Physics 12c: Introduction to Statistical Mechanics

David Simmons-Duffin

Course website

About these notes

These notes are not original. Their primary purpose is to remind me what to say during lectures. They draw heavily from several sources, including Kittel and Kroemer, Statistical Physics lecture notes by David Tong[1] and Matt Schwartz[2], lecture notes from John Preskill's 2016 Physics 12c course, and possibly other sources. Last updated: September 2, 2020.

---

[1]http://www.damtp.cam.ac.uk/user/tong/statphys.html
[2]http://users.physics.harvard.edu/~schwartz/teaching

Contents

List of definitions

1   Counting

1.1   Macrostates and microstates

Classical mechanics and quantum mechanics give precise predictions for almost any physical observable, within their respective realms of validity. For example, to make a prediction using quantum mechanics, one must determine the initial state $|\psi(t=0)\rangle$ of a system and the Hamiltonian H that generates time evolution. Then one simply evolves the state forward in time using the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle. \tag{1.1}$$

This is already an interesting problem for an individual Hydrogen atom. In the approximation that the proton is stationary and (and ignoring the spin of the proton), we can write the state as a linear combination of position and spin eigenstates for the electron

$$|\psi_{\text{single-electron}}(t)\rangle = \sum_{s=\uparrow,\downarrow}\int d^d\mathbf{x}\,\psi_s(\mathbf{x},t)|\mathbf{x},s\rangle, \tag{1.2}$$

where $s=\uparrow,\downarrow$ is the spin of the electron, and $d=3$ is the number of spatial dimensions. The Schrodinger equation becomes a partial differential equation (PDE) for two functions $\psi_{\uparrow,\downarrow}(\mathbf{x},t)$ of $d+1=4$ variables (that is, the number of spatial dimensions plus the time).

Now let us use this framework to describe the weather. The number of molecules in the atmosphere is approximately 10^{44}. So all we need to do is solve the Schrödinger equation, which is a PDE for a function $\psi_{s_1,\dots,s_{10^{44}}}(\mathbf{x}_1,\dots,\mathbf{x}_{10^{44}},t)$ of $3\times 10^{44}+1\approx 3\times 10^{44}$ variables.

Even to achieve a more modest goal of exactly modeling a handful of a typical substance, we would need to solve the Schrödinger equation for a wavefunction with $O(N_A)=O(10^{23})$ variables, where N_A is Avagodro's number.

Of course, these are impossible computational tasks. But that doesn't stop us from doing physics. The reason is that the complete wavefunction of a system, including the positions and spins of all the constituent particles, contains much more information than we usually care about. We can still make predictions without all that detailed information.

Let us introduce some terminology.

Definition (microstate/state). A *microstate* is a complete specification of the physical state of a system. We often use the term *state* to mean a microstate.

In the context of quantum mechanics, a microstate is a complete specification of the wavefunction of the system. In the context of classical mechanics, a microstate is a complete list of the positions and momenta (and other degrees of freedom) of all the constituent particles $\{\mathbf{x}_1,\dots,\mathbf{x}_N,\mathbf{p}_1,\dots,\mathbf{p}_N\}$.

Definition (macrostate). A *macrostate* is a partial specification of the state of a system, typically involving a small number of quantities. A single macrostate corresponds to a large number of microstates.

For example, we might specify a macrostate of a gas by giving its total energy U and its volume V. A more complicated example of a macrostate is a description of the atmosphere using fluid dynamics in terms of a density field $\rho(\mathbf{x}, t)$ and velocity field $\mathbf{v}(\mathbf{x}, t)$ satisfying the Navier-Stokes equations. Although this is much more information that the two numbers $\{U, V\}$, it is vastly less information than the full quantum wavefunction $\psi(\mathbf{x}_1, \ldots, \mathbf{x}_{10^{44}})$, so it is still correct to call $\{\rho(\mathbf{x}, t), \mathbf{v}(\mathbf{x}, t)\}$ a macrostate of the atmosphere.

For practical purposes, we usually only measure macroscopic quantities, i.e., we measure the macrostate of a system. The idea of statistical mechanics is to avoid studying the precise dynamics of individual microstates by instead describing the *statistics* of microstates using probability theory.

Statistics and probability theory create a bridge between microscopic laws and macroscopic phenomena. We will see that they explain many different physical processes and have lots of applications.

1.2   Large numbers and sharp distributions

Using statistics is a good idea because of the incredibly enormous numbers involved. The fact that these numbers are so huge means that statistical predictions can be ridiculously good.

As an example, let us study a toy model. Consider a system with N sites on a line, where each site can be in one of two states $\uparrow, \downarrow$. For concreteness, imagine that each site contains a small magnet which can point either up or down, with magnetization $m_i = m s_i$ with $s_i = \pm 1$, $(i = 1, \ldots, N)$. We refer to s_i as the "spin" at site i. A state is specified by giving the spins at each site. An example state for the case $N = 60$ is[3]

$$\downarrow\downarrow\downarrow\downarrow\uparrow\uparrow\downarrow\uparrow\downarrow\downarrow\uparrow\uparrow\downarrow\uparrow\uparrow\uparrow\uparrow\downarrow\downarrow\uparrow\uparrow\uparrow\downarrow\uparrow\uparrow\downarrow\uparrow\downarrow\uparrow\uparrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\uparrow\uparrow\downarrow\downarrow\uparrow\downarrow\uparrow\uparrow\uparrow\downarrow\downarrow\uparrow\downarrow\uparrow\uparrow\uparrow\downarrow\downarrow\uparrow\uparrow\downarrow\downarrow\downarrow \qquad (1.3)$$

The total number of states is 2^N.

Let us imagine coupling these spins to some random environment and letting them evolve for a while. For simplicity, we assume the spins do not interact strongly with each other and that they have no particular preference for pointing up or down. After a while, let us measure the total magnetization, given by the sum of the magnetic moments at each individual site:

$$M = m \sum_{i=1}^{N} s_i. \qquad (1.4)$$

---

[3]Mathematica code: `StringJoin[Table[RandomChoice[{"\uparrow","\downarrow"}],60]]`

What will we observe?

For convenience, let us get rid of the factors of m and discuss a slightly more convenient quantity, the "spin excess"

$$s = \frac{M}{2m} = \frac{1}{2}\sum_{i=1}^{N} s_i = \frac{1}{2}(N_\uparrow - N_\downarrow), \tag{1.5}$$

where $N_\uparrow$ and $N_\downarrow$ are the number of spins pointing up and down, respectively. Using the fact that $N = N_\uparrow + N_\downarrow$, we find

$$N_\uparrow = \frac{1}{2}N + s, \qquad N_\downarrow = \frac{1}{2}N - s. \tag{1.6}$$

Note that s can take $N + 1$ different values $s \in \{-\frac{N}{2}, -\frac{N}{2} + 1, \ldots, \frac{N}{2} - 1, \frac{N}{2}\}$. If N is large, the number of possible values of s (macrostates) is much smaller than the total number of microstates of the system, which is 2^N. This means that some macrostates correspond to a huge number of microstates. We see that the most extreme macrostates $s = \pm\frac{N}{2}$ have one microstate each. The next most extreme $\pm\left(\frac{N}{2} - 1\right)$ have N microstates each. You may be able to guess that most of the microstates are bunched near middle values of s. Let us count them more precisely.

The number of states with spin excess s is given by the number of ways to choose $N_\uparrow$ objects from a set of size N:

$$g(N, s) = \binom{N}{N_\uparrow} = \binom{N}{N_\downarrow} = \frac{N!}{N_\uparrow!N_\downarrow!} = \frac{N!}{(\frac{1}{2}N + s)!(\frac{1}{2}N - s)!}. \tag{1.7}$$

Another way to derive this result is using generating functions. First let us fix a given state. For each site, associate a symbolic factor x to $\uparrow$ and a factor x^{-1} to $\downarrow$. After multiplying the factors associated to all the sites, the result is $x^{N_\uparrow - N_\downarrow} = x^{2s}$. Now, to count all possible states, we can associate to each site an expression $x + x^{-1}$. We claim that the coefficient of x^{2s} in

$$Z(x) = \underbrace{(x + x^{-1})(x + x^{-1}) \cdots (x + x^{-1})}_{N \text{ factors}}$$

$$\tag{1.8}$$

is the number of states with spin excess s. To see this, note that each configuration (1.3) is in one-to-one correspondence with a term in the expansion of $Z(x)$

$$\begin{aligned}
Z(x) = {}& xx \cdots x \\
& + xx^{-1}x \cdots x \\
& + xxx^{-1} \cdots x + (2^N - 3 \text{ more terms}).
\end{aligned} \tag{1.9}$$

Each configuration with spin excess s contributes x^{2s}, so the coefficient of x^{2s} is the total number of states with spin excess s. We can compute (1.8) with the binomial theorem

$$Z(x) = (x + x^{-1})^N$$
$$= \sum_s \frac{N!}{(\frac{1}{2}N + s)!(\frac{1}{2}N - s)!}x^{2s}$$
$$= \sum_s g(N, s)x^{2s}. \tag{1.10}$$

Matching coefficients of x^{2s}, we recover (1.7).

Let us study the behavior of $g(N, s)$ in the limit of large N. To evaluate it, we can use Stirling's approximation[4]

$$N! \approx (2\pi N)^{1/2} N^N e^{-N + \frac{1}{12N} + \cdots}$$
$$\approx e^{(N + \frac{1}{2}) \log N - N + \log \sqrt{2\pi}}, \tag{1.11}$$

You will derive Stirling's approximation on the first problem set. Plugging this into $g(N, s)$, we have

$$\log g(N, s) \approx (N + \tfrac{1}{2}) \log N - N + \log \sqrt{2\pi}$$
$$- (N_\uparrow + \tfrac{1}{2}) \log N_\uparrow + N_\uparrow - \log \sqrt{2\pi}$$
$$- (N_\downarrow + \tfrac{1}{2}) \log N_\downarrow + N_\downarrow - \log \sqrt{2\pi} \tag{1.12}$$

now adding and subtracting $\frac{1}{2} \log N$ and using that $N = N_\uparrow + N_\downarrow$,

$$\log g(N, s) \approx (N_\downarrow + \tfrac{1}{2} + N_\uparrow + \tfrac{1}{2}) \log N - (N_\uparrow + \tfrac{1}{2}) \log N_\uparrow - (N_\downarrow + \tfrac{1}{2}) \log N_\downarrow$$
$$- \tfrac{1}{2} \log N - \log \sqrt{2\pi}$$
$$= -(N_\uparrow + \tfrac{1}{2}) \log \frac{N_\uparrow}{N} - (N_\downarrow + \tfrac{1}{2}) \log \frac{N_\downarrow}{N} - \tfrac{1}{2} \log N - \log \sqrt{2\pi}. \tag{1.13}$$

We have already anticipated that the largest number of microstates will be concentrated at small s. Thus, let us make the approximation that s/N is small. We can then Taylor expand the logarithms,

$$\log \frac{N_{\uparrow,\downarrow}}{N} = \log \frac{1}{2} \left(1 \pm \frac{2s}{N}\right) = -\log 2 \pm \frac{2s}{N} - \frac{2s^2}{N^2} + \dots \tag{1.14}$$

---

[4]In this course, $\log x$ will always denote the correct definition of the logarithm, which is log base e.

