolution $|\phi\rangle|\psi\rangle\rightarrow|\phi\rangle|\phi\rangle$ is not reversible: cloning a state necessarily involves erasing the information stored in $|\psi\rangle$, so entropy must increase. This entropy increase provides a fundamental limit on how efficient error-correcting quantum computers can become.

In summary, von Neumann entropy lets us understand both the information loss by measurement and by losing entanglement. Entanglement is the quantum analog of correlations in a classical system. Discarding this information is the reason quantum systems become non-deterministic and entropy increases. We don't have to discard the information though. In fact, figuring out how to exploit the information stored in entanglement is critical to the function of quantum computers.

8 Black hole entropy (optional)

This section requires some advanced appreciation of general relativity. It's not a required part of the course, but some students might find this discussion interesting.

Using general relativity, you can prove some interesting results about black holes. General relativity is described by Einstein's equations, which are like a non-linear version of Maxwell's equations. Instead of $\partial_\mu F_{\mu\nu}=J_\nu$ we have

$$R_{\mu\nu}-\frac{1}{2}Rg_{\mu\nu}=T_{\mu\nu} \tag{56}$$

The right-hand side of this equation, $T_{\mu\nu}$ is the energy-momentum tensor, which is the source for gravitational radiation like the current J_μ is the source for electromagnetic radiation. The object $R_{\mu\nu}$ is called the Ricci curvature, it is constructed by taking 2 derivatives on the metric $g_{\mu\nu}$ in various combinations: $R_{\mu\nu}=\partial_\mu\partial_\alpha g_{\alpha\beta}+\partial_\mu g_{\nu\alpha}\partial_\alpha g_{\alpha\gamma}+\cdots$. So $g_{\mu\nu}$ plays the role that the vector potential A_μ plays in Maxwell's equations where $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu$.

If we set $J_\mu=0$, there is a spherically-symmetric static solution to Maxwell's equations, $A_0=\frac{e}{4\pi\epsilon_0 r}$ and $\vec{A}=0$. This is the Coulomb potential. It has one free parameter e. The Coulomb potential is singular at $r=0$ indicating that there is some charge localized there. In fact, this solution corresponds to a current which is zero everywhere but the origin: $\vec{J}=0$ and $J_0=e\delta(\vec{x})$. In Newtonian gravity, the spherically-symmetric static solution is the Newtonian potential $\Phi=-G\frac{M}{r}$. with G Newton's constant.

In general relativity, the spherically-symmetric static solution to Einstein's equations is

$$g_{00}=\Big(1-\frac{r_s}{r}\Big)c^2,\quad g_{rr}=\frac{1}{1-\frac{r_s}{r}},\quad g_{\theta\theta}=r^2,\quad g_{\phi\phi}=r^2\sin^2\theta \tag{57}$$

and $g_{ij}=0$ for $i\neq j$. This solution, called the Schwarzschild solution, describes a black hole. This solution is unique up to a single parameter r_s called the **Schwarzschild radius**. Its uniqueness implies black holes have no "hair", meaning that every black hole is identical to an external observer (up to possible conserved charges like electric charge which can be seen through electric field lines ending at the black hole). Note that the solution is singular not only at $r=0$ but also at $r=r_s$.

In the non-relativistic limit, general relativity reduces to Newtonian gravity. The precise correspondence is that $g_{00}=1+2\Phi$. Matching on to $\Phi=-G\frac{M}{r}$ lets us relate the parameter r_s in the solution to the black hole mass M:

$$r_s=2\frac{MG}{c^2} \tag{58}$$

The spherical surface at $r=r_s$ is called the event horizon. It turns out that nothing inside the event horizon can ever escape. The size (surface area) of the event horizon is

$$A=4\pi r_s^2=16\pi\frac{M^2G^2}{c^4} \tag{59}$$

Classically, things only fall in to a black hole, so their energy only goes up, and therefore the area of the event horizon only increases.

Because the potential is singular on the event horizon, unusual things can happen. One such thing is that due to quantum field theory the infinite potential energy can be turned into photons produced that radiate outwards. Stephen Hawking showed that the spectrum of these photons is identical to a hot gas (a blackbody, to be covered in Lecture 12) at temperature

$$T=\frac{\hbar c^3}{8\pi GMk_B} \tag{60}$$

This Hawking temperature is inversely proportional to the mass: very small black holes are very hot, and very large black holes are cold. This unusual behavior is associated with a negative heat capacity. Indeed, the specific heat of a black hole is

$$c_S=\frac{1}{M}\frac{\partial M}{\partial T}=-\frac{1}{T}=-\frac{8\pi Gk_B}{\hbar c^3}M<0 \tag{61}$$

As things fall into a black hole, its mass goes up and its temperature goes down. A solar mass black hole has a temperature $T=10^{-8}K$. A supermassive black hole, like Sagittarius $A^\star$ in the center of our galaxy is about 1 million solar masses and has $T=10^{-14}K$.

If nothing falls into a black hole, the black hole will completely evaporate due to Hawking radiation in finite time

$$t_{\text{evap}}=5120\pi\frac{G^2M^3}{\hbar c^4} \tag{62}$$

As a black hole evaporates, its mass goes down and its temperature goes up. The bigger the black hole, the longer it takes to evaporate. A solar-mass black hole would take 10^{74} years to evaporate. An atomic mass black hole would evaporate in 10^{-98} seconds.

You probably know that the universe is filled with cosmic microwave background (CMB) radiation at a temperature of $3K$. A black hole radiating at this temperature has mass $M_{3K}=10^{22}$kg, around the mass of the moon. Black holes less massive than M_{3K} will be hotter than the CMB and therefore radiate more energy than they absorb, eventually evaporating. Black holes more massive than M_{3K} will be colder than the CMB; these will absorb CMB radiation, and their mass will increase over time, and their temperature will decrease. Thus these black holes will only grow and never evaporate. If a black hole has mass exactly equal to M_{3K}, it can be in equilibrium with the CMB. However this equilibrium is unstable: a small fluctuation will send it's temperature increasing or decreasing. This lack of a stable equilibrium is a typical behavior of systems with negative heat capacity (including all gravitational systems, like stars or galaxies).

Black holes also have entropy. Since $\frac{\partial S}{\partial E}=\frac{1}{T}$, taking the energy of a black hole as its rest mass, $E=Mc^2$ the entropy is

$$S=\int\frac{dE}{T}=\frac{8\pi G}{\hbar c}\int MdM=\frac{4\pi G}{\hbar c}M^2=\frac{c^3}{4\hbar G}A \tag{63}$$

Note that black holes have entropy proportional to their surface area. String theory even provides a way of counting microstates for certain supersymmetric black holes that agrees with this formula.

So black holes have entropy, but no hair, and they evaporate in finite time into pure uncorrelated heat. This means that if some data falls into a black hole, it is lost forever. In this way, black holes destroy information and radiate it out as heat, much like Landauer or Bennett's SetToZero operation. There is one important difference though. When we "SetToZero" a bit, the information is not destroyed, just lost by being embedded irretrievably in correlations in the heat. We know this because the law of physics are reversible. When a black hole destroys information it really destroys it – it cannot be stored in correlations of the outgoing radiation because nothing can get out of a black hole, including information. This is the **black hole information paradox**.

To see this another way, information can fall into a black hole well before the radiation is emitted. Since black holes have no hair, that information cannot be accessed in any way by an external observer. For example, suppose we just throw some books into a black hole, at such a rate that the energy input exactly equals the thermal radiation rate. Then the black hole's horizon stays constant so the information must be going out in the radiation. However, this is impossible since once something passes the black hole horizon, it can never affect anything outside the horizon. Thus the information really seems to be lost as it falls into the black hole.

The basic conflict is that the laws of gravity and quantum mechanics are deterministic and reversible – if we know the exact starting state, we should be able to predict the exact final state. The precise statement is that in quantum mechanics and gravity, as well as in string theory, time evolution is unitary. Information cannot be lost in a closed, reversible, unitary theory.

The conflict between unitarity and black hole evaporation can be understood clearly with von Neumann entropy. Say the initial state is a wavefunction describing two electrons moving towards each other at super high energy. This is a pure state. They then collide to form a black hole. The black hole then evaporates and the information leaves as heat. The entropy goes up, so the outgoing state is mixed. Thus black holes mediate the evolution from a pure state into a mixed state. This is in conflict with Schrödinger's equation, or more generally, any theory with unitary evolution (such as string theory). If unitarity can be violated by black holes, then it would contribute through virtual effects in quantum field theory to unitarity violation in every other process, in conflict with observation.

9 Summary

We have seen a lot of different ways of thinking about entropy this lecture. The Gibbs entropy is

$$S = -k_B \sum P_i \ln P_i \tag{64}$$

Here P_i is the probability of finding the system in a microstate i and the sum is over all possible microstates i consistent with some macroscopic parameters (volume pressure etc.). In equilibrium, this definition is equivalent to $S = k_B \ln \Omega$ with Ω the number of microstates, but Eq. (64) can be used in any situation where the probabilities are well-defined, including time-dependent non-equilibrium systems.

This Gibbs entropy is proportional to the (Shannon) information entropy:

$$H = -\sum P_i \log_2 P_i \tag{65}$$

In this equation, P_i is the probability of certain data showing up. H has the interpretation as the minimal number of bits needed to encode the data, on average. Information entropy quantifies our ignorance of the system. The more entropy, the less information we have.

An important result from information theory is Landauer's principle: erasing information dissipates heat. This connects information to thermodynamics. When 1 bit of information is erased, the information entropy goes up by 1 bit. Doing so is necessarily accompanied by the release of heat which increases the thermodynamic entropy by the same amount. While the information is technically somewhere encoded in the heated-up molecules, we accept that we will never recover this information and forget about it. Thus we increase our ignorance of the state of the whole system and the entropy goes up.

The broader lesson from this lecture is the modern view of entropy is not as a measure of disorder but as a measure of ignorance. Indeed, the information-theoretic point of view unifies all the different forms of entropy and is the cleanest way to resolve the various entropy-related paradoxes (Gibbs paradox, Maxwell's demon, etc.). It's not that there are two kinds of entropy that we must add: counting microstates and information, but rather that *all* entropy measures the lack of information. When we count the number of microstates Ω these are the states that give the same macroscopic parameters. Thus, given only the information E, V, N etc, Ω measures the things we don't know, namely which microstate we have, consistent with our information. The Gibbs and Shannon entropy formulas are equivalent and therefore both measure ignorance as well. Thus, if you get one thing out of this lecture it should be that **entropy** $=$ **ignorance**.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 7: Ensembles

1 Introduction

In statistical mechanics, we study the possible microstates of a system. We never know exactly which microstate the system is in. Nor do we care. We are interested only in the behavior of a system based on the possible microstates it could be, that share some macroscopic property (like volume V, energy E, or number of particles N). The possible microstates a system could be in are known as the **ensemble** of states for a system. There are different kinds of ensembles.

So far, we have been counting microstates with a fixed number of particles N and a fixed total energy E. We defined Ω as the total number microstates for a system. That is

$$\Omega(E,V,N)=\sum_{\substack{\text{microstates k}\\ \text{with same N,V,E}}} 1 \tag{1}$$

Then $S=k_B\ln\Omega$ is the entropy, and all other thermodynamic quantities follow from S. For an isolated system with N fixed and E fixed the ensemble is known as the **microcanonical ensemble**. In the microcanonical ensemble, the temperature is a derived quantity, with $\frac{1}{T}=\frac{\partial S}{\partial E}$. So far, we have only been using the microcanonical ensemble.

For example, a gas of identical monatomic particles has $\Omega(E,V,N)\sim\frac{1}{N!}V^NE^{\frac{3}{2}N}$. From this we computed the entropy $S=k_B\ln\Omega$ which at large N reduces to the Sackur-Tetrode formula. The temperature is $\frac{1}{T}=\frac{\partial S}{\partial E}=\frac{3}{2}\frac{Nk_B}{E}$ so that $E=\frac{3}{2}Nk_BT$. Also in the microcanonical ensemble we observed that the number of states for which the energy of one degree of freedom is fixed to ε_i is $\Omega(E-\varepsilon_i)$. Thus the probability of such a state is $P_i=\frac{\Omega(E-\varepsilon_i)}{\Omega(E)}\sim e^{-\varepsilon_i/k_BT}$. This is the Boltzmann distribution.

Within the context of the microcanonical ensemble, we also derived the Boltzmann distribution using the principle of maximum entropy. This approach is very general. It uses nothing about the system other than that the total number of degrees of freedom N is large and the total energy is E. To use the maximum entropy principle we counted the number of ways that N particles could be allocated into groups of size n_i with energies ε_i, so that $\sum n_i=N$ and $\sum n_i\varepsilon_i=E$. We found that in the most probable allocation of particles to groups, the probability of finding a particle with energy ε_i was

$$P_i=\frac{1}{Z}e^{-\beta\varepsilon_i} \tag{2}$$

where $Z=\sum_i e^{-\beta\varepsilon_i}$ and $\beta=\frac{1}{k_BT}$.

Sometimes we don't know the total energy, but we know the temperature. This situation is in fact much more common than knowing the energy. For example, in the room you are in, what is the total energy of the air molecules? I bet you don't have a clue. But I bet you have a good idea of what the temperature is.

When we fix temperature instead of energy, we have to allow the energy to fluctuate. For example, think of two systems in thermal contact. The thermal contact allows energy to flow in and out of each system, so energy of each system is not fixed. We call a set of microstates with N, V and T fixed but variable E the **canonical ensemble**. In the canonical ensemble, the primary object is not the number of states Ω or the entropy S but rather the **partition function**

$$Z(\beta)=\sum_{\substack{\text{microstates k}\\ \text{with same N,V}}} e^{-\beta E_k} \tag{3}$$

In the partition function, energies of the microstates summed over can vary. Thus the left-hand side, $Z(\beta)$, cannot depend on energy. Instead, it depends on temperature. Once Z is known, it is straightforward to compute the average energy $\langle E\rangle$ and other thermodynamic quantities, as we will see.

In both the microcanonical and canonical ensembles, we fix the volume. We could instead let the volume vary and sum over possible volumes. Allowing the volume to vary gives the **Gibbs ensemble**. In the Gibbs ensemble, the partition function depends on pressure rather than volume, just as the canonical ensemble depended on temperature rather than energy.

In the microcanonical, canonical, and Gibbs ensembles, the number of particles N in the system is fixed. In some situations, we want the number of particles to vary. For example, chemical reactions change the number of each molecule type. So in chemistry we can't fix N. Instead we fix something called the chemical potential, μ. Chemical potential is like a pressure for particle number. Chemical potential is a very important concept, but very difficult to grasp, so we will spend a lot of time understanding it in this lecture and beyond.

When N can vary we use the **grand canonical ensemble**. The main object of interest in the grand canonical ensemble is the **grand partition function**

$$\mathcal{Z} = \sum_{\substack{\text{microstates} k \\ \text{withsame} V}} e^{-\beta E_k} e^{\beta \mu N_k} \tag{4}$$

The grand canonical ensemble is used in chemistry, quantum statistical mechanics, and much of condensed matter physics.

2 Canonical ensemble

In the microcanonical ensemble, we calculated properties of its system by counting the number of microstates at fixed energy. Then, for example, temperature is a derived quantity, $\frac{1}{k_B T} = \frac{\partial \ln \Omega}{\partial E}$. In the canonical ensemble, we fix the temperature T, and the (average) energy becomes the derived quantity.

In order to fix the temperature, it is a useful conceptual trick to imagine our system of interest in thermal contact with a heat reservoir. This means the system and heat reservoir can exchange energy through heat, but no work can be done by the system on the reservoir or vice versa. The point of the reservoir is to make concrete the idea of fixing the temperature and letting the energy fluctuate.

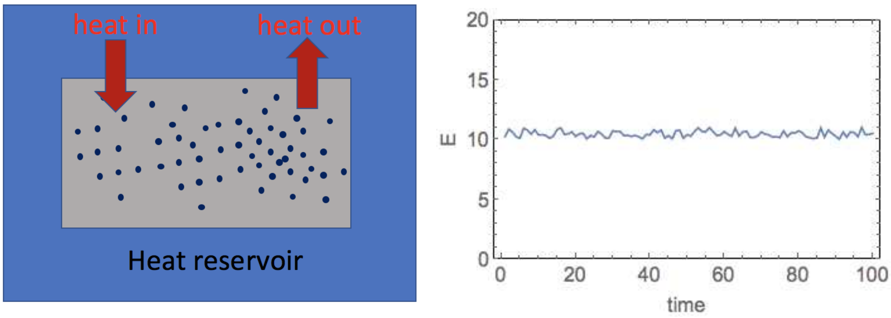


Figure 1. When a system is in thermal contact with a heat reservoir, its temperature is fixed. Its energy fluctuates around its average value.

We do not allow particles to go from the system to the reservoir, only energy. The number of particles in the system can be small – we can have a single atom even – it won't matter. This is important because the canonical ensemble will allow us to discuss systems with a limited number of quantum states, in contrast to the microcanonical ensemble where we really did need to expand at large N to make progress. Although the system can be small, the reservoir must be large, so that it has much much more energy than the system. But this is not a constraint, just a conceptual trick, since the reservoir does not actually need to exist. As we will see, the actual computation of the partition function for the canonical ensemble does not involve the reservoir at all.

We would like to know what is the probability of finding the system in a **fixed microstate $\boldsymbol{k}$** with energy E_k? To be clear: every momentum and position of every particle in k is fixed.

Since the system + reservoir is a closed system, the total energy of the system + reservoir is fixed at E_{tot}. Once we have fixed the microstate k of the system, the total number of states is determined only by properties of the reservoir. More precisely, the probability of finding the system in microstate k is proportional to the number of ways of configuring the system + reservoir with the system in microstate k. Since the total energy is fixed, this number is the same as the number of ways of configuring the reservoir with energy $E_{\text{res}} = E_{\text{tot}} - E_k$:

$$P_k = P_{\text{res}}(E_{\text{res}}) = C \times \Omega_{\text{res}}(E_{\text{res}}) = C \times \Omega_{\text{res}}(E_{\text{tot}} - E_k) \tag{5}$$

for some constant C. $\Omega_{\text{res}}(E_{\text{res}})$ is the number of microstates of the reservoir with energy E_{res}.

Now let us use the fact that $E_k \ll E_{\text{res}} \approx E_{\text{tot}}$, which comes from our assumption of a heat reservoir. We can then expand the logarithm of the number of reservoir states around $E_k = 0$:

$$\ln\Omega_{\text{res}}(E_{\text{tot}} - E_k) = \ln\Omega_{\text{res}}(E_{\text{tot}}) - E_k \frac{\partial \ln\Omega_{\text{res}}(E)}{\partial E}\bigg|_{E = E_{\text{tot}}} + \cdots \tag{6}$$

Next we can use that $\frac{\partial \ln\Omega_{\text{res}}(E)}{\partial E} = \beta = \frac{1}{k_B T}$ in equilibrium[1], so

$$\ln\Omega_{\text{res}}(E_{\text{tot}} - E_k) = \ln\Omega_{\text{res}}(E_{\text{tot}}) - \beta E_k \tag{7}$$

Exponentiating both sides gives

$$\Omega_{\text{res}}(E_{\text{tot}} - E_k) = \Omega_{\text{res}}(E_{\text{tot}}) e^{-\beta E_k} \tag{8}$$

Then by Eq. (5) we have

$$P_k = \frac{1}{Z} e^{-\beta E_k} \tag{9}$$

for some constant $Z = \frac{1}{C \times \Omega_{\text{res}}(E_{\text{tot}})}$. In the canonical ensemble, we will compute Z not using Ω_{res} but by the shortcut that the probabilities sum to 1, $\sum P_k = 1$.

The formula for P_k we found in Eq (9) is the Boltzmann distribution. Note how much quicker the derivation of the Boltzmann distribution is in the canonical ensemble than in the microcanonical ensemble. In the microcanonical ensemble, we had to count all the states, take the logarithm, expand at large N, express E in terms of T, expand for small ε and simplify. Alternatively, we could use the maximum entropy principle, which still required us to split N particles into m groups, work out the combinatoric factors, take N large, insert Lagrange multipliers, then maximize entropy. In the canonical ensemble, we just hook the system up to a reservoir then "bam!" out pops Boltzmann.

The constant Z is called the **partition function**. Using $\sum P_k = 1$, we find

$$Z = \sum_{\text{microstates } k} e^{-\beta E_k} \tag{10}$$

where we sum over all microstates k. Often many microstates have the same energy. We denote the number of states with energy E by g_E, called the **degeneracy**. Then

$$Z = \sum_{\text{energies } E} g_E e^{-\beta E} \tag{11}$$

If the energies are continuous then the degeneracy becomes a function and we can write

$$Z = \int g(E) dE e^{-\beta E} \tag{12}$$

where $g(E)$ is called the **density of states**: $g(E)dE$ gives the number of states with energies between E and $E + dE$. The set of energies of a system along with the density of states is called the **spectrum** of a theory.

[1] This requires taking the thermodynamic limit (large N) for the reservoir. We do not have to take large N for the system since we are purposefully avoiding ever using Ω_{sys}.

The partition function is an amazingly powerful object. If we know it exactly, we can calculate any thermodynamic property of the system. For example,

$$\langle E\rangle = \sum_k E_k P_k = \frac{1}{Z}\sum_k E_k e^{-\beta E_k} = \frac{1}{Z}\left[-\partial_\beta \sum_k e^{-\beta E_k}\right] = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} \tag{13}$$

So

$$\boxed{\langle E\rangle = -\frac{\partial \mathrm{lnZ}}{\partial \beta}} \tag{14}$$

Thus, knowing the partition function, we can get the expected value for the energy of the system by simply differentiating.

How do we extract the entropy from the partition function? Using the Gibbs formula for entropy, we find

$$S = -k_B \sum_k P_k \ln P_k = -k_B \sum_k \frac{e^{-\beta E_k}}{Z} \ln \frac{e^{-\beta E_k}}{Z} = k_B \sum_k \frac{e^{-\beta E_k}}{Z} (\beta E_k + \ln Z) \tag{15}$$

We notice the sum in Eq. (13) as the first term on the right . The second term also simplifies using that $Z = \sum e^{-\beta E_k}$ and $k_B \beta = \frac{1}{T}$. Therefore

$$S = \frac{\langle E\rangle}{T} + k_B \ln Z \tag{16}$$

We also write this as

$$F \equiv -k_B T \ln Z = \langle E\rangle - TS \tag{17}$$

where F is the **free energy**. Calling the (log of the) partition function the free energy suggests that it has a physical interpretation, which it does. However, we can't do everything all at once. We'll study free energy in Lecture 8. In this lecture, we will not attempt to interpret Z itself but rather calculate it and use it to derive thermodynamic properties.

An important point about the canonical ensemble is that we use it to derive results about the system only, independent of how it is kept at finite temperature. The partition function is a sum over microstates of the system. P_k is the probability of finding the system in microstate k when it is in equilibrium at a temperature T *no matter what it is in contact with*. We need it to be in contact with something to exchange energy and keep it at finite temperature, but the details of those surroundings are totally irrelevant (except for the temperature). You can see this directly since the surroundings do not appear in the definition of Z.

One reason the canonical ensemble is very important is because it usually easier to compute than the microcanonical ensemble. While the microcanonical ensemble requires a constrained sum (constrained to only include microstates k with $E_k = E$), the sum in the canonical ensemble is unconstrained. As we saw in Section 6 of Lecture 4 a constrained system can be viewed as an unconstrained system with a Lagrange multiplier. We replace ln $\Omega(E)$ which is constrained at energy E, with $\ln Z(\beta) = \ln \Omega - \beta\langle E\rangle$ (i.e. Eq. (16)), with β the Lagrange multiplier. Then setting $\frac{\partial \mathrm{lnZ} - E\beta}{\partial \beta} = \langle E\rangle - E = 0$ enforces the Lagrange multiplier constraint. So, in a sense the canonical ensemble is powerful because of the power of Lagrange multipliers.

Note that we write $\langle E\rangle$ for the expected value of energy, rather than E since $\langle E\rangle$ is calculated rather than fixed from the beginning. The thing $\langle E\rangle$ which we compute in Eq. (14) is a function of β, so $\langle E\rangle$ is a derived quantity rather than one fixed from the start. In a real system connected to a heat bath, the temperature would be fixed, but the energy could fluctuate in time around $\langle E\rangle$ as little bits of energy flow in and out of the heat bath. The canonical ensemble of states is a much bigger set than the microcanonical ensemble – any possible state with any possible energy is included. If the actual system is isolated so its energy does not fluctuate, then we can simply impose the constraint $\langle E\rangle = E$ (equivalently, we impose the Lagrange-multiplier partial derivative condition). This takes us from an unconstrained canonical-ensemble system to a constrained microcanonical-ensemble system. In fact, this is mostly how we will use the canonical ensemble, to compute equilibrium properties of an isolated system. In such cases, we use $\langle E\rangle$ and E interchangeably, and we can use the extra constraint $\langle E\rangle = E$ to solve for a relation between T and E. The same relation between E and T can be derived from the microcanonical ensemble or from the canonical ensemble (as we will check when we can).

3 Example 1: monatomic ideal gas

For an ideal monatomic gas with positions q and momenta p, the energy depends only on momenta $E=\sum_j \frac{\vec{p}^2}{2m}$. So

$$Z \approx \int \frac{d^{3N}q d^{3N}p}{(\Delta q)^{3N}(\Delta p)^{3N}} \exp\left[-\beta \sum_j \frac{\vec{p}_j^2}{2m}\right] \tag{18}$$

Here Δp and Δq are the size of phase space regions that we consider minimal. Classical mechanics gives no indication of what we should take for Δq and Δp, and no results that we derive will depend on our choices. As mentioned before, in quantum mechanics, we know to set $\Delta q \Delta p = h$ (see Lecture 10) so let's take this value. Also, recall that for entropy to be extrinsic, we have to count any state in which the same positions and momenta are occupied as the same state. Thus we need to divide the integration by $N!$ for identical particles. This gives

$$Z = \frac{1}{N!}\int \frac{d^{3N}q d^{3N}p}{h^{3N}} \exp\left[-\beta \sum_j \frac{\vec{p}_j^2}{2m}\right] \tag{19}$$

The q integrals trivially give a factor of V^N. The p integrals are the product of $3N$ Gaussian integrals. Each one gives

$$\int_{-\infty}^{\infty} dp e^{-\beta\frac{p^2}{2m}} = \sqrt{\frac{2\pi m}{\beta}} \tag{20}$$

So that

$$Z = \frac{1}{N!}\left(\frac{V}{h^3}\right)^N \left(\frac{2\pi m}{\beta}\right)^{\frac{3}{2}N} \tag{21}$$

Mostly we are interested in this at large N, where $N! \to e^{-N}N^N$ gives

$$\boxed{Z_{\text{monatomic gas}} = e^N \left(\frac{V}{Nh^3}\right)^N \left(\frac{2\pi m}{\beta}\right)^{\frac{3}{2}N}} \tag{22}$$

Once we have Z it is easy to compute the (average) energy:

$$\langle E \rangle = -\frac{\partial \text{lnZ}}{\partial \beta} = -\frac{3}{2}N\frac{\partial}{\partial \beta}\ln\left(\frac{2\pi m}{\beta}\right) = \frac{3}{2}Nk_BT \tag{23}$$

For an isolated system, we then set $\langle E\rangle = E$. This is then in agreement with the result from the equipartition theorem (the 3 kinetic degrees of freedom each get $\frac{1}{2}k_BT$ of energy per molecule on average).

Note that this analysis of the ideal gas in the canonical ensemble was a much easier way to compute the average energy than in the microcanonical ensemble, where we had to look at the surface area of a $3N$-dimensional sphere.

3.1 Heat capacity and entropy

Recall that the heat capacity C_V is the amount of heat required to change the temperature at constant volume: $C_V = \left(\frac{Q}{\Delta T}\right)_V = \left(\frac{\partial E}{\partial T}\right)_V$. Recalling that $\beta = \frac{1}{k_BT}$ we have

$$C_V = \frac{\partial\langle E\rangle}{\partial T} = \frac{\partial \langle E\rangle}{\partial \beta}\frac{\partial \beta}{\partial T} = -\frac{1}{k_BT^2}\frac{\partial}{\partial\beta}\left[-\frac{\partial \text{lnZ}}{\partial \beta}\right] = \frac{1}{k_BT^2}\frac{\partial^2\text{lnZ}}{\partial \beta^2} \tag{24}$$

This equation lets us compute the heat capacity directly from the partition function.

Let's check for the monatomic ideal gas. Using Eq. (22) we find that

$$C_V = \frac{1}{k_B T^2}\frac{\partial^2}{\partial\beta^2}\ln\left(\beta^{-\frac{3}{2}N}\right) = \frac{3}{2}N\frac{1}{\beta^2 k_B T^2} = \frac{3}{2}N k_B \tag{25}$$

in agreement with our previous results.[2]

Plugging the partition function for the monatomic ideal gas, Eq. (22) into Eq. (16) for the entropy, we get

$$S = \frac{3}{2}N k_B + k_B \ln\left[e^N\left(\frac{V}{Nh^3}\right)^N\left(\frac{2\pi m}{\beta}\right)^{\frac{3}{2}N}\right] \tag{28}$$

which reduces to:

$$S = N k_B\left[\ln\frac{V}{Nh^3} + \frac{3}{2}\ln[2\pi m k_B T] + \frac{5}{2}\right] \tag{29}$$

Substituting $T = \frac{2}{3Nk_B}E$ this gives back the Sackur-Tetrode equation that we computed with the microcanonical ensemble.

4 Example 2: vibrational modes

Let's work out the canonical ensemble for another system, the vibrational modes of a diatomic molecule. For a diatomic molecule, motion along the axis of the molecule is governed by a potential $V(x)$. The equilibrium position x_0 is where the force vanishes: $F = -V'(x_0) = 0$. Expanding the potential near its minimum (the equilibrium position), $V(x) = V(x_0) + \frac{1}{2}(x - x_0)^2 V''(x_0) + \cdots$ we see that for small deviations from equilibrium, the potential is quadratic. Thus for small displacements it is going to be well modeled by a simple harmonic oscillator with spring constant $k = V''(x_0)$. The oscillation frequency is $\omega = \sqrt{\frac{k}{m}}$.

I assume you studied the quantum mechanics of a simple harmonic oscillator in your QM course. The oscillator has Hamiltonian

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2 \tag{30}$$

The energy eigenstates are

$$\psi_n(x) = \frac{1}{\sqrt{2^n n!}}\left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{m\omega x^2}{2\hbar}} H_n\left(\sqrt{\frac{m\omega}{\hbar}}x\right) \tag{31}$$

where $H_n(z) = (-1)^n e^{z^2}\frac{d^n}{dz^n}(e^{-z^2})$ are the Hermite polynomials. You can check that

$$\left(-\frac{\hbar^2}{2m}\partial_x^2 + \frac{1}{2}m w^2 x^2\right)\psi_n = E_n\psi_n \tag{32}$$

[2] It's interesting to write the calculation another way. Note that

$$\frac{\partial}{\partial\beta}\left[-\frac{\partial \ln Z}{\partial\beta}\right] = \frac{\partial}{\partial\beta}\left[\frac{1}{Z}\sum E e^{-\beta E}\right] = -\frac{1}{Z^2}\left(\frac{\partial Z}{\partial\beta}\right)\sum E e^{-\beta E} - \frac{1}{Z}\sum E^2 e^{-\beta E} \tag{26}$$

using that $-\frac{1}{Z}\frac{\partial Z}{\partial\beta} = \langle E\rangle$ we see that this is $-\langle E^2\rangle$. Thus,

$$C_V = -\frac{1}{k_B T^2}\frac{\partial}{\partial\beta}\left[-\frac{\partial \ln Z}{\partial\beta}\right] = \frac{\langle E^2\rangle - \langle E\rangle^2}{k_B T^2} \tag{27}$$

In other words, the heat capacity is given by the RMS energy fluctuations. This tells us that how a system changes when heated can be determined from properties of the system in equilibrium (the RMS energy fluctuations). In other words, to measure the heat capacity, we do not ever have to actually heat up the system. Instead, we can let the system heat up itself through thermal fluctuations away from the mean. This is a special case of a very general and powerful result in statistical physics known as the **fluctuation dissipation theorem**. Another example was our computation of how the drag coefficient in a viscous fluid related to the fluctuations determined by random walks (Brownian motion): if you drag something, the energy dissipates in the same way that statistical fluctuations dissipate.

where

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right), \qquad n = 0, 1, 2, \cdots \tag{33}$$

So a harmonic oscillator at rest has energy $E_0 = \frac{1}{2}\hbar\omega$. Each successive mode has $\hbar\omega$ more energy than the previous mode.

Note that for the simple harmonic oscillator, there is only one degree of freedom, so $N = 1$. If we fix the energy E, then we know the exact state of the system, $\Omega = 1$. Thus the microcanonical ensemble is not much use: it doesn't let us answer any questions we would like to ask. For example, we want to know what the typical energy in a vibrational mode is at fixed temperature? If we fix the energy ahead of time, we obviously can't answer this question.

So let us work in the canonical ensemble and compute the partition function for the system. We need to evaluate

$$Z_{\text{vib}} = \sum_n e^{-\beta E_n} = e^{-\frac{\beta}{2}\hbar\omega}\sum_{n=0}^{\infty} e^{-n\beta\hbar\omega} = \frac{1}{2\sinh\left(\frac{\beta}{2}\hbar\omega\right)} \tag{34}$$

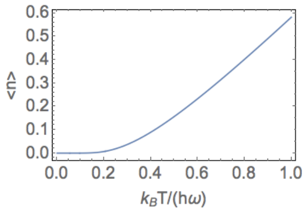
where $\sinh(x) = \frac{1}{2}(e^x - e^{-x})$ is the hyperbolic sine function. In the last step, we have performed the sum over n using $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$ with $x = e^{-\beta\hbar\omega}$ and simplified.[3]

With the exact partition function known, we can start computing things. The energy is

$$\langle E\rangle = -\frac{\partial \ln Z}{\partial\beta} = -\frac{\hbar\omega}{2}\coth\left(\frac{\beta}{2}\hbar\omega\right) = \hbar\omega\left(\frac{1}{e^{\beta\hbar\omega} - 1} + \frac{1}{2}\right) \tag{35}$$

(Feel free to use Mathematica or similar software to do these sums and take derivatives.) Comparing to Eq. (33) we see that the average excitation number is

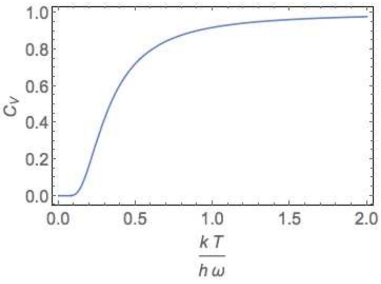
$$\langle n\rangle = \frac{1}{e^{\frac{\hbar\omega}{k_B T}} - 1} = \tag{36}$$



For $k_B T \lesssim \hbar\omega$, $\langle n\rangle \approx 0$ and only the ground state is occupied; from (35), the energy flatlines at its zero point: $E_0 = \frac{1}{2}\hbar\omega$. At higher temperatures, the $\langle E\rangle$ and $\langle n\rangle$ grow linearly with the temperature.

The heat capacity is

$$C_V = \frac{\partial E}{\partial T} = k_B\left(\frac{\hbar\omega}{k_B T}\right)^2 \frac{e^{-\frac{\hbar\omega}{k_B T}}}{\left(1 - e^{-\frac{\hbar\omega}{k_B T}}\right)^2} = \tag{37}$$

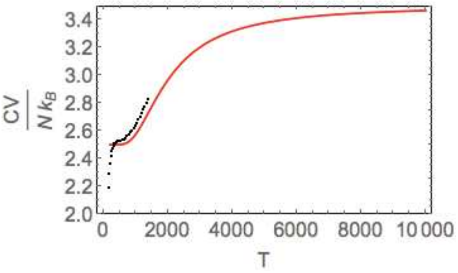


Note that heat capacity is very small until the first energy state can be excited, then it grows linearly.

[3] More explicitly, we define $x = e^{-\beta\hbar\omega}$ so that $e^{-n\beta\hbar\omega} = x^n$ and $e^{-\frac{\beta}{2}\hbar\omega} = \sqrt{x}$. Then $Z_{\text{vib}} = \sqrt{x}\sum_{n=0}^{\infty} x^n = \frac{\sqrt{x}}{1-x} = \frac{1}{\frac{1}{\sqrt{x}} - \sqrt{x}} = \frac{1}{e^{\frac{\beta}{2}\hbar\omega} - e^{-\frac{\beta}{2}\hbar\omega}} = \frac{1}{2\sinh\left(\frac{\beta}{2}\hbar\omega\right)}$.

For H_2, the vibrational mode has $\tilde{\nu}_{\text{vib}} = 4342\text{cm}^{-1}$ corresponding to $T_{\text{vib}} = \frac{ch\tilde{\nu}}{k_B} = \frac{\hbar\omega}{k_B} = 6300 K$. So at low energies, the vibrational mode cannot be excited which is why the heat capacity for hydrogen is $C_V = \frac{5}{2} N k_B T$ rather than $\frac{7}{2} N k_B T$. We discussed this in Lecture 4, but now we have explained it and can make more precise quantitative predictions of how the heat capacity changes with temperature. Including the kinetic contribution, and a factor of N for the N molecules that can be excited in the vibration mode we see

$$C_V = \frac{5}{2} N k_B + N k_B \left(\frac{T_{\text{vib}}}{T}\right)^2 \frac{e^{-\frac{T_{\text{vib}}}{T}}}{\left(1 - e^{-\frac{T_{\text{vib}}}{T}}\right)^2} = \tag{38}$$



This shows how the heat capacity goes up as the vibrational mode starts to be excitable. Note that although the temperature for the vibrational mode is 6300 K, the vibrational mode starts to be excited well below that temperature. The dots are data. We see good agreement! Can you figure out why the heat capacity dies off at low temperature? What do you think explains the small offset of the data from the theory prediction in the plot? We'll eventually produce a calculation in even better agreement with the data, but we need to incorporate quantum indistinguishability to get it right, as we will learn starting in Lecture 10.

We can also compute the partition function for a gas including both kinetic and vibrational motion. Since each momentum and each vibrational excitation is independent for the N particles, we have

$$Z = \frac{1}{N!}\left[\int \frac{d^3q d^3p}{h^3} \sum_n e^{-\beta\left(\frac{\vec{p}^2}{2m} + E_n\right)}\right]^N = Z_{\text{monatomic gas}} \times (Z_{\text{vib}})^N \tag{39}$$

This is a very general result: the partition function of a system with multiple independent modes of excitation is the product of the partition functions for the separate excitations. This follows simply from the fact that the total energy is the sum of each excitation energy and the exponential of a sum is the product of the exponentials.

5 Gibbs ensemble

In the microcanonical ensemble, we computed the number of states at a given energy $\Omega(V, N, E)$ and used it to derive the entropy $S(V, N, E) = k_B \ln\Omega(V, N, E)$. In the canonical ensemble, we computed the partition function by summing over Boltzmann factors, $Z(N, V, \beta) = \sum_k e^{-\beta E_k}$. In both cases we have been holding V and N fixed. Now we want to try varying V.

First, let's quickly recall from Section 2 of Lecture 4 why the temperature is the same in any two systems in thermal equilibrium. The quickest way to see this is to recall that entropy is extensive, so a system with energy $E_1 = E$ and another with energy $E_2 = E_{\text{tot}} - E$ has total entropy

$$S_{12}(E_{\text{tot}}, E) = S_1(E) + S_2(E_{\text{tot}} - E) \tag{40}$$

Then the state with maximum entropy is the one where

$$0 = \frac{\partial S_1(E)}{\partial E} + \frac{\partial S_2(E_{\text{tot}} - E)}{\partial E} = \frac{\partial S_1(E_1)}{\partial E_1}\bigg|_{E_1 = E} - \frac{\partial S_2(E_2)}{\partial E_2}\bigg|_{E_2 = E_{\text{tot}} - E} = \frac{1}{T_1} - \frac{1}{T_2} \tag{41}$$

where the $\frac{\partial S(x)}{\partial x}\Big|_{x = x_0}$ notation means evaluate the partial derivative at the point $x = x_0$. The minus sign in the second term comes from using the chain rule $\frac{\partial S_2(E_{\text{tot}} - E)}{\partial E} = \frac{\partial E_2}{\partial E}\frac{\partial S_2(E_2)}{\partial E_2} = -\frac{\partial S_2(E_2)}{\partial E_2}$ with $E_2 = E_{\text{tot}} - E$. This is how we showed that $\frac{1}{T} = \frac{\partial S}{\partial E}$ is the same in the two systems in Lecture 4.

Now let's consider an ensemble that lets V vary. This is sometimes called the **Gibbs ensemble**. In the Gibbs ensemble you have two systems in equilibrium that can exchange energy *and* volume. Exchanging volume just means we have a moveable partition in between them. So the total volume is conserved

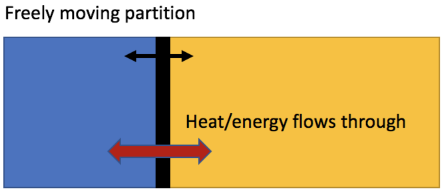


Figure 2. An ensemble where volume is allowed to vary

Now we just apply the same formal argument as in Eqs. (40) and (41): the entropy is the sum of the entropy of the two sides, and the total volume is fixed: $V_1 + V_2 = V_{\text{tot}}$. This implies that

$$S_{12}(E_{\text{tot}}, V_{\text{tot}}, E, V) = S_1(E, V) + S_2(E_{\text{tot}} - E, V_{\text{tot}} - V) \tag{42}$$

And so maximizing entropy, by demanding both the partial derivative with respect to E and the one with respect to V vanish give that the temperature $\frac{1}{T} = \frac{\partial S}{\partial E}$ is the same on both sides (from the E derivative) and that

$$0 = \frac{\partial S_{12}(E_{\text{tot}}, V_{\text{tot}}, E, V)}{\partial V} = \frac{\partial S_1(E, V_1)}{\partial V_1}\bigg|_{V_1 = V} - \frac{\partial S_2(E, V_2)}{\partial V_2}\bigg|_{V_2 = V_{\text{tot}} - V} = \frac{P_1}{T_1} - \frac{P_2}{T_2} \tag{43}$$

In the last step we have *defined* P by $\left(\frac{\partial S}{\partial V}\right)_E = \frac{P}{T}$. The T is added for convenience (so that P will be pressure, as opposed to pressure × temperature). What Eq. (43) shows is that this quantity P is the same for any two systems in equilibrium.

To show that P is the same as what we ordinarily call pressure, all we have to do is compute $\frac{\partial S}{\partial V}$ in some sample system, such as a monatomic ideal gas. Using the entropy of a monatomic ideal gas in the canonical ensemble, Eq. (29), we find

$$\left(\frac{\partial S}{\partial V}\right)_E = \frac{\partial}{\partial V} N k_B \left[\ln\frac{V}{N} + \frac{3}{2}\ln\left(\frac{4\pi m E}{3 N h^2}\right) + \frac{5}{2}\right] = \frac{N k_B}{V} = \frac{P}{T} \tag{44}$$

This explains why we defined P as $\left(\frac{\partial S}{\partial V}\right)_E = \frac{P}{T}$ rather than say $\left(\frac{\partial S}{\partial V}\right)_E = P$.

To summarize, we have established that $\left(\frac{\partial S}{\partial V}\right)_E$ is the same for any two systems in equilibrium. We also already know that $\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_V$ is the same for any two systems in equilibrium. We conclude that the quantity $T\left(\frac{\partial S}{\partial V}\right)_{E,N}$ is the same for any two systems in equilibrium, and give this quantity the name **pressure**:

$$P \equiv T\left(\frac{\partial S}{\partial V}\right)_E \tag{45}$$

There is a unique value for pressure among systems in equilibrium. This is of course consistent with the familiar observation that two gases will equilibrate with equal pressure, since a pressure difference would cause the partition to move. But this way of deriving it, we never had to talk about gas molecules or forces. Using Eq. (45) you can compute the pressure for a solid or photon gas or Bose-Einstein condensate or whatever. For example, one of the important applications of Eq. (45) is the *degeneracy pressure* of a system of identical fermions. Degeneracy pressure is present even at $T = 0$ and is responsible for keeping solids and neutron stars from collapsing. Degeneracy pressure will be studied in lectures 10, 13 and 15.

Note that Eq. (45) works for S computed any way you like it, as $S = k_B \ln\Omega$ in the microcanonical ensemble or $S = \frac{\langle E\rangle}{T} + k_B \ln Z$ in the canonical ensemble. The entropy is the entropy, however you compute it, and the pressure is the pressure.

Now let us consider the total derivative of $S(E, V)$:

$$dS = \left(\frac{\partial S}{\partial E}\right) dE + \left(\frac{\partial S}{\partial V}\right) dV = \frac{1}{T} dE + \frac{P}{T} dV \tag{46}$$

or

$$\boxed{dE = TdS - PdV} \tag{47}$$

This equation is none other than

$$\Delta E = Q - W \tag{48}$$

The change in energy is the heat $Q = TdS$ absorbed minus the work done by changing volume.

You might still be asking, how do we know that the quantity "P" really is "pressure"? Well, we showed that it is for a monatomic ideal gas. And everything in equilibrium has the same "pressure" ($P = \frac{F}{A}$ so if the pressures aren't equal there's a net force, things change, and it's not equilibrium) and the same "P". Thus, by the law of syllogism, it must be "P"= "pressure" for any system. It's the same argument about how we know that T is "temperature".

The Gibbs ensemble is usually just considered a variation on the canonical ensemble. You could in principle try to define a partition function for this ensemble by summing $Z_{\text{GE}} = \sum e^{-\beta E_k - \beta P V_k}$, but then you'd have to be able to compute the volume V_k for a microstate. I don't know of any examples where this is done. The point of the Gibbs ensemble is that thinking of volume varying between systems gives a nice general way to think about pressure, as conjugate to volume and analogous to temperature. This leads to the relation $P = T\left(\frac{\partial S}{\partial V}\right)_E$. This is a generalization of the conventional definition $P = \frac{F}{A}$. This relation is exact. Although we derived it using the Gibbs ensemble, it holds independent of how S is computed or whether the system is isolated.

6 Grand canonical ensemble

Now let's consider systems where the number of particles is not fixed. A basic example for this is from chemistry: in a chemical reaction, the number of each molecule species is not conserved. For example, when iron rusts the reaction is

$$3\ \text{Fe} + 4\ \text{H}_2\text{O} \quad \rightarrow \quad \text{Fe}_3\text{O}_4 + 4\ \text{H}_2 \tag{49}$$

Although the number of each type of *atom* is conserved, the number of each type of *molecule* is not. Other examples where number is not conserved are the photons that come off of a hot object (blackbody radiation) or radioactive decay, such as $n \rightarrow p^+ + e^-$.An ensemble where the total number N of things defining each system (H_2O or H_2 molecules) can change, but there are one or more conserved quantities (number of hydrogen or oxygen atoms) among systems, is called the **grand canonical ensemble**.

Consider first the case where two systems have a single conserved quantity. For example, system 1 might be H_2O molecules and system 2 might be H_2 molecules. Just like how maximizing entropy in the Gibbs ensemble implied equal pressures, the result of maximizing entropy in the grand canonical ensemble implies

$$\frac{\partial S_1(E, V, N)}{\partial N} = \frac{\partial S_2(E, V, N)}{\partial N} \tag{50}$$

for any two systems in equilibrium that can share constituents of N. As with pressure, we rescale by some convention (multiply by $-T$ in this case) and give the derivative a name

$$\mu \equiv -T\left(\frac{\partial S}{\partial N}\right)_{E,V} \tag{51}$$

This quantity is called the **chemical potential**. In equilibrium, the chemical potential of any two systems that can exchange particles is the same ($\mu_1 = \mu_2$). The minus sign is a convention. It makes the chemical potential negative in most circumstances.

A useful way to think about chemical potential is as a pressure for number density. For example, suppose you have an atom that has two states, a ground state 0 and an excited state 1. In equilibrium, there will be some concentrations $\langle n_0\rangle$ and $\langle n_1\rangle$ of the two states, and the two chemical potentials μ_1 and μ_2 will be equal. Since the excited states have more energy, in equilibrium we would have $\langle n_1\rangle < \langle n_0\rangle$. Say we then add to the system some more atoms in the ground state. This would push more atoms into the excited state to restore equilibrium. This pushing is due to the "number density pressure" of the chemical potential. Adding to n_0 pushes up μ_0, so $\mu_0 \neq \mu_1$ anymore; the number densities then change until equilibrium is restored.

While there is only one kind of temperature there are lots of chemical potentials: one for every thing we count in our ensemble. The more general formula is

$$\frac{\partial S_1(E, V, N_1, N_2, \cdots)}{\partial N_1} = -\frac{\mu_1}{T}, \qquad \frac{\partial S_1(E, V, N_1, N_2, \cdots)}{\partial N_2} = -\frac{\mu_2}{T}, \qquad \cdots \tag{52}$$

If Eq. (49), we would have 4 types of molecules (Fe ,H_2O, Fe_3O_4 4 H_2) so there 4 N's and 4 corresponding μ's. The 3 conserved quantities (H,O,Fe) then give 3 linear constraints among those chemical potentials. An explicit example is given below (see Eq. (84) below).

You should not think of the chemical potential as being connected to the grand canonical ensemble in any essential way. The chemical potential is property of the system, like pressure or temperature, relevant no matter what statistical system we use to perform the calculation. To see how chemical potential is embedded in the microcanonical ensemble, recall our microcanonical maximum entropy calculation, where we imposed $\Sigma n_i = N$ and $\sum n_i\varepsilon_i = E$ as constraints. Then we maximized entropy by maximizing

$$\frac{S}{k_B} = \ln\Omega = -N\sum_{i=1}^{m} f_i \ln f_i - \alpha\left(\sum n_i - N\right) - \beta\left(\sum n_i\varepsilon_i - E\right) \tag{53}$$

Since $\frac{\partial \ln\Omega}{\partial E} = \beta$, we identified this Lagrange multiplier β with the usual $\beta = \frac{1}{k_B T}$. Since $\frac{\partial \ln\Omega}{\partial N} = \alpha$ we can now identify $\mu = -\alpha k_B T$ as the chemical potential. Thus given Ω in the microcanonical ensemble, we compute the chemical potential as

$$\mu = -\alpha k_B T = -k_B T\left(\frac{\partial \ln\Omega(E, V, N)}{\partial N}\right) = -T\left(\frac{\partial S}{\partial N}\right)_{E,V} \tag{54}$$

in agreement with Eq. (51).

As in Eq. (47) we can now consider the total derivative of energy, letting E, V and N all vary:

$$dS = \left(\frac{\partial S}{\partial E}\right)dE + \left(\frac{\partial S}{\partial V}\right)dV + \left(\frac{\partial S}{\partial N}\right)dN = \frac{1}{T}dE + \frac{P}{T}dV - \frac{\mu}{T}dN \tag{55}$$

That is,

$$\boxed{dE = TdS - PdV + \mu dN} \tag{56}$$

This implies that

$$\left(\frac{\partial E}{\partial N}\right)_{S,V} = \mu \tag{57}$$

So the chemical potential represents the change in energy when a particle is added at constant V and S. This is *almost* intuitive. Unfortunately the constant-S constraint makes Eq. (57) hard to interpret. Don't worry though, we'll come up with better ways to understand μ in Section 7.

6.1 Grand partition function

As in Section 2 let us now hook a small system up to a reservoir to derive the Boltzmann factor. This time the reservoir should have large energy and large particle number, and both energy and particle number can flow between the system and reservoir. As before, think about picking one microstate k of the system with energy E_k and N_k particles. Once E_k and N_k are fixed, the total number of microstates is determined only by the states in the reservoir. Eq. (7) becomes

$$\ln\Omega_{\text{res}}(E_{\text{tot}} - E_k, N_{\text{tot}} - N_k) = \ln\Omega_{\text{res}}(E_{\text{tot}}, N_{\text{tot}}) - \beta E_k + \beta\mu N_k \tag{58}$$

where Eq. (54) was used. This leads to a Boltzmann factor

$$P_k = \frac{1}{\mathcal{Z}} e^{-\beta E_k + \beta\mu N_k} \tag{59}$$

where

$$\mathcal{Z}(V, \beta, \mu) = \sum_k e^{-\beta E_k + \beta\mu N_k} \tag{60}$$

is called the **grand partition function**.

The grand partition function lets us calculate the expected number of particles

$$\langle N\rangle = \sum_k N_k P_k = \frac{1}{\mathcal{Z}} \sum_k N_k e^{-\beta E_k + \beta\mu N_k} = \frac{1}{\beta}\frac{1}{\mathcal{Z}}\frac{\partial\mathcal{Z}}{\partial\mu} = \frac{1}{\beta}\frac{\partial\ln\mathcal{Z}}{\partial\mu} \tag{61}$$

We can also calculate the usual things the partition function lets us calculate, such as the average energy. Taking a derivative with respect to β we get

$$\frac{1}{\mathcal{Z}}\frac{\partial\mathcal{Z}}{\partial\beta} = \sum_k (-E_k + \mu N_k) e^{-\beta E_k + \beta\mu N_k} = -\langle E\rangle + \mu\langle N\rangle \tag{62}$$

so that

$$\langle E\rangle = -\frac{\partial\ln\mathcal{Z}}{\partial\beta} + \frac{\mu}{\beta}\frac{\partial\ln\mathcal{Z}}{\partial\mu} \tag{63}$$

Particle number and chemical potential are conjugate, like pressure and volume. If you know N for a system then you can calculate μ by $\frac{\partial E}{\partial N}$. This is like how if you know the energy for a system, you can calculate temperature from $\frac{1}{T} = \frac{\partial S}{\partial E}$. If you know the chemical potential instead of N, then you can compute average number by $\langle N\rangle = \frac{1}{\beta}\frac{\partial\ln\mathcal{Z}}{\partial\mu}$. This is like how if you know temperature and not the energy, you can compute the average energy form $\langle E\rangle = -\frac{\partial\ln Z}{\partial\beta}$.

Finally, let's compute the entropy, in analogy to Eq. (16). We start with Eq. (15), which goes through to the grand canonical ensemble with $Z \to \mathcal{Z}$ and $\beta E \to \beta(E - \mu N)$:

$$S = k_B \sum \frac{e^{-\beta E_k + \beta\mu N_k}}{\mathcal{Z}} [\beta(E_k - \mu N_k) + \ln\mathcal{Z}] \tag{64}$$

$$= \frac{\langle E\rangle}{T} - \mu\frac{\langle N\rangle}{T} + k_B \ln\mathcal{Z} \tag{65}$$

Thus,

$$\boxed{-k_B T \ln\mathcal{Z} = \langle E\rangle - TS - \mu\langle N\rangle} \tag{66}$$

This will be a useful relation.

7 Chemical potential

One of the most important uses of chemical potential is for phase transitions. In Lecture 9, we"ll see that different phases of matter are characterized by different chemical potentials so that phase boundaries are where the chemical potentials are equal. Unfortunately, we don't have analytical forms for the chemical potential in most real systems, like liquids. So in order to understand phase transitions and other sophisticated uses of chemical potential, we first need to build up some intuition from simple solvable examples.

Remember, μ is independent of which ensemble we use to we calculate it. Although it is a natural object for the grand canonical ensemble, we will work in this section in the microcanonical ensemble since it is somewhat more intuitive. (We'll come back to the grand canonical ensemble and the grand partition function in the next lecture and use it a lot in quantum statistical mechanics.)

7.1 Ideal gas

For a monatomic gas, using the Sackur-Tetrode equation, we find

$$\mu = -T\left(\frac{\partial S}{\partial N}\right)_{E,V} = -T\frac{\partial}{\partial N}\left\{Nk_B\left[\ln\frac{V}{N}+\frac{3}{2}\ln\left(\frac{4\pi mE}{3Nh^2}\right)+\frac{5}{2}\right]\right\} \tag{67}$$

$$= -k_BT\left[\ln\frac{V}{N}+\frac{3}{2}\ln\left(\frac{4\pi mE}{3Nh^2}\right)\right] \tag{68}$$

Note that the $\frac{5}{2}$ has dropped out. Using $E=\frac{3}{2}Nk_BT$ for this gas, we can write this relation in an abbreviated form

$$\mu = k_BT\ln\left[n\left(\frac{h^2}{2\pi mk_BT}\right)^{3/2}\right] = k_BT\ln n\lambda^3 \tag{69}$$

where $n=\frac{N}{V}$ is the number density and

$$\lambda = \frac{h}{\sqrt{2\pi mk_BT}} \tag{70}$$

is called the **thermal de Broglie wavelength** or just **thermal wavelength**.

Recall that the de Broglie wavelength is $\lambda=\frac{h}{p}$, with p the momentum. Since the momenta of particles in a gas vary according to the Maxwell-Boltzmann distribution, they have a variety of different de Broglie wavelengths. The thermal wavelength is the wavelength of a typical particle in a thermal system, more precisely, one with momentum $p=\sqrt{2\pi mk_BT}=\sqrt{\frac{2\pi}{3}}p_{\mathrm{rms}}$ with p_{rms} the RMS momentum of a gas a temperature T. The de Broglie wavelength is measure of the length scale at which quantum effects become important. If a gas at temperature T is more dense than 1 particle per thermal de Broglie wavelength-cubed, $n>\frac{1}{\lambda^3}$, then quantum statistical mechanics must be used. We will make this correspondence precise in Lecture 10.

To get a feel for typical numbers, consider air at room temperature. The molar mass of air is $29.0\frac{g}{\mathrm{mol}}$ and density is $\rho=1.27\frac{\mathrm{kg}}{m^3}$. So $\lambda=\frac{h}{\sqrt{2\pi m_{N_2}k_BT}}=1.87\times10^{-11}m$ while $n=(3.35\times10^{-9}m)^{-3}$. In other words, the typical distance between atoms in air is $d=3.3\mathrm{nm}$ and the thermal wavelength is much smaller, $\lambda=0.02\mathrm{nm}$. So in air $n\lambda^3=1.7\times10^{-7}\ll1$. This means that $\mu=k_BT\ln n\lambda^3=-0.39\mathrm{eV}$. Note that the chemical potential of air is negative. Since $\mu=k_BT\ln n\lambda^3$, the chemical potential will always be negative in regions where $n<\frac{1}{\lambda^3}$ and classical statistical mechanics applies.

Solving Eq. (69) for n gives

$$n = \frac{1}{\lambda^3}\exp\left(\frac{\mu}{k_BT}\right) \tag{71}$$

This says that the number density is related exponentially to the chemical potential. Thus if we double the number of particles, $n\to2n$, the chemical potential goes up by $\mu\to\mu+\ln2$. As the system gets denser and denser, the chemical potential rises towards 0. When the chemical potential is negative, the gas is not too crowded and can considered dilute, with interparticle interactions ignored (as in a classical ideal gas).

To get additional intuition for chemical potential, suppose you have a system with a concentration gradient. Then the part with more particles will be at higher chemical potential and the part with lower gradient at lower chemical potential, according to Eq. (71). This is why it is called a potential – it is like potential energy, but for particle number. Number density then flows down the chemical potential until equilibrium is established.

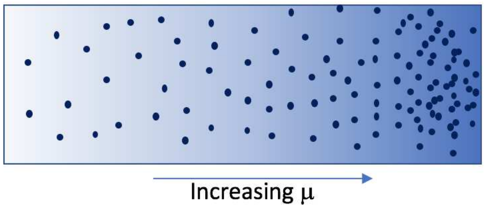


Figure 3. Chemical potential is higher in more dense regions. It is like potential energy for particle number. Particles move from high to low μ, until the μ's are all equal.

7.2 Ground-state energy

There is not a unique way to define the ground state energy; it depends on what we include. For example, we could include chemical bond energy in the ground-state energy, or the rest mass mc^2, or the gravitational potential energy relative to the surface of the earth mgh or relative to the center of the earth $G\frac{Mm}{r}$, and so on. There is always going to be some arbitrariness in the definition of energy. However, once we define the ground-state energy for one thing, we can talk without ambiguity about the ground-state energy of everything.

So let's say that we have established a convention and the ground state energy for a molecule is ε. To be concrete, think of of ε as the chemical-bond energy of the molecule. Then there is a contribution $N\varepsilon$ to the total energy of the N molecules. For a monatomic gas, the total energy is offset from the kinetic energy: $E = E_{\text{kin}} + N\varepsilon$. The ground state energy doesn't affect the number of states, so the functional form of entropy $S(E_{\text{kin}}) = k_B \ln\Omega$ is the same with or without the offset; either way, it counts the number of ways a total amount of kinetic energy is distributed among the momenta of the gas molecules. Since we want to write entropy a function of the total energy, we should the substitute $E_{\text{kin}} = E - N\varepsilon$. Then, in the microcanonical ensemble we have

$$S = Nk_B\left[\ln\frac{V}{N} + \frac{3}{2}\ln\left(\frac{4\pi m(E - N\varepsilon)}{3Nh^2}\right) + \frac{5}{2}\right] \tag{72}$$

Of course, if there is just one type of molecule present, we can just choose $\varepsilon = 0$, but including ε explicitly will allow us to discuss systems of molecules with different ground state energies (chemical bond energies, rest masses, etc.)

One can also derive Eq. (72) from the canonical ensemble. Let the partition function without the energy offset (i.e. $\varepsilon = 0$) be Z_0 and so $\langle E_0\rangle = -\frac{\partial \ln Z_0}{\partial\beta}$ and $S_0(E_0) = \frac{\langle E_0\rangle}{T} + k_B \ln Z_0$. Then with the energy offset we get $Z = Z_0 e^{-\beta N\varepsilon}$ and so $\langle E\rangle = -\frac{\partial \ln Z}{\partial\beta} = \langle E_0\rangle + N\varepsilon$. Then

$$S = \frac{\langle E\rangle}{T} + k_B \ln Z = \frac{\langle E_0\rangle}{T} + \frac{N\varepsilon}{T} + k_B \ln Z_0 - k_B\beta N\varepsilon = \frac{\langle E_0\rangle}{T} + k_B \ln Z_0 = S_0(E_0) = S_0(E - N\varepsilon) \tag{73}$$

Thus using the canonical ensemble we again find that the entropy S with the offset has the functional form as S_0 without the offset; it is only the energy at which we evaluate S that changes. This is in agreement with Eq. (72).

From the entropy, we can compute the chemical potential

$$\mu = -T\left(\frac{\partial S}{\partial N}\right)_{E,V} = k_B T \ln n\lambda^3 + \varepsilon \tag{74}$$

with λ in Eq. (70), so that for an ideal gas with ground-state energy ε

$$n = \frac{1}{\lambda^3}\exp\left(\frac{\mu - \varepsilon}{k_B T}\right) \tag{75}$$

Thus if we were to change the zero-point energy offset $\varepsilon \rightarrow \varepsilon + \Delta\varepsilon$, we could compensate for this by shifting $\mu \rightarrow \mu + \Delta\varepsilon$. In other words, the chemical potential is measured *relative* to the ground state: only the difference $\mu - \varepsilon$ appears. This is just like how only potential energy *differences* are physical, and why we call the chemical potential a *potential*.

With the energy offset, we can refine our observation about the chemical potential being negative for a classical ideal gas. That observation held when the gas was dilute: $n < \frac{1}{\lambda^3}$, with λ the thermal de Broglie wavelength. Now we see that a more precise statement is that for a classical ideal gas, $\mu - \varepsilon < 0$, i.e. chemical potential is less than the ground state energy ε. Eq. (74), $\mu = k_B T \ln n\lambda^3 + \varepsilon$ says that the chemical potential gets two contributions: one from the density and one from the energy. The density contribution is of entropic origin and depends on how many molecules are in the system. The energetic contribution is due to the internal structure of the molecule and independent of whatever else is going on. Equilibrium, where chemical potentials are equal, comes from a balance between these two contributions. This should become clearer with some examples.

For a monatomic gas that is not ideal, so there are other quadratic degrees of freedom such as vibrational modes with a characteristic temperature. At large T the partition function with f_v vibrational modes is

$$Z = e^N \left(\frac{V}{Nh^3}\right)^N \left(\frac{2\pi m}{\beta}\right)^{\frac{3}{2}N} \left(\frac{1}{\beta\hbar\omega}\right)^{f_v N} \tag{76}$$

and $\langle E\rangle = \frac{3+2f_v}{2} N k_B T$ as per the equipartition thoerem. Then

$$S = N k_B \left[\ln\frac{V}{N} + \frac{3}{2}\ln\left(\frac{4\pi m\,(E - N\varepsilon)}{(3+2f_v)\,Nh^2}\right) + f_v \ln\frac{2\,(E-N\varepsilon)}{(3+2f_v)N\hbar\omega} + \frac{5+2f_v}{2}\right] \tag{77}$$

adding in the ground state energy, this leads to

$$n = \left(\frac{T}{T_v}\right)^{f_v} \frac{1}{\lambda^3} \exp\left(\frac{\mu - \varepsilon}{k_B T}\right) \tag{78}$$

Or equivalently

$$\mu = \varepsilon + k_B \mathrm{T}\ln\frac{N}{V} - \frac{3}{2} k_B T \ln\frac{2\pi m k_B T}{h^2} - f_v k_B T \ln\frac{T}{T_v} + \cdots \tag{79}$$

where $T_{\mathrm{vib}} = \frac{\hbar\omega}{k_B}$ as in Section 4. Other quadratic degrees of freedom modify the chemical potential in a similar way.

7.3 Chemical reactions

Chemical potentials are useful in situations where particles turn into other types of particles. When there is more than one type of particle in the system (as there typically are when we consider problems involving chemical potentials), we need a different μ for each particle. So Eq. (56) becomes

$$dE = TdS - PdV + \sum \mu_j dN_j \tag{80}$$

As a concrete example, consider the Haber process for the production of ammonia

$$3\ \mathrm{H}_2 + \mathrm{N}_2 \leftrightharpoons 2\ \mathrm{NH}_3 \tag{81}$$

Note that the number of each individual molecule is not conserved, but because the number of hydrogen atoms and nitrogen atoms is conserved, the relative coefficients (3,1 and 2) in Eq. (81) are fixed. In chemistry, the **concentrations** or **molar number densities** of molecule j are denoted as $[j] = \frac{n_j}{N_A}$, with $n_j = \frac{N_j}{V}$ and $N_A = 6 \times 10^{23} \frac{1}{\mathrm{mol}}$ Avogadro's number. In equilibrium, there will be some relationship among the concentrations $[\mathrm{H}_2]$ of hydrogen, $[\mathrm{N}_2]$ for nitrogen and $[\mathrm{NH}_3]$ for ammonia that we can compute using chemical potentials.

First, we note the reaction Eq. (81) implies that if $[\mathrm{N}_2]$ goes down by one molecule, $[H_2]$ must go down by 3 molecules and $[\mathrm{NH}_3]$ must go up by 2 molecules. That is

$$d[\mathrm{H}_2] = 3d[\mathrm{N}_2], \quad d\,[\mathrm{NH}_3] = -2d[\mathrm{N}_2], \tag{82}$$

As the concentrations change, at fixed volume and fixed total energy, the entropy changes as

$$dS = \frac{\partial S}{\partial[\mathrm{H}_2]} d[\mathrm{H}_2] + \frac{\partial S}{\partial[\mathrm{N}_2]} d[\mathrm{N}_2] + \frac{\partial S}{\partial[\mathrm{NH}_3]} d[\,\mathrm{NH}_3] \tag{83}$$

Thus, using that $dS = 0$ in equilibrium, Eq. (82) and the definition of the chemical potentials as $\mu_i = \frac{\partial S}{\partial N_i}$, we find[4]

$$3\mu_{H_2} + \mu_{N_2} = 2\,\mu_{\mathrm{NH}_3} \tag{84}$$

This constraint among the chemical potentials is a generalization of $\mu_1 = \mu_2$ in equilibrium for two systems that can exchange particles. Here there are 3 systems that can exchange particles.

Now, from Eq. (75) we know how to relate the number of particles to the chemical potential for a monatomic ideal gas:

$$[X] = \frac{1}{\lambda^3}\exp\left[-\frac{\varepsilon_X - \mu_X}{k_B T}\right] \tag{85}$$

where ε_X is the ground state energy for molecule X. To get the μ's to drop out, we can take the ratio of concentrations to appropriate powers:

$$\frac{[\mathrm{NH}_3]^2}{[\mathrm{H}_2]^3[\mathrm{N}_2]} \approx \frac{\lambda_{H_2}^9 \lambda_{N_2}^3}{\lambda_{\mathrm{NH}_3}^6} \times \exp\left[-\frac{2\varepsilon_{\mathrm{NH}_3} - 3\varepsilon_{H_2} - \varepsilon_{N_2}}{k_B T}\right] \underbrace{\exp\left[-\frac{2\,\mu_{\mathrm{NH}_3} - 3\mu_{H_2} - \mu_{N_2}}{k_B T}\right]}_{=1} \tag{86}$$

The second exponential is just 1 because of Eq. (84), which is why we chose the powers of $[H_2]$ and [NH$_3$] that we did on the left hand side. The $\approx$ means that we are approximating everything as monatomic ideal gases (not a great approximation but it's a start)

The sum of energies is just the net energy change in the reaction, $\Delta\varepsilon$. For the Haber process, which is exothermic, $\Delta\varepsilon = -92.4\frac{\mathrm{kJ}}{\mathrm{mol}}$. So

$$\frac{[\mathrm{NH}_3]^2}{[\mathrm{H}_2]^3[\mathrm{N}_2]} \approx \frac{\lambda_{H_2}^9 \lambda_{N_2}^3}{\lambda_{\mathrm{NH}_3}^6}\exp\left[-\frac{\Delta\varepsilon}{k_B T}\right] \quad \text{(assuming monatomic gases)} \tag{87}$$

This is special case (for monatomic ideal gases) of the law of mass action. It says that the relative concentrations of reacting molecules in equilibrium are determined by the Boltzmann factor dependent on the change in energy associated with the reaction. This formula arises from a balance between entropic contributions to the chemical potentials on both sides (though their number densities) and energetic contributions (in the exponential factor).

We have written explicitly the reminder that this formula assumes that the reactants and products are monatomic gases. This is not a bad assumption in some cases. More generally though, for chemicals reacting, we will need to add corrections to the right hand side. These corrections will be included in the next lecture, where the law of mass action is derived in full.

7.4 Example: matter antimatter asymmetry (optional)

For another example, consider the process of a proton-antiproton annihilation. Antiprotons p^- are anti-particles of protons. They have the same mass as protons but opposite electric charge. Protons and anti-protons can annihilate into photons

$$p^+ + p^- \leftrightharpoons \gamma + \gamma \tag{88}$$

The reverse reaction is photons converting into proton-antiproton pairs. These annihilations and conversions happen constantly when the temperature is well above the threshold energy for pair production

$$k_B T \gg \varepsilon = 2 m_p c^2 = 2\,\mathrm{GeV} \tag{89}$$

We don't care so much about the details of why or how this process occurs, just that it does occur. This threshold temperature is around 2×10^{13} K. So in most systems of physical interest (stars, your body, etc.) this doesn't happen. It did happen however, in the early universe, until 0.01 seconds after the big bang.