Keeping only the terms up to quadratic order in s^2/N^2 (we justify this in a moment), our expression becomes

$$- \left(\frac{N}{2} + s + \frac{1}{2} \right) \left(-\log 2 + \frac{2s}{N} - \frac{2s^2}{N^2} \right)$$
$$- \left(\frac{N}{2} - s + \frac{1}{2} \right) \left(-\log 2 - \frac{2s}{N} - \frac{2s^2}{N^2} \right) - \frac{1}{2} \log N - \log \sqrt{2\pi}. \tag{1.15}$$

Because of symmetry under $s \leftrightarrow -s$, the odd-order terms in s will drop out. We only need to keep the constant and quadratic terms which are

$$\log g(N, s) \approx -\frac{2s^2}{N} \left(1 - \frac{1}{N} \right) + N \log 2 - \frac{1}{2} \log \frac{N\pi}{2}$$
$$\approx -\frac{2s^2}{N} + N \log 2 - \frac{1}{2} \log \frac{N\pi}{2}. \tag{1.16}$$

In the last step, we discarded a term of the form $\frac{2s^2}{N^2}$. To justify this, we should make sure that this term is subleading compared to all the others — in other words, it should be smaller than constants like $-\frac{1}{2} \log \frac{\pi}{2}$ in the large N limit. We confirm this in a moment. The result can be written as

$$g(N, s) \approx g(N, 0) e^{-\frac{2s^2}{N}}, \tag{1.17}$$

where

$$g(N, 0) \approx 2^N \left(\frac{2}{\pi N} \right)^{1/2}. \tag{1.18}$$

1.3   Probability and expectation values

Definition (expectation value). Given a probability distribution $P(s)$, the *expectation value* (or average value) of a function $f(s)$ is defined as

$$\langle f \rangle = \sum_s f(s) P(s). \tag{1.19}$$

Here, the probability function $P(s)$ must be normalized

$$\sum_s P(s) = 1, \qquad \text{or equivalently} \qquad \langle 1 \rangle = 1. \tag{1.20}$$

Note that expectation values are linear $\langle af + b \rangle = a\langle f \rangle + b$, where a and b are constants independent of s.

1.4   Spin system again

For our spin system, we made the physical assumption that the spins were allowed to freely interact with a random environment for a while. Mathematically, we will assume that when we observe the spin system, *every spin configuration (microstate) is equally likely.* Since a spin excess s corresponds to $g(N, s)$ different microstates, the probability of observing a given spin excess is

$$P(N, s) = \frac{g(N, s)}{2^N} \approx \left(\frac{2}{\pi N}\right)^{1/2} e^{-\frac{2s^2}{N}}. \tag{1.21}$$

We divide by the total number of states so that our probability distribution is normalized.

$$\sum_{s=-\frac{N}{2}}^{\frac{N}{2}} P(N, s) = 1. \tag{1.22}$$

For convenience, we will abbreviate $\sum_{s=-\frac{N}{2}}^{\frac{N}{2}} \to \sum_s$.

We can now compute expectation values of various measurements. Firstly, we have

$$\langle s \rangle = \sum_s s P(N, s) = 0. \tag{1.23}$$

We don't have to do any computation to get this result — it follows from the fact that $sP(N, s)$ is odd under $s \leftrightarrow -s$. A more interesting question is

$$\langle s^2 \rangle = \sum_s s^2 P(N, s). \tag{1.24}$$

To compute this, we can approximate it as an integral

$$\langle s^2 \rangle = \left(\frac{2}{\pi N}\right)^{1/2} \int_{-\infty}^{\infty} ds\, s^2 e^{-\frac{2s^2}{N}} \tag{1.25}$$

Here, we used another trick. Even though the maximal values of s are $\pm\frac{N}{2}$, we extended the integral from $-\infty$ to ∞. This is OK because the integrand is extremely tiny at extreme values of s, so we don't make a very big error by changing the integral there. We can now change variables $s \to \sqrt{\frac{N}{2}}x$,

$$\langle s^2 \rangle = \left(\frac{2}{\pi N}\right)^{1/2} \left(\frac{N}{2}\right)^{3/2} \int_{-\infty}^{\infty} dx\, x^2 e^{-x^2} \tag{1.26}$$

To do the integral over x, note that

$$I(a) = \int_{-\infty}^{\infty} e^{-ax^2} = \sqrt{\frac{\pi}{a}}. \tag{1.27}$$

The integral we want is

$$-\frac{dI(a)}{da}\bigg|_{a=1} = \frac{1}{2}\frac{\sqrt{\pi}}{a^{3/2}}\bigg|_{a=1} = \frac{\sqrt{\pi}}{2}. \tag{1.28}$$

In the end, we find

$$\langle s^2 \rangle = \frac{N}{4}. \tag{1.29}$$

The root mean square spin excess is

$$s_* = \langle s^2 \rangle^{1/2} = \frac{\sqrt{N}}{2}, \tag{1.30}$$

This is the width of the gaussian distribution (1.21). In terms of s_*, we can write the probability distribution (1.21) as

$$P(N,s) = \frac{1}{\sqrt{2\pi}s_*} e^{-\frac{1}{2}\left(\frac{s}{s_*}\right)^2} \tag{1.31}$$

The width s_* sets the typical size of s in the probability distribution $P(N,s)$. Having computed it, we can revisit some of the approximations we made in computing $g(N,s)$. One approximation was that $s/N \ll 1$, so that we could expand the logarithm in (1.14). This is clearly true if $s \approx \sqrt{N}$ and N is large. Thus, (1.14) is a good approximation in the region $s \approx s_*$ where the probability is largest. It is a bad approximation where the probability is small (very very small, as we will see in a moment), but who cares about that! Another approximation we made was to discard the term $\frac{2s^2}{N^2}$ but keep the constant $-\frac{1}{2}\log\frac{\pi}{2}$ in (1.16). Again, we see that this is justified near the peak of the probability distribution, since $\frac{s_*^2}{N^2} \sim \frac{1}{N} \ll 1$.

In the large N limit, the width $s_* = \frac{\sqrt{N}}{2}$ is large. However, it is small compared to the full possible range of s. To be more precise, let us define a fractional spin excess

$$t = \frac{s}{N} \in \left[-\tfrac{1}{2}, \tfrac{1}{2}\right]. \tag{1.32}$$

The fractional width is $\langle t^2 \rangle^{1/2} = \frac{1}{2\sqrt{N}}$. In the large-$N$ limit, this is tiny. For example, suppose $N = 10^{22}$, which is typical for a macroscopic system. The fractional width is $t_* = \langle t^2 \rangle^{1/2} \approx 10^{-11}$, so most states have extremely tiny fractional spin excess. Furthermore, if we measure t, we are overwhelmingly likely to find it close to t_*. You

might think: Hey, 10^{-6} is also pretty small. Perhaps we might sometimes measure $t \in [10^{-6}, 2 \times 10^{-6}]$! You would be wrong. The probability of this happening is

$$\int_{10^{-6}}^{2 \times 10^{-6}} ds P(N, s) \approx 10^{-6} e^{-\frac{1}{2}\left(\frac{10^{-6}}{10^{-11}}\right)^2} \approx 10^{-6} e^{-\frac{10^{10}}{2}} \approx 10^{-2 \times 10^9}. \qquad (1.33)$$

If you do observe a spin excess of 10^{-6}, here is a list of more likely scenarios:

- Your measurement equipment is broken.

- Your eyes are not working (i.e. your personal measurement equipment is broken).

- You accidentally drove to the wrong university and measured the wrong spin system.

- You are actually living in the matrix and the machines are doing an experiment to see if you ever studied statistical mechanics.

The point is that the distribution (1.21) becomes *ridiculously* sharply peaked near $s = 0$ when N is large, at least if we measure deviations relative to macroscopic quantities (i.e. quantities that scale linearly with N). For all intents and purposes, even though we have very little knowledge of the spin system and its connection to the environment, we can confidently predict $t = 0$, based on counting.

1.5   Connection to random walks

The fact that deviations away from mean values go like $\sqrt{N}$, and fractional deviations go like $1/\sqrt{N}$, at large N, is a generic phenomenon about sums of random contributions. Another way to derive it is to think of the spins as specifying a random walk in spin excess space. We start at the left of the spin chain with spin excess $S_0 = 0$. Each time we move past a site, we change the spin excess in the direction specified by that site

$$S_n = S_{n-1} + \frac{1}{2} s_n. \qquad (1.34)$$

The total spin excess is the endpoint of the random walk $S_N = s$. Taking an expectation value of the above, we have

$$\langle S_n \rangle = \langle S_{n-1} \rangle + \frac{1}{2} \langle s_n \rangle. \qquad (1.35)$$

However, $\langle s_n \rangle = 0$, so we find the recursion relation

$$\langle S_n \rangle = \langle S_{n-1} \rangle, \qquad (1.36)$$

which has solution $\langle S_n \rangle = \langle S_0 \rangle = 0$. This is just the statement that the average spin excess vanishes.

Now consider the expectation value of the square of (1.34),

$$\langle S_n^2 \rangle = \langle (S_{n-1} + \tfrac{1}{2} s_n)^2 \rangle = \langle S_{n-1}^2 \rangle + \langle S_{n-1} s_n \rangle + \tfrac{1}{4} \langle s_n^2 \rangle. \tag{1.37}$$

The middle term vanishes because the value of s_n does not depend on S_{n-1}. We say that the two quantities are uncorrelated. We have $\langle S_{n-1} s_n \rangle = \langle S_{n-1} \rangle \langle s_n \rangle = 0$. More precisely,

$$\begin{aligned}
\langle S_{n-1} s_n \rangle &= \sum_{s_1, \ldots, s_n} P(s_1, \ldots, s_n) S_{n-1}(s_1, \ldots, s_{n-1}) s_n \\
&= \sum_{s_1, \ldots, s_n} P(s_1) \cdots P(s_n) S_{n-1}(s_1, \ldots, s_{n-1}) s_n \\
&= \left(\sum_{s_1, \ldots, s_{n-1}} P(s_1) \cdots P(s_n) S_{n-1}(s_1, \ldots, s_{n-1}) \right) \left(\sum_{s_n} P(s_n) s_n \right) \\
&= \langle S_{n-1} \rangle \langle s_n \rangle \\
&= 0. \tag{1.38}
\end{aligned}$$

In the second line, we used the fact that the probability of a configuration of spins is the product of the probabilities for each individual spin ($P(s_i) = 1/2$ for $s_i = \pm 1$). This is true in our simple model. However, in more general cases, like in real magnets, we might not be able to factorize $P(s_1, \ldots, s_n)$ and the above computation would not be correct.

However, the last term is interesting because $s_n = \pm 1$, so s_n^2 is always 1. Thus, we find

$$\langle S_n^2 \rangle = \langle S_{n-1}^2 \rangle + \tfrac{1}{4}. \tag{1.39}$$

The solution to this recursion relation is

$$\langle S_n^2 \rangle = \frac{n}{4}, \tag{1.40}$$

so in particular $\langle s^2 \rangle = \frac{N}{4}$, and the RMS value is $\langle s^2 \rangle^{1/2} = \frac{\sqrt{N}}{2}$, which agress with our earlier computation. Thus, $\sqrt{N}$ fluctuations are typical in quantities that get a large number of independent random contributions.

The sharpness of probability distributions in the large-N limit is to a large extent responsible for the incredible success of statistical methods for macroscopic systems. It allows us to turn probabilistic statements into precise predictions.

[End of lecture 3]

2 Entropy and Temperature

2.1 Fundamental assumption of statistical mechanics

Consider a closed (isolated) physical system with some space of states. The fundamental assumption of statistical mechanics is

Claim (Fundamental assumption of statistical mechanics). All accessible states are equally likely.

The term "accessible" requires some explanation:

- **Conservation laws.** Accessible states should be consistent with conservation laws: they should have the correct values of conserved quantities like total energy, momentum, and angular momentum.

  In this course, we will encounter other types of conserved quantities, some of which are fundamental and some of which are approximate. An example of an approximate conserved quantity is particle number. In relativistic quantum field theory (QFT), all particles can be created or destroyed. However, particle creation/destruction might not occur at the energy scales relevant for a given problem. For example, chemical processes do not involve enough energy to create or destroy electrons, protons, or neutrons. For this reason, electron number, proton number, and neutron number are good approximately conserved quantities in chemistry. By contrast, photons can be created or destroyed in chemical processes, so photon number is not a good approximately conserved quantity.

  Let us also comment about what it means to have the "correct" value of a conserved quantity, e.g. energy. If a system is in an exact energy eigenstate $|E\rangle$, then only that eigenstate is accessible. However, we typically don't find macroscopic systems in exact energy eigenstates. In practice, we usually only know the total energy of a system with a resolution ΔU that depends on our measuring device. In this case, states with energies in the window $[U - \Delta U, U + \Delta U]$ are accessible.

- **Time scale.** Which states are accessible also depends on the time scale being considered. Glass is a famous example of a substance that exhibits dynamics on very long time scales — its relaxation time can be millions of years. However, for practical purposes on human time scales, we shouldn't consider states where glass flows into other configurations.

2.1.1 Intuition for the fundamental assumption

The intuition behind the fundamental assumption is that interactions within the system will tend to randomize the state. However, those interactions must still

respect conservation laws, so the system will wander through the space of states in a random way while still respecting, e.g. conservation of energy.

Two properties of typical systems support this picture

- **Chaos.** Chaos is the exponential sensitivity of a system to changes in initial conditions. It is sometimes called the "butterfly effect:" the flapping of a butterfly's wings is a tiny perturbation, but its effects can amplify exponentially over time. Because of chaos, even if we have precise knowledge of the initial state of a system, that knowledge rapidly degrades when we try to predict its future evolution.

  For an illustration of chaos, I like this simulation of a double pendulum with 1 million closeby initial conditions: https://twitter.com/rickyreusser/status/1231990500023926784. Although all 1 million instances start out very close to one another, their trajectories quickly diverge from each other. After a short time, the pendulums are spread out over all possible configurations (consistent with conservation of energy).

  Here's another fun one, showing chaos in a simulation of four bodies interacting via gravity: https://twitter.com/simon_tardivel/status/1215728659010670594.

- **Egodicity.** An ergodic system is one for which the average of an observable over the set of accessible states is the same as the average over time for a particular state.

  The idea behind ergodicity in classical systems is that a trajectory through phase space $\mathbf{q}(t), \mathbf{p}(t)$ will eventually pass close by any other accessible point in phase space. For example, a gas molecule bouncing around a room will eventually go everywhere and have every momentum.

  Most systems are not ergodic in the strict mathematical sense. There can be regions of phase space that are never reached. (Example: A circular billiards table on which a billiard ball traces out a path in the shape of an n-pointed regular star centered at the center of the table. In this case, there is an n-gon region in the center of the table that the ball never reaches.) Or it can take an exponentially long time to visit every point in phase space. However, we are usually not interested in whether a system passes close by every microstate. For studying the dynamics of macrostates, ergodicity is often an excellent approximation.

All this intuition does not amount to a proof, and in fact the fundamental assumption is only an approximation. There are some properties of a state that are very difficult to measure, and therefore not important for our purposes, but take a very long time to randomize. An example is the "computational complexity" of a state.

2.1.2   Mathematical statement of fundamental assumption

Mathematically, the fundamental assumption says that, in an isolated system, the probability of observing a state s is $P(s) = 1/g$, where g (which is independent of s) is the number of accessible states. Expectation values of observables $X(s)$ are given by

$$\langle X \rangle = \sum_s X(s) \left(\frac{1}{g} \right) = \frac{1}{g} \sum_s X(s), \tag{2.1}$$

where the sum $\sum_s$ runs over accessible states.

2.2   Thermal contact and most probable configuration

Consider two systems S_1, S_2 and let us bring them into contact so that energy can be exchanged between them. This is called "thermal contact." Together, the two systems form a larger closed system $S = S_1 + S_2$ with constant energy $U = U_1 + U_2$.

Let the number of states of the system S_1 with energy U_1 be $g_1(U_1)$ and let the number of states of S_2 with energy U_2 be $g_2(U_2) = g_2(U - U_1)$. The number of states of S such that S_1 has energy U_1 is

$$g_1(U_1)g_2(U - U_1) \tag{2.2}$$

The total number of states of the combined system is

$$g(U) = \sum_{U_1} g_1(U_1)g_2(U - U_1) \tag{2.3}$$

According to the fundamental assumption, after some amount of time, every state of the combined system is equally likely. In particular, the probability distribution for observing that subsystem S_1 has energy U_1, given that the combined system has energy U, is

$$P(U, U_1) = \frac{g_1(U_1)g_2(U - U_1)}{g(U)}. \tag{2.4}$$

Note that due to (2.3), the distribution is correctly normalized, i.e. $\sum_{U_1} P(U, U_1) = 1$.

2.2.1   Example: spin system

Let us look at a typical example of such a probability distribution. Consider the spin system from the last lecture, and let us turn on a magnetic field B so that the energy U is proportional to the spin excess $U = 2mBs$.[5] Using (1.17) from the

---

[5]In most of this lecture, we use s to label microstates. Here, we are briefly using it to label the spin-excess which specifies a macrostate. We hope this does not cause confusion.

previous lecture, we can write the number of microstates as a function of U,

$$g(N, U) = g(N, 0)e^{-\frac{U^2}{2N(mB)^2}} \tag{2.5}$$

Let us bring two such systems into thermal contact, and suppose that the total energy is U. The number of states where S_1 has energy U_1 is

$$g(N_1, U_1)g(N_2, U_2) = g(N_1, 0)g(N_2, 0) \exp\left(-\frac{U_1^2}{2N_1(mB)^2} - \frac{U_2^2}{2N_2(mB)^2}\right)$$
$$= g(N_1, 0)g(N_2, 0) \exp\left(-\frac{1}{2(mB)^2}\left(\frac{U_1^2}{N_1} + \frac{(U - U_1)^2}{N_2}\right)\right). \tag{2.6}$$

The quantity in the exponent has a maximum when

$$0 = \frac{U_1}{N_1} - \frac{U - U_1}{N_2}$$
$$U_1 = \frac{N_1}{N_1 + N_2}U \equiv \widehat{U}_1 \tag{2.7}$$

The largest number of microstates corresponds to the case where the total energy U is distributed between the two systems in amounts proportional to their size: system S_1 has $\frac{N_1}{N_1+N_2}$ of the energy and system S_2 has $\frac{N_2}{N_1+N_2}$ of the energy.

Let us write $U_1 = \widehat{U}_1 + \delta U_1$ and substitute to find

$$g(N_1, 0)g(N_2, 0) \exp\left(-\frac{1}{2(mB)^2}\left(\frac{U^2}{N_1 + N_2} - \delta U_1^2\left(\frac{1}{N_1} + \frac{1}{N_2}\right)\right)\right) \tag{2.8}$$

We can compute $g(U)$ by integrating over δU_1 (exercise). Dividing, we find

$$P(U, U_1) = \sqrt{\frac{2}{\pi}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)} e^{-\frac{\delta U_1^2}{2(mB)^2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}. \tag{2.9}$$

Suppose that N_1 and N_2 are both large. We'll will write $N_1, N_2 = O(N)$, where N is some large number. Let us also allow take $U = O(N)$ in the large-N limit, so that the total energy scales with the size of the system. The probabiity distribution $P(U, U_1)$ has width $\delta U_1 \sim O(\sqrt{N})$.[6] Meanwhile, $\widehat{U}_1$ scales like $O(N)$. Thus, the

---

[6]We could obtain this by computing $\langle \delta U_1^2 \rangle^{1/2}$, but a quicker way is to look at the exponent in (2.9). Typical values of δU_1 will be such that the quantity in the exponent is not too negative (otherwise it will be very suppressed). Let us suppose it is $O(1)$,

$$\frac{\delta U_1^2}{2(mB)^2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right) = O(1) \tag{2.10}$$

Ignoring the factors $2(mB)^2$, which are not important for determining the large-N scaling of δU_1, we find

$$\frac{\delta U_1^2}{N} = O(1), \tag{2.11}$$

or $\delta U_1 = O(\sqrt{N})$ as claimed.

fractional deviation $\delta U_1/U_1 \sim O(1/\sqrt{N})$ is small. We should think of (2.9) as being extremely sharply-peaked as a function of U_1.