[4] Technically speaking, you need a conserved total N to define the chemical potential. Because the numbers of atoms are conserved, there are chemical potentials μ_H and μ_N for them. So what we are doing implicitly above is defining the chemical potentials for the molecules in terms of the atomic chemical potentials $\mu_{H_2} \equiv 2\mu_H, \quad \mu_{N_2} \equiv 2\mu_N$ and $\mu_{\mathrm{NH}_3} \equiv \mu_N + 3\mu_H$ from which Eq. (84) follows.

Note that while the above reaction conserves the number of protons minus the number of antiprotons, it does not conserve the number of photons. Indeed, other reactions can easily change photon number, such as

$$\gamma + e^- \leftrightharpoons e^- + \gamma + \gamma \tag{90}$$

A more down-to-earth example is the light in your room – photons are constantly being produced in the lightbulbs. Eq. (90) implies that

$$\mu_\gamma + \mu_{e^-} = \mu_{e^-} + 2\mu_\gamma \tag{91}$$

In other words, that

$$\mu_\gamma = 0 \tag{92}$$

This is a general property of particles that are not associated with any conserved quantity: their chemical potential vanishes. Then the reaction in Eq. (88) gives

$$\mu_{p^+} + \mu_{p^-} = 0 \tag{93}$$

In addition, it is natural to suppose that all the protons and antiprotons came from processes like $\gamma + \gamma \to p^+ + p^-$ that produce or remove the same number of protons and antiprotons. This would make the proton/antiproton concentrations equal[5], and their chemical potentials equal too (and hence $\mu_{p^+} = \mu_{p^-} = 0$ by Eq. (93)).

Now, the energy change in the reaction $\gamma + \gamma \to p^+ + p^-$ is $\Delta\varepsilon = 2m_p c^2$. Thus, as in Eq. (87), using $T = 3K$ (the temperature of outer space), and treating protons and antiprotons as monatomic ideal gases (an excellent approximation in fact) with a a thermal wavelength $\lambda = \frac{h}{\sqrt{2\pi m\, k_B(3K)}} \approx 1\,\text{nm}$, we find

$$[p^-] = [p^+] = \frac{1}{\lambda^3} e^{-\frac{\Delta\varepsilon}{2k_B T}} = \left(\frac{2\pi m_p\, k_B T}{h}\right)^{3/2} e^{-\frac{m_p c^2}{k_B T}} = 4 \times 10^{-843112945335} \frac{1}{m^3} \approx 0 \tag{94}$$

So this first pass calculation says there shouldn't be any protons *or* antiprotons around at all!

To refine our calculation, it's important to note that we used equilibrium physics, but since the universe is expanding, equilibrium is not always a good approximation. At some point as the universe expands and cools, the protons and antiprotons become so dilute that they cannot find each other to annihilate. This is called "freeze-out". The freeze-out temperature is set by when the rate for $p^+ p^- \to \gamma\gamma$ is equal to the expansion rate of the universe. The rate for $p^+ p^- \to \gamma\gamma$ depends on the proton's scattering cross section, which is approximately its cross-sectional area $\sigma \sim \frac{1}{m_p^2}$, the number density $[p^+]$ and the velocity, which we can take to be given by the Maxwell-Boltzmann average $\langle \frac{1}{2} m_p \vec{v}^2 \rangle = \frac{3}{2} k_B T$. Putting these factors together, the annihilation rate (events per unit time) is:

$$\Gamma_{\text{annihilate}} = n\sigma v = (2\pi m_p\, k_B T)^{3/2} e^{-\frac{m_p c^2}{k_B T}} \frac{1}{m_p^2} \sqrt{\frac{3k_B T}{m_p}} \tag{95}$$

The expansion rate requires some general relativity. The result is

$$\Gamma_{\text{expansion}} = \frac{k_B^2 T^2}{M_{\text{Pl}}} \tag{96}$$

where $M_{\text{Pl}} = G_N^{-1/2} = 10^{19}$ GeV is Planck's constant. Setting these equal results in a freezeout temperature

$$T_f = 2.4 \times 10^{11} K \tag{97}$$

At this temperature, the proton concentration is not so small:

$$[p^+] = [p^-] \approx \left(\frac{2\pi m_p\, k_B T_f}{h}\right)^{3/2} e^{-\frac{m_p c^2}{k_B T_f}} = 10^{23} \frac{1}{m^3} \tag{98}$$

[5] Actually, in any unitary quantum field theory the equilibrium concentrations of particles and antiparticles must be the same. This follows from an unbreakable symmetry known as CPT invariance that combines switching particles and antiparticles (C), flipping the particles spins (P) and time-reversal invariance (T).

However, as the universe continues to expand from T_f down to $3K$ its size scales with temperature so the proton number density gets diluted down to

$$[p^+] = [p^-] \approx 10^{23}\frac{1}{m^3}\left(\frac{3K}{T_f}\right)^3 = 1.68 \times 10^{-10}\frac{1}{m^3} \tag{99}$$

This is the honest-to-goodness prediction of cosmology for the density of protons left over from the big bang.

Unfortunately, the prediction $[p^+] = 10^{-10}\frac{1}{m^2}$ is in stark disagreement with data: the average number density of protons in the universe is actually $[p^+] = 0.26\frac{1}{m^3}$. This is a problem. In fact, this is one of the great unsolved problems in fundamental physics, called the mystery of **baryogenesis** or the **matter-antimatter asymmetry**.

One possible solution is to set the initial conditions so that $[p^+] \neq [p^-]$ to start with. Once these are set, if all the processes are symmetric in p^+ and p^- then $[p^+] \neq [p^-]$ will persist. Note however, that the universe is currently $10^{26}\, m$ wide, and growing. There are 10^{80} more protons than antiprotons in the observable universe today. When the universe is only $10^{-35}m$ across, this would correspond to a shocking number density of $10^{185}\frac{1}{m^3}$. So it would be a little strange to set this enormous asymmetry in the early universe. Moreover, the current cosmological model involves inflation, which produces exponential growth at early times, so whatever initial asymmetry we set would be completely washed away when inflation ends. In other words, it's possible, but would be very unsettling, to solve the baryogenesis problem by tuning the initial conditions.

Another option is to start off symmetric but have processes that are not symmetric between particles and antiparticles. In turns out in the Standard Model of particle physics, there are none: for every way of producing an electron or proton, there is also a way of producing a positron or antiproton with exactly the same rate. In fact, this equality is guaranteed by symmetries (lepton number and baryon number). Moreover, if you made a modification so that the symmetries were violated, then effectively protons could turn into antiprotons. Thus, since protons and antiprotons have the same mass (and value of ε) their chemical potentials would push them towards the same concentrations, which by Eq. (94) is zero. The story is again a little more complicated, since there is inflation, and reheating and the expansion of the universe is not quite quasi-static, and there is actually a super-tiny violation of the symmetry between protons and antiprotons within the Standard Model. Even when you include all these things, it doesn't work, you still get no matter out once the universe cools.

So we are stuck. Why is there so much matter in the universe? Why is there more matter than antimatter? Nobody knows.-

8 Summary

In this lecture, the main conceptual tool in statistical mechanics was introduced: the ensemble. An ensemble is an imaginary collection of systems with different microscopic arrangements and the same fixed macroscopic properties. The properties we consider come in conjugate pairs (E, T), (N, μ) and (V, P), and the ensembles pick one variable from each pair to be independent. The main ensembles are

- The microcanonical ensemble: N, V and E are fixed.
- The canonical ensemble: N, V and T are fixed.
- The grand canonical ensemble: V, T and the chemical potential μ are fixed.
- The Gibbs ensemble (not used much): N, P and T are fixed.

Each ensemble has an natural object that can be used to extract dependent quantities. In the microcanonical ensemble, the object is the number of microstates $\Omega(N, V, E)$ or equivalently the entropy $S = k_B \ln \Omega$. Then

$$\frac{1}{T} = \frac{\partial S}{\partial E}, \quad P = T\frac{\partial S}{SV}, \quad \mu = -T\frac{\partial S}{\partial N} \tag{100}$$

A shortcut for all these relations is the differential relation

$$dE = TdS - PdV + \mu dN \tag{101}$$

In all the ensembles, we just have to *imagine* that the systems are connected to other systems to maintain constant T, P, μ etc. The properties of the systems we calculate hold whether the systems are actually in equilibrium with something else or not. The ensembles as conceptual tools to calculate properties of systems based on known information, independent of their surroundings.

In the canonical ensemble, the natural object is the partition function $Z(T, N, V) = \sum_k e^{-\beta E_k}$ where the sum is over microstates k with E_k their energy. One should think of states in the canonical ensemble as being in thermal equilibrium with something else, like a heat bath. So energy can flow in and out. Since temperature is fixed, not energy, there are microstates included in the canonical ensemble with different energies, and the energies can be arbitrarily large. The average of the energies of all the microstates is

$$\langle E\rangle = \frac{1}{Z}\sum_k E_k e^{-\beta E_k} = -\frac{1}{Z}\frac{\partial Z}{\partial\beta} = -\frac{\partial \mathrm{lnZ}}{\partial\beta} \tag{102}$$

This average $\langle E\rangle$ means two things: first, it means we average over all the possible microstates k with all values of E_k, since E can fluctuate due to the thermal contact. Second, it means the average over time of the energy of the actual microstate of the system in thermal contact with a bath. The time average is the ensemble average, due to ergodicity. We can only ever measure the time average, not the ensemble average, but we compute the ensemble average.

In the grand canonical ensemble, the number of particles is not fixed. For example, particles might evaporate, or chemical reactions could occur. The chemical potential μ is the conjugate variable to N, like T is conjugate to E: two systems in equilibrium that can exchange N have the same μ, just like two systems in equilibrium that can exchange E are at the same T.

In the Gibbs ensemble, volume is not fixed, for example, a balloon with gas in it. When two systems in equilibrium can exchange volume, they are at the same pressure. So P is conjugate to V.

The conjugate pairs are not exactly equivalent. Any individual microstate of a system can have N, V and E well defined. The other properties, T, μ, P and S are not defined for a single microstate, only for an ensemble. For example, if we have a gas and specify the positions and momenta of all the molecules, it is a single point in phase space, with no entropy and no temperature. Entropy and temperature are *ensemble* properties. Pressure is an ensemble property too. You might think an isolated system has a pressure. But to measure the pressure, we need to see the force on the walls of the box. This force is microscopically very irregular, so we must average over time to get something sensible. This time average is the same as the ensemble average due to ergodicity. So again we measure the time average but compute the ensemble average, just like temperature.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 8: Free energy

1 Introduction

Using our various ensembles, we were led to a rather simple differential relation between energy and other state variables, Eq. (56) of Lecture 7:

$$dE = TdS - PdV + \mu dN \tag{1}$$

The terms on the right are different ways the energy can change: the first term is when heat comes in, since $dS = \frac{dQ}{T}$, the second term is work going out from expanding volume, the third term is the energy associated with bond formation or increasing particle creation.

Holding any two of the four differentials fixed we can derive from this equation lots of other formulas

$$\left(\frac{\partial S}{\partial E}\right)_{V,N} = \frac{1}{T}, \quad \left(\frac{\partial S}{\partial N}\right)_{E,V} = -\frac{\mu}{T}, \quad \left(\frac{\partial S}{\partial V}\right)_{E,N} = \frac{P}{T}, \quad \left(\frac{\partial E}{\partial S}\right)_{V,N} = T, \quad \left(\frac{\partial E}{\partial V}\right)_{S,N} = -P, \tag{2}$$

and so on. Some of these are more useful than others. It's important to label what you are holding fixed because S, E, V, N are not all independent. Any one of these variables is a function of the others

$$E = E(S, V, N), \qquad S = S(E, V, N), \qquad V = V(E, S, N), \qquad N = N(E, S, V) \tag{3}$$

The variables T, P, μ are then derived quantities, calculated by taking partial derivatives with respect to the independent variables using one of these functional forms.

We often use monatomic ideal gases to check these general relations. For example,, the entropy of a monatomic ideal gas is

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

Then

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{V,N} = \frac{3}{2}\frac{Nk_B}{E} \tag{5}$$

which is consistent with $E = \frac{3}{2}Nk_BT$ for a monatomic ideal case. Similarly,

$$P = T\left(\frac{\partial S}{\partial V}\right)_{E,N} = T\frac{Nk_B}{V} \tag{6}$$

which is the ideal gas law, and so on.

Since partial derivatives commute, we can derive additional relations with more derivatives:

$$\left(\frac{\partial T}{\partial V}\right)_{S,N} = \left(\frac{\partial}{\partial V}\left(\frac{\partial E}{\partial S}\right)_{V,N}\right)_{S,N} = \left(\frac{\partial}{\partial S}\left(\frac{\partial E}{\partial V}\right)_{S,N}\right)_{V,N} = -\left(\frac{\partial P}{\partial S}\right)_{V,N} \tag{7}$$

and so on. These type of relations among derivatives are known as **Maxwell relations**. Keep in mind that the general relations among E, S, V, N, T, P, μ hold for *any* system. Starting with the specification of a system in any ensemble (micro, canonical, or grand canonical), we can compute these quantities and the relations will hold. The first 3 equations in Eq. (2) are the *definitions* of T, μ and P. The others follow mathematically from these using multivariate calculus.

When we have a function like $S(V, E, N)$, we can see how the system responds when we change V, E and N. In physical situations, we are often much more interested in knowing how our system responds to changes in temperature (when we heat it) or pressure (when we compress it), or, for chemical reactions in particular, how to characterize equilibrium properties of a system when T and P are held fixed. So we would like the system to be described by functions that depend explicitly on T and P rather than E and V. In this lecture, we construct new variables (free energies) that depend on T and P. The four new potentials we introduce are

$$\textbf{Helmoltz free energy}: \quad F \equiv E - TS \tag{8}$$

$$\textbf{Enthalpy}: \quad H \equiv E + PV \tag{9}$$

$$\textbf{Gibbs free energy}: \quad G \equiv E + PV - TS \tag{10}$$

$$\textbf{Grand free energy}: \quad \Phi \equiv E - TS - \mu N \tag{11}$$

These are general definitions of these new variables, holding for arbitrary systems. In this lecture, we will motivate and construct these quantities and explore their physical significance.

As a warning, this business of dependent and independent variables is going to feel a bit awkward at first. That is because we are used to having a clear distinction between independent and dependent variables, while in statistical mechanics it is very handy to mix up the dependent and independent variables depending on context. Although awkward at first, this approach is very powerful and worth mastering. We'll try to be as clear about what is going on as possible.

As another piece of advice going forward: sequester in your mind all the complicated subtleties with entropy (ergodicity, Boltzmann's H theorem, Lodschmidt's paradox, Landauer's principle, etc.). If you really reach down deep to understand the foundations of the second law, all these things are important. However to *use* statistical mechanics, and thermodynamics, in physics, chemistry, astronomy and so on, we define systems based on macroscopic quantities (P, V, N, T, ...), entropy is extensive ($S_{\text{tot}} = S_1 + S_2$), the second law holds without subtlety ($\Delta S_{\text{tot}} \geqslant 0$), and μ, T and P are constant for systems in equilibrium.

2 Euler identity

Before getting started with all the new potentials, there is actually a very nice relation we can derive among all the different interdependent variables using only Eq. (1) and the fact that entropy is extensive.

From Eq. (1) we see that $S = S(E, V, N)$ (which we already knew since that's how we set up the microcanonical ensemble in the beginning). E, V, N and S are all extensive quantities. When we double the size of a system, they all double. This is in contrast to P, μ and T which are intensive quantities. When we double the system, they do not change.

It's actually very hard to make an extensive function. Any extensive function $f(x)$ must satisfy $f(cx) = cf(x)$ for any c. Differentiating with respect to c gives

$$x f'(cx) = f(x) \tag{12}$$

This holds for any c, so we can set $c = 1$ so $x f' = f$. The solution to this differential equation is then $f(x) = ax$ for some a. That is, f must be a linear function of its argument to be extensive.

Now let's generalize to a function of multiple variables. Extensivity requires

$$S(cE, cV, cN) = cS(E, V, N) \tag{13}$$

Differentiating both sides with respect to c gives

$$\left(\frac{\partial S}{\partial cE}\right)_{V,N} \frac{\partial cE}{\partial c} + \left(\frac{\partial S}{\partial cV}\right)_{E,N} \frac{\partial cV}{\partial c} + \left(\frac{\partial S}{\partial cN}\right)_{V,E} \frac{\partial cN}{\partial c} = S(E, V, N) \tag{14}$$

which simplifies to

$$S = \left(\frac{\partial S(cE, cV, cN)}{\partial cE}\right)_{V,N} E + \left(\frac{\partial S(cE, cV, cN)}{\partial cV}\right)_{E,N} V + \left(\frac{\partial S(cE, cV, cN)}{\partial cN}\right)_{V,E} N \tag{15}$$

$$= \frac{E}{T} + \frac{P}{T}V - \frac{\mu}{T}N \tag{16}$$

or equivalently

$$\boxed{E = TS - PV + \mu N} \tag{17}$$

This is known as the **Euler equation**. You can check it yourself for a monatomic ideal gas.

A related result comes from taking the total derivative of both sides

$$dE = d(\mathrm{TS}) - d(PV) + d(\mu N) = TdS + SdT - VdP - PdV + \mu dN + Nd\mu \tag{18}$$

Subtracting Eq. (1) we get

$$SdT - VdP + Nd\mu = 0 \tag{19}$$

This is called the **Gibbs-Duhem equation**. It says that T, P and μ are not independent – changing T and P makes μ change in a certain way. The Gibbs-Duhem equation generates a whole new set of partial derivative identifies such as

$$\left(\frac{dP}{dT}\right)_{\mu} = \frac{S}{V}, \quad \left(\frac{dT}{d\mu}\right)_{P} = -\frac{N}{S} \tag{20}$$

and so on.

The Euler equation and the Gibbs-Duhem equation hold for almost all statistical mechanical systems. Keep in mind, however, that the extensivity of entropy is not guaranteed by definition, and in some situations, where there are long-range interactions like gravity, as in stars, entropy is not extensive. Indeed, famously, for a black hole, entropy scales as the *area* of the black hole's event horizon, not the volume of the hole. That being said, for the vast majority of statistical mechanical systems we will consider, the Euler equation holds. To be safe, we will avoid using the Euler equation, but rather check that it holds in situations where entropy is indeed extensive.

3 Helmholtz Free Energy

We define the **Helmholtz free energy** as

$$F \equiv E - TS \tag{21}$$

Free energy is a concept particularly useful at constant temperature.

We can take the differential of F:

$$dF = dE - TdS - SdT = -PdV + \mu dN - SdT \tag{22}$$

We used $d(TS) = TdS + SdT$, which follows from the chain rule, in the first step and Eq. (1) in the second. Note that the dS has dropped out. The right-hand-side of Eq. (22) suggests that we should think of F as a function of V, N and T

$$F = F(V, N, T) \tag{23}$$

We say that V, N, T are the *natural* arguments of the Helmholtz free energy. We can then generate a whole new set of Maxwell relations, by picking two terms in Eq. (22):

$$\left(\frac{\partial F}{\partial V}\right)_{N,T} = -P, \quad \left(\frac{\partial F}{\partial N}\right)_{V,T} = \mu, \quad \left(\frac{\partial F}{\partial T}\right)_{V,N} = -S, \quad \left(\frac{\partial T}{\partial V}\right)_{F,N} = -\frac{P}{S} \tag{24}$$

and so on.

It is perhaps also worth pointing out to the mathematically-oriented crowd that the operation of replacing the dependent variable S in $E(S, V, N)$ with a new dependent variable given by a derivative, $T = \frac{\partial E}{\partial S}$, is known as a **Legendre transform**. Another example of a Legendre transform is going from a Lagrangian $L(\dot{q}, q)$ that depends on velocity $\dot{q}$ to a Hamiltonian $H(p, q)$ that depends on momentum $p = \frac{\partial L}{\partial \dot{q}}$. The Hamiltonian does not depend on $\dot{q}$ just as F does not depend on S. If you're not excited by the mathematics of it, the fact that we call it a Legendre transform is of no consequence. All the physics you need comes from the definition $F = E - TS$.

Before continuing, let me try to address a common pitfall. You might ask why is there no dE in Eq. (22)? Similarly, you could ask why is there no dT in Eq. (1)? The answer is that the non-trivial content in Eq. (1) is precisely that there is no dT (and no $d\mu$ or dP either). Of course T does vary, so if we have a nonzero dN, dV and dS then dT is probably nonzero as well. The point is that dT is not an *independent* variation. Eq. (1) says that we don't need to know what dT is to compute dE, we just need dS, dV and dN. Similarly, the content of Eq. (22) is that we don't need to know dE or dS, it is enough to know dV, dN and dT. E is a dependent variable in Eq. (22), so its variation is determined by the variation of the other, independent variables. Of course, we can always change variables and write $F(P, \mu, E)$ or whatever we want. However, there are not simple formulas for $\left(\frac{\partial F}{\partial P}\right)_{\mu, E}$ so writing F this way is not useful; it does not lead to any sort of simplification.

3.1 Free energy for work

The Helmholtz free energy is one of the most useful quantities in thermodynamics. Its usefulness stems from the fact that dV, dN and dT are readily measurable. This is in contrast to $E(S, V, N)$ which depends on entropy that is hard to measure and in contrast to $S(E, V, N)$ which depends on energy that is hard to measure. Helmholtz free energy is particularly powerful for systems at constant temperature where $dF = dE - TdS$. In previous courses you have studied mechanical systems using energy. Mechanical systems all fixed degrees of freedom, so $S = 0$ and $F = E$. Free energy is a generalization of energy whose importance is revealed by working at finite T.

Consider an isolated system in volume V kept at constant T by being in contact with a heat bath. Imagine that the system starts in some state at V and T and ends at the same V and T. We do not need to know anything about what happens between start and end other than that the only interaction with the surroundings is through heat exchange. We would like to know what is the maximum amount of work W the system can do in this situation? In general, energy is conserved, so we know $W + \Delta E_{\text{system}} - Q = 0$ where Q is the heat drawn in from the heat bath. The change in entropy of the bath is $\Delta S_{\text{bath}} = -\frac{Q}{T}$. The work causes no change in entropy, so by the second law of thermodynamics, $\Delta S_{\text{system}} \geqslant \frac{Q}{T}$, with the equality holding only if the transformation of the system is reversible. Then the free energy change is

$$\Delta F_{\text{system}} = \Delta E_{\text{system}} - T \Delta S_{\text{system}} \leqslant (-W + Q) - T\frac{Q}{T} = -W \tag{25}$$

That is,

$$W \leqslant -\Delta F_{\text{system}} \tag{26}$$

where the inequality becomes an equality if and only if the transformation is done reversibly. So the free energy of the system is depleted to do the work. This is why free energy is called free: it is the *energy available to do work*. In an insulated system (not heat exchange), the energy of the system is used for work, but in an isothermal system, it is the free energy that is used for work.

Note that it was important that the system be kept at constant volume. If the volume were to change, it would have to do work on the surrounding heat bath. Such work would be immediately dissipated as heat, causing an additional entropy increase. We will consider the more physically common situation of a constant pressure process in Section 5.

For another perspective, consider the case where no work is done, or where work is not relevant, such as when gases mix together, or some chemical reactions occur, or a system settles down after some perturbation. When $W = 0$, we get from Eq. (26) that $\Delta F_{\text{system}} \leqslant 0$. Thus, in a system kept at constant temperature and volume, interacting with the surroundings only through an exchange of heat (i.e. no work), the Helmholtz free energy never increases. As the system settles down towards equilibrium, F will decrease until equilibrium is reached when it stops decreasing (if it could decrease more by a fluctuation, it would, and then it could never go up again). Therefore, in an isolated system kept at constant temperature and volume, the *equilibrium is the state of minimum Helmholtz free energy*.

To be extra clear, let us emphasize that free energy refers to the *free energy of the system only*, $F = F_{\text{system}}$. So to find the equilibrium state we minimize the free energy of the system, ignoring the heat bath. Indeed, this is why free energy is powerful: it lets us talk about the system alone.

For a concrete example, consider a system of two gases separated by a partition, initially with different pressures P_1 and P_2 and different volumes V_1 and V_2 with $V_1 + V_2 = V$, in thermal contact with a heat bath. Then, since $F = F(T, N, V)$ and T and N are fixed, the minimization condition is

$$0 = dF = \left(\frac{\partial F}{\partial V_1}\right) dV_1 + \left(\frac{\partial F}{\partial V_2}\right) dV_2 = -P_1 dV_1 - P_2 d(V - V_1) = (P_2 - P_1) dV_1 \tag{27}$$

Thus the pressures are equal at equilibrium – if $P_1 \neq P_2$ then changing V would lower F. Of course, we knew this already; previously, we derived the pressure equality from maximization of the total entropy at constant energy. Here we are deriving it from minimization of free energy of the system at constant temperature. The two are equivalent. Indeed, the second law of thermodynamics is equivalent to the minimization of free energy. In general however, it is much easier to deal with systems at constant temperature than at constant entropy, and to minimize the free energy of the system rather than to maximize the *total* entropy.

In summary,

- **Free energy is to a constant $\boldsymbol{T}$ system what $\boldsymbol{E}$ is to a mechanical system.**
- **Helmholtz free energy is the available energy to do work at constant $\boldsymbol{T}$ and $\boldsymbol{V}$**.
- **In a system kept at constant $\boldsymbol{T}$ and $\boldsymbol{V}$, interacting with the surroundings only through an exchange of heat (i.e. no work), the Helmholtz free energy never increases.**
- **In an isolated system at constant $\boldsymbol{T}$ and $\boldsymbol{V}$, Helmholtz free energy is minimized in equilibrium**.
- **Free energy refers to the free energy *of the system only* $\boldsymbol{F} = \boldsymbol{F}_{\text{system}}$**.

Another important point to keep in mind is that free energy, like entropy, is a property of the ensemble. You cannot talk about the free energy or entropy of an individual microstate.

3.2 Free energy and the partition function

Next, consider how to compute free energy from the partition function in the canonical ensemble. Recall that in the canonical ensemble, $S = \frac{\langle E \rangle}{T} + k_B \ln Z$. So for an isolated system where $\langle E \rangle = E$ we immediately get that

$$\boxed{F = -k_B T \ln Z} \tag{28}$$

So the free energy *is* (- $k_B T$ times the logarithm of) the partition function. Equivalently, we can write Eq. (28) is

$$e^{-\beta F} = Z = \sum e^{-\beta E} \tag{29}$$

which shows that the free energy is the same as the energy if there is only one microstate. So the free energy gives a physical interpretation to the partition function.

For a monoatomic ideal gas recall that

$$Z = e^N \left(\frac{V}{N}\right)^N \left(\frac{2\pi m}{h^2 \beta}\right)^{\frac{3}{2}N} \tag{30}$$

So

$$F = -k_B T \ln Z = -N k_B T \left[\ln \frac{V}{N} + \frac{3}{2} \ln\left(\frac{2\pi m k_B T}{h^2}\right) + 1 \right] \tag{31}$$

This is much more complicated than $E = \frac{3}{2} N k_B$. The complication is because F has information about both energy and entropy, $F = E - TS$. As a check, we note that

$$\left(\frac{\partial F}{\partial T}\right)_{V,N} = -N k_B \left[\ln \frac{V}{N} + \frac{3}{2} \ln\left(\frac{2\pi m k_B T}{h^2}\right) + \frac{5}{2} \right] = -S \tag{32}$$

which, using that $E = \frac{3}{2} N k_B T$ for a monatomic ideal gas, agrees with Eq. (4). Also,

$$P = -\left(\frac{\partial F}{\partial V}\right)_{T,N} = \frac{N k_B T}{V} \tag{33}$$

in agreement with the ideal gas law.

3.3 Spring and gas

To build some more intuition for free energy, let's consider the change in F in various contexts. First, consider the free expansion of a gas from $V_1 \to V_2$. No work is done in free expansion. The distribution of molecular speeds is the same, they just have more room to move around in, so the internal energy of the gas doesn't change, $\Delta E = 0$. The entropy change is $\Delta S = N k_B \ln \frac{V_2}{V_1}$, as we saw in the discussion entropy of mixing. So the change in free energy is

$$\Delta F = \Delta E - T \Delta S = -N k_B T \ln \frac{V_2}{V_1} \tag{34}$$

In this case, the transition is completely driven by entropic considerations.

Now consider a piston on a spring, with force constant k, immersed in a heat bath (e.g. freely moving in air at room temperature). The spring only has 1 degree of freedom, the position of the piston. Let x be the displacement of the piston from its equilibrium position. Its energy is

$$E(x, \dot{x}) = \frac{1}{2} m \dot{x}^2 + \frac{1}{2} k x^2 \tag{35}$$

The first term is kinetic energy, the second is potential energy. The entropy of the spring is zero and doesn't change as the spring moves. As the spring moves the piston, it dissipates energy into the bath. Note however that it does no *work* on the bath: all the energy transferred goes to heating up the bath; there is no sense in which the energy transferred is useful in any sense, so it cannot be doing work. Thus this system qualifies for our general results about F decreasing. Since $S = 0$ for the piston+spring system, $F = E$. So minimizing F just corresponds to minimizing E. This free energy is minimized for $x = \dot{x} = 0$, the equilibrium position of the spring. Alternatively, we could define the system as the spring plus the heat bath. Then, $\Delta E = 0$ so $\Delta F = -T \Delta S_{\text{bath}} = -Q$. Thus free energy is minimized when the maximum amount of heat is transferred to the bath, i.e. the entire energy of the spring.

Now let's put the spring and the gas together. We'll start the gas off at volume V_1 and the spring at $x = 0$:

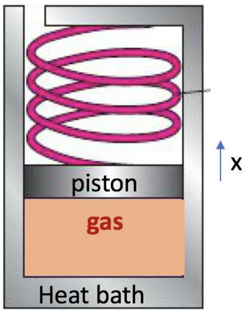


Figure 1. Spring-loaded piston against gas in a heat bath

We define the system as the piston+gas, and call the bath the surroundings. What is the equilibrium state?

To find the equilibrium, we want the free energy to be stationary when we vary x, so we need

$$\frac{\partial F}{\partial x} = \frac{\partial E}{\partial x} - T\frac{\partial S}{\partial x} \tag{36}$$

As the gas does work on the spring, it pulls heat in from the bath to stay at constant temperature. Since T is constant, the energy of the gas does not change. Thus, the energy in the gas does not depend on x. So the energy change of the system $\frac{\partial E}{\partial x}$ comes only from the spring, where we get $\frac{\partial E}{\partial x} = \mathcal{F}_{\text{piston}} = -kx$. We use the curly $\mathcal{F}$ for force to distinguish it from free energy. $\mathcal{F}_{\text{piston}}$ is the force acting on the gas by the piston. The piston doesn't have any entropy, so $\frac{\partial S}{\partial x}$ comes entirely from the gas. Now, the volume of the gas is $V = V_1 + Ax$ where A is the area of the piston. So

$$\frac{\partial S}{\partial x} = \frac{\partial S}{\partial V}\frac{\partial V}{\partial x} = \frac{P}{T}A = \frac{\mathcal{F}_{\text{gas}}}{T} \tag{37}$$

In the last step we used that pressure times area is force. So Eq. (36) becomes

$$\frac{\partial F}{\partial x} = \mathcal{F}_{\text{piston}} - \mathcal{F}_{\text{gas}} \tag{38}$$

Setting the variation of the free energy to zero, $\frac{\partial F}{\partial x} = 0$, implies $\mathcal{F}_{\text{piston}} = \mathcal{F}_{\text{gas}}$. This of course makes complete sense – we can compute the equilibrium point of the spring by demanding that the forces are equal, or we could find the equilibrium by minimizing the free energy. The result is the same.

3.4 Energy (non)-minimization

It is common lore to think of energy being minimized by physical systems: a ball rolls down to the bottom of a hill and stays there. But energy is conserved, so where does this common sense lore come from? It comes from free energy! All those systems in which energy is minimized are really minimizing free energy. You may never have thought about the gas in air surrounding the ball, but if it weren't for the gas, or the molecules in the dirt that can heat up due to friction, the ball would just roll right back up the hill.

The gas-spring example hopefully illustrates the point that there is no tendency to minimize energy. Total energy is conserved, in spontaneous motion, adiabatic motion, or whatever. The energy E in the definition $F = E - TS$ is not the total energy but rather the energy of the system. We assume the system is in thermal equilibrium, so energy in the form of heat can leave the system into the surroundings, or enter the system from the surroundings. Thus ΔE may not be zero, but the energy of system itself does not tend towards a minimum.

The tendency to minimize free energy is entirely because of a tendency to maximize entropy. This is clearest if we write

$$\Delta F = \Delta E - T\Delta S = -T\left[\underbrace{\left(-\frac{\Delta E}{T}\right)}_{\Delta S_{\text{surr}}} + \underbrace{\Delta S}_{\Delta S_{\text{sys}}}\right] \tag{39}$$

The first term $\frac{\Delta E}{T} = \frac{Q}{T}$ is the entropy change of the surroundings. The second term is the entropy change of the system. Their total is maximized, so entropy is maximized, and free energy is minimized. When we think of a spring with friction slowly stopping, we think it is minimizing energy. Indeed, it is minimizing energy, but that is because it is maximizing $-\frac{E}{T} = S$.

In summary, energy minimization is really free energy minimization, which really is entropy maximization.

4 Enthalpy

We use the symbol H for **enthalpy**. It is defined as

$$H \equiv E + PV \tag{40}$$

So, using Eq. (1),

$$dH = dE + PdV + VdP = TdS + VdP + \mu dN \tag{41}$$

Thus,

$$\left(\frac{\partial H}{\partial S}\right)_{P,N} = T, \qquad \left(\frac{\partial H}{\partial P}\right)_{S,N} = V, \qquad \left(\frac{\partial H}{\partial N}\right)_{P,S} = \mu \tag{42}$$

Enthalpy is a concept useful at constant pressure. It is a function $H = H(S, P, N)$.

Recall that we had two heat capacities related to how much temperature rises when heat dQ is put in at constant V or constant P

$$C_V = \left(\frac{\partial Q}{\partial T}\right)_V, \quad C_P = \left(\frac{\partial Q}{\partial T}\right)_P \tag{43}$$

For constant V, no work is done, since $W = PdV$. So $C_V = \left(\frac{\partial E}{\partial T}\right)_V$. For constant P, the gas has to expand when heat is absorbed to keep the pressure constant, so work is done and the energy goes down. The total energy change is $\Delta E = Q - P\Delta V$. In other words, $\Delta H = Q$. Thus,

$$C_P = \left(\frac{\partial H}{\partial T}\right)_P \tag{44}$$

Thus **enthalpy plays the role at constant pressure that energy does at constant volume**.

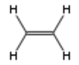
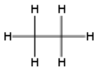
For a monatomic ideal gas, since $PV = Nk_BT$ and $E = \frac{3}{2}Nk_BT$ we get immediately that

$$H = E + PV = \frac{5}{2}Nk_BT \tag{45}$$

so, from Eq. (44) $C_P = \frac{5}{2}Nk_B$ in agreement with what we found in Lecture 5.

4.1 Enthalpy of chemical bonds

Enthalpy is especially useful in chemistry, where pressure is nearly always constant but volume is not under control. Constant pressure is very common – chemicals in a solution are at constant pressure, as are biological reactions in cells. Since chemistry (and biology) takes place at constant pressure and temperature, when a chemical reaction occurs the heat released is $Q = \Delta H$. Thus when you measure the heat released or absorbed in a reaction, you are measuring the enthalpy.

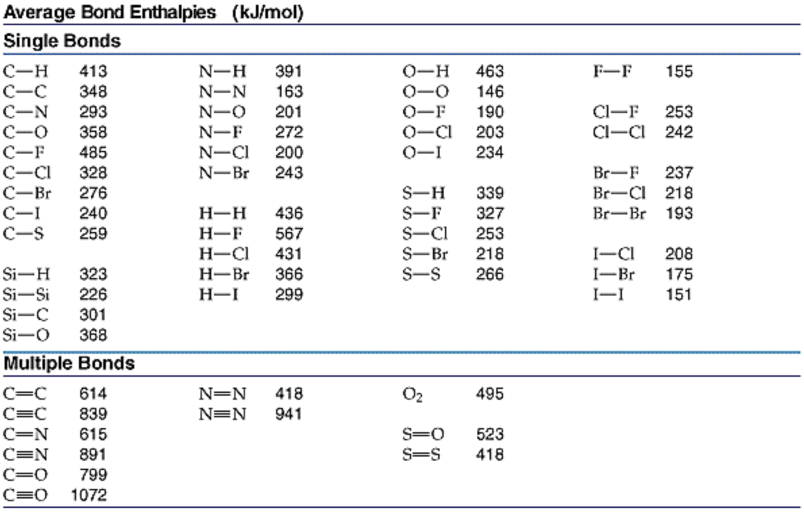
There are different ways to compute the enthalpy of a reaction, to different degrees of approximation. Consider for example the hydrogenation of ethene= (C_2H_4)=  into ethane (C_2H_6)=  :

$$H_2 + C_2H_4 \rightarrow C_2H_6 \tag{46}$$

First of all, you can just look up the enthalpy for this reaction. That is called the **standard enthalpy of reaction**, where standard means under some reference conditions, like $298K$ and 1 atm and denoted by °. For this reaction, the reaction enthalpy is $\Delta_r H^\circ = -136.3 \pm 0.3 \frac{\text{kJ}}{\text{mol}}$. Since the enthalpy change is negative, the reaction is exothermic (heat is released).

If we don't know the reaction enthalpy, we can compute it by taking the differences of the enthalpies of the products and the reactants. So in this case, we look up that the **enthalpy of formation** of each. The enthalpy of formation is the enthalpy change in forming something from its constituent atoms (C, S, Si) or diatomic molecules (H_2,O_2,N_2,F_2). So the enthalpy of formation of C or H_2 is 0, by definition. The enthalpy of formation for C_2H_4 is $\Delta_f H^\circ = 52.4 \pm 0.5 \frac{\text{kJ}}{\text{mol}}$ and for C_2H_6 is $\Delta_f H^\circ = -84 \pm 0.4 \frac{\text{kJ}}{\text{mol}}$. It takes energy to make ethane, but energy is released when ethane is formed. Combining these, the enthalpy of the reaction is $\Delta_r H^\circ = -84 \frac{\text{kJ}}{\text{mol}} - 52.4 \frac{\text{kJ}}{\text{mol}} = -136.4 \frac{\text{kJ}}{\text{mol}}$ in agreement with the directly measured enthalpy of reaction. That we can add and subtract enthalpies to get the reaction enthalpy is known as **Hess's law**. There are lots of laws in chemistry.

If we don't have the enthalpy's of formation handy, our last resort is to compute the enthalpy change ourselves by adding up the enthalpy of each bond. There are only a finite number of relevant bonds for organic chemistry, so we can just make a table of them:



Average Bond Enthalpies (kJ/mol)

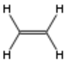
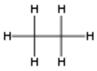
Single Bonds

C—H	413	N—H	391	O—H	463	F—F	155
C—C	348	N—N	163	O—O	146		
C—N	293	N—O	201	O—F	190	Cl—F	253
C—O	358	N—F	272	O—Cl	203	Cl—Cl	242
C—F	485	N—Cl	200	O—I	234		
C—Cl	328	N—Br	243			Br—F	237
C—Br	276			S—H	339	Br—Cl	218
C—I	240	H—H	436	S—F	327	Br—Br	193
C—S	259	H—F	567	S—Cl	253		
		H—Cl	431	S—Br	218	I—Cl	208
Si—H	323	H—Br	366	S—S	266	I—Br	175
Si—Si	226	H—I	299			I—I	151
Si—C	301						
Si—O	368						

Multiple Bonds

C=C	614	N=N	418	O_2	495
C≡C	839	N≡N	941		
C=N	615			S=O	523
C≡N	891			S=S	418
C=O	799				
C≡O	1072				

Figure 2. Average enthalpies of common covalent bolds. Numbers are listed as positive, as enthalpies required to break the bonds.

So C_2H_4=  has one $C = C$ double bond, $H_b = -614 \frac{\text{kJ}}{\text{mol}}$ and 4 CH bonds each with $H_b = -413 \frac{\text{kJ}}{\text{mol}}$ giving it an enthalpy of $H_{C_2H_4} = -2266 \frac{\text{kJ}}{\text{mol}}$. C_2H_6 =  has 6 CH bonds and one C-C single bond, giving it $H_{C_2H_6} = -2826 \frac{\text{kJ}}{\text{mol}}$. H_2 has $H_{H_2} = -436 \frac{\text{kJ}}{\text{mol}}$. So the net enthalpy change in the reaction is $H_{C_2H_6} - H_{C_2H_4} - H_{H_2} = -124 \frac{\text{kJ}}{\text{mol}}$. This is not too far off from the reaction value of $\Delta_r H^\circ = -136.3 \frac{\text{kJ}}{\text{mol}}$ but not terribly close either. The approximation that all covalent bonds in any material have the same enthalpy is apparently not terrific.

Also keep in mind that bond enthalpies are tabulated for gases. For liquids and solids, intermolecular interactions cannot be neglected and the bond enthalpies are not enough. For liquids and solids, there are still useful tabulated enthalpies of formation and reaction that can be used and will give better estimates than adding the bond enthalpies.

The bottom line is

- Using enthalpies of reaction is best (when available).
- Using enthalpies of formation and Hess' law is the most convenient compromise, giving nearly identical to reaction enthalpies and requiring less tabulated data.

- Using bond enthalpies is a last resort, only recommended when formation enthalpies are not available.

4.2 Expansion work

A natural question to ask is whether the difference between enthalpy and energy actually even matters. The difference between enthalpy and energy is $\Delta(\mathrm{PV})$. To see how big this is, relative to the enthalpy or energy, let's consider some examples.

For a first example, consider chemical reactions in solids. Clam shells have a layer of calcite and a layer of aragonite. These minerals are both naturally occurring forms of calcium carbonate $\mathrm{CaCO_3}$. They both have the same chemical composition, but different crystal structures. Aragonite forms from calcite at high pressure, but at typical pressures on the surface of the earth, aragonite is unstable and turns into calcite on the 10 million year timescale. The two minerals have different densities: aragonite is more dense, at $2.93 \frac{g}{\mathrm{cm}^3}$ than calcite at $2.71 \frac{g}{\mathrm{cm}^3}$. Thus when aragonite converts into calcite, its volume expands, doing work and an enthalpy $\Delta H = 0.21 \frac{\mathrm{kJ}}{\mathrm{mol}}$ is released at $P = 1\,\mathrm{atm}$. The volume change per unit mass is

$$\frac{\Delta V}{m} = \left(\frac{1}{2.71 \frac{g}{\mathrm{cm}^3}} - \frac{1}{2.93 \frac{g}{\mathrm{cm}^3}}\right) = 0.028 \frac{\mathrm{cm}^3}{g} \tag{47}$$

Now 1 mole of $\mathrm{CaCO_3}$ weights 100g, and using $P = 1\,\mathrm{bar} = 10^5 \mathrm{Pa} = 10^5 \frac{J}{m^3} = 10^{-4} \frac{\mathrm{kJ}}{\mathrm{cm}^3}$ we get

$$P\Delta V = 10^{-4} \frac{\mathrm{kJ}}{\mathrm{cm}^3} \left(0.028 \frac{\mathrm{cm}^3}{g}\right) \times \frac{100\,g}{\mathrm{mol}} = 2.8 \times 10^{-4} \frac{\mathrm{kJ}}{\mathrm{mol}} \tag{48}$$

So in this case $P\Delta V \ll \Delta H$ and therefore the enthalpy and energy changes are nearly identical. The work done is only a small fraction, 0.1%, of the enthalpy change.

Volume change is more important when the total number of molecules is not the same on both sides of a reaction. For example, consider the enthalpy change in the formation of ammonia gas, $\mathrm{NH_3}$, through the reaction of hydrogen and nitrogen gases:

$$3H_2 + N_2 \rightarrow 2\mathrm{NH_3} \tag{49}$$

This reaction converts 4 molecules into 2, so $\Delta n = -2$, so the volume will go down. Using the ideal gas law at room temperature

$$\Delta(\mathrm{PV}) = (\Delta n) RT = -2 \times 8.3 \frac{J}{\mathrm{mol}\,K} \times 298\,K = -4.9 \frac{\mathrm{kJ}}{\mathrm{mol}} \tag{50}$$

Let us compare this to the enthalpy change. Computing the enthalpy change by adding the bond enthalpies gives $\Delta H = -97 \frac{\mathrm{kJ}}{\mathrm{mol}}$ for this reaction, not far off from the measured reaction enthalpy change of $\Delta_r H° = -91.88 \frac{kJ}{\mathrm{mol}}$. So we find that, $\Delta(\mathrm{PV}) = -4.5 \frac{\mathrm{kJ}}{\mathrm{mol}}$ is around 5% of the total enthalpy change in this case. 5% is small, but not so small that it can be neglected. Indeed, a 5% change in the energetics can have important effects on reaction kinetics.

In summary, as a rule of thumb, enthalpies and energies are pretty similar for solids and liquids, differing at the less than a percent level, but for gases the difference can be relatively large. The difference between the energy and enthalpy change is essentially given by $\Delta H - \Delta E = (\Delta n) RT$ with Δn the change in the number of moles of *gas*. This is equal to $2.48 \frac{\mathrm{kJ}}{\mathrm{mol}} \Delta n$ at room temperature and pressure. These enthalpy changes are included in tabulated enthalpies of formation and reaction (which are different for liquids and gases) and are also included in bond enthalpies (defined for gases and computed by averaging the formation enthalpies of various molecules with that bond).

5 Gibbs free energy

Gibbs free energy is defined as

$$G \equiv H - TS = E + PV - TS \tag{51}$$

The differential of G is

$$dG = dE + d(PV) - d(TS) = VdP - SdT + \mu dN \tag{52}$$

so $G = G(P, N, T)$ and

$$\left(\frac{\partial G}{\partial T}\right)_{P,N} = -S, \qquad \left(\frac{\partial G}{\partial P}\right)_{T,N} = V, \qquad \left(\frac{\partial G}{\partial N}\right)_{P,T} = \mu \tag{53}$$

Gibbs free energy is the constant pressure version of Helmholtz free energy; it is $E - TS$ with E replaced by H. Recall that Helmholtz free energy is useful at constant volume and constant temperature. At constant pressure, as in chemistry and biology, enthalpy and Gibbs free energy are used. Gibbs free energy gives the maximum amount of work that can be done at constant pressure and temperature. At constant pressure and constant temperature, Gibbs free energy has its minimum value at equilibrium.

A powerful function of the Gibbs free energy is that it tells which direction a reaction will go. In a chemical reaction at constant T and P, the amount of heat released is given by the enthalpy change $\Delta_r H$. To get the sign right, note that if $\Delta_r H = H_{\text{products}} - H_{\text{reactants}}$ is positive, $\Delta_r H > 0$, then the products have more enthalpy than the reactants, so the surroundings must put in energy and heat is withdrawn. The entropy change in the surroundings (the air or solution or whatever heat bath is fixing T and P in the first place) is $\Delta S_{\text{surround}} = \frac{Q_{\text{out}}}{T} = -\frac{\Delta_r H}{T}$. The entropy change in the system is $\Delta_r S$ so the total entropy change is $\Delta S = \Delta S_{\text{sys}} + \Delta S_{\text{surround}} = \Delta_r S - \frac{\Delta_r H}{T} = -\frac{\Delta_r G}{T}$. Since total entropy always increases the reaction can only proceed if the Gibbs free energy change is *negative*: $\Delta_r G < 0$. Thus the sign of the Gibbs free energy change indicates which way the reaction will spontaneously proceed.

For an ideal monatomic gas, we want to express G as a function of P, N and T. We start by using Eq. (31) to write

$$G = F + PV = -Nk_B T\left[\ln\frac{V}{N} + \frac{3}{2}\ln\left(\frac{2\pi m T}{h^2}\right) + 1\right] + PV \tag{54}$$

Then we convert V to P using the ideal gas law $V = \frac{N k_B T}{P}$ so

$$G = -Nk_B T\left[\ln\frac{k_B T}{P} + \frac{3}{2}\ln\left(\frac{2\pi m T}{h^2}\right)\right] \tag{55}$$

Then we find

$$\mu = \left(\frac{\partial G}{\partial N}\right)_{P,T} = \frac{G}{N} = k_B T \ln n\lambda^3 \tag{56}$$

with $\lambda = \frac{h}{\sqrt{2\pi m k_B T}}$ the thermal wavelength, in agreement with our previous result. This explicit calculation confirms that for a monatomic ideal gas the chemical potential is the Gibbs free energy per molecule.

Note that we found that $G = \mu N$ for the ideal gas. This relation is actually very general. It follows immediately from the definition $G = E + PV - TS$ and the Euler relation in Eq. (17) (which relied on the extensivity of entropy). When there are multiple species, $G = \sum \mu_i N_i$. Thus, similar to how the partition function and the Helmholtz free energy were equivalent, the chemical potential and the Gibbs free energy are equivalent.

5.1 Partial pressure

If we have a gas with N_1 molecules of type 1 and N_2 molecules of type 2, then they both exert pressure on the walls of the container. We call the pressure due to molecules of type i the **partial pressure** for that type and denote it P_i. The ideal gas law $PV = N k_B T$ holds for a single gas or for a mixture of gases. So say there are two gases with $N_1 + N_2 = N$. Then by the ideal gas law, the partial pressure of gas one is

$$P_i \equiv \frac{N_i}{V} k_B T \quad \text{(ideal gases)} \tag{57}$$

Since the ideal gas law $PV = Nk_BT$ holds for the whole mixture, we can divide by it giving

$$\boxed{\frac{P_i}{P} = \frac{N_i}{N}} \tag{58}$$

This relation is called **Dalton's law**. We derived it for ideal gases where the pressure is linear in the number density but it also holds empirically for many interacting systems such as liquids.

Next we can define the **partial Gibbs free energy** for one monatomic ideal gas in a mixture as that for a single gas, with N replaced by N_i and P replaced by P_i:

$$G_i = -N_i k_B T\left[\ln\left(\frac{k_B T}{P_i}\right) + \frac{3}{2}\ln\left(\frac{2\pi m T}{h^2}\right)\right] \tag{59}$$

This definition is useful because it makes Gibbs free energy extensive:

$$G = G_1 + G_2 + \cdots \tag{60}$$

To check this, consider mixing two gases with the same mass, like He^3 and He^4 gas, with $N_1 = N_2 = \frac{N}{2}$. Then $\frac{N}{N_i} = 2$

$$G_1 + G_2 = -Nk_B T\left[\ln\left(2\frac{k_B T}{P}\right) + \frac{3}{2}\ln\left(\frac{2\pi m T}{\hbar^2}\right)\right] \tag{61}$$

$$= -Nk_B T\left[\ln\frac{k_B T}{P} + \frac{3}{2}\ln\left(\frac{2\pi m T}{\hbar^2}\right)\right] - Nk_B T\ln 2 \tag{62}$$

The first term on the second line is the Gibbs free energy for a gas of N identical particles of mass m. The final term can be written as$-Nk_BT\ln 2 = -T\Delta S$, where $\Delta S = Nk_B\ln 2$. This extra term is exactly the entropy of mixing. It is part of the Gibbs free energy since the mixed system comprises two different types of particles, and so has more entropy (and less Gibbs free energy), then a single homogeneous gas.

5.2 Law of mass action

Now we want to generalize away from a monatomic ideal gas to an arbitrary mixture of general ideal gases. We can write the partition function for a single molecule of an ideal gas as

$$\zeta = Z_{\text{single molecule}} = \sum_{\varepsilon} e^{-\beta\varepsilon} = \int \frac{d^3x\, d^3p}{h^3} \sum_{\text{internal states j}} e^{-\beta\left(\frac{\vec{p}^2}{2m} + \varepsilon_j\right)} \tag{63}$$

The energy ε_j includes all contributions to the energy other than momentum. For example, including chemical binding energy ε_b, kinetic degrees of freedom and a vibrational mode of frequency ω

$$\zeta = Ve^{-\beta\varepsilon_b}\left(\frac{2\pi m}{\beta h^2}\right)^{\frac{3}{2}} \frac{1}{2\sinh\left(\frac{\beta}{2}\hbar\omega\right)} = \frac{V}{\lambda^3} e^{-\beta\varepsilon_b} \frac{1}{2\sinh\left(\frac{\beta}{2}\hbar\omega\right)} \tag{64}$$

The volume factor comes from the d^3x integral. This integral is always there, so ζ will always scale linearly with V. A key element of ζ is the binding energy ε_b which must be treated with a consistent zero-point for all the molecules involved in the mixture. In general ζ is complicated for molecules, and we will not try to compute it. As we'll see, we don't need to, because we can measure it. The binding energy part is included in the standard enthalpy of formation that we can look up.

If there are N_1 molecules, then the possible total energies are sums of the energies of the individual molecules. The partition function is then the product of the single particle partition functions divided by the identical particle factor N_1!:

$$Z_1 = \frac{1}{N_1!} \sum_{\varepsilon_1 \cdots \varepsilon_{N_1}} e^{-\beta(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{N_1})} = \frac{1}{N_1!}(Z_{\text{singlemolecule}})^{N_1} = \frac{1}{N_1!}\zeta_1^{N_1} \tag{65}$$

If there are two gases, we multiply their partition functions. So

$$Z = Z_1 Z_2 = \frac{1}{N_1!}\zeta_1^{N_1} \frac{1}{N_2!}\zeta_2^{N_2} \tag{66}$$

That the partition function is multiplicative in this way follows from assuming the gases are non-interacting: the energy of each molecule is independent of the other molecules. We are not making any other assumption about the molecules though, such as how many degrees of freedom there are.

The Helmholtz free energy of the two gas mixture is then

$$F = -k_B T \ln Z \approx -k_B T\left[N_1\left(\ln\frac{\zeta_1}{N_1}+1\right)+N_2\left(\ln\frac{\zeta_2}{N_2}+1\right)\right] \tag{67}$$

where Stirling's relation $N! \approx e^{-N}N^N$ was used in the $\approx$ step. The chemical potential for gas 1 is then

$$\mu_1 = \left(\frac{\partial F}{\partial N_1}\right)_{T,V,N_2} = -k_B T \ln\frac{\zeta_1}{N_1} \tag{68}$$

and similarly for gas 2. To make sure we haven't messed anything up, the Gibbs free energy is

$$G = F + PV = -k_B T\left[N_1\left(\ln\frac{\zeta_1}{N_1}+1\right)+N_2\left(\ln\frac{\zeta_2}{N_2}+1\right)\right] + \underbrace{(N_1+N_2)k_B T}_{\text{from ideal gas law}} \tag{69}$$

$$= \mu_1 N_1 + \mu_2 N_2 \tag{70}$$

so that $G = \sum_i \mu_i N_i$ as expected.

It is helpful to express the chemical potential in terms of the total number of particles and the fraction $\frac{N_i}{N}$. We can easily do this by rewriting the logarithm in Eq. (68):

$$\mu_i = -k_B T \ln\frac{\zeta_i}{N_i} = -k_B T \ln\frac{\zeta_i}{N} + k_B T \ln\frac{N_i}{N} \tag{71}$$

$$= -k_B T \ln\frac{k_B T \zeta_i}{PV} + k_B T \ln\frac{N_i}{N} \tag{72}$$

Thus,

$$\boxed{\frac{N_i}{N} = \exp\left[\frac{\mu_i - G_{0i}}{k_B T}\right]} \tag{73}$$

where, using $N = \frac{PV}{k_B T}$,

$$G_{0i} = -k_B T \ln\frac{\zeta_i}{N} \tag{74}$$

G_{0i} is the Gibbs free energy for a molecule in a mixture. For a single molecule $N_i = N = 1$ and so $G_{0i} = \mu_i = 1 \times G$.

Recalling that $G = G(N, P, T)$ we should make sure to have ζ_i depend on these quantities, so for the mocule with vibrational mode and ground state energy,

$$G_{0i}(P,T) = -k_B T \ln\left[\frac{kT}{P}\frac{1}{\lambda^3}e^{-\beta\varepsilon_b}\frac{1}{2\sinh\left(\frac{\beta}{2}\hbar\omega\right)}\right] \tag{75}$$

Note that, critically, G_{0i} depends on the total pressure P not the partial pressure P_i. If G_{0i} is evaluated at the partial pressure P_i then it would be the partial Gibbs free energy per particle: $G_{0i}(P_i, T) = \mu_i$. The difference between $N_i G_{0i}$ and $G_i = N_i \mu_i$ is exactly the entropy of mixing terms, as in Eq. (62).

Eq. (73) is like the relation $n = \frac{1}{\lambda^3}\exp\left(\frac{\mu-\varepsilon}{k_B T}\right)$ we derived last lecture for a monatomic ideal gas. Indeed,if you substitute $\zeta_i = \frac{V}{\lambda^3}e^{-\beta\varepsilon_b}$ for a monatomic ideal gas, you can check that Eq. (73) reduces to $n_i = \frac{1}{\lambda^3}\exp\left(\frac{\mu_i - \varepsilon_b}{k_B T}\right)$ exactly. The difference is Eq. (73) is more general than monatomic ideal gases. We have also written the formula in terms of the fraction $\frac{N_i}{N}$ rather than $n_i = \frac{N_i}{V}$ and we have G_{0i} written with pressure explicit rather than volume (so we can hold P fixed).

From here we get a relation between the fractions of reactants. Let us use the notation

$$x_j \equiv \frac{N_j}{N} \tag{76}$$

for the molar fraction of a reactant. Then, for something like

$$2A+3B\leftrightharpoons 7C \tag{77}$$

for which $2\mu_A+3\mu_B=7\mu_C$ we would find that the μ_i drop out from the combination

$$\frac{x_C^7}{x_A^2 x_B^3}=\exp\left[-\frac{\Delta_r G_0}{k_B T}\right] \tag{78}$$

this is the **law of mass action**. In the general form, the powers on the left hand side are determined by the stoichiometric coefficients in the reaction equation and $\Delta_r G_0$ is the change in Gibbs free energy per particle (i.e. $\Delta_r G_0=7G_{0A}-2G_{0B}-3G_{0C}$ for this example).

Recall that we derived the law of mass action in the previous lecture for monatomic ideal gases, where the exponent had the energy change $\Delta\varepsilon$ per particle and the left hand side had the concentrations $[A]=\frac{N_j}{V}$ rather than the molar fractions. That previous formula is a special case of this more general mass action formula, as you can check.

To understand the law of mass action, recall that the *total* Gibbs free energy change in a reaction at equilibrium is zero, $\Delta G=0$; otherwise, G could be minimized by moving molecules from one side to another. The overall ΔG has a part *per molecule*, which is the Gibbs reaction energy $\Delta_r G$ and a part that depends on the concentrations, encoded in the x_i fractions. This second part is entirely entropic, driven by the entropy of mixing, Thus the law of mass action says that in equilibrium these two contributions to ΔG exactly cancel. It thereby lets us figure out the equilibrium concentrations from the Gibbs reaction energy per particle.

By Dalton's law, we also have $x_i=\frac{P_i}{P}$ so Eq. (78) can also be thought of as an equation for the equilibrium partial pressures. Chemists also prefer to use moles rather than particles, so we use $\Delta_r G$, the **Gibbs reaction energy** (in $\frac{\text{kJ}}{\text{mol}}$) and use RT instead of $k_B T$. The ratio of fractions is also given a name, **the equilibrium constant** and the symbol K:

$$K\equiv\frac{x_C^7}{x_A^2 x_B^3}=\frac{\left(\frac{N_C}{N}\right)^7}{\left(\frac{N_A}{N}\right)^2\left(\frac{N_B}{N}\right)^3}=\frac{1}{P^4}\frac{P_C^7}{P_A^2 P_B^3}=e^{-\frac{\Delta_r G_0}{k_B T}}=e^{-\frac{\Delta_r G}{RT}} \tag{79}$$

Let's do an example. In the reaction $H_2O(g)\rightleftharpoons H_2(g)+\frac{1}{2}O_2(g)$ at $T=5000K$ the Gibbs reaction energy is $\Delta_r G=118\frac{\text{kJ}}{\text{mol}}$. If we start with 1 mol of H_2O it will decompose into α moles of H_2 and $\frac{\alpha}{2}$ moles of O_2 leaving $1-\alpha$ moles of H_2O. The total number of moles is $1-\alpha+\alpha+\frac{\alpha}{2}=1+\frac{\alpha}{2}$. Then

$$\frac{x_{H_2}x_{O_2}^{1/2}}{x_{H_2O}}=\frac{\left(\frac{\alpha}{1+\frac{\alpha}{2}}\right)\sqrt{\frac{\alpha}{2(1+\frac{\alpha}{2})}}}{\frac{1-\alpha}{1+\frac{\alpha}{2}}}=\exp\left[-\frac{118\frac{\text{kJ}}{\text{mol}}}{8.3\frac{J}{\text{mol}\,K}\times 5000K}\right]=0.058 \tag{80}$$

Solve numerically for α gives $\alpha=0.17$. Thus at 5000K, 17% of the water molecules will be decomposed into H_2 and O_2. At room temperature $\Delta_r G=228\frac{\text{kJ}}{\text{mol}}$ and $\alpha=10^{-27}$.

This example indicates an important qualitative point about using the Gibbs free energy: generally $\Delta_r G$ is of order hundreds of $\frac{\text{kJ}}{\text{mol}}$ while $RT=2.5\frac{\text{kJ}}{\text{mol}}$ at room temperature. So the factor $\exp\left(-\frac{\Delta_r G}{RT}\right)$ is almost always either very very small, if $\Delta_r G>0$ or very very large if $\Delta_r G<0$. Thus for **exergonic** reactions ($\Delta_r G<0$), the reaction strongly favors the products, while for **endergonic** reactions ($\Delta_r G>0$), the reactants are favored. Thus to a good approximation, we can use the rule of thumb that

- If we mix some chemicals, the sign of $\Delta_r G$ tells which way the reaction will go.

This rule of thumb works only when the concentrations are not exponentially small so that the $\exp\left(-\frac{\Delta_r G}{RT}\right)$ dominates. Of course, if the system is in equilibrium, it will not proceed in any direction. Or if the concentration of products (or reactants) is small enough, the reaction will proceed in the only direction it can. In such a case, when the reaction proceeds against the direction of $\Delta_r G$, the total Gibbs free energy of the system is still decreasing: the entropy of mixing associated with the concentration imbalance dominates over the $\Delta_r G$ effect from the reaction itself (remember, we pulled the concentration-dependence out in Eq.(71)).

In summary, the law of mass action always tells us which way the reaction will proceed given some initial concentrations. The rule-of-thumb only determines the reaction direction when the concentrations are not exponentially small.

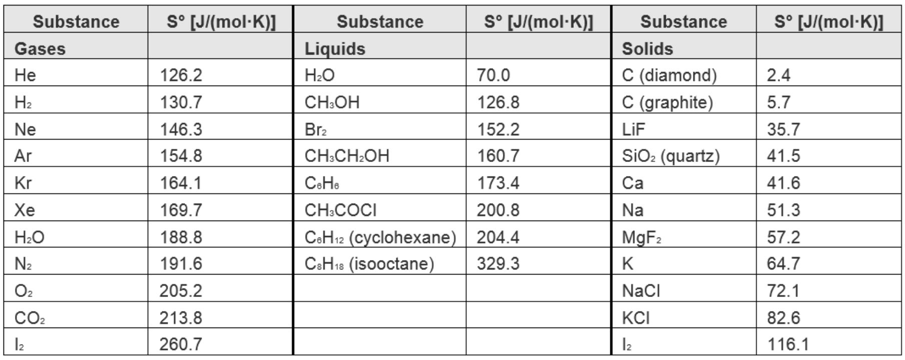
5.3 Direction of chemical reactions

We saw that the direction a reaction proceeds is determined by the sign of $\Delta_r G$. What do we know about this sign? Although G_{0i} and $\Delta_r G$ are in principle computable from the partition function for a single molecule, this is never actually done. One can easily look up $\Delta_r G$ under standard conditions $P = 1\,\text{bar}$ and $T = 298K$. This standard value of the Gibbs reaction energy is denoted $\Delta_r G°$. It is more useful however to look up $\Delta_r H$ and $\Delta_r S$ and compute $\Delta_r G$ via:

$$\Delta_r G = \Delta_r H - T\Delta_r S \tag{81}$$

The reason this formula is useful is because both $\Delta_r H$ and $\Delta_r S$ are generally *weakly* dependent on temperature (less then any power of T), while $\Delta_r G$ *depends strongly on* T because of the explicit factor of T in Eq. (81). So Eq. (81) lets us compute $\Delta_r G$ at any temperature, while $\Delta_r G°$ only gives the value at one reference temperature.

The dominant contribution to $\Delta_r H$ is from bonds breaking and forming, with a subleading contribution from $\Delta(\text{PV}) \approx \Delta n_{\text{gas}} R T$. What determines $\Delta_r S$? Or more practically, how can we measure $\Delta_r S$? One way is to ask at what temperature the reaction is in equilibrium. Then we know $\Delta_r G = 0$ and we can compute $\Delta_r S = \frac{\Delta_r H}{T}$. Alternatively, we could note the heat Q given off in the reaction and use $\Delta S_{\text{surroundings}} = \frac{Q}{T}$; if we can reverse the reaction adiabatically and isothermally doing work W, then $\Delta S = \frac{Q - W}{T}$. Conveniently, chemists have measured the entropy of enough reactions to tabulate **standard molar entropies.** For example,



Substance	S° [J/(mol·K)]	Substance	S° [J/(mol·K)]	Substance	S° [J/(mol·K)]
Gases		Liquids		Solids	
He	126.2	H_2O	70.0	C (diamond)	2.4
H_2	130.7	CH_3OH	126.8	C (graphite)	5.7
Ne	146.3	Br_2	152.2	LiF	35.7
Ar	154.8	CH_3CH_2OH	160.7	SiO_2 (quartz)	41.5
Kr	164.1	C_6H_6	173.4	Ca	41.6
Xe	169.7	CH_3COCl	200.8	Na	51.3
H_2O	188.8	C_6H_{12} (cyclohexane)	204.4	MgF_2	57.2
N_2	191.6	C_8H_{18} (isooctane)	329.3	K	64.7
O_2	205.2			NaCl	72.1
CO_2	213.8			KCl	82.6
I_2	260.7			I_2	116.1

Figure 3. Standard molar entropies for some compounds at 298 K

Then the change in entropies can be computed by taking the difference of the standard molar entropies of the products and reactants. Note from the table that

- More complex molecules have higher entropies.
- Gases have higher entropies than liquids which have higher entropies than solids.

These observations are consistent with our understanding of entropy as measuring disorder.

Let's consider an example. **Limestone** is a commonly occurring mineral, calcium carbonate $CaCO_3$. By itself, limestone is not so useful, but can be converted to **lime**, calcium oxide CaO in a kiln. Lime is an extremely useful mineral, used in making steel, mortar and cement and in agriculture. To make lime from limestone involves the reaction

$$CaCO_3 \leftrightharpoons CaO + CO_2 \tag{82}$$

The enthalpy change in this reaction is $\Delta_r H = 178\,\frac{\text{kJ}}{\text{mol}}$.The entropy change is $\Delta_r S = 161\frac{J}{K\,\text{mol}}$. At room temperature, $T = 298K$, so

$$\Delta_r G = \Delta_r H - (298K)\,\Delta_r S = 130\,\frac{\text{kJ}}{\text{mol}} \tag{83}$$

Since the reaction increases the Gibbs free energy, it does not spontaneously occur.

On the other hand, if we heat the limestone in a lime kiln, to $T = 1500K$, then

$$\Delta_r G = \Delta_r H - (1500K)\,\Delta_r S = -64\,\frac{\mathrm{kJ}}{\mathrm{mol}} \tag{84}$$

Thus the kiln allows the reaction to occur. Note that there is also some temperature dependence to $\Delta_r H$ from $\Delta(\mathrm{PV}) = \Delta n k_B T = RT = 8.3\frac{J}{K\,\mathrm{mol}}$. Since $\Delta(\mathrm{PV}) \ll \Delta_r S$ we can neglect the temperature dependence of $\Delta_r H$ compared to $T\Delta_r S$ term.

Normally heating up a system makes a reaction go faster. Taking milk out of the fridge makes it go bad faster. But milk would go bad eventually even if left in the fridge. With the lime kiln, heating the system does not just make the reaction go faster. *It changes the direction of the reaction.* The reverse reaction naturally occurs at room temperature – lime will eventually turn back into limestone if left in the presence of $\mathrm{CO_2}$.

Note that we have assumed that all of the T dependence is in the explicit factor of T in the definition $\Delta_r G = \Delta_r H - T\Delta_r S$. What about the temperature dependence of $\Delta_r H$ and $\Delta_r S$ themselves? We can write $\Delta_r H = \Delta_r E + \Delta_r(\mathrm{PV})$. The $\Delta_r E$ contribution is from breaking bonds, which is independent of T. Solids and liquids have the same volume, so $\Delta_r(\mathrm{PV}) = (\Delta n_{\mathrm{gas}})RT \approx 8\frac{J}{\mathrm{mol}\,K}\Delta n_{\mathrm{gas}}T$. Although this depends linearly on T, just like the $T\Delta_r S$ factor in $\Delta_r G$, the numerical values for $\Delta_r S \approx 161\frac{J}{\mathrm{mol}\,K}$ in this example are much bigger. Thus we can neglect the temperature dependence of $\Delta_r H$ when solids and liquids are involved; for gases, it is a subleading effect. The temperature dependence of $\Delta_r S$ is generally logarithmic, since $S \sim N k_B \ln T$ and generally very small.

In the above discussion of $\mathrm{CaCO_3}$ we used only the sign of $\Delta_r G$ to determine the reaction direction. The law of mass action tells us the relative concentrations in equilibrium

$$K = \frac{x_{\mathrm{CaO}}\, x_{\mathrm{CO_2}}}{x_{\mathrm{CaCO_3}}} = \exp\left[-\frac{\Delta_r G}{k_B T}\right] \tag{85}$$

For $T = 298K$, $K = \exp\left(-\frac{130}{0.0083 \times 293}\right) = 10^{-25}$ and $T = 1500K$, $K = \exp\left(\frac{65}{8.3 \times 1.5}\right) = 185$, so we see that the equilibrium concentrations are hugely different at these two temperatures.

Suppose the system is in equilibrium at some temperature. Then if we add more CaO to the system, x_{CaO} will go up. To keep K the same, the system will adjust to remove CaO and increase $\mathrm{CaCO_3}$. This is an example of a phenomenological observation called

- **Le Chatelier's principle**: a system will work to restore equilibrium by counteracting any change.