2.2.2 Back to the general case

Generically, for macroscopic systems S_1 and S_2, the product $g_1(U_1)g_2(U - U_1)$ will be sharply peaked. Typically, $g_1(U_1)$ is a rapidly-changing function of U_1. Similarly $g_2(U_2)$ is a rapidly-changing function of U_2. The product $g_1(U_1)g_2(U - U_1)$ involves a competition between the two factors. A local maximum $\widehat{U}_1$ can occur when there is a delicate balance between the rate of increase of $g_1(U_1)$ and the rate of decrease of $g_2(U - U_1)$. If we move U_1 slightly away from this point, the balance is destroyed and the product decreases rapidly away from its maximum.

Because $P(U, U_1)$ is sharply peaked, we are overwhelmingly likely to observe a value of U_1 where $P(U, U_1)$ is maximized — the *most probable configuration*.

In summary, before the systems are in thermal contact, we will observe a value of U_1 where $g_1(U_1)$ alone is maximized. After thermal contact, the two systems will evolve to the most probable configuration $U_1 = \widehat{U}_1$, where $g_1(U_1)g_2(U - U_1)$ is maximized. The final configuration is called *thermal equilibrium*.

2.3 Entropy and temperature

The condition to have a maximum of $P(U, U_1)$, or equivalently $g_1(U_1)g_2(U - U_1)$, is

$$
\begin{aligned}
0 &= \frac{\partial}{\partial U_1} \left(g_1(U_1)g_2(U - U_1) \right) \\
&= \frac{\partial g_1}{\partial U_1} g_2 - g_1 \frac{\partial g_2}{\partial U_2}
\end{aligned}
\tag{2.12}
$$

We can write this as

$$
\frac{\partial \log g_1}{\partial U_1} = \frac{\partial \log g_2}{\partial U_2}.
\tag{2.13}
$$

Definition (entropy). The *entropy* of a system is the log of the number of accessible states

$$
\sigma = \log g.
\tag{2.14}
$$

The condition for thermal equilibrium (i.e. maximizing $P(U_1)$) becomes

$$
\frac{\partial \sigma_1}{\partial U_1} = \frac{\partial \sigma_2}{\partial U_2}.
\tag{2.15}
$$

Definition (temperature). The *temperature* of a system is defined by

$$
\frac{1}{\tau} = \frac{\partial \sigma}{\partial U}.
\tag{2.16}
$$

We use $1/\tau$ on the left-hand side so that τ has units of energy. (Note that the entropy σ is unitless.) With this definition, the condition for thermal equilibrium becomes

$$\tau_1 = \tau_2. \tag{2.17}$$

For historical reasons, the temperature that we usually measure with a thermometer is defined by

$$T = \frac{\tau}{k_B}, \tag{2.18}$$

where $k_B = 1.381 \times 10^{-23} J/^\circ K$ is Boltzmann's constant. Similarly, the entropy is usually defined[7]

$$S = k_B \log g, \tag{2.19}$$

so that we have

$$\frac{1}{T} = \frac{\partial S}{\partial U}. \tag{2.20}$$

These conventions date from a time when the relationship between temperature, entropy, and the number of states was not understood. From a modern point of view, k_B is a silly constant. Entropy should be unit-less, and temperature should have units of energy. In this course, we will almost always work with τ and σ instead of T and S.

2.4　Comments on entropy

The logarithm of the number of states, the entropy $\sigma = \log g$, has some nice properties compared to the number of states itself. Firstly, entropy is *extensive*. If we have two non-interacting (or weakly-interacting) systems, then the entropy of the combined system is the sum of the entropies of the individual systems,

$$\sigma = \log(g_1 g_2) = \sigma_1 + \sigma_2. \tag{2.21}$$

If we take N copies of a system, the entropy of the combined system is N times the entropy of an individual system. In particular, for typical physical systems, the entropy scales like the number of degrees of freedom of the system. For example, in a gas at fixed temperature and pressure, the entropy is proportional to the volume.

Another nice property of the logarithm is that σ is insensitive to many details that we have left unspecified, unlike g. For example, in our definition of accessible states, we said that there might be some energy resolution, so that accessible states have energy in a window $U \pm \Delta U$. If we double the size of the window $\Delta U' = 2\Delta U$,

---

[7]This expression is engraved on Boltzmann's gravestone in the form $S = k \log W$.

then the number of accessible states will roughly double, $g \to g' = 2g$.[8] Thus, the actual value of g depends sensitively on our precise energy resolution. However, the entropy will change by $\sigma \to \sigma' = \sigma + \log 2$. If $\sigma \sim 10^{22}$ is large, the difference between σ and σ' is insignificant. Thus, we can reliably discuss the entropy of a system without having to give a detailed definition of the space of accessible states.

All of this supports the mantra that when you have a ridiculously large number, it's better to think about it's logarithm. You'll notice that in our examples with spin systems and gaussian integrals, we were manipulating logarithms.

2.5 Entropy increases

Suppose system S_1 has energy U_{10} before thermal contact. The initial number of states of the combined system is

$$(g_1 g_2)_0 = g_1(U_{10}) g_2(U - U_{10}). \tag{2.22}$$

After thermal contact, the systems will evolve to a configuration with $U_1 = \widehat{U}_1$ where the number of states is maximized

$$(g_1 g_2)_{\max} = g_1(\widehat{U}_1) g_2(U - \widehat{U}_1) \tag{2.23}$$

Clearly we have

$$(g_1 g_2)_{\max} \geq (g_1 g_2)_0 \tag{2.24}$$

with equality if $\widehat{U}_1 = U_{10}$ (i.e. the system S_1 was already at the equilibrium temperature of the combined system).

This is the statement that entropy increases. The entropy of a closed system will remain constant or increase when a constraint internal to the system is removed. Some examples of operations that increase entropy are adding particles, adding energy, increasing volume, decomposing molecules.

It is extremely unlikely to observe a system spontaneously lower its entropy. For example, suppose we have a gas of particles in a room. A low entropy configuration would be one where all the particles are contained in the top left corner. (This is low entropy because the number of such states is much smaller than the number of states where the particles roam freely.) However, the chances of the room evolving into this state are ridiculously tiny. Suppose the particles are non-interacting. The probability at any given time of observing a particle in $1/2$ of the room is $1/2$. The probability of observing all the particles simultaneously in $1/2$ the room is $1/2^N \sim 1/2^{10^{23}}$ — i.e. it will never happen.

---

[8]This is true assuming the density of states is roughly constant as a function of energy over scales of size ΔU.

Consider two systems brought into thermal contact that have not yet reached equilibrium. They will exchange some energy $U_1 \to U_1 + \delta U$ and $U_2 \to U_2 - \delta U$ in such a way that entropy increases. Thus, we have

$$\frac{\partial \sigma_1}{\partial U_1} \delta U - \frac{\partial \sigma_2}{\partial U_2} \delta U \geq 0$$

$$\left(\frac{1}{\tau_1} - \frac{1}{\tau_2} \right) \delta U \geq 0. \tag{2.25}$$

In other words, δU will be positive and S_1 will get hotter if $\tau_2 > \tau_1$, and vice versa if $\tau_1 > \tau_2$. Thus, energy flows in the way one would expect: from hotter to cooler.

[End of lecture 4]

2.6   The arrow of time

An interesting property of the laws of nature is that they are time-reversal invariant. For example, if we have a solution $x(t)$ to

$$F(x) = m \frac{d^2 x}{dt^2}, \tag{2.26}$$

then $x(-t)$ is another solution.

Actually, the Standard Model of particle physics is not invariant under time-reversal alone. Instead, it is invariant under a symmetry called CRT that simultaneously reverses time (T), reflects a spatial direction (R), and exchanges particles with antiparticles (C for "charge conjugation"). It was discovered in 1964 that if you don't do CR as well as T, then certain processes involving nuclear decay are not invariant. Anyway, applying the symmetry CRT, a solution of the equations of the Standard Model with time evolving one way can be transformed into another solution with time evolving the other way.

In this case, one might ask: how do we know which direction of time is the right one? The fact that entropy increases gives our answer. If we took a solution with time reversed, then all microscopic physical laws would be obeyed, but we'd see weird things happen like eggs unbreaking, shockwaves miraculously pushing bombs back together and causing them to bounce off the ground and attach to airplane wings, etc.. It is possible to unambiguously determine the arrow of time as long as a system does not start out in a maximum entropy configuration. One of the great mysteries of fundamental physics is why the initial entropy of the universe is so low.

2.7   Heat capacity

Definition (heat capacity). The *heat capacity C* is defined by

$$C = \frac{\partial U}{\partial \tau}. \tag{2.27}$$

It is the change in energy over change in temperature.

Heat capacity is *extensive*: it scales with the size of a system. This is hopefully intuitive — the bigger a system is, the more energy it can store. However, let us understand it from the formula (2.27). Energy U is extensive: the total energy of system $\mathcal{S}_1 + \mathcal{S}_2$ is the sum $U = U_1 + U_2$. Meanwhile, temperature is intensive: if we combine two systems with temperature τ, the resulting system still has temperature τ. Consider a system $\mathcal{S}$ with heat capacity $C_{\mathcal{S}}(\tau) = \frac{\partial U_{\mathcal{S}}}{\partial \tau}$. Now consider N copies of $\mathcal{S}$, which we denote $N\mathcal{S}$. The heat capacity is

$$C_{N\mathcal{S}}(\tau) = \frac{\partial U_{N\mathcal{S}}}{\partial \tau} = N \frac{\partial U_{\mathcal{S}}}{\partial \tau} = N C_{\mathcal{S}}(\tau). \tag{2.28}$$

Even more roughly, we can say: in a system of size N, U scales like $O(N)$ and τ scales like $O(1)$. Thus $\frac{\partial U}{\partial \tau}$ scales like $O(N/1) = O(N)$.

A nice property of heat capacity is that it is directly measurable. To measure it, you add some energy and see how the temperature changes. From the heat capacity, one can compute entropy by integrating. Note that

$$\frac{\partial \sigma}{\partial \tau} = \frac{\partial \sigma}{\partial U} \frac{\partial U}{\partial \tau} = \frac{C}{\tau}. \tag{2.29}$$

Thus, the difference in entropy of a system at temperatures τ_1 and τ_2 is

$$\sigma_1 - \sigma_2 = \int_{\tau_1}^{\tau_2} \frac{C(\tau)}{\tau} d\tau. \tag{2.30}$$

For example, setting $\tau_1 = 0$, we have

$$\sigma(\tau_0) = \int_0^{\tau_0} \frac{C(\tau)}{\tau} d\tau + \sigma(0). \tag{2.31}$$

The entropy at zero temperature is the log of the ground state multiplicity $\sigma(0) = \log g(0)$. For most systems, $g(0)$ is small, so $\sigma(0)$ is effectively zero compared to macroscopic entropies at nonzero temperature (which are $O(N)$, where N is the number of particles). In such systems, we can drop the last term above. For glassy systems, $\sigma(0)$ can still have macroscopic size — such systems have frozen-in disorder, so that there are many different (near) ground states where the atoms are distributed in many different ways.

Equation (2.31) is an awesome formula! You might think that the entropy, i.e. the log of the number of accessible states, is a pretty abstract concept. Furthermore, it's not easy to imagine computing it directly: how can we count the microstates of a system of 10^{23} particles? But using (2.31) you can actually figure out what it is!

Another property we should mention is that heat capacity is usually positive: adding energy usually increases temperature. However, there are some exotic systems that possess negative heat capacity. The most famous example is a black hole:

adding energy increases the size of the black hole, which lowers the temperature (and lowers the rate of Hawking radiation). Conversely, removing energy decreases the size of the black hole and raises the temperature. This leads to a runaway effect: the black hole will radiate more and more, its temperature rising and rising, until it decays away to nothing.

2.8   The laws of thermodynamics

The above observations about temperature and entropy have been codified into "laws" of thermodynamics[9]

0th law: If two systems are in thermal equilibrium with a third system, then they are in thermal equilibrium with each other. This is the definition of equilibrium $\tau_1 = \tau_2$, together with the transitive property of equality.

1st law: Heat is a form of energy. This is the principle of conservation of energy.

2nd law: Entropy increases.

3rd law: Entropy approaches a constant value as the temperature approaches zero. This is simply the ground-state entropy $\sigma(0) = \log g(0)$.

From the point of view of statistical mechanics, most of these aren't really fundamental laws — they're consequences of statistics and properties of typical systems. Honestly, I never learned which law had which number, just like I never bothered to learn the difference between class 1, class 2, and class 3 levers. I'm supposedly a professional physicist, and it's almost never been a problem, although people talk a lot about the 2nd law when discussing black hole physics (for reasons that we'll see later), so I kind of know which one that is. To avoid hypocrisy, I will try not to test you on which law is which.

---

[9]Here's a song about the laws of thermodynamics: https://www.youtube.com/watch?v=VnbiVw_1FNs.

3 The Boltzmann distribution

3.1 Microcanonical ensemble

Definition (ensemble). A probability distribution $P(s)$ on the space of states is called an *ensemble*.

An ensemble is needed to compute expectation values of observables

$$\langle X \rangle = \sum_s P(s) X(s). \tag{3.1}$$

In the last lecture, we studied the ensemble for an isolated system with a known fixed energy U. According to the fundamental assumption of statistical mechanics, it is given by

$$P_{\text{microcanonical}}(s) = \begin{cases} \frac{1}{g(U)} & \text{if } s \text{ has energy } U, \\ 0 & \text{otherwise,} \end{cases} \tag{3.2}$$

where $g(U)$ is the number of states with energy U — i.e., the number of accessible states.

Definition (microcanonical ensemble). For historical reasons, the probability distribution (3.2) is called the *microcanonical ensemble*.

3.2 Canonical ensemble

However, many systems we encounter can exchange energy with their environment, and thus should not be thought of as having fixed energy. A basic question in thermal physics is how a system $\mathcal{S}$ behaves when exposed to an environment at some temperature. To model the environment, let us introduce:

Definition (reservoir). A *reservoir* $\mathcal{R}$ (or "heat bath") is an extremely large system. In particular, a reservoir has an very large heat capacity $C = \frac{\partial U}{\partial \tau} \ggg 1$.

Because of its large heat capacity, a reservoir can absorb essentially any amount of energy without changing its temperature. We would like to know how $\mathcal{S}$ behaves when brought into thermal contact with a reservoir $\mathcal{R}$ at temperature τ. We are interested in this question even when $\mathcal{S}$ is not macroscopic. Some examples of a system $\mathcal{S}$ and reservoir $\mathcal{R}$ are

- $\mathcal{S}$ is an atom in a crystal and $\mathcal{R}$ is the crystal itself

- $\mathcal{S}$ is a few air molecules in the room and $\mathcal{R}$ is all the air in the room

- $\mathcal{S}$ is a dark matter particle and $\mathcal{R}$ is the dark matter halo in the galaxy

- More generally, $\mathcal{S}$ is a small subpart of a larger system $\mathcal{R}$

What is the probability that the system $\mathcal{S}$ will be in a state s with energy E_s? Let the total energy of the combined system $\mathcal{S} + \mathcal{R}$ be E_{tot}. By the fundamental assumption of statistical mechanics,

$$
\begin{aligned}
P(s) &\propto g_{\mathcal{S}}(\mathcal{S} \text{ is in state } s) \times g_{\mathcal{R}}(E_{\text{tot}} - E_s) \\
&= g_{\mathcal{R}}(E_{\text{tot}} - E_s).
\end{aligned}
\tag{3.3}
$$

The first factor $g_{\mathcal{S}}(\mathcal{S} \text{ is in state } s)$ is simply 1 because we are demanding that the microstate of $\mathcal{S}$ be exactly s. In the second factor $g_{\mathcal{R}}(E_{\text{tot}} - E_s)$, we have used conservation of energy.

The constant of proportionality in (3.3) is the inverse of the number of accessible states of $\mathcal{S} + \mathcal{R}$. So that we don't have to discuss this quantity, consider the ratio of probabilities associated to two microstates s, s',

$$
\frac{P(s)}{P(s')} = \frac{g_{\mathcal{R}}(E_{\text{tot}} - E_s)}{g_{\mathcal{R}}(E_{\text{tot}} - E_{s'})}.
\tag{3.4}
$$

Let us write the number of states of $\mathcal{R}$ in terms of its entropy,

$$
= \frac{e^{\sigma_{\mathcal{R}}(E_{\text{tot}} - E_s)}}{e^{\sigma_{\mathcal{R}}(E_{\text{tot}} - E_{s'})}}.
\tag{3.5}
$$

We should imagine that $E_s \ll E_{\text{tot}}$, so that it is a good idea to expand the entropy in a Taylor series

$$
\begin{aligned}
\sigma_{\mathcal{R}}(E_{\text{tot}} - E_s) &= \sigma_{\mathcal{R}}(E_{\text{tot}}) - E_s \frac{\partial \sigma_{\mathcal{R}}}{\partial U} + \frac{1}{2} E_s^2 \frac{\partial^2 \sigma_{\mathcal{R}}}{\partial U^2} + \dots \\
&= \sigma_{\mathcal{R}}(E_{\text{tot}}) - \frac{E_s}{\tau} + \frac{1}{2} E_s^2 \frac{\partial^2 \sigma_{\mathcal{R}}}{\partial U^2} + \dots.
\end{aligned}
\tag{3.6}
$$

In the second line, we have assumed that $\mathcal{R}$ is at temperature τ and we have used the definition of temperature.

We claim that in the limit of a very large reservoir, the second and higher derivative terms above can be dropped. Let us first consider the second derivative term. It is given by

$$
\frac{\partial^2 \sigma_{\mathcal{R}}}{\partial U^2} = \frac{\partial}{\partial U} \left(\frac{1}{\tau} \right) = -\frac{1}{\tau^2} \frac{\partial \tau}{\partial U} = -\frac{1}{\tau^2} \frac{1}{C},
\tag{3.7}
$$

where C is the heat capacity of the reservoir. Remember that a reservoir has extremely large heat capacity, so this term is negligible.

Let us understand this result in a cruder way that will help with the higher-order terms in (3.6). The largeness of C comes from the fact that energy U is extensive and temperature τ is intensive. Thus, $C = \frac{\partial U}{\partial \tau}$ scales like $O(N)$. By the same large-N scaling analysis, we expect $\frac{\partial^k \sigma_{\mathcal{R}}}{\partial U^k}$ to scale like $O(N/N^k) = O(N^{1-k})$:

the N in the numerator comes from the fact that $\sigma_\mathcal{R}$ is extensive, and the N^k in the denominator comes from the fact that U is extensive and we have k factors of U. Thus, in the limit of a large reservoir, the higher-order terms in (3.6) should be dropped as well. On your homework, you will make this result rigorous in the case where the environment can be modeled as N non-interacting copies of a single system.

Thus, let us drop the higher-derivative terms in (3.6). The quantity $\sigma_\mathcal{R}(E_\text{tot})$ cancels between the numerator and denominator, and we are left with

$$\frac{P(s)}{P(s')} = \frac{e^{-E_s/\tau}}{e^{-E_{s'}/\tau}} \qquad \Longrightarrow \qquad P(s) = Ae^{-E_s/\tau}, \tag{3.8}$$

where A is independent of s. Using the fact that probabilities must sum to 1, we can solve for A,

$$\sum_s Ae^{-E_s/\tau} = 1 \qquad \Longrightarrow \qquad A = \frac{1}{Z(\tau)}, \tag{3.9}$$

where

$$Z(\tau) \equiv \sum_s e^{-E_s/\tau}. \tag{3.10}$$

Overall, we find

$$P(s) = \frac{e^{-E_s/\tau}}{Z(\tau)}, \tag{3.11}$$

Note that almost all reference to the reservoir has dropped out, with the only remaining dependence on $\mathcal{R}$ through its temperature τ. Equation (3.11) is the most important equation in statistical mechanics. It's so important that the numerator, denominator, and ratio on the right-hand side all have their own names:

Definition (Boltzmann factor). The quantity $e^{-E_s/\tau}$ in the numerator of (3.11) is called a *Boltzmann factor*.

Definition (partition function). The quantity $Z(\tau)$ in (3.10) is called the *partition function*.

Definition (canonical ensemble). For historical reasons, the probability distribution (3.11) is called the *canonical ensemble*.

3.3   Why Taylor expand the entropy?

If you look over the above derivation, all we did was Taylor expand the entropy and keep the first two terms. One question is: why didn't we Taylor expand the number

of states in (3.4) and keep the first two terms of *that* expansion instead? Instead of an exponential, that procedure would have given a linear function of E_s, which is the wrong answer. The reason is that Taylor expanding a logarithm gives a more efficient way to approximate quickly-growing functions.