6 Osmotic pressure

As a final example, let us return to the topic of osmotic pressure, introduced in the discussion of entropy of mixing. Osmotic pressure is the pressure resulting from a concentration imbalance on the two sides of a semi-permeable barrier. For example, if you put a raisin in water, the higher sugar content of the raisin as compared to the water forces the water to be drawn into the raisin. The result is that the raisin swells up, almost back to the size of a grape. The drying out of a grape is also osmosis: water flows out of the grape into the air over time, and the grape desiccates. Grocers spray water on their fruits and vegetables to increase the local concentration of water so their produce looks more appealing. How about your fingers getting wrinkled when you stay in the bath too long. Is this osmosis?

Say we have water in a U-shaped tube with a semi-permeable membrane in the middle. The membrane allows water to pass but not sugar. Now put some sugar on one side. As water flows into the sugar side, it will increase the pressure on that side, lowering the pressure on the other side. Thus the sugar water will move up. This is a physical effect due to the entropy of mixing.

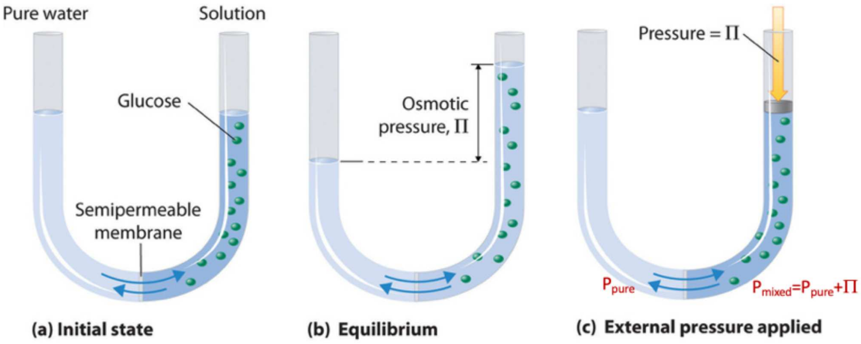


Figure 4. Osmotic pressure arises when concentrations are different on two sides of a semi-permeable barrier. Adding glucose to one side causes a pressure imbalance. This can be compensated by applying an external pressure Π called the osmotic pressure.

One way to quantify the effect is by the pressure you need to apply on the sugary side to restore the balance. This pressure is called the **osmotic pressure** and denoted by the symbol Π.

To compute the osmotic pressure, let us start with some definitions. A **solution** is a mixture of **solvent** and **solute**, with the solvent being the major component and the solute being a small addition. For example, in sugar water, water is the solvent and sugar the solute.

We want to compute the chemical potential of the solvent (water) on both sides, which we can derive from the Gibbs free energy accounting for the entropy of mixing. Let us say that initially there is the same number N_w of solvent (water) molecules on both sides of the barrier and we then add N_s solute (sugar) molecules to one side.

Recall that entropy of mixing is the extra entropy that a mixed substance has compared to a pure substance with the same properties (N, T, P). A great thing about the entropy of mixing is that it only depends on whether the things mixing are indistinguishable, not any other other properties of those things (internal degrees of freedom, etc.). For N particles in a volume V, the entropy is $S=k_B\ln\left(\frac{1}{N!}V^N\right)\approx -k_B N\ln\frac{N}{V}$. For our N_w molecules of solvent (water) and N_s molecules of solvent, the entropy of mixing is

$$\Delta S_{\text{mix}} = -k_B\left[\underbrace{\left(N_w\ln\frac{N_w}{V}+N_s\ln\frac{N_s}{V}\right)}_{\text{mixed system}} - \underbrace{(N_w+N_s)\ln\left(\frac{N_w+N_s}{V}\right)}_{\text{pure system with } N_w+N_s \text{particles}}\right] \tag{86}$$

$$= -k_B\left[N_w\ln\frac{N_w}{N_w+N_s}+N_s\ln\frac{N_s}{N_w+N_s}\right]>0 \tag{87}$$

Note that V dropped out. Indeed, the entropy of mixing only depends on the $\frac{1}{N!}$ factor.

In the limit that $N_s \ll N_w$ we can expand in $\frac{N_s}{N_w}$ to find

$$\Delta S_{\text{mix}} = k_B N_s\left(1-\ln\frac{N_s}{N_w}\right)+\mathcal{O}\left(\frac{N_s^2}{N_w}\right) \tag{88}$$

Let's write $G_w(T,P)$ for the Gibbs free energy of pure solvent (water). Since $N_s \ll N_w$, the enthalpy of the solute gives a negligible contribution to the total Gibbs free energy, and the only contribution that matters is from the entropy of mixing. So the total Gibbs free energy on the mixed side, generalizing Eq. (62), is

$$G_{\text{mixed}} = G_w - TS_{\text{mix}} = G_w - k_B T N_s\left(1-\ln\frac{N_s}{N_w}\right) \tag{89}$$

Thus the chemical potential of the solvent on the mixed side is then

$$\mu_w^{\text{mixed}}(T,P) = \left(\frac{\partial G_{\text{mixed}}}{\partial N_w}\right)_{T,P,N_s} = \mu_w(T,P) - k_B T\frac{N_s}{N_w} \tag{90}$$

where $\mu_w(T,P)=\frac{G_w}{N_w}$ is the chemical potential of the pure solvent.

Equilibrium requires the chemical potential of the solvent to be the same on both sides of the barrier. So

$$\mu_w(T,P_{\text{pure}})=\mu_w^{\text{mixed}}(T,P_{\text{mixed}})=\mu_w(T,P_{\text{mixed}})-k_BT\frac{N_s}{N_w} \tag{91}$$

The osmotic pressure we are trying to compute is the pressure difference $\Pi=P_{\text{mixed}}-P_{\text{pure}}$. In the limit $N_s\ll N_w$ we must have $\Pi\ll P_{\text{pure}}$ so that we can expand

$$\mu_w(T,P_{\text{mixed}})=\mu_w(T,P_{\text{pure}}+\Pi)=\mu_w(T,P_{\text{pure}})+\Pi\left(\frac{\partial\mu_w}{\partial P}\right)_{T,N_w} \tag{92}$$

Since $G_w=\mu_wN_w$,

$$\left(\frac{\partial\mu_w}{\partial P}\right)_{T,N_w}=\frac{1}{N_w}\left(\frac{\partial G_w}{\partial P}\right)_{T,N_w}=\frac{V}{N_w} \tag{93}$$

Combining the last 3 equations gives

$$\mu_w(T,P_{\text{pure}})=\mu_w(T,P_{\text{pure}})+\Pi\frac{V}{N_w}-k_BT\frac{N_s}{N_w} \tag{94}$$

So that

$$\boxed{\Pi=k_BT\frac{N_s}{V}} \tag{95}$$

This is known as **van 't Hoff's formula**.

You might notice that this formula for the pressure increase is the same as the ideal gas law $P=k_BT\frac{N}{V}$. Indeed, if you used gases rather than liquids and solids, and added some gas to the right hand side at constant volume, of course the pressure would go up. The increase in pressure, at fixed T and V would be exactly the partial pressure of the new gas added, $P^{\text{new}}=k_BT\frac{N_{\text{new}}}{V}$. Van 't Hoff's formula is easy to understand for gases. The amazing thing however is that it applies for liquids and solids too. To see why, recall that an ideal gas is one where we can neglect intermolecular interactions. This leads to features like the energy is extensive, scaling linearly with the number density $E\sim n$. If pairwise intermolecular interactions dominated then we would expect $E\sim n^2$. In water or other liquids, interactions are important, so there are non-extensive contributions to the energy, like surface tension. Now, interactions between a solute (sugar) and solvent (water) can certainly be strong. The key point however is that as long as the solute is dilute, the solute/solvent interactions will be rare: any N_s^2 scaling in thermodynamic properties will be subdominant to the ideal gas N_s scaling: if we write $E\sim c_0+c_1N_s+c_2N_s^2$, we can always make N_s small enough so that $E\sim c_0+c_1N_s$ is a good approximation. So for dilute solutes, the N_s dependence is always linear, and approaches ideal, extensive, scaling behavior. This is the ultimate origin of the universality of the van 't Hoff formula even for mixtures of liquids, and solids, and non-ideal gases. The key assumption is that $N_s\ll N_w$: the solute is dilute.

You can repeat this exercise when the solute contains a number of different substances (sugar, salt, etc). The result is that the osmotic pressure is

$$\Pi=k_BT\sum_s\frac{N_s}{V} \tag{96}$$

This indicates that osmotic pressure is a **colligative property**, meaning it doesn't matter what the solute is, just the total concentration of solutes.

For example, when your blood pressure is 120/80 it means that the pressure measured is 120 mmHg=0.16 atm coming out of your heart and 80 mmHg=0.1 atm going in. This doesn't mean that your blood is at much lower than atmospheric pressure, since it is measured by a device in the atmosphere – these numbers are *relative* to atmospheric pressure; thus the actual pressure in your veins is 1.1 to 1.16 atm. The osmotic pressure in blood is 7.65 atm at $37°C=310K$. This doesn't mean that blood is under 7.65 atm of pressure; it means that if a blood vessel were immersed in pure water, 7.65 atm would have to be applied to prevent the influx of water. Blood vessels are not surrounded by water, but by solution, also at an osmotic pressure around 7.65, so no solution flows.

We can use these numbers to find the net concentration n_s (in mol/L) of all dissolved solvents:

$$n_s = \frac{N_s}{N_A \cdot V} = \frac{\Pi}{RT} = \frac{7.65\,\text{atm}}{0.08314\frac{L\,\text{atm}}{\text{mol}\,K}310\,K} = 296\,\frac{\text{mmol}}{L} \tag{97}$$

where mmol is a millimole, and we have divided by Avogadro's number N_A to get an expression in terms of mols. This quantity n_s is called the **osmolarity** and often written in millimoles per kg of solution (water): $n_s = 296\,\frac{\text{mmol}}{L} = 296\frac{\text{mmol}}{\text{kg}}$. For example, we can compare to the osmolarity of $50g$ of sugar (glucose, molar mass 180 $\frac{\text{g}}{\text{mol}}$) mixed into 1 liter of water:

$$n_s = \frac{50g}{180\,\frac{g}{\text{mol}}}\frac{1}{L} = 277\frac{\text{mmol}}{L} \tag{98}$$

So mixing a little more than $50g$ of sugar (about 2 tablespoons) with a liter of water and drinking it will not increase your blood pressure. More sugar (or salt) than this concentration will cause your blood pressure to go up.

Osmotic pressure really is an *entropic effect*. It is a powerful example of the importance of entropy. You might have been concerned from time to time that the way we were defining entropy had some arbitrariness to it, due to the choice to coarse-grain or add indistinguishable particle factors. Osmotic pressure indicates that entropy is real and unambiguous. It has physical observable consequences and makes quantitative predictions that can be tested experimentally. Add salt and the pressure goes up and work can be done.

Finally, it is worth pointing out that equilibrium properties, such as from equating the chemical potentials on two sides of a semipermeable barrier, do not tell us anything about the microscopic mechanism by which equilibrium is achieved. Diffusion (Lecture 2) involving the random walks of molecules, is one way. Another is convection, when large macroscopic currents, such as temperature gradients, push the molecules around together. There is also advection, whereby the motion of one type of molecule pulls another along with it. Capillary action is another transport mechanism relevant for liquids where surface tension draws water into a straw or a paper towel. Imbibation is a transport phenomenon in solids or colloids whereby the a material expands as liquid is absorbed (such as a sponge or seed). What is the microscopic mechanism for osmosis? Probably some combination of the above. The point, however, is that the microscopic mechanism is irrelevant. Equilibrium thermodynamics lets us compute the main result, the osmotic pressure, independently of the microscopic mechanism of osmosis.

Test your understanding: In going from step (a) to step (b) in Fig. 4, an amount of water is mixed with the saltwater and entropy increases. When we apply a pressure Π to restore the original levels in the U-tube as in step (c), we are unmixing the amount of water, undoing the entropy of mixing. What compensates for this decrease in entropy so that the second law still holds? Is anything different about the water in the left tube in panels (a) and (c) of Fig. 4?

7 Grand free energy

There is one more free energy that is used sometimes, called the grand free energy Φ. We're not going to use it until Lecture 11, because it's harder to interpret physically, but I include the relevant formulas here since they are closely related to the other free energies from this lecture.

Recall that the grand canonical partition function is

$$\mathcal{Z}(V, T, \mu) = \sum_k e^{-\beta\varepsilon_k + \beta\mu N_k} \tag{99}$$

where the sum is over microstates k with N_k particles and energy ε_k. We also showed that

$$-k_B T \ln\mathcal{Z} = \langle E\rangle - TS - \mu\langle N\rangle \tag{100}$$

We then define the **grand free energy** Φ like the free energy F, but computed using the grand canonical ensemble.

$$\Phi = -k_B T \ln\mathcal{Z} \tag{101}$$

Thus,

$$\Phi(T,V,\mu) = \langle E\rangle - TS - \mu\langle N\rangle \tag{102}$$

Recall that $F(V,T,N)$ and now we have $\Phi(V,T,\mu)$, so the grand free energy has traded N for μ in the free energy. This is similar to how we used $F(V,T,N)=E-TS$ to trades entropy for temperature using $E(S,V,N)$. Indeed, we can also write

$$\Phi(T,V,\mu) = F - \mu N \tag{103}$$

Using Eq. (1) again, we find

$$\boxed{d\Phi = -SdT - PdV - Nd\mu} \tag{104}$$

An important property of Φ is that it is an extensive (like the other energies, internal energy, Helmholtz free energy, Gibbs free energy) function of only a single extensive variable V. Thus it must be proportional to V. Since $\frac{\partial\Phi}{\partial V}\Big|_{\mu,T} = -P$ we then have

$$\Phi = -PV \tag{105}$$

This is a useful relation, similar to $G=\mu N$ for the Gibbs free energy. Like $G=\mu N$, $\Phi=-PV$ follows from the Euler relation in Eq. (17). We'll use the grand free energy in quantum statistical mechanics. It is not used in chemistry or for pure thermodynamics computations, since we can just use PV. I only include it here since it is a free energy and naturally part of the "free energy" lecture. We'll only need the definition Eq. (101) and the relation (104) in future applications.

8 Summary

Free energy is a very important, very general concept in physics. The basic kind of free energy is

- Helmholtz free energy: $F = E - TS$

It represents what we intuitively think of as energy in a thermal system. For example, we intuitively think of energy being minimized, as a ball rolls to the bottom of a hill and stops. But energy, E, is conserved. What is really being minimized is F. The minimization happens actually because entropy S is being maximized. Thus, F is a very general, very useful quantity that can be minimized to find the equilibrium state of the system.

Many physical, biological and chemical systems are at constant pressure. In such circumstances, there is an extra contribution to the free energy and it is more convenient to use

- Enthalpy $H = E + PV$ instead of E
- Gibbs free energy $G = H - TS$ instead of F

The extra PV factor in the definition of enthalpy seamlessly includes the effect of work done when the volume changes at constant pressure. In chemistry and biology, enthalpy and Gibbs free energy rule. When a reaction occurs, we care more about the enthalpy change than the energy change. The biggest difference between enthalpy and energy is for reactions which change in the number of moles of molecules in the gaseous phase, Δn. Then $\Delta H - \Delta E = \Delta \mathrm{PV} = (\Delta n)RT$. This difference comes from the extra work that has to be done when the gas expands against the constant pressure. Enthalpy changes have been measured for many reactions. If they are not known, they can be approximated using differences in molecular enthalpies or approximate bond enthalpies.

We used free energy to derive a number of important results from chemistry. If a reaction can lower the Gibbs free energy, it will occur. If it only increases G, it will not occur. The equilibrium state minimizes G. One important consequence is the law of mass action which determines the equilibrium concentrations of chemicals based on their relative Gibbs free energies. Another is Le Chatelier's principle: a system will work to restore equilibrium by counteracting any change.

Finally, we discussed osmotic pressure. Osmotic pressure is an entropic effect. If you put a grape in a glass of water, water will be drawn into the grape, increasing its pressure. The pressure increase Π is given by van 't Hoff's formula $\Pi = k_B T\frac{N_s}{V}$. This is, not coincidentally, the form for the partial pressure of an ideal gas. The solute acts like an ideal gas because it's dilute in the solvent: any contributions scaling like N_s^2 are subdominant and can be neglected.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 9: Phase Transitions

1 Introduction

Some phases of matter are familiar from everyday experience: solids, liquids and gases. Solid H_2O (ice) melting into liquid H_2O (water) is an example of a phase transition. You may have heard somewhere that there are 4 phases of matter: solid, liquid, gas and plasma. A plasma is an ionized gas, like the sun. I don't know why plasmas get special treatment though – perhaps it's the old idea of the "four elements." In fact, there are thousands of phases. For example, ferromagnetic is a phase, like a permanent iron magnet. When you heat such a magnet to high enough temperature, it undergoes a phase transition and stops being magnetic. Conductors, insulators, and semi-conductors are also phases of matter. At very very high temperatures, nuclei break apart into a quark-gluon phase. Solids generally have lots of phases, determined by crystal structure or topological properties. For example, diamond and graphite are two phases of carbon with different lattice structure:

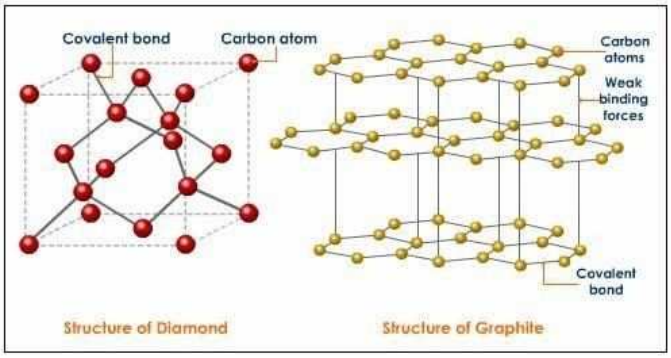


Figure 1. Two of the phases of solid carbon

It's actually quite hard to come up with a precise definition of a phase. Some textbooks say

- A **phase** is a uniform state of matter.

This is an intuitive definition, but not very precise. Taken literally, it is too general: a gas at a different temperature is a different uniform state of matter. So is it a different phase? We don't want it to be. We want "gas" to be the phase.

A more technically precise definition is

- A **phase** is a state of matter whose properties vary smoothly (i.e. it is an analytic function of P, V, T etc).

You might first think that this definition makes liquid H_2O (water) and gasesous H_2O (steam) in the same phase, since we can boil water and it slowly becomes steam. Although this does sound smooth, it is not. For example, consider the temperature of water as heat is added. As you heat it the temperature rises. But when it hits the boiling point, the temperature does not rise anymore, instead the heat goes into vaporizing the water. Then onces it's all gas, its temperature changes again. So the density of H_2O change discontinuously and non-analytically as a function of temperature $\rho(T = 99.9°C)$ is very different from $\rho(T = 100.1°C)$.

Connecting phase to smoothness properties allows to shift focus from phases themselves to the transformations between phases called **phase transitions**. Phase transitions are an incredibly important area of physics.

Physicists take two different approaches to phase transitions. On the one hand, we can treat each phase as its own statistical mechanical system. For example, the ensemble we use to describe ice is very different from the one we use to describe water vapor – neglecting interactions is an excellent approximation for many gases, but a horrible approximation for solids. This makes the discontinuities in density, entropy, etc, across a phase transition inherent in our description. It lets us derive very useful formulas for phase transformations, such as how the pressure and temperature of the phase transformation are related.

The second approach is to construct a statistical mechanical system that describes a substance on both sides of a phase transition. For example, if we knew the exact partition function for N molecules of H_2O, we should be able to see the gas, liquid and solid states arise from different limits. This partition function would necessarily have non-smooth properties across the values of temperature and pressure associated with the phase transition. It is these kinks in otherwise smooth functions that make phase transitions so interesting. Where do they come from? Water itself is too complicated to write down an exact partition function, but there are plenty of simpler systems that we can solve to understand phase transitions. Pursuing these simpler systems leads to concepts you may have heard of like critical phenomena, the renormalization group, mean-field theories, etc. Most of these topics, unfortunately, we will not have time to pursue – phase transitions are the focus of much of modern condensed matter physics and could easily occupy a year-long graduate course.

In this lecture, we will start with discussing the familiar phases of solid, liquid and gas, and understand transitions between them using statistical thermodynamics. Then we will discuss some of the broader, more general aspects of phase transitions.

2 Solids, liquids and gases

It is not hard to figure out what we mean by solid, liquid, or gas. Both solids and liquids are essentially incompressible. The **compressibility** is defined

$$\beta_T = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{1}$$

So solids and liquids have $\beta_T \approx 0$. This means no matter how much pressure we put on, we cannot make solids or liquids much denser. Gases have much larger values of β_T. Indeed, from the ideal gas law $PV = Nk_BT$, we see that for a gas $\beta_T = \frac{1}{P} \neq 0$. Compressibility is an example of an **order parameter**, something whose value characterizes the phase.

Solids and liquids of the same substance often have approximately the same density, while the density of gases is much lower. Liquids and solids are called **condensed matter**. Solids differ from liquids and gases in that they are rigid. More precisely, they do not deform under a sheer stress, i.e. they have zero shear modulus S_s. Thus shear stress is an order parmeter for the liquid-solid phase transition. Liquids and gases are collectively called **fluids**.

It may be helpful to say a few more words about liquids. Liquids generally have around the same density as solids, so the atoms are all in contact with each other. Instead of having strong covalent or metallic bonds, like in a solid (cf. Lecture 14), liquids have weaker ionic or hydrogen bonds that keep the molecules close. Although the attractive force in liquids is weak, typical thermal velocities in a liquid are not enough to overcome it.

For example, in water, H_2O, the H-O bonds are covalent, with the shared electrons localized closer to the O than the H. This makes the H slightly positively charged and the O negatively charged. When a two water molecules approach each other, the H from one is then weakly attracted to the O from the other, forming a hydrogen bond. The O in each water molecule can form 2 covalent bonds and 2 hydrogen bonds, giving liquid water a tetrahedral formation. Molecules on the surface of water must have fewer than 2 hydrogen bonds per oxygen, on average. Thus there is any energy cost to having a surface. The Gibbs free energy per area of surface $\gamma = \frac{G}{A}$ is called the **surface tension**:

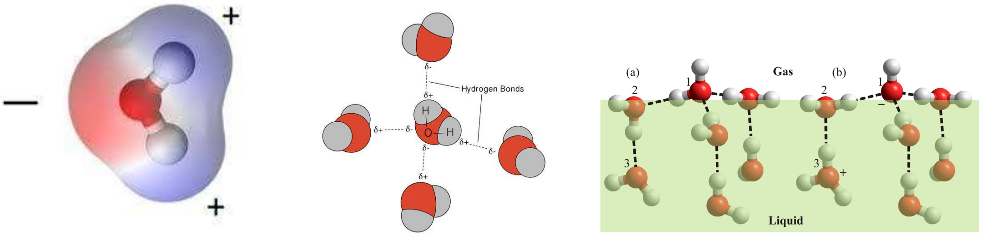


Figure 2. The H and O atoms in water have small charges (left) forming attractive hydrogen bonds (middle). Surfaces cannot saturate all the bonds, therefore there is any energy cost $G = \gamma A$ to having a surface, with γ the surface tension.

All liquids have a surface tension. The surface tension of water happens to be particularly large (e.g. $\gamma_{\text{water}} = 73\frac{\text{mN}}{m}$ compared to say, $\gamma_{\text{CO}_2} = 17\frac{\text{mN}}{m}$), but not the largest (mercury has $\gamma_{\text{Hg}} = 486\frac{\text{mN}}{m}$). One consequence of the surface tension is that a liquid will not expand to fill all the available volume. Liquids form droplets. Even in zero-gravity, water still forms droplets, as does mercury:

Figure 3. (Left) water in zero gravity. (Right) a ball of liquid mercury in zero gravity being hit by a piece of metal.

Whether a solid turns into a liquid when it is heated depends not only on the types of interactions among the molecules, but also on the pressure. At low pressures, when solids are heated, the atoms that break free of their covalent bonds fly off into gas, and the solid sublimates. Only if the pressure is sufficiently high will the molecules stick around close enough to each other for the weak attractive interactions to dominate and the liquid phase to form.

Two phases can be in equilibrium with each other. Consider an unopened bottle of water (constant volume and temperature). The liquid water in the bottle and water vapor in the bottle are in equilibrium. Water molecules are constantly evaporating from the water into the air, and water molecules are condensing into the water at exactly the same rate. Thus volume and temperature are not good ways of characterizing phases. What happens if we open the bottle? Eventually, all the water will evaporate. An open bottle is at constant temperature and pressure. Under these conditions generally a single phase dominates. Thus, phases of matter can be characterized by temperature and pressure.

Why are there single phases at constant T and P? For a single pure substance, the chemical potential is a function of P and T alone: $\mu = \mu(P, T)$. For example, $\mu = k_B T \ln\left(\frac{P}{k_B T \lambda^3}\right)$ for a montatomic ideal gas, where λ is the thermal wavelength, $\lambda = \frac{h}{\sqrt{2\pi m k_B}}$; for a more general ideal gas $\mu = k_B T \ln\left(\frac{P}{k_B T}\zeta\right)$ with ζ the single-particle partition function. There is no N dependence in the chemical potential at fixed P and T since the Gibbs free energy $G(N, P, T) = N\mu$ is extensive.

When there are two phases present the Gibbs free energy is $G=N_1\mu_1(T,P)+N_2\mu_2(T,P)$. Note however one important difference between the two-phase case and the two-chemicals case: when two phases are present there is no entropy of mixing. Solids and liquids coagulate because of the surface tension, so even if multiple phases are present they are always separated and the mixing entropy is tiny if not completely absent (maybe $\Delta S\sim k_B\log(a\text{ few})$ for a few chunks of ice in ice water, but this is negligible compared to the entropies of the ice and water separately $S\sim Nk_B$). Another way to see that the mixing entropy is absent is that the pressures are the *same* in two phases; they don't become partial pressures that add up to the total pressure (partial pressure will be relevant of water vapor is mixed with something else, like air, but not for pure water/ice equilibrium that we are discussing here).

At constant temperature and pressure, equilibrium is determined by minimizing the Gibbs free energy G. Recall that

$$dG=VdP-SdT+\mu_1dN_1+\mu_2dN_2 \tag{2}$$

So if ΔN particles go from phase 1 to phase 2 at constant T and P then

$$\Delta G=\mu_1(-\Delta N)+\mu_2(\Delta N)=(\mu_2-\mu_1)\Delta N \tag{3}$$

Thus, to minimize G, particles move from higher chemical potential to lower chemical potential. This will keep happening until there are no more particles to change phase. Thus at a fixed value of T and P, as long as $\mu_1\neq\mu_2$, only one phase is allowed. In the law of mass action, the chemical potentials had N dependence which lead to equilibrium situations with different amounts of substances. That N dependence all came from the entropy of mixing. Since here entropy of mixing is absent, one phase completely annihilates the other.

Note that if we express the chemical potential in terms of V rather than P, for a single ideal gas, it takes the form $\mu=k_BT\ln\left(\frac{N}{V}\zeta\right)$. This expression *does* depend on N. So as N changes the chemical potential changes too. Thus, at constant volume, equilibrium can be achieved with finite amounts of two phases. We'll come back to this situation when discussing vapor pressure below.

Returning to constant pressure, it is only when $\mu_1(T,P)=\mu_2(T,P)$ that there is no change in the Gibbs free energy with ΔN, and so only then can the two phases can be present at once. Since $\mu_1(T,P)$ and $\mu_2(T,P)$ are functions, setting them equal generates a curve in the T/P plane. This curve is the **phase boundary**. On the phase boundary two phases are in equilibrium.

A diagram of the phases as a function of pressure and temperature is called a **phase diagram**. Here are some example phase diagrams for carbon dioxide, argon and water.

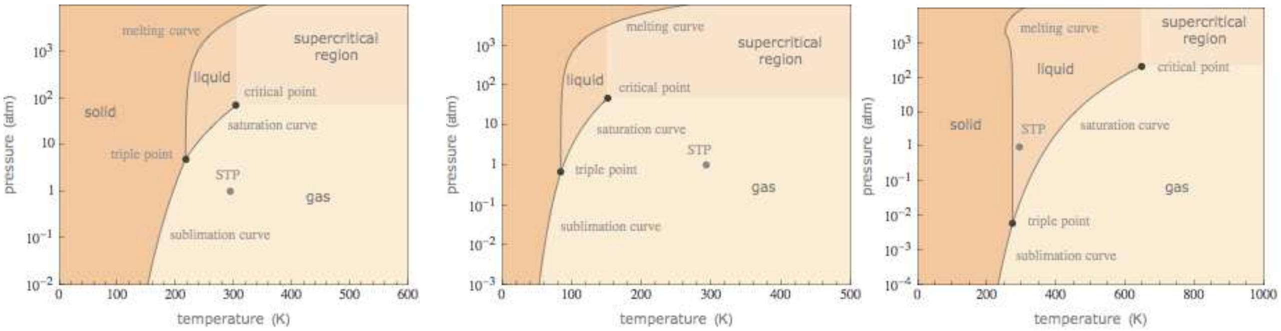


Figure 4. Phase diagrams for CO_2, Ar, and H_2O. The point STP in these plots refers to $T=20°C=293K$ and $P=1$ atm.

The thick lines in the phase diagram are the **phase boundaries**, determined by $\mu_1=\mu_2$. A **phase transition** is the transformation as a phase boundary is crossed. We define

- **Melting**: transition from solid to liquid.

- **Freezing/Fusion**: transition from liquid to solid.
- **Boiling/vaporization**: transition from liquid to gas.
- **Condensation**: transition from gas to liquid.
- **Sublimation**: transition from solid to gas.
- **Deposition**: transition from gas to solid.

Note in the CO_2 phase diagram that as temperature is increased at $P=1\,\mathrm{atm}$, CO_2 goes from solid directly to gas: it sublimates. This is why smoke comes off dry ice, but there is no liquid. Liquid CO_2 requires at least 5 atmospheres of pressure.

In a pure phase (off the phase boundary) there is one type of substance and so $G=\mu N$ with N fixed. Then, $\mu=\frac{G}{N}$ and so

$$\left(\frac{\partial\mu}{\partial T}\right)_P=\frac{1}{N}\left(\frac{\partial G}{\partial T}\right)_P=-\frac{S}{N} \tag{4}$$

Since $S>0$ this implies that the chemical potential always decreases as the temperature goes up. Moreover if there are two phases with different entropies, then as the temperature is raised, the chemical potential of the one with the larger molar entropy (entropy per mole of particles) will go down more. Thus the higher entropy state is preferred at larger temperature. This explains why solids melt when you heat them and liquids boil: the phase transitions are driven by entropy and $S_{\text{solid}}<S_{\text{liquid}}<S_{\text{gas}}$. To be fair, we haven't shown that $S_{\text{solid}}<S_{\text{liquid}}<S_{\text{gas}}$, instead, we can deduce it from phase diagram since when heated solids melt and liquids vaporize.

The first derivative of G with respect to temperature is discontinuous across a phase boundary: infinitesimally below it $\left(\frac{\partial G}{\partial T}\right)_P=-S_{\text{liquid}}$ and infinitesimally above it $\left(\frac{\partial G}{\partial T}\right)_P=-S_{\text{gas}}$. In a pure phase, G is a smooth function (it and all its derivatives are continuous). Therefore, G changes non-smoothly, the phase changes. We'll come back to this in Section 4.

Let's observe some features of the CO_2 phase diagram. First note that liquid/solid phase boundary and the liquid/gas phase boundary intersect. The point where they intersect is a special value of P,T called the **triple point**. At the triple point, three phases are in equilibrium together: $\mu_{\text{solid}}=\mu_{\text{liquid}}=\mu_{\text{gas}}$.

Note also that the phase boundaries do not extend up forever. They end at a point in P,T space called the **critical point**. Critical points are super interesting experimentally and theoretically. An implication of the phase boundary ending is that one can go around the phase transition line. That is, one can smoothly transform a liquid into a gas, without crossing a phase boundary. This is one reason why it is hard to give precise definitions of phases. For example, we said that solids and liquids have small compressibilities β_T. But we didn't say how small. As you move around the critical point from the liquid side, the compressibility gets larger. At some point we don't consider it small and the phase is somewhere between a liquid and a gas – like a very dense gas. So let's not even try to give precise general definitions to different phases. Instead, we'll study transitions between phases. These transitions *are* precisely defined since the phase boundary is precise.

3 Phase boundaries

Suppose we are close to a phase boundary, but on one side of it. Then a single phase completely dominates, and N is fixed ($dN=0$). Since $\frac{G}{N}=\mu$, then $dG=Nd\mu$ and from Eq. (2), $dG=VdP-SdT$, we find

$$d\mu=\frac{V}{N}dP-\frac{S}{N}dT \tag{5}$$

Note the consistency with Eq. (4) at constant P. This holds in any pure phase region, even arbitrarily close to the phase boundary. In particular, it holds on both sides of the phase boundary, as we approach the phase boundary. But on the phase boundary, $\mu_1 = \mu_2$, so the way pressure and temperature must change as we move along the phase boundary is determined by setting $d\mu_1 = d\mu_2$ which gives

$$\frac{V_1}{N_1} dP - \frac{S_1}{N_1} dT = \frac{V_2}{N_2} dP - \frac{S_2}{N_2} dT \tag{6}$$

That is

$$\frac{dP}{dT} = \frac{\Delta\left(\frac{S}{N}\right)}{\Delta\left(\frac{V}{N}\right)} \tag{7}$$

This equation, called the **Clapeyron equation**, determines the shape of the phase boundary.

3.1 Latent heat

The Clapeyron equation involves the change in the molar entropy, $\frac{S}{N}$. Recall that $G = H - TS$ so $\mu = \frac{G}{N} = \frac{H}{N} - T\frac{S}{N}$. Since $\mu_1 = \mu_2$ in equilibrium then $T\Delta\left(\frac{S}{N}\right) = \Delta\left(\frac{H}{N}\right)$ at the phase boundary. So we can can also write the Clapeyron equation as

$$\frac{dP}{dT} = \frac{1}{T} \frac{L}{\Delta\left(\frac{1}{n}\right)} \tag{8}$$

where $n = \frac{N}{V}$ is the number density and

$$L = \Delta\left(\frac{H}{N}\right) \tag{9}$$

is called the **latent heat**.

Latent heat is the change in enthalpy per molecule, like a reaction enthalpy, $\Delta_r H$ but for a phase transition at **saturation** (on the phase transition boundary). It is a heat because as you heat something at saturation, the heat put in changes the entropy by $\Delta S = \frac{Q}{T}$. Since $\Delta S = \frac{\Delta H}{T}$ we have simply that $Q = \Delta H$: the heat put in is the change in enthalpy. The latent heat is the heat put in per molecule to change the phase.

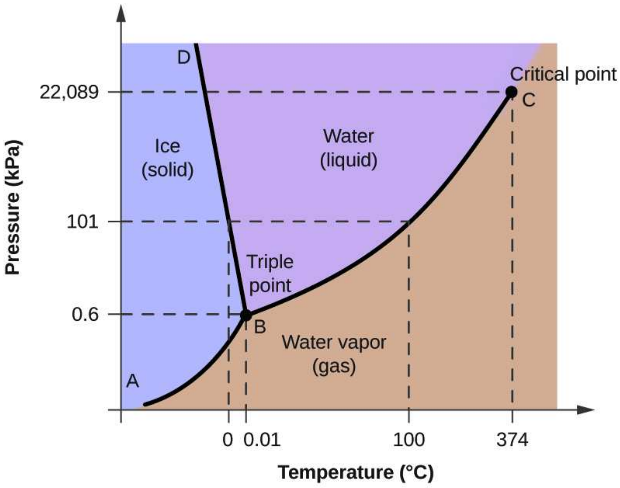
For example, when you boil water, you put more and more heat in, and more water evaporates, but the temperature doesn't change. The heat you put in is providing the energy it takes to break up the hydrogen bonds in the water. The enthalpy of formation of liquid water is $\Delta_f^\circ H = -286\frac{\text{kJ}}{\text{mol}}$ and for water vapor is $\Delta_f^\circ H = -242\frac{\text{kJ}}{\text{mol}}$. The difference between these is the latent heat of vaporization of water at 1 atm: $L_{\text{vap}} = 44\frac{\text{kJ}}{\text{mol}}$. This is positive since it takes heat to boil water. Note that $44\frac{\text{kJ}}{\text{mol}}$ is a large number: water has a lot of energy stored in its hydrogen bonds and it is hard to separate them. For comparison, consider the heat capacity of liquid water $C_P = 75\frac{J}{\text{mol}\,K}$, which implies that to heat water from $0°C$ up to $100°C$ takes only $\Delta H = 7.5\frac{\text{kJ}}{\text{mol}}$ of energy. That is, it takes 6 times as much energy to vaporize water as it does to heat it up to its boiling point from its freezing point: water does not want to evaporate.

That the latent heat of vaporization of water is so large is the reason that sweating is such an efficient form of cooling. Say it's $105°F$ outside, which is higher than body temperature. If you didn't sweat, your body would just heat up until it reached equilibrium with the air. Instead, liquid water forms on your skin, and it draws heat from your body, evaporating into air and cooling your body at the same time. Air conditioners exploit the latent heat of vaporization as well, as we will explore on a problem set. Note that latent heat, which exploits the different chemical potential for different phases, allows temperature differences to *increase* spontaneously. This is not in conflict with the second law of thermodynamics because the total entropy is increasing: the evaporated water has a much larger entropy increase than the entropy decrease from cooling your body.

It's also worth pointing out, for completeness, that the latent heat of fusion for melting ice is $L_{\text{fuse}} = 6.0\frac{\text{kJ}}{\text{mol}}$, which is not particularly large. Indeed, it is a much smaller energy than L_{vap}. Note that $L_{\text{fuse}} > 0$ since it takes heat to melt ice.

Using the Clapeyron equation, we can deduce some simple features of phase boundaries. Consider first the solid to liquid transition (i.e. melting). This involves breaking covalent bonds, so $L_{\text{fuse}} > 0$.[1] In general, the density change in going from solid to liquid is usually very small and slightly negative: most solids are slightly more dense than their liquid forms. So $\Delta\left(\frac{1}{n}\right) \gtrsim 0$. Therefore by Eq. (8), $\frac{dP}{dT}$ is generally very large and positive. That is, the liquid/solid phase boundary is usually quite steep and goes slightly to the right in the T/P plane. This can be clearly seen in the phase diagrams in Fig. 4 above.

A well known exception to the density decreasing on melting is water. Water expands when it freezes due to the unusual importance of hydrogen bonding in the liquid. We can see this in the phase diagram in Fig. 4, or more clearly if we zoom in with a logarithmic T axis:



2

Figure 5. Phase diagram for water.

Note that the solid liquid boundary goes up and to the left, so $\frac{dP}{dT} < 0$. Since $L > 0$ and $T > 0$ this must mean $\Delta\left(\frac{1}{n}\right) < 0$, as is indeed the case for water. It is the tetrahedral structure of solid ice that makes it not particularly dense. Some other materials with tetrahedral structure (like silicon or gallium) also have denser liquids than solids.

For the transition between liquid and gas (vaporization), the enthalpy change is again positive $\Delta H > 0$, so $L > 0$. In water this is because hydrogen bonds are broken. In general it's because there are attractive interactions among molecules in liquids sticking them together, and it takes energy to separate molecules that are attracted to each other. In addition $\Delta(\text{PV}) = RT$ upon vaporization which also contributes to the latent heat. Gases are usually much less dense than liquids, so $\Delta\left(\frac{1}{n}\right)$ is positive and generally much larger than for the solid-liquid transition. Since the gas density is much less than the liquid density $n_{\text{gas}} \ll n_{\text{liquid}}$, we can write

$$\Delta\frac{1}{n} = \frac{1}{n_{\text{gas}}} - \frac{1}{n_{\text{liquid}}} \approx \frac{1}{n_{\text{gas}}} = \frac{V_{\text{gas}}}{N_{\text{gas}}} = \frac{k_B T}{P} \tag{10}$$

where the ideal gas law $PV = Nk_B T$ was used in the last step. It is conventional to use molar units for latent heat (L is in $\frac{\text{kJ}}{\text{mol}}$), so we replace k_B with the ideal case constant and have

$$\frac{dP}{dT} = \frac{PL}{RT^2} \tag{11}$$

[1] The only exception is helium. In a small region of the ^{3}He phase diagram, liquid helium solidifies upon heating and $L_{\text{fus}} < 0$.

This is known as the **Clausius Clapeyron** equation. Here L is the **latent heat of vaporization**. The pressure at the liquid-gas transition is called the **vapor pressure**.

Since the latent heat is dominated by the enthalpy change of breaking bonds, we expect it to be a slowly varying function of temperature. If we assume L is independent of T, then we can integrate the Clausius-Clapeyron equation. Writing it as

$$\frac{1}{P}dP = \frac{L}{RT^2}dT \tag{12}$$

we can integrate both sides to give

$$P = C \times \exp\left[-\frac{L}{RT}\right] \tag{13}$$

with C an integration constant. Starting at any point $P = P^\circ$ and $T = T^\circ$ as a boundary condition we then have

$$P = P^\circ \exp\left[-\frac{L}{R}\left(\frac{1}{T} - \frac{1}{T^\circ}\right)\right] \tag{14}$$

For example, at sea level ($P^\circ = 1$ bar) water boils at $T^\circ = 373\ K$. At 1000 m, the atmospheric pressure is $P = 0.9$ bar. Using the latent heat of vaporization of water is $L = 42\frac{\text{kJ}}{\text{mol}}$ and $R = 8.3\frac{J}{\text{mol}}$, we get

$$T = \left(\frac{1}{T^\circ} - \frac{R}{L}\ln\frac{P}{P^\circ}\right)^{-1} = 370.1K \tag{15}$$

which is three degrees lower.

Keep in mind that you cannot extrapolate the Clausius-Clapeyron equation too far. Eventually, the temperature dependence of the latent heat becomes important. For small changes in T and P it is usually makes predictions in excellent agreement with observation.

3.2 Vapor pressure

The **vapor pressure** is the pressure of a pure substance at saturation (on the phase boundary). Generally, vapor pressure refers to the pressure of a gas (vapor), so we use vapor pressure to describe liquid-gas and solid-gas phase transitions. At saturation, two phases can exist in equilibrium. Conversely, if two phases are in equilibrium, the pressure of the gas must be the vapor pressure at that temperature. When there is a mixture of liquids or gases, the partial pressure of each must equal the appropriate vapor pressure in equilibrium.

For example, if we have a sealed bottle of water, there will be some water vapor in the bottle, above the water. If there is only water vapor (no air), then the vapor pressure at room temperature can be determined from the Clausius-Clapeyron equation. Using Eq. (14) at the boiling point $T^\circ = 373K$, $P^\circ = 1$ atm and $L = 42\frac{\text{kJ}}{\text{mol}}$, we find that at room temperature, $T = 298K$ that $P = 0.034\,\text{atm} = 3.142\text{kPa}$. This is consistent with Fig. 5. Note that our calculation implies that the vapor pressure of water at room temperature is much lower than atmospheric pressure. This may seem unintuitive, since it implies that water should not evaporate. Indeed, it would not, if there were no air. In fact, air is only around 1% water at sea level, so the partial pressure of water in air is 0.01 atm which is about 3 times smaller than the vapor pressure. So water does evaporate into air. If you seal a bottle of water at room temperature, the water will start to evaporate and the partial pressure of water in the bottle will increase. It will go up from 0.01 atm to 0.03 atm and then stop, since it matches the vapor pressure. At this point, the total pressure inside the bottle has gone up from 1 atm to 1.02 atm. This small pressure increase is responsible for the pfft you sometimes hear when opening a bottle of water, even if it's not carbonated.

What happens if we mix some solute into the water? For example, how does the vapor pressure of water change when salt is added, and how does its boiling point T_{boil} change? To an excellent approximation, the salt stays in the water, so that only the water is in equilibrium with its vapor. Let $\mu_w(P,T)$ be the chemical potential of pure liquid water, $\mu_{\text{gas}}(P,T)$ be the chemical potential of the pure water vapor, and $\mu_w^{\text{mixed}}(P,T)$ be the chemical potential of water in the saltwater mix. The boiling point $T_{\text{boil}}=T_0$ for pure water at a vapor pressure P_0 satisfies

$$\mu_w(P_0,T_0)=\mu_{\text{gas}}(P_0,T_0) \tag{16}$$

For saltwater, the equilibrium condition is $\mu_w^{\text{mixed}}(P,T)=\mu_{\text{gas}}(P,T)$.

Recall from the discussion of osmotic pressure that the saltwater has higher entropy than pure water, due to entropy of mixing, so it has lower Gibbs free energy, $G=G_0-TS_{\text{mix}}$ and therefore lower chemical potential. For small solute concentrations, we found

$$\mu_w^{\text{mix}}(P,T)=\mu_w(P,T)-k_BT\frac{N_s}{N_w} \tag{17}$$

where N_s is the number of salt molecules and N_w the number of water molecules.

We first ask how the vapor pressure changes at fixed temperature $T=T_0$. Expanding around P_0 by writing $P=P_0+\Delta P$ we get

$$\mu_w^{\text{mixed}}(P,T_0)=\mu_w(P_0+\Delta P,T_0)-k_BT_0\frac{N_s}{N_w}=\mu_w(P_0,T_0)+\Delta P\left(\frac{\partial\mu_w}{\partial P}\right)_T-k_BT_0\frac{N_s}{N_w} \tag{18}$$

Similarly,

$$\mu_{\text{gas}}(P,T_0)=\mu_{\text{gas}}(P_0,T_0)+\Delta P\left(\frac{\partial\mu_{\text{gas}}}{\partial P}\right)_T \tag{19}$$

Now, $\left(\frac{\partial\mu}{\partial P}\right)_T=\frac{V}{N}$ so setting $\mu_w^{\text{mixed}}(P,T_0)=\mu_{\text{gas}}(P,T_0)$ and using Eq. (16) we get

$$\Delta P\left(\frac{V_w}{N_w}-\frac{V_{\text{gas}}}{N_{\text{gas}}}\right)=k_BT_0\frac{N_s}{N_w} \tag{20}$$

The molar volume of the liquid is much lower than the gas (the gas density $n=\frac{N}{V}$ is much larger), so we can drop $\frac{V_w}{N_w}$ compared to $\frac{V_{\text{gas}}}{N_{\text{gas}}}$. Using the ideal gas law, $\frac{V_{\text{gas}}}{N_{\text{gas}}}=\frac{k_BT_0}{P_0}$ we then have

$$(P-P_0)\left(-\frac{k_BT_0}{P_0}\right)=k_BT_0\frac{N_s}{N_w} \tag{21}$$

or

$$\boxed{\Delta P=-\frac{N_s}{N_w}P_0} \tag{22}$$

This is known as **Raoult's law**. It says the vapor pressure decreases when a solute is added proportional to the molar fraction of the solute.

The decrease in vapor pressure can be understood physically. When there is pure water and pure water vapor, there is an equilibrium between molecules evaporating from the solution and condensing into it. When solvent is added, the number density of water on the surface goes down slightly, so fewer water molecules evaporate per unit time while the same number are condensing. Since more condense than evaporate, the gas pressure goes down until equilibrium is reestablished, at a lower vapor pressure.

Knowing how the pressure changes, we can then find the temperature change from the Clausius-Clapeyron equation $\frac{dP}{dT}=\frac{PL}{RT^2}$, Eq. (11). Recall that $\frac{dP}{dT}$ is the slope of the phase boundary. For small ΔP and ΔT we can use $\frac{dP}{dT}=\frac{\Delta P}{\Delta T}$. We want to move back by $-\Delta P$ to restore the original pressure, so

$$\Delta T=-\Delta P\frac{RT^2}{P_0L}=\frac{N_s}{N_w}\frac{RT_0^2}{L} \tag{23}$$

Since $\Delta P<0$ at fixed T, the boundary shifts down/right, thus $\Delta T>0$ at fixed P. The signs of these equations are easiest to undrestand by looking at the liquid-vapor boundary in Fig. 5. A version of this diagram with salt-water included is shown in Fig. 6.

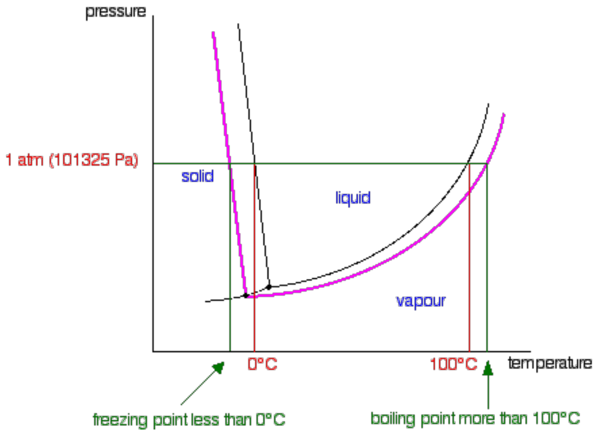


Figure 6. Black curve is the phase boundary for pure water, pink is for salt water. So at 1 atm, the boiling point goes up and the freezing point goes down.

For example, if you add a tablespoon of salt (0.547 mol) to 2 liters of water (111 mol), the vapor pressure at $T = 373K$ goes down from 1 bar by $\Delta P = -0.005$ bar. Using the latent heat of vaporization of water $L = 42\frac{\text{kJ}}{\text{mol}}$ we get that $\Delta T = 0.14K$. So the boiling point goes up, but by less than a degree. Thus raising the temperature is not the reason we add salt to water when cooking!

Adding salt to water also lowers its freezing point. The formula is the same as Eq. (23) with the opposite sign, since the salt is in the water, not the ice.[2] See also Fig. 6. Say we put 1 cup of salt (8.7 moles) out per square meter of ice that is 1 mm thick ($1L$ total, 55 moles). The latent heat of fusion for water is $L = 6.0\frac{\text{kJ}}{\text{mol}}$, about $1/7^{\text{th}}$ of the latent heat of vaporization. Then

$$\Delta T = -\frac{8.7}{55}\frac{8.3\frac{J}{\text{mol}\,K}(273K)^2}{6.0\,\frac{\text{kJ}}{\text{mol}}} = -16\,K \tag{24}$$

Thus if you salt your sidewalk, it won't freeze until the temperature drops to $T = -16°C = 3°F$.

So salt raises the boiling point and lowers the freezing point. The easiest way to understand this is that the salt disolves in the liquid phase of water. This increases the entropy of the liquid, but not the gas or solid. So it lowers the Gibbs free energy of the liquid relative to the other phases, and makes the liquid relatively preferable.

3.3 Chemical potential phase diagrams

As we saw, phase boundaries are determined by the condition that the chemical potentials of the two phases agree. If we are off a phase boundary, the phase with the lower chemical potential will dominate. To see this, recall that $G = \mu N$ so we are just minimizing Gibbs free energy to find the dominant phase. So at a given pressure, a given phase will dominate over the range of temperatures for which its chemical potential is lowest. This gives us a different perspective on phase transitions which is sometimes useful.

We know that solids will dominate at low temperature and gases at high temperature. Thus the solid phase has the lowest chemical potential at $T = 0$ and gas has the lowest chemical potential at high T. We also know that

$$\left(\frac{\partial \mu}{\partial T}\right)_P = -\frac{S}{N} < 0 \tag{25}$$

So the slopes of the chemical potential curves at constant pressure are always negative. Moreover, since $S_{\text{solid}} < S_{\text{liquid}} < S_{\text{gas}}$, the gas has the steepest chemical potential curve, followed by liquid, then solid. Finally, since since the entropy of a solid at $T = 0$ is zero or nearly zero by the 3rd law of thermodynamics, the solid line starts off horizontal. So a μ/T diagram will look something like this:

[2] This shouldn't be obvious, since we used $n_{\text{gas}} \ll n_{\text{liquid}}$, while we can't use $n_{\text{solid}} \ll n_{\text{liquid}}$. We actually used this limit twice: once to drop a term in Eq. (20) and once in Eq. (10). To derive Eq. (23) directly, you can avoid both expansions. Try it yourself!

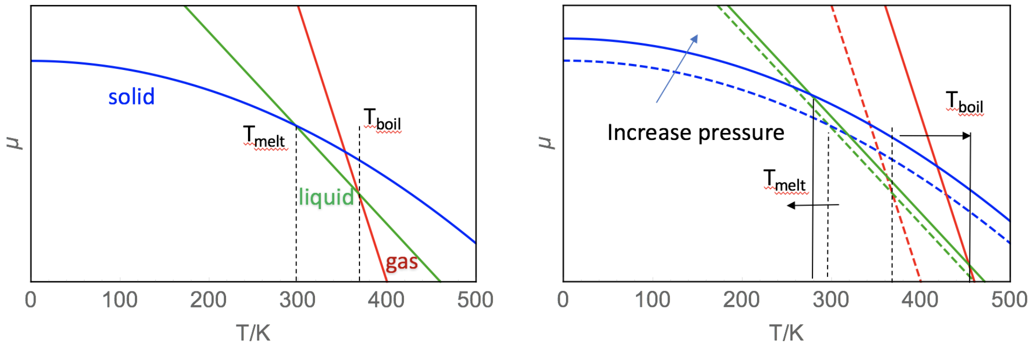


Figure 7. Phase diagram in the chemical potential/temperature plane for H_2O. Left shows the curves at some pressure P. Right shows the effect of increasing the pressure, whereby the dashed lines move to the solid lines.

What happens when we change the pressure? At a given temperature, if we change the pressure then

$$\left(\frac{\partial \mu}{\partial P}\right)_T = \frac{V}{N} = \frac{1}{n} > 0 \tag{26}$$

So the less dense the phase, the more its chemical potential changes, and increasing the pressure always drives the chemical potential up. Thus the gas curve shifts up the most as pressure increases. For water, as shown in the figure, the liquid is denser than the solid so its chemical potential changes less. We see therefore that at higher pressure, the melting temperature for water is lower and the boiling temperature is higher. The increase in temperature of boiling at higher pressure in qualitative agreement with Eq. (14).

Here's another example

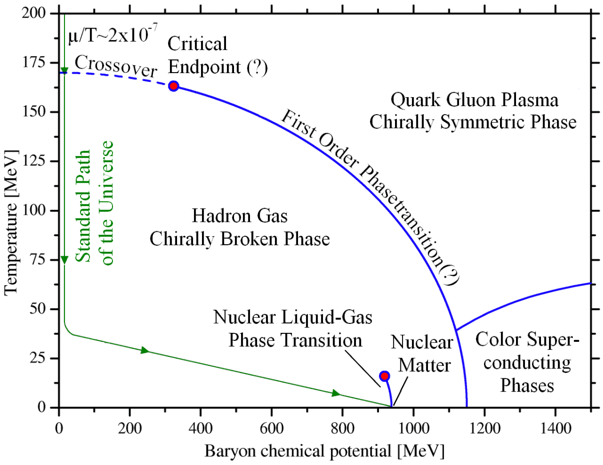


Figure 8. Phase diagram for subatomic matter

This figure shows the phase diagram for subatomic matter: quarks and gluons, as a function of temperature and baryon chemical potential. Baryon chemical potential is the chemical potential for quarks; since quark number is conserved, it can be nonzero. Some information about this phase diagram we know from nuclear physics, some from astrophysics (e.g. neutron stars), some from collisions of ionized lead and ionized gold at particle accelerators, some from cosmology, some from theoretical calculations and simulations. We even have some insight into this phase diagram from string theory. A lot is still unknown. An open question about this phase diagram is whether there is a critical point between the quark-gluon plasma phase and the hadron gas phase (the red dot). This could have implications for the earliest moments of the universe, just after the big bang.

4 General phase transitions

We saw with the liquid/solid/gas phase transitions that $\left(\frac{\partial G}{\partial T}\right)_P$ changes discontinuously at the phase boundary. When this happens we say the transition is of the first order.

- **First-order phase transition**: $\left(\frac{\partial G}{\partial T}\right)_P$ changes discontinuously at the phase boundary

The "first" in "first order" refers to the first derivative of G. It is possible for $\left(\frac{\partial G}{\partial T}\right)_P$ to be continuous, but higher derivatives of G to be discontinuous:

- **nth order phase transition**: $\left(\frac{\partial^n G}{\partial T^n}\right)_P$ changes discontinuously at the phase boundary

This classification of phase transitions is known as the **Ehrenfest classification**. It is common to call any transition with $n > 1$ a **second order transition**. So in common parlance

- **Second-order phase transition**: $\left(\frac{\partial G}{\partial T}\right)_P$ is continuous across the phase boundary

This is because first order transitions are special: they have latent heat, which is a barrier to changing phases. Second order transitions are smooth and have no barrier. They occur at points in phase diagrams where the phases merge into one, and the latent heat vanishes.

In modern treatments, we use first and second order a little more loosely. It doesn't have to be the Gibbs free energy that determines the order, it can be the Helmholtz free energy, or just the energy, depending on what is appropriate for the problem. And it doesn't have to be derivative with respect to temperature, but can be some other derivative of the energy. The thing whose continuity we are questioning is called the **order parameter**. For example, in the Ehrenfest classification, the entropy $S = -\left(\frac{\partial G}{\partial T}\right)_P$ is the order parameter. For boiling water, density is a natural order parameter and a little more intuitive than the entropy. At fixed N and m the density $\rho = \frac{Nm}{V}$ is equivalent to using $V = \left(\frac{\partial F}{\partial P}\right)_T$.

4.1 Paramagnetism

Paramagnetic means that a material is attracted to an applied magnetic field, like iron filings to a magnet. Many elements (gold, potassium, calcium,...) are paramagnetic. Most compounds and stable molecules are weakly paramagnetic. Paramagnetic materials generally have unpaired electrons that are free to align the external field, while diamagnetic materials have closed orbits. The opposite of paramagnetic is **diamagnetic**, which means something is repelled by an applied magnetic field (google "levitating frog" for example). A third type of magnetism is **ferromagnetism**, whereby a material can produce a coherent magnetic field, i.e. as in an ordinary iron refrigerator magnet. If you heat an iron magnet above 1043 K it will lose its magnetism and become paramagnetic. The transition between paramagnetic and ferromagnetic phases of iron is an example of a phase transition.

The ferromagnetism in iron comes about when all the magnetic spins in iron are aligned. The spins themselves can be thought of as little magnets that attract each other.

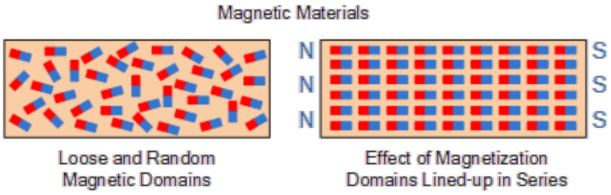

Magnetic Materials
N
N
N
S
S
S
Loose and Random Magnetic Domains
Effect of Magnetization Domains Lined-up in Series

Figure 9. When a magnetic material is cooled, its spins spontaneously align as it enters the ferromagnetic phase.

In this crude model, each pair has energy $-\varepsilon$ if they are aligned and $+\varepsilon$ if they are not aligned. So the difference between the fully aligned (ferromagnetic) state and disordered (paramagnetic) state is roughly $E = 2N\varepsilon$. On the other hand, the disordered state has much higher entropy. The entropy of the fully-aligned magnetic state is zero. But the disordered state has $\Omega = 2^N$ configurations and entropy $S = Nk_B \ln 2$. Thus the free energy in the magnetic state is

$$F_{\text{magentic}} = E - TS = -N\varepsilon \tag{27}$$

and the free energy of the disordered state is

$$F_{\text{disordered}} = E - TS = N(\varepsilon - Tk_B \ln 2) \tag{28}$$

By minimizing the free energy we see that the transition from paramagnetic to ferromagnetic state occurs when these free energies cross each other, which is at

$$k_B T_c \approx \frac{2\varepsilon}{\ln 2} \tag{29}$$

T_c is known as the **Curie temperature**.

What is a good order parameter for the transition? Entropy S should work, as it goes from $S = Nk_B \ln 2$ down to zero. Unlike the liquid/gas transition however, the entropy change as the magnet is cooled will be smooth: at temperatures near T_c the magnetic will have some spins aligned, be slightly ferromagnetic and have some intermediate entropy. Thus the ferromagnetic phase transition is second order.

Another order parameter we could consider is the magnetization M. The direction of the magnetic field is a vector $\vec{M}$, and we can define $M = |\vec{M}|$. Above T_c, $M = 0$ exactly. As T is lowered, M is nonzero. This function $M(T)$ is going to be continuous across T_c. However all of the derivatives of $M(T)$ cannot be continuous – if they were then the function would have to be $M(T) = 0$ exactly for all T (mathematically, this is a property of analytic functions).

It is interesting to think of the direction of $\vec{M}$ rather than its magnitude as the order parameter. In the magnetic phase, the direction of $\vec{M}$ picks up some definite value (there can be domains inside the magnetic with different directions, but let's just focus on one domain for now). One super interesting thing about $\vec{M}$ picking up a direction is that we cannot know ahead of time which direction it would be. At high temperature no direction is preferred. Indeed, all the spins are constantly flipping around in 3D. So the theory at high temperature is rotationally invariant. At low temperature, it is not. But fundamentally, the interactions between spins are like $\vec{s}_1 \cdot \vec{s}_2$ involving a rotationally-invariant dot product. So the Hamiltonian is rotationally invariant and it is only the state that violates the symmetry. When this happens we say that the rotational symmetry is **spontaneously broken**.

Symmetries are an extraordinarily powerful tool in physics. In this context, they help specify phases and phase transitions. For another example, note that a liquid is invariant under translations. Microscopically, of course the molecules are in particular positions. But knowning the position of some of the molecules does not tell us anything about where molecules far away are. On the other hand, a solid is *not* translationally invariant. Once you know where one atom is, you can pinpoint all the rest throughout the crystal. We say the solid has long-range order. So when a liquid freezes, translational symmetry is spontaneously broken and long-range order results.

There are lots of consequences of spontaneous symmetry breaking. One is that it tells us that there must be arbitrarily low energy excitations of the system. This very general result is known as **Goldstone's theorem**. For example, in a solid, we know we can push it and it will move. It can move arbitrarily slowly, so we can push it with arbitrarily low energy to have an effect. However, as we push it, what we actually do is push the atoms on the side where our hand is. These atoms push the next atoms, and so on, all throughout the solid. So really, to move the solid by pushing, we are setting up a wave of very low energy. As we move an atom, the system works to restore the lattice to how it was. This doesn't happen in a liquid where the translation symmetry is unbroken. The excitations of a solid are called **phonons** and have a massless dispersion relation $\omega(k) \to 0$ as $k \to 0$. Phonons are covered in Lecture 11. The excitations in a magnet from the spontaneously broken rotational symmetry also have $\omega(k) \to 0$ as $k \to 0$. They are called **spin waves**.

To understand spontaneous symmetry breaking, the emergence of long-range order, etc. requires techniques of condensed matter physics that take us well beyond the course material.

4.2 Critical phenomena

The final topic I want to mention concerns that dot at the end of the liquid/gas phase boundary denoted as the **critical point**. Critical points are super-interesting places with a lot of unusual properties.

Recall that as you heat up a liquid, it will eventually vaporize. At the point of the phase transition, the heat will go into latent heat of vaporization and the temperature will not change. Eventually, all the liquid is vaporized and the heat will start raising the temperature again. We can see this fairly clearly in a phase diagram in the pressure-volume plane, as this one for CO_2:

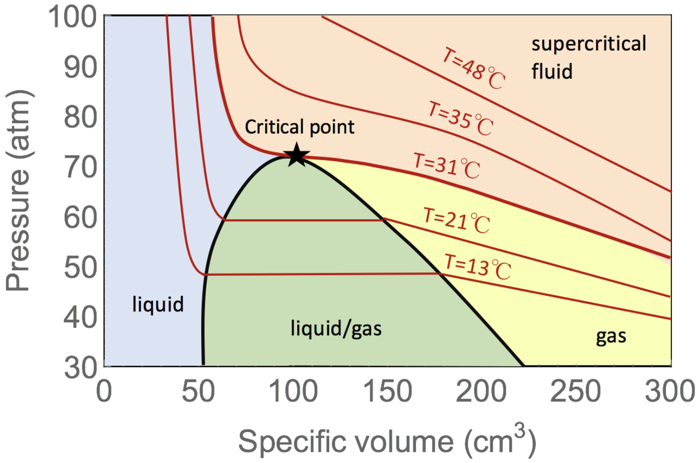


Figure 10. P-V phase diagram for CO_2. The x-axis is the specific volume, $v = \frac{V}{N}$. The green region has liquid and gas.

The lines in this plot are lines of constant temperature, or **isotherms**. Say we start on left in the liquid region, along the $T = 13°C$ isotherm. As we decompress the CO_2 isothermally, we will move along the isotherm, lowering pressure and increasing the specific volume (lowering the density). When we hit the phase boundary, the liquid starts vaporizing. During the phase transition pressure and temperature are fixed, so we move horizontally, $\frac{\partial P}{\partial v} = 0$. After the phase transition is complete, CO_2 is all gas, and the decompression lowers the pressure and density once again. So the isotherm gives a decreasing function $P(v)$ outside of the coexistence region and is flat in the coexistence region.

A natural order parameter for the transition is the number density n, or equivalently the specific volume, $v = \frac{1}{n}$. We see from the figure that the specific volume changes discontinuously from liquid to gas along the $T = 13°C$ isotherm, so the transition at this temperature is 1st order.

Now consider what happens as the temperature is increased. As you can see from the figure, at a higher temperature, the difference in specific volume between the liquid and gas phases at constant temperature is smaller. The latent heat of vaporization is smaller too. Eventually, there is no difference between the liquid and gas phases and the latent heat vanishes: it takes no energy to convert a liquid to a gas. This happens at the critical temperature T_c, which intersects the liquid/gas coexistence region at the critical specific volume v_c and critical pressure P_c, that is, at the critical point. The specific volume changes smoothly from liquid to gas if we pass through the critical point, so the phase transition at this point is second order.

At temperatures above the critical temperature, the material is both gas and liquid, or neither gas nor liquid, depending on how you look at it. We call it a **supercritical fluid**. The "super" in this context just means "beyond" – in contrast to the "super" in superfluids or superconductors which are truly exotic phases of matter. A supercritical fluid is in between a gas or a liquid. For example, supercritical CO_2 is used to decaffeinate coffee – its viscosity and diffusivity are like those of a gas, so it penetrates the beans easily, and its density is like that of a liquid, so a lot of it can get in. It happens also to bind well to caffeine (this property is much more important that its supercritical fluid properties). Supercritical CO_2 is also used in dry cleaning.

On any isotherm in the liquid/gas region, $\left(\frac{\partial P}{\partial v}\right)_T = 0$. At the critical point, the length of the horizontal part of the isotherm has gone to zero, but it is still flat. Moreover, since the isotherms are decreasing on either side of the critical point, we know that $\left(\frac{\partial^2 P}{\partial v^2}\right)_T = 0$ as well: the critical point is a point of inflection. It also happens to be true that, but is not so easy to show, that all of the derivatives of P vanish, $\left(\frac{\partial^n P}{\partial v^n}\right)_T = 0$. Mathematically, this means that $P(v)$ is a non-analytic function at the critical point. Second order phase transitions are super interesting because somehow this crazy mathematical behavior, where all the derivatives vanish, arises out of functions like the entropy S or the partition function Z that depend smoothly on temperature, pressure, volume, etc.

It's not just the derivatives $\left(\frac{\partial^n P}{\partial v^n}\right)_T$ that vanish. We could equally well have looked at the phase diagram in the $T - v$ plane:

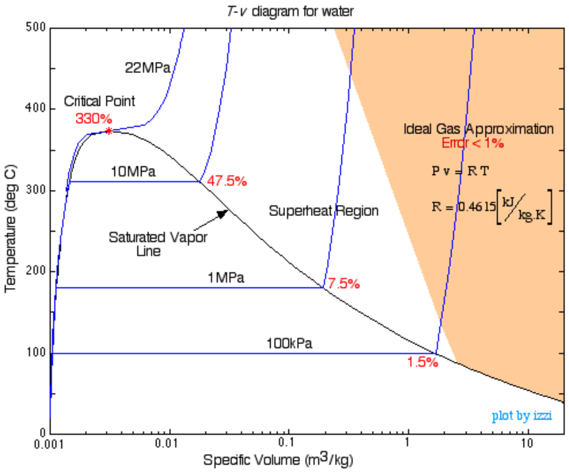


Figure 11. TV diagram for water

In this case, at subcritical pressure, water increases its specific volume when heated until it boils. At the critical point, the latent heat vanishes and the water and steam become the same. The critical point is also a point of inflection for $T(v)$, and in fact, $\left(\frac{\partial^n T}{\partial v^n}\right)_P = 0$, so $T(v)$ is a non-analytic function.

Away from the critical point, the various thermodynamic quantities that we have discussed, latent heat, enthalpy of formation, heat capacity, isothermal compressibility, etc., help us distinguish one material from another. But these are all related to derivatives of thermodynamic quantities. Because all the derivatives vanish, $\left(\frac{\partial^n P}{\partial v^n}\right)_T = \left(\frac{\partial^n T}{\partial v^n}\right)_P = 0$, all the dimensionful physical quantities we use to characterize a material either vanish or are infinite at this point. For example, here is a plot of heat capacity of propane near the critical point:

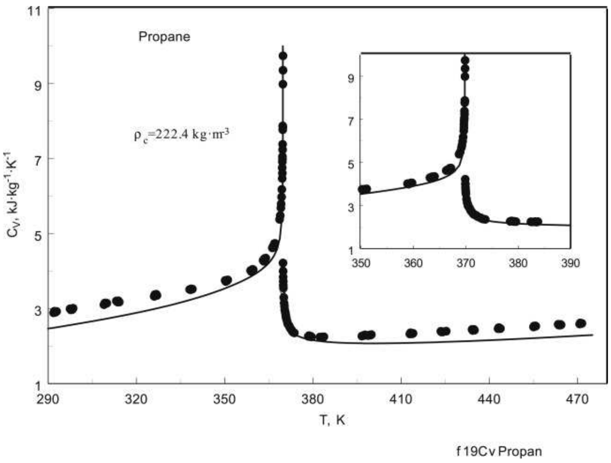


Figure 12. Heat capacity of propane near the critical point, showing the singularity.

One way to understand the disappearance of scale more physically is to think about water. The difference between liquid water and an ideal gas is that water has hydrogen bonds (see Fig. 2). In water vapor, the molecules are generally too far apart for hydrogen bonds to matter. The relative importance of hydrogen bonds in water versus gas determines the surface tension γ and everything else that makes water a liquid. However, as the density or pressure on the gas is increased, the relative importance hydrogen bonds in the vapor phase increases too. Consequently the surface tension of the liquid/gas boundary goes down. So typical droplet sizes grow. At the critical point, the surface tension vanishes and droplets of any size can form: the single dimensional scale (the surface tension γ) has vanished. This can be seen through the phenomenon of critical opalescence.