As an example, consider the model function $g(U) = U^N$, where $N = 10^{22}$ is enormous. If we Taylor expand $g(U)$ itself, then we have

$$g(E_{\text{tot}} - E_s) = E_{\text{tot}}^N \left(1 - N \frac{E_s}{E_{\text{tot}}} + \frac{N(N-1)}{2} \frac{E_s^2}{E_{\text{tot}}^2} + \dots \right) \tag{3.12}$$

The problem with this series compared to (3.6) is that even though each subsequent term comes with an extra $1/E_{\text{tot}}$ (which goes like $1/N$ and is small), it also comes with an extra factor of N (which is large). In the large-N limit, we cannot drop all the higher-derivative terms. Instead, the best we can do is drop subleading quantities in N in each term. This leads to

$$g(E_{\text{tot}} - E_s) \sim E_{\text{tot}}^N \left(1 - N \frac{E_s}{E_{\text{tot}}} + \frac{N^2}{2} \frac{E_s^2}{E_{\text{tot}}^2} + \dots \right)$$

$$= E_{\text{tot}}^N e^{-E_s \frac{N}{E_{\text{tot}}}} = g(E_{\text{tot}}) e^{-E_s/\tau} \tag{3.13}$$

so we get an exponential again. Here, we used the fact that $\tau = g(E_{\text{tot}})/g'(E_{\text{tot}}) = E_{\text{tot}}/N$ in this example. Taking the logarithm first gives a more efficient way to sum all these terms.

3.4 The partition function

The partition function $Z(\tau)$ is the most important quantity in statistical mechanics. Almost everything you might ever want to know about a system is hidden in its partition function (suitably generalized), and the -1'st law of thermodynamics is that the first thing you should try to do when you encounter a new system is compute its partition function. An example of a quantity hidden in the partition function is the expectation value of the energy

$$U = \langle E \rangle = \sum_s E_s P(s)$$

$$= \frac{1}{Z(\tau)} \sum_s E_s e^{-E_s/\tau}$$

$$= \frac{1}{Z(\tau)} \tau^2 \frac{\partial}{\partial \tau} Z(\tau)$$

$$= \tau^2 \frac{\partial}{\partial \tau} \log Z(\tau) \tag{3.14}$$

It is conventional to define $\beta = 1/\tau$ and write

$$\langle E \rangle = -\frac{\partial}{\partial \beta} \log Z(\beta). \tag{3.15}$$

We can also compute the expectation value of the square of the energy

$$\langle E^2 \rangle = \frac{1}{Z(\beta)} \frac{\partial^2}{\partial \beta^2} Z(\beta) = \frac{\partial^2}{\partial \beta^2} \log Z(\beta) + \left(\frac{1}{Z(\beta)} \frac{\partial Z(\beta)}{\partial \beta} \right)^2. \tag{3.16}$$

In particular, the standard deviation of the energy is

$$\Delta E^2 = \langle (E - \langle E \rangle)^2 \rangle = \langle E^2 \rangle - \langle E \rangle^2 = \frac{\partial^2}{\partial \beta^2} \log Z(\beta). \tag{3.17}$$

As an example, consider a two-state system with ground-state energy 0 and excited state energy ε. The partition function is

$$Z(\tau) = 1 + e^{-\varepsilon/\tau}. \tag{3.18}$$

The average energy is

$$\begin{aligned} U = \langle E \rangle &= \frac{\tau^2}{Z(\tau)} \frac{\partial Z(\tau)}{\partial \tau} \\ &= \frac{\varepsilon e^{-\varepsilon/\tau}}{1 + e^{-\varepsilon/\tau}} \end{aligned} \tag{3.19}$$

When τ/ε is small, this is close to zero (in fact, it dies incredibly quickly as τ gets small). When τ/ε is large, this asymptotes to $\varepsilon/2$. The heat capacity is

$$C = \frac{\partial U}{\partial \tau} = \frac{\varepsilon^2}{\tau^2} \frac{e^{-\varepsilon/\tau}}{(1 + e^{-\varepsilon/\tau})^2}. \tag{3.20}$$

This has the surprising feature that it spikes near $\tau/\varepsilon \sim 0.5$, a phenomenon known as the *Schottky anomaly*. It is "anomalous" because in most materials, the heat capacity is monotonically increasing with temperature. The reason the Schottky anomaly appears here is that our system has a maximum energy ε. As $\tau \to \infty$, the largest $U = \langle E \rangle$ can ever get is $\varepsilon/2$ (coming from an equal probability of being in the excited and ground state). Typical macroscopic systems have no maximum energy (or a maximum energy that is way bigger than typical temperatures). In that case, as we increase τ, the probability becomes nontrivial for more and more high energy states and $\langle E \rangle$ can keep increasing.

Consider now a pair of two-state systems, each with zero ground state energy, and with excited state energies $\varepsilon_1, \varepsilon_2$. The partition function is

$$\begin{aligned} Z_{1+2}(\tau) &= 1 + e^{-\varepsilon_1/\tau} + e^{-\varepsilon_2/\tau} + e^{-(\varepsilon_1+\varepsilon_2)/\tau} \\ &= (1 + e^{-\varepsilon_1/\tau})(1 + e^{-\varepsilon_2/\tau}) \\ &= Z_1(\tau) Z_2(\tau). \end{aligned} \tag{3.21}$$
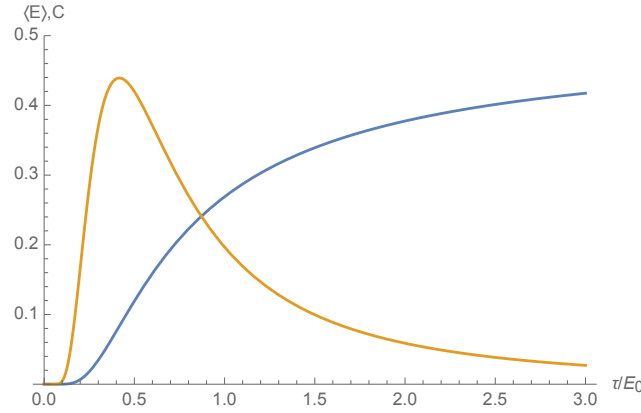
Figure 1: $\langle E \rangle$ for a two-state system as a function of τ/E_0.

This is a special case of a more general result: If $\mathcal{S}_1, \mathcal{S}_2$ are non-interacting systems, then the partition function of the combined system is the product of the partition functions

$$Z_{\mathcal{S}_1 + \mathcal{S}_2}(\tau) = Z_{\mathcal{S}_1}(\tau) Z_{\mathcal{S}_2}(\tau). \tag{3.22}$$

[End of lecture 5]

3.5   Example: spin system

As an example, the partition function of our spin system from the last lecture can be computed as follows. An individual spin has energies $E = \pm mB$. The partition function for a single spin is thus

$$Z_{\text{single-spin}}(\tau) = e^{-mB/\tau} + e^{mB/\tau} = 2\cosh(mB/\tau) \tag{3.23}$$

Thus, the partition function for N non-interacting spins is

$$Z_{\text{spin-system}}(\tau) = (2\cosh(mB/\tau))^N = (2\cosh(\beta mB))^N. \tag{3.24}$$

The expected value of the energy is

$$\langle E \rangle = -\frac{\partial}{\partial \beta} \log Z = -\frac{\partial}{\partial \beta} \left(N \log 2 + N \cosh(\beta mB) \right)$$

$$= -NmB \tanh(\beta mB) = -NmB \tanh(mB/\tau). \tag{3.25}$$

At low temperatures, we have $\langle E \rangle = -NmB$ as all the spins align with the magnetic field. At high temperatures, we have $\langle E \rangle = 0$ as thermal fluctuations dominate over the effects of the applied field.

The standard deviation in energy is

$$\Delta E^2 = -\frac{\partial}{\partial \beta}\left(-NmB\tanh(\beta mB)\right) = N(mB)^2\frac{1}{\cosh^2(\beta mB)}. \tag{3.26}$$

Note that the fluctuations are of size $\Delta E^2 \sim N$, whereas the energy itself is also $O(N)$. Thus, the fractional fluctuations are small, of order $\Delta E/\langle E\rangle \sim 1/\sqrt{N}$.

There is a more general result relating the size of fluctuations in energy to the heat capacity. Recall that

$$C = \frac{\partial\langle E\rangle}{\partial\tau} = -\left(\frac{\partial\tau}{\partial\beta}\right)^{-1}\frac{\partial\langle E\rangle}{\partial\beta} = \beta^2\frac{\partial^2}{\partial\beta^2}\log Z = \frac{\Delta E^2}{\tau^2} \tag{3.27}$$

We know that heat capacity scales linearly with the size of the system N, so this shows that ΔE^2 also scales like N.

3.6   Thermal density matrix

3.6.1   Density matrices

In quantum mechanics, a state $|\psi\rangle$ is a vector in a Hilbert space $\mathcal{H}$. The expectation value of an observable $\widehat{X}$ in the state $|\psi\rangle$ is computed by[10]

$$\langle\widehat{X}\rangle = \langle\psi|\widehat{X}|\psi\rangle. \tag{3.28}$$

More generally, we may not know the precise state of a system, but only a probability distribution on the space of states. For example, we may have a lab partner that randomly decides whether to prepare an up state or down state with probabilities p and $1-p$. In this case, the expectation value of an observable is

$$\langle\widehat{X}\rangle = p\langle\uparrow|\widehat{X}|\uparrow\rangle + (1-p)\langle\downarrow|\widehat{X}|\downarrow\rangle. \tag{3.29}$$

This is not the same thing as having a single quantum state

$$|?\rangle = \sqrt{p}|\uparrow\rangle + \sqrt{1-p}|\downarrow\rangle \qquad \text{(wrong)}. \tag{3.30}$$

The expectation value of $\widehat{X}$ in such a state would involve cross-terms $\langle\uparrow|\widehat{X}|\downarrow\rangle$ that are not present in (3.29). Instead, the quantities p and $(1-p)$ represent classical probabilities, on top of the inherent quantum mechanical probabilities associated with a measurement.

A probability distribution on the space of states can be represented by a density matrix. In the case described above, the density matrix is

$$\widehat{\rho} = p|\uparrow\rangle\langle\uparrow| + (1-p)|\downarrow\rangle\langle\downarrow|. \tag{3.31}$$

---

[10]In these few sections, I will try to be careful about putting hats on quantum mechanical operators, to distinguish them from C-numbers. However, I will be less careful about this later.