We say that the theory near the critical point is **scale-invariant** or **conformal**, since no dimensionful quantity is available to characterize the material. That is, choosing units relative to the critical values:

$$\hat{T} = \frac{T}{T_c}, \quad \hat{P} = \frac{P}{P_c}, \quad \hat{v} = \frac{v}{v_c}, \quad \hat{n} = \frac{1}{\hat{v}} = \frac{n}{n_c} = \frac{\rho}{\rho_c} \tag{30}$$

all thermodynamic quantities, such as the free energy, become *independent of any other property of the material*. For example, P_c and T_c for neon, argon, krypton, xenon, N_2, O_2, CO and CH$_4$ are

	Ne	Ar	Kr	Xe	N$_2$	O$_2$	CO	CH$_4$
$T_c\,(°C)$	-228.7	1122.3	-63.8	16.6	-147	-118.4	-140	-82.1
$P_c\,$(atm)	26.9	48	54.3	58	33.5	50.1	34.5	45.8

(31)

These temperatures and pressures are dimensionful quantities, with no apparent relation among them. Now, we rescale the temperatures, pressures, specific volumes and specific densities by these critical values and look at $\hat{T}$ and $\hat{n}$ near the critical points for the different materials:

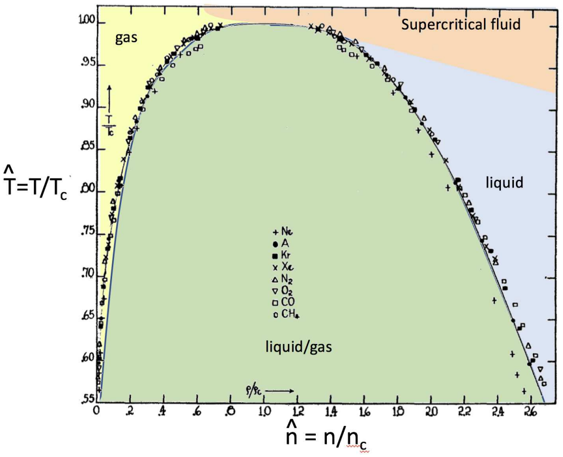


Figure 13. Reduced temperature versus reduced number density for a variety of different substances at saturation. Adapted from E.A. Guggenheim, J. Chem. Phys. 13, 253 (1945).

Remarkably, in the vicinity of the critical point, the $\hat{T} - \hat{n}$ phase boundaries all have exactly the same shape! This implies that all of the derived quantities, such as heat capacity, isothermal compressibility, etc, should be related in every material near its critical point. This powerful observation is known as the **law of corresponding states**.

The shape in Fig. 13 is fit by a functions $\hat{n}(\hat{T})$ on the liquid and gas side of the form

$$\hat{n}_\ell(\hat{T}) = 1 + \frac{3}{4}(1 - \hat{T}) + \frac{7}{4}(1 - \hat{T})^{1/3}, \qquad \hat{n}_g(\hat{T}) = 1 + \frac{3}{4}(1 - \hat{T}) - \frac{7}{4}(1 - \hat{T})^{1/3} \tag{32}$$

The fractional exponent makes these functions non-analytic, so that all the derivatives are singular at the critical point $\hat{T} = 1$, $\hat{n}^{(k)}(1) = \infty$. This means all the derivatives of $\hat{T}(\hat{n})$ vanish at the critical point $\hat{n} = 1$, $\hat{T}^{(k)}(1) = 0$.

The exponent $1/3$ in Eq. (32) is an example of a **critical exponent**. It characterizes the approach of the density towards the critical point. If we used an order parameter other then density, for example, the heat capacity, it would approach the critical point with a different scaling behavior $C_V \sim (1 - \hat{T})^{-\alpha}$. Because of the law of corresponding states, the critical exponents can be calculated with any material for which the order parameter applies. In fact, the material can be a made-up theoretical one: the universality is so strong that the material doesn't even have to exist. An important example is the ising model, which treats a material as a lattice of spins with interaction energies taking the values ± 1. Computing the critical exponents of the ising model agrees with measured values of the critical exponents in water to one part in 1000!

5 Summary

This lecture introduced the notion of a phase of matter. Unfortunately, it has hard to define what a phase is precisely, since one often smoothly go between two phases. For example, at very high temperature, liquid water and gaseous water are pretty hard to tell apart. There are two ways out.

First, we can use the term "phase" to partition states of matter only into regions that cannot be smoothly transformed into each other. With this definition, liquid and gaseous water are the same phase, but solid is different. The solid phase has less symmetry than the liquid/gas phase (discrete translational symmetry instead of full continuous translational and rotational symmetry). More generally, one can go further and define phases by their symmetries. This approach can be very powerful and mathematically rigorous. However, it sidesteps the fact that liquid and gaseous water clearly are different!

Second, we can take a more local point of view, and define phases as regions where a particular theoretical model is accurate. The model amounts to a set of assumptions and the specification of a partition function. For example, we could neglect interactions to make an ideal gas. Then the smooth transition between phases goes through a region where the assumptions break down. When two models are accurate in the same thermodynamic region (P, V, T), there is a phase boundary. This is a more general treatment of phases than the first option, but requires approximations. Indeed, the business of physicists is making approximations, and such approximations necessarily break down. This second approach is typically what physicists mean by a phase.

When you have a phase boundary, the chemical potentials/Gibbs free energies of the two phases must agree. Indeed, the equality of chemical potentials is a neccessary condition for equilibrium. However, the entropy, which is the derivative of μ with resect to T, might not be continous. When S changes abrubtly, we say the phase transition is first order. So the chemical potential has a kink for a first order phase transition. If S is continous, but $\frac{\partial S}{\partial T} = \frac{\partial^2 \mu}{\partial T^2}$ is not continous, we say the transition is second order.

Phase boundaries are fascinating places. The slope of the phase boundary is given by the Clausius equation $\frac{dP}{dT} = \frac{1}{T} \frac{L}{\Delta\left(\frac{1}{n}\right)}$, where the latent heat L is the change in (molar) enthalpy between the two phases $L = \Delta\left(\frac{H}{N}\right)$. For liquid-gas transitions, the Clausius equation simplies to the Clausius-Clapyron equation $\frac{dP}{dT} = \frac{PL}{RT^2}$ Latent heat is the energy cost of a phase transition. When you boil water, you break up all the hydrogen bonds to make liquid gas. The latent heat quantifies this. As you move along a phase boundary, the latent heat can get smaller and smaller until it finally vanishes. The vanishing point is called the critical point. The phase transition is first order when $L \neq 0$ and becomes second order when $L = 0$. The latent heat characterizes the energy scale of the phase transition, so when $L = 0$ there is no scale, and so critical points are characterized by long-range correlations.

We are often interested in how phase boundaries change when we modify the phases. For example, adding salt to water changes its boiling point. Studying phase boundaries has allowed us to quantify these effects, through equations like Raoult's law.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 10: Quantum Statistical Mechanics

1 Introduction

So far, we have been considering mostly classical statistical mechanics. In classical mechanics, energies are continuous, for example momenta $\vec{p}$ and hence kinetic energy $E = \frac{\vec{p}^2}{2m}$ can be any real number. Positions $\vec{q}$ can also take any continuous values in a volume $V = L^3$. Thus the number of states Ω, the entropy S, the partition function Z, all involve $\int dq dp$ which is formally infinite. We regulated these infinities by artificially putting limits Δq and Δp on the positions and momenta that are allowed, and we found that physical predictions we made did not depend on Δp and Δq. Quantum statistical mechanics will allow us to remove these arbitrary cutoffs.

In quantum mechanics there are two important changes we must take into account. First, position, momentum and energy are in general quantized. We cannot specify the position and momentum independently, since they are constrained by the uncertainty principle $\Delta p \Delta q \leqslant \frac{\hbar}{2}$. For example, the momenta modes of a particle in a box of size L are quantized as $p_n = \frac{2\pi}{L}\hbar n$, with n an integer. There is no extra degree of freedom associated with position. When summing over states, we sum over $n = \frac{L p_n}{2\pi\hbar}$ only, not over q. However, in the continuum limit we can replace the sum over n with an integral, and suggestively write $L = \int dq$. That is, the quantum sum gets replaced by an integral as

$$\sum_n \to \int dn = \frac{L}{2\pi\hbar}\int dp = \int \frac{dq dp}{h} \qquad (1)$$

Recall that in classical statistical mechanics when we first computed Ω in the microcanonical ensemble, we had to arbitrarily break up position and momentum into sizes Δq and Δp. Eq. (1) justifies the replacement $\Delta p \Delta q \to h$ that we previously inserted without proof in the Sackur-Tetrode equation and elsewhere.[1] Note that the integral measure has a factor of h not $\hbar$. We'll do conversions between sums and integrals like in Eq. (1) in more detail in deriving the density of states over the next several lectures.

Quantization is important when we are *not* in the continuum limit. One regime where quantum effects are clearly important is when the temperature is of order the lowest energy of the system, $k_B T \sim \varepsilon_0$. For example, classically, each vibrational mode of a molecule contributes $N k_B$ to the heat capacity, but at low temperature, the measured heat capacity does not show this contribution. In quantum statistical mechanics, the vibrational contribution to the heat capacity is cut off at temperatures below $k_B T \lesssim \hbar\omega_0$, with ω_0 the vibrational frequency, in agreement with measurements. We discussed this effect in Lecture 7.

The second quantum effect we need to account for has to do with identical particles. In quantum mechanics, identical particles of half-integer spin, like the electron, can never occupy the same state, by the Pauli exclusion principle. These particles obey **Fermi-Dirac statistics**. Identical particles of integer spin, like the photon, can occupy the same state but the overall wavefunction must be symmetric in the interchange of the two particles. Integer-spin particles obey **Bose-Einstein statistics.**

We discussed indistinguishability before in Lecture 6, in the context of the second law of thermodynamics and the Gibbs paradox. In that lecture, we found that we could decide if we want to treat particles as distinguishable or indistinguishable. If we want to treat the particles as distinguishable, then we must include the entropy increase from measuring the identity of all the particles to avoid a conflict with the second law of thermodynamics. Alternatively, if

[1] Eq. (1) makes the appearance of dq seem like a trick. This is an artifact of using momentum eigenstates for our basis. If we used something symmetric in p and q like coherent states, the product $dpdq$ would appear more naturally.

we never plan on actually distinguishing them, we can treat them as indistinguishable. We do this by adding a factor of $\frac{1}{N!}$ to the number of states Ω, i.e. instead of $\Omega \sim V^N$ we take $\Omega \sim \frac{1}{N!}V^N$; then there is automatically no conflict with the second law of thermodynamics. This kind of classical indistinguishable-particle statistics, with the $N!$ included, is known as **Maxwell-Boltzmann statistics.** Quantum identical particles is a stronger requirement, since it means the multiparticle wavefunction must be totally symmetric or totally antisymmetric. In a classical system, the states are continuous, so there is exactly zero chance of two particles being in the same state. Thus, the difference among Fermi-Dirac, Bose-Einstein and Maxwell-Boltzmann statistics arises entirely from situations were a single state has a nonzero change of being multiply occupied.

2 Bosons and fermions

The wavefunction $\psi(x,s)$ of an electron depends on the coordinate x and the spin $s=\pm\frac{1}{2}$. When there are two electrons, the wavefunction depends on two coordinates and two spins. We can write it as $\psi(x_1,s_1,x_2,s_2)$. Thus $|\psi(x_1,s_1,x_2,s_2)|^2$ gives the probability of finding one electron at x_1 with spin s_1 *and* one electron at x_2 with spin s_2. Every electron is identical and therefore[2]

$$|\psi(x_1,s_1,x_2,s_2)|^2=|\psi(x_2,s_2,x_1,s_1)|^2 \tag{2}$$

If the modulus of two complex numbers is the same, they can only differ by a phase. We thus have

$$\psi(x_1,s_1,x_2,s_2)=\eta\psi(x_2,s_2,x_1,s_1) \tag{3}$$

for some phase $\eta=e^{i\theta}$ with $|\eta|=1$. If we swap the particles back, we then find

$$\psi(x_1,s_1,x_2,s_2)=\eta\psi(x_2,s_2,x_1,s_1)=\eta^2\psi(x_1,s_1,x_2,s_2) \tag{4}$$

So that $\eta^2=1$ and $\eta=\pm1$. We call particles with $\eta=1$ **bosons** and those with $\eta=-1$ **fermions**.

Particles come with different spins $s=0,\frac{1}{2},1,\frac{3}{2},\cdots$. A beautiful result from quantum field theory is that particles with integer spins, $s=0,1,2,\cdots$ are bosons and particles with half-integrer spins $s=\frac{1}{2},\frac{3}{2},\cdots$ are fermions. This correspondance is known as the **spin-statistics theorem**.

As an aside, I'll give a quick explanation of where the spin-statistics theorem comes from. First we need to know what spin means. The spin determines how the state $|\psi\rangle$ changes under rotations. For example, say a particle is localized at $x=0$, so $\psi(\vec{x})=\langle x|\psi\rangle=\delta^3(x)$, and has spin pointing in the z direction, so $j_z=s$. Now rotate it around the x axis by an angle θ. Doing so means it will pick up a phase determined by the spin

$$|\psi\rangle\rightarrow e^{is\theta}|\psi\rangle \tag{5}$$

In other words, the spin s is the coefficient of θ in the phase. For example, write the electric field vector $\vec{E}$ as a complex number $z=E_x+iE_y$. Then under a rotation by angle θ, $z\rightarrow e^{i\theta}z$, so $s=1$ according to Eq. (5). This tells us that photons (the quanta of the electric field) have spin 1. Under a 360° rotation ($\theta=2\pi$), we are back to where we started. But that doesn't mean that $|\psi\rangle\rightarrow|\psi\rangle$. Indeed, we see directly that if s is a half-integer, then $|\psi\rangle=-|\psi\rangle$ while if s is an integer then $|\psi\rangle\rightarrow|\psi\rangle$. Now consider a two particle state, $\psi(x_1,\uparrow,x_2,\uparrow)$ with $s=\uparrow$ denoting spin up in the z direction. Now we rotate by 180° around the center between x_1 and x_2, this interchanges the particles, but adds a phase $e^{is\pi}$ for particle 1 and another phase $e^{is\pi}$ for particle 2. So the interchange adds a factor of $e^{2\pi is}=\pm1$ total to the wavefunction. This is exactly the factor η in Eq. (3). Thus $\eta=e^{2\pi is}$ which is 1 for integer spins and -1 for half-integer spins. In other words, integer spin particles are bosons and half-integer spin particles are fermions.

[2] Electrons are identical by definition, since if every electron were not identical, we could give the different "electrons" different names. For example, there's a particle called the muon which is similar to the electron but heavier. Thus the electron/muon wavefunction will have $\psi(x_1,x_2)\neq\psi(x_2,x_1)$. Unless there are an infinite number of possible quantum numbers that we can use to tell all electrons apart, some of them will be identical and satisfy $|\psi(x_1,s_1,x_2,x_2)|^2=|\psi(x_2,s_2,x_1,s_1)|^2$. The identical particles for which quantum statistics applies are, by definition, these identical ones.

The main implication of the spin-statistics theorem for statistical mechanics is the Pauli exclusion principle: no two fermions (like electrons) with the same quantum numbers (like spin) can occupy the same state. To see this, say the wavefunction for an state is $\psi_0(x,\, s)$. Then if two particles were in the same state with the same spin then $\psi(x_2, s_2, x_1, s_1) = \psi_0(x_1, s_1)\psi_0(x_2, s_2)$. This is symmetric under interchange. So if $\psi(x_2, s_2, x_1, s_1) = -\psi(x_1, s_1, x_2, s_2)$, as for fermions, then ψ must vanish. Instead, for fermions we can only have $\psi(x_2, s_2, x_1, s_1) = \psi_0(x_1, s_1)\psi_1(x_2, s_2)$ with two *different* single particle wavefunctions ψ_0 and ψ_1. More generally, each identical fermion must be in a different state, so when we pile N fermions into a system, they start filling the energy levels from the bottom up.

2.1 Three types of statistics

There are 3 types of statistics we will discuss

- In (classical) Maxwell-Boltzmann statistics any particle can be in any state and an overall factor of $\frac{1}{N!}$ for indistinguishability is added *ad hoc* to the number of states $\Omega = \frac{1}{N!}\sum_k 1$ (microcanonical ensemble) or to partition function: $Z = \frac{1}{N!}\sum_k e^{-\beta E_k}$ (canonical ensemble).
- In Bose-Einstein statistics multiple particles can occupy each single-particle state. But when they do, there is only one combined state with the multiple particles in it. No $N!$ is added: $Z = \sum e^{-\beta E_k}$.
- In Fermi-Dirac statistics, no two particles can occupy the same state. No $N!$ is added and $Z = \sum e^{-\beta E_k}$.

Let's compare the situations with an example. Suppose there are 2 particles and 3 possible single-particle states with energies ε_1, ε_2 and ε_3. These energies do not have to be different, but let's suppose they are for simplicity (the single-particle states could have the same energies but different quantum numbers, such as different spin or position or vibrational excitation along a different axis). For Maxwell-Boltzmann statistics, the particles are treated as distinguishable for counting states, and the $\frac{1}{N!}$ is added at the end. Denoting the two particles A and B, the possible states and canonical partition function (with $\beta = 1$ for simplicity) in this case is:

ε_1	ε_2	ε_3
AB		
A	B	
A		B
B	A	
	AB	
	A	B
B		A
	B	A
		AB

$$, \qquad Z_{\text{MB}} = \frac{1}{2!}[e^{-2\varepsilon_1} + e^{-2\varepsilon_2} + e^{-2\varepsilon_3} + 2e^{-\varepsilon_1-\varepsilon_2} + 2e^{-\varepsilon_1-\varepsilon_3} + 2e^{-\varepsilon_2-\varepsilon_3}] \tag{6}$$

The $\frac{1}{2!}$ is the identical particles factor. Note that we treat the two particles as different when constructing the partition function, then divide by $N!$.

For Bose-Einstein statistics, the possible states and canonical partition function are

ε_1	ε_2	ε_3
AA		
A	A	
A		A
	AA	
	A	A
		AA

$$, \qquad Z_{\text{BE}} = e^{-2\varepsilon_1} + e^{-2\varepsilon_2} + e^{-2\varepsilon_3} + e^{-\varepsilon_1-\varepsilon_2} + e^{-\varepsilon_1-\varepsilon_3} + e^{-\varepsilon_2-\varepsilon_3} \tag{7}$$

Note that we treat the particles as identical, so $AA = BA = AB$ is one state. No $N!$ is added.

For Fermi-Dirac statistics, the possible states and canonical partition function are

ε_1	ε_2	ε_3
A	A	
A		A
	A	A

$$Z_{\text{FD}} = e^{-\varepsilon_1 - \varepsilon_2} + e^{-\varepsilon_1 - \varepsilon_3} + e^{-\varepsilon_2 - \varepsilon_3} \tag{8}$$

Note that there are 9 two-particle states for the Maxwell-Boltzmann case, 6 for the boson case and only 3 for the fermion case. The 3 two-particle states in the fermion case are the ones where each particle has a different energy. If there were m possible energy levels instead of 3 for the two particles, then there would be

$$n_{\text{fermi}} = \left(\begin{array}{c} m \\ 2 \end{array} \right) = \frac{1}{2} m^2 - \frac{1}{2} m \tag{9}$$

possible two-fermion states. The number of boson two-particle states would include all the possible fermion ones plus the m states where both particles have the same energy, so

$$n_{\text{bose}} = n_{\text{fermi}} + m = \frac{1}{2} m^2 + \frac{1}{2} m \tag{10}$$

The number of classical two-particle states is $n_{\text{class}} = m^2$. To compare to the fermion or boson cases, note that in the limit that $k_B T \gg \varepsilon_i$ for all the energies, the Boltzmann factors all go to 1 so the partition function Z just counts the number states. Thus the effective number of states in Maxwell-Boltzmann statistics is

$$n_{\text{MB}} = Z_{\text{MB}}(k_B T \gg \varepsilon_j) = \frac{1}{2!} n_{\text{class}} = \frac{1}{2} m^2 \tag{11}$$

Thus we see that for 2 particles when the number of thermally accessible energies m is large, the number of two-particle states is the same in all three cases.

For N particles and m energy levels, the same analysis applies. The difference between the fermion and boson cases is due to the importance of states where more than one particle occupy the same energy level. The number of such states is always down by at least a factor of m from the number of states where all the particles have different energies. Thus when m gets very large, so that there are a large number of thermally accessible energy levels, then the distinction between bosons and fermions starts to vanish. In other words, $m \gg N$ almost all the microstates will have all the particles at different energies. For any particular set of N energies for the N particles there would be $N!$ classical assignments of the particles to the energy levels, but only one quantum assignment. This explains why the $\frac{1}{N!}$ Gibbs factor for identical particles in Maxwell-Boltzmann statistics is consistent with quantum statistical mechanics. We will make more quantitative the limit in which quantum statistics is important in Section 4.

3 Non-interacting gases

Next we compute the probability distributions for the different statistics, assuming only that the particles are non-interacting. We call the non-interacting systems *gases*, i.e. Bose gas and Fermi gas. Neglecting interactions allows us to write the total energy as the sum of the individual particle energies. For photons, neglecting interactions is an excellent approximation since Maxwell's equations are linear – the electromagnetic fields does not interact with itself. For fermions, like electrons, even though there are electromagnetic interactions, because fermions like to stay away from each other (due to the Pauli exclusion principle), the interactions are generally weak. Thus the non-interacting-gas approximation will be excellent for many bosonic and fermionic systems, as we will see over the next 5 lectures.

Let us index the possible energy levels each particle can have by $i = 1, 2, 3, \cdots$. The energy of level i is ε_i. For example, a particle in a box of size L has quantized momenta, $p_n = \frac{n\pi}{L}\hbar$ and energies $\varepsilon_n = \frac{p_n^2}{2m}$. In any microstate with a given N and E there will be some number n_i of particles at level i, so $\sum n_i = N$ and $\sum n_i \varepsilon_i = E$. It is actually very difficult to work with either the microcanonical ensemble (fixed N and E) or the canonical ensemble (fixed N and T) for quantum statistics. For example, with 2 particles, the canonical partition function for a Bose gas would be

$$Z_{\text{BE}} = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} e^{-\beta(\varepsilon_i + \varepsilon_j)} \tag{12}$$

You can check that with only 3 energy levels, this reproduces Eq. (7). Note that we must have $j \geqslant i$ in the second sum to avoid double counting the (i, j) and (j, i) states. For N particles, we would need N indices and can write

$$Z_{\text{BE}} = \sum_{i_1 \leqslant i_2 \leqslant \cdots \leqslant i_N} e^{-\beta(\varepsilon_{i_1} + \varepsilon_{i_2} + \cdots + \varepsilon_{i_N})} \tag{13}$$

Doing the sum this way over N indices is very difficult; even in the simplest cases it is impossible to do in closed form. For Fermi-Dirac statistics the partition function can be written the same way with the sum over indices with strict ordering: $i_1 < i_2 < \cdots < i_N$. It is also very hard to do.

Conveniently, the grand canonical ensemble comes to the rescue. In the grand canonical ensemble, we sum over N and E, fixing instead μ and T. The grand partition function is

$$\mathcal{Z} = \sum_{N, E} e^{-\beta(E - N\mu)} \tag{14}$$

Since $\sum n_i = N$ and $\sum n_i \varepsilon_i = E$ the grand partition function can also be written as

$$\mathcal{Z} = \prod_{\text{single particle states } i} \sum_{n_i = 0}^{\infty} e^{-n_i \beta(\varepsilon_i - \mu)} \tag{15}$$

where n_i is the number of particles in state i. To see that this agrees with Eq. (14) consider that each term in the product over i picks one factor from each term in the sum over n_i. Since all possible values of ε_i and n_i are included once, we reproduce the sum over all possible N and E with $\sum n_i = N$ and $\sum n_i \varepsilon_i = E$. For example, if $n_i = 0, 1, 2, \cdots$ are allowed (as in a Bose system), then Eq. (15) means (with $\beta = \mu = 1$ for simplicity):

$$\mathcal{Z} = 1 + e^{\mu - \varepsilon_1} + e^{2\mu - 2\varepsilon_1} + \cdots + e^{2\mu - 2\varepsilon_2} + e^{2\mu - \varepsilon_1 - \varepsilon_2} + \cdots + e^{\mu - \varepsilon_3} + \cdots + e^{9\mu - 6\varepsilon_1 - 2\varepsilon_2 - \varepsilon_3} + \cdots \tag{16}$$

$$= (1 + e^{\mu - \varepsilon_1} + e^{2\mu - 2\varepsilon_1} + \cdots)(1 + e^{\mu - \varepsilon_2} + e^{2\mu - 2\varepsilon_2} + \cdots)(1 + e^{\mu - \varepsilon_3} + e^{2\mu - 2\varepsilon_3} + \cdots)\cdots \tag{17}$$

The second line shows all terms with $N = 2$ and energies ε_1, ε_2, ε_3 and agrees with the explicit enumeration in Eq. (7). The equivalence of Eq. (15) and Eq. (14) is important. So please make sure you understand it, working out your own checks as needed.

Eq. (15) can be written more suggestively as

$$\mathcal{Z} = \prod_{\text{single particle states } i} \mathcal{Z}_i \tag{18}$$

with

$$\mathcal{Z}_i = \sum_{\text{possible occupancies } n_i} e^{-n_i \beta(\varepsilon_i - \mu)} \tag{19}$$

This factor $\mathcal{Z}_i$ is the one-state grand-canonical partition function, i.e. the grand-canonical partition function for a system with only one single-particle state, although the state can be multiply-occupied. Eq. (18) means that the full grand partition function is a product of the grand partition functions for the separate single-particle states: each degree of freedom can be excited independently. This is very powerful, since it means the probability of finding n_i particles occupying state i is completely independent of whatever else is happening. This is true at fixed μ but wouldn't be true at fixed N – try to factor Eq. (7) in this way; it doesn't work.

The grand canonical ensemble makes it very easy to write an expression for the expected occupation number of state i:

$$\langle n_i\rangle = \frac{1}{\mathcal{Z}_i}\sum_{n_i} n_i e^{-n_i\beta(\varepsilon_i-\mu)} = \frac{1}{\beta}\frac{\partial \ln \mathcal{Z}_i}{\partial\mu} \tag{20}$$

We can also express this as

$$\langle n_i\rangle = -\frac{\partial\Phi_i}{\partial\mu} \tag{21}$$

where

$$\Phi_i = -\frac{1}{\beta}\ln\mathcal{Z}_i \tag{22}$$

is the grand free energy for state i.

3.1 Non-interacting Bose gas

For a Bose system, there can be any number $n_i = 0,1,2,\cdots$ of bosons in a given state i. The single-state grand-canonical partition function for the state i with energy ε_i, from Eq. (19), is then

$$\mathcal{Z}_i = \sum_{n=0}^{\infty} e^{-n\beta(\varepsilon_i-\mu)} \tag{23}$$

This is a simple geometric series: there can be 0 particles, 1 particle, 2 particles, etc in the state, and so the total amount of energy in the state is an integer multiple of ε_i. Performing the sum, we find

$$\mathcal{Z}_i = \sum_{n=0}^{\infty} e^{-n\beta(\varepsilon_i-\mu)} = \frac{1}{1-e^{-\beta(\varepsilon_i-\mu)}} \tag{24}$$

The full partition function from Eq. (18) is then

$$\mathcal{Z} = \prod_i \mathcal{Z}_i = \prod_i \frac{1}{1-e^{-\beta(\varepsilon_i-\mu)}} \tag{25}$$

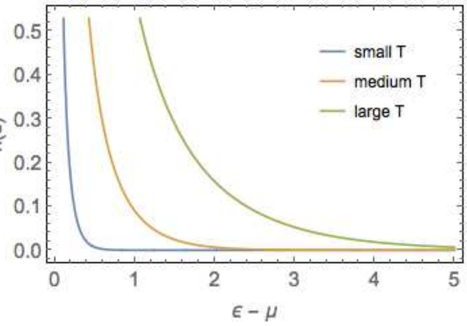
Note again that the product is over all the different states i, with ε_i the energy of that state.

The grand free energy for each state is

$$\Phi_i = -\frac{1}{\beta}\ln\mathcal{Z}_i = \frac{1}{\beta}\ln(1-e^{-\beta(\varepsilon_i-\mu)}) \tag{26}$$

Then the occupation number for each state is

$$\boxed{\langle n_i\rangle = -\frac{\partial\Phi_i}{\partial\mu} = \frac{1}{e^{\beta(\varepsilon_i-\mu)}-1}} = \tag{27}$$



This is known as the **Bose-Einstein distribution**.

Note that if $\varepsilon_i < \mu$ then $\langle n_i\rangle$ is negative, which is impossible. This means that $\mu < \varepsilon_i$ for all ε_i, i.e. the chemical potential for any Bose system will be less than all of the energies. Often we set the ground state energy to zero in which case

- $\mu < 0$ for bosons

Negative chemical potential is consistent with the formula we derived in Lecture 7 for μ for a classical monatomic ideal gas: $\mu = k_B T \ln n\lambda^3 \approx -0.39\,\mathrm{eV} < 0$ with n the number density and λ the thermal wavelength. We'll use the chemical potential for bosons in Lectures 11 and 12.

3.2 Non-interacting Fermi gas

For fermions, no two particles can occupy the same single-particle state. So the one-state grand-canonical partition function is simply

$$\mathcal{Z}_i = \sum_{n=0,1} e^{-\beta(\varepsilon_i - \mu)n} = 1 + e^{-\beta(\varepsilon_i - \mu)} \tag{28}$$

The full grand-canonical partition function is then

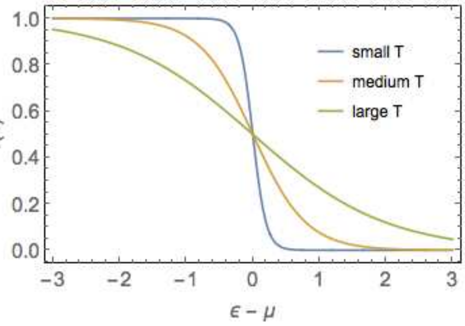
$$\mathcal{Z} = \prod_i \mathcal{Z}_i = \prod_i \, [1 + e^{-\beta(\varepsilon_i - \mu)}] \tag{29}$$

The grand free energy for state i is

$$\Phi_i = -\frac{1}{\beta} \ln \mathcal{Z}_i = -\frac{1}{\beta} \ln[1 + e^{-\beta(\varepsilon_i - \mu)}] \tag{30}$$

So that the occupation number for state i is

$$\boxed{\langle n_i \rangle = -\frac{\partial \Phi_i}{\partial \mu} = \frac{1}{e^{\beta(\varepsilon_i - \mu)} + 1} \equiv f(\varepsilon)} = \tag{31}$$



This is known as the **Fermi-Dirac distribution** or **Fermi function**, $f(\varepsilon)$. We see that at low temperature, states with energy greater than μ are essentially unoccupied: $f(\varepsilon > \mu) \approx 0$ and states with energies less than μ are completely filled.

The chemical potential at $T = 0$ is called the **Fermi energy**, $\varepsilon_F = \mu(T = 0)$. At $T = 0$, all states below the Fermi energy are filled and the ones above the Fermi energy are empty. Thus the chemical potential for Fermi systems is positive at low temperature and corresponds to the highest occupied energy level of the system. We'll use the Fermi energy extensively to discuss fermionic systems in the quantum regime, in Lectures 13, 14 and 15.

As the temperature increases, the chemical potential decreases. This follows from $\frac{\partial \mu}{\partial T} = -\frac{S}{N}$ and $S > 0$. Eventually, the chemical potential will drop below all of the energy levels of the system. That is, it becomes negative, as with a Bose system or a classical gas.

3.3 Maxwell-Boltzmann statistics

Maxwell-Boltzmann statistics are the statistics of classical indistinguishable particles. The partition function is computed by summing over states with any particles in any state, and the sum is divided by $N!$ to account for indistinguishability. Although it is often an excellent approximation, no physical system actually obeys Maxwell-Boltzmann statistics *exactly*. Nevertheless, counting states this way turns out to be very much easier than using Bose or Fermi statistics and gives the same answer in the continuum/classical limit.

For Bose or Fermi gases, we were able to write the partition function as the product of partition functions for each state, as in Eq. (18): $\mathcal{Z}=\prod_i \mathcal{Z}_i$. With Maxwell-Boltzmann statistics, it's not clear yet if we can do this since we must divide by $N!$ where N is the *total* number of particles, not just the number of particles in state i. To avoid this subtlety, let's instead compute the grand-canonical partition function directly from the canonical partition function. Indeed, we can always write

$$\mathcal{Z}=\sum_N Z_N e^{\beta N\mu} \tag{32}$$

where Z_N is the canonical partition function with N particles. We could have tried this for Bose or Fermi statistics, but for those cases the canonical partition function Z_N with N fixed is hard to evaluate.

The canonical partition function Z_N with Maxwell-Boltzmann statistics is

$$Z_N=\frac{1}{N!}\sum_{\substack{\text{distinguishable}\\ N\text{particle microstates } k}} e^{-\beta E_k} \tag{33}$$

In this sum the particles are treated as distinguishable and indistinguishability is enforced entirely through the $N!$ factor.[3]

Since each particle is independent and the total energy E is unconstrained, we can pick any state i for any particle and therefore

$$Z_N=\frac{1}{N!}\sum_{i_1}\cdots\sum_{i_N} e^{-\beta(\varepsilon_{i_1}+\cdots+\varepsilon_{i_N})} \tag{34}$$

Here the sum over i_j is over the possible states i for particle j. Since the sums are all the same, we can simplify this as in Eq. (15):

$$Z_N=\frac{1}{N!}\left(\sum_i e^{-\beta\varepsilon_i}\right)\cdots\left(\sum_i e^{-\beta\varepsilon_i}\right)=\frac{1}{N!}Z_1^N \tag{35}$$

where the single-particle canonical partition function is

$$Z_1=\sum_i e^{-\beta\varepsilon_i} \tag{36}$$

The grand partition function for Maxwell-Boltzmann statistics is then

$$\mathcal{Z}=\sum_N \frac{1}{N!}Z_1^N e^{\beta N\mu}=\exp[Z_1 e^{\beta\mu}]=\exp\left[\sum_i e^{-\beta(\varepsilon_i-\mu)}\right]=\prod_i \exp[e^{-\beta(\varepsilon_i-\mu)}] \tag{37}$$

This has the same form as Eq. (18) after all

$$\mathcal{Z}=\prod_i \mathcal{Z}_i \tag{38}$$

with

$$\mathcal{Z}_i=\exp[e^{-\beta(\varepsilon_i-\mu)}] \tag{39}$$

In fact, if we write

$$\mathcal{Z}_i=\sum_n \frac{1}{n!}e^{-n\beta(\varepsilon_i-\mu)} \tag{40}$$

We see that $\mathcal{Z}_i$ is exactly the grand partition function for a single state using Maxwell-Boltzmann statistics.

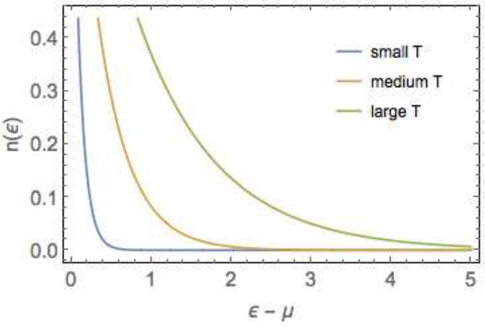
[3] Note that dividing by $N!$ is a little too much. $N!$ is the number of permutations when the particles are in different states, such as $(A,B,0)$ and $(B,A,0)$ in the table in Eq. (6). But it divides by too much when there is more than one particle in the same state, like the state $(AB,0,0)$ in the top row in the table in Eq. (6); there is only one state like this, not two. This approximation is valid when the number of accessible states is much larger than N, as at high temperature or in the classical continuum limit. In such limits, the number of configurations with more than one particle in a state is negligible.

The grand free energy for state i is then

$$\Phi_i = -\frac{1}{\beta}\ln\mathcal{Z}_i = -\frac{1}{\beta}e^{-\beta(\varepsilon_i-\mu)} \tag{41}$$

and the expected number of particles in state i is

$$\boxed{\langle n_i\rangle = -\frac{\partial\Phi_i}{\partial\mu} = e^{-\beta(\varepsilon_i-\mu)}} = \tag{42}$$



This is known as the **Maxwell-Boltzmann distribution function**. It looks a lot like the Bose-Einstein distribution function.

Unlike the Bose-Einstein case, taking $\mu \geqslant \varepsilon_i$ does not give negative occupation numbers. It can however, give occupation numbers greater than 1. When states are multiply occupied then using classical statistics is no longer justified and we must use Bose or Fermi statistics. Thus for Maxwell-Boltzmann statistics, we should have $\mu < \varepsilon_i$ for all ε_i. Or, setting the ground state energy $\varepsilon_0 = 0$ we should have $\mu < 0$. Note that this is consistent with our understanding of μ from classical statistical thermodynamics. For example, in Lecture 7 we saw that for a monatomic ideal gas with $\varepsilon_0 = 0$ then $\mu = k_B T\ln(n\lambda^3)$ with $\lambda = \frac{h}{\sqrt{2\pi m k_B T}}$ the thermal wavelength. As long as we are at low density $n < \lambda^{-3}$, then $\mu < 0$. When $n \approx \lambda^{-3}$ then quantum effects become important and the sign of μ will depend on whether the gas is bosonic or fermionic, i.e. our classical computation of the monatomic ideal gas is no longer valid. We next discuss this quantum/classical transition in more detail.

4 Quantum and classical regime

Note that all of the statistics predict that the expected number of particles in state i is

$$\langle n_i\rangle = \frac{1}{e^{\beta(\varepsilon_i-\mu)} + c} \tag{43}$$

with $c = -1, 0$ or 1 for Bose-Einstein, Maxwell-Boltzmann and Fermi-Dirac statistics, respectively. The classical limit is when the probability two particles being in the same state is irrelevant. This limit requires that the $\langle n_i\rangle$ are all small, $\langle n_i\rangle \ll 1$, which, from Eq. (43) implies

$$e^{\beta(\varepsilon_i-\mu)} \gg 1 \tag{44}$$

In this limit, all the statistics gives the same results:

$$\langle n_i\rangle \approx e^{-\beta(\varepsilon_i-\mu)} \ll 1 \tag{45}$$

Let's think a little about the why the condition $e^{\beta(\varepsilon_i-\mu)} \gg 1$ is appropriate for the classical limit. Since $\beta\to\infty$ means $T\to 0$, this seems like a low temperature limit. It certainly can be a low temperature limit: taking $k_B T \ll \varepsilon$ gaurantees the state with energy ε doesn't have even one particle in it, much less two. But the classical limit can be achieved other ways as well. The key point you have to keep in mind is that while the grand canonical ensemble works at fixed μ, our intuition is for fixed N. If N is fixed, then μ is a dependent variable and has strong temperature dependence, stronger even than β. For example, recall that for a classical ideal gas (Lecture 7, Section 7.2)

$$\mu = \varepsilon_0 + k_B T\ln\frac{N}{V} - \frac{3}{2}k_B T\ln\frac{2\pi m k_B T}{h^2} - f_v k_B T\ln\frac{T}{T_v} + \cdots \tag{46}$$

where ε_0 is ground state energy, f_v are quadratic vibrational degrees of freedom, and other possible contributions are in the $\cdots$. So $\mu\sim T\ln T$ at large T when N is fixed. Therefore, for Maxwell-Boltzmann statistics at fixed N we find for a monatomic ideal gas

$$e^{-\beta(\varepsilon_i-\mu)}=\frac{N}{V}\left(\frac{h^2}{2\pi mk_BT}\right)^{3/2}\left(\frac{T_v}{T}\right)^{f_v}e^{-\frac{\varepsilon_i}{k_BT}} \tag{47}$$

At large T and fixed N, this goes to zero, so that any individual single-particle state is unlikely to be multiply occupied. More physically, at high T, more and more states become accessible so it is less and less likely for any state to be multiply occupied. In addition, at high T particles are moving fast and their de Broglie wavelengths shrink: they become more like classical particles. So we can have $\langle n_i\rangle\ll 1$ either when $T\to 0$ at fixed μ or when $T\to\infty$ at fixed N. Another classical limit is $V\to\infty$ to make more and more states. There are many ways to get $\langle n_i\rangle\ll 1$.

Quite generally, increasing the temperature at fixed N decreases the chemical potential, since $\frac{\partial\mu}{\partial T}=-\frac{S}{N}<0$. For a Bose system, $\mu<\varepsilon_i$ for all ε_i so $\varepsilon_i-\mu>0$ and grows with temperature. In fact, it grows with temperature *faster* than β decreases with temperature: roughly $\varepsilon-\mu\sim\frac{S}{N}T\sim fT\ln T$ for some f (e.g. $S\sim Nk_B\ln T$ in an ideal gas) so $\beta(\varepsilon-\mu)\sim f\ln T$ and $\langle n_i\rangle\sim\frac{1}{T^f}$ at large T. Thus we see more broadly that high temperature gives classical statistics, as you would expect, despite the fact that a superficial reading of $e^{-\beta(\varepsilon_i-\mu)}\ll 1$ makes it seem otherwise.

I suspect a number of you are cursing the chemical potential at this point. Why don't we just use the microcanonical or canonical ensembles, at fixed N, rather than the grand canonical ensemble with its unintuitive μ? As I have been saying, it is extremely difficult to use the microcanonical or canonical ensembles for Bose-Einstein of Fermi-Dirac statistics due to the challenging combinatorics of counting microstates at fixed N. As we will see, having μ around is not bad at all once you get used to it. μ is essential to understanding Bose-Einstein condensation (Lecture 12) and metals (Lectures 13 and 14).

We have seen that many limits (large T, small T, large V) lead to classical statistics. The challenge is to find a regime where the quantum statistics dominate. For quantum statistics to be relavant, we need $\langle n_i\rangle\not\ll 1$. Since the higher energy states are going to be less populated than the ground state, a reasonable condition is that $\langle n_0\rangle\approx 1$. Using Eq. (47), the ground state with $\varepsilon_i=0$ has $\langle n_0\rangle=1$ when

$$\frac{N}{V}=\left(\frac{2\pi mk_BT}{h^2}\right)^{3/2}=\frac{1}{\lambda^3} \tag{48}$$

So we want $\lambda\approx\left(\frac{V}{N}\right)^{1/3}$. In other words

- **quantum statistics are important when the thermal de Broglie wavelength is of order or bigger than the interparticle spacing**

For example, at room temperature $\lambda_{\text{air}}\approx 1.87\times 10^{-11}m$ while the average interparticle spacing in air is $\left(\frac{V}{N}\right)^{1/3}=\left(\frac{k_BT}{P}\right)^{1/3}=3\times 10^{-9}$, so particles in air are around a factor of 100 times too far for quantum effects to be important. If we keep the pressure at 1 atm, we would have to cool air to $T=0.57K$ for quantum effects to matter.

For a classical monatomic ideal gas $\mu=k_BT\ln(\lambda^3n)$. Thus in the classical regime when $\lambda^3\frac{N}{V}\ll 1$ the chemical potential is a large negative number, $\mu\ll 0$. As temperature is decreased at constant N and V then λ goes up and λ^3n increases. As λ^3n gets close to 1 then $\mu\to 0$ from below (more precisely, $\mu\to\varepsilon_0$ with ε_0 the ground state energy, but we usually set $\varepsilon_0=0$). As $\mu\to 0$, whether the system is bosonic or fermionic becomes important. For a bosonic system, μ can get closer and closer to 0, but can never reach it. As we will see in Lecture 12, as $\mu\to 0$, bosons accumulate in the ground state, a process called Bose-Einstein condensation. For a fermionic system, μ goes right through zero and ends up at a positive value the Fermi energy $\mu=\varepsilon_F>0$ at $T=0$. Thus the closeness of μ to 0 or equivalently the closeness of λ^3n to 1 indicates the onset of quantum statistics.

5 Summary

In this lecture we have seen how statistical mechanics is modified for systems of identical particles. There are three types of statistics we use

• Fermi-Dirac statistics applies to particles of half-integer spin like electrons in metals or white-dwarf stars or half-integer spin nuclei like protons or neutrons. Fermi-Dirac statistics applies not just to elementary particles, but also to atoms that have an odd number of fermions, such as ^{3}He, which comprises 2 protons, 1 neutron and 2 electrons. In Fermi-Dirac statistics, no two identical particles can occupy the same single-particle state.

• Bose-Einstein statistics applies to particles of integer spin, like photons, phonons, vibrational modes, ^{4}He atoms (=2 neutrons, 2 protons and 2 electron) or ^{95}Rb atoms (35 protons, 60 neutrons, 35 electrons). In Bose-Einstein statistics, any number of bosons can occupy the same state.

• Maxwell-Boltzmann statistics are a kind of phony classical statistics for which calculations are generally easier than with bosons or fermions. We treat the particles as all distinguishable, then throw in a factor of $\frac{1}{N!}$ to the partition function to account for indistinguishability.

The main results of this lecture were the expressions for the partition functions and probability distributions for the various statistics. We observed that for ideal gases (non-interacting particles) with any statistics, the grand-canonical partition function can be written as

$$\mathcal{Z} = \prod_{\text{single particle states } i} \mathcal{Z}_i \tag{49}$$

where $\mathcal{Z}_i$ is the grand-canonical partition function for a single state. For the various statistics we found

$$\mathcal{Z}_i = \frac{1}{1 - e^{-\beta(\varepsilon_i - \mu)}} \qquad (\text{Bose} - \text{Einstein}) \tag{50}$$

$$\mathcal{Z}_i = 1 + e^{-\beta(\varepsilon_i - \mu)} \qquad (\text{Fermi} - \text{Dirac}) \tag{51}$$

$$\mathcal{Z}_i = \exp[e^{-\beta(\varepsilon_i - \mu)}] \quad (\text{Maxwell} - \text{Boltzmann}) \tag{52}$$

From these, we deduce the probabilities $\langle n_i \rangle$ for finding n_i particles in a given state i with energy ε_i. The general formula is $\langle n_i \rangle = \frac{\partial}{\partial \mu}\left(\frac{1}{\beta}\ln \mathcal{Z}_i\right)$. We found

$$\langle n_i \rangle = \frac{1}{e^{\beta(\varepsilon_i - \mu)} - 1} \qquad (\text{Bose} - \text{Einstein}) \tag{53}$$

$$\langle n_i \rangle = \frac{1}{e^{\beta(\varepsilon_i - \mu)} + 1} \qquad (\text{Fermi} - \text{Dirac}) \tag{54}$$

$$\langle n_i \rangle = e^{-\beta(\varepsilon_i - \mu)} \quad (\text{Maxwell} - \text{Boltzmann}) \tag{55}$$

The general rule of thumb is that quantum statistics are relevant when states have a decent chance of being multiply occupied. At room temperature, there are so many possible states for the momenta and position of the molecules, that the chance of two being in the same state is utterly negligible. That's why classical statistical mechanics works fine most of the time. The way to find out if quantum statistics are important is to look at the thermal de Broglie wavelength

$$\lambda = \sqrt{\frac{h^2}{2\pi m k_B T}} \tag{56}$$

When the interparticle spacing $(V/N)^{1/3}$ is smaller than λ then quantum statistics is important.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 11: Phonons and Photons

1 Introduction

In this lecture, we will discuss some applications of quantum statistical mechanics to Bose systems.

To review, we found in the last lecture that the grand partition function for Bose-Einstein particles can be written as

$$\mathcal{Z} = \prod_{\text{single particle states } i} \mathcal{Z}_{\varepsilon_i} \tag{1}$$

where the product is over possible states i for a single particle and

$$\mathcal{Z}_\varepsilon = \frac{1}{1 - e^{-\beta(\varepsilon - \mu)}} \tag{2}$$

is the grand-canonical partition function for a single state with energy ε (computed by summing over all the possible numbers n of bosons that could simultaneously be in that state). Using the grand canonical ensemble, we found that the expected number of particles in a given mode is determined by the Bose-Einstein distribution function

$$\langle n_i \rangle = \frac{1}{e^{\beta(\varepsilon_i - \mu)} - 1} \tag{3}$$

Here μ is the chemical potential, which (for bosons) has to be lower than all the energy levels ε_i in the system.

In this lecture, we will study two important cases of Bose systems where quantum effects are important: the Debye theory of solids, where the bosons are the excitations of vibrational modes known as **phonons**, and the theory of blackbody radiation, where the bosons are quanta of light known as **photons**. Both types of bosons have $\mu = 0$, since there is no conserved particle number. In the next lecture we will study Bose-Einstein condensation, where the bosons are atoms and $\mu \neq 0$.

In order to study any of these systems, we want to perform the product over states i in Eq. (1). Generally, we do this by first taking the logarithm (to get the grand free energy Φ), so that the product becomes a sum, then performing the sum as an integral over energy

$$\Phi = -\frac{1}{\beta} \ln \mathcal{Z} = -\frac{1}{\beta} \sum_i \ln \mathcal{Z}_{\varepsilon_i} = -\frac{1}{\beta} \int d\varepsilon\, g(\varepsilon) \ln \mathcal{Z}_\varepsilon \tag{4}$$

where $g(\varepsilon)$ is the measure on the integral over energies ε, known as the **density of states**. Thus, the first step to studying any of these systems is working out the density of states, for which we need the distribution of energy levels of the system. Many quantities can be derived from Φ recalling that

$$d\Phi = S dT + P dV + N d\mu \tag{5}$$

and that $\Phi = -PV$ (we showed this at the end of Lecture 8).

2 Debye model of solids

Our first example is the theory of phonons, or vibrations of a lattice of atoms in a solid. This will let us compute the heat capacity of a solid and other useful thermodynamic quantities.

2.1 Classical model of solids

The model of a solid we will use is called a **harmonic solid**. It is envisioned as a lattice of nuclei connected by covalent bonds which act as springs

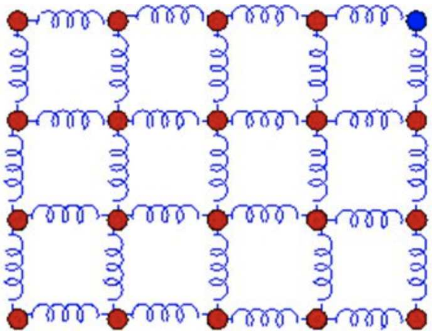


Figure 1. An harmonic solid treats atoms as connected by springs.

If you knock one side of the solid, a sound wave will propagate through by vibrating each successive molecule. The quanta of sound-wave excitations in solids are called phonons.

The solid is made up of N atoms, each of which can move in 3 directions, so there are $3N$ springs (up to boundary conditions, which will not matter for N large). There will therefore be $3N$ normal modes of vibration. If $\vec{A}(\vec{x},t)$ describes the displacement from equilibrium of the atom at position $\vec{x}$ at time t it will satisfy the wave equation

$$[\partial_t^2 - c_s^2\vec{\nabla}^2]\vec{A}(\vec{x},t) = 0 \tag{6}$$

where c_s is the speed of sound in the solid. (Recall from 15c that this comes from the force on atom at position n coming from the difference of forces form the two sides $m\,\ddot{x}_n = k(x_{n+1} - x_n) - k(x_n - x_{n-1}) \sim k\partial_n^2 x_n \sim a^2 k\nabla^2 x_n$ with a the lattice spacing.)

In general, the speed of sound is $c_s = \sqrt{\frac{dP}{d\rho}}$ (you should have derived this result in 15c). For solids, the density is $\rho = m\frac{N}{V}$, so $\frac{dP}{d\rho}$ is determined by the relation between pressure and volume. We represent this with the bulk modules $B = -V\left(\frac{\partial P}{\partial V}\right)_{T,N}$, so that $c_s = \sqrt{\frac{dP}{d\rho}} = \sqrt{\frac{dP}{dV}\frac{dV}{d\rho}} = \sqrt{\frac{B}{\rho}}$. The bulk modulus is the restoring force, like k for a spring: $c_s = \sqrt{\frac{B}{\rho}} \sim \sqrt{\frac{k}{m}}$. This restoring force is determined not by the the things vibrating to transmit the wave (the atoms) but by the things generating the force (the electrons). To compute c_s we need to compute B which necessarily involves the electrons. We will compute B for solids in Lecture 13. To study the vibrations of a solid we can simply treat c_s as a constant.

Given the wave equation and c_s, normal mode oscillations of the system have the form

$$\vec{A}(\vec{x},t) = \vec{A}_i \cos(\vec{k}_i \cdot \vec{x})\cos(\omega_i t) \tag{7}$$

for constant wavevectors $\vec{k}_i$ and angular frequencies ω_i. Plugging $\vec{A}(\vec{x},t)$ in to the wave equation implies $\vec{k}_i$ and ω_i are related by

$$\omega_i = c_s|\vec{k}_i| \tag{8}$$

Each of the normal modes is a standing plane wave whose displacement depends only on $\vec{k}_i \cdot \vec{x}$, i.e. it is constant in the plane normal to $\vec{k}_i$. Displacement can be in any of 3 directions, so $\vec{A}$ is a vector. The prefactor $\vec{A}_i$ in Eq. (7) is the polarization vector of the sound wave. For example, sound waves can be longitudinal (with $\vec{A}_i \propto \vec{k}_i$), like sound waves in air. They can also be transverse (with $\vec{A}_i \cdot \vec{k}_i = 0$), like water waves or excitations of a string, in either of the 2 transverse directions. To visualize the three polarizations think of moving a square grid, like a window screen, in any of 3 dimensions.

The allowed wavevectors are classically quantized. The atoms on the edges have forces only one one side, so they must have open-string boundary conditions $\partial_x A(\vec{x},t) = 0$. In one dimension, we would impose that $\partial_x A(0,t) = \partial_x A(L,0) = 0$ which gives $k = \frac{\pi n}{L}$ for $n = 0,1,2,3,\cdots$. In 3 dimensions the condition is

$$\vec{k}_n = \frac{\pi}{L}\vec{n} \tag{9}$$

with $\vec{n}$ a vector of whole numbers (e.g. $\vec{n} = (3, 0, 2)$). Then the frequencies are

$$\omega_n = \frac{\pi}{L} c_s n, \quad n \equiv |\vec{n}| \tag{10}$$

If the solid has some other crystal structure, the normal modes will be different, but we focus on this regular cubic lattice for simplicity. By the way, the boundary conditions aren't actually important for the bulk properties.[1]

So far, this has all been classical – 15c material. Springs and balls propagate waves just like we have described, with the wavenumbers in Eq. (9). The amplitude $|\vec{A}_i|$ can take any real value – sound waves can have arbitrarily small amplitudes.

2.2 Phonons

To transition to quantum mechanics, we associate energy to each wave proportional to its frequency. A natural way to do this is to treat each mode as an independent simple harmonic oscillator. This is called the **Debye model**. Each oscillator indexed by $\vec{n}$ can be excited m times, so the energy of that oscillator is

$$\varepsilon_{\vec{n}}(m) = \hbar\omega_n \left(m + \frac{1}{2} \right) \tag{11}$$

Again, $\vec{n}$ tells us which mode and m tells us which excitation of that mode. There are 3 possible excitations for any $\vec{n}$, corresponding to the 3 polarizations of the sound waves (as with the classical theory, $\vec{n}$ specifies a plane of atoms that vibrates and the 3 polarizations tell us which direction it is vibrating in).

Having no excitations at all corresponds to $m = 0$ for all $\vec{n}$ and the energy is $E_0 = \sum_{\vec{n}} \frac{1}{2}\hbar\omega_n$. Since this ground state energy can never change, thermodynamic properties of the system will not depend on it (only on energy differences), so we will set $E_0 = 0$ for convenience.

The labels $\vec{n}$ for the normal modes are the states labeled i in our general discussion of Bose systems. The excitations m correspond to m **phonons**. The classical amplitude $\vec{A}_i$ in Eq. (7) is proportional to the number of phonons in that mode, with $\vec{n}$ determining the wavevector $\vec{k}$ according to Eq. (9).

What is the chemical potential for phonons? When you heat a solid for example by putting it in air, lots of phonons are produced from simply banging air molecules into the solid. That is, processes like air → air + phonon can happen. Since the sum of the chemical potentials on both sides of any reaction are the same in equilibrium, this implies that $\mu = 0$ for phonons. There is no conservation of phonon number.

Using $\vec{n}$ for i in Eq. (2) we want to compute the grand free energy, proportional to

$$\sum_i \ln \mathcal{Z}_i = 3 \sum_{\vec{n}} \ln \frac{1}{1 - e^{-\beta\varepsilon_{\vec{n}}}} \tag{12}$$

The factor of 3 comes from the 3 polarizations.

Now for a large number N of atoms, the modes will be closely spaced and effectively continuous. If the integers $\vec{n}$ are very large, we can perform the sum as an integral. In general, we are interested in sums $\sum_{\vec{n}} F(n)$ for various functions $F(n)$. For example, $F(n) = 1$ will give us $3\sum_{\vec{n}} 1 = N_{\text{modes}}$, the number of different single-particle states. $F(n) = \ln \frac{1}{1 - e^{-\beta\varepsilon_{\vec{n}}}}$ gives the partition function, and so on. Denoting the highest possible $\vec{n}$ of the 3D normal modes by $\vec{n}_{\max}$ we have

$$3 \sum_{\vec{n}} F(n) = 3 \int_{\vec{n} > 0}^{\vec{n}_{\max}} d^3\vec{n} F(n) = \frac{3}{8} \int_{-\vec{n}_{\max}}^{\vec{n}_{\max}} d^3\vec{n} F(n) = \frac{3}{8} \int_0^{n_D} 4\pi n^2 dn F(n) \tag{13}$$

[1] A common alternative choice is periodic boundary conditions $A(0, t) = A(L, t)$ which gives $k = 2\pi \frac{n}{L}$ for the $\sin(kx)$ modes or $\cos(kx)$ modes. Equivalently, we write the modes as exponentials $A(x, t) = A_0 \exp(ik \cdot x - i\omega t)$ so that $k = 2\pi \frac{n}{L}$ with $n = 0, \pm 1, \pm 2, \cdots$. The factor of 2π instead of π and having twice as many modes cancel to give the same statistical mechanical results at large L.

The factor of 8 comes from the restriction that the components of $\vec{n}$ are all whole numbers, so the vector lies in one octant 3D space; since the integrand will only depend on $n = |\vec{n}|$ it is convenient to integrate over the whole sphere in spherical coordinates with a factor of $\frac{1}{8}$. The upper limit on integration over n we call $n_D = |\vec{n}_{\max}|$. n_D is finite because there are a finite number of normal modes ($3N$) in a solid. The relation between n_D and $3N$ is easiest to determine by simply performing the integral in Eq. (13) and setting it equal to $3N$. Doing so gives $n_D = \left(\frac{6}{\pi}N\right)^{1/3}$. However, we won't use this result for n_D. It is more convenient to change variables first from n to frequency and then to fix the upper limit of integration in frequency space.

From Eq. (10) the frequency is related to n as $\omega_n = \frac{\pi}{L}c_s n$ so we can write

$$3\sum_{\vec{n}} F(n) = \frac{3}{8}\int_0^{n_D} 4\pi n^2 dn F(n) = \frac{3}{8}\frac{V}{c_s^3\pi^3}4\pi\int_0^{\omega_D}\omega^2 d\omega F\left(\frac{L\omega}{\pi c_s}\right) = \int_0^{\omega_D} g(\omega) d\omega F\left(\frac{L\omega}{\pi c_s}\right) \tag{14}$$

where the **density of states** is

$$\boxed{g(\omega) = \frac{3V}{2\pi^2 c_s^3}\omega^2} \tag{15}$$

The density of states gives the integration measure for integrating over energies: the number of states between ω and $\omega + d\omega$ is $g(\omega)d\omega$.

The upper limit ω_D on the frequency integral is called the **Debye frequency**. It is finite since there are a finite number of atoms hence a finite number of normal modes. To find the relation between N and ω_D, we perform the sum/integral with $F(n) = 1$ to count the states:

$$3N = 3\sum_{\vec{n}} = \frac{3V}{2\pi^2 c_s^3}\int_0^{\omega_D}\omega^2\, d\omega = \frac{V}{2\pi^2 c_s^3}\omega_D^3 \tag{16}$$

so that

$$\omega_D = c_s\left(6\pi^2\frac{N}{V}\right)^{1/3} \tag{17}$$

Using this relation, the density of states can be written in a simpler form:

$$g(\omega) = \frac{9N}{\omega_D^3}\omega^2 \tag{18}$$

Then the grand-canonical partition function becomes

$$\ln\mathcal{Z} = \int g(\varepsilon) d\varepsilon \ln\mathcal{Z}_\varepsilon = -\frac{9N}{\omega_D^3}\int_0^{\omega_D}\omega^2\, d\omega \ln(1 - e^{-\beta\hbar\omega}) \tag{19}$$

Mathematica can do this integral in terms of polylogarithms, but the final form is not illuminating, so we'll leave it as an integral.

The energy in the Debye model is computed as

$$E = -\frac{\partial}{\partial\beta}\ln\mathcal{Z} = \frac{9N\hbar}{\omega_D^3}\int_0^{\omega_D}\frac{\omega^3}{e^{\beta\hbar\omega} - 1}d\omega \tag{20}$$

Again, this integral can be evaluated in an ugly closed form. We get nicer expressions by expanding, which we can do before evaluating the integral. For large T (small β),

$$E\left(T \gg \frac{\hbar\omega_D}{k_B}\right) \approx \frac{9N\hbar}{\omega_D^3}\int_0^{\omega_D}\frac{\omega^3}{\beta\hbar\omega}d\omega = 3Nk_BT + \cdots \tag{21}$$

The leading behavior is consistent with the classical equipartition theorem: each of the $3N$ modes gets $\frac{1}{2}k_BT$ of potential energy and $\frac{1}{2}k_BT$ of kinetic energy. For small T, we find

$$E\left(T \ll \frac{\hbar\omega_D}{k_B}\right) \approx \frac{9N\hbar}{\omega_D^3}\int_0^{\omega_D}\frac{\omega^3}{e^{\beta\hbar\omega}}d\omega = \frac{3\pi^4}{5}N\hbar\omega_D\left(\frac{T}{T_D}\right)^4 + \cdots \tag{22}$$

where

$$T_D = \frac{\hbar\omega_D}{k_B} = \frac{\hbar}{k_B}c_s\left(6\pi^2\frac{N}{V}\right)^{1/3} \tag{23}$$

is the **Debye temperature**. The Debye temperature is the temperature corresponding to the highest energy phonon mode. When $T = T_D$, the Boltzmann factor $e^{-\frac{\hbar\omega_D}{k_B T}} \approx 1$, and none of the phonon modes are exponentially suppressed. Thus T_D can be thought of as the temperature above which all of the phonon modes are excited. Typical Debye temperatures for metals are in the $T_D \sim 100\,K - 500\,K$ range.

The heat capacity is for low temperature

$$C_V(T \ll T_D) = \frac{\partial E}{\partial T} = \frac{12\pi^4}{5} N k_B \left(\frac{T}{T_D}\right)^3 \tag{24}$$

The scaling of the heat capacity as T^3 is a testable prediction of the Debye model. At high temperature

$$C_V(T \gg T_D) = \frac{\partial E}{\partial T} = 3 N k_B \tag{25}$$

consistent with the law of Dulong and Petit. Roughly speaking, the Debye temperature is the temperature above which the solid can be treated classically, using the equipartition theorem, and below which quantum statistical mechanics is important.

Here is a comparison of the heat capacity of various metals to the prediction from the Debye model:

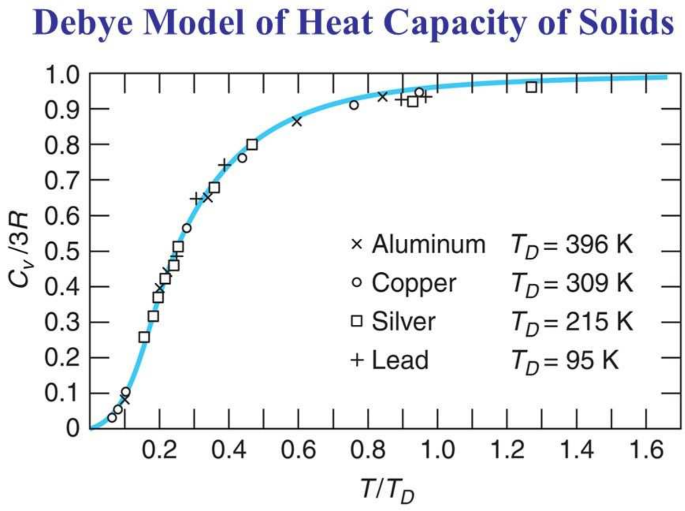


Figure 2. Comparison of the specific heat of various metals to the prediction of the Debye model (solid curve) and the law of Dulong and Petit (horizontal line at $\frac{C_V}{3R} = 1$).

We see that the experimental data fits very well to the Debye model: the T^3 behavior and constant high T behavior are evident. The Debye temperature in the figure comes from fits to the curve and tends to be lower than that predicted by Eq. (23). For example, in aluminum, $c_S = 5100\frac{m}{s}$ and Eq. (23) gives $T_D = 591\,K$ while the fit gives $396\,K$. The reason for the discrepancy is that the Debye model is only a model – it ignores dispersion in the sound waves (really $\omega(k) \sim \sin^2\!\left(k\frac{L}{N}\right)$ for a regular lattice) and it treats all the bond strengths as the same, which isn't really true. Nevertheless, the Debye model gives an excellent prediction for the shape of the heat capacity, with only one free parameter T_D. It works so well because the shape is essentially determined by the boson distribution function, which is correct even when more details of the solid are incorporated.

Note that the heat capacity in the Debye model goes to zero as $T \to 0$. This is a requirement of the 3rd law of thermodynamics ($S \to$ finite as $T \to 0$). To see that, consider a system with entropy S_1 at a temperature T_1 and cool it to $T = 0$ at constant volume. By Clausius' formula the entropy can be computed as

$$S_1 = \int \frac{dQ}{T} = \int_0^{T_1} \frac{1}{T}\frac{dQ}{dT} dT = \int_0^{T_1} C_V \frac{dT}{T} \tag{26}$$

where $C_V = \left(\frac{\partial Q}{\partial T}\right)_V$ was used. For this integral to be finite as $T_1 \to 0$ we must have that $C_V \to 0$ as $T \to 0$. Thus the small T behavior of C_V that we found in the Debye model is consistent with the 3rd law of thermodynamics (C_V=constant, the classical prediction, would not be).

3 Blackbody radiation

A blackbody is an object that can absorb and emit electromagnetic waves of any frequency. Recall that radiation is produced from accelerating charges. A charge oscillating at a frequency ω will emit electromagnetic radiation; conversely light of frequency ω can induce oscillation of charged particles. So to make a blackbody, all we need is some free charge carriers. That's pretty common. Any material when heated will have molecules moving, banging into each other, and vibrating with various energies up to $k_B T$. Naturally, these vibrating molecules will produce and absorb electromagnetic waves. When studying blackbody radiation, we don't care so much about the mechanism by which radiation can thermalize with a material body. We are instead interested in the properties of the equilibrium system.

We will begin with the thermodynamics of classical electromagnetic radiation, where we will encounter the ultraviolet catastrophe. Quantum statistical mechanics averts this catastrophe and makes quantitative and easily verifiable predictions about blackbody radiation.

3.1 Classical blackbodies

Let's start with classical electromagnetism. The energy density in an electromagnetic field is

$$\frac{E}{V} = \frac{1}{2}\varepsilon_0 \vec{E}^2 + \frac{1}{2}\frac{1}{\mu_0}\vec{B}^2 \tag{27}$$

Electromagnetic waves also exert radiation pressure

$$\vec{P} = \frac{1}{c}\vec{S} = \varepsilon_0 \vec{E} \times \vec{B} \tag{28}$$

where $\vec{S}$ is the Poynting vector that gives the energy flux (power/area). Radiation pressure is a vector. It gives the momentum per unit area of the classical radiation. Unlike pressure due to thermal motion of a gas, radiation pressure not be uniform in all directions (the radiation pressure from the sun comes only in the direction away from the sun). If we confine radiation to a box with reflecting walls, then when the radiation bounces off a wall, it its momentum/area changes by $-2\vec{P}$. If we allow for uniform radiation going towards or away from the wall in the x, y or z direction, it has a $\frac{1}{6}$ chance of going towards the wall, so the pressure from electromagnetic radiation is then is then $P = \frac{1}{3}|\vec{P}| = \frac{1}{3c}|\vec{S}|$.

In the absence of charged particles, light is described by electromagnetic fields $\vec{E}$ and $\vec{B}$ satisfying Maxwell's equations with no sources:

$$\vec{\nabla} \cdot \vec{E} = 0 \qquad \vec{\nabla} \times \vec{E} = -\frac{\partial}{\partial t}\vec{B} \qquad \vec{\nabla} \cdot \vec{B} = 0 \qquad \vec{\nabla} \times \vec{B} = \mu_0 \epsilon_0 \frac{\partial}{\partial t}\vec{E} \tag{29}$$

These combine to produce wave equations for $\vec{E}$ and $\vec{B}$

$$\left[\frac{\partial^2}{\partial t^2} - c^2 \vec{\nabla}^2\right]\vec{E}(\vec{x}, t) = 0, \qquad \left[\frac{\partial^2}{\partial t^2} - c^2 \vec{\nabla}^2\right]\vec{B}(\vec{x}, t) = 0 \tag{30}$$

with $c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$ the speed of light. Solutions to these equations are electromagnetic waves. A useful basis of electromagnetic waves are linearly-polarized standing waves

$$\vec{E}(\vec{x}, t) = \vec{E}_0 \cos(\vec{k} \cdot \vec{x})\cos(\omega t) \tag{31}$$

$$\vec{B}(\vec{x}, t) = \vec{B}_0 \cos(\vec{k} \cdot \vec{x})\cos(\omega t) \tag{32}$$

Plugging into Maxwell's equations (or the wave equations), we find a number of constraints

$$\omega = c|\vec{k}|, \qquad \vec{k} \cdot \vec{E}_0 = 0, \qquad \vec{k} \cdot \vec{B}_0 = 0, \qquad \omega \vec{B}_0 = \vec{k} \times \vec{E}_0, \tag{33}$$

For any vector $\vec{k}$ there are two independent vectors orthogonal to it, thus the second constraint, $\vec{k}\cdot\vec{E}_0=0$ there are two choices for $\vec{E}_0$ for a given $\vec{k}$ (the two polarizations). The third, longitudinal, polarization that was present in sound waves (phonons) is absent for light (the only solution with $\vec{E}_0\propto\vec{k}$ is $\vec{E}_0=0$). The 4th constraint implies that $\vec{B}_0$ is completely fixed by $\vec{k}$ and $\vec{E}_0$. In particular,

$$|\vec{E}|=|\vec{E}_0|=c|\vec{B}_0|=c|\vec{B}| \tag{34}$$

So the energy density for a plane wave is from Eq. (27)

$$\frac{E}{V}=\varepsilon_0|\vec{E}|^2 \tag{35}$$

and the radiation pressure is, from Eq. (28)

$$\vec{P}=\varepsilon_0|\vec{E}|^2\hat{\vec{k}} \tag{36}$$

where $\hat{\vec{k}}$ is a vector in the $\vec{k}$ direction with unit magnitude, $|\vec{k}|=1$. If this is all unfamiliar to you, see my 15c lecture notes, on the canvas site. In particular, Lectures 13 and 14.

Now say we have a box of light which is an incoherent sum of plane waves of all different directions and magnitudes. Since the pressure is $P=\frac{1}{3}|\vec{P}|=\frac{\varepsilon_0}{3}|\vec{E}|^2$ we then hvae

$$E=3PV \tag{37}$$

This is the equation of state for electromagnetic radiation. Our derivation was purely classical, and does not depend on thinking of electromagnetic waves has having quantized energies or being made of particles called photons.

Next, we want to compute how the energy density $\frac{E}{V}=3P$ depends on temperature. We can do this using basic thermodynamics. We start with the free energy, which satisfies the general Maxwell-Relation

$$\left(\frac{\partial F}{\partial V}\right)_T=-P \tag{38}$$

For an ideal gas $P(V,N,T)$ can depend on volume as $P=\frac{N}{V}k_BT$. For classical electromagnetic radiation there is no concept of N, so we have simply $P=P(V,T)$. For the pressure to be intensive it cannot depend on V and therefore P only depends on T. We can then trivially integrate Eq. (38) to get $F=-PV$. The equation of state Eq. (37) then implies the relation

$$F=-PV=-\frac{E}{3} \tag{39}$$

So for classical electromagnetic radiation, the free energy is negative and directly proportional to the total energy. Since $F=E-TS$ and $S=-\left(\frac{\partial F}{\partial T}\right)_V$ we then have

$$F=-3F+T\left(\frac{\partial F}{\partial T}\right)_V \tag{40}$$

This is a differential equation whose solution is $F\propto T^4$. Then $E\propto T^4$ as well, and since E is extensive ($E=3PV$ with P independent of V), we know $E=\text{const}\times VT^4$. We write this more conventionally as

$$\boxed{\frac{E}{V}=4\sigma\times\frac{1}{c}T^4} \tag{41}$$

for some constant 4σ (the 4 is a convention, chosen so that Eq. (62) below is simple). That the energy density is proportional to T^4 is known as **Stefan's law**. Thermodynamics alone does not let us compute the proportionality constant σ.

The Gibbs free energy is

$$G=F+PV=(-PV)+PV=0 \tag{42}$$

Since $G=\mu N$ this says that the chemical potential for light is zero

$$\mu=0 \tag{43}$$

We can also understand this as we did for phonons: since radiation can be produced by vibrating atoms, atom → atom + radiation is allowed, and, since the sum of chemical potentials on the two sides of a reaction are equal, we conclude that the the chemical potential for electromagnetic radiation must vanish.

What about the frequency dependence of the light? Our classical model of light in a box just had a bunch of different frequency plane waves bouncing around incoherently. We did not attempt to apportion energy into the different wavenumbers. How could we do so? The only classical handle we have is the equipartition function – we allocate $\frac{1}{2}k_B T$ of energy for each $\vec{k}$ and each polarization. Since $\omega = c|\vec{k}|$ and the volume of wavevectors goes like $d^3k \sim k^2 dk \sim \omega^2 d\omega$ we expect more and more energy at higher and higher frequencies. That is

$$I(\omega) \approx \omega^2 k_B T \tag{44}$$

This is known as the **Rayleigh-Jeans Law**. Instead, the observed energy spectrum dies off at high frequencies:

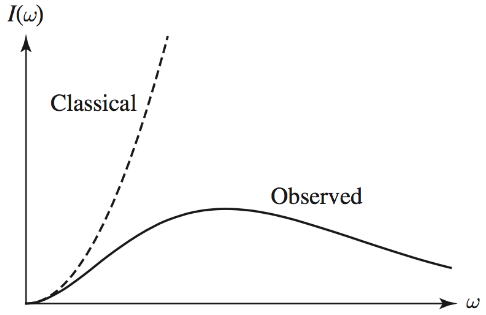


Figure 3. The classical prediction for the intensity of radiation coming from a blackbody disagrees with experimental observation at frequencies $\omega \gtrsim \frac{k_B T}{\hbar}$.

This is known as the **ultraviolet catastrophe.** This catastrophe of 19th century classical physics was resolved in 1900 by Max Planck who first quantized the energy in the different modes.

3.2 Quantum theory

To compute the thermodynamic properties of blackbody radiation using quantum statistical mechanics, we first need to work out the energy spectrum and the density of states. The quanta of light are called photons. The amplitude of an electromagnetic wave with wavevector $\vec{k}$ is proportional to the number of photons with that wavenumber (analogously to phonons). Since we can have varying amplitudes at each $\vec{k}$, states can be multiply occupied so photons must be bosons. This also follows from quantum electrodynamics (QED), where photons are seen to have spin 1 and therefore are bosons by the spin-statistics theorem.

To model a blackbody, let's treat it as a cavity with walls of L. It doesn't really matter if the blackbody is a cube or round or what the boundary conditions are. To see this, note that the boundary conditions only affect the longest-wavelength fluctuations – the ones that can reach the sides. The higher frequency/smaller wavelength modes that we care most about (e.g. for the ultraviolet catastrophe) are determined by the dynamics of the interior of the body.

For concreteness, let's take Neumann boundary conditions on the edge of a 3D box, i.e. force the spatial derivatives of the wavefunctions to vanish there. Then the independent modes of the electric field have the form

$$\vec{E}(x,y,z) = \vec{E}_0 \cos\left(\frac{\pi n_1}{L}x\right)\cos\left(\frac{\pi n_2}{L}y\right)\cos\left(\frac{\pi n_3}{L}z\right) \tag{45}$$

with n_1, n_2 and n_3 taking whole number values. The vector $\vec{E}_0$ in front is the polarization, as in the classical theory. Thus the normal modes in the box are described by wavevectors

$$\vec{k} = \left(\frac{\pi n_1}{L}, \frac{\pi n_2}{L}, \frac{\pi n_3}{L}\right) = \frac{\pi}{L}\vec{n} \tag{46}$$

with $\vec{n}=(n_1,n_2,n_3)$ a vector of whole numbers. The frequency is $\omega_k=c|\vec{k}|$ and so the energy is

$$\varepsilon_{\vec{n}}=\hbar\omega_k=\hbar c\,|\vec{k}|=\frac{\pi}{L}\hbar c n \tag{47}$$

with $n=|\vec{n}|$. The mode with $\vec{n}=(0,0,0)$ is zero, so it cannot store energy. All the physical modes have at least one component of $\vec{n}$ nonzero.

Note that the allowed frequencies in a 1D box are quantized *classically*, $\omega=\frac{\pi}{L}cn$, $n=0,1,2,3\cdots$, just like in a solid. In a solid there are 3N normal modes and an upper limit on ω $(\omega\leqslant\omega_D)$ but in a blackbody, there is no upper bound. Quantum mechanics doesn't quantize the system, it just lets us convert the frequencies into energies $\varepsilon=\hbar\omega$ by introducing Planck's constant.

To get a feel for the numbers involved, first consider the number of photons we might expect in a box at given temperature. The total flux of energy from the sun hitting the earth is $\Phi=1.4\frac{\text{kW}}{m^2}$ (see Eq. (67) below) peaked in visible (optical frequency) light. If we make a m^3 box of sunlight, we would have $\Phi\frac{1}{c}m^3=5\times10^{-6}J$ of energy. A typical optical photon has energy $\varepsilon=\frac{\pi}{500\text{nm}}\hbar c$ so there are around $N=10^{13}$ visible photons in a cubic meter of sunlight. We will compute N more precisely below, but this estimate at least gives a sense of the numbers involved: typical photon number densities are high, but not as number densities of atoms in air $(\sim N_A\sim10^{24}/m^3)$.

3.3 Equation of state

Using Eq. (47), $\varepsilon_{\vec{n}}=\frac{\pi}{L}\hbar c|\vec{n}|$, the grand-canonical partition function for the photon gas is

$$\ln\mathcal{Z}=2\sum_{\vec{n}}\ln\mathcal{Z}_n=2\sum_{\vec{n}}\ln\frac{1}{1-e^{-\beta\varepsilon_n}} \tag{48}$$

The average energy is computed from the grand partition function

$$\langle E\rangle=-\frac{\partial\ln\mathcal{Z}}{\partial\beta}=\sum_{\vec{n}}\varepsilon_n\frac{2}{e^{\beta\varepsilon_n}-1} \tag{49}$$

This matches the general formula $\langle E\rangle=2\sum_i\varepsilon_i\langle n_i\rangle$, with $\langle n_i\rangle=\frac{1}{e^{\beta\varepsilon_n}-1}$ the Bose-Einstein distribution, with an extra factor of 2 for two polarizations.

We can compute the pressure as $P=-\left(\frac{\partial\Phi}{\partial V}\right)_T$ where the grand free energy is $\Phi=-\frac{1}{\beta}\ln\mathcal{Z}$. All the volume dependence comes from $\varepsilon_n=\frac{\pi}{L}\hbar c n$ with $L=V^{1/3}$. So

$$P=\frac{2}{\beta}\sum_n\frac{\partial\ln\mathcal{Z}_n}{\partial V}=\frac{2}{\beta}\sum_n\frac{\partial\ln\mathcal{Z}_n}{\partial\varepsilon_n}\frac{\partial\varepsilon_n}{\partial L}\frac{\partial L}{\partial V}=\frac{2}{\beta}\sum_{\vec{n}}\frac{-\beta}{e^{\beta\varepsilon_{\vec{n}}}-1}\frac{-\varepsilon_n}{L}\frac{L}{3V}=\frac{1}{3V}\langle E\rangle \tag{50}$$

That is,

$$\langle E\rangle=3PV \tag{51}$$

This equation of state is identical to what we derived classically, as it should be since that derivation used only thermodynamics.

3.4 Planck distribution

To proceed further, we need to sum over $\vec{n}$. As usual, we do this by going to the continuum limit and integrating over energy. As in the phonon case (cf. Eq. (14)), we compute the density of states by going from n to ω (we use $\omega=\frac{\varepsilon}{\hbar}$ conventionally for blackbodies) and the sum to an integral:

$$2\sum_{\vec{n}}1=2\times\frac{1}{8}\int_0^\infty d^3\vec{n}=2\times\frac{\pi}{2}\frac{V}{c^3\pi^3}\int_0^\infty\omega^2d\omega=\int_0^\infty g(\omega)d\omega \tag{52}$$

where the density of states is

$$\boxed{g(\omega)=\frac{V}{c^3\pi^2}\omega^2} \tag{53}$$

This differs from the Debye model by a factor of $\frac{2}{3}$ for the number polarizations and the replacement $c_s\to c$ for the speed of light. There is also no upper limit to ω for photons. Otherwise the density of the state is the same since photons and phonons are both massless particles-in-a-box.

Then we have

$$\langle E\rangle = 2\sum_{\vec{n}}\varepsilon_{\vec{n}}\frac{1}{e^{\beta\varepsilon_{\vec{n}}}-1} = \int_0^\infty \hbar\omega\frac{1}{e^{\beta\hbar\omega}-1}g(\omega)d\omega = \frac{V\hbar}{c^3\pi^2}\int_0^\infty\frac{\omega^3}{e^{\beta\hbar\omega}-1}d\omega \tag{54}$$

It is helpful to write

$$\frac{\langle E\rangle}{V} = \int_0^\infty d\omega u(\omega,T) \tag{55}$$

where $u(\omega,T)d\omega$ is the energy density as a function of frequency:

$$u(\omega,T)d\omega = \hbar\omega\frac{1}{e^{\beta\hbar\omega}-1}g(\omega)d\omega = \frac{\hbar}{\pi^2c^3}\frac{\omega^3}{e^{\frac{\hbar\omega}{k_BT}}-1}d\omega \tag{56}$$

That is, $u d\omega$ is the amount of energy density in photons with frequencies between ω and $\omega+d\omega$ To compute the total energy density in the blackbody, we just perform the integral:

$$\frac{\langle E\rangle}{V} = \int_0^\infty d\omega u(\omega,T) = \frac{\pi^2k_B^4}{15\hbar^3c^3}T^4 = 7.5\times 10^{-16}\frac{J}{m^3K^4} \tag{57}$$

Now we can write this as in Eq. (41):

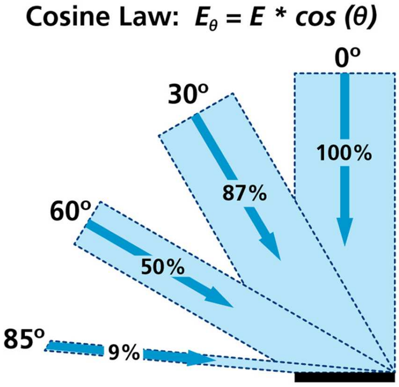
$$\frac{\langle E\rangle}{V} = \frac{4\sigma}{c}T^4 \tag{58}$$

where

$$\sigma = \frac{\pi^2k_B^4}{60\hbar^3c^2} = 5.67\times 10^{-8}\frac{W}{m^2K^4} \tag{59}$$

is known as **Stefan's constant**. Thus quantum statistical mechanics has computed the proportionality factor that thermodynamics alone was unable to provide.

Rather than the total or differential energy, we are often interested in the power: how much energy leaves the blackbody per unit time. A related quantity is the **flux** of radiation, which is the power per unit area. The flux is given by the energy density times the velocity. For example, the flux in an electromagnetic wave is $\Phi = c\frac{E}{V}$, since electromagnetic waves travel at the speed of light. Similarly, you might think that the flux from a blackbody is $c\frac{\langle E\rangle}{V}$. This is not quite right, however. First of all, half of the photons emitted from the surface go into the blackbody, not out. So we are off by at least a factor of 2. Second of all, although there are photons leaving each point of the surface going off in any outward direction, the outward flux of photons shrinks due to a geometric factor associated with the propagation direction:



(60)

Think of the flux of photons passing through a spherical shell far away from the blackbody at some time. At any point on the shell, it is only the perpendicular component of the flux from the surface that hits the point. So we need to multiply by a projection factor of cosθ when doing the angular integral over the outgoing directions. (The factor of $\cos\theta$ actually has a name, Lambert's cosine law). Thus, instead of 4π of the sphere, the effective solid angle reached from the surface of the black body is

$$\Omega_{\text{eff}} = \int_0^{2\pi} d\phi \int_0^{\pi/2} \sin\theta d\theta \cos\theta = \pi \tag{61}$$

The 0 to $\frac{\pi}{2}$ limits on the θ integral (upper hemisphere only) enforce the first point, that the radiation is outgoing. Thus we need to rescale our flux by $\frac{\Omega_{\text{eff}}}{\Omega_{\text{total}}} = \frac{\Omega_{\text{eff}}}{4\pi}$. So the total flux is

$$\boxed{\Phi = \frac{\Omega_{\text{eff}}}{4\pi} c \frac{E}{V} = \sigma T^4} \tag{62}$$

This relation between flux and temperature is known as the **Stefan-Boltzmann law**.

Similarly, we can compute the **intensity**, meaning the differential flux per unit angular frequency, by using the differential energy density $u(\omega, T)$ instead of the energy density. We then get

$$\boxed{I(\omega, T) d\omega \equiv \frac{\Omega_{\text{eff}}}{4\pi} c u(\omega, T) d\omega = \frac{\hbar}{4\pi^2 c^2} \frac{\omega^3}{e^{\frac{\hbar\omega}{k_B T}} - 1} d\omega} \tag{63}$$

This is the intensity of blackbody radiation as a function of frequency. It has units of power per area per frequency. This function is known as the **Planck distribution** or **Planck radiation formula**.

To convert the intensity from angular frequency ω to wavelength $\lambda = \frac{2\pi}{\omega} c$ we need a Jacobian factor of $\left|\frac{d\omega}{d\lambda}\right| = \frac{2\pi}{\lambda^2} c$. This gives the intensity as a function wavelength:

$$I(\lambda, T) d\lambda = I(\omega, T) \left|\frac{d\omega}{d\lambda}\right| d\lambda = \frac{2c^2 h\pi}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1} d\lambda \tag{64}$$

This intensity has units of power per area per wavelength. It describes the wavelength-dependence of the emitted power from a blackbody, known as the **blackbody spectrum**:

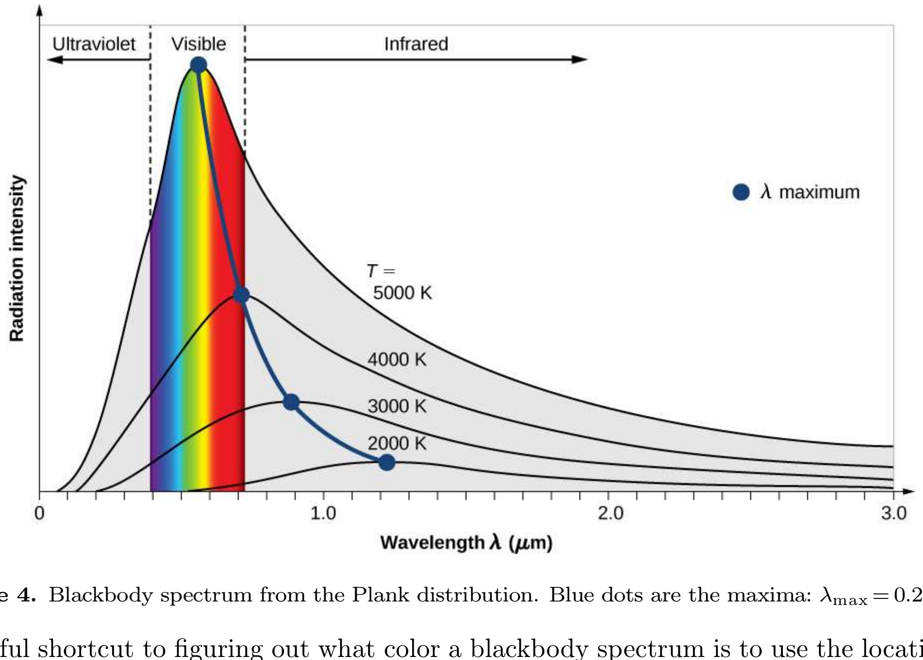


Figure 4. Blackbody spectrum from the Plank distribution. Blue dots are the maxima: $\lambda_{\max} = 0.201 \frac{hc}{k_B T}$.

A useful shortcut to figuring out what color a blackbody spectrum is to use the location of the peak. We can compute this by solving $\frac{dI}{d\lambda} = 0$. Pulling out the dimensionful factors, $\frac{dI}{d\lambda} = 0$ gives a transcendental equation that can be solved numerically (please check!). The result is

$$\lambda_{\text{peak}} = 0.201 \frac{hc}{k_B T} = \frac{2.89\,\text{mm} \cdot K}{T}$$

The relationship between the peak wavelength and temperature is known as **Wein's displacement law**: $T\cdot\lambda_{\text{peak}} = 2.89\,\text{mm}\cdot K = \text{constant}$.

For an example, the surface of the sun is 5800 K, so the peak of its blackbody spectrum is $\lambda_{\max} = 2.89$ mm $\cdot\, K\frac{1}{5780\,K} = 499$ nm $=$ 🟩 is right in the middle of the visibile spectrum. What luck! Stars that are cooler peak at large wavelength (lower frequency, lower energy). For example, Antares (a red giant) peaks in red wavelengths.

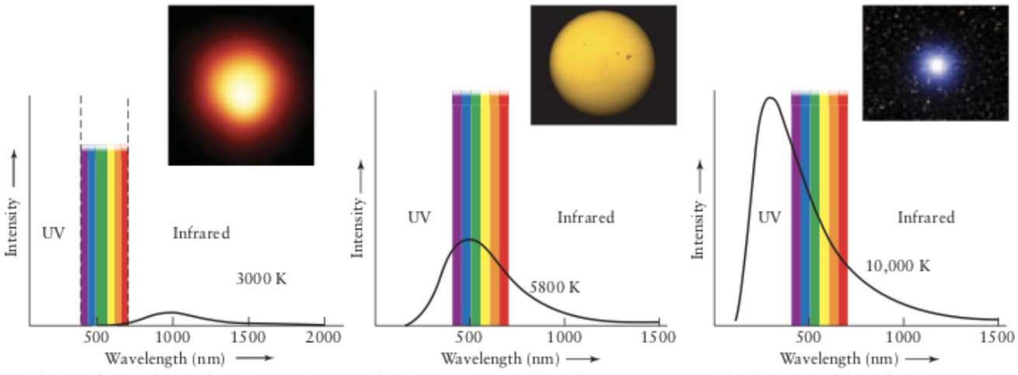


Figure 5. A cold star like Antares (left) might have a surface temperature of 3000K. It looks red. The sun (middle) has a surface temperature around 5800 K, making it yellower. Sirius (right) has a surface temperature around 9,900K, making it look blue to white. To figure out the apparent color you need to integrate over the spectrum and account for scattering off the sky as well.

We can compute other features of solar radiation from our formulas. The total radiation flux is

$$\Phi_{\odot} = \sigma \times (5780\,K)^4 = 6.32 \times 10^7 \frac{W}{m^2} \tag{65}$$

The sun's radius is $R_{\odot} = 7 \times 10^8 m$ thus, the total power radiated is

$$P_{\odot} = 4\pi R^2 \Phi = 3.84 \times 10^{26} W \tag{66}$$

The earth is $R_{\text{sun}\to\text{earth}} = 1.5 \times 10^{11} m$ from the sun. At this distance, we receive

$$\Phi_{\text{sun}\to\text{earth}} = \frac{1}{4\pi R_{\text{sun}\to\text{earth}}^2} P_{\odot} = 1358 \frac{W}{m^2} \tag{67}$$

This solar flux $1.35\,\text{kW}/m^2$ is critical for nearly every aspect of life on earth. About 1 $\frac{\text{kW}}{m^2}$ of this flux penetrates the atmosphere and makes it to the earth's surface on a clear day.

3.5 Entropy and photon number

The grand-canonical partition function for a blackbody is, from Eq. (48)

$$\ln\mathcal{Z} = -2\sum_{\vec{n}} \ln\left(1 - e^{-\beta\frac{\pi\hbar c}{L}|\vec{n}|}\right) = -\int_0^\infty \ln(1 - e^{-\beta\hbar\omega}) g(\omega) d\omega \tag{68}$$

Plugging in the density of states in Eq. (53) gives

$$\ln\mathcal{Z} = -\frac{V}{c^3\pi^2}\int_0^\infty \omega^2 d\omega \ln(1 - e^{-\beta\hbar\omega}) = \frac{\pi^2}{45 c^3\hbar^3\beta^3} V \tag{69}$$

We can double check the energy density using Eq. (49)

$$\langle E\rangle = -\frac{\partial \ln\mathcal{Z}}{\partial\beta} = \frac{\pi^2 V}{15 c^3\hbar^3\beta^4} = 4\frac{\sigma}{c} T^4 V \tag{70}$$

The heat capacity is

$$C_V = \frac{\partial\langle E\rangle}{\partial T}\bigg|_V = 16\sigma T^3 V \tag{71}$$

This goes to zero at $T \to 0$ in agreement with the 3rd law of thermodynamics.

The entropy is (see Eq. (26))

$$S = \int_0^T dT' \frac{C_V(T')}{T'} = \frac{16}{3}\frac{\sigma}{c} T^3 V = \frac{32\pi^5}{45}\left(\frac{k_B T}{h c}\right)^3 k_B V \tag{72}$$

The expected number of photons in the volume V is

$$\langle N \rangle = \int d\omega g(\omega) \frac{1}{e^{\beta\hbar\omega} - 1} = \frac{V}{c^3 \pi^2} \int_0^\infty d\omega \frac{\omega^2}{e^{\beta\hbar\omega} - 1} = \frac{2V}{c^3 \pi^2 \hbar^3 \beta^3} \zeta_3 = 16\pi V \left(\frac{k_B T}{h c}\right)^3 \zeta_3 \tag{73}$$

where $\zeta_3 = \zeta(3) = \sum_1^\infty \frac{1}{x^3} = 1.202$. Note that S and N are proportional to each other:

$$S = \frac{2\pi^4}{45 \xi_3} k_B N = 3.61 k_B N \tag{74}$$

Thus each photon has $3.61 k_B$ of entropy (on average). The information carried by each thermal photon is $H = \frac{S}{N k_B \ln 2} = 5.2\,\text{bits}$. Note that this entropy per photon is independent of temperature.

To get a feel for these expressions, let's plug in some numbers. The cosmic microwave background (CMB) is a photon gas (blackbody) at 2.73 K. The photon spectrum therefore has a peak at $\lambda_{\text{max}} = 1.05\text{mm}$ (microwave). The energy density of the CMB is $\frac{\langle E \rangle}{V} = 0.26 \frac{\text{MeV}}{m^3}$. This is much smaller than the energy density of matter, which is around 0.2 proton$/m^3$ giving $\frac{\langle E \rangle}{V} \approx 200 \frac{\text{MeV}}{m^3}$. In fact, both of these are much smaller than the *total* energy density of the universe, which we can deduce from the rate of expansion $\frac{E}{V} = \frac{3c^2 H_0^2}{8\pi G} = 5200 \frac{\text{MeV}}{m^3}$. Thus matter is only 4% of the energy density of the universe and photons are a measly 0.005%. The missing 96% is called "dark", since we don't see it. It consists of dark energy and dark matter.

The entropy density of the CMB is $\frac{S}{V} = (3 \times 10^9) \frac{k_B}{m^3}$. This is much *greater* than the entropy density of matter: 1 proton$/m^3$ gives around k_B / m^3 of entropy. So the entropy density of photons is a billion times greater than the entropy density for matter. The observable universe is around 10^{27} meters across, giving a total entropy of the universe of $S = 10^{90} k_B$ almost entirely contained in the photons of the CMB.

The number density of CMB photons is $n_\gamma = \frac{N}{V} = (8 \times 10^8) \frac{1}{m^3}$. That is, on average, every cubic meter of empty space has around a billion photons in it left over from the big bang, and about 1/5th of a proton, $n_B \sim 0.2 \frac{1}{m^3}$. The baryon-to-photon ratio is

$$\eta = \frac{n_B}{n_\gamma} = 6 \times 10^{-10} \tag{75}$$

One important unsolved mystery is why this is not much smaller (the baryogenesis problem, discussed in Lecture 7). In any case, this ratio η provides a useful standard cosmological parameter. Since both n_γ and n_B scale like T^3, the baryon-to-photon ration has remained essentially constant since the CMB photons were produced, 3 minutes after the big bang.

4 Summary

This lecture introduced and studied two important bosonic systems using quantum statistical mechanics: phonons, which account for properties of solids, and photons, which explain blackbody radiation.

For either case, the first step was to work out the density of states. The density of states $g(\varepsilon)$ gives the number of states with energies between ε and $\varepsilon + d\varepsilon$. From it we can integrate $\int_0^{\varepsilon_{\max}} g(\varepsilon) d\varepsilon = N_{\text{states}}$ to get the total number of states. We also use the density of states to compute the grand-canonical partition function as $\ln\mathcal{Z} = \int g(\varepsilon) d\varepsilon \ln\mathcal{Z}_\varepsilon$ where $\mathcal{Z}_\varepsilon = (1 - e^{-\beta(\varepsilon - \mu)})^{-1}$ is the partition function for a single state of energy ε. Much of the difficult work in statistical mechanics is in computing the density of states $g(\varepsilon)$. From $\mathcal{Z}$ we deduce the heat capacity and other thermodynamic quantities.

In the Debye model, the possible states are the normal modes of excitation of a solid. These are each treated as simple harmonic oscillators. The excitation number of each oscillator is the number of phonons in that mode. The density of states is then $g(\omega) = 3\frac{V}{2\pi^2 c_s^3}\omega^2$ with the factor of 3 accounting for the 3 polarizations of sound waves in the solids. Since there are a finite number of normal modes, there is a maximum energy. Using $\int_0^{\varepsilon_D} g(\varepsilon) d\varepsilon = N_{\text{states}} = 3N_{\text{atoms}}$, one can trade ε_D for N_{atoms}. We then computed the heat capacity at high temperature, $C_V \approx 3Nk_B$, consistent with the Law of Dulong and Petit, and a low temperature, $C_V \approx \frac{12\pi^4}{5} N k_B \left(\frac{T}{T_D}\right)^3$, where $T_D = \frac{\hbar\omega_D}{k_B}$ is the Debye temperature. The heat capacity in both regimes is in good agreement with data for a variety of solids.

Blackbody radiation refers to the equilibrium of electromagnetic radiation in a cavity at temperature T. In quantum statistical mechanics, blackbody radiation is computed using a grand canonical ensemble of photons. The density of states for photons is $g(\omega) = 2\frac{V}{2c^3\pi^2}\omega^2$. This is similar to the density of states in the Debye model , up to the fact that there are two polarizations of photons but 3 polarizations of photons, and the different speeds. An important result is Stefan's law for the energy density: $\frac{E}{V} = 4\frac{\sigma}{c} T^4$. The scaling with T can be deduced classically, but the numerical constant in front $\sigma = \frac{\pi^2 k_B^4}{60\hbar^3 c^2}$ can only be determined using quantum statistical mechanics.

The most important result from the theory of blackbody radiation is the intensity spectrum

$$I(\lambda, T) d\lambda = \frac{2c^2 h\pi}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1} d\lambda \tag{76}$$

Intensity is the power radiated per unit area per unit wavelength. Integrating it over λ gives the total radiated power per unit area, aka the total flux, $\Phi = \sigma T^4$. For example, from this we can compute the total power radiated from the sun given its surface temperature: $\Phi_{\odot} = \sigma \times (5780K)^4 = 6.32 \times 10^7 \frac{W}{m^2}$.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 12: Bose-Einstein Condensation

1 Introduction

Bose-Einstein condensation is a quantum phenomenon in Bose gases in which a large number of bosons simultaneously occupy the ground state of a system. Bose-Einstein condensates were predicted in 1925 by Bose and discovered experimentally 70 years later by Weimann, Cornell and Ketterle who shared the Nobel prize for their discovery in 2001. More precisely, these scientists constructed an experiment where the phase transition to Bose-Einstein condensation could be clearly seen and measured. Many quantum phenomena such as superconductivity, superfluids or lasers can also be understood as Bose-Einstein condensation.

Your first thought might be, of course a lot of bosons are in the ground state! After all, there is no quantum effect preventing them from being in the ground state (no Pauli exclusion), and naturally particles want to be in the state of lowest energy. This is a good thought; let's follow through. How many bosons do you expect in the ground state? Well, say they obey Maxwell-Boltzmann statistics, so that $n_i \sim e^{-\varepsilon_i / k_B T}$. This function is pretty flat for $\varepsilon_i \ll k_B T$, so we would expect that if there are say 100 states below $k_B T$ then each one should have roughly the same number of particles in it – nothing too special about the ground state. Thus, if you want a sizable fraction, say 1/2 the particles, to be in the ground state, you would have to get $k_B T$ down below the energy of the first excited state ε_1. This argument is correct for Maxwell-Boltzmann statistics, and we'll flesh it out more in a moment. The amazing thing is that with Bose-Einstein statistics the argument completely fails – you can find more than half of the particles in the ground state even for temperatures with $k_B T \gg \varepsilon_1$.

Bose-Einstein condensation is tricky to explain, so we'll approach it different ways. First, we'll try to understand what it is about Bose-Einstein statistics that allows condensation to happen through a simple system that we can solve in the canonical ensemble. Then we'll do the general case using the grand-canonical ensemble, first numerically, and then through analytic expansions.

In this lecture, it will be helpful to set the ground state energy to zero: $\varepsilon_0 = 0$. By setting it to zero, we mean that we list all energies as *relative* to the ground state. Shifting all the energies as well as the chemical potential in this way will have no effect on the physics and can be done without loss of generality.

2 Two-state system: canonical ensemble

Consider a system with N particles but only two possible energy states: $\varepsilon_0 = 0$ and $\varepsilon_1 = \varepsilon$. Because there are only two states, we can study this system in the canonical ensemble for both Maxwell-Boltzmann statistics and Bose-Einstein statistics. We'll see that even in this two-state system the ground state occupancy can be much larger than $\frac{N}{2}$ even when $k_B T \gg \varepsilon$ with Bose-Einstein statistics. With the large number of states present in any realistic system, the canonical ensemble will not be tractable and we will have to resort to the more abstract grand canonical ensemble.

Let's start with the simplest case of our 2-state system, $N = 1$. If there is only one particle, then for any statistics, the partition function is

$$Z_1 = \sum_k e^{-\beta E_k} = 1 + e^{-\beta \varepsilon} \tag{1}$$

The probability of finding the particle in the ground state is $P_{\text{ground}} = \frac{1}{Z_1} e^{-\beta \varepsilon_0} = \frac{1}{Z_1}$ and so the expected fraction of particles in the ground state is

$$\frac{\langle N_{\text{ground}} \rangle}{N} = 1 \cdot P_{\text{ground}} + 0 \cdot P_{\text{not}-\text{ground}} = \frac{1}{Z_1} = \frac{1}{1 + e^{-\beta \varepsilon}} \tag{2}$$

Again, this holds for any statistics, since there is only one particle.

Now say there are N particles. With Maxwell-Boltzmann statistics, the probability of finding any particle in the ground state is independent of the probability of finding any other particle anywhere. This implies that the N particle partition function is related to the 1 particle one by

$$Z_N^{\text{MB}} = \frac{1}{N!}(Z_1)^N = \frac{1}{N!}(1+e^{-\beta\varepsilon})^N \tag{3}$$

Of the 2^N microstates, there is only one microstate with all the particles in the ground state, so

$$P_{\text{all ground}} = \frac{1}{Z_N^{\text{MB}}}\left(\frac{1}{N!}e^{-\beta\varepsilon_0}\right) = \frac{1}{Z_N^{\text{MB}}}\frac{1}{N!} \tag{4}$$

Note that we include the $\frac{1}{N!}$ in the probabilty for the same reason we do in Z_N^{MB}, to account for identifal particles. To confirm that this is correct, let us verify that the probabilities sum to one. There are N states with 1 particle in the excited state, $\left(\begin{smallmatrix} N \\ 2 \end{smallmatrix}\right)$ states with 2 particles in the excited state and so on. So the sum of probabilties is

$$\sum_k P_k = \frac{1}{Z_N^{\text{MB}}}\frac{1}{N!}\left[1 + Ne^{-\beta\varepsilon} + \left(\begin{array}{c} N \\ 2 \end{array}\right)e^{-2\beta\varepsilon} + \cdots + \left(\begin{array}{c} N \\ N \end{array}\right)e^{-N\beta\varepsilon}\right] = \frac{1}{N!Z_N^{\text{MB}}}(1+e^{-\beta\varepsilon})^N = 1 \tag{5}$$

where Eqs. (1) and (3) were used in the last step. To compute the expected number in the ground state, we multiply each term in this sum by the ground state occupancy:

$$\langle N_{\text{ground}}^{\text{MB}}\rangle = \frac{1}{N!Z_N^{\text{MB}}}\left(N\cdot 1 + (N-1)\cdot Ne^{-\beta\varepsilon} + (N-2)\cdot\left(\begin{array}{c} N \\ 2 \end{array}\right)e^{-2\beta\varepsilon} + \cdots + 0\cdot\left(\begin{array}{c} N \\ N \end{array}\right)e^{-N\beta\varepsilon}\right) \tag{6}$$

$$= \frac{N}{e^{-\beta\varepsilon}+1} \tag{7}$$

You can check this sum in Mathematica. Note that at large T ($\beta\to 0$), $\langle N_{\text{ground}}^{\text{MB}}\rangle$ goes to $\frac{N}{2}$: half the particles are in the ground state, half in the excited state.

With Bose-Einstein statistics there is only one state with m particles in the ground state and $N-m$ particles in the excited state. So there are only $N+1$ possible states all together and

$$Z_N^{\text{BE}} = 1 + e^{-\beta\varepsilon} + e^{-2\beta\varepsilon} + \cdots + e^{-N\beta\varepsilon} = \frac{1-e^{-(N+1)\beta\varepsilon}}{1-e^{-\beta\varepsilon}} \tag{8}$$

Then the expected number in the ground state is

$$\langle N_{\text{ground}}^{\text{BE}}\rangle = \frac{1}{Z_N^{BE}}[N\cdot 1 + (N-1)\cdot e^{-\beta\varepsilon} + (N-2)\cdot e^{-2\beta\varepsilon} + \cdots + 0\cdot e^{-N\beta\varepsilon}] \tag{9}$$

$$= \frac{1}{e^{-\beta\varepsilon}-1} + \frac{N+1}{1-e^{-(N+1)\beta\varepsilon}} \tag{10}$$

Let us look at Eqs. (7) and (10) numerically for $N=100$:

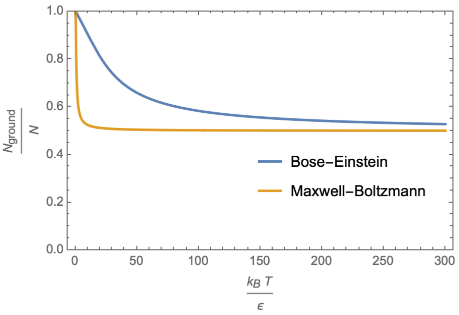


Figure 1. The fractional population of the ground state in a two state system with $N=100$.

This plot demonstrates Bose-Einstein condensation. With Maxwell-Boltzmann statistics, the temperature has to be very low to get the lowest state to have an appreciable filling fraction. At temperature $k_BT \gtrsim \varepsilon$ both the ground state and the first excited state are around equally populated so $\langle N_{\text{ground}}\rangle = \frac{N}{2}$. In contrast, with Bose-Einstein statistics, a significant fraction of the the particles are in the ground state even well above the temperature $k_BT = \varepsilon$. For example, with at $k_BT = 10\varepsilon$ we find 90% of the atoms are in the ground state for Bose-Einstein statistics, but only 52% for Maxwell-Boltzmann statistics. This demonstrates Bose-Einstein condensation.

Bose-Einstein condensation is a phase transition whereby the ground state become highly occupied. What is the critical temperature for this to occur? There is only two dimensionless numbers we can work with $\frac{k_BT}{\varepsilon}$ and N. We want N to be large, $N \gg 1$. Then we can expand in two limits $\frac{k_BT}{\varepsilon} \gg N$ (i.e. $\beta\varepsilon N \ll 1$) and $\frac{k_BT}{\varepsilon} \ll N$ (i.e. $\beta\varepsilon N \gg 1$). Expanding Eq. (10) in the first limit gives

$$\frac{\langle N^{\text{BE}}_{\text{ground}}\rangle}{N} = \frac{1}{2} + \frac{N\varepsilon}{12k_BT} + \cdots \qquad \left(\frac{k_BT}{\varepsilon} \gg N\right) \tag{11}$$

This is the true high-temperature limit, where the classical behavior $\frac{\langle N^{\text{BE}}_{\text{ground}}\rangle}{N} \to \frac{\langle N^{\text{MB}}_{\text{ground}}\rangle}{N} \approx \frac{1}{2}$ is approached as $T \to \infty$. In the second limit $N \gg \frac{k_BT}{\varepsilon} \gg 1$, the temperature is large, but not too large, and the expansion gives a different result

$$\frac{\langle N^{\text{BE}}_{\text{ground}}\rangle}{N} = 1 - \frac{k_BT}{N\varepsilon} + \cdots \qquad \left(N \gg \frac{k_BT}{\varepsilon}\right) \tag{12}$$

This limit shows the growth of $\frac{\langle N^{\text{BE}}_{\text{ground}}\rangle}{N}$ toward 1 as $T \to 0$. The comparison of these approximations to Eq. (10) looks like

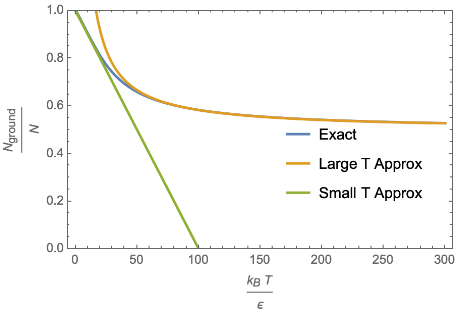


Figure 2. The approximations to the Bose-Einstein curve in Eqs. (11) and (12).

The crossover point is roughly where the approximations used for our expansions break down. The first term is the same order as the second term in Eq. (11) when $k_BT = \frac{N\varepsilon}{6}$. For Eq. (12) the crossover is at $k_BT = N\varepsilon$. Thus we find a critical temperature $T_c \sim \frac{N\varepsilon}{6k_B} \sim \frac{N\varepsilon}{k_B}$ for this two state model indicating the onset of Bose-Einstein condensation. (The crossover point in this 2-state example is not at a precise temperature, as you can see from the plot. T_C will become precise when we consider a realistic system with a large number of states in the next section.)

To emphasize how strange Bose-Einstein condensation is, remember that at $k_BT \gg \varepsilon$ we should be able to use $e^{-\beta\varepsilon} \approx 1$ independent of N. That is, the ground state and first excited state should have pretty similar thermodynamic properties and occupancy at high temperature. This is *not* what we are finding. Instead, for $N = 10$ million, with a temperature 1 million times the excited state energy, 90% of the atoms are in the ground state and only 10% are in the excited state.

Now, this 2-state model is not a realistic approximation to any physical system. It turns out to be very difficult to calculate $\langle N^{\text{BE}}_{\text{ground}}\rangle$ in the canonical ensemble for a realistic system that has an infinite number of states. The difficulty is that we have to count the number of ways of allocating N particles to the states, and then to perform the sum over occupancies of the ground state times Boltzmann factors. It turns out to be much easier to compute the general case using the grand-canonical ensemble with μ instead of N, as we will now see.

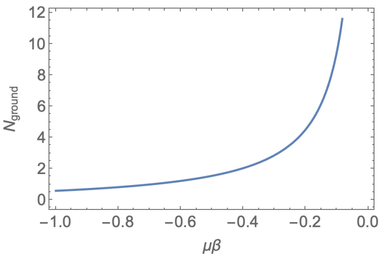
3 Grand canonical ensemble

With Bose-Einstein statistics, we determined that using the grand canonical ensemble the expected number of particles in a state i is

$$\langle N_i\rangle = \frac{1}{e^{\beta(\varepsilon_i-\mu)}-1} \tag{13}$$

with ε_i the energy of the state i. So the expected number of particles in the ground state ($\varepsilon_i=0$) is

$$\langle N_{\text{ground}}\rangle = \frac{1}{e^{-\beta\mu}-1} = \tag{14}$$



Recall that for Bose gases, μ will always be lower then all the energies (i.e. $\mu<0$ when $\varepsilon_0=0$). We can see this explicitly from the plot since there is a singularity at $\mu=0$. This singularity, that $N_{\text{ground}}\to\infty$ as $\mu\to0$ is not physical. Recall that in the grand canonical ensemble we do not fix N, so the same distribution has to allow for arbitrarily large N. If we know N then we have to trade μ for N by imposing the constraint $\sum_i \langle N_i\rangle = N$. This is not so easy, but we can do it.

To replace μ by N, we first invert Eq. (14) to solve for μ in terms of $\langle N_{\text{ground}}\rangle$:

$$e^{-\beta\mu} = \frac{1}{\langle N_{\text{ground}}\rangle}+1 \tag{15}$$

Our strategy will then be to find the ground state occupancy by using the constraint the the total number of particles is N. That is, we will compute

$$N = \sum_{i=0}^{\infty}\langle N_i\rangle = \sum_{i=0}^{\infty}\frac{1}{e^{\beta\varepsilon_i}e^{-\beta\mu}-1} = \sum_{i=0}^{\infty}\frac{1}{e^{\beta\varepsilon_i}\left(\frac{1}{\langle N_{\text{ground}}\rangle}+1\right)-1} \tag{16}$$

Once we work out the ε_i we can do the sum numerically or analytically and therefore find $\frac{\langle N_{\text{ground}}\rangle}{\mathrm{N}}$.

Before beginning the calculation, let us quickly ask about the non-ground state occupancies. Since

$$\mu = -\frac{1}{\beta}\ln\left[\frac{1}{\langle N_{\text{ground}}\rangle}+1\right] < -\frac{1}{\beta}\ln\left[\frac{1}{N}+1\right] < 0 \tag{17}$$

μ is always negative and gets closest to zero when $\langle N_{\text{ground}}\rangle$ is largest. Conversely, $\langle N_{\text{ground}}\rangle$ is largest when μ is closest to zero as can be seen in Eq. (14).

Can excited states have a large number of particles in them? The explosion of particles in the ground state arose because if $\mu\to0$ then $e^{-\beta\mu}\to1$ and $\langle\ N_{\text{ground}}\rangle = \frac{1}{e^{-\beta\mu}-1}\to\infty$. For the first excited state,

$$\langle N_1\rangle = \frac{1}{e^{\beta(\varepsilon_1-\mu)}-1} = \frac{1}{e^{\beta\varepsilon_1}\left(\frac{1}{\langle N_{\text{ground}}\rangle}+1\right)-1} \tag{18}$$

Since $\frac{1}{\langle N_{\text{ground}}\rangle}+1>1$ and $e^{\beta\varepsilon_1}>1$, this can never get too large. Indeed, as $\mu<0$ there is always a gap between μ and any energy other than the ground state, so $e^{\beta(\varepsilon_1-\mu)}>e^{\beta\varepsilon_1}>1$. Since μ cannot get arbitrarily close to ε_i condensation cannot happen in any excited state. The ground state is special.[1]

[1] Technically speaking, this is true only in equilibrium. In a laser, photons condense into an excited state. But lasers must be pumped – they are not in equilibrium.

3.1 Exact numerical solution

Now let us calculate N, and hence $\frac{\langle N_{\text{ground}}\rangle}{N}$. Bose-Einstein condensation is relevant at low temperature, where particles are non-relativistic. So consider a non-relativistic gas of monatomic bosonic atoms in a 3D box of size L. The allowed wavevectors of the system are $\vec{k}_n = \frac{\pi}{L}\vec{n}$ just like for photons or phonons, and the momenta are $\vec{p}_n = \hbar\vec{k}_n$ as always. In a non relativistic system, the energies are

$$\varepsilon_n = \frac{\vec{p}_n^{\,2}}{2m} = \frac{\hbar^2\pi^2}{2mL^2}\vec{n}^2 \tag{19}$$

We have set the ground state to $\varepsilon_0 = 0$ (rather than $\varepsilon_0 = mc^2$), since the absolute energy scale will be irrelevant. Another useful number is the gap to the first excited state

$$\varepsilon_1 = \frac{\hbar^2\pi^2}{2mL^2}(1,0,0)^2 = \frac{\hbar^2\pi^2}{2mL^2} \tag{20}$$

So that

$$\varepsilon_n = \varepsilon_1 n^2 \tag{21}$$

Then, from Eq. (16) we get

$$N = \sum_{n_x,n_y,n_z=0}^{\infty} \frac{1}{e^{\beta\varepsilon_1(n_x^2+n_y^2+n_z^2)}\left(\frac{1}{\langle N_{\text{ground}}\rangle}+1\right)-1} \tag{22}$$

This formula lets us compute N given $\langle N_{\text{ground}}\rangle$ and $\beta\varepsilon_1$. For example, if $\langle N_{\text{ground}}\rangle = 80$ and $T = 10\frac{\varepsilon_1}{k_B}$ then doing the sum numerically gives $N = 167.5$. This means with 167.5 particles at $T = 10\frac{\varepsilon_1}{k_B}$, then 80 will be in the ground state.

What we really want is to specify N and T and find $\langle N_{\text{ground}}\rangle$. Do get this function, we need to do the sum in Eq. (22) and then solve for $\langle N_{\text{ground}}\rangle$ in terms of $\beta\varepsilon_1$ and N. Unfortunately, we cannot do the sum exactly, but at least we can do it numerically. For $N = 100$ we find the numerical solution for $\frac{\langle N_{\text{ground}}\rangle}{N}$ has the form (see the Mathematica notebook on canvas)

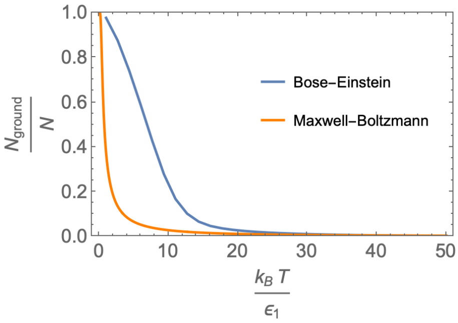


Figure 3. Exact numerical result for the ground state occupancy in a Bose system with $N = 100$.

I added to the plot the prediction using Maxwell-Boltzmann statistics. For MB statistics, we drop all the factors of ±1. So, $\langle N_{\text{ground}}\rangle = e^{\beta\mu}$ and so Eq. (22) becomes

$$N = \sum_{n_x,n_y,n_z} \frac{1}{e^{\beta\varepsilon_1(n_x^2+n_y^2+n_z^2)}\left(\frac{1}{\langle N_{\text{ground}}\rangle}\right)} = \langle N_{\text{ground}}\rangle \sum_{n_x,n_y,n_z} e^{-\beta\varepsilon_1(n_x^2+n_y^2+n_z^2)} \tag{23}$$

These sums can be done numerically with the result is plotted alongside the BE result in Fig. 3. If we turn the sum into integrals, then $\frac{\langle N_{\text{ground}}^{\text{MB}}\rangle}{N} \approx \left(\frac{4\varepsilon_1}{\pi k_B T}\right)^{3/2}$ which looks a lot like the exact result that is plotted.

For different N the curve shifts, but looks qualitatively the same. After a little fiddling (inspired by the analytic result below), we see that if we plot the ground state occupancy as a function of $\frac{k_B T}{N^{2/3}\varepsilon_1}$ the result is essentially independent of N for the Bose-Einstein case:

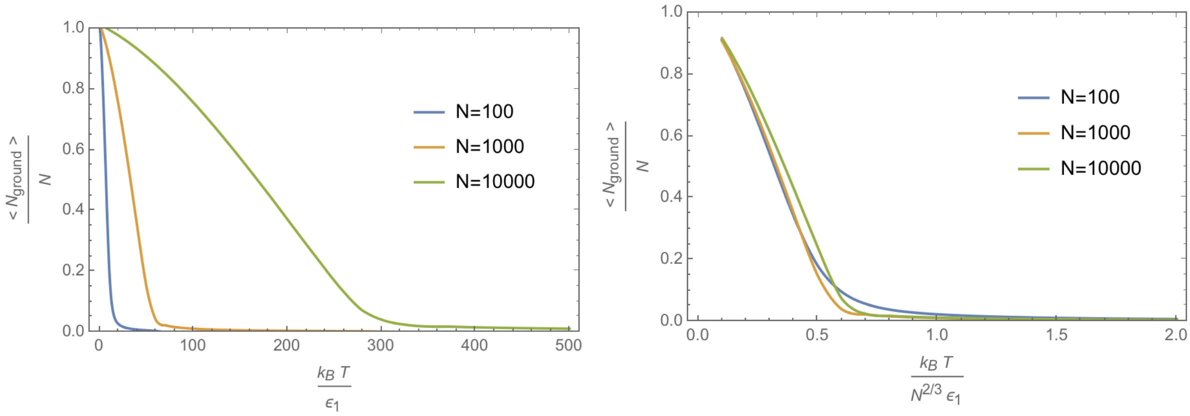


Figure 4. Ground state occupancy in a Bose system for different N as a function of T (left) and of $\frac{T}{N^{2/3}}$ (right). Pulling out a factor of $N^{2/3}$ makes the ground state occupancy essentially independent of N.

The kink in the graph, at around $T_c \approx 0.8\frac{N^{2/3}\varepsilon_1}{k_B}$ indicates the phase transition. Above this temperature, the ground state is basically empty, having only its fair share of particles. Below this temperature, $\langle N_{\text{ground}}\rangle$ starts growing linearly with T. The rescaling of our numerical result from the left to the right plot indicates that the critical temperature T_c scales like $N^{2/3}$, a result that we will next confirm analytically.

3.2 Approximate analytical solution

Having determined the exact solution numerically, let us proceed to use an analytical approach to determine some scaling relations and the transition temperature.

As with the phonon or photon gas, we first transform the sum to an integral via

$$\sum_{\vec{n}} \rightarrow \frac{1}{8}\int_0^\infty 4\pi n^2 \, dn \tag{24}$$

where the $\frac{1}{8}$ comes from $\vec{n}$ being a vector of whole numbers, as in the phonon or photon case. We want to convert n to ε, which we can do using using Eq. (21), $\varepsilon = \varepsilon_1 n^2$ so

$$d\varepsilon = 2\varepsilon_1 n \, dn \tag{25}$$

So

$$\sum_{\vec{n}} \rightarrow \frac{\pi}{2}\int_0^\infty n n \, dn = \frac{\pi}{2}\int_0^\infty \sqrt{\frac{\varepsilon}{\varepsilon_1}}\frac{d\varepsilon}{2\varepsilon_1} = \frac{\pi}{4\varepsilon_1^{3/2}}\int_0^\infty \sqrt{\varepsilon}\, d\varepsilon \tag{26}$$

As before we write this as

$$\sum_{\vec{n}} \rightarrow \int g(\varepsilon) d\varepsilon \tag{27}$$

where

$$g(\varepsilon) = \frac{\pi}{4\varepsilon_1^{3/2}}\sqrt{\varepsilon} \tag{28}$$

is the density of states.

At this point, we would like to integrate over ε to find N

$$N = \int_0^\infty g(\varepsilon)\langle n_\varepsilon\rangle = \frac{\pi}{4\varepsilon_1^{3/2}}\int_0^\infty d\varepsilon \sqrt{\varepsilon}\frac{1}{e^{\beta(\varepsilon-\mu)}-1} \tag{29}$$

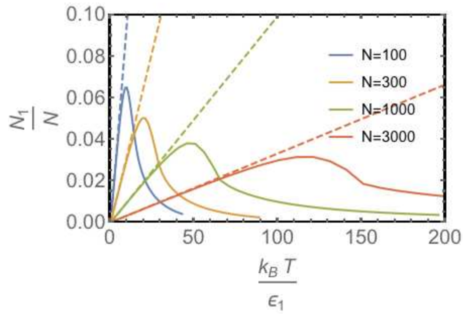
This is a little too quick, however. The problem is that converting a sum to an integral can only be justified if we do not care at all about the discreteness. For Bose-Einstein condensation we *do* care about the discreteness: the ground state, as we have seen, is qualitatively different from the other states.

Although discreteness is important for the ground sates, for the excited states, even the first excited state, there is no issue – the chemical potential μ can never approach any of their energies and so their occupancy numbers will never be unusually large. So let us proceed by taking the continuum limit for all *but* the ground state. Moreover, since $e^{-\beta\mu}\approx 1$ when $\langle N_{\text{ground}}\rangle \gtrsim 1$ which is the region of interest, we can simply set $\mu=0$ for the excited state calculation and Eq. (29) becomes

$$\langle N_{\text{excited}}\rangle \approx \frac{\pi}{4\varepsilon_1^{3/2}}\int_{\varepsilon_1}^{\infty} d\varepsilon\sqrt{\varepsilon}\frac{1}{e^{\beta\varepsilon}-1} = \left(\frac{\pi k_B T}{4\varepsilon_1}\right)^{3/2}\zeta_{3/2}\left[1+\mathcal{O}\left(\sqrt{\frac{\varepsilon_1}{k_B T}}\right)+\cdots\right] \tag{30}$$

where $\zeta_{3/2}=\zeta\left(\frac{3}{2}\right)\approx 2.61$ with $\zeta(z)$ the Riemann Zeta function. The first term on the right comes from integrating from 0 to ∞. The second term on the right, scaling like $\sqrt{\frac{\varepsilon_1}{k_B T}}$ relative to the first, comes from the region $0<\varepsilon<\varepsilon_1$. Typically $k_B T \gg \varepsilon_1$ and these corrections are small.

When does the approximation that $\mu=0$ break down? That is $\mu\approx 0$ is a good approximation the BEC regime (low T), where the ground state is anomalously filled. It breaks down at higher T when $\langle N_{\text{ground}}\rangle$ gets small, as you can see from Eq. (14): as $\langle N_{\text{ground}}\rangle \to 0$ then $\mu\to-\infty$. We can also check this numerically, for example, by looking at the exact numerical solution for $\langle N_1\rangle$, using Eq. (18) and comparing to the $\mu=0$ approximation, where $\langle N_1\rangle = \frac{1}{e^{\beta(\varepsilon_1-\mu)}-1}$:



As

Figure 5. Comparing $\langle N_1\rangle$ computed numerically (solid) to $\langle N_1\rangle$ in the $\mu=0$ approximation (dashed).

So the $\mu=0$ approximation breaks down when $\langle N_{\text{ground}}\rangle\approx 0$ and so $\langle N_{\text{excited}}\rangle \approx N$. Another way to see that the approximation is breaking down is that if we continue to apply $\mu=0$ at higher temperatures then Eq. (30) would imply $\langle N_{\text{excited}}\rangle > N$. Indeed, setting Eq. (30) equal to N we find that our approximation breaks down when

$$\langle N_{\text{excited}}\rangle = \zeta_{3/2}\left(\frac{\pi k_B T}{4\varepsilon_1}\right)^{3/2} > N \tag{31}$$

This transition is where $\langle N_{\text{ground}}\rangle \to 0$. In other words, it occurs at the **critical temperature** where $\langle N_{\text{ground}}\rangle$, the order parameter for BEC goes from 0 to finite value. Setting $\langle N_{\text{excited}}\rangle = N$ therefore gives us a formula for the BEC critical temperature:

$$N=\zeta_{3/2}\left(\frac{\pi k_B T_c}{4\varepsilon_1}\right)^{3/2} = 2.612\left(\frac{m k_B T_c}{8\hbar^2\pi}\right)^{3/2} V \tag{32}$$

where with ε_1 from Eq. (20) was used. Solving for T_c gives

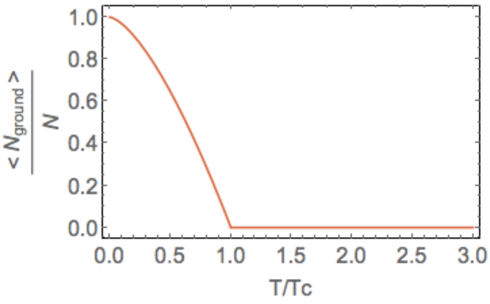
$$\boxed{T_c=\frac{4\varepsilon_1}{\pi k_B}\left(\frac{N}{\zeta_{3/2}}\right)^{2/3} = 3.31\frac{\hbar^2}{k_B m}\left(\frac{N}{V}\right)^{2/3}} \tag{33}$$

This lets us write Eq. (30) a

$$\frac{N_{\text{excited}}}{N} = \left(\frac{T}{T_c}\right)^{3/2}, \qquad T<T_c \tag{34}$$

Thus, the fraction in the ground state is

$$\frac{\langle N_{\text{ground}}\rangle}{N} = \frac{N-\langle N_{\text{excited}}\rangle}{N} = 1-\left(\frac{T}{T_c}\right)^{3/2} = \tag{35}$$



Comparing to our exact numerical results from Section 3.1 we find good agreement:

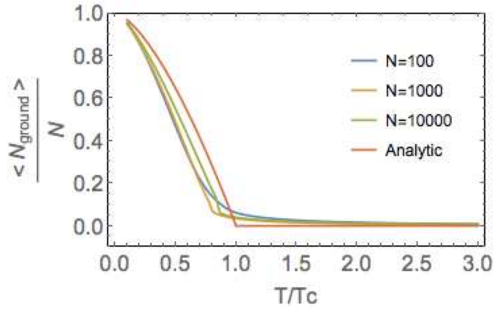


Figure 6. Comparison of the analytic result $1-\left(\frac{T}{T_c}\right)^{3/2}$ with the exact numerical results for $N=10^2, 10^3, 10^4$.

4 Experimental evidence

Even though the transition temperature to the BEC is well above the first excited state energy, one still needs to get a system of bosons very very cold to produce a BEC. It took 70 years from when BEC's were first conjectured theoretically (by Satyendra Nath Bose in 1924) to when technology to cool atoms was sufficiently advanced that the BEC could be produced and detected (Cornell, Weimann and Ketterle in 1995). The tricky part is that normally when you cool a gas of atoms, they solidify. A solid is not a BEC. So you need to cool the gas while keeping the density low. However, since $T_c = 3.31\frac{\hbar^2}{k_B m}\left(\frac{N}{V}\right)^{2/3}$ if you keep the density low you need very very low (ultracold) temperatures.

Since $T_c = 3.31\frac{\hbar^2}{k_B m}\left(\frac{N}{V}\right)^{2/3}$ depends inversely on the mass of the atoms lighter atoms allows the critical temperature to be higher. You might therefore think hydrogen is the easiest element to cool to see a BEC form. Unfortunately, in the mid 1990s, lasers weren't available that could operate at frequencies conducive to cooling hydrogen. Cornell and Weimann, and Ketterle, used Rubidium atoms, ^{87}Rb to form the BEC. Rubidium has a convenient set of energy levels that were will suited to the available laser cooling technologies at the time. They were able to cool around 2000 atoms using magneto-optical traps (MOTs) along with laser and evaporative cooling techniques to the nanokelvin temperature scales required for the BEC. They found that below $T_c \approx 170\,\text{nK}$ Bose-Einstein condensation can be seen. This critical temperature is in excellent agreement with the general formula $T_c = 3.31\frac{\hbar^2}{k_B m}\left(\frac{N}{V}\right)^{2/3}$ when you plug in the density they were able to achieve.

In order to see the BEC, one needs to be able to distinguish atoms in the ground state from atoms in the excited states. The basic feature that makes this possible is that the ground state atoms all have smaller wavenumbers and hence slower velocities than the other states. So if you remove the magnetic trap, the atoms will start to spread, with the ground state atoms spreading more slowly. Thus the atoms' positions after a short time indicate their initial velocities. The scientists photographed (i.e. illuminated the system with a resonant laser pulse) the system after $t = 100\,\text{ms}$ and found a high density of atoms that had not moved very far:

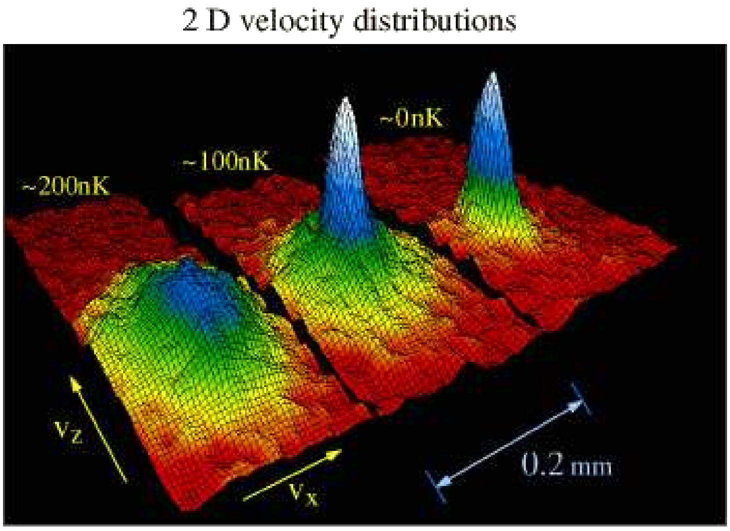


Figure 7. Observation of a Bose-Einstein condensate

In this figure, from Weimann and Cornell's group at the JILA laboratory in Boulder, we see the distribution of atoms in the BEC at different temperatures. The critical temperature is $T_c \approx 170\,\mathrm{nK}$. Above this temperature (left), the distribution is pretty smooth, consistent with a Maxwell-Boltzmann velocity distribution. Below the critical temperature you can clearly see the higher density corresponding at an anomalously large occupancy of the ground state, consistent with expectations from a BEC. The right image shows the BEC at even lower temperature, where the occupancy is even higher. Direct observation of thermodynamic properties of the system, such as the energy density and heat capacity (particularly by Ketterle's group at), further confirmed that this system was a BEC.

Since their discovery BECs have continued to exhibit some amazing and unusual properties. Because of their coherence (all the atoms are in the *same* state), they can manifest quantum phenomena at larger scales than electrons and are in many ways more controllable than electrons. A number of groups at Harvard and MIT study BECs.

For example, Prof. Greiner (Harvard) uses lasers to localize rubidium atoms in an optical lattice. By controlling the spacing and lattice properties, he can fine tune the system, essentially choosing whatever Hamiltonian he wants. Once the rubidium atoms form a BEC, they exhibit strongly correlated behavior that can be connected to the properties of the Hamiltonian. Such an approach may lead the way to building quantum computers with longer coherence times, or to understanding what material properties might be most likely to produce room-temperature superconductors.

Another example is from Prof. Hau's lab (also at Harvard). In 1999 Prof. Hau constructed a BEC of sodium atoms. Normally, a laser tuned to a hyperfine splitting of the sodium levels would be absorbed and so sodium appears opaque to this frequency. However, Hau was able to entangle the ground and excited states of sodium using photons of a different laser in such a way that she could adjust the transparency of sodium to the first laser. The result is that she could manipulate the dispersion relation of light propagating through the sodium BEC and achieve arbitrarily small group velocities. In her first paper on the subject, she slowed light down to $17\frac{m}{s}$ with this technique. Subsequently, she was able to get light to stop completely.

The BECs produced in laboratories have carefully controlled properties, and very restricted interactions. A BEC of rubidium or sodium is unstable to condense in solid form and be held apart with some careful tricks using magnetic fields and optical traps. Helium is a noble gas that naturally has very weak interactions and will only solidify at very high pressure. At low temperature and pressure, it forms a Bose-Einstein condensate called a **superfluid**. Liquid ^{4}He is a BEC of helium atoms. Although ^{3}He is fermionic, pairs of ^{3}He atoms are bosonic, so liquid ^{3}He can be thought of as forming due to the pairing of helium atoms. Superfluids have zero viscosity. This is closely related to the bosons all being in the same state, but in this case, the state is not the zero-momentum state but one of non-zero momentum since there is a density current flowing through the fluid.

Bose Einstein condensation is also related to **superconductors**. In a superconductors, like solid mercury at $T < 4.2K$, there is no resistivity. In the BCS theory, the superconductivity is explained through the condensation of pairs of electrons called **cooper pairs**. These pairs act like bosons and form a condensate at low temperature. Because electrons are charged, so are the cooper pairs. The condensation of charged pairs screens the magnetic field in the superconductors, allowing the charged current to flow with zero resistance. Thus superconductors are much like superfluids with charged bosons instead of neutral ones.

5 Summary

Bose-Einstein condensation is a phenomenon whereby systems of bosons tend towards configurations where most of the bosons are in the identical state. Typically, you might think that at a temperature T, all the states with energy $\varepsilon \lesssim k_B T$ would be occupied. Instead, what happens is that a single state, which can have energy $\varepsilon_0 \ll k_B T$ gets an order-one fraction of the particles, while other states, even ones with $\varepsilon \ll k_B T$ as well, are hardly occupied at all.

There are various ways to understand Bose-Einstein condensation. We looked at a toy model using the canonical ensemble where we saw the condensation happen. A more powerful tool is the grand canonical ensemble. In the grand canonical ensemble, the chemical potential μ is key. The condensation happens for the state with $\varepsilon_0 \approx \mu$ for which the expected occupation $\langle n_i \rangle = \frac{1}{e^{\beta(\varepsilon_i - \mu)} - 1}$ blows up.

We found that there is a phase transition. Above a critical temperature $T_c = \frac{4\varepsilon_1}{\pi k_B}\left(\frac{N}{\zeta_{3/2}}\right)^{2/3}$, all states with $\varepsilon \lesssim k_B T$ are evenly occupied. However, below T_c, the occupation number of the ground state grows like $\frac{\langle N_{\text{ground}} \rangle}{N} = 1 - \left(\frac{T}{T_c}\right)^{3/2}$.

Bose-Einstein condensates include superfluids, superconductors and lasers. They were predicted in 1924 but not seen in the lab until 1995.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 13: Metals

1 Introduction

Fermions satisfy the Pauli-exclusion principle: no two fermions can occupy the same state. This makes fermionic systems act very differently from bosonic systems.

We introduced the non-interacting Fermi gas in Lecture 10, and will continue to discuss it here. We assume the fermions do not interact with each other. This means that the single-particle states of a system are not affected by which states are occupied. So we will fix the states, then fill them from the bottom up, then see what happens when we heat it up. You can think of the possible states as bunch of little shelves and we stack the fermions onto them one-by-one. The exclusion principle implies that once a state is occupied, it cannot be occupied again: the occupancy number of each state is either 0 or 1. This leads to the single-particle grand partition function of state i

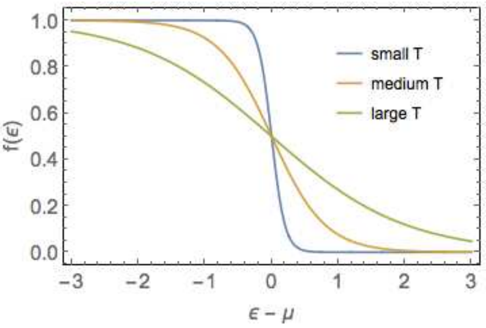
$$\mathcal{Z}_i = \sum_{n=0,1} e^{-\beta(\varepsilon_i - \mu)n} = 1 + e^{-\beta(\varepsilon_i - \mu)} \tag{1}$$

and grand free energy for state i:

$$\Phi_i = -\frac{1}{\beta}\ln \mathcal{Z}_i = -\frac{1}{\beta}\ln[1 + e^{-\beta(\varepsilon_i - \mu)}] \tag{2}$$

The occupation number for state i we then found to be given by

$$\langle n_i \rangle = -\frac{\partial \Phi_i}{\partial \mu} = \frac{1}{e^{\beta(\varepsilon_i - \mu)} + 1} \equiv f(\varepsilon_i) = \tag{3}$$



This is known as the **Fermi-Dirac distribution** or **Fermi function**, $f(\varepsilon)$. As $T \to 0$ the function approaches a step function where all the states below μ are occupied and the staes above μ are unoccupied. The value of μ at $T = 0$ is called the **Fermi energy** and denoted by ε_F. Sometimes the Fermi energy is also called the **Fermi level**.