It has matrix elements

$$\begin{pmatrix} \langle\uparrow|\widehat{\rho}|\uparrow\rangle & \langle\uparrow|\widehat{\rho}|\downarrow\rangle \\ \langle\downarrow|\widehat{\rho}|\uparrow\rangle & \langle\downarrow|\widehat{\rho}|\downarrow\rangle \end{pmatrix} = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}. \tag{3.32}$$

The expectation value of an observable can be written

$$\begin{aligned} \langle\widehat{X}\rangle_{\widehat{\rho}} &= \mathrm{Tr}(\widehat{\rho}\widehat{X}) \\ &= p\mathrm{Tr}(|\uparrow\rangle\langle\uparrow|\widehat{X}) + (1-p)\mathrm{Tr}(|\downarrow\rangle\langle\downarrow|\widehat{X}) \\ &= p\langle\uparrow|\widehat{X}|\uparrow\rangle + (1-p)\langle\downarrow|\widehat{X}|\downarrow\rangle. \end{aligned} \tag{3.33}$$

In general, given a probability distribution $P(s)$ on a space of orthonormal states, the corresponding density matrix is given by

$$\widehat{\rho} = \sum_s P(s)|s\rangle\langle s|, \tag{3.34}$$

and the expectation value of an observable is

$$\langle\widehat{X}\rangle_{\widehat{\rho}} = \mathrm{Tr}(\widehat{X}\widehat{\rho}) = \sum_s P(s)\mathrm{Tr}(\widehat{X}|s\rangle\langle s|) = \sum_s P(s)\langle s|\widehat{X}|s\rangle. \tag{3.35}$$

This is a more general version of (3.1). Specifically, (3.1) applies when the states $|s\rangle$ are eigenstates of $\widehat{X}$ with eigenvalues $X(s)$.

3.6.2   Thermal density matrix

In the canonical ensemble, the probability distribution $P(s)$ is given by (3.11). This leads to

Definition (thermal density matrix). The *thermal density matrix* at temperature τ is given by

$$\widehat{\rho}_\tau = \frac{\sum_s e^{-E_s/\tau}|s\rangle\langle s|}{Z(\tau)}. \tag{3.36}$$

The quantity in the numerator can be written in a basis-independent way as

$$\sum_s e^{-E_s/\tau}|s\rangle\langle s| = e^{-H/\tau}, \tag{3.37}$$

where H is the Hamiltonian. Equation (3.37) is simply the definition of the exponential of a diagonalizable operator: we just exponentiate all the eigenvalues. See the discussion around (A.10) for details.

The partition function itself can be written

$$Z(\tau) = \sum_s e^{-E_s/\tau} = \mathrm{Tr}(e^{-H/\tau}), \tag{3.38}$$

which is the statement that the trace of an operator is the sum of its eigenvalues. Overall, the density matrix associated to the canonical ensemble is

$$\widehat{\rho}_\tau = \frac{e^{-H/\tau}}{\mathrm{Tr}(e^{-H/\tau})}. \tag{3.39}$$

This incredibly beautiful formula has abundant applications, from statistical physics to condensed matter physics to particle physics to string theory. We will meet some of them later in this course.

4   Entropy in general ensembles

4.1   Gibbs/Shannon entropy

We defined entropy in the microcanonical ensemble as the log of the number of accessible states. Suppose that instead we have a general ensemble, i.e. a general probability distribution $P(s)$ on the space of states. How should we define the entropy? After answering this question in general, we will be particularly interested in the case of the canonical ensemble.

    We will look for a generalized definition of entropy satisfying two conditions:

- If there are g equally likely accessible states, then we should have $\sigma = \log g$, as before.

- σ should be extensive.

It turns out that these conditions single out a unique answer!

    The argument is as follows. Suppose that we don't have just one copy of our system $\mathcal{S}$, but instead a large number N of identical copies. We will compute the entropy of N copies, and then divide by N.

    The systems are in states

$$s_{i_1}, s_{i_2}, \ldots, s_{i_N}. \tag{4.1}$$

If N is large enough, the number of systems in state s_i will be $P(s_i)N$. That is, the string above will contain $P(s_1)N$ s_1's, $P(s_2)N$ s_2's, etc.. By taking a large number of systems, we have turned probabilities into eventualities. A string where s_1, s_2, etc. do not appear in the correct proportions should be considered inaccessible in the large-N limit. To determine the entropy, we count the number of accessible configurations g_{total} of the total system and apply the familiar definition $\sigma_{\text{total}} = \log g_{\text{total}}$. The entropy of a single system is then $\sigma = \sigma_{\text{total}}/N$.

    The problem of counting the number of strings (4.1) with $P(s_1)N$ s_1's, $P(s_2)N$ s_2's, etc. was on the homework. The answer is

$$e^{N\sigma}, \tag{4.2}$$

where σ is $\frac{\log 2}{\log e}$ times the Shannon entropy of the probability distribution,[11]

$$\sigma = -\sum_s P(s) \log P(s). \tag{4.3}$$

This is how we will define the entropy of a general ensemble.

---

[11]In information theory, it is conventional to define entropy using $\log_2$. In physics, it is more appropriate to use the natural logarithm.

Definition (Gibbs/Shannon entropy). The *Gibbs/Shannon entropy* of a general ensemble is $P(s)$ is given by (4.3).

Formula (4.3) was written down by Gibbs, and was rediscovered many years later by Shannon in the context of information theory.

We see that entropy is a function of a probability distribution. When we discuss "the" entropy of a macroscopic system, we are using the fact that many different probability distributions all give similar answers. For example, if I know the average energy within a window $U \pm \Delta U$, this gives me a probability distribution on the space of states. For macroscopic systems, the Gibbs/Shannon entropy of that distribution does not depend strongly on ΔU.

Definition (von-Neumann entropy). In the quantum mechanical context, the entropy can be written

$$\sigma = -\mathrm{Tr}(\widehat{\rho} \log \widehat{\rho}), \tag{4.4}$$

where $\widehat{\rho}$ is the density matrix. This is called the *von-Neumann entropy*

4.2   Sanity checks

Let us check that the Gibbs/Shannon entropy gives the right answer for the microcanonical ensemble. Recall that

$$P_{\mathrm{microcanonical}}(s) = \begin{cases} \frac{1}{g(U)} & \text{if } s \text{ has energy } U \\ 0 & \text{otherwise} \end{cases}. \tag{4.5}$$

The microcanonical Gibbs/Shannon entropy is then

$$-\sum_{s \text{ with energy } U} \frac{1}{g(U)} \log \frac{1}{g(U)} - \sum_{\text{other } s} 0 = \log g(U). \tag{4.6}$$

Here, it is important that we define $0 \times \log 0 = \lim_{p \to 0} p \log p = 0$.

Another sanity check is that the entropy is additive, even for a combination of two systems with different probability distributions. This is not something we used in our derivation in section 4, but it will happen naturally anyways. Consider systems $\mathcal{S}_1$ and $\mathcal{S}_2$ with probability distributions $P_1(s_1)$ and $P_2(s_2)$ on their respective space of states. States of the combined system $\mathcal{S}_1 + \mathcal{S}_2$ are labeled by pairs (s_1, s_2), with

probabilities $P(s_1, s_2) = P_1(s_1)P_2(s_2)$. The combined Gibbs/Shannon entropy is

$$
\begin{aligned}
\sigma_{12} &= -\sum_{s_1, s_2} P(s_1, s_2) \log P(s_1, s_2) \\
&= -\sum_{s_1, s_2} P_1(s_1)P_2(s_2)(\log P_1(s_1) + \log P_2(s_2)) \\
&= -\sum_{s_1} P_1(s_1) \log P_1(s_1) \left(\sum_{s_2} P_2(s_2) \right) - \sum_{s_2} P_2(s_2) \log P_2(s_2) \left(\sum_{s_1} P_1(s_1) \right) \\
&= \sigma_1 + \sigma_2.
\end{aligned}
\tag{4.7}
$$

4.3 Entropy of the canonical ensemble

Let us finally compute the Gibbs/Shannon entropy of the canonical ensemble:

$$
\begin{aligned}
\sigma &= -\sum_s P(s) \log P(s) \\
&= -\sum_s \frac{e^{-E_s/\tau}}{Z} \log \frac{e^{-E_s/\tau}}{Z} \\
&= \sum_s \frac{e^{-E_s/\tau}}{Z} \left(\frac{E_s}{\tau} + \log Z \right) \\
&= \frac{\langle E \rangle}{\tau} + \log Z \\
&= \frac{U}{\tau} + \log Z.
\end{aligned}
\tag{4.8}
$$

This gives a definition of entropy appropriate for a system in thermal contact with a reservoir.

4.4 Comments on Shannon Entropy

In information theory, Shannon entropy is a measure of the amount of information in a message. Consider a data file given by a collection of 0's and 1's. How much can we compress the file? We would like to perform lossless compression, so that the original file can be completely recovered from its compressed version. It is reasonable to define the amount of information in the file as the minimal number of bits it can be compressed to.

If the data is completely random and unstructured, then there is no way to compress it. This is because the possible number of random data files with n bits is 2^n. A hypothetical compression algorithm would map each of these files to a string of $m < n$ bits. However, the number of such strings is only $2^m < 2^n$, so this cannot be done in a 1-1 manner.

However, if the data has some structure, compression is possible. An extreme example is if we know the file is just a sequence of exactly n 0's. This can be encoded using zero bits. A slightly less extreme example is if we know the file is a constant sequence of length n, containing either all 0's or all 1's. Such files can be encoded using 1 bit: 0 if the file is all 0's or 1 if the file is all 1's.

[End of lecture 6]

In general, when data has structure, the presence of that structure decreases the number of possible files, making compression possible. For example, an XML file always starts with `<xml>` and ends with `</xml>`. In total, this is 88 bits of information that we already know is there. Consequently, the number of valid XML files of length n is less than $2^{n-88} < 2^n$.

A more interesting example of structured data is human language. A naive way to store a document of English text uses ASCII encoding (https://en.wikipedia.org/wiki/ASCII), where each character is stored using 8 bits. For example, `a` is ASCII code 01100001, `b` is ASCII code 01100010, etc.. This is not very efficient for English text because the English language contains only 26 letters — perhaps together with punctuation and spaces, we can say that there are $2^5 = 32$ symbols. Thus, we should be able to compress English using no more than 5 bits per character.

However, not every random sequence of characters is a good English sentence. For example, the letter `e` is more likely to appear than `z`. We can exploit this extra structure to make a better compression scheme.