At any temperature $T > 0$, some fermions will have energies higher than ε_F, and therefore some states with energy less than ε_F will not be occupied. We call the unoccupied states **holes**. You can picture a finite-temperature Fermi gas as having fermions constantly bouncing out of lower-energy states into higher energy states, leaving holes, then other fermions falling out of excited states into those holes. The electron/hole picture gives a useful way to think about finite temperature metals and gives a powerful way to understand the physics of semiconductors, as we will see in this and the next lecture.

We call the temperature corresponding to the Fermi energy the **Fermi temperature**:

$$T_F \equiv \frac{\varepsilon_F}{k_B} \tag{4}$$

This is like how the Debye temperature was related to the Debye energy by $T_D = \frac{\varepsilon_D}{k_B}$ in Lecture 11. Similarly, we can say roughly that for $T < T_F$ quantum effects are important. When the temperature of a system is well below the Fermi temperature, $T \ll T_F$, then $\mu \approx \varepsilon_F$ and the Fermi distribution looks very much like a step function $f(\varepsilon) \approx \theta(\varepsilon_F - \varepsilon)$. In such situations, the lowest states are almost all occupied and we say the system is **degenerate**. So, in a degenerate system, quantum effects are important. Electrons in metals have $T_F > 10000K$, so they are degenerate systems at room temperature and quantum statistical mechanics is necessary to understand their properties.

Generally, Fermi temperatures in metals are very high (for example, T_F for electrons in copper is around 80,000K, as we will see below). So high in fact, that it is generally impossible to heat a system above the Fermi temperature without changing the nature of the system (i.e. melting the metal). Thus, quantum statistics are *always* important for electrons in metals.

The typical picture of a highly degenerate fermionic system is having a relatively small set of energy levels that are relevant, namely those near ε_F. Energy levels with $\varepsilon \ll \varepsilon_F$ require a lot of energy, of order ε_F, to be excited into an empty state, since the first available state is above ε_F. This is nearly impossible when $T \ll T_F$. The set of low-energy states that do not participate in the thermal activity are called the **Fermi sea**. The Fermi sea can be very deep (a lot of states), but its depth is largely irrelevant to the thermal properties of the material. Similarly, the levels with energies much higher than ε_F are impossible to excite when $T \ll T_F$. So only those levels near ε_F matter. Thus degenerate gases have something very close to a symmetry between states above and below ε_F: a symmetry between electrons and holes. We'll use this picture to understand some properties of metals.

2 Free electron gas

The possible single-particle states of the electrons in a free electron gas are the same as for a boson in a box. The allowed wavenumbers are

$$\vec{k} = \frac{\pi}{L}\vec{n}, \quad \vec{n} = \text{triplet of whole numbers} \tag{5}$$

In the non-relativistic limit energies are determined by the usual relation in quantum mechanics

$$\varepsilon_n = \frac{\hbar^2 \vec{k}^2}{2m_e} = \frac{\pi^2\hbar^2}{2m_e L^2} n^2 \tag{6}$$

with $n = |\vec{n}|$. So

$$n = \frac{L}{\pi}\sqrt{\frac{2m_e}{\hbar^2}}\sqrt{\varepsilon}, \quad dn = \frac{L}{2\pi}\sqrt{\frac{2m_e}{\hbar^2}}\frac{d\varepsilon}{\sqrt{\varepsilon}} \tag{7}$$

For electrons, there are two spins. So we compute the density of states via

$$2\sum_n \rightarrow 2 \times \frac{1}{8}\int 4\pi n^2 dn = \frac{V}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\int \sqrt{\varepsilon}\, d\varepsilon \tag{8}$$

As before, the $\frac{1}{8}$ accounting for the modes only counting the first octant (i.e. the $\vec{n}$ are whole numbers) while the integral over the sphere includes all 8 octants. We'll also study the relativistic and ultrarelativistic limits when we discuss white dwarf stars in Lecture 15, but in metals the electrons are generally non-relativistic, as we will check shortly. Thus $g(\varepsilon) = \frac{V}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\sqrt{\varepsilon}$. This scaling $g(\varepsilon) \sim \sqrt{\varepsilon}$ is the same as for the Bose gas from last lecture, only the constant in $g(\varepsilon)$ is different.

Recall that as $T \rightarrow 0$, the Fermi function becomes a step function: $\langle n_\varepsilon \rangle \rightarrow \begin{cases} 1, \varepsilon < \varepsilon_F \\ 0, \varepsilon > \varepsilon_F \end{cases}$. The Fermi energy ε_F is the chemical potential at $T = 0$ or equivalently the energy below which all the states are occupied at $T = 0$. So in particular, the number of of states at $T = 0$ is

$$N = \int_0^\infty \langle n_\varepsilon \rangle g(\varepsilon) d\varepsilon \xrightarrow{T=0} \frac{V}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\int_0^{\varepsilon_F} \sqrt{\varepsilon}\, d\varepsilon = \frac{V}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\frac{2}{3}\varepsilon_F^{3/2} \tag{9}$$

Therefore

$$\boxed{\varepsilon_F = \frac{\hbar^2}{2m_e}\left(3\pi^2\frac{N}{V}\right)^{2/3}} \tag{10}$$

Using this, it's convenient to write the density of states as

$$g(\varepsilon) = \frac{V}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\sqrt{\varepsilon} = \frac{3N}{2\varepsilon_F^{3/2}}\sqrt{\varepsilon} \tag{11}$$

The **Fermi temperature** is then

$$T_F = \frac{\varepsilon_F}{k_B} \tag{12}$$

This temperature is the characteristic temperature above which electrons can be readily excited into unoccupied states. Typically $T \ll T_F$ and so the probability of an electron getting excited into a unoccupied state (i.e. $\varepsilon \gtrsim \varepsilon_F$) is very small: $P \sim e^{-\frac{\varepsilon}{k_B T}} \approx e^{-\frac{T_F}{T}} \ll 1$.

The energy of the Fermi gas at zero temperature is

$$E_0 = E(T=0) = \frac{3N}{2\varepsilon_F^{3/2}}\int_0^{\varepsilon_F} \varepsilon\sqrt{\varepsilon}d\varepsilon = \frac{3}{5}N\varepsilon_F \tag{13}$$

So the average energy of each electron at zero termperature is $\langle\varepsilon\rangle = \frac{E_0}{N} = \frac{3}{5}\varepsilon_F$.

An important concept in electron gases, or Fermi gases in general, is that of **degeneracy pressure**. Pressure is defined as $P = -\frac{\partial E}{\partial V}$. Normally we associate pressure with the kinetic motion of molecules bombarding the walls of a container. This exerts a force on the walls, so that if the volume increases, the energy would go down. Because the force is due to thermal motion, this ordinary pressure is absent at $T=0$. Degeneracy pressure, on the other hand, is a contribution to $-\frac{\partial E}{\partial V}$ in Fermi gases that persists even at $T=0$.

The way to understand degeneracy pressure is by thinking about how the energy levels change as the volume is shrunk, for example under isothermal compression. For a classical gas, isothermal compression does not change the energy of the gas (i.e. $E = C_V k_B T$ is volume independent). What it does is increase the density of molecules, so the number of collisions on the wall of a container goes up, and hence the pressure increases ($P = \frac{N}{V}k_B T$). For a Bose gas, as the volume shrinks, all the energy levels go up, but the occupancies of different levels adjust so that the total energy is the same. Thus, Bose gases behave essentially like classical gases when compressed. For a degenerate Fermi gas, when the volume goes down, the occupancies of states cannot adjust. Instead, the energy levels go up and the total energy simply increases. This contribution to $-\frac{\partial E}{\partial V}$ is called degeneracy pressure.

To see that this pressure really is different from kinetic motion on container walls, consider the $T=0$ limit. As we saw before, $T=0$ is often a good approximation when $T \ll T_F$. Since we have been working at $T=0$ already, all we have to do is differentiate Eq. (13):

$$P = -\frac{\partial E}{\partial V} = -\frac{\partial}{\partial V}\left[\frac{3}{5}N\varepsilon_F\right] = -\frac{3}{5}N\frac{\partial}{\partial V}\left[\frac{\hbar^2}{2m_e}\left(3\pi^2\frac{N}{V}\right)^{2/3}\right] = \frac{2}{3}\frac{E_0}{V} \tag{14}$$

In this limit, the classical pressure goes to zero but the degeneracy pressure persists. Degeneracy pressure is essential to understand white dwarfs and neutron stars, as we'll see in Lecture 15.

3 Sommerfeld free electron model for metals

Sommerfeld proposed that the electrons in a metal can be described by a free electron gas. The basic idea is that each atom in a metal has some number of loosely bound electrons that it contributes to the gas. The alkali metals (first column of periodic table, e.g. potassium) are **monovalent**, contributing one electron each. Some metals contribute more than one electron per atom (we'll see why in Lecture 15). The weakly bound electrons, called the **valence electrons**, are assumed to be free at leading order, so they just bounce around like particles in a box, with the size of the box determined by the size of the metal. This is called the **free electron model**, Before getting into the predictions of the free electron model, it's worth a few comments about why on earth this model should be at all reasonable. At first thought, it seems crazy that we could both ignore the attractive interactions of the electrons to the positively charged nuclei and ignore the repulsive Coulomb interactions among the electrons themselves. On second thought, however, we realize that because of the exclusion principle, electrons like to stay away from each other, so the electron-electron forces may indeed be small. As for the nuclei, we can't forget that metals have not only fairly high nuclear charge ($Z = 29$ for copper), but also a whole lot of other electrons more tightly bound to the nucleus than the electron being contributed to the electron gas. Thus the valence electrons are forced to be far away from the nuclei (by the exclusion principle) and moreover, the other electrons screen the nuclear charge. With these considerations, the free electron model at least does not seem horribly wrong, and we can start to study it.[1]

First, let's compute the Fermi energy ε_F and fermi temperature $T_F = \frac{\varepsilon_F}{k_B}$ in the free electron model. Consider copper, which has a density $\rho = 9.0\frac{g}{\text{cm}^3}$ and an atomic weight of 63.6 $\frac{\text{g}}{\text{mol}}$, so the number density of atoms is $n = 9.0\frac{g}{\text{cm}^3}/63.6\frac{g}{\text{mol}} = 0.14\frac{\text{mol}}{\text{cm}^3}$. Treating copper as monovalent (Cu is [Ar]$3d^{10}4s^1$ as we explain next lecture), this is also the number density of valence electrons. So $n_e = 0.14\frac{\text{mol}}{\text{cm}^3} = 8.4 \times 10^{28}\frac{\text{electrons}}{m^3}$. The Fermi energy is then

$$\varepsilon_F = \frac{\hbar^2}{2m_e}(3\pi^2 n)^{2/3} = 7\,\text{eV} \tag{15}$$

Note that this is much less than the electron rest mass $m_e c^2 = 511\text{keV}$, so the non-relativistic limit is justified. The Fermi energy is also much higher than $k_B T = 25\,\text{meV}$ at room temperature. The corresponding Fermi temperature is

$$T_F = \frac{\varepsilon_F}{k_B} = 82000\,K \tag{16}$$

Compared to room temperature $T_0 = 298K$. We see that $\frac{T_F}{T_0} \approx 270$. Since the Fermi temperature is much higher than room temperature, copper at room temperature is highly degenerate and quantum effects dominate.

Let's next look at the degeneracy pressure in a metal. From Eq. (14) we can write

$$P = \frac{2}{3}\frac{E_0}{V} = \frac{2}{5}\frac{N}{V}\varepsilon_F = \frac{2}{5}\frac{N}{V}k_B T_F \tag{17}$$

Now the Fermi temperature in metals is always well above room temperature, as in Eq. (16), $\frac{T_F}{T_0} \approx 270$. The density of metals is also typically much higher than in gases, by a factor of 1000 or so. Thus the degeneracy pressure is of order $1000 \times 270 \times 1\,\text{atm} = 10^5\,\text{atm}$ – enormous! The electrons really *really* want to get out of the metal; in a larger volume, their energy levels would go down and the degeneracy pressure would be relieved. Fortunately, the electrons cannot leave since the outward force is compensated by strong attractive forces holding the metal together. Thus we are not going to be able to measure degeneracy pressure directly in a metal without understanding those (rather complicated) attractive foces.

Instead of pressure, a more reasonable quantity to look at is the bulk modulus, $B = -V\left(\frac{dP}{dV}\right)_T$. The bulk modulus measures how much a solid will compress when a pressure is applied. It tells how springy a material is. Indeed, as mentioned in Lecture 11, the bulk modulus determines the speed of sound in solids, as $c_s = \sqrt{\frac{dP}{d\rho}} = \sqrt{\frac{B}{\rho}}$ with ρ the density. Thus B plays the role of the spring constant k in Hooke's law. In solids, this constant should be determined by the restoring force when atoms are moved, which is due to the electrons. Thus we expect that B can be computable from a good model of electrons in solids.

[1] A more formal justification of the free electron model is given by **Fermi liquid theory**. Basically, Fermi liquid theory uses renormalization-group methods to show that when interactions are included in the free electron model, the electron mass changes to an "effective mass" which is slightly different from m_e, but otherwise the interactions can be ignored.

For a free electron gas at $T=0$,

$$B=-V\frac{dP}{dV}=V\frac{d^2}{dV^2}\left[\frac{3}{5}N\frac{\hbar^2}{2m_e}\left(3\pi^2\frac{N}{V}\right)^{2/3}\right]=\frac{2}{3}\frac{N}{V}\varepsilon_F=\frac{10}{9}\frac{E_0}{V}=\frac{5}{3}P \tag{18}$$

Let's compute this for potassium. Solid potassium has an valence electron number density of $n=1.4\times10^{28}\frac{1}{m^3}$. This gives a Fermi energy of $\varepsilon_F=3.4\times10^{-19}J$ and a bulk modulus of $B=3.2\times10^9\frac{N}{m^2}$. The experimental value is $B=3.1\times10^9\frac{N}{m^2}$. Thus our prediction from the free electron gas is in excellent agreement with the measured value. Equivalently using the density of potassium is $\rho=856\frac{\text{kg}}{m^3}$ we predict a speed of sound $c_s=\sqrt{\frac{B}{\rho}}=1933\frac{m}{s}$; the measured value is $c_s=2000\frac{m}{s}$. Apparently, the elasticity of metals is determined by the degeneracy pressure of the electrons within it. This is rather a profound conclusion: when I squeeze a penny in my hand, the penny is fighting back not with the classical Coulomb force, but with the quantum Pauli exclusion principle.

Next, we want to compute the temperature dependence of the energy and heat capacity. To do that, we need to first eliminate the chemical potential. As with bosons, we can trade μ for N by integrating over the density of states. For fermions, using Eq. (11), we find

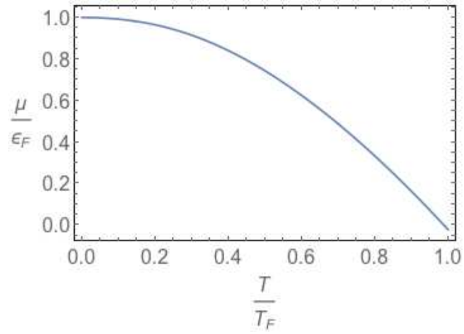
$$N=\int_0^\infty\frac{g(\varepsilon)}{e^{\beta(\varepsilon-\mu)}+1}d\varepsilon=\frac{3N}{2\varepsilon_F^{3/2}}\left(-\frac{\sqrt{\pi}}{2\beta^{3/2}}\right)\text{Li}_{3/2}(-e^{\beta\mu}) \tag{19}$$

where $\text{Li}_{3/2}$ is a polylogarithm function. Thus, we get an equation relating μ to ε_F at a given T:

$$-\frac{4}{3\sqrt{\pi}}(\beta\varepsilon_F)^{3/2}=\text{Li}_{3/2}(-e^{\beta\mu}) \tag{20}$$

Unfortunately, the polylogarithm of an exponential is really hard to work with. So let us start by examining it numerically. Solving this transcendental equation numerically we find a perfectly smooth function $\mu(\varepsilon_F,T)$:

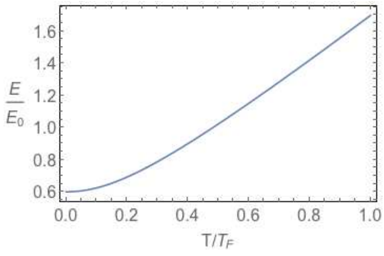
$$\mu(\varepsilon_F,T)= \tag{21}$$



Plugging in some numbers: at room temperature for copper $\frac{T}{T_F}=\frac{298}{82000}=0.0036$ and Eq. (20) gives $\mu=0.9999\varepsilon_F$. At the melting temperature of copper, $T_{\text{melt}}=1358K$, Eq. (20) gives $\mu=0.9996\varepsilon_F$. Thus in the entire range for which copper is a solid, $T=0$ to $T=T_{\text{melt}}$ μ only varies from $1.0\varepsilon_F$ to $0.9997\varepsilon_F$. This justifies the common practice of using the chemical potential and Fermi energy interchangeably when studying metals.

We can also compute the energy of the free electron gas:

$$E=\frac{3N}{2\varepsilon_F^{3/2}}\int_0^\infty\frac{\varepsilon\sqrt{\varepsilon}}{e^{\beta(\varepsilon-\mu)}+1}d\varepsilon=\frac{-15\sqrt{\pi}E_0}{8(\beta\varepsilon_F)^{5/2}}\text{Li}_{5/2}(-e^{\beta\mu})= \tag{22}$$



where $E_0 = \frac{3}{5} N\varepsilon_F$ as in Eq. (13). Thus the total energy, like the chemical potential, is a smooth function of T the nasty polylogarithm in the analytic formula.

Before doing a careful small T expansion, let's think a little about what we expect. The key to small T behavior of the energy is understanding how many states can be excited. Because N is so large, the energy levels near the Fermi surface are essentially continuous: there are generally *a lot* of electrons and a lot of states. Most of the electrons are in the Fermi sea, and it takes a lot of energy to excite an electron out of deep in the sea. Since $k_B T \ll T_F$ for any metal, only the highest energy electrons have any hope of being excited. All of the thermal activity comes from a small number of electrons close to the Fermi surface moving to states slightly above it. The electrons deep in the sea would require energies of order $\varepsilon_F \gg k_B T$ to move into a free state, for which there are exponentially small probabilities, $P \sim e^{-T/T_F}$. In other words, the states that are involved in the thermal activity are essentially only those with energies $\varepsilon_F - k_B T \lesssim \varepsilon \lesssim \varepsilon_F + k_B T$

At a temperature T some states will be excitable and gain energy $\sim k_B T$. So the total energy should be $E \approx E_0 + N_{\text{excitable}} k_B T$. This scaling is the essential content of the equipartition theorem. As a check on this logic, for a *bosonic* gas, we found the density of states to scale like $g(\varepsilon) \sim \varepsilon^2$ so the total number of excitable states are those with $\varepsilon < k_B T$ which is $N_{\text{excitable}} \sim \int_0^{k_B T} g(\varepsilon) d\varepsilon \sim T^3$. Thus $E \sim T^4$ for bosons, as we found through explicit calculation in the Debye model. For a fermionic gas, only those states with energies ε between $\varepsilon_F - k_B T$ and $\varepsilon_F + k_B T$ can be excited, so $N_{\text{excitable}} \sim k_B T$ and therefore $E \approx E_0 + k_B^2 T^2$. Plugging in E_0 from Eq. (13) we get

$$E = \frac{3}{5} N\varepsilon_F + a N \frac{k_B^2 T^2}{\varepsilon_F} + \cdots \tag{23}$$

for some a. The factor of ε_F in the denominator of the T^2 term was included by dimensional analysis – it is the only scale that could possibly appear. Eq. (23) is in qualitative agreement with Eq. (22). This heuristic argument is helpful to understand the quadratic dependence on T, however, it cannot give us the coefficient a. We now turn to computing this coefficient a.

3.1 Small T expansion

In this section, we will discuss how to expand the chemical potential and total energy at small T. These quantities are determined by Eqs. (19) and (22):

$$N = \int_0^\infty g(\varepsilon) f(\varepsilon) d\varepsilon, \qquad E = \int_0^\infty \varepsilon g(\varepsilon) f(\varepsilon) \tag{24}$$

with $g(\varepsilon) = \frac{3N}{2\varepsilon_F^{3/2}} \sqrt{\varepsilon}$ and $f(\varepsilon) = \frac{1}{e^{\beta(\varepsilon-\mu)}+1}$ the Fermi distribution. Although we can do the integrals analytically, the resulting polylogaritms are not easy to expand around $T = 0$. Of course, you can just look up the polylogarithm limits, but we will learn more about the system by finessing the small T expansion using some tricks developed by Sommerfeld. We'll start with the chemical potential, then do the energy. Don't worry too much if you don't follow everything here, we'll do the calculation in a slicker way in Section 3.2. The reason I include this calculation is that it introduces some useful tricks which you can add to your toolbox to use in other contexts.

The first trick is to integrate by parts:

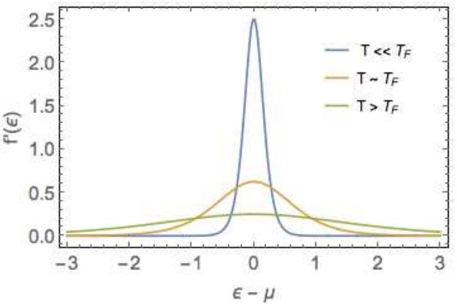
$$N = \frac{3N}{2\varepsilon_F^{3/2}} \int_0^\infty f(\varepsilon) \sqrt{\varepsilon} d\varepsilon = \underbrace{\frac{N}{\varepsilon_F^{3/2}} \varepsilon^{3/2} f(\varepsilon) \Big]_0^\infty}_{=0} - \frac{3N}{2\varepsilon_F^{3/2}} \int_0^\infty [f'(\varepsilon)] \left[\frac{2}{3} \varepsilon^{3/2}\right] d\varepsilon \tag{25}$$

The first term on the right vanishes at both boundaries, so it is zero. Then we have

$$1 = \frac{1}{\varepsilon_F^{3/2}} \int_0^\infty [-f'(\varepsilon)] \varepsilon^{3/2} d\varepsilon \tag{26}$$

The derivative of the Fermi distribution is a sharply peaked Guassian-like function centered on μ with width $\sqrt{k_B T}$:

$$-f'(\varepsilon) = \frac{\beta}{2}\frac{1}{1+\cosh(\beta(\varepsilon-\mu))} = \tag{27}$$



The second trick is to observe (as we can from the plot) that when $T \ll T_F$, the support of $f'(\varepsilon)$ is for energies $|\varepsilon - \mu| \lesssim k_B T$. In this region we can Taylor expand the function multiplying $f'(\varepsilon)$ in Eq. (26) around $\varepsilon = \mu$ giving

$$1 = \frac{1}{\varepsilon_F^{3/2}} \int_0^\infty [-f'(\varepsilon)] \left[\mu^{3/2} + \frac{3}{2}\sqrt{\mu}(\varepsilon - \mu) + \frac{3}{8\sqrt{\mu}}(\varepsilon-\mu)^2 + \cdots \right] d\varepsilon \tag{28}$$

The third trick is to note that since $f'(\varepsilon)$ is exponentially suppressed at small ε we can extend the lower limit of integration to $-\infty$. Once the integral is from $-\infty$ to ∞ the integral is much easier to do. Changing variables to $x = \beta(\varepsilon - \mu)$ makes the integrals doable by Mathematica

$$1 = \left(\frac{\mu}{\varepsilon_F}\right)^{3/2} \int_{-\infty}^{\infty} \frac{dx}{2+2\cosh(x)} \left[1 + \frac{3}{2\beta\mu}x + \frac{3}{8\beta^2\mu^2}x^2 + \cdots \right] = \left(\frac{\mu}{\varepsilon_F}\right)^{3/2} \left[1 + \frac{\pi^2}{8\beta^2\mu^2} + \cdots \right] \tag{29}$$

Now we can solve perturbatively for μ. Substituting $\mu = \varepsilon_F + aT^2 + \cdots$ into Eq. (29) and expanding to order T^2 leads us directly to

$$\mu = \varepsilon_F \left[1 - \frac{\pi^2}{12}\frac{k_B^2}{\varepsilon_F^2}T^2 + \cdots \right] \tag{30}$$

As a check on this, plugging in $T = 1358K$ (the melting point of copper), this gives $\mu = 0.9997\varepsilon_F$ in excellent agreement with our exact numerical calculation.

We can do a similar expansion for the energy. The only difference is that instead of integrating the Fermi function against $g(\varepsilon)$ we integrate against $\varepsilon g(\varepsilon)$. After integrating by parts and dropping the boundary term, the equation for the total energy, Eq. (22) becomes

$$E = \frac{N}{\varepsilon_F^{3/2}} \int_0^\infty [-f'(\varepsilon)] \left[\frac{3}{5}\mu^{5/2} + \frac{3}{2}\mu^{3/2}(\varepsilon-\mu) + \frac{9}{8}\sqrt{\mu}(\varepsilon-\mu)^2 + \cdots \right] d\varepsilon \tag{31}$$

$$= N\left(\frac{\mu}{\varepsilon_F}\right)^{3/2} \mu \int_{-\infty}^{\infty} \frac{dx}{2+2\cosh(x)} \left[\frac{3}{5} + \frac{3}{2\beta\mu}x + \frac{9}{8\beta^2\mu^2}x^2 + \cdots \right] \tag{32}$$

$$= N\left(\frac{\mu}{\varepsilon_F}\right)^{3/2} \mu \left[\frac{3}{5} + \frac{3\pi^2}{8\beta^2\mu^2} + \cdots \right] \tag{33}$$

Substituting the form of μ in Eq. (30) we find that to second order in T

$$E = \frac{3}{5}N\varepsilon_F + N\frac{\pi^2 k_B^2 T^2}{4\varepsilon_F} + \cdots \tag{34}$$

Thus the unknown coefficient a in Eq. (23) is now known: $a = \frac{\pi^2}{4}$.

What did we learn from all these tricks? First of all, we got the answer $a=\frac{\pi^2}{4}$ without having to expand $\mathrm{Li}_{5/2}(e^{-\beta\mu})$. The second thing we learned was that the expansion was difficult mainly because of the $\varepsilon\geqslant 0$ constraint – we couldn't just integrate $\int_{-\infty}^{\infty}\varepsilon f(\varepsilon)d\varepsilon$ from the start since the integrand blows up at $\varepsilon\to\infty$, so we had to do this integration by parts trick to get the integral in a form where we could extend the limits of integration. The physical reason that the extension works is that the very low energy states and the very high energy states are irrelevant, as we anticipated. In fact, there's a more physical and less mathematical way to isolate these relavant states called the electron/hole picture that we will discuss next.

3.2 Electrons and holes

The electron and hole picture gives a quick way to understand the leading temperature dependence of the energy. Since $T\ll T_F$ for any metal, or equivalently, $k_BT\ll\varepsilon_F$, it is very hard to excite electrons deep in the Fermi sea. As mentioned by Eq. (23) only states within k_BT of $\varepsilon_F\sim\mu$ can contribute to the energy shift $E-E_0$ and to the heat capacity. Thus overall, both excited states (above μ) and states below μ that do not have electrons in them (holes) are rare.

The idea of the electron/hole picture is to replace the original Fermi gas, with its inert Fermi sea, by a symmetric thermal system of electron excitations and the holes. Of course, in reality the electrons and holes are related, since each electron excitation comes from a hole. However, with a large N of electrons, we can treat the electrons and holes as their own statistical mechanical system and ignore the (very small) correlations among them.

To construct the gas, we first define the zero-point of energy in the new system to be at the fermi level. Now, each excited electron in the original system in a state with energy $\varepsilon>\mu$ contributes positive energy $\Delta=\varepsilon-\mu>0$ to this new gas. Each hole, from a state of energy $\varepsilon<\mu$ in the orignal system, represents the *absence* of a state with negative energy $-\Delta=\varepsilon-\mu<0$. Thus each hole contributes *positive* energy $\Delta=\mu-\varepsilon$ to the new gas. The positive energy can be understood if we think about an excitation of the original gas from energy $\mu-\Delta_h$ to $\mu+\Delta_e$. This excitation contributes a total of $\Delta_h+\Delta_e$ to the energy, with Δ_e coming from the electron and Δ_h from the hole.

Recall that the probabilities of occupation for each state are independent in a Fermi gas (this is the main reason we use the grand canonical ensemble at fixed μ rather than the canonical ensemble at fixed N). So the probability of finding an electron with excitation energy $\Delta_e=\varepsilon-\mu$ is

$$f_e(\Delta_e)=f(\varepsilon)=f(\mu+\Delta_e) \tag{35}$$

where $f(\varepsilon)=\frac{1}{e^{\beta(\varepsilon-\mu)}+1}$ is the Fermi function.

What is the probability of finding a hole of energy Δ_h? Since any state has either a hole or an electron, the probability of finding a state of energy ε to be empty is $1-f(\varepsilon)$. Since $\varepsilon=\mu-\Delta_h$ for holes, the probability of finding a hole with energy Δ_h is then

$$f_h(\Delta_h)=1-f(\mu-\Delta_h) \tag{36}$$

Now a useful property of the Fermi distribution $f(\varepsilon)=\frac{1}{e^{\beta(\varepsilon-\mu)}+1}$ is that it is symmetric around its midpoint at $\varepsilon=\mu$. More precisely,

$$f(\mu+\Delta)=1-f(\mu-\Delta) \tag{37}$$

And therefore

$$\underbrace{f_h(\Delta)}_{\text{prob. of hole with energy }\Delta>0}=1-f(\mu-\Delta)=f(\mu+\Delta)=\underbrace{f_e(\Delta)}_{\text{prob. of electron with energy }\Delta>0} \tag{38}$$

Thus holes and electrons have identical probability distributions!

This gives us an easy way to calculate the total energy: we just calculate the energy of the electron excitations and double. The electron excitation contributions are like the original electron contributions but shifted to $\mu=0$. In this region, the density of states is approximately constant

$$g(\mu+\Delta)\approx g(\mu-\Delta)\approx g(\varepsilon_F)=\frac{3N}{2\varepsilon_F} \tag{39}$$

So, the contribution to the total energy from the electron excitations about μ is

$$E_e = \int_\mu^\infty (\varepsilon - \mu) g(\varepsilon) f(\varepsilon) d\varepsilon \approx g(\varepsilon_F) \int_0^\infty \frac{\Delta}{e^{\beta\Delta}+1} d\Delta = \left(\frac{3N}{2\varepsilon_F}\right) \frac{\pi^2}{12\beta^2} = \frac{\pi^2 N}{8\varepsilon_F} k_B^2 T^2 \tag{40}$$

The hole contribution is identical

$$E_h = \int_0^\mu (\mu - \varepsilon) g(\varepsilon) f(\varepsilon) d\varepsilon \approx g(\varepsilon_F) \int_0^\infty \frac{\Delta}{e^{\beta\Delta}+1} d\Delta = \left(\frac{3N}{2\varepsilon_F}\right) \frac{\pi^2}{12\beta^2} = \frac{\pi^2 N}{8\varepsilon_F} k_B^2 T^2 \tag{41}$$

The only additional approximation we must do for holes is take the upper limit of integration on Δ to ∞, since $\varepsilon > 0$ means $\Delta_h < \mu$. Due to the exponential expression at large Δ, taking the upper limit to ∞ has essentially no effect on the integral.

Thus the total energy is

$$E = E_0 + E_e + E_h = E_0 + \frac{\pi^2 N}{4\varepsilon_F} k_B^2 T^2 \tag{42}$$

This is in perfect agreement with Eq. (34). (To get subleading terms in Eq. (34) you can include subleading terms in Eqs. (39)-(41).)

This calculation was certainly simpler than Sommerfeld's from Section 3.1 (although you may not have believed this one without that one as a check). The two are actually equivalent and make the same approximations. In this version, we sidestepped the fact that the energy of the holes is bounded by $\Delta < \varepsilon_F$ by setting the hole contribution equal to electron contribution, and integrating to ∞. In the Sommerfeld calculation, we integrated by parts and dropped the boundary terms to avoid the $\varepsilon_F < \infty$ limit. The electron/hole picture is nice not only because it is simpler but because it gives us a powerful language for discussion fermionic systems. This language is particularly useful for solid-state physics, as we'll see in the next lecture.

3.3 Heat capacity

Having computed the temperature dependence of the energy of a free electron gas, it is now trivial to find the heat capacity

$$C_V^{\text{electons}} = \frac{\partial E}{\partial T} = \frac{\pi^2}{2} N k_B \frac{T}{T_F} + \cdots \tag{43}$$

We computed this as the leading term in the $\frac{T}{T_F}$ expansion, but since $T_F \gtrsim 50000K$ for metals, this expansion can work for any temperature a metal can have – up to the melting point. In particular, it holds even if $T \gg 298K$.

We can compare this to the heat capacity due to phonons (oscillations of the atoms), for example through the law of Dulong and Petit: $C_V^{\text{phonons}} \approx 3Nk_B$. Recall that we derived this behavior in Lecture 11 as the high-temperature limit of the Debye model. Typical Debye temperatures for metals are around 100 K, so $T \gg T_D \sim 100K$ is reasonable at room temperature. Note that high temperature for phonons ($T \gg T_D$) is consistent with low temperature for electrons ($T \ll T_F$), and so we can use both limits to see

$$\frac{C_V^{\text{electrons}}}{C_V^{\text{phonons}}} = \frac{\pi^2}{6} \frac{T}{T_F} \ll 1 \tag{44}$$

This explains why the electrons do not contribute appreciably to the total heat capacity $C_V = C_V^{\text{phonons}} + C_V^{\text{electrons}} \approx C_V^{\text{phonons}}$. Historically, there was a puzzle for some time about why the law of Dulong and Petite should be correct, since the equipartition theorem says that electrons should contribute to the heat capacity just like the atoms. We now understand that because of Fermi-Dirac statistics the valence electrons have very high energy compared to room temperature, so only a small fraction of the electrons can be thermally excited (those near the Fermi level). Thus, at room temperature the electronic contribution is very small and this historical puzzle is thereby resolved by quantum statistics.

At very low temperatures $T \ll T_D$, the law of Dulong and Petite no longer applies. In the Debye model we found that for $T \ll T_D$

$$C_V^{\text{Debye}} = \frac{12\pi^4}{5} N k_B \left(\frac{T}{T_D}\right)^3 \tag{45}$$

where the formula for the Debye temperature is

$$T_D = \frac{\hbar \omega_D}{k_B} = \frac{\hbar}{k_B} c_s \left(6\pi^2 \frac{N}{V} \right)^{1/3} \tag{46}$$

with c_s the speed of sound. Thus the full heat capacity at low temperature is the sum of the electronic and phononic parts

$$C_V = C_V^{\text{electrons}} + C_V^{\text{phonons}} = \frac{\pi^2}{2} N k_B \frac{T}{T_F} + \frac{12\pi^4}{5} N k_B \left(\frac{T}{T_D} \right)^3 + \cdots \tag{47}$$

For copper $T_D = 315K$ and $T_F = 80,000K$. Thus we have to go to very low temperatures to see the electronic contribution. The two contributions are comparable when

$$\frac{T}{T_F} \sim \left(\frac{T}{T_D} \right)^3 \quad \Rightarrow \quad T \sim \sqrt{\frac{T_D^3}{T_F}} \tag{48}$$

For copper this is $T \sim 19K$.

We can see the electronic contribution by measuring the heat capacity at small T and plotting $\frac{C_V}{T}$ as a function of T^2:

$$\frac{C_V}{T} = \frac{\pi^2}{2} N k_B \frac{1}{T_F} + \frac{12\pi^4}{5} N k_B \frac{1}{T_D^3} (T^2) \tag{49}$$

The y-intercept of such a plot gives the electronic contribution and the slope gives the phonon contribution. Here is such a plot for copper:

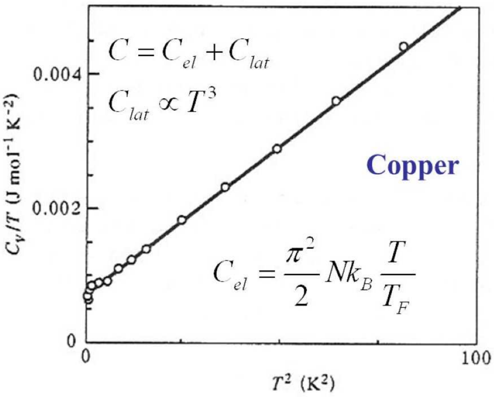


Figure 1. Heat capacity of copper at low temperatures compared to the prediction from the Sommerfeld free electron gas model.

Measurements like this demonstrate an electronic contribution in good agreement with the prediction from the Sommerfeld free electron model.

An obvious limitation of the free electron model is that it is the same model for any metal. Thus it cannot tell us what is a conductor, what is an insulator, or account for any of the interesting structure of the elements in the periodic table. To do that we need to make some improvements.

4 Nearly free electron gas

To improve on the free electron model, we can consider how electrons become free. If the atoms are far apart, then the electrons would have been in orbitals around their atom. There is thus some kind of attractive potential $V(r)$ trying to keep each electron bound to one atom. As atoms move closer together, their potentials start to overlap. In the limit that the potentials overlap a lot, the sum of the potentials can be very weakly position dependent:

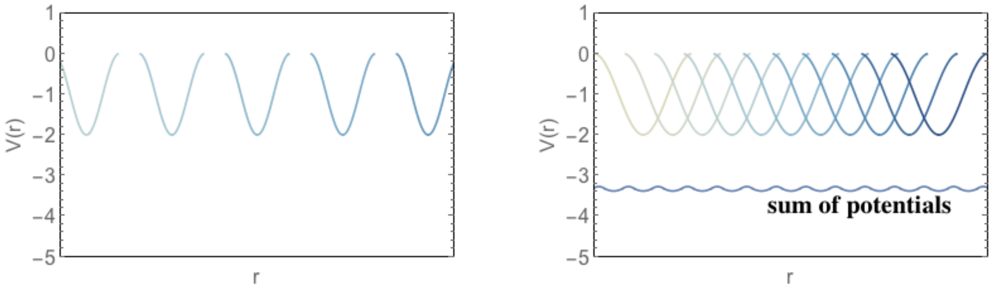
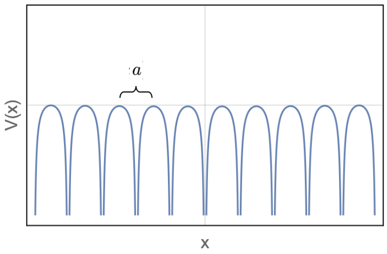


Figure 2. When atoms are far apart (left), the atomic potentials do not overlap. When atoms become close (right), potentials overlap with the net result of a nearly free potential: the sum of the potentials is nearly flat.

Most metals have a very regular lattice strcture. More precisely, these potentials are periodic, meaning that there exists a vector $\vec{a}$ such that $V(\vec{x}+\vec{a}) = V(\vec{x})$. It is interesting therefore to consider a weak periodic potential as a model of the atomic system. This is known as the **nearly free electron model**. It is hard to calculate properties of periodic potentials in three dimensions exactly, but in one dimension we can actually find some exact solutions that are qualitatively similar to 3D systems and to real metals.

In one dimension, we are interested in finding the energies of a system with a periodic potential. By periodic potential, we mean that the full potential is the sum over potentials $V_1(x)$ representing atoms, shifted by the lattice spacing a. That is, the potential has the form

$$V(x) = \sum_{n=-\infty}^{\infty} V_1(x-na) = \tag{50}$$



Summing n from $-\infty$ to ∞ is an approximation, but allows us to find a nice simple form for the result.

Infinite periodic potentials have been studied in great detail and a lot of general results are known about them. A general result known as **Bloch's theorem** is that energy eigenstates can be written as

$$\psi_k(x) = e^{ika} u_k(x) \tag{51}$$

where a is the lattice spacing, k is the wavenumber and $u_k(x)$ is a function with the periodicity of the lattice, i.e. $u_k(x+a) = u_k(x)$ in this case. These solutions are known as **Bloch waves**. Bloch waves are like smooth plane waves e^{ikx} at scales $x \gtrsim a$ but have a bumpier structure, given by the Fourier transform of $u_k(x)$ at length scales less than a. Bloch's theorem holds in any number of dimensions.

Plugging Eq. (51) into the Schrodinger equation $(\frac{\hbar^2}{2m}\vec{\nabla}^2\psi + V(x)\psi = \varepsilon\psi)$, you can show that for a given wavenumber k (the phase in Eq. (51)), the energies ε satisfy

$$\frac{\cos\left(\sqrt{\frac{2m_e\varepsilon}{\hbar^2}}\,a + \delta_k\right)}{|t_k|} = \cos(ka) \tag{52}$$

This is an implicit formula for the dispersion relation $\varepsilon(k)$ of the system. The quantities $|t_k|$ and δ_k are the magnitude and phase of the transmission coefficient $t_k = |t_k|e^{i\delta_k}$, which can also be calcualted from $V(x)$ (t_k is the transmission coefficient for scattering a wave with wavenumber k past $V_1(x)$). The appearance of t can be understood because the waves at a given k have to scatter past each individual potential in a regular way to become energy-eigenstate standing waves. We're not going to derive Eq. (52), but the derivation can be found in many solid-state physics books (e.g. Hook and Hall Chapter 4 or Ashcroft and Mermin Chapter 8).

To be concrete, consider a special case where the potentials are square wells. This case is known as the **Kronig-Penney** model and an exact solution is known. The solution is particularly simple in the limit where the width of the wells goes to zero and the strength of the wells to infinity so that each potential is a δ-function, i.e. $V_1(x) = \lambda\delta(x)$. In this case, the magnitude and phase of the transmission coefficient are given by

$$|t_k| = \cos\delta_k = \frac{1}{\sqrt{1 + \frac{m_e^2\lambda^2}{\hbar^4 k^2}}} \tag{53}$$

You can find the derivation of the Kronig-Penney model in many places, including Wikipedia. We're not going to derive it here.

Let's look at the the energies of the Kronig-Penney model as a function of k. We do this by finding solutions to Eq. (52) using the expressions in Eq. (53) for $|t_k|$ and δ_k. As usual, we begin by finding the solutions numerically. The result looks like this:

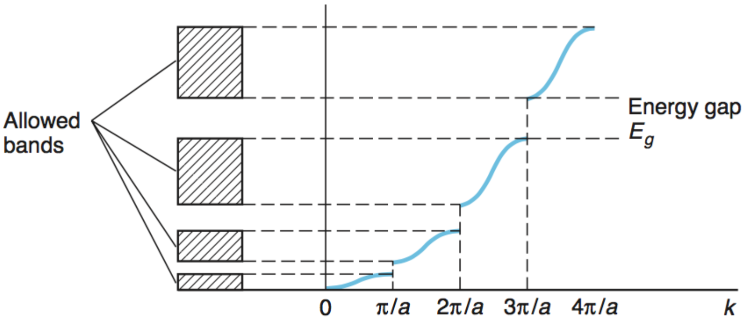


Figure 3. Allowed and forbidden region in the Kronig-Penney model. Only have the bands, for $k > 0$ are shown. There are mirror image bands for $k < 0$.

In this figure, you can see that the energies break up into a series of **bands**. The energy difference between the top of one band and bottom of the next is known as the **bandwidth**. As k is increased, there are certain energies that never appear. These are energies in the **band gap**. The Kronig-Penney model is the simplest exactly solvable model where the band structure appears. Bands are critical to understanding properties of metals and semiconductors.

We can see fairly easily from the general result in Eq. (52) why there are bands, i.e. why there are not solutions for every ε. Because t is a trasmission coefficient, $|t| < 1$ and thus the cosine on the left is bigger than the one on the right. So when the left cosine is bigger than $|t|$ then there can be no solution. This happens around the regions where $k = \frac{\pi}{a}, \frac{2\pi}{a}, \frac{3\pi}{a}, \cdots$ The band gaps are centered around these regions. The wavenumbers $k = \frac{\pi}{a}$ correspond to periodic eigenfunctions. If $\lambda = 0$ (no potential), these are just the normal particle-in-a-box modes. If $\lambda < 0$ the energy of these modes is lowered. If $\lambda > 0$ the energy is raised. Either way, there is no longer an energy corresponding to these modes with period of the lattice, hence the band gap.

As the potential is removed, $V \to 0$, we should recover the free-particle solutions. You can actually see this from the figure: as the gaps are removed, the bands merge into a parabola. The parabola is none other than the dispersion relation for the free electron gas, Eq (6): $\varepsilon = \frac{\hbar^2 k^2}{2m_e}$. To see this analytically, note that with the potential removed, the transmission coefficient is $t = 1$. From Eq. (52) we can see that when $|t| = 1$ and $\delta = 0$ then $\sqrt{\frac{2m_e E}{\hbar^2}} = k$, agreeing with the dispersion relation for the free electron gas.

Although this model was only one-dimensional, periodic potentials predict bands also in three dimensions, and bands are observed experimentally in metals. What is critically important in determining properties of the solid is whether the Fermi level is in the middle of a band or in the band gap. For example, suppose each potential corresponds to one atom and each atom donates one electron to the nearly-free electron gas. In this case, there are N electrons and the maximum k allowed (in 1D) is $k = N\frac{\pi}{L}$ with L the size of the system. Using $L = Na$ we see $k \leqslant \frac{\pi}{a}$, which, matching onto Figure 3 would be the highest occupied electron level right at the edge of the band. Remember though that electrons have two spin states, so each level is actually two-fold degenerate. This means that if each atom donates one electron, the lowest band will actually be only half filled. With a half-filled band the electrons can easily be excited, gaining energy and momentum. If you put an electric field on such a system, the electrons can adjust, picking up momentum in one direction:

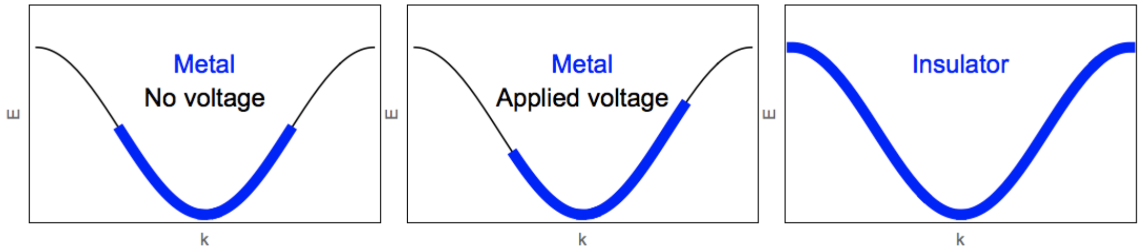


Figure 4. The thin curve in these plots shows the band: how the energy levels E depend on wavenumber k. The thick curve are the levels that are occupied. For metals in which atoms donate one electron each half the band is filled. With a half-filled band, an applied electric field can easily make the energy levels adjust, showing that such materials are conductors. Insulators have a completely filled band and cannot easily adjust to an applied field.

Thus mono-valent (one electron donated per atom) materials (most metals) are **conductors**. That metals conduct is closely related to their shininess: the conduction electrons move freely, so they easily absorb and re-emit radiation.

If each atom donates two electrons into the band, the band would be completely filled. In that case, a tiny amount of energy cannot excite the electron and the material does not conduct electricity well. Such materials are known as insulators or more precisely **band insulators**.

In three dimensions, the relationship $\varepsilon(\vec{k}) = \varepsilon_F$ defines a surface instead of two points. This surface is called the **Fermi surface**. The Fermi surface is an abstract surface in momentum space, not the real surface of a metal in position space. It describes the boundary between occupied and unoccupied energy levels at $T = 0$. For the free-electron gas where $\varepsilon = \frac{\hbar^2 \vec{k}^2}{2m_e}$ this is the surface of constant $|\vec{k}|$ namely a sphere. In metals such as sodium which have very weakly bound valence electrons, so the free-electron gas model is an excellent approximation, the Fermi surface is indeed spherical. For other metals, the Fermi surface can have a different shape. Here are the Fermi surfaces for sodium, copper and graphene

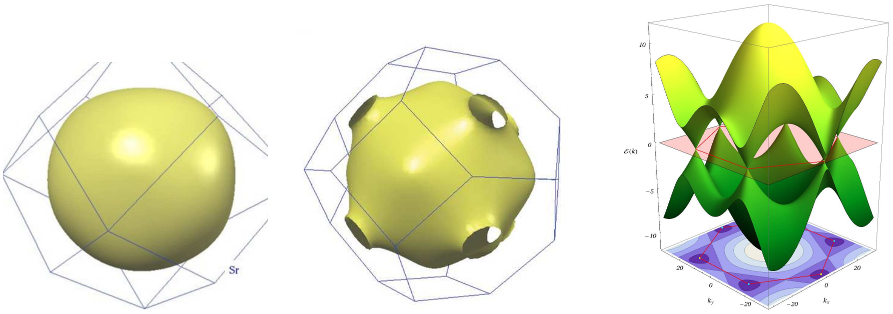


Figure 5. Fermi surfaces for sodium (left), copper (middle) and graphene (right).

In copper the things that look like holes in the Fermi surface are not really holes but rather the regions where $\frac{d\varepsilon}{dk}$ blows up. This is the 3D analog of a the 1D bandgap region in the Kronig-Penney model: when you go from one band to the next, ε jumps as k changes by a little bit, so $\frac{d\varepsilon}{dk}$ is large.

Graphene is a 2D sheet of carbon. So the values of $\vec{k}$ with $\varepsilon(\vec{k}) = \varepsilon_F$ are not a set of points, as in 1D, or a surface, as in 3D, but a curve. The curve is given by an intersection of a plane (the energy plane) with the surface. In the figure, the Fermi energy is indicated by the red plane. We see that as $\varepsilon \to \varepsilon_F$ the allowed k go to points along the surface of a cone. These cones are called **Dirac cones**. In an ultrarelativistic electron gas, the dispersion relation is $\varepsilon(\vec{k}) = c|\vec{k}|$. Because of the absolute value, the ultrarelativistic dispersion relation has the same cuspy behavior and Dirac cone structure we can see in the graphene Fermi surface. Thus electrons in graphene near the cusp act like ultrarelativistic particles, even though their energies can be $\varepsilon \lesssim \varepsilon_F \ll m_e c^2$. These effectively massless Fermion excitations give graphene some of its most interesting properties: it is the world's most conductive material. Normal metals have to move massive electrons around, but graphene can move around effectively massless ones. Graphene is also 200 times stronger than steel. By the way, graphene was discovered in 2004 by Geim and Novoselov by peeling scotch tape off a block of graphite (you make graphene every time you write with a pencil). Geim and Novoselov won the Nobel Prize for their discovery in 2010.

5 Summary

In this lecture we studied the statistical mechanics of fermions. A key concept in thermal Fermi systems is the Fermi level, defined as the energy below with all the states would be occupied and above which none of the states are occupied, at zero temperature. It is equal to the chemical potential at zero temperature.

The basic tool we use for studying fermionic systems is the free electron gas. This is a gas of non-interacting fermions quantized as non-relativistic particles-in-a-box. The density of states is $g(\varepsilon) = \frac{V}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\sqrt{\varepsilon}$. The free-electron gas predicts the bulk modulus andspeed of sound of metals in agreement with data. It also predicts the heat capacity of metals, and in particular that this heat capacity is linear in T at low temperature. This is in contrast to the heat capacity of phonons which scales like T^3 at low temperature (see Lecture 11). We computed the low-temperature limit of the heat capacity two ways. First, we applied a set of subtle mathematical tricks developed by Sommerfeld to expand for $T \ll T_F$. Second, we treated the free electron gas as a gas of electrons and holes (absence of electrons) with a symmetric spectrum around the Fermi level. This second method is much faster and motivates the broad use of holes as a theoretical tool. We'll come back to holes in Lecture 14.

A step beyond the free electron gas is the nearly free electron gas. Our treatment of the nearly-free gas was somewhat heuristic; the required calculations are not hard, and covered in any solid-state physics book, but they are beyond what we can do in one lecture. The important qualitative results we found were that

- The addition of periodic potentials generates bands, that is groups of energy levels separated by gaps.
- The band structure can explain insulators, when the Fermi level is in the middle of a band, and conductors, when the Fermi level is in a gap between bands.

Unfortunately, it is very difficult to connect the nearly free electron gas to actual metals. That is, we cannot determine which materials will conduct and which will insulate. To understand real materials, a more flexible method is the **tight-binding model** (Lecture 14), where the bands emerge from energy levels of atomic orbitals ($4s, 3d$, etc).

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 14: Semiconductors

1 Introduction

We saw how the Sommerfeld free electron model can explain many properties of metals. It postulates that to a leading approximation, electrons in metals roam freely in the metal like particles in a box. These electrons are fermions so this is an example of a free-electron gas. The free electron gas gives a number of very reasonable predictions. It predicts Fermi energies for metals in the $\sim$10 eV range and Fermi temperatures in the $\sim$50,000 K range. Thus Fermi gases in metals are highly degenerate, with very few states excitable at room temperature.

We used the free electron model to show that the heat capacity of electrons in metals is linear in T. Thus electrons dominate over the phonon contribution to C_V at low T but at high T, the electrons' contribution is negligible. The low T prediction is in agreement with data and the high T prediction explains why the electronic contribution can be ignored in the law of Dulong and Petit. The free electron model also gives very good predictions for the bulk modulus of metals.

In computing the heat capacity of metals we noted that there is an alternative language for discussing properties of degenerate Fermi gases where instead of thinking of all the energy levels as being excitations of the ground state, we think of the excited electrons as excitations above the Fermi level ε_F and the states that get excited as holes below the Fermi level. Although the holes are the absence of electrons, an excitation of a hole gives a positive contribution to the energy of the gas. Thus we can think of a Fermi gas as a collection of excitations and holes. This picture gives a powerful way to understand properties of semiconductors, as we will see in this lecture.

We also saw how the nearly free electron model, where the free theory is perturbed with a weak periodic potential leads to the emergence of bands. The band structure is crucial to understanding properties of real metals. Unfortunately, the free-electron model is very bad at incorporating differences among elements. A much more flexible approach is the **tight-binding model** which we explore in this lecture. The tight-binding model constructs allowed electronic states by combining atomic orbitals, similar to molecular orbital theory. After explaining the relevant concepts, we will use the tight-binding model to understand one of the most important technological innovations of the 20th century: semiconductors. Semiconductors are essential to every aspect of everyday life: they allow for very efficient and powerful computers. In Section 3 we use the band picture developed from the tight-binding model to understand how pn junctions and transistors work, and then discuss how transistors lead to computers.

Although this lecture seems long, it has very few equations, so it should be relatively quick to read. In fact, the lecture is really two lectures in one: Section 2 explains how to understand why different elements form different kinds of solids. The rest of the lecture explains why their band structure gives semiconductors such amazing properties. Section 2 is a bit more chemistry than physics, but it is important science that you should know. That being said, if you already have a qualitative understanding of the electronic properties of different elements then by all means feel free to skip it. In any case, please try to understand the material in this lecture, even if it takes multiple passes at the reading. As this lecture synthesizes chemistry, physics and technology, I think it contains a good amount material that could stick with you after the course is over.

2 The periodic table

To understand metals and semiconductors, we need a better understanding of the electron orbitals in elements than you might have gotten from your intro quantum mechanics class. In this section we'll first review the hydrogen atom then describe how to generalize to the other elements.

2.1 Hydrogen atom review

In quantum mechanics, you (hopefully) solved the Schrödinger equation to find the energy states of the hydrogen atom. These satisfy $\left(-\frac{1}{2m}\vec{\nabla}^2 - \frac{e^2}{4\pi\epsilon_0 r}\right)\psi_{nm\ell\sigma} = \varepsilon_n\psi_{nm\ell\ell\sigma}$. The eigenfunctions are separable: $\psi_{n\ell m\sigma} = R_{n\ell}(r)Y_{\ell m}(\theta, \phi)$ with $R(r)$ given by Laguerre polynomials and $Y_{\ell m}(\theta, \phi)$ are spherical harmonics. The energy levels depend only on the principle quantum number $n \geqslant 1$: $\varepsilon_n = -\text{Ry}\frac{1}{n^2}$ where $\text{Ry} = \frac{m_e e^4}{8h^2\varepsilon_0^2} = 13.6\text{eV}$ is the Rydberg constant. The other quantum numbers are the the angular momentum ℓ and the projection m of angular momentum on the z axis. ℓ is a whole number from 0 to $n-1$ and $m = -\ell, -\ell+1, \cdots, \ell$. Thus there are $n-1$ values of ℓ for every n and $2\ell+1$ values of m for every ℓ. The final quantum number is the spin $\sigma = \pm\frac{1}{2}$. The energy levels of hydrogen don't depend on ℓ, m or σ, only on n.

The quantum number ℓ gives the shape of the orbital, and m its orientation. We associate letters to ℓ values: $\ell = 0$ is the letter s, $\ell = 1$ is the letter p, $\ell = 2$ is the letter d, $\ell = 3$ is the letter f. (By the way, these letters originated from properties of associated spectral lines: sharp, principle, diffuse and fundamental.) The s orbitals are spherically symmetric. There is only one s orbital for each n since $2\ell+1 = 1$ when $\ell = 0$. The p orbitals have two lobes, one with $\psi < 0$ and one with $\psi > 0$. There are three p orbitals for each n since $2\ell+1 = 3$ when $\ell = 1$. The three p orbitals have different orientations, with the lobes pointing in the x, y or z direction. There are five d orbitals and seven f orbitals. Here are some orbital shapes of the s, p and d orbitals

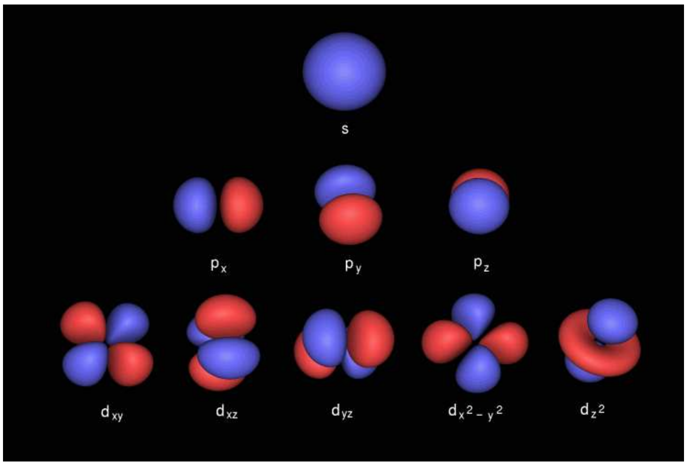


Figure 1. The shape of hydrogen atom orbitals are determined by spherical harmonics. These plots are like radiation patterns: the distance from the origin at the angle θ, ϕ is given by $Y_{\ell m}(\theta, \phi)$ with the color denoting the sign.

The radial wavefunction is determined mostly by the principle quantum number n. For $n = 1$, $R_{10}(r) \sim \exp\left(-\frac{r}{a_0}\right)$ with $a_0 = \frac{\hbar}{m_e c\alpha} = 5 \times 10^{-11}m$ the Bohr radius. Thus the size of the hydrogen atom itself is around a_0. With $n = 2$, $R_{20}(r) \sim \left(2 - \frac{r}{a_0}\right)\exp\left(-\frac{r}{2a_0}\right)$. So the $n = 2$ wavefunction also dies exponentially, but has a node at $r = 2\alpha_0$ after which it flips sign. In general, the radial dependence looks like $R_{n\ell}(r) \sim \exp\left(-\frac{r}{na_0}\right)$. So for bigger n the orbitals get bigger and bigger, with the atomic radius scaling as $\langle r\rangle \sim na_0$. The ℓ dependence of the radial wavefunctions is subleading – it doesn't affect the exponential, only a polynomial prefactor, e.g. $R_{21} \sim r\exp\left(-\frac{r}{2a_0}\right)$.

The spin σ doesn't affect the shape of the wavefunction. The only relevant fact is that since σ has two possible values, there are two degenerate orbitals (same energy) at each $n\ell m$. Thus for $n = 1$ there are two possible states, $1s^1$ and $1s^2$. For $n = 2$ there are two $2s$ states and six $2p$ states. For $n = 3$ there are ten $3d$ states, and so on.

2.2 Hydrogenic atoms

That may be where you left off in quantum mechanics. What do the orbitals of other elements look like? If there is only one electron and the nucleus has charge Z then we have the exact solution: the potential is Z times as big and states for the electron have energies $\varepsilon_n = -\text{Ry}\,\frac{Z^2}{n^2}$ and sizes $\langle r\rangle \sim \frac{n}{Z}\alpha_0$. The challenge comes when there is more than one electron, since the previous electrons screen the nuclear charge, and can do so asymmetrically. The screening also breaks the degeneracy of the energy levels and distorts the wavefunctions away from the hydrogenic form.

To a first approximation, we can treat the orbitals of the additional electrons as similar to those of one-electron atoms. In fact, the orbital shapes for the electrons in multiple-electron atoms are often similar to those of one-electron atoms, even if the energies of those orbitals are very different. Thus we label the multi-electron atoms using the notation for hydrogen-atom orbitals. For example, we write boron as $1s^2 2s^2 2p^1$, meaning the $n=1, \ell=0$ orbital has two electrons in it (the $1s^2$ part), the $n=2, \ell=0$ orbital has two electrons, and the $n=2, \ell=1$ orbital has 1 electron.

Now consider carbon ($Z=6$) which has another electron to come in. It will then be $1s^2 2s^2 2p^2$. But are the two electrons going to be the same p orbital with different spins, or different p orbitals? To figure out the answer, suppose the first $2p$ electron is in a p_x orbital. Since its electron cloud points in the x direction, the nuclear charge in x will be screened more in x than in the y or z directions. This means that the p_y and p_z orbitals will have lower energy than the other-spin p_x orbital, since they are less screened. So after p_x then either p_y or p_z gets filled. That is, for carbon, the two $2p$ electrons will have different values of m. In nitrogen, $1s^2 2s^2 2p^3$, the three $2p$ orbitals will be in the three different directions, p_x, p_y and p_z. It is not until oxygen, $1s^2 2s^2 2p^4$, that a second electron goes into the p_x state. Note here that x, y and z are arbitrary, the point is only that different m states are filled before two spin states go in to the same m. This effect leads to **Hund's rule:** every orbital is singly occupied with one electron before any is doubly occupied. When all the orbitals of a given ℓ are filled, we say the **shell** (all the m and s values for a given ℓ) is **closed**.

To understand the periodic table, we need more than Hund's rule. We have to go beyond the hydrogenic atom spectrum and understand how the electron screening moves the energy levels up and down. It turns out to be very complicated. The energy levels as a function of Z look like

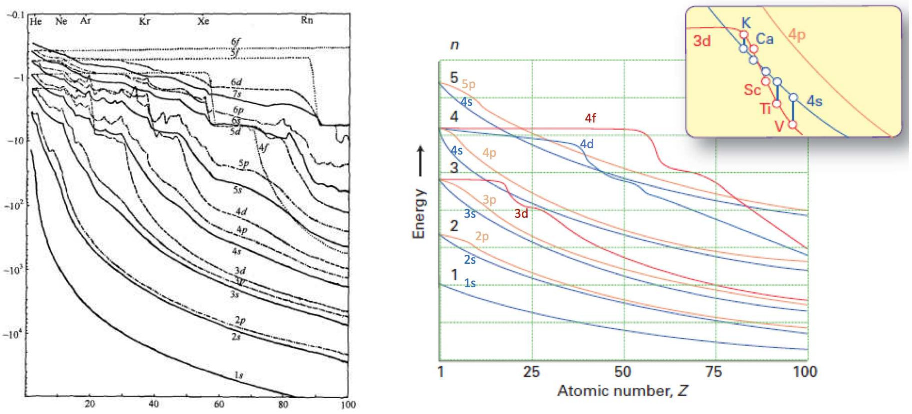


Figure 2. Energies of orbitals as a function of atomic number, including electron screening effects. Left is more realistic, from an old book, and right is more of a cartoon. Note that the 3d orbitals have energies above the 4s orbitals around $Z=19$ (potassium).

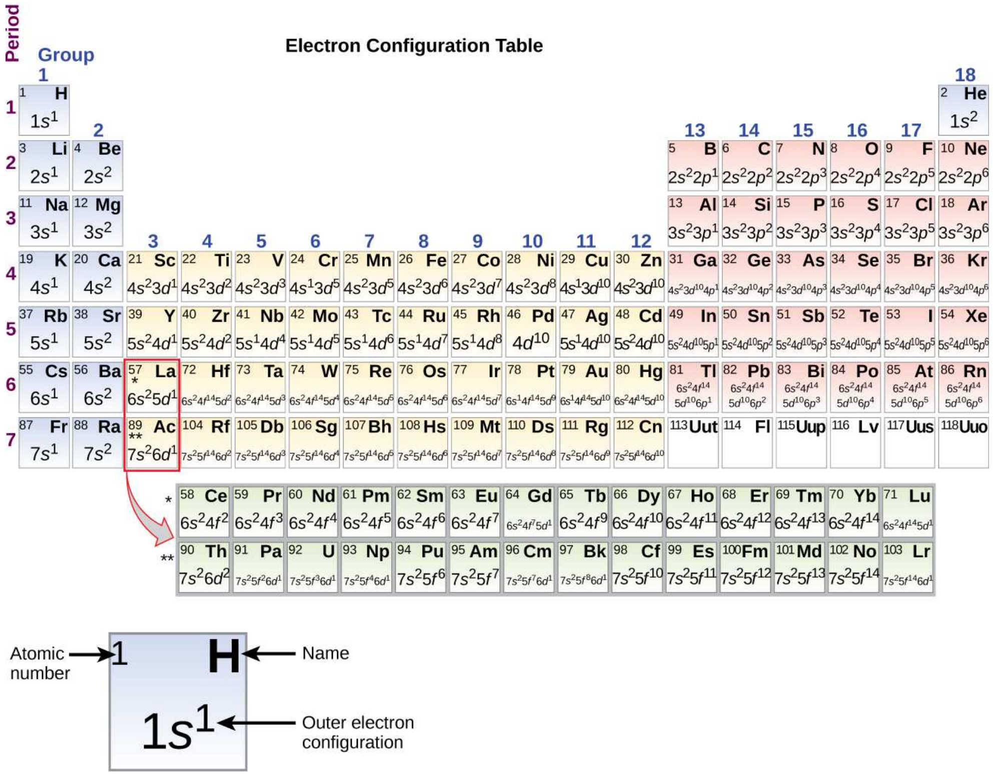
We see a few things from this plot. First of all, there is always a large gap between the p orbitals and the s orbitals in the shell above it. This gap is very important. It makes the elements with filled p orbitals and no extra electrons very stable. Such elements are called the noble gases and sit at the far right of the periodic table. Note that the noble gases do not have *all* the orbitals filled for a given n filled, just all the p orbitals filled for a given n.

The first row of the periodic table fills $n=1$. The second row fills $n=2$. The third row fills $n=3$, $\ell\leqslant 1$, but the $3d$ orbitals do not get filled until the 4th row. After argon, which is [Ne]$3s^23p^6$ with [Ne] meaning all the closed shells of neon (Ne$=1s^22s^22p^6$), one might think that $3d^1$ is next. Indeed, from the hydrogen atom, we know that the $3d$ orbitals have lower energy than the $4s$ orbitals, since the energy only depends on n. However, we see from Fig. 2 that by the time $n=4$, the degeneracy associated with the principle quantum number is pretty badly broken: the $4s^1$ level has lower energy than the $3d^1$ level. The result is that the two 4s orbitals fill first (K and Ca), then the ten $3d$ orbitals get filled (transition metals Sc to Zn), then the 4p orbitals fill until krypton, Kr$=$[Ar]$4s^23d^{10}4p^6$. In other words, to a reasonable approximation, the energy levels are grouped not quite by n but in sets

- $(1s)$, $(2s2p)$, $(3s3p)$, $(4s4p3d)$, $(5s5p4d)$, $(6s6p\,4f\,5d)$, $\cdots$

The filling of the levels in these groups gives the "periodicity" of the periodic table. Each group is a row.

Here is a periodic table showing the valence electron configurations (valence here means the electrons outside of the last nobel gas)"



Electron Configuration Table

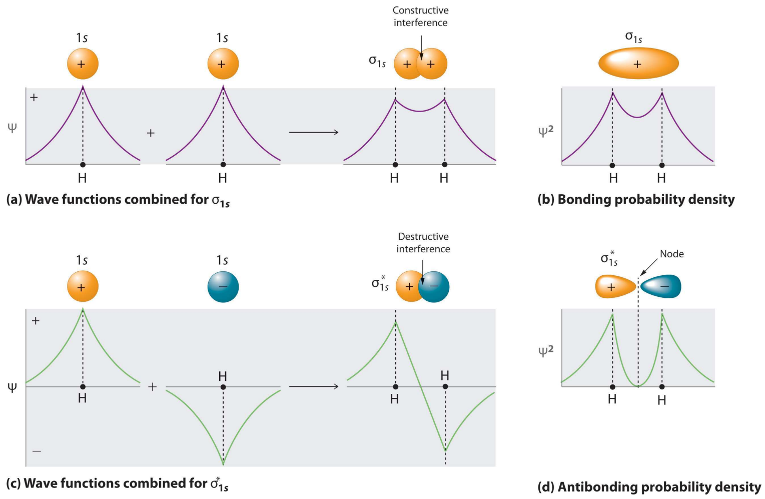
Period	Group 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H $1s^1$																	2 He $1s^2$
2	3 Li $2s^1$	4 Be $2s^2$											5 B $2s^22p^1$	6 C $2s^22p^2$	7 N $2s^22p^3$	8 O $2s^22p^4$	9 F $2s^22p^5$	10 Ne $2s^22p^6$
3	11 Na $3s^1$	12 Mg $3s^2$											13 Al $3s^23p^1$	14 Si $3s^23p^2$	15 P $3s^23p^3$	16 S $3s^23p^4$	17 Cl $3s^23p^5$	18 Ar $3s^23p^6$
4	19 K $4s^1$	20 Ca $4s^2$	21 Sc $4s^23d^1$	22 Ti $4s^23d^2$	23 V $4s^23d^3$	24 Cr $4s^13d^5$	25 Mn $4s^23d^5$	26 Fe $4s^23d^6$	27 Co $4s^23d^7$	28 Ni $4s^23d^8$	29 Cu $4s^13d^{10}$	30 Zn $4s^23d^{10}$	31 Ga $4s^23d^{10}4p^1$	32 Ge $4s^23d^{10}4p^2$	33 As $4s^23d^{10}4p^3$	34 Se $4s^23d^{10}4p^4$	35 Br $4s^23d^{10}4p^5$	36 Kr $4s^23d^{10}4p^6$
5	37 Rb $5s^1$	38 Sr $5s^2$	39 Y $5s^24d^1$	40 Zr $5s^24d^2$	41 Nb $5s^14d^4$	42 Mo $5s^14d^5$	43 Tc $5s^14d^6$	44 Ru $5s^14d^7$	45 Rh $5s^14d^8$	46 Pd $4d^{10}$	47 Ag $5s^14d^{10}$	48 Cd $5s^24d^{10}$	49 In $5s^24d^{10}5p^1$	50 Sn $5s^24d^{10}5p^2$	51 Sb $5s^24d^{10}5p^3$	52 Te $5s^24d^{10}5p^4$	53 I $5s^24d^{10}5p^5$	54 Xe $5s^24d^{10}5p^6$
6	55 Cs $6s^1$	56 Ba $6s^2$	57 La * $6s^25d^1$	72 Hf $6s^24f^{14}5d^2$	73 Ta $6s^24f^{14}5d^3$	74 W $6s^24f^{14}5d^4$	75 Re $6s^24f^{14}5d^5$	76 Os $6s^24f^{14}5d^6$	77 Ir $6s^24f^{14}5d^7$	78 Pt $6s^14f^{14}5d^9$	79 Au $6s^14f^{14}5d^{10}$	80 Hg $6s^24f^{14}5d^{10}$	81 Tl $6s^24f^{14}$ $5d^{10}6p^1$	82 Pb $6s^24f^{14}$ $5d^{10}6p^2$	83 Bi $6s^24f^{14}$ $5d^{10}6p^3$	84 Po $6s^24f^{14}$ $5d^{10}6p^4$	85 At $6s^24f^{14}$ $5d^{10}6p^5$	86 Rn $6s^24f^{14}$ $5d^{10}6p^6$
7	87 Fr $7s^1$	88 Ra $7s^2$	89 Ac ** $7s^26d^1$	104 Rf $7s^25f^{14}6d^2$	105 Db $7s^25f^{14}6d^3$	106 Sg $7s^25f^{14}6d^4$	107 Bh $7s^25f^{14}6d^5$	108 Hs $7s^25f^{14}6d^6$	109 Mt $7s^25f^{14}6d^7$	110 Ds $7s^25f^{14}6d^8$	111 Rg $7s^25f^{14}6d^9$	112 Cn $7s^25f^{14}6d^{10}$	113 Uut	114 Fl	115 Uup	116 Lv	117 Uus	118 Uuo
*				58 Ce $6s^24f^2$	59 Pr $6s^24f^3$	60 Nd $6s^24f^4$	61 Pm $6s^24f^5$	62 Sm $6s^24f^6$	63 Eu $6s^24f^7$	64 Gd $6s^24f^75d^1$	65 Tb $6s^24f^9$	66 Dy $6s^24f^{10}$	67 Ho $6s^24f^{11}$	68 Er $6s^24f^{12}$	69 Tm $6s^24f^{13}$	70 Yb $6s^24f^{14}$	71 Lu $6s^24f^{14}5d^1$	
**				90 Th $7s^26d^2$	91 Pa $7s^25f^26d^1$	92 U $7s^25f^36d^1$	93 Np $7s^25f^46d^1$	94 Pu $7s^25f^6$	95 Am $7s^25f^7$	96 Cm $7s^25f^76d^1$	97 Bk $7s^25f^86d^1$	98 Cf $7s^25f^{10}$	99 Es $7s^25f^{11}$	100 Fm $7s^25f^{12}$	101 Md $7s^25f^{13}$	102 No $7s^25f^{14}$	103 Lr $7s^25f^{14}6d^1$	

Figure 3. Periodic table showing valence electron configurations

2.3 Molecular orbital theory

Now that we understand the periodic table, based on orbitals of atoms in isolation, let's talk about bonds. Why do chemical bonds form?

Recall from quantum mechanics that if we have a potential well, there will be a bound state with some energy ε_0 and wavefunction $\psi(x)$. If we put two of these wells next to each other, then instead of two states with energy ε_0, one state will have energy $\varepsilon_-=\varepsilon_0-\Delta$ and the other energy $\varepsilon_+=\varepsilon_0+\Delta$. The state with lower energy is the symmetric combination $\psi_+=\frac{1}{\sqrt{2}}(\psi_1+\psi_2)$ and the state with high energy is the antisymmetric combination $\psi_-=\frac{1}{\sqrt{2}}(\psi_1-\psi_2)$. This is a very generic feature of quantum mechanical systems: when two systems are brought together, their energies will split, with some going lower and and some going higher.[1]



(a) Wave functions combined for σ_{1s}

(b) Bonding probability density

(c) Wave functions combined for σ^*_{1s}

(d) Antibonding probability density

Figure 4. When two 1D systems are brought together, the energy eigenstates are approximately the sum and difference of the two separate energy eigenstates.

The same thing happens when we bring two atoms together. Consider two hydrogen atoms. Each separately has two 1s orbitals (the two spin states) with energies ε_0. When we bring them together, two states get higher in energy and two get lower in energy, so their new energies are $\varepsilon_\pm = \varepsilon_0 \pm \Delta$ just like in 1D quantum mechanics. We call the lower energy states the **bonding orbitals** and the higher energy states the **antibonding orbitals**. This idea of constructing orbitals for the molecules by taking linear combinations of the atomic orbitals is called **molecular orbital theory**. It works pretty well and can explain a lot of chemistry.

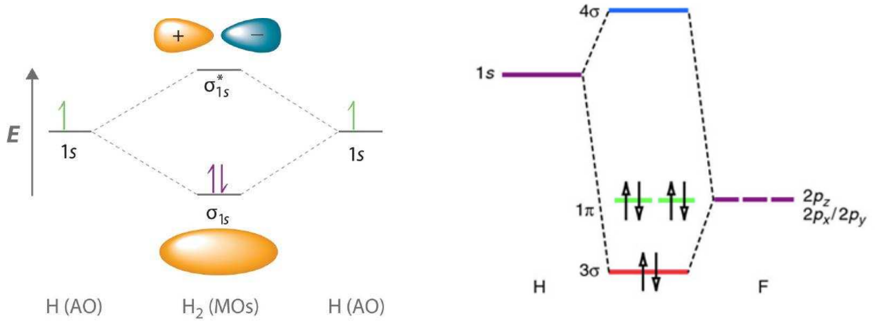


Figure 5. Left shows H_2. When the two H are apart, each has one electron (green arrows) in a 1s orbital. When they are close, the orbitals combine into a bonding (σ_{1s}) and antibonding ($\sigma^\star_{1s}$) orbital. The two elecrons go into the bonding orbital (purple arrows) since it has lower energy than $1s$. Right shows HF: when different elements bind, the molecular orbitals are asymmetric combinations of the two atomic orbitals.

So for two H atoms, there are 2 total electrons and the energy is lowered if both electrons are shared in the bonding orbitals. This is a **covalent bond**. On the other hand, if we bring two helium atoms together, with a total of 4 electrons, the two bonding orbitals would get filled with two electrons, but then the other two electrons must go into the antibonding orbitals. This is not energetically favorable, so it doesn't happen: He doesn't bond with itself.

[1] $\varepsilon_\pm = \langle\psi_\pm|H|\psi_\pm\rangle = \frac{1}{2}[\langle\psi_1|H|\psi_1\rangle + \langle\psi_2|H|\psi_2\rangle] \pm \frac{1}{2}[\langle\psi_1|H|\psi_2\rangle + \langle\psi_2|H|\psi_1\rangle] = \varepsilon_0 \mp \Delta$ with $\Delta = -\mathrm{Re}[\langle\psi_1|H|\psi_2\rangle]$.

For fluorine $F = 1s^2 2s^2 2p^5$, consider the three $2p$ orbitals with 5 electrons. If all three $2p$ orbitals are shared among an F_2 molecule, there would be 3 bonding and 3 antibonding orbitals to hold 10 total electrons. Thus 6 go into bonding and 4 into antibonding. Filling antibonding orbitals is not energetically favorable. So instead, each atom in F_2 leaves four of its five $2p$ electrons in their unshared orbitals and only shares one valence electron each. Then the problem is reduced to sharing a single orbital, say p_z, and one electron from each F can go into the bonding orbital. F_2 forms covalent bonds this way.

What about when two different atoms combine? The valence electrons (most weakly bound electrons) in each atom are the last ones to be added. The valence electrons of different elements have different binding energies, so the analog is a 1D asymmetric double-well quantum mechanics problem with wells of different depths, i.e. different binding energies E_1 and E_2. The depth of each well is the binding energy, or equivalently the

- **Ionization energy**: the amont of energy required to remove an electron from a neutral atom.

When different elements are brought together, the eigenstates are again linear combinations $\psi_+ \sim c_1\psi_1 + c_2\psi_2$ and $\psi_- = c_1\psi_1 - c_2\psi_2$ but the linear combinations are not symmetric ($c_1 \neq c_2$). The bonding orbital will be more like the lower-energy atomic orbital, and the antibonding orbital will be more like the higher-energy atomic orbital (see the right diagram in Fig. 5). Consequently the electrons will be localized close to the element with the lower-energy orbital.

The ionization energies of the elements can be measured. They look like this:

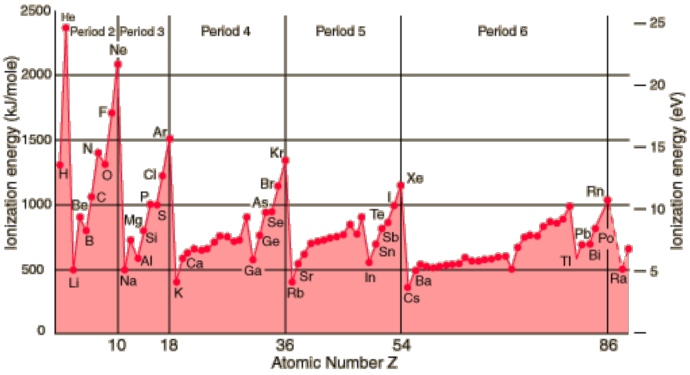


Figure 6. Ionization energies of the various elements

We see from this figure that the noble gases have the highest ionization energies so they are the most stable. Elements with one electron above a filled p orbital, namely those in the first column of the periodic table (H and the **Alkali metals**, Li, Na, K Rb) have low ionization energies: they want to give up an electron. The valence electrons are very weakly bound to these elements. On the other hand, elements that are one electron short of a filled p orbital (the **halogens**, F, Cl, Br, I) would become very stable if an electron were added. Their valence electrons are tightly bound and will not be given up easily. These have high ionization energies (but not as high as the noble gases which really really do not want to give up electrons).

So now if we bring an Alkali metal like sodium Na with binding energy $\varepsilon_1 \sim 5\text{eV}$ together with a halogen like Cl, with binding energy $\varepsilon_2 \sim 13\text{eV}$, the symmetric bonding molecular orbital will have energy $\varepsilon_- < \varepsilon_1$ and be localized near the Cl, while the antisymmetric anti-bonding molecular orbital will have energy $\varepsilon_+ > \varepsilon_2$ and be localized near Na. The two electrons will therefore both go into the bonding orbital and live close to the Cl. This type of bond where an electron is essentially donated from one atom to another is called an **ionic bond**. From molecular orbital theory, an ionic bond is rather an extreme form of covalent bond than something qualitatively different.

So molecular orbital theory explains some basic rules of thumb in chemistry such as the **octet rule**: atoms tend to prefer to have 8 electrons in the valence shell (ignoring the d and f orbitals). For example, Na gives up an electron to get a full shell and Cl receives one to get a full shell. You sometimes see this written as $\text{Na}\cdot + \cdot\ddot{\underset{\cdot\cdot}{\text{Cl}}}\!:\rightarrow \text{Na} \quad :\ddot{\underset{\cdot\cdot}{\text{Cl}}}\!:$. The octet rule is due to the trend that ionization energies increase until a p orbital is filled. Note that even in ionic bonds atoms never completely give up their electrons. Cl is neutral and does not form a bound state with an additional electron. Atoms towards the left side of the periodic table have fewer than half an octet filled and need to bond with something on the right side. Atoms toward the right side of the periodic table have more electrons to share, so it's easier for them to form covalent bonds and stable molecules.

Two concepts closely related and ionization energy are

- **electron affinity**: the energy change when a neutral atom attracts an electron to become a negative ion.

For example $\text{Cl} + e^- \to \text{Cl}^-$ releasing 3.6 eV of energy. Thus chlorine has an electron affinity of 3.6 eV. In other words while ionization energy is the change when losing an electron, electron affinity is the change when an electron is added to form a negative ion.

The other concept is.

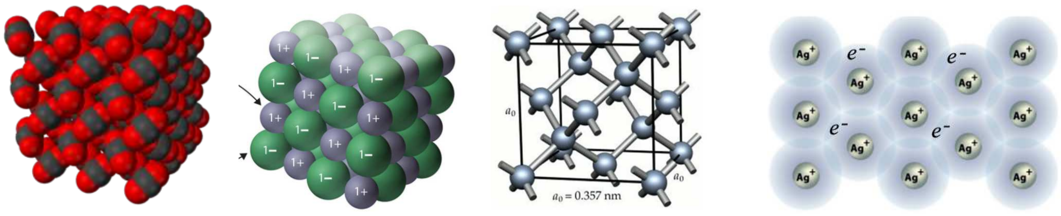
- **electronegativity**: how close an atom likes to pull bonding electrons towards itself

Electronegativity is vaguely defined and there are many competing definitions (such as the Pauling scale) that do not concern us. Qualitatively speaking, electronegativities are roughly proportional to ionization energies.

Elements on the far right (He, Ne, etc.) are noble gases. They have all filled shells and don't have any desire to lose or attract electrons. They have large ionization energies ($\sim$20eV for Ne) and no electron affinities at all since they cannot form anions. Noble gases don't bond, so electronegativity is not defined for them. Atoms next to them, the halogens (F, Cl, etc.), have high ionization energies ($\sim$12eV for Cl), high electron affinities ($\sim$3eV for Cl). They need one electron to fill a shell, so they desperately want it and have high electronegativities. Atoms on the far left (the alkali metals Li, Na etc.) will have filled orbitals if they lose one electron, so they have low ionization energies (5.3 eV for Li), no electron affinity, and low electronegativities. As you go from left to right in the table, the ionization energy, electron affinity and electronegativity increases.

2.4 Solids

We're almost back to statistical mechanics. The final concept from chemistry we need to understand is how to think about solids. Every element or compound forms a solid when cold enough. But these solids can have very different forms and properties depending on what the constituent atoms or molecules are. Almost all solids can be characterized as molecular, ionic, covalent-network or metallic:



Molecular solid
e.g. CO_2

Ionic solid
e.g. NaCl

Covalent network solid
e.g. diamond

Metallic solid
e.g. gold

Figure 7. Different types of solids

Molecules that are themselves quite stable, like CO_2 or H_2O or noble gases, have little interest in sharing electrons. Their solid forms are **molecular solids**: conglomerations of molecules with some regularity but no exactly periodic lattice structure. The solids are held together by various relatively weak attractive interactions, such as hydrogen bonds, dipole-dipole interactions, or dispersive forces. Ice is a molecular solid.

Elements that form ionic bonds, like table sald, NaCl, are happy sharing an electron pair between them. Their solid forms are **ionic solids**: the electronegative element (Cl for NaCl) draws the electron in close, leaving essentially pairs of Na^+ and Cl^- ions. These ions attract each other electrically holding the solid together.

Some elements, like carbon, like to form covalent bonds with their neighbors, leading to **covalent network solids**. The bonds in these solids are neighbor-to-neighbor and so the electrons are localized between the atoms, not distributed throughout the solid. A covalent network solid is in a sense one giant molecule. These solids are generally insulators, but some conduct. For example, diamond insulates, but graphite conducts. The difference is in the bonding structure. In diamond, each carbon is covalently bonded to 4 neighbors, tying up all its valence electrons in bonds and forming a rigid geometric lattice. This is what makes diamond so inflexible. In graphite, each carbon atom is bonded to only 3 neighbors, essentially in a plane, so 1 valence electron is mobile and can conduct electricity. The planes are only weakly attached to each other, which is why pencils work and graphene (2D graphite) exists.

The final solid form is the **metallic solid**. In a metallic solid, the electrons are loosely bound to the atoms and are easily shared among the atoms. It is these solids to which the free electron gas model applies.

Which elements form which type of solids? Let's go through elements from right to left on the periodic table. The noble gases are very stable and have no electron affinities, so they form molecular solids. Atoms which like to form diatomic molecules, such as the halogens (F_2, Cl_2, Br_2), oxygen O_2 and nitrogen N_2 will also form molecular solids, made out of these diatomic molecules. Carbon forms covalent networks, like diamond, graphite and graphene. Phosphorous is highly reactive and forms molecular solids with whatever it has reacted with. Sulfur forms octoatomic molecules S_8 that solidify. Selenium (Se) forms covalent networks. These are all the non-metallic elemental solids. Most of the rest of the periodic table have spare valence electrons and low electronegativities so they form elemental metallic solids. Elements on the boundary between metals and non-metals have intermediate properties. These are called **metalloids** and include the semiconductors:

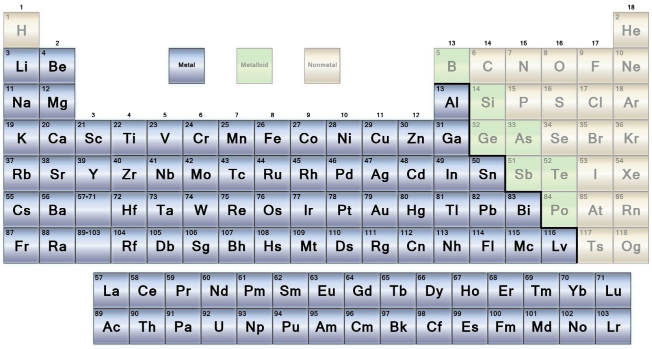

Metal
Metalloid
Nonmetal
1 H 18 He
2 Li Be 13 B 14 C 15 N 16 O 17 F Ne
Na Mg 3 4 5 6 7 8 9 10 11 12 Al Si P S Cl Ar
K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge As Se Br Kr
Rb Sr Y Zr Nb Mo Tc Ru Rh Pd Ag Cd In Sn Sb Te I Xe
Cs Ba 57-71 Hf Ta W Re Os Ir Pt Au Hg Tl Pb Bi Po At Rn
Fr Ra 89-103 Rf Db Sg Bh Hs Mt Ds Rg Cn Nh Fl Mc Lv Ts Og
La Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu
Ac Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr

Figure 8. Metals, non-metals and metalloids in the periodic table.

2.5 Tight-binding model

Just as molecules can be understood by combining the orbitals of separate atoms using molecular orbital theory, we can understand metallic solids by combining a regular array of atoms. This is known as the **tight-binding model**. The tight-binding model is basically molecular orbital theory applied to an array of atoms.

The analog quantum mechanics problem is a series of N particles-in-a-box or potential wells. When two wells with energies ε_0 are brought together, we get one eigenstate with lower energy and one with higher energy $\varepsilon_\pm = \varepsilon_0 \pm \Delta$. When three wells are brought together, the lowest energy lowers more, the highest energy raises more, the energies are roughly $\varepsilon_0 - 1.5\Delta, \varepsilon_0, \varepsilon_0 + 1.5\Delta$ and so on. Using the energy levels of the hydrogen atom, the basic picture looks like this

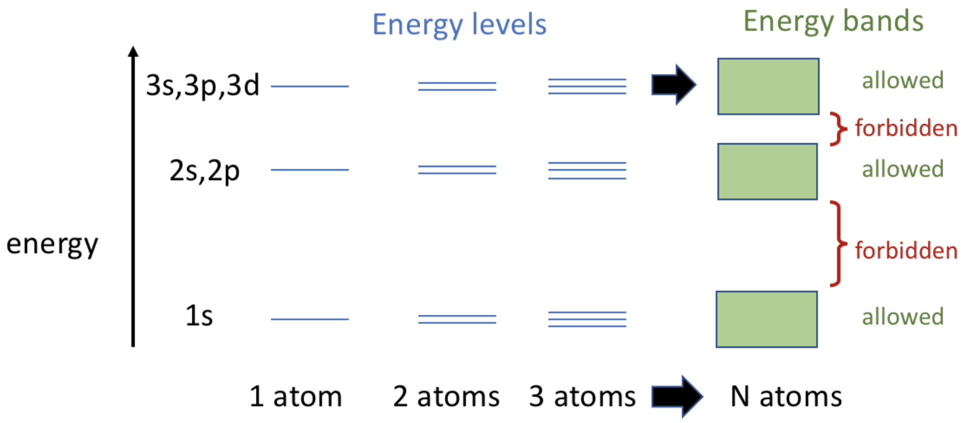


Figure 9. As orbitals mix, energy bands form

The discrete energy levels turn into **bands**. Half of each band comprises bonding orbitals with energies below the energy ε_0 of the isolated atom. The other half of each band comprises the anti-bonding orbitals with energies greater than ε_0. Every energy level is 2-fold degenerate because of the two electron spins. So if there are N atoms, there are $2N$ energy levels in each band. The number of bands is set by the number of original orbitals.

For elements above hydrogen, the $2p$ and $2s$ orbitals are not degnerate even for a single molecule (recall Fig. 2). Thus, when the orbitals combine, the center of the bands are offset but the bands may overlap. For example, in sodium: the bands look like

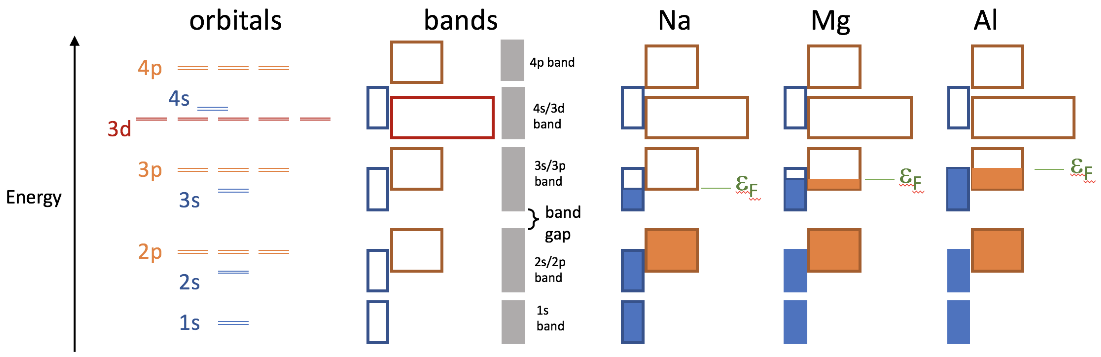


Figure 10. Bands of metals on the 3rd row of the periodic table. In Na the $3s/3p$ band is $1/8$ filled. In Mg it is $1/4$ filled and in Al it is $3/8$ filled. In all three metals, the Fermi energy is within a band(not in a band gap).

Sodium is $\mathrm{Na} = 1s^2 2s^2 2p^6 3s^1$, so each atom has one valence electron in the 3s orbital. When the bands form, the N valence electrons from the $3s^1$ orbitals get distributed over the N energy levels of the binding band. Thus the $3s$ band is half filled, as indicated.

Magnesium is $\mathrm{Mg} = 1s^2 2s^2 2p^6 3s^2$, you might think the $3s$ band would be completely filled, with the low energy and high energy states all populated, making Mg an insulator. However, we see from the figure that the $3p$ band overlaps the $3s$ band. So once the $3s$ band gets filled up to the level of the $3p$ band, the electrons start filling the $3p$ band. In this way the energy of magnesium is lower than the energy it would have if the $3s$ bands were completely filled. Thus there really is a hybrid $3s/3p$ band that is less than halfway filled for Mg.

Aluminium is $\mathrm{Al} = 1s^2 2s^2 2p^6 3s^2 3p^1$. It is similar to Mg, contributing 3 electrons per atom to the hybrid $3s/3p$ orbital that can hold 8 electrons total. Aluminum just has one more electron going into this hybrid band than magnesium.

You might think you can keep going this way. Silicon is $\mathrm{Si} = [\mathrm{Ne}]3s^2 3p^2$, so it could in principle fill the $3s/3p$ band up halfway. Halfway up isn't great though, because above halfway the anti-bonding orbitals would have to start being filled, and those have more energy than the free atoms, so it's not energetically favorable to fill them. Silicon has another option though. It has 4 electrons to share and needs 4 electrons to form a closed shell. So it can form covalent bonds with its 4 nearest neighbors. Thus instead of forming a metallic solid, Silicon forms a covalent network solid, like its upstairs neighbor carbon. Compared to carbon, silicon's electrons are farther out ($3s^2 3p^2$ rather than $2s^2 2p^2$), so they are relatively more weakly bound than in carbon. Silicon is conflicted: it is on the boundary between having metallic bonds and having covalent bonds. It is a metalloid.

To the right of silicon, the electrons would have to fill the anti-bonding part of the band. This is definitely unfavorable compared to forming covalent bonds. So the elements between Si and Ar, namely P, S and Cl are non-metallic. For example, the halogen at end of the row is chlorine $[\mathrm{Ne}]3s^2 3p^5$. If chlorine tried to share its 7 s/p electrons in a metallic solid, the band would be half-filled after putting 4 electrons in, so the last 3 electrons would be in antibonding orbitals. On the other hand, if we just look at the singly-occupied $3p$ orbital, say $3p_z$ with a spin-up electron in it, this looks a lot like hydrogen with a single electron in a $1s$ orbital. It can form bonding orbitals using just this orbital, lowering the energy compared to separated chlorine atoms. Thus Cl forms diatomic molecules and molcular solids.

As you go down in the periodic table, the $3d$ and then $4d$ orbitals become relevant. These can hold a lot of electrons, so the bands are very wide and a lot of elements form metallic solids using these orbitals.

2.6 Summary

This section was a bit long, and it may not have been clear to you why all this chemistry was included in a physics class. The point was to understand why some elements form metals, some form insulators, and some form semiconductors (see next section). It is perhaps worth a quick summary of the main logical steps leading from QM to the classification of solids:

- The large degeneracy of energy levels of the hydrogen atom is broken in other elements. The closely spaced energies are in sets: $(1s)$, $(2s2p)$, $(3s3p)$, $(4s4p3d)$, $(5s5p4d)$, $(6s6p4f5d)$, $\cdots$
- When two atoms are brought together, the two valence orbitals from the atoms in isolation combine into a bonding orbital, with lower energy than both, and an anti-bonding orbital, with higher energy than both.
- When N orbitals combine, the many bonding and antibonding orbitals merge into bands, with the energy level of the isolated-atom orbitals falling right in the middle of the band.
- For elements on the left and in the center of the periodic table, it's energetically favorable to pool electrons, forming metals with the bonding-orbital part of the band filled.
- As you move to the right on the periodic table, electrons become more strongly bound (higher ionization energies), and it is no longer energetically favorable to form bands. Instead, single electrons are shared with nearest neighbors in covalent bonds.
- The column in the periodic table with carbon, silicon and germanium would have half-filled bands in metals, which is energetically neutral. They are on the fence between being metals and covalent network solids.

We will next apply this understanding to explain doped semiconductors and their use in computers.

3 Semiconductors

Now that we understand the origin of real bands in actual elements, we can start classifying materals. First some more terminology. If the Fermi level is between two bands we call the band above it the **conduction band** and the band below it the **valence band**. The **bandgap** is the energy difference between the valence and conduction band. In a metal, the Fermi level is within a band, so the valence and conduction bands overlap. In an **insulator** there is a big bandgap. A **semiconductor** is somewhere between an insulator and a conductor: it has a bandgap, typically of order 1 eV.

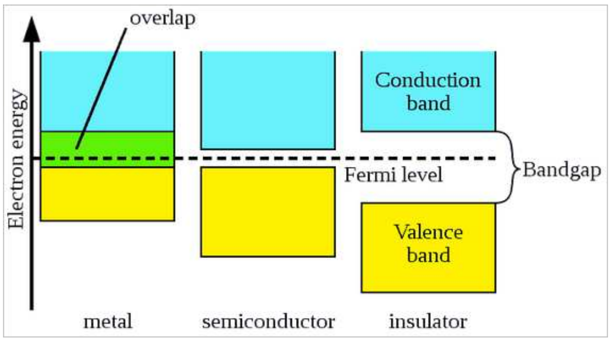


Figure 11. Metals have no bandgap, intrinsic semiconductors a small bandgap and insulators have a large bandgap.

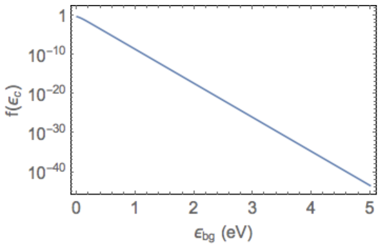
In intrinsic semiconductors, or insulators, the Fermi level is right in the center of the bandgap. This follows from charge conservation and the symmetry of the Fermi distribtion. Charge conservation implies the same total number of electrons and holes (since electron-hole pairs come from real electrons being excited from the valence to conduction band). Recall that the Fermi distribution

$$f(\varepsilon)=\frac{1}{e^{\beta(\varepsilon-\varepsilon_F)}+1} \tag{1}$$

is symmetric around ε_F (that is, $f(\varepsilon_F+\Delta)=1-f(\varepsilon_F-\Delta)$). Also recall that the relevant energies have $\Delta\sim k_BT\ll\varepsilon_F$ so that $g(\varepsilon)\approx g(\varepsilon_F)\times\Theta(\varepsilon)$ where $\Theta(\varepsilon)=0$ if $\varepsilon_C<\varepsilon<\varepsilon_V$, with ε_C the energy at the bottom of the conduction band and ε_V the energy at the top of the valence band, and $\Theta=1$ outside of this region. Since the Fermi distribution is symmetric, it is only possible for $N_e=\int_{\varepsilon_F}^{\infty}d\varepsilon f(\varepsilon)g(\varepsilon)$ and $N_h=\int_0^{\varepsilon_F}d\varepsilon f(\varepsilon)g(\varepsilon)$ to be equal if the same number of electron and hole states are removed from the density of states by Θ, i.e. if $\varepsilon_C-\varepsilon_F=\varepsilon_F-\varepsilon_V$. Since $\varepsilon_C-\varepsilon_V=\varepsilon_{\text{bg}}$, this implies that $\varepsilon_F=\varepsilon_V+\frac{\varepsilon_{\text{bg}}}{2}$, which is right in the middle of the band. So the Fermi level is is the center of the bandgap.[2] If the system is modified so that the distribution of available levels is no longer symemtric between electrons and holds, the Fermi level can move. This happens in doped semiconductors (see Section 3.1).

Now, where does the bandgap size ~1 eV typical for a semiconduction come from? At room temperature $k_BT=0.025\,\text{eV}$. We can then compute the probability of finding an electron at the base of the conduction band, at energy $\varepsilon_C=\varepsilon_F+\frac{1}{2}\varepsilon_{\text{bg}}$ as

$$f\left(\varepsilon_F+\frac{1}{2}\varepsilon_{\text{bg}}\right)=\frac{1}{\exp\left(\frac{\varepsilon_{\text{bg}}}{2k_BT}\right)+1}\approx\exp\left(-\frac{\varepsilon_{\text{bg}}}{0.05\text{eV}}\right)= \tag{2}$$



[2] Is is not at the *exact* center of the bandgap, because the density of states is not *exactly* symmetric around ε_F. However, corrections are of order $\left(\frac{\varepsilon_{\text{bg}}}{\varepsilon_F}\right)^2\approx 0.01$ which is small enough to neglect.

We see that this function is exponentially falling. For $\varepsilon_{\text{bg}}=1\text{eV}$, the base of the conduction band only has a 10^{-9} chance of being occupied. At $\varepsilon_{\text{bg}}=2$ eV the probability is already down to 10^{-18}, and at $\varepsilon_{\text{bg}}=3\text{eV}$ it's down to 10^{-27}. Considering there are of order $N_A\sim10^{24}$ electrons around, we conclude that a bandgap of order $\varepsilon_{\text{bg}}\gtrsim2.5$ eV a material will no longer conduct and become an insulator.

The bandgap in NaCl, which forms an ionic solid, is $\varepsilon_{\text{bg}}=8.9\text{eV}$. NaCl is an insulator. Diamond (a solid form of carbon) at room temperature has $\varepsilon_{\text{bg}}=5.4\text{eV}$. Diamond is also an insulator. Silicon, below diamond, has $\varepsilon_{\text{bg}}=1.08\text{eV}$ making it on the boundary between conductor and insulator: silicon is a semiconductor. An important characteristic of a semiconductor is number density of conduction electrons (in the conduction band). To compute this, we can treat the electrons in the conduction band as having the usual non-relativistic dispersion relation: $\varepsilon=\varepsilon_F+\frac{1}{2}\varepsilon_{\text{bg}}+\frac{\hbar^2\vec{k}^2}{2m_e}$ where $\varepsilon_C=\varepsilon_F+\frac{1}{2}\varepsilon_{\text{bg}}$ is the energy at the bottom of the conduction band. Then the density of states for conduction electrons is as in the free-electron gas, $g(\varepsilon)=\frac{V}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\sqrt{\varepsilon-\varepsilon_F-\frac{1}{2}\varepsilon_{\text{bg}}}$.[3] Approximating $f(\varepsilon)\approx\exp\left(-\frac{\varepsilon-\varepsilon_F}{k_BT}\right)$ as in Eq. (2) we then get

$$n_e=\frac{1}{V}\int_{\varepsilon_F+\frac{1}{2}\varepsilon_{\text{bg}}}^{\infty}g(\varepsilon)f(\varepsilon)d\varepsilon=2\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{3/2}e^{-\frac{\varepsilon_{\text{bg}}}{2k_BT}}\tag{3}$$

Similarly the density of holes is

$$n_h=\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\int_{-\infty}^{\varepsilon_F-\frac{1}{2}\varepsilon_{\text{bg}}}\sqrt{\varepsilon_F-\frac{1}{2}\varepsilon_{\text{bg}}-\varepsilon}\,e^{-\frac{\varepsilon_F-\varepsilon}{k_BT}}d\varepsilon=2\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{3/2}e^{-\frac{\varepsilon_{\text{bg}}}{2k_BT}}\tag{4}$$

So, in a pure intrinsic semiconductor the densities of electrons and holes are the same: $n_e=n_h=n_I$. n_I is called the **intrinsic carrier concentration**. Plugging in the numbers for silicon we get

$$n_I=2\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{3/2}e^{-\frac{\varepsilon_{\text{bg}}}{2k_BT}}=\left(2.4\times10^{19}\frac{1}{\text{cm}^3}\right)e^{-\frac{1.08\text{eV}}{0.05\text{eV}}}=10^{10}\frac{1}{\text{cm}^3}\tag{5}$$

If the Fermi level were not in the middle of the band gap (as it won't be in doped semiconductors), the expressions $\varepsilon_F\pm\frac{1}{2}\varepsilon_{\text{bg}}$ in Eqs. (3) and (4) would change to $\varepsilon_F+a\varepsilon_{\text{bg}}$ and $\varepsilon_F-(1-a)\varepsilon_{\text{bg}}$, but we would still find

$$n_en_h=n_I^2=4\left(\frac{mk_BT}{2\pi\hbar^2}\right)^3e^{-\frac{\varepsilon_{\text{bg}}}{k_BT}}\tag{6}$$

This equation is sometimes called the law-of-mass-action for semiconductors, as it relates the equilibrium electron and hole concentrations much like the law of mass action does in chemistry.

Let's try to understand a little better why carbon is an insulator and silicon is a semiconductor. Carbon is $C=[\text{He}]2s^22p^2$. Its 4 valence electrons allow it to form 4 covalent bonds in a covalent network solid filling up it shells. This is more stable than forming metallic bonds. The atomic spacing in diamond is $a_C=0.154\text{nm}$. Its electrical resistivity is $r_C=10^{14}\Omega m$. Silicon is $\text{Si}=[\text{Ne}]3s^23p^2$. Silicon, like carbon, has a valence of 4, and forms covalent bonds, but the valence electrons of silicon are farther out, so the bonds are weaker. These weaker bonds are why the atomic spacing in silicon is about twice as large as diamond, $a_{\text{Si}}=0.235\text{nm}$. Thus while silicon is sort of covalently bonded, it could equally well be thought of as having metallic bonds. The tight-binding model lets us interpolate between covalent and metallic. The resistivity of silicon is $r_{\text{Si}}=0.001\Omega m$, many orders of magnitude smaller than diamond (but many orders of magnitude larger than metals, such as copper with a resistivity of $r_{\text{Cu}}=1.7\times10^{-8}\Omega m$).

Below silicon on the periodic table is germanium: $\text{Ge}=[\text{Ar}]3d^{10}4s^24p^2$. Its electrons are even farther out, lowering its bandgap to $\varepsilon_{\text{bg}}=0.67\text{eV}$, the lattice spacing is $a_{\text{Ge}}=0.243\text{nm}$ and the resistivity $r_{\text{Ge}}=0.0005\Omega m$, about half that of silicon. So silicon and germanium are semiconductors. Another important semiconductor is gallium arsenide GaAs, with $\varepsilon_{\text{bg}}=1.43\text{eV}$.

[3] Technically, the mass here is an "effective mass" determined by the curvature of the dispersion relation at the base of the conduction band. For simplicity, we'll just take the effective mass to be the electron mass.

3.1 Doping

What makes semiconductors very important is that the valence and conduction bands and the bandgap are relativity easy to manipulate by adding impurities. This is called **doping**.

The properties of doped semiconductors are best understood in the language of electrons and holes. Recall from the last lecture that electrons and holes helped us understand the heat capacity of metals in the free electron model. The power of the picture comes from the symmetry of the Fermi function $f(\varepsilon) = \frac{1}{e^{\beta(\varepsilon - \varepsilon_F)} + 1}$, namely that $f(\varepsilon_F + \Delta) = 1 - f(\varepsilon_F - \Delta)$. A gas of electrons has as many excited states above ε_F as there are holes below ε_F. Not only is the number of states the same, but the probability of finding an electron state at $\varepsilon_F + \Delta$ is the same as the probability of finding a hole at $\varepsilon_F - \Delta$. A hole contributes a positive amount to the energy of the electron gas, since the electron that should have been in the hole is missing. If a electron in the valence band at energy ε is excited into the conduction band at energy ε', of the $\varepsilon' - \varepsilon$ total energy, the part $\varepsilon_F - \varepsilon$ of it is attributed to the hole and the rest $\varepsilon' - \varepsilon_F$ is attributed to the electron.

Now we introduce another useful property of holes: they have positive charge. When an electron is excited out of an atom, it leaves a positive ion in its place. An analogy is a line of parked cars with a spot at the front. As a car moves out of its spot into the spot in front of it, it leaves a hole. Then another car behind it can fill the hole, leaving a different hole. In this way the hole moves backwards in the line, although it is really the cars that are moving. In the electron gas picture, the electrons are not associated with individual atoms, so the holes should not be associated with individual atoms either. Instead, we should think of the states near ε_F in a semiconductor as gas of negatively-charged electrons and positively-charged holes each of which has a minimum energy of excitation $\varepsilon \gtrsim \frac{\varepsilon_{\text{bg}}}{2}$.

Ok, now to doping. Let's take silicon as the semiconductor and consider adding a small amount of phosphorous. Phosphorous is the element to the right of silicon and therefore has one more valence electron (5 instead of 4): $P = [\text{Ne}]3s^2 3p^3$. The extra electron is weakly bound and, since this new electron is the 5th of the possible 8 that the 3s/3p orbital could hold, it would have to contribute to an antibonding orbital if P formed a metal. Therefore, its energy is close to the conduction band of Si, but slightly lower since it is energetically favorable *not* to fill antibonding orbitals. More quantitatively, the new levels in P-doped Si are centered at 45 meV below the conduction band. When we dope Si with P, the extra valence electrons from the phosphorous fill the new levels and are easily excited into the conduction band. Thus adding phosphorous increases the carrier concentration and makes silicon more conductive. P-doped Si is an example of an **n-type semiconductor** since the new particles added are **n**egatively charged, i.e. electrons.

We can also dope silicon by adding some of the element on its left, aluminum: $\text{Al} = [\text{Ne}]3s^2 3p^1$. Aluminum has one fewer electron than silicon. Thus it provides new places for the electrons in Si to go. The energies of these new states are slightly above the valence band, so we call these states "acceptor" sites. For aluminum, the new acceptor sites are centered 57 meV above the valence band. Thus, Al-doped Si is also more conductive than pure silicon because electrons can move from the valence band into these new acceptor sites. Note that Al does not contribute new electrons, so it is the valence electrons from Si that move into the new sites. Alternatively, we can say that Al contributes new holes. These new holes can move down into the valence band. The hole picture is nice because of the symmetry where we swap n-type ↔ p-type semiconductors and electron ↔ hole. Al-doped-Si is called a **p-type semiconductor** because holes are **p**ositively charged.

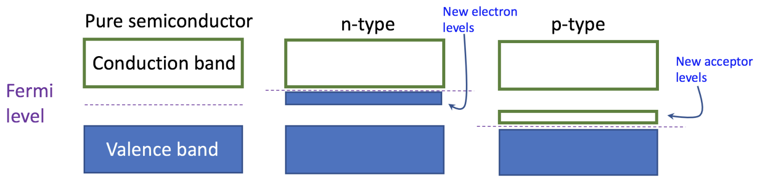


Figure 12. Doping of semiconductors changes the band structure.

FYI, typical doping fractions range from 1 part per thousand to 1 part per billion.

4 Diodes

Now it gets interesting. What happens if we put a p-type semiconductor next to an n-type one? The extra electrons from the P atoms will diffuse across the junction to find the Al atoms. Eventually, enough charge will be transferred that a substantial charge difference will build up, inhibiting further charge transfer. We call this the accumulated voltage the **junction potential** and denote it by V_{junc} The result looks like this:

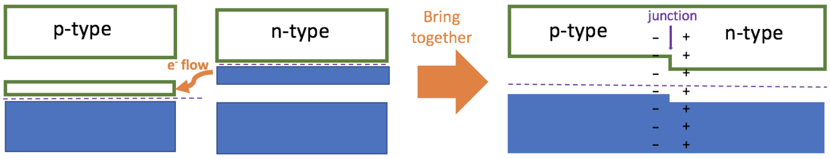


Figure 13. When p-type and n-type semiconductors are brought together, electrons move from the n side to the p side. Charge builds up in the interface forming a p-n junction.

To be more quantitative, consider first the pure n-type or pure p-type semiconductor. In these, the extra donor/acceptor sites significantly increases the carrier concentrations. On the n-type side, the donor (electron) concentrations are typically $n_e^{n\text{-type}} = 1.0\times 10^{16}\frac{1}{\text{cm}^3}$. Because of the law of mass action, Eq. (6), the concentration of holes on the n-type side is then $n_h^{n\text{-type}} = \frac{n_I^2}{n_e^{n\text{-type}}} = 2.2\times 10^4\frac{1}{\text{cm}^3}$ with $n_I = 1.5\times 10^{10}\frac{1}{\text{cm}^3}$ the intrinsic carrier concentration (set by the band gap) as in Eq. (5). On the p-type side, the acceptor (hole) concentrations are typically $n_h^{p\text{-type}} = 1.0\times 10^{15}\frac{1}{\text{cm}^3}$ and so $n_e^{p\text{-type}} = \frac{n_I^2}{n_p^{n\text{-type}}} = 2.2\times 10^5\frac{1}{\text{cm}^3}$. Now, when we put the two sides together the difference in electron/hole concentrations must be compensated for by the production of a junction potential. So we should have

$$n_h^{n\text{-type}} = n_h^{p\text{-type}} e^{-\frac{\varepsilon_{\text{junc}}}{k_B T}} \tag{7}$$

This implies

$$\varepsilon_{\text{junc}} = k_B T \ln \frac{n_h^{p\text{-type}} n_e^{n\text{-type}}}{n_I^2} = 25\,\text{meV}\times \ln \frac{10^{15}10^{16}}{10^{20}} = 0.61\,\text{eV} \tag{8}$$

The voltage is then $V_{\text{junc}} = \frac{\varepsilon_{\text{junc}}}{e} = 0.61V$. Commercial silicon p-n junctions are typically engineered by varying the doping levels to have junction potentials of this size, around 0.6 V.

Now what happens when we apply an external voltage to the p-n junction? If the negative terminal is connected to the n side (**forward bias**), it supplies electrons that diffuse towards the junction. At the junction they help neutralize the charge barrier, lowering the junction potential. Alternatively, we can think of holes diffusing out of the positive terminal into the p side, to neutralize the junction potential. Once the voltage overcomes the junction potential ($\gtrsim 0.6V$), the natural diffusion of electrons from the n side to the p side (or holes from p to n) can resume. The electrons then annihilate the holes contributed from the battery and a steady-state current flows through the system.

If the applied voltage is in the other direction, however, so that the negative terminal connects to the p side, it will draw holes from the n side and send electrons into the p side. The electrons on the p side make the junction potential larger and no current passes through. The junction voltage simply increases. This is called **reverse bias**. If enough reverse-bias voltage is applied, above the **breakdown voltage** (typically $30V$-50V), then the electrons are forced across the junction and a current flows (in the opposite direction to the forward bias case).

Thus this p-n junction acts a **rectifier** or **diode**: it allows current to flow in only one direction: $V \gtrsim 0.6V$ is required in forward bias to conduct but $V \gtrsim 50V$ is required for reverse bias to conduct:

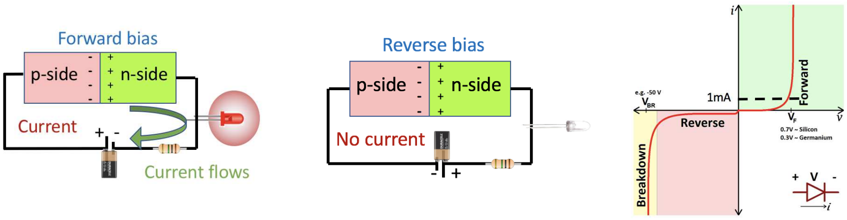
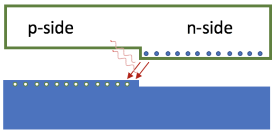


Figure 14. Current only flows through a p-n junction when a negative voltage is applied to the n side (forward bias). Flipping the battery (reverse bias) dues not induce current. Right shows the current induced for an applied voltage for a p-n junction.

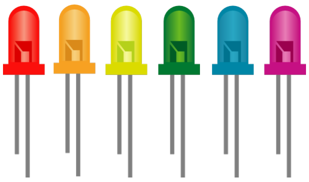
4.1 Light-emitting diodes

In forward bias mode, with an applied voltage larger than the junction potential ($V_{\text{applied}} \gtrsim 0.6V$), the voltage continuously pulls holes out of the p-side and supplies electrons to the n side so that the current continues to flow. The flowing electrons enter on the n side into its conduction band. Similarly, the flowing holes enter the valence band on the p side. As the conduction electrons cross the junction from the n side to the p side they would need to pick up energy to stay in the conduction band (since the p-side conduction band is higher than the n-side one). As there is nowhere to get this energy from, the electrons instead fall down into the valence band on the p-side, annihilating holes. A falling electron loses roughly $\Delta\varepsilon \sim \varepsilon_{\text{bg}}$ of energy from this transition. Like any electronic transition, this energy leaves the system through a photon of wavelength $\lambda = \frac{hc}{\Delta\varepsilon}$. For silicon, the bandgap is around $\varepsilon_{\text{bg}} = 1.1$ eV which corresponds to a wavelength of $\lambda = 1130\text{nm}$, in the infrared. In summary, at voltages above around $0.6V$ of the junction potential, a p-n junction emits monochromatic infrared light. Such a device is called a **light-emitting diode** or **LED**.



$=$

$=$



(9)

The earliest LEDs were indeed in the infrared and provided the technology for the first remote controls – that little red dot on your old TV remote is an LED.

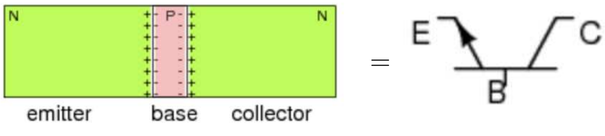
Since a 1.1 eV bandgap is in the infrared, in order to get to higher frequencies (visible)some fancy material-science engineering is required. As we noted before, you can't have a semiconductor with too large of a bandgap or it will just be an insulator. Moreover, the bigger the bandgap, the more unstable the diode becomes. An effective method for growing blue LEDs was finally found by Akasaki, Amano and Nakamura in the early 1990s. They received the Nobel Prize for their work in 2014. The reason that blue was so important is that blue is that there are many chemicals called phosphors that can be used to convert blue light to other colors. So the blue LED technology led to the ubiquitous extremely energy efficient LED lights found today. An LED lightbulb with the same luminescence as a 100 W incandescent lightbulb can use as little as 5 W: a 95% efficiency gain.

LEDs are extremely efficient since essentially all of the energy goes into light. This is in contrast to incandescent lights which use blackbody radiation: only a small fraction of the energy of an incandescent bulb is in the visible spectrum (you know how to calculate this!). Much of the energy of an incandescent bulb is in the infrared and dissipated as heat. That's why incandescent bulbs get very hot, but LEDs do not.

5 Transistors

It is hard to overestimate the importance of transistors in nearly every aspect of technology. Early computation used vacuum tubes or even mechanical relays as transistors. These worked, but were slow and clunky. It wasn't until Shockley, Bardeen and Brattain showed that transistors could be made out of semiconductors in 1947 (Nobel prize 1956) that modern computing took off. These solid-state transistors allowed for the miniaturization of computers and the efficiency improvements that we have seen over the last 70 years.

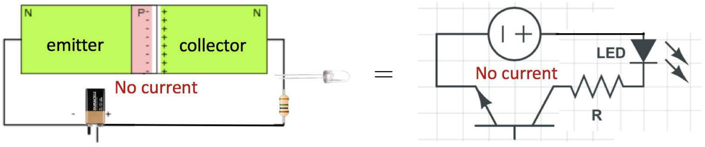
A **bipolar junction npn transistor** is made by combining two n-type semiconductors on either side of a very thin p-type semiconductor. We call the middle part the base, and the two sides the emitter and collector. The emitter is generally much more heavily doped than the collector. The transistor looks like this



(10)

The symbol on the right is how we represent an npn transistor in a circuit diagram.

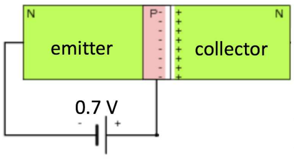
To a first approximation, this thing is just two diodes sandwiched together. So if we try to connect the emitter to the collector, with say a battery and an LED, nothing happens:



(11)

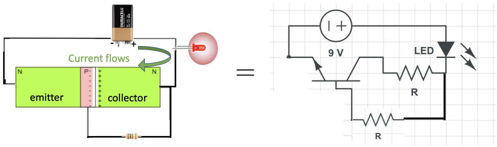
Turning the battery around doesn't help either: no current flows with either positive or negative voltage.

Current would flow if we connect the base to the emitter or collector. If we apply a voltage forward bias between the emitter and base we can eliminate their junction potential:



(12)

The forward bias on the emitter NP junction allows the emitter to conduct freely, in either direction. Thus, now if we run current between the emitter and collector, with say an LED on one side, we would find we find a light:



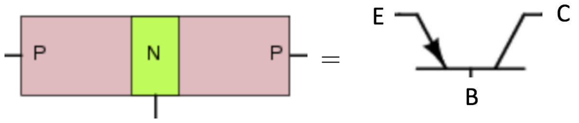
(13)

The point is that adding applying a 0.7 V voltage between the base and the emitter controls whether current can flow or not into the circuit.[4]

[4] Real transistors are slightly asymmetric between collector and emitter. A typical voltage drop between base and emitter is $V_{be} \approx 0.6 - 0.7V$, and between base and collector is $V_{bc} \approx 0.5 - 0.6V$, so that the collector is at slightly lower potential than the emitter ($V_{ce} \approx 0.1V$).

Note that for this to work the base has to be thin. The thinner it is, the more likely electrons are to be swept across the base and deposited into the collector rather than exiting out the bottom into the forward-biasing circuit. Typically more than 99% of the current flows into the collector.

The other kind of transistor is a pnp transistor which has two p-type semiconductors sandwiching an n type semiconductor in the middle. These are drawn as



(14)

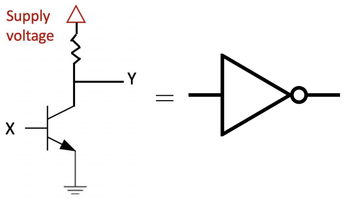
The arrow on the emitter shows which way current can flow when the transistor is turned on. Remember the convention in electronics is that current arrows leave the positive terminal and flow into the negative terminal. So in npn transistors, current flows from collector to emitter when a positive voltage is applied to the base; in pnp transistor, current flows from emitter to collector when the base is connected to negative voltage (ground).

5.1 From transistors to logic gates

We have shown that a transistor is a kind of gate. If you put a positive voltage to the base of a npn transistor, it opens the gate, allowing current to flow from collector to emitter. If you put a negative voltage to the base of a pnp transistor, it opens the gate, allowing current to flow from emitter to collector. What can we do with these things? The next step to getting from transistors to computers is building the basic set of boolean logic gates. These are things like NOT, AND, NAND, OR, NOR or XOR.

Let's start with NOT. A NOT gate, also called an **inverter**, takes 0 and turns it into 1 and takes 1 and turns it into 0. It can be simply built from a npn transistor by connecting X to the base and Y to the collector. We also hook a supply voltage up to the collector and the emitter to the ground. So it looks like this

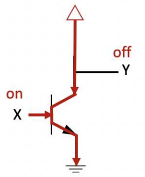
NOT =



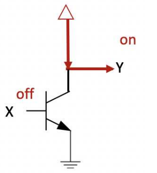
(15)

So if X is 0 (off) then the current will not be able to pass through the transistor and is diverted to Y. Thus $X = 0$ implies $Y = 1$. If X is 1 (on) then the gate is open and the current can pass right through the transistor to ground, giving Y nothing. Thus $X = 1$ implies $Y = 0$. This is a not gate:

$1 \to 0$:



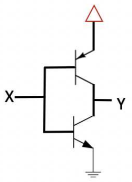
$0 \to 1$:



(16)

By the way, this gate also shows how transistors act like amplifiers: a small voltage (0.7V) applied at X can be turned into an arbitrarily large voltage at Y (that of the supply voltage). The efficiency of amplification led to transistor radios in the 1950s: a weak radio signal comes in that we can idealize as a binary pattern 010010101 at X. Adding two NOT gates in series with a large supply voltage produces the same pattern but amplified at Y. Previous radios used vacuum tubes for amplification and were heavy, delicate pieces of furniture. The method of amplification using solid state transistors allowed radios to be small and portable, ushering in the consumer electronics revolution.

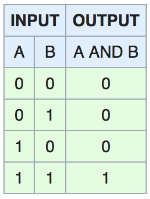
This style of logic gate construction is called NMOS (n-type metal-oxide semiconductor). It uses only npn transistors.[5] An alternative circuit design with identical function are the CMOS (complimentary metal-oxide semiconductor) circuits that combine npn and pnp transistors. For example, a CMOS inverter gate looks like



(17)

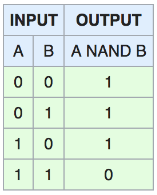
Note that the CMOS does not need a resistor to prevent shorting: in the off position, no current is drawn. Thus CMOS has very little static power consumption which is one of its main advantages. CMOS circuits replaced NMOS as the standard sometime in the 1990s. The integrated circuits in your computer and your phone undoubtedly use CMOS.[6]

The other standard logic gates are all 2 → 1 gates, so they take two inputs and give one out. AND for example, gives $A \wedge B$ meaning that if A and B are both 1, then it gives 1 otherwise it gives 0. The definitions of these various operations are



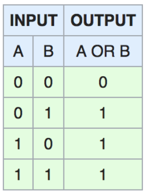
INPUT		OUTPUT
A	B	A AND B
0	0	0
0	1	0
1	0	0
1	1	1

AND



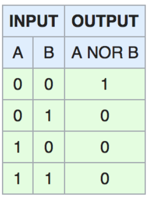
INPUT		OUTPUT
A	B	A NAND B
0	0	1
0	1	1
1	0	1
1	1	0

NAND



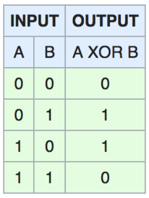
INPUT		OUTPUT
A	B	A OR B
0	0	0
0	1	1
1	0	1
1	1	1

OR



INPUT		OUTPUT
A	B	A NOR B
0	0	1
0	1	0
1	0	0
1	1	0

NOR



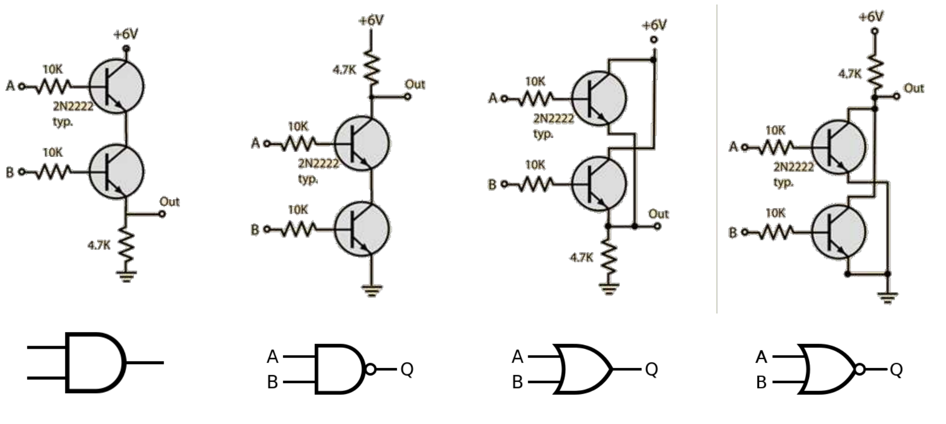
INPUT		OUTPUT
A	B	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

XOR

(18)

Example NMOS circuit diagrams to build some of these with transistors are:

AND NAND OR NOR



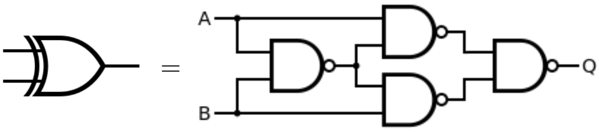
(19)

Let's look at the NAND gate. The inputs A and B both have to be on for current to flow to ground. If either is off, then the current flows to "Out". This is consistent with the NAND logic table. The AND gate is the same, but the output is after the second emitter, so both have to be on for current to flow out. For OR, if the based of either transistor gets current, then the current will pass through to "Out", consistent with OR logic.

[5] npn transistors are faster than pnp transistors, since electrons have larger mobility than holes.

[6] Another technical point is that CMOS circuits generally use field-effect transistors (FETs) rather than the bipolar junction transistors (BJTs) that we have been discussing. FETs, like BJTs, are solid-state devices made form p-type and n-type semiconductors. Their design allows them to control the current flow through the transistor using voltage across the collector-emitter channel, so that, in contrast to a BJT, no current flows through the base.

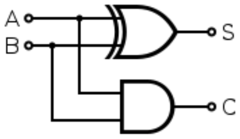
The XOR gate is an "exclusive OR", meaning it gives 1 if *and only if* 1 of the inputs is 1 and other other is zero. We can construct an XOR gate out of NAND gates:



(20)

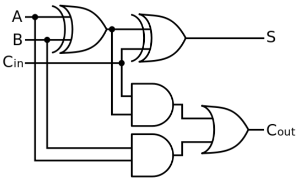
Can you figure out why this combination of NAND gates acts as an XOR gate?

Note from the logic table in Eq. (18) for XOR is the same as for the addition of 1 bit binary numbers. Although the single XOR gate only adds two 1 bit numbers modulo 2, which is maybe not so interesting, XOR gates can be combined to make more complicated adders. To add more than 2 bits, we need to carry over the second binary digit. We do this with a carry block in what's called a **half adder**



(21)

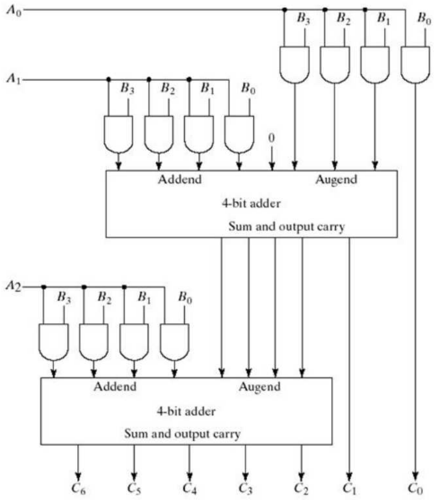
In this diagram, A and B are added with an XOR into S (for sum). The S bit is 0 if $A = B = 0$, but also if $A = B = 1$. If $A = B = 1$ the answer should be 2, and we need a bit in the second "digit". The "carry" bit C carries this information: the AND sets bit C only if $A = B = 1$. Now we just hook this up with another bit to add 3 bits in a **full adder**:



(22)

Can you figure out why this adds the 3 bits? You can check your undestanding at http://www.falstad.com/circuit/e-fulladd.html.

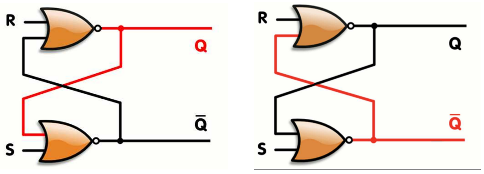
From here, you can add more and more bits together by sewing in successive adders. You can make more complicated integrated circuits to multiply numbers and do other operations with them. For example, here's a circuit that multiplies 3 bits $A = (A_2, A_1.A_0)$ by 4 bits $B = (B_0, B_1, B_2, B_3)$ to give 7 bits in C



(23)

So to multiply 7 × 13 you would set $A = 111$ and $B = 1101$. The output should be the 7 bits in $C = 1011011$.

There's one final element needed to make a computer: memory. There are many ways to make memory, but the basic form is a **latch** or **flip-flop**. For example, we can make such a thing using cross-coupled NOR gates:



(24)

This object called a or **SR latch** (for Set-Reset) has two stable states, shown in red. Either $Q=1$ and $\bar{Q}=0$ (left) or $Q=0$ and $\bar{Q}=1$ (right). To change from the left state to the right, we can put a current into the R input. Since R feeds into a NOR gate, it only gives current out if both inputs are 0 so when we change R, the current stops. Once the current stops, it no longer flows into the bottom NOR gate. Since $S=0$ this then makes the current go out of that gate and link up with the top NOR. At this point the current can be removed from R and the circuit stays in the right state. To go from right to left, we can pulse S. In this way, we can store a bit, and toggle it back and forth through the Set and Reset switches: Set makes it 1 (left), Reset makes it 0 (right).

Now that we have memory and can compute things, it's just a matter of putting them together in ever more complex combinations. Note that NAND has 2 transistors, XOR has 8, AND and OR have 3, so the full adder has 25 transistors. So a 4×4 bit adder should have roughly 400 transistors.

Adding 4 bit numbers doesn't seem like much, but you can actually reuse circuit elements fairly easily, storing the output then feeding it back in. In early computers, the transistors were not the solid-state devices we discussed here but rather vacuum tubes. The Mark 1 computer that you've walked by a thousand times in the Science Center but never looked at has around 1000 mechanical relay transistors. It operated on 24 bit numbers and took about $0.3s$ to add two such numbers. An iPhone X has 4 billion transistors and can add two 24 bit numbers in around $\frac{1}{\text{GHz}}=10^{-9}s$.

6 Summary

The first half of this lecture introduces molecular orbital theory and explained how to think about the periodic table from the point of view of the types of solids formed. A summary of the main results of the first have was given in Section 2.6. There's a lot of physical chemistry in this first half. So if you never plan on taking a physical chemistry or inorganic chemistry class, you are strongly encouraged to study this first half in depth.

The second half of the lecture focused on semiconductors and why they are so important for technology. Semiconductors are like insulators, in that their Fermi level is in a band gap, but the band gap is small, typically of order 1 eV, so it's pretty easy for electrons in the band below the Fermi level (the valence band) to get excited into the band above the Fermi level (the conduction band). Indeed, at room temperature 10^{-9} of the electrons are excited. This is in constrast to an insulator like diamond where 10^{-45} of the electrons are excited. By applying external voltage was can easily manipulate the electrical properties of semiconductors.

The key use of semiconductors is in making diodes, which are materials in which current can only flow in one direction. Diodes are made by doping semiconductors by adding atoms with extra valence electrons, ($n-\text{type}$) or fewer valence electrons ($p-\text{type}$). Light-emitting-diodes make very energy efficient lightbulbs.

Diodes can be sandwiched together in the npn or pnp order to form transistors. Their important property is that the conduct only when a voltage is applied to the base (middle part). So that they acts as if statements: *if* a voltage is applied, let current flow. Combining transistors in various ways one can make logic gates, and then computers. A cell phone typically has billions of transistors.

Matthew Schwartz
Statistical Mechanics, Spring 2025

Lecture 15: Stars

1 Introduction

There are at least 100 billion stars in the Milky Way. Not everything in the night sky is a star – there are also planets and moons as well as nebulae ("cloudy" objects including distant galaxies, clusters of stars, and regions of gas) – but it's mostly stars. These stars are almost all just points with no apparent angular size even when zoomed in with our best telescopes. An exception is Betelgeuse (Orion's shoulder). Betelgeuse is a red supergiant 1000 times wider than the sun. Even it only has an angular size of 50 milliarcseconds: the size of an ant on the Prudential Building as seen from Harvard square. So stars are basically points and everything we know about them experimentally comes from measuring light coming in from those points.

Since stars are pointlike, there is not too much we can determine about them from direct measurement. Stars are hot and emit light consistent with a blackbody spectrum from which we can extract their **surface temperature** T_s. We can also measure how bright the star is, as viewed from earth . For many stars (but not all), we can also figure out how far away they are by a variety of means, such as parallax measurements.[1] Correcting the brightness as viewed from earth by the distance gives the **intrinsic luminosity**, L, which is the same as the power emitted in photons by the star. We cannot easily measure the mass of a star in isolation. However, stars often come close enough to another star that they orbit each other. For such stars, we can measure their mass using Kepler's laws. Finally, by looking at details of the stellar spectra we can find evidence of metals. To an astronomer, **metallicity** is the amount of any element heavier than helium in a star.

With temperature and luminosity data, we can make a scatter plot of stars in the galaxy to see if there are any patterns. This was done by Hertzsprung and Russell first around 1910:

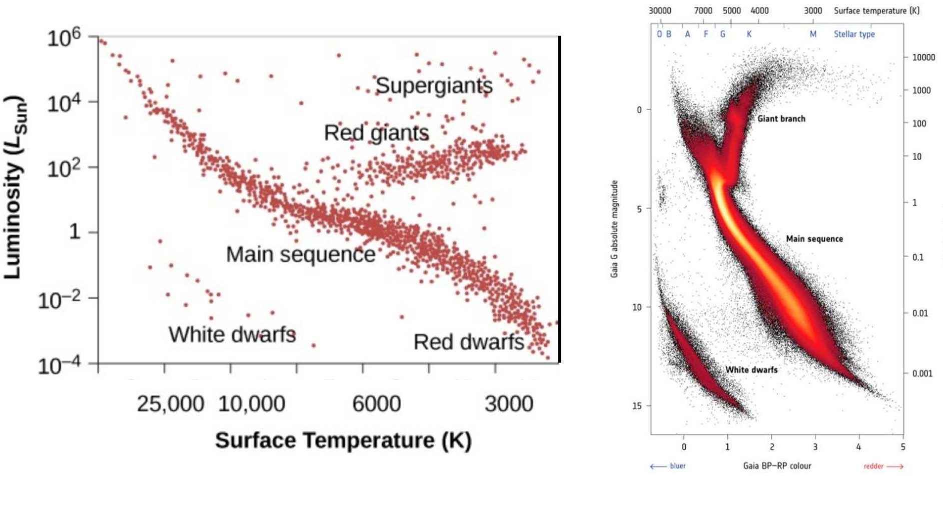


Figure 1. Hertzsprung–Russell diagram of luminosity and surface temperature. Left is an early estimate. Right is more recent from the Gaia experiment.

We see from the Hertzsprung–Russell (HR) diagram that most stars fall in two swaths: a diagonal swath, known as the **main sequence** (90% of stars), which includes stars like our sun, and a horizontal swath with luminosities around 100 times brighter than our sun called the **horizontal**

[1] Unfortunately, we cannot use the redshift of the blackbody spectrum to find distance, since a redshifted blackbody spectrum looks like a blackbody spectrum at a different temperature. This follows from dimensional analysis, since the Planck spectrum only depends on T through the combination $\frac{h\nu}{k_B T}$ so rescaling ν can be compensated by rescaling T.

branch (labeled “Red giants” or “giant branch” in the figure). A first question from this data is “why do stars fall along swaths rather than being distributed throughout the L/T plane?” The basic answer is the **Vogt-Russel theorem**: all stars with the same chemical composition and mass are the same. In particular, if the chemical composition is the same, then the mass indexes a one-parameter family of points in *any* plot, including a HR diagram. The main sequence swath describes stars with hydrogen-burning cores, like our sun. The red giants composing the horizontal branch are burning helium. There is another region heading upwards from the horizontal branch not clearly distinguishable in this figure known as the **asymptotic giant branch**. It contains supergiants which burn heavy elements. Finally, the stars in the lower left are called white dwarfs.

We’ll first discuss how the branches on the HR diagram all fit together in the standard picture of stellar evolution. Then we’ll do some simple stellar thermodynamics. Stellar physics is generally quite complicated, with lots of coupled equations, for convection, radiation, etc. These equations must be solved numerically to make precise quantitative predictions, but analytic approximations can be made as well. In fact, we can get a basic picture of how things work using just some physical reasoning and order-of-magnitude estimates. Some stars (white dwarfs and neutron stars) are simple enough that we will actually be able to describe them accurately using quantum statistical mechanics.

2 Overview of stellar evolution

In this section we discuss the standard narrative for star formation. Much of stellar physics is calibrated relative to the sun. The sun is denoted with the symbol $\odot$. The mass of the sun is $M_\odot = 2\times 10^{30}$ kg and the radius of the sun is $R_\odot = 7\times 10^5$km.

After the big bang, the universe cooled into around 75% hydrogen and 25% helium, with trace amounts of lithium. Some patches of the universe started out denser and hotter than others (we think this is due to primordial quantum density fluctuations during inflation), and the denser patches became more dense over time due to gravitational attraction. When a region of gas becomes dense enough, the density increase accelerates rapidly, a process called a **Jeans instability**. After the instability is reached the gas continues to compress until stars form, and if the temperature gets high enough, hydrogen fusion begins.

Hydrogen fusion in stars refers to a collection of processes whose net effect is to turn hydrogen H into ^{4}He. This fusion generates thermal pressure, keeping the stars from collapsing further under their own gravity. Thus for most of its lifetime, a star is in **hydrostatic equilibrium** and equilibrium statistical mechanics can be used.

High temperatures and densities are required for fusion. If the amount of matter that coalesced to form a star is not greater than around 0.08 $M_\odot$, the star will not burn and instead is just a ball of hydrogen, like Jupiter. These failed stars are called **brown dwarfs**. The main difference between a brown dwarf and a planet is how they formed: planets form from supernova remnants while brown dwarfs form from collapse of interstellar gas.

The region of the star where temperatures are high enough for hydrogen to fuse into helium is called the **core**. The core of the sun is a ball with radius around $0.25R_\odot$. For stars that are lighter than around $0.4M_\odot$ convection currents move the helium out of the core and allow hydrogen to fall in. The relatively low mass of these stars, right above the threshold for hydrogen fusion, make them small and cool. They are called **red dwarfs**. 73% of the stars in the Milky Way are red dwarfs. These stars burn hydrogen slowly, in principle until it is all used up. However, it takes hundreds of billions of years for a red dwarf to burn all its fuel and since the universe is only 15 billion years old, none of red dwarfs have yet to run out of hydrogen.

In stars with masses above $0.4M_\odot$, like our sun (about 20% of stars), the density in their cores gets high enough to prevent much convection. Thus the helium produced there basically stays put. After the hydrogen in the core is all burned up (so only helium is left), the thermal pressure from fusion stops and the star starts to contract again. The subsequent increase in density and temperature allows fusion to commence in the shell slightly outside of the core. Helium produced from this shell falls into the core, increasing the temperature further. As heat flows out from the core, it pushes the outer layers farther away. The net effect is that 1) the core contracts to around 1/3 its original size and its temperature goes up from $15\times 10^6 K$ to $100\times 10^6 K$ and 2) the outer layers are pushed out (to $\sim 100 R_\odot$) and cool. The increased core temperature can allow helium to fuse into carbon, nitrogen and oxygen. Thus after core hydrogen is used up, the equilibrium configuration of the star is qualitatively different: it has a core of helium and metals and is larger with a cooler surface temperature. By the Vogt-Russel theorem, these stars form a different line in the HR diagram. The star has left the main sequence to become a **red giant**.

In the red giant phase the star continues to burn the remaining hydrogen, expanding and cooling, and fuses helium in the core. This stage is relatively fast, and less than 0.5% of stars are red giants. Eventually all the helium in the core is used up and core nuclear fusion stops. Then, once again, the thermal pressure stops and the star's core can contract under its own weight and heat up, allowing helium fusion to continue in shells around it. It has entered the asymptotic giant branch. Again matter is pushing outward. This time the matter contains H, He, C, N and O. There is a lot of convection in this phase, causing the heavy elements to be dredged up from the core in a series of bursts. The result is the formation of a set of shells of matter called a **planetary nebula**:

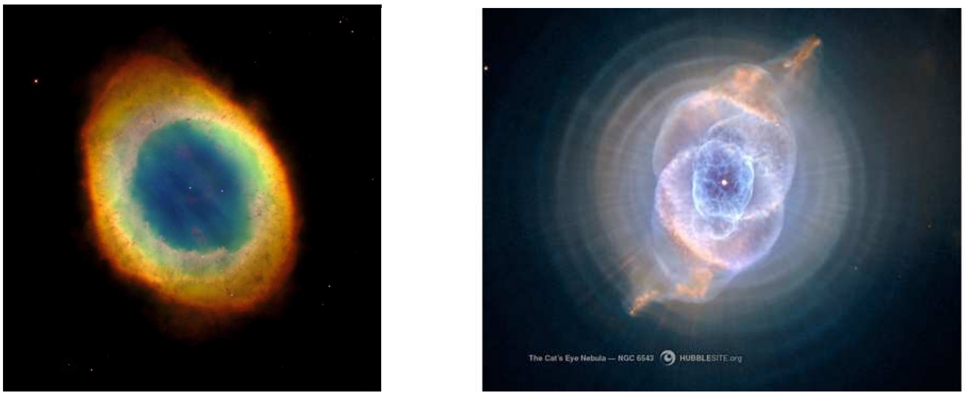


Figure 2. The ring nebula (left) and cats eye nebula (right) are planetary nebula. The dot in the middle of each is a white dwarf. Planetary nebula have nothing to do with planets.

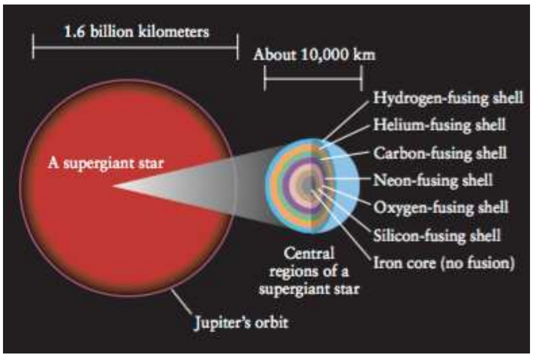
Planetary nebula are colorful because UV radiation from the core ionizes the atoms in the gas. This makes them some of the most beautiful objects in the sky. They have nothing to do with planets ("planetary" is a historical misnomer – at low resolution, they look a bit like round, green or blue planets).

Two things can now happen to the core of an asymptotic giant branch star, depending on its mass. If the star's mass is not large enough for its weight to overcome the electron degeneracy pressure in the core, the core cannot contract enough for fusion to continue. In this case, the core becomes a star called a **white dwarf** held together by gravity and stabilized by electron degeneracy pressure. About 4% of stars are white dwarfs. The white dwarf is a kind of star that cannot be understood without quantum statistical mechanics. A main-sequence star is stable against gravitational collapse because of the thermal pressure from active nuclear fusion. In a white dwarf, there is no nuclear fusion and little thermal and radiation pressure, so only degeneracy pressure can be stabilizing the star.

In Section 4 we will show that the degeneracy pressure limits the mass of a white dwarf. The limit is called the **Chandrasekhar limit**, $M_{\mathrm{WD}} \lesssim 1.4 M_\odot$. The corresponding bound on the original star's mass is $M \lesssim 8 M_\odot$. If $M \gtrsim 8 M_\odot$ then the electron degeneracy pressure in the core cannot prevent further contraction. These high-mass stars continue to collapse leading to additional stages of nuclear fusion. The next set of reactions include carbon fusion, which produces heavier elements like Ne, Na and Mg, and neon fusion, which produces more O and Mg. Once the carbon and neon are used up, the cycle advances: core contraction, shell expansion, and core temperature/density increase, leading to oxygen fusion. Then silicon fusion. Eventually iron is produced. Iron is the endpoint of nuclear fusion since elements above iron have endothermic fusion reactions and exothermic fission reactions[2]. Stars where these processes are happening are called **supergiants**.

[2] After iron, the Coulombic repulsion of the protons which scales like Z^2 begins to outweigh the nuclear attraction which scales like Z, with Z the atomic number.

Each stage of evolution of a supergiant generates a new shell of core material. This leads to a star that has different elements in different places, like shells of an onion:



(1)

All this nuclear activity makes supergiant stars very bright. Some of the brightest stars in the sky are supergiants, including Betelgeuse, Rigel, and Antares.

Once iron is formed and fusion stops, the core is only around the size of the earth and the supergiant is the size of our solar system. The thermal pressure stops abruptly and the gravitational pressure then commences to collapse the star. This happens very rapidly, raising the temperature to 5 billion K in about a tenth of a second. After around $0.25s$ the core can have a density of $4 \times 10^{17}\frac{\text{kg}}{m^3}$. That's like packing the entire mass of the earth into Jefferson Hall. The high temperature unleashes high energy photons which break apart the nuclei into protons and neutrons. The high density of electrons and protons allows them to fuse:

$$p^+ + e^- \rightarrow n + \nu \tag{2}$$

As the neutrinos produced by this fusion leave the star, the pressure drops further allowing additional contraction. The star begins collapsing catastrophically, at as much as $1/4$ the speed of light. Eventually, the core of neutrons cannot collapse further due to the strong interactions and the neutron degeneracy pressure. The incoming waves of matter then bang into the neutron core and bounce outward. This leads to an enormous explosion called a **core-collapse supernova**. After the explosion, the core may be obliterated. Or there may be a remnant **neutron star** (made of neutrons stabilized by degeneracy pressure) or, if the core mass is too large, a **black hole**.

Here's a summary of the story:

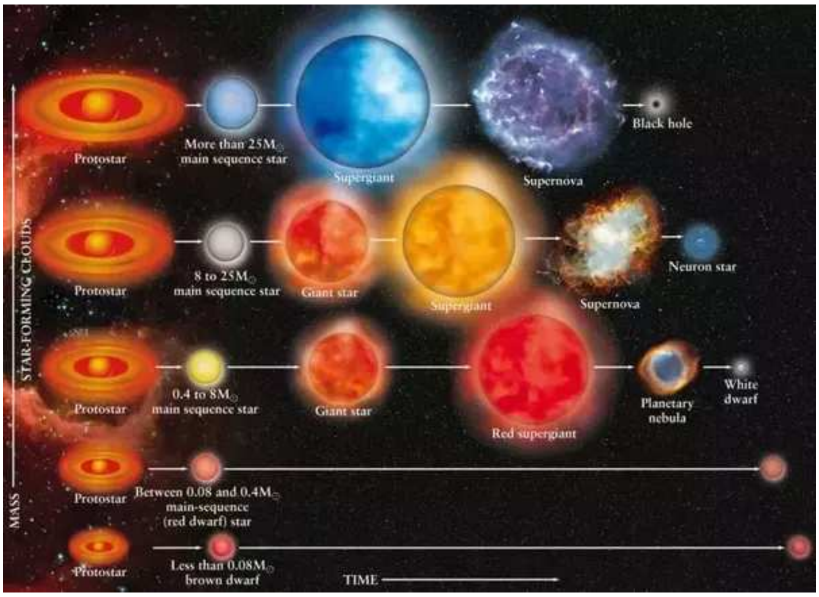


Figure 3. The fate of a star depends on its mass.

During a supernova, elements heavier than iron can be produced. Although the production of these elements by fusion is not energetically favorable, there is so much energy around in a supernova that it can still happen – like thermal motion pushing a ball up a hill. That being said, it is not clear if enough heavy elements can be produced this way to explain their abundances. In particular, it seems hard to get enough gold. Interestingly, the discovery of a neutron star mergers through gravitational waves suggest that neutron star mergers may be another important source of gold, potentially resolving this deficit (see Section 6).

After the supernova the explosion remnants are expelled off into the galaxy. Eventually they coalesce along with more primordial hydrogen to form a new generation of stars. These new stars, called **population I stars**, include our sun. Technically population I stars include stars of second, third, fourth and further generation. Our sun is believed to be a third generation star, although it, like most stars, probably formed form the coalescence of a number of differents stars as well as primordial hydrogen. About 2% of stars are population I. Population I stars have higher metallicities than first generation (population II) stars (yes, the notation is backwards) and are often surrounded by other objects with high metallicity, such as planets. Thus astronomers searching for exosolar planets focus on population I stars. Even population II stars have some heavy elements. Astronomers believe there should exist stars with only hydrogen and helium. They call these population III (generation zero?). These stars are rare, but there should still be many in our galaxy. The reason they are rare is because the original hydrogn/helium from the big bang was relatively smooth and prone for forming large stars with short lifetimes (100 million years). So there would have had to be a small patch that happened to be isolated enough not to have mixed with other supernova remnants.

3 Stellar thermodynamics

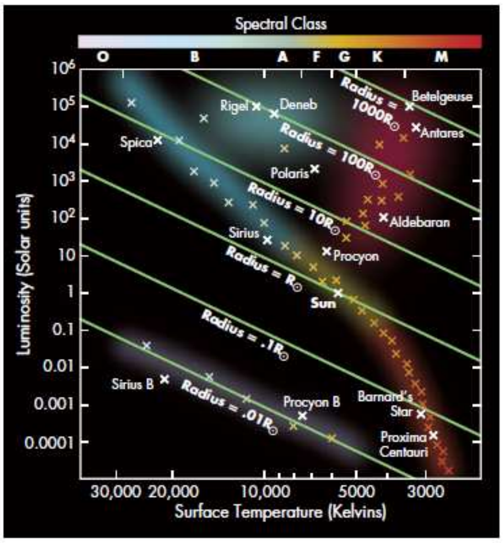
In this section, we will try to back up some of the qualitative arguments from the overview with equations. We will discuss some of the equations of stellar structure and discuss polytropes and the Eddington model, which is a very good approximation to stars in the main sequence. This will allow us to compute the density profile in the stars and the core temperature, and we will use some of the equations we derive here in the calculation of the Chandrasekhar limit in Section 4.

3.1 Luminosity

Let's begin with luminosity and temperature, as shown in the HR diagram. Knowning the intrinsic luminosity and surface temperature, we can find the radius R with the Stefan-Boltzmann law for the flux $\Phi = \sigma T_s^4$:

$$L = \text{surface area} \times \text{flux} = 4\pi R^2 \Phi = 4\pi R^2 \sigma T_s^4 \tag{3}$$

with $\sigma = 5.6 \times 10^{-8} \frac{W}{m^2 K^4}$ is Stefan's constant. So stars with the same R should fall along diagonal lines in the log-log HR diagram:



(4)

We that see that most the the main sequence stars have radii R within two orders of magnitude of the sun: $0.1R_\odot < R < 10R_\odot$. Red giants are bigger, and the white dwarfs are smaller, hence their names.

3.2 Core temperature

Although the surface temperature is what we readily see, it is the core temperature that is critical to allow nuclear fusion to occur. There are many ways to estimate the core temperature. Consider for example, a model that treats the star as two parts, a core, where fusion is occurring, and an outer shell where fusion is not occurring. The core has some pressure P that keeps the shell from collapsing in. The pressure of the shell is due to its weight as set by the gravitational pull of the core. Since $P = \frac{F}{A}$ we get

$$P_{\text{grav}} = \frac{F_{\text{grav}}}{4\pi r_{\text{core}}^2} = G\frac{M_{\text{shell}} M_{\text{core}}}{4\pi r_{\text{core}}^4} \tag{5}$$

The pressure in the core is a thermal pressure. If the core is made up of a gas of hydrogen the ideal gas law gives

$$P_{\text{therm}} = \frac{N_{\text{core}} k_B T_{\text{core}}}{V_{\text{core}}} = \frac{M_{\text{core}}}{m_p} \frac{1}{\left(\frac{4}{3}\pi r_{\text{core}}^3\right)} k_B T_{\text{core}} \tag{6}$$

where m_p is the mass of a proton. Setting $P_{\text{grav}} = P_{\text{therm}}$ gives

$$k_B T_{\text{core}} = \frac{1}{3} \frac{G M_{\text{shell}} m_p}{r_{\text{core}}} \tag{7}$$

Using dimensional analysis, we can estimate that $\frac{M_{\text{shell}}}{r_{\text{core}}} \approx \frac{M_\odot}{R_\odot}$. We then find

$$T_{\text{core}} = \frac{G M_\odot m_p}{3 k_B R_\odot} = 7.7 \times 10^6 K \tag{8}$$

Current best theoretical models put T_{core} around $15 \times 10^6 K$, so this back-of-the-envelope calculation is off by about a factor of 2. You get similar answers using the virial theorem ($E_{\text{grav}} = -2E_{\text{kin}}$), as you did on a problem set, or by equating the escape velocity to the average thermal velocity.

How hot does the core have to be for hydrogen fusion to occur? The first stage in hydrogen fusion is

$$p^+ + p^+ \to D^+ + e^+ + \nu \tag{9}$$

where D^+ is the deuteron (^{2}H nucleus). This process occurs through the weak interaction and is generally much much slower than the rate for $D + D \to {}^4$He which occurs through the strong interaction. For hydrogen fusion to occur, the protons have to be able to approach closely enough to fuse. Classically, they can only approach each other until their kinetic energy is used up fighting the potential energy barrier. The condition for this is

$$\frac{e^2}{4\pi\varepsilon_0 r_{\min}} = \frac{1}{2} m_p v^2 \tag{10}$$

Taking the velocity to be that of a thermal ideal gas, $\frac{1}{2} m_p v^2 = \frac{3}{2} k_B T$, and asking for the protons to get within the a proton radius of each other, $r_{\min} = r_p \approx 10^{-15} m \approx \frac{h}{m_p c}$, Eq. (10) then gives an estimate of the temperature

$$T_{\min}^{\text{classical}} = \frac{1}{3k_B} \frac{e^2}{4\pi\varepsilon_0 r_p} = 4.2 \times 10^9 K \tag{11}$$

This temperature is 300 times higher than we estimated for the sun's core temperature, suggesting that fusion should not occur. Of course, it is not necessary for the average proton to get over the barrier; we can exploit the far tail of the thermal distribution. Even then, the probability of having enough energy is given by the Boltzmann factor $P \sim e^{-300}$, which amounts to zero collisions in the Sun. According to this classical estimate, protons in the sun simply do not collide.

Fortunately, it is not necessary for protons to actually pass over the Coulomb barrier. Instead, they can tunnel through it quantum mechanically. To estimate the temperature where tunnelling can occur, we can use that a particle moving with velocity v has a wavefunction that extends to the size of its de Broglie wavelength $\lambda = \frac{h}{m_p v}$. Replacing $r_{\min}$ in Eq. (10) by λ leads to $\frac{e^2}{4\pi\varepsilon_0}\frac{m_p v}{2\pi\hbar} = \frac{1}{2} m_p v^2$. Taking $v = v_{\text{rms}} = \sqrt{\frac{3k_BT}{m_p}}$ this leads to

$$T_{\min} = \left(\frac{e^2}{4\pi\varepsilon_0}\right)^2 \frac{m_p}{3\pi^2\hbar^2 k_B} = 19.7\times 10^6 K \tag{12}$$

This is now within a factor of 2-3 of our estimate for the sun's core temperature. Thus, the chance of proton's overlapping enough to possibly allow fusion is order 1 and many protons in the sun can pass this threshold. A more accurate calculation puts the hydrogen fusion ignition temperature at $10\times 10^6 K$, so again our back-of-the-envelope calculation is not bad.

Going from a lower bound on the temperature for fusion to occur to the rate for fusion is very difficult. The reason is that 3 things must happen

1. protons must tunnel through electromagnetic Coulomb barrier
2. after tunneling, the strong force must to hold them together long enough for
3. the weak force to allow fusion to occur

Despite these complications involving all the forces of nature simultaneously, the calculations can be done and seem to be in excellent agreement with observations of the sun.

One prediction of hydrogen fusion in the sun, from Eq. (9), is that a boatload of neutrinos should be produced. We first measured these **solar neutrinos** in the 1960s, and only around 1/3 of the predicted flux was observed. This was called the **solar neutrino problem**. At the time the missing neutrinos were attributed to theorists bungling the (extremely complicated) nuclear physics calculations. It turns out the theorists' calculations (mostly John Bahcall) were actually nearly perfect. The missing neutrinos were due to quantum mechanical oscillations among electron, muon and tauon neutrinos, possible if and only if neutrinos have mass. When muon neutrinos were observed around the year 2000 with exactly the right solar flux, the solar neutrino problem was solved and neutrino mass was unambiguously established!

After deuterium is formed, nuclear fusion proceeds rapidly to form ^{4}He, as the $D + D \to {}^4$He, reactions involve the strong force, which has cross sections orders of magnitude larger than the $p + p \to D$ cross section. Careful calculations along these lines indicate the following table of ignition threshold temperatures for various nuclear reactions:

process	**reactions**	**minimum T**
hydrogen fusion	H→He	$10\times 10^6 K$
helium fusion	He → C,O	$100\times 10^6 K$
carbon fusion	C→O, Ne, Mg, Na	$500\times 10^6 K$
neon fusion	Ne → O, Mg	$1200\times 10^6 K$

(13)

Working backwards from these numbers and using Eq. (8) we see that for helium fusion to occur, the core of the sun would have to shrink by a factor of 10 of so, so this is what happens as a star enters the red giant phase.

3.3 Stellar structure equations

We want to do a little better than the rough estimates above. Stars are equilibrium objects and governed by a set of relatively simple equations called the **equations of stellar structure**. We will discuss two of these and show how they can be used to compute the density and temperature distributions within a star.

The mass $m(r)$ within the sphere at distance r is determined by integrating the density

$$m(r) = \int_0^r dr' \, 4\pi r'^2 \rho(r') \tag{14}$$

Taking $\frac{d}{dr}$ of both sides leads to the 1st stellar structure equation called the conservation-of-mass equation:

$$\boxed{\frac{dm(r)}{dr} = 4\pi r^2 \rho(r)} \qquad \text{(mass conservation)} \tag{15}$$

Next, we know that the gravitational force keeping a star together must balance the internal pressure pushing it apart. The force acting on the spherical shell at distance r that has mass $dm = 4\pi r^2 \rho dr$ is determined by the mass $m(r)$ within that shell. So,

$$F_g = -G\frac{m(r) \times 4\pi r^2 \rho(r) dr}{r^2} \tag{16}$$

The pressure $P(r)$ is also a function of r. There is pressure pushing out from r and pushing in from $r + dr$, so the net force on the shell is

$$F_p = 4\pi r^{\,2}\Big[\, P(r + dr) - P(r)\Big] = 4\pi r^2 \frac{dP}{dr} dr \tag{17}$$

Setting the internal pressure F_p equal to the gravitational pressure F_g gives the 2nd stellar structure equation called the equation of hydrostatic equilibrium

$$\boxed{\frac{dP(r)}{dr} = -\frac{Gm(r)\rho(r)}{r^2}} \qquad \text{(hydrostatic equilibrium)} \tag{18}$$

There are two more stellar structure equations, related to energy production and transport in the star. These depend on the rate of heat generation, the ability of the star to absorb radiation (its opacity), and the rates of conductivity and convection. We're going to skip these last two equations because they involve more complicated physics and because we can actually learn a lot from just the first two.

Multiplying both sides of the hydrostatic equilibrium equation by $\frac{r^2}{\rho(r)}$ and differentiating we get

$$\frac{d}{dr}\left(\frac{r^2}{\rho(r)} \frac{dP(r)}{dr}\right) = -G\frac{dm}{dr} = -4\pi G r^2 \rho(r) \tag{19}$$

where the mass conservation equation was used in the second step. This single equation couples the pressure and density.

To solve Eq. (19) we need to know something about the pressure or density. For example, if the density is constant, its solution is

$$P(r) = P_c - G\frac{\rho^2}{6} r^2 \tag{20}$$

with P_c the pressure at $r = 0$. Thus the pressure decays quadratically from its central value in the constant density approximation. This is not a great approximation, but a decent start.

Alternatively, we might postulate that a star is like an ideal gas dominated by adiabatic convection, so $PV^\gamma = \text{constant}$, with $\gamma = \frac{5}{3}$ for a monatomic gas. Then

$$P \propto \left(\frac{1}{V}\right)^\gamma \propto \rho^{5/3} \tag{21}$$

This is called the **adiabatic convection model**. We can then plug this in to Eq. (19) and solve. Better approximations come from understanding the sources of stellar pressure. Pressure in a star can be gas pressure, radiation pressure, or degeneracy pressure. We discuss the first two here and degeneracy pressure in Section 4.

For the gas pressure, we treat the star as an ideal gas, so $P_{\text{gas}} = \frac{N}{V} k_B T$. To use this in Eq. (19) we must relate $\frac{N}{V}$ to the mass density ρ. That is, we need to know how many independent ideal gas particles there are for every m_p of mass. The answer depends on whether the gas is ionized or not. Recall that ionization energies of atoms are in the 10 eV range (13eV for H and 79eV for He). The cores of stars are around $10^7 K \sim \text{keV}$, thus we can safely assume full ionization: all the electrons are stripped form the atoms. If the gas is 75% hydrogen and 25% ^{4}He by mass (the cosmological abundance), then for every $^4\text{He}^{2+}$ nucleus there are 12 H^+ nuclei and $2 + 12 = 14$ free electrons. So $\frac{N}{V\rho} = \frac{1 + 12 + 14}{4m_p + 12m_p} = \frac{1}{0.59m_p}$. In general, we write $\frac{N}{V} = \frac{\rho}{\mu m_p}$ with a new parameter μ so that

$$P_{\text{gas}} = \frac{\rho}{\mu m_p} k_B T \tag{22}$$

With ionized hydrogen and helium we found $\mu = 0.59$. Including the metal content of the sun, this goes up slightly, to $\mu = 0.62$.

Radiation pressure is determined by blackbody radiation:

$$P_{\text{rad}} = \frac{4\sigma}{3c} T^4 \tag{23}$$

with σ Stefan's constant. Note that matter and radiation have different scalings with temperature, so bigger, hotter stars have relatively more radiation pressure than smaller, cooler stars.

3.4 Polytropic Stellar Models

To determine how much gas and radiation pressure are present, we need the other stellar structure equations related to energy production and transport. It turns out that for hydrogen burning stars, these equations imply that the ratio $\frac{P_{\text{rad}}}{P_{\text{gas}}}$ is roughly independent of r throughout the star. This observation was made by Eddington in 1926. If $\frac{P_{\text{rad}}}{P_{\text{gas}}}$ is constant in a star, then both P_{rad} and P_{gas} are proportional to the total pressure $P_{\text{tot}} = P_{\text{rad}} + P_{\text{gas}}$. Writing $P_{\text{rad}} = \beta P_{\text{tot}}$ then $P_{\text{gas}} = (1 - \beta) P_{\text{tot}}$ and so $\frac{P_{\text{rad}}}{P_{\text{gas}}} = \frac{\beta}{1-\beta}$. We can then solve Eqs. (22) and (23) for the temperature in terms of density

$$\frac{4\sigma}{3c} T^4 = \frac{\beta}{1-\beta} \frac{\rho}{\mu m_p} k_B T \tag{24}$$

so that

$$T = \left(\frac{3ck_B}{4\mu m_p \sigma} \frac{\beta}{1-\beta} \right)^{1/3} \rho^{1/3} \tag{25}$$

Then

$$P = \frac{1}{1-\beta} P_{\text{gas}} = \frac{1}{1-\beta} \frac{k_B}{\mu m_p} \left(\frac{3ck_B}{4\mu m_p \sigma} \frac{\beta}{1-\beta} \right)^{1/3} \rho^{4/3} = \left(\frac{3ck_B^4 \beta}{4\mu^4 m_p^4 \sigma (1-\beta)^4} \right)^{1/3} \rho^{4/3} \tag{26}$$

This is called the **Eddington Solar Model**. The Eddington solar model gives an excellent approximation to our sun and most main sequence stars.

The scaling $P \sim \rho^{4/3}$ is similar to the adiabatic convection model $P \sim \rho^{5/3}$ in Eq. (21). Other models also give power laws (we show in Section 4 that a degenerate ultrarelativistic electron gas as in a white dwarf gives $P \sim \rho^{4/3}$). So in many cases, the equation of state amounts to

$$P = K \rho^{1 + \frac{1}{n}} \tag{27}$$

for some n and some K. Such models are called **polytropic models**. For the adiabatic convection model, $n = \frac{3}{2}$; for the Eddington model, $n = 3$ and

$$P = K\rho^{4/3}, \quad K = \left(\frac{3ck_B^4 \beta}{4\mu^4 m_p^4 \sigma (1-\beta)^4} \right)^{1/3} \quad \text{(Eddington model polytrope)} \tag{28}$$

It's helpful to make ρ and r dimensionless. Writing ρ_c for the density at $r = 0$, we can introduce a dimensionless radius ξ and a dimensionless function $\theta(\xi)$ with the transformations

$$\rho(r) = \rho_c [\theta(\xi)]^n, \quad \xi = \frac{r}{r_0}, \qquad r_0 = \sqrt{\frac{(n+1)K}{4\pi G \rho_c^{1-\frac{1}{n}}}} \tag{29}$$

The choice of r_0 is fixed so that when we plug these relations into Eq. (19) with the polytropic form in Eq. (27) we get a non-dimensional equation

$$\boxed{\frac{1}{\xi^2}\frac{d}{d\xi}\left(\xi^2\frac{d\theta}{d\xi}\right)=-\theta^n} \tag{30}$$

This is called the **Lane-Emden Equation** and is the basis of much stellar modeling.

The Lane-Emden equation can only be solved analytically for only a few values of n, but it's easy to solve numerically for any n by integrating. Assuming the density is non-singular (doesn't go to ∞) at $r=0$ with core density ρ_c gives the boundary conditions $\theta(0)=1$ and $\theta'(0)=0$. We then find numerical solutions for various n

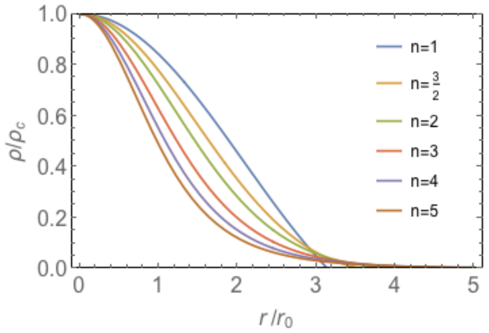


Figure 4. Density profile of a star, $\rho=\theta(\xi)^n$ with $\xi=\frac{r}{r_0}$ according for polytropes of various n.

We see that the profiles are actually qualitatively similar. For $n<5$ the curves all cross $\rho=0$, which tells us the size of the star. That is, we define $\xi_{\max}$ as the scale where $\theta(\xi_{\max})=0$, then the radius of the star is, $R=r_0\xi_{\max}$. For $n\geqslant 5$ there is no zero-crossing ($\xi_{\max}=\infty$) and the model is not a good one for any real star. In the Eddington model, $n=3$, the zero crossing $\theta(\xi_{\max})=0$ occurs at $\xi_{\max}=6.9$.

The density determines the total mass. We can write this in terms of a dimensionless integral:

$$M=4\pi\int_0^R\rho(r)r^2dr=4\pi\int_0^{\xi_{\max}}\rho_c\theta(\xi)^n(r_0\xi)^2d(r_0\xi)=4\pi\rho_c\frac{R^3}{\xi_{\max}^3}\int_0^{\xi_{\max}}[\theta(\xi)]^n\xi^2d\xi=C_nR^3\rho_c \tag{31}$$

where

$$C_n=\frac{4\pi}{\xi_{\max}^3}\int_0^{\xi_{\max}}[\theta(\xi)]^n\xi^2d\xi \tag{32}$$

is a dimensionless number that we can evaluate numerically for each n. For example, for $n=3$ we get $C_3=0.0077$. Values for $\xi_{\max}$ and C_n for some values of n are given in Table 1. These are computed in a mathematica notebook on the canvas site:

n	1	$\frac{3}{2}$	2	3	4	5
$\xi_{\max}$	π	3.7	4.4	6.9	15.0	∞
C_n	12.6	9.3	7.0	0.077	1.5	∞

Table 1. Values for the zero crossing ξ_n and the mass integral C_n for various polytropes.

Now, let's specialize to the Eddington model, with $n=3$. So we use $\xi_3=6.9$ and $C_3=0.077$. Plugging in Eqs. (29) and Eq. (28) to $R=r_0\xi_{\max}$ gives an expression for R in terms of the core density ρ_c:

$$R=r_0\xi_{\max}=\sqrt{\frac{K}{\pi G\rho_c^{2/3}}}\xi_{\max}=\xi_{\max}\left(\frac{3ck_B^4\beta}{4\pi^3\mu^4G^3m_p^4\sigma(1-\beta)^4}\right)^{1/6}\left(\frac{1}{\rho_c}\right)^{1/3}\qquad(n=3) \tag{33}$$

Since $R\sim\rho_c^{-1/3}$ for $n=3$ and $M\sim R^3\rho_c$ by Eq. (31) we see that, for $n=3$, the mass is conveniently independent of core density ρ_c:

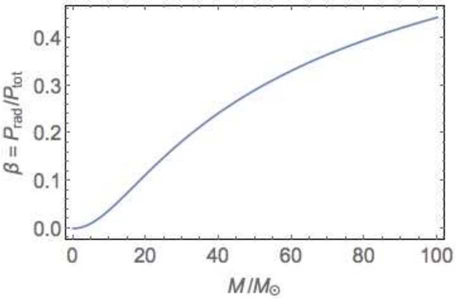
$$M=C_3R^3\rho_c=C_3\xi_{\max}^3\left(\frac{K}{\pi G}\right)^{3/2}\qquad (n=3) \tag{34}$$

Plugging in $\xi_{\max}=6.9$, $C_3=0.077$ and K from Eq. (28), we get

$$M=0.077\times(6.9)^3\times\sqrt{\frac{3ck_B^4\beta}{4\pi^3G^3\,\mu^4m_p^4\,\sigma(1-\beta)^4}}=17.9M_\odot\frac{\sqrt{\beta}}{(1-\beta)^2\mu^2} \tag{35}$$

Thus remarkably, we get a prediction for the mass of the sun within striking distance of the sun's actual mass from this model alone; we did not have to specify the radius R or the core density ρ_c.

Recall that μ refers to the chemical composition of a star. The sun has $\mu=0.62$ so we'll use this value; using this and setting $M=M_\odot$ we can solve Eq. (35) for $\beta_\odot=4.6\times10^{-4}$. Recalling that $\beta=\frac{P_{\text{rad}}}{P_{\text{tot}}}$, this indicates that there is very little radiation pressure in the sun. As the mass increases, so does β:



(36)

Thus radiation pressure is more important for heavier stars.

Plugging in $R=R_\odot$ and $M=M_\odot$ to Eq. (34) we find the sun's core density to be

$$\rho_c=\frac{M_\odot}{C_3R_\odot^3}=76.2\times10^3\frac{\text{kg}}{m^3} \tag{37}$$

Our result is about half as dense as a more accurate model predicts, $\rho_{c\odot}=156\times10^3\frac{\text{kg}}{m^3}$, but not bad. Note that the density of the sun's core is around 100 times greater than the average solar density, $\rho_{\text{avg}}=\frac{M_\odot}{\frac{4}{3}\pi R_\odot^3}=14\times10^2\frac{\text{kg}}{m^3}$.

We can invert Eq. (34) to find

$$K=\pi G\left(\frac{M}{C_n\xi_{\max}^3}\right)^{2/3} \tag{38}$$

Then the core pressure is

$$P_c=K\rho_c^{4/3}=\pi G\left(\frac{M_\odot}{C_3\xi_{\max}^3}\right)^{2/3}\left(\frac{M_\odot}{C_3R_\odot^3}\right)^{4/3}=\pi G\frac{M_\odot^2}{\xi_{\max}^2C_3^2R_\odot^4}=1.24\times10^{16}\,\text{atm} \tag{39}$$

More accurate models give $P_c=2.38\times10^{16}\,\text{atm}$, so again, we are about a factor of 2 off.

The temperature in the Eddington model is given by Eq. (22) with $P_{\text{gas}}=(1-\beta)P$:

$$T(r)=(1-\beta)\frac{\mu m_p}{k_B}\frac{P(r)}{\rho(r)} \tag{40}$$

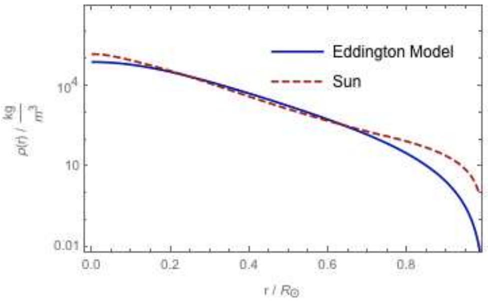
So the core temperature is

$$T_c=(1-\beta)\frac{\mu m_p}{k_B}\frac{P_c}{\rho_c}=(1-\beta)\frac{\mu}{k_B}\pi G\frac{M_\odot m_p}{\xi_{\max}^2C_3R_\odot}=12.2\times10^6K \tag{41}$$

which is quite close to the more accurate value of $T_c=15.8\times10^6\,K$.

The advantage of a complete model like this one, in constrast to simple dimensional analysis, is that it predicts the shape of the density, temperature and pressure profiles. The density is the $n=3$ curve in Fig. 4. Putting in ρ_c, we can plot our prediction and compare to a more complete and accurate model (the "Standard Solar Model" labeled "Sun" below). We get:

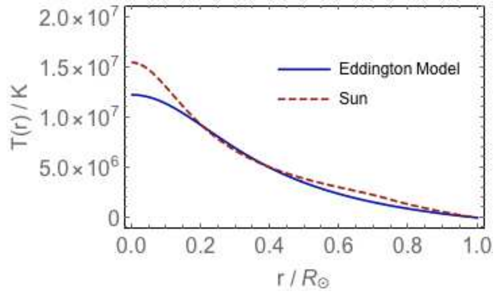
$$\rho(r) = \tag{42}$$



Other than the factor of 2 difference in ρ_C, which you can't really see in a log plot, this isn't bad.

The temperature scales like $\rho^{1/3}$ from Eq. (25), so we can compute its profile too

$$T(r) = T_c\left(\frac{\rho(r)}{\rho_c}\right)^{1/3} = \tag{43}$$



Given the rate for hydrogen fusion, which is strongly temperature and density dependent, these profiles (or more accurate ones) can be used to determine the lifetime (and age) of the sun and other stars.

Whether or not the Eddington Model is correct, we can draw some general lessons from this exercise. Since *some* equation of state exists for the star, there is some unique solution to the stellar structure equations and therefore some precise relation between the radiation pressure and the thermal gas pressure in a star. This leads to a set of one-parameter families of solutions: given a single dimensionful parameter, such as the mass, then the temperature, size, luminosity, pressure and density are all determined. This is the origin of the Vogt-Russel theorem mentioned in the introduction. Each *type* of star, with its appropriate nuclear physics and equations of state, leads to a 1D curve in the HR diagram. Thus there is the main sequence, described reasonably well by the Eddington Solar Model. For the horizontal branch (red giants), different physics is relevant, but those stars all lie on a curve. For the asymptotic giant branch, the stars line on a different curve. The width of the curves in the HR diagram are due to the stars aging – over time, their chemical composition slowly changes and the physics slowly varies, until the star explodes when the star moves to a different HR branch.

4 White dwarfs

We have discussed thermal pressure in a star and radiation pressure. Both are important in main sequence and giant stars. A third type of pressure, degeneracy pressure, is critical to determining the endpoint of stellar evolution. In particular, white dwarf stars, which have no nuclear reactions going on, are prevented from collapse due to electron degeneracy pressure alone.

Despite their lack of nuclear fusion, white dwarfs are very hot. Remember, they are remnants of the cores of giant stars that have collapsed so far that only degeneracy pressure keeps them from imploding all the way down to a black hole. The collapse to form the white dwarf starts from a very hot core, with temperatures at least $10^7\,K$ and then gets hotter as it collapses. Thus the *surface* temperature of a white dwarf is comparable the *core* temperature of a main sequence star $\sim 10^7$ K. This much much hotter than the surface temperature of the sun $\sim 5800K$. White dwarfs do not have active nuclear fusion, but the thermal motion of the charged particles (electrons and protons) causes them to radiate electromagnetically. They are so hot that the blackbody spectrum in the visible region is essentially flat, which is why they are white. Without fusion, gravity makes the white dwarf contract until it becomes extremely dense: a typical white dwarf has a density of $\rho \approx 10^{10}\frac{\text{kg}}{m^3}$. This is more than 1 million times denser than the sun $\rho_\odot \approx 1400\frac{\text{kg}}{m^3}$. A typical white dwarf has the mass of the sun and the size of the earth.

The chemical composition of white dwarfs can vary, depending on what collapsed to form them. There are helium white dwarfs, from the collapse of a star that never got hot enough to fuse helium. Most white dwarfs are mostly carbon and oxygen, from red giants or supergiants that have fused helium but could not compress enough to fuse carbon. They do not have much hydrogen, since hydrogen is all burned up at earlier stages of stellar evolution. At a temperature of $T = 10^7K$ the typical thermal energy is $k_B T = 850\,\text{eV}$. This is well above the typical ionization energies of atoms ($\sim$10eV), so white dwarfs should be thought of as fully ionized. Thus for each proton there is one free electron. The elements that can compose a white dwarf: helium, carbon or oxygen, all have roughly equal numbers of neutrons and protons, so there is around 1 electron for every $2m_p$ of stellar mass. With a density of $\rho = 10^{10}\frac{\text{kg}}{m^3}$ the number density of electrons is then

$$n_e = \frac{\rho}{2m_p} = 2 \times 10^{36}\frac{1}{m^3} \tag{44}$$

For example, recall from Lecture 13, that the Fermi energy for an non-relativistic electron gas is $\varepsilon_F = \frac{\hbar^2}{2m_e}\left(3\pi^2\frac{N}{V}\right)^{2/3}$. For the white dwarf, this evaluates to $\varepsilon_F = 450\,\text{keV}$ corresponding to a Fermi temperature of $T_F = \frac{\varepsilon_F}{k_B} = 5 \times 10^9 K$. This is much much higher than typical white dwarf temperature of $10^7 K$, so we can assume the white dwarf is *completely degenerate*. Note also that ε_F is the same order as the electron rest mass $m_e c^2 = 511$ keV. Thus the electron gas in a white dwarf can be relativistic.

We begin by considering the limit where the electrons are ultrarelativistic ($\varepsilon \gg m_e c^2$). We will justify this approximation after doing a more difficult calculation using a relativistic, but not necessarily ultrarelativistic, electron gas in Section 5.

For an electron gas, the allowed momenta are the usual

$$\vec{p} = \hbar\frac{\pi}{L}\vec{n}, \quad \vec{n} = \text{triplet of whole numbers} \tag{45}$$

If the gas is ultrarelativistic the dispersion relation is

$$\varepsilon = c\,|\vec{p}| = c\hbar\frac{\pi}{L}n \tag{46}$$

We get the density of states by the usual replacement

$$2\sum_n \to 2 \times \frac{1}{8}\int 4\pi n^2 dn = \frac{V}{c^3\pi^2\hbar^3}\int \varepsilon^2 d\varepsilon \tag{47}$$

so

$$g(\varepsilon) = \frac{V}{c^3\pi^2\hbar^3}\varepsilon^2 \tag{48}$$

Then

$$N = \int_0^{\varepsilon_F} g(\varepsilon)\,d\varepsilon = \frac{V}{c^3\pi^2\hbar^3}\int_0^{\varepsilon_F}\varepsilon^2 d\varepsilon = \frac{V}{3c^3\pi^2\hbar^3}\varepsilon_F^3 \tag{49}$$

and thus the Fermi energy is

$$\varepsilon_F = c\hbar\left(3\pi^2\frac{N}{V}\right)^{1/3} \tag{50}$$

Next, we compute the total energy of the white dwarf. Assuming complete degeneracy, we need to integrate the momenta up to p_F:

$$E=\int_0^{\varepsilon_F}\varepsilon g(\varepsilon)d\varepsilon=\frac{V}{c^3\pi^2\hbar^3}\int_0^{\varepsilon_F}\varepsilon^3 d\varepsilon=\frac{V}{4c^3\pi^2\hbar^3}\varepsilon_F^4 \tag{51}$$

With the total energy, we can now compute the pressure

$$P_{\text{degen}}=-\frac{\partial E}{\partial V}=\frac{c\hbar}{12\pi^2}\left(\frac{3\pi^2 N}{V}\right)^{4/3} \tag{52}$$

This is the **degeneracy pressure**. It is entirely due to the electrons being in excited states due to the Pauli exclusion principle (and not to thermal motion, since we are working in the $T=0$ limit).

Scaling out the mass (recalling that there are 2 electrons for each m_p of mass, so $\rho=2m_p n_e$), we can write

$$P_{\text{degen}}=\frac{c\hbar}{12\pi^2}\left(\frac{3\pi^2}{2m_p}\right)^{4/3}\rho^{4/3} \tag{53}$$

This gas is therefore a polytrope as in Eq. (27) with index $n=3$, just like the Eddington model, and with $K=\frac{c\hbar}{12\pi^2}\left(\frac{3\pi^2}{2m_p}\right)^{4/3}$. By the way we have implicitly used that the mass density ρ has the same shape as the electron density n_e, which follows from local charge neutrality.

Conveniently, we have already studied this polytropic form. Recall that the Lane-Emden equation describes hydrostatic equilibrium: the gravitational attraction is exactly counterbalanced by pressure. So using the Lane-Emden solution, we should be able to immediately find the Chandrasekhar mass. In the Lane-Emden equation we use $\rho(r)=\rho_c\theta[\xi(r)]^3$ where

$$\xi=\frac{r}{\sqrt{\frac{K}{\pi G\rho_c^{2/3}}}}\leqslant\xi_{\max}=6.89 \tag{54}$$

so that

$$R=\xi_{\max}\sqrt{\frac{K}{\pi G\rho_c^{2/3}}}=\frac{\xi_{\max}}{2}\left(\frac{3c^2\hbar^3}{16\pi\rho_c^2 m_p^4 G_N^3}\right)^{1/6} \tag{55}$$

Plugging this into Eq. (34) gives a bound on the mass that can be supported by degeneracy pressure alone

$$M\leqslant C_3R^3\rho_c=\frac{C_3\xi_{\max}^3}{32}\sqrt{\frac{3c^3\hbar^3}{\pi G_N^3 m_p^4}}=0.77\sqrt{\frac{c^3\hbar^3}{G_N^3 m_p^4}}=1.41M_\odot \tag{56}$$

This is known as the **Chandrasekhar limit**. If a white dwarf has a mass heavier than this limit, its gravitational attraction will overwhelm its degeneracy pressure and it will collapse.

Most white dwarfs in our galaxy whose masses can be measured (those in binary systems), have masses around $0.5M_\odot$. One of the most massive white dwarfs known is Sirius B. It is the companion of Sirius A, the “dog star”, a main sequence star that is the brightest in the sky. To find the Sirius binary system, follow Orion’s belt. Because Sirius B is in a binary system, we can measure its mass to be $1.02M_\odot$. This is, of course, within the Chandrasekhar bound.

While the ultrarelativistic limit has let us compute the Chandrasekhar limit relatively quickly, the radius R has completely dropped out of the result. This is an artifact of the $n=3$ polytropic form (it also dropped out in the Eddington model in Eq. (35)). Thus we cannot compute the density of the star and verify that we are in the ultrarelativistic limit. Moreover, it is also questionable to apply the ultrarelativistic limit for the entire energy integral because the lowest energy states are necessarily non-relativistic. We next re-calculate the bound not assuming the ultrarelativistic limit. This will confirm and justify Eq. (56).

5 Complete Chandrasekhar limit calculation

The calculation from Section 4 assumed the electrons were ultrarelativistic. This let us reuse results from the $n=3$ polytrope that describes our sun to quickly get the Chandrasekhar bound. In this section, we perform a more complete calculation, not assuming an ultrarelativistic dispersion relation. This will allow us to justify the ultrarelativistic limit from Section 4 and also allow us to determine the radius and density of a white dwarf.

If you are exhausted and already believe the ultrarelativistic limit, feel free to only skim this section. The main result is shown in Fig. 6. Please at least make you understand what is being plotted in this figure.

5.1 Relativisitic case

For a relativistic but not necessarily ultrarelativistic electron, the dispersion relation is

$$\varepsilon(p)=\sqrt{m_e^2c^4+c^2p^2} \tag{57}$$

with m_e the electron mass. Because of this awkward square root, we will integrate using the momentum rather than the energy. So we replace

$$2\sum_n \to 2\times\frac{1}{8}\int 4\pi n^2 dn=\frac{V}{\pi^2\hbar^3}\int p^2 dp \tag{58}$$

That is, the density of momentum states is

$$g(p)=\frac{V}{\pi^2\hbar^3}p^2 \tag{59}$$

Correspondingly, instead of the Fermi energy, it will be useful to us the **Fermi momentum** p_F, defined so that

$$N=\int_0^{p_F} g(p)\,dp=\frac{V}{\pi^2\hbar^3}\int_0^{p_F} p^2 dp=\frac{V}{3\pi^2\hbar^3}p_F^3 \tag{60}$$

Thus,

$$p_F=\hbar\left(3\pi^2\frac{N}{V}\right)^{1/3} \tag{61}$$

The Fermi energy is related to the Fermi momentum by

$$\varepsilon_F=\varepsilon(p_F)-\varepsilon(0)=\sqrt{m_e^2c^4+c^2p_F^2}-m_ec^2 \tag{62}$$

Next, we compute the total energy of the white dwarf. Assuming complete degeneracy, we need to integrate the momenta up to p_F:

$$E=\int_0^{p_F}\sqrt{m_e^2c^2+p^2}\,g(p)dp=\frac{Vc}{\pi^2\hbar^3}\int_0^{p_F}\sqrt{m_e^2c^2+p^2}\,p^2dp \tag{63}$$

Changing variables to $x=\frac{p}{m_e c}$ gives

$$E=\frac{Vm_e^4c^5}{\pi^2\hbar^3}\int_0^{x_F}\sqrt{1+x}x^2\,dx=\frac{Vm_e^4c^5}{8\pi^2\hbar^3}f(x_F) \tag{64}$$

where

$$x_F=\frac{p_F}{m_ec}=\frac{\hbar}{m_ec}\left(3\pi^2\frac{N}{V}\right)^{1/3} \tag{65}$$

and

$$f(x)=8\int_0^{x_F}\sqrt{1+x}x^2\,dx=\sqrt{1+x^2}(x+2x^3)-\ln\left(x+\sqrt{1+x^2}\right) \tag{66}$$

With the energy, we can now compute the pressure

$$P_{\text{degen}}=-\frac{\partial E}{\partial V}=\frac{m_e^4c^5}{8\pi^2\hbar^3}\left[-f(x_F)-f'(x_F)V\frac{\partial x_F}{\partial V}\right] \tag{67}$$

This simplifies to

$$\boxed{P_{\text{degen}}=\frac{m_e^4c^5}{8\pi^2\hbar^3}\left[\frac{1}{3}x_F^3\sqrt{1+x_F^2}(2x_F^3-3)+\ln\left(x_F+\sqrt{1+x_F^2}\right)\right]} \tag{68}$$

This is the formula for the degeneracy pressure for a relativistic electron gas.

As a check, we can expand in the ultra-relativistic limit, $x_F \gg 1$ giving

$$P_{\text{degen}}^{\text{ultra}-\text{rel}} = \frac{m_e^4 c^5}{12\pi^2\hbar^3} x_F^4 = \frac{c\hbar}{12\pi^2}\left(\frac{3\pi^2 N}{V}\right)^{4/3} \tag{69}$$

in agreement with Eq. (52). We can also work out the non-relativistic limit, by expanding to leading order in x_F:

$$P_{\text{degen}}^{\text{non}-\text{rel}} = \frac{m_e^4 c^5}{15\pi^2\hbar^3} x_F^5 = \frac{\hbar^2}{5m_e}(3\pi^2)^{2/3}\left(\frac{N}{V}\right)^{5/3} \tag{70}$$

As in the ultrarelativistic case, we can use $N = \frac{\rho}{2m_p}$ to write this as a polytrope

$$P_{\text{degen}}^{\text{non}-\text{rel}} = K\rho^{5/3}, \quad K = \frac{\hbar^2}{10 m_e m_p}\left(\frac{3\pi^2}{2m_p}\right)^{2/3} \tag{71}$$

So this polytrope has index $n = \frac{3}{2}$, like an adiabatic ideal gas.

5.2 Equilibrium

For the white dwarf to be in equilibrium, the outward pressure which tries to increase the volume must be compensated by the inward pressure from gravity. To compute the inward pressure, we need the volume dependence of the gravitational energy. If we assume uniform density, the gravitational energy is

$$E_{\text{grav}}^{\text{const dens.}} = -\int dr \frac{G}{r} \times \overbrace{\left(\frac{4}{3}\pi\rho r^3\right)}^{\text{mass inside shell at } r} \times \overbrace{(4\pi r^2\rho)}^{\text{mass of shell at } r} dr = -\frac{16}{15}\pi^2\rho^2 G R^5 = -\frac{3}{5}G\frac{M^2}{R} \tag{72}$$

where $M = \frac{4}{3}\pi R^3 \rho$ was used in the last step.

Of course, ρ is not constant. To get an improved calculation of the gravitational energy, we can use the density profile from the Lane-Emden equation using one of the polytropic form limits. The energy integral can be reduced to a scaleless integral similar to Eq. (31):

$$E_{\text{grav}} = -\frac{16\pi^2}{3}G\int_0^R dr\, r^4 \rho(r)^2 = -\frac{16\pi^2}{3}G\rho_c^2\frac{R^5}{\xi_{\max}^5}\int_0^{\xi_{\max}}[\theta(\xi)]^{2n}\xi^4 d\xi \tag{73}$$

Then using Eq. (31), $M = C_n R^3 \rho_c$, we can write

$$E_{\text{grav}} = -D_n G\frac{M^2}{R}, \quad D_n = \frac{16\pi^2}{3\xi_{\max}^5 C_n^2}\int_0^{\xi_{\max}}[\theta(\xi)]^{2n}\xi^4 d\xi \tag{74}$$

For $n = 3$ (ultra-relativistic case) we have as before that $\xi_{\max} = 6.9$ and $C_3 = 0.077$ and now find $D_3 = 0.68$. You can work out the numbers for the non-relativistic case yourself, as we won't use them here. It's perhaps illuminating to compare the constant and ultrarelativistic density profiles. Normalizing to the same total mass M:

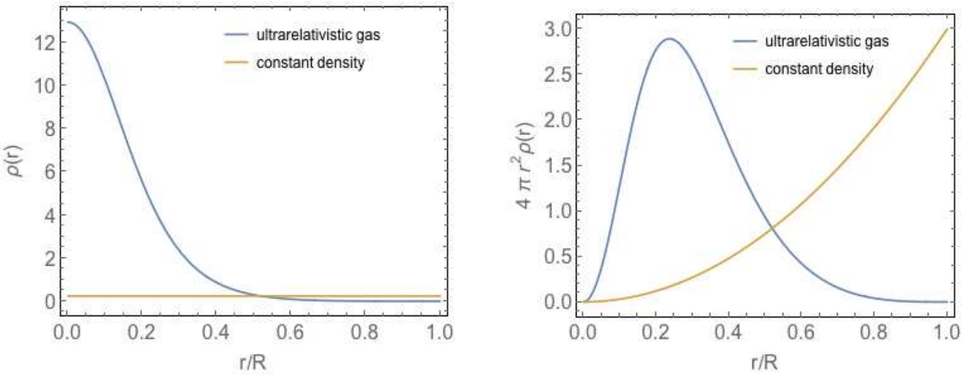


Figure 5. Density profiles for the constant-density assumption and an ultrarelativistic degenerate electron gas assumption (the $n = 3$ polytrope). Right shows $4\pi r^2 \rho(r)$ normalized to integrate to 1.

Writing in general $E_{\text{grav}} = -DG\frac{M^2}{R}$, the gravitational pressure is, using $V = \frac{4}{3}\pi R^3$ and $\frac{\partial V}{\partial R} = 3\frac{V}{R}$

$$P_{\text{grav}} = -\frac{\partial E_{\text{grav}}}{\partial V} = \frac{\partial}{\partial R}\left(DG\frac{M^2}{R}\right)\frac{\partial R}{\partial V} = -\left(DG\frac{M^2}{R^2}\right)\left(\frac{R}{3V}\right) = -\frac{DGM^2}{4\pi R^4} \tag{75}$$

Thus degeneracy pressure can hold off the collapse if

$$P_{\text{degen}} \geqslant D\frac{GM^2}{4\pi R^4} \tag{76}$$

As a consistency check, we compare to the ultra-relativistic case. Using the ultra-relativistic form for P_{degen} in Eq. (69) and replacing $\frac{N}{V} = \frac{M}{2m_p}\frac{1}{\frac{4}{3}\pi R^3}$ and $D = D_3 = 0.68$ we need to solve

$$P_{\text{degen}}^{\text{ultra}-\text{rel}} = \frac{c\hbar}{12\pi^2}\left(3\pi^2\frac{M}{2m_p}\frac{3}{4\pi R^3}\right)^{4/3} \geqslant D_3\frac{GM^2}{4\pi R^4} \tag{77}$$

Here we see that R drops out and so

$$M \leqslant \frac{9\sqrt{3\pi}}{64D_3^{3/2}}\sqrt{\frac{c^3\hbar^3}{G_N m_p^4}} = 1.41\,M_\odot \tag{78}$$

Which is the same as we found using Eq (56).

Back to the general relativistic case, with the same replacement $\frac{N}{V} = \frac{3M}{8m_p R^3}$ we can write Eq. (65) as

$$x_F = \frac{\hbar}{m_e cR}\left(\frac{9\pi M}{8m_p}\right)^{1/3} \tag{79}$$

In terms of this function x_F, the equilibrium condition in Eq. (76) with Eq. (68) becomes:

$$\boxed{\frac{1}{3}x_F^3\sqrt{1+x_F^2}(2x_F^3-3) + \ln\left(x+\sqrt{1+x^2}\right) = D\left(\frac{8\pi^2\hbar^3}{m_e^4c^5}\right)\frac{GM^2}{4\pi R^4}} \tag{80}$$

5.3 Mass radius relation

To study Eq. (80) we first put in some numbers. Using $V = \frac{4}{3}\pi R^3$ and $M = 2m_p N$ we can write x_F in Eq. (65) as

$$x_F(R) = \frac{\hbar}{m_e c}\left(3\pi^2\frac{\frac{M}{2m_p}}{\frac{4}{3}\pi R^3}\right)^{1/3} = \frac{\hbar}{m_e cR}\left(\frac{9\pi M}{8m_p}\right)^{1/3} = 0.97\left(\frac{R_E}{R}\right)\left(\frac{M}{M_\odot}\right)^{1/3} \tag{81}$$

where $M_\odot = 1.98 \times 10^{30}$kg is the mass of the sun and $R_E = 6370$km is the radius of the earth. Plugging in m_e and the other constants, we can also write the right-hand side of Eq. (80) as

$$D\left(\frac{8\pi^2\hbar^3}{m_e^4c^5}\right)\frac{1}{4\pi}G\frac{M^2}{R^4} = 0.69D\left(\frac{M}{M_\odot}\right)^2\left(\frac{R_E}{R}\right)^4 \tag{82}$$

Because $f(x)$ is an order-1 function and x_F and the right hand side of (80) are all the same order when $R \sim R_E$ and $M \sim M_\odot$, we expect that the solutions will also have this form.

Note that we are not making any claims about whether M is close to $M_\odot$ or R is close to R_E, we are just writing x_F in a suggestive way. In other words, it so happens that $M_\odot$ and R_E are close to combinations of constants that appear in x_F:

$$\sqrt{\frac{c^3\hbar^3}{G_N^3 m_p^4}} = 1.8M_\odot, \qquad \sqrt{\frac{\hbar^3}{cG_N m_e^2 m_p^2}} = 0.78\,R_E \tag{83}$$

For example, Sirius B, the white dwarf partner of the bright Sirius A, has a mass of $1.02M_\odot$. Now we can calculate its radius by numerically solving Eq. (80). We get $R=0.54R_E$, confirming our estimates of density from the beginning of this section. Most white dwarfs are around $0.5M_\odot$; for these, we find $R\approx R_E$.

Using the value of $D=D_3=0.68$ we computed using the ultrarelativistic polytrope, we can solve Eq. (80) numerically to get the mass-radius relation:

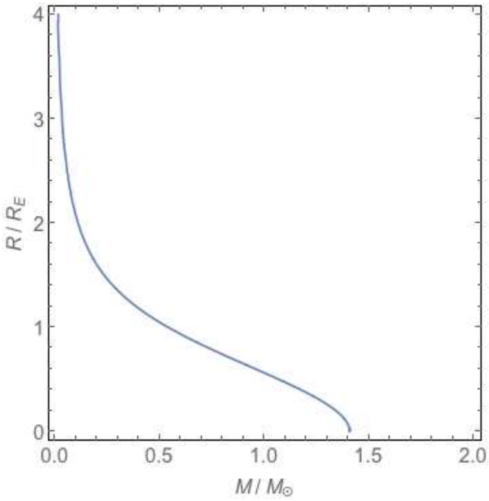


Figure 6. The relationship between mass and size of a white dwarf determined by equating the degeneracy pressure to the gravitational pressure. Axes are normalized to the sun's mass and the earth's radius.

We see from the plot that for $M\gtrsim 1.4M_\odot$ there are no solutions. Having included full relativistic corrections, we can also check our assumptions. Note that the bound comes from small R. As $R\to 0$ the gas gets denser and denser, so the electrons become faster and faster. We can see this also from Eq. (79) where $x_F\sim\frac{1}{R}$ so as $R\to 0$, x_F gets large. Expanding at large x_F is exactly the ultrarelativistic limit. Therefore the bound is as in Eq. (78):

$$M\leqslant\frac{9\sqrt{3\pi}}{64D_3^{3/2}}\sqrt{\frac{c^3\hbar^3}{G_N m_p^4}}=1.41M_\odot \tag{84}$$

This is called the **Chandrasekhar limit**. For masses of white dwarfs greater than the Chandrasekhar limit, there is no solution to equating the degeneracy pressure with the gravitational pressure. The gravitational pressure is just too great, and the white dwarf will collapse (ultimately to a neutron star or a black hole).

6 Neutron stars

When a star's total mass is above around $8M_\odot$, the mass around the core will be larger than the Chandrasekhar limit. In this case after helium fusion ends and the core temperature is not hot enough for carbon fusion to occur, the gravitational pressure will overwhelm the electron degeneracy pressure and the core will collapse past the white dwarf stage. As mentioned in Section 2, the further compression heats up the core, allowing the remaining stages of fusion to occur, making elements up to iron. After these stages finish, the star collapses again, reaching such high densities that the protons smash into the electrons producing neutrons and neutrinos. This $p^+ + e^- \to n+\nu$ reaction is energetically favorable because the neutrinos stream out of the star relieving some of the pressure. Ultimately, only neutrons are left.

Neutrons are fermions, and the core of neutrons in a neutron star is a Fermi gas, like a white dwarf. We can determine the characteristic size and mass of a neutron star by substituting in the appropriate scale, namely, replacing m_e by $m_n = m_p$. For a white dwarf, recall the characteristic mass and radius as in Eq. (83):

$$M_{\mathrm{WD}} \approx \sqrt{\frac{c^3\hbar^3}{G_N^3 m_p^4}} = 3.6 \times 10^{30}\mathrm{kg} = 1.8 M_\odot, \qquad R_{\mathrm{WD}} \approx \sqrt{\frac{\hbar^3}{c G_N m_e^2 m_p^2}} = 4970\,\mathrm{km} = 0.78 R_E \tag{85}$$

We see that the mass scale does not depend on m_e, so it is unaffected by $m_e \to m_p$ and so should be around the same for a neutron star. More precisely, if we recall that in a white dwarf there is 1 electron for every $2m_p$ of mass so that $n = \frac{\rho}{2m_p}$ in a neutron star we have more simply $n = \frac{\rho}{m_p}$. This amounts to replacing $m_p \to \frac{m_p}{2}$ and $m_e \to m_p$ and thus $M_{\mathrm{NS}} \approx 4 M_{\mathrm{WD}}$ and so

$$M_{\mathrm{NS}} \lesssim 4 \times 1.4 M_\odot = 5.6 M_\odot \tag{86}$$

This is our estimate for the Chandrasekhar bound for neutron stars. The actual bound should be a bit lower since the binding energy of the neutrons, effects from general relativity, and rotational energy cannot be neglected. One estimate, called the **Tolman–Oppenheimer–Volkoff limit**, is $M_{\mathrm{NS}} \lesssim 3 M_\odot$, but this estimate is controversial. Determining a precise upper bound on the mass of a neutron star is still an open theoretical question. The largest observed neutron stars to date is around $2.7 M_\odot$.

The characteristic size of a neutron star is determined from the size of the white dwarf with the $m_p \to \frac{m_p}{2}$ and $m_e \to m_p$ replacements:

$$R_{\mathrm{NS}} \approx \sqrt{\frac{4\hbar^3}{c G_N m_p^4}} = 5\,\mathrm{km} \tag{87}$$

Thus, although neutron stars have similar masses to white dwarfs, they are much much denser. The whole mass of the sun is being squeezed into $5\,\mathrm{km} = 3\,\mathrm{miles}$ – roughly the size of Cambridge. The density of such a star is

$$\rho_{\mathrm{NS}} \approx \frac{M_\odot}{\frac{4}{3}\pi R_{\mathrm{NS}}^3} = 3 \times 10^{18}\frac{\mathrm{kg}}{m^3} \tag{88}$$

Compare this to the density of water $\rho_{\mathrm{water}} = 10^3 \frac{\mathrm{kg}}{m^3}$, to the core of the sun $\rho_{c\odot} = 10^5 \frac{\mathrm{kg}}{m^3}$, or to the density of a white dwarf, $\rho_{\mathrm{WD}} \approx 10^9 \frac{\mathrm{kg}}{m^3}$. None of these are even close. In fact, the density of a neutron star is larger than the density of a proton: $\rho_{\mathrm{proton}} \approx \frac{m_p}{\frac{4}{3}\pi(10^{-15}m)^3} = 10^{17}\frac{\mathrm{kg}}{m^3}$! That is, while a white dwarf is like a gigantic metal, a neutron star is like a gigantic nucleus, with atomic number 10^{57}.

Such a giant nucleus has a lot of strange properties. By angular momentum conservation, a slowly rotating star will collapse to a extremely rapidly rotating neutron star: like an ice skater bringing her arms in by a factor of a million. Neutron stars can be spinning as fast as 1000 times per second. The spinning also concentrates the magnetic field of the star, to as much as 10^{15} Gauss. Compare this to the earth's magnetic field (0.5 G) or the fields in magnets at the Large Hadron Collider (5×10^4 G). Such large magnetic fields can act as a dynamo generating large electric fields near the star's surface. These enormous electric fields then create electron/positron pairs which are thrown out of the star, radiating electromagnetically. The result is a beam of light, spinning around as the neutron star spins, like an out-of-control lighthouse beacon. If the beam is aligned to hit us, we see this is a periodic signal. Such neutron stars are known as **pulsars**.

In July, 1054 AD, a Chinese astronomer observed a new "guest star", brighter than any other star in the sky. It lasted for about a month and then faded. Arab astronomers observed the same object, and perhaps Native American astronomers as well. We now know that this was a core-collapse supernova. 1000 years later, the supernova remnants are visible as the **crab nebula**, in the constellation Taurus. There is a neutron star in the middle of the crab nebula called the crab pulsar. The pulsar has a mass of $1.4 M_\odot$ and period of $0.3s$.

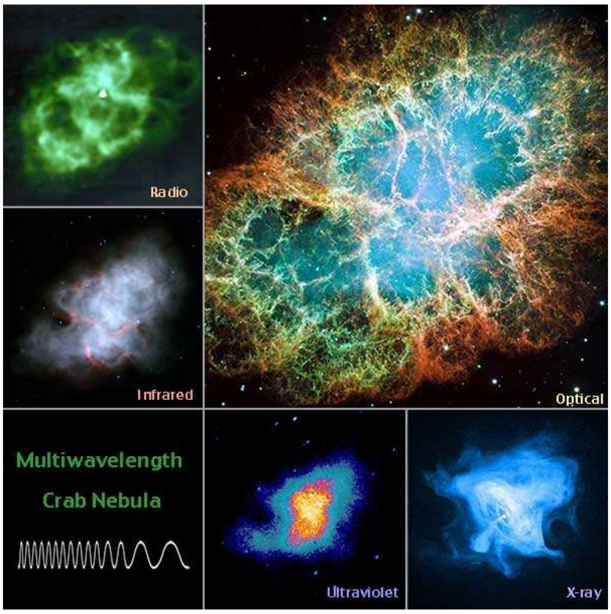


Figure 7. The crab nebula is around 10 light years across. In its center is a neutron star pulsar. The neutron star cannot be seen in the optical spectrum, but is clearly visible in the radio and x-ray bands.

In 2017, a binary neutron star system was discovered through its gravitational wave signal as the neutron stars merged. The neutron stars had masses around $1.5M_{\odot}$. One interesting mystery that such mergers might explain is where all the gold in the universe came from. It turns out it is very difficult to explain how gold might have come from supernovae. However, the nuclear physics of neutron star mergers seems like it gives a better explanation. This science is all very recent, and while early results seem promising, more data and more careful calculations and simulations are needed to draw any definite conclusions.

7 Summary

This was a long lecture and covered a lot of Stellar physics. The introductory sections explained how stars are classified and typical lifecycle: hydrogen collapses gravitationally until it gets hot and dense enough for fusion to occur. Hydrogen is fused into helium, and a succession of fusion processes allow elements to turn into heavy elements up until iron which is the most stable element. Fusion requires a high enough temperature to occur, which in turn requires a high enough density. After one element is burned up, if the subsequent ignition threshold cannot be reached, fusion stops, thermal radiation pressure ceases, then the star contracts and explodes into a supernova.

Stellar physics is complicated, but some approximate results can be derived without too much work. A useful model that seems phenomenologically accurate is the polytropic stellar model ,where one assumes that the equation of state relating pressure and density is a power law: $P = K\rho^{1+\frac{1}{n}}$ for some K and n. A special case is the Eddington solar model where $n = 3$. Using only this assumption and equations of stellar structure, we computed the density profile $\rho(r)$. Given the total mass of the sun, we the predict its core density to be $\rho_c \sim 10^4\frac{\text{kg}}{m^3}$ which is off by about a factor of two from the predictions from more accurate numerical simulations. We also predicted the core temperature of the sun to be 12.2×10^6K, which is close to its actual core temperature of $T_c = 15.8\times10^6K$. A general lesson is that if we know the equation of state, then inputing the total mass M lets us predict the temperature, core density, luminosity, etc. This justifies the Vogt-Russel theorem: stars fall into categories based on their equation of state (what's going on in the star). Each category has only one free pamameter, the mass. This explains why stars populations fall along lines or thin bands in the Hertzsprung-Russel diagram.

The main application of statistical mecahnics to stellar physics came in our discussion of white dwarfs and neutron stars. These objects have exhausted the nuclear fuel and are held in equilibrium by a balance of gravitational attraction and repulsion caused by degeneracy pressure. Degeneracy pressure is purely quantum effect. Quantum statistical mechanics is therefore required to understand white dwarfs.

White dwarfs are like very dense metals. A typical white dwarf has the mass of the sun and the size of the earth. The degeneracy pressure in a white dwarf is due to electrons. It scales like $P_{\text{degen}} \sim \rho^{4/3}$, so that white dwarfs are an $n = 3$ polytope, just like the sun in the Eddington model. Recycling previous results, we then found that for the degeneracy pressure to not be overwhelmed by gravity, the mass of the white dwarf had to be smaller than 1.44 times the sun's mass: $M \lesssim 1.41 M_{\odot}$. This is called the Chandrasekhar limit. We first did the calculation quickly, assuming that the electrons were relativistic. We then did it more carefully, without making the relativistic assumption, and arrived at the same Chandrasekhar limit

Neutron stars are essentially all neutrons. They are like big nuclei. A typical neutron star might have the mass of the sun but be only 3 miles across. For neutron stars the bound comes from simple replacments $m_e \rightarrow m_p$ in the white dwarf calculation. This results in $M \lesssim 5.6 M_{\odot}$ or else gravity will overwhelm the (neutron) degeneracy pressure. There are other effects in neutron stars, such as from general relativity, that make the more accurate bound closer to $M \lesssim 3 M_{\odot}$. No neutron star has been observed with mass bigger than 2.7 $M_{\odot}$.