Let us understand how this works in a toy language with only two characters `a` and `b`. First, suppose that `a`,`b` occur equally often. From the point of view of letter frequency, our text file is like a random set of 0's and 1's, and we cannot compress it. (This statement ignores possible correlations between letters, which we discuss in a moment.) Now suppose instead that `a` occurs 90% of the time and `b` occurs 10% of the time. To compress a sequence, it's a good idea to use fewer bits for `a` than for `b`. One way to accomplish this is the following scheme:

Sequence	Code
aa	0
a	10
b	11

(4.9)

Because `a` occurs very often, we use 0 to stand for a pair of `a`'s. For example, the sequence `aaaaaabaaa`, which is 10 characters, becomes 00011010, which is 8 bits. The probabilities of all possible two-letter sequences are

Sequence	aa	ab	ba	bb
Probability	81%	9%	9%	1%
Code	0	1011	1110	1111

(4.10)

Thus, the average number of bits per character in scheme (4.9) is

$$\frac{1}{2}\left(1 \times 0.81 + 4 \times 0.09 + 4 \times 0.09 + 4 \times 0.01\right) = 0.785. \tag{4.11}$$

Given a sequence of symbols a_i occurring with probabilities P_i, Shannon's answer for the number of bits per symbol needed to compress the sequence is the Shannon entropy

$$H = \text{minimal \# of bits} = -\sum_i P_i \log_2 P_i. \tag{4.12}$$

This is Shannon's **source coding theorem**. In our two-character example, H is given by

$$H = -(0.9 \log_2 0.9 + 0.1 \log_2 0.1) = 0.468. \tag{4.13}$$

So our the compression scheme (4.9) that uses 0.785 bits per character is not optimal. Can you improve it?

The probabilities of occurrences of various letters in English text are

	a	b	c	d	e	f	g	h	i	j	k	l	m
%	8.17	1.49	2.20	4.25	12.7	2.23	2.02	6.09	6.97	0.15	1.29	4.03	2.41

	n	o	p	q	r	s	t	u	v	w	x	y	z
%	6.75	7.51	1.93	0.10	5.99	6.33	9.36	2.76	0.98	2.56	0.15	1.99	0.08

$$\tag{4.14}$$

For example, a randomly-selected letter is 12.7% likely to be e and 0.08% likely to be z. According to the source coding theorem, we should be able to encode English using

$$H = -\left(0.0817 \log_2 0.0817 + 0.0149 \log_2 0.0149 + \cdots + 0.0008 \log_2 0.0008\right) = 4.17 \tag{4.15}$$

bits per letter. This is slightly better than the 5 bits we estimated above.

However, letters aren't just randomly distributed according to some probabilities. They form words and sentences. This additional structure makes language more compressible. By thinking about sequences of words rather than letters, Shannon estimated that the information entropy of English text has $H = 2.62$ bits/character.

The Hutter prize is a 50,000€ competition to encode a 100MB snapshot Wikipedia using the smallest possible number of bits. The current record is 16MB, which is a compression ratio of 100/16=6.25, so each character is represented by $\frac{8\,\text{bits}}{6.25} = 1.28\,\text{bits}$. This is better than Shannon's estimate of 2.62 bits. Thus, correlations between words also play a role in reducing the entropy of English text.[12]

---

[12]Most people speak approximately 14 characters per second. Using the Hutter prize result $H = 1.28$ bits/letter to estimate the information content, this is 18 bits/second. In conversation, you convey roughly $18\text{bit/s} \times \text{byte/8bit} \approx 8\,\text{kB/hour}$ of information.

4.4.1   Optimal compression and entropy

An optimal compression algorithm for files with structured data of length N is as follows: make a list of all valid files of length N. ("Valid" means consistent with the known structure of the files.) Map each file to its index in the list, encoded as a binary number. The amount of information in a file is the length of that binary number. As you showed on your homework, if we perform this exercise for a file of length N whose content is drawn from an alphabet a_i with probabilities P_i, we obtain NS bits, where S is the Shannon entropy of the probability distribution P_i.

Another way to think about this algorithm is as follows: our knowledge of the structure of the data is a "macrostate." It is a coarse measurement that tells us something, but not everything, about the contents of the file. The set of valid files is the set of "accessible states." The number of bits we need to encode a structured file is the log of the number of accessible states.

The more we know about the structure of the file, the smaller the number of accessible states and the smaller the information entropy. Conversely, the less we know about the structure of the file, the larger the number of accessible states and the larger the information entropy. Thus, information entropy is a measure of ignorance about the contents of some binary data. Likewise, thermodynamic entropy is a measure of ignorance about the microstate of a system. Hopefully this analogy explains why tools of information theory are useful in statistical mechanics.

Here's a simulation that demonstrates a connection between information entropy and thermodynamic entropy in a beautiful way:
https://twitter.com/i/status/1182208172100116480. On the left is a simulation of two types particles that are initially un-mixed, and then mix over time. On the right is the size of the resulting image, compressed using the png compression algorithm. As the particles mix, our ignorance about the precise microstate of the system increases. This is because there are more accessible states when the particles are mixed than when they are not mixed. (We discuss entropy of mixing in section **??**.) At the same time, the image gets more complicated and requires a larger png. This is roughly because there are more possible images of mixed particles than unmixed particles. If the compression algorithm were lossless and optimal (which is not really true in this case), then the size of the resulting file would equal the thermodynamic entropy of the mixture.

5   The thermodynamic limit and free energy

5.1   The thermodynamic limit

Definition (thermodynamic limit). The limit $N \to \infty$, where N represents the size of a system, is called the *thermodynamic limit.*

You will show on the problem set that for a system with N degrees of freedom in the canonical ensemble, $\Delta E / \langle E \rangle = O(1/\sqrt{N})$. Consequently, in the thermodynamic limit, all uncertainty about the energy of a system goes away. The canonical ensemble becomes effectively the same as the microcanonical ensemble (with some relationship between $\langle E \rangle$ and τ).

To see this more precisely, consider the partition function

$$Z_{\mathrm{c}}(\tau) = \sum_s e^{-E_s/\tau}. \tag{5.1}$$

Here, we include the subscript 'c' to emphasize that Z_{c} is the partition function in the canonical ensemble. Let us organize the states into energy levels with degeneracy $g(U) = e^{\sigma_{\mathrm{mc}}(U)}$, where $\sigma_{\mathrm{mc}}(U)$ is the entropy of the *microcanonical* ensemble. The sum (5.1) becomes

$$Z_{\mathrm{c}}(\tau) = \sum_U e^{\sigma_{\mathrm{mc}}(U) - U/\tau}. \tag{5.2}$$

The summand involves a competition between two effects: (1) the increasing number of states $e^{\sigma_{\mathrm{mc}}(U)}$ as energy increases, and (2) the exponentially dying Boltzmann factors.[13] The crossover point where these effects balance out is the solution U_* of the equation

$$0 = \left. \frac{\partial}{\partial U}(\sigma_{\mathrm{mc}}(U) - U/\tau) \right|_{U=U_*}. \tag{5.3}$$

In the thermodynamic limit, the summand in (5.2) becomes sharply peaked at the "most probable configuration" $U = U_*$, so we can approximate it by the value of the summand at that point:

$$Z_{\mathrm{c}}(\tau) \approx e^{\sigma_{\mathrm{mc}}(U_*) - U_*/\tau} = Z_{\mathrm{mc}}(\tau), \qquad \text{(thermodynamic limit)} \tag{5.4}$$

The right-hand side of (5.4) is the partition function of the microcanonical ensemble (i.e. of a system with $e^{\sigma_{\mathrm{mc}}(U_*)}$ states, all with energy U_*). Applying the formula $\langle E \rangle = \tau^2 \frac{\partial \log Z}{\partial \tau}$, we see that $\langle E \rangle \approx U_*$ in the thermodynamic limit.

---

[13]Student question: how do we know that the Boltzmann factor eventually balances the growth of $g(U)$ — i.e. that $g(U)$ doesn't grow superexponentially? Answer: A system where $g(U)$ grows superexponentially forever would have an infinite partition function. If this happens, it means that some approximation we made in describing the system is breaking down.

Because all thermodynamic observables are encoded in the partition function, equality of the canonical and microcanonical partition functions in the thermodynamic limit implies equality of other quantities too. As an example, let us apply formula (4.8) for the entropy:

$$
\begin{aligned}
\sigma_{\mathrm{c}} &= \frac{\langle E \rangle}{\tau} + \log Z \\
&\approx \frac{U_*}{\tau} + \log \left(e^{\sigma_{\mathrm{mc}}(U_*) - U_*/\tau} \right) \qquad \text{(thermodynamic limit)} \\
&= \sigma_{\mathrm{mc}}(U_*).
\end{aligned}
\tag{5.5}
$$

5.2   Free energy

Definition (free energy). The *free energy* is defined by

$$
F = U - \tau\sigma = -\tau \log Z.
\tag{5.6}
$$

In section (5.1), we showed that in the thermodynamic limit, a system will evolve toward the "most probable configuration," where $\sigma(U) - U/\tau = -\frac{F}{\tau}$ is maximized (equation 5.3). Equivalently, in the thermodynamic limit, a system at fixed temperature will seek to minimize F.

The fact that F is minimized in the most probable configuration is not a new principle. Recall that the entropy of a reservoir is given by

$$
\sigma_{\mathcal{R}}(U_{\mathrm{tot}} - U) = \sigma_{\mathcal{R}}(U_{\mathrm{tot}}) - U/\tau.
\tag{5.7}
$$

Thus, we have

$$
\begin{aligned}
F &= U - \tau\sigma \\
&= \tau\sigma_{\mathcal{R}}(U_{\mathrm{tot}}) - \tau\sigma_{\mathcal{R}}(U_{\mathrm{tot}} - U) - \tau\sigma \\
&= -\tau\sigma_{\mathcal{S}+\mathcal{R}}(U_{\mathrm{tot}}, U) + \tau\sigma_{\mathcal{R}}(U_{\mathrm{tot}}) \\
&= -\tau\sigma_{\mathcal{S}+\mathcal{R}}(U_{\mathrm{tot}}, U) + \text{independent of } U
\end{aligned}
\tag{5.8}
$$

F is just $-\tau$ times the entropy $\sigma_{\mathcal{S}+\mathcal{R}}(U_{\mathrm{tot}}, U) = \sigma + \sigma_{\mathcal{R}}(U_{\mathrm{tot}} - U)$ of the combined system and reservoir, plus a term that doesn't depend on U. So the statement that F is minimized in the thermodynamic limit is just the usual statement that the entropy of the combined system+reservoir is maximized.[14]

Free energy plays a similar role for systems in contact with a heat bath as energy for damped systems. (Damping really means coupled to a zero-temperature heat bath.) The term U causes the system to seek low energies, and the term $-\tau\sigma$ causes

---

[14]Student question: what do we mean that the energy evolves in time toward the most probable configuration? Answer: imagine the system is very weakly coupled to a reservoir, so that it is in thermal equilibrium with itself, with energy as a function of time.

the system to seek high entropy. These effects balance each other at the minimum of the free energy.

A nice example is a polymer that can straighten out or curl up. There are many more configurations where the polymer is curled up than where it is straight. Thus, the polymer wants to curl up to increase its entropy. We can make it energetically favorable for the polymer straighten out by pulling on it with an external force F_{ext}. The free energy is then

$$F = -\ell F_{\text{ext}} - \tau \sigma(\ell), \tag{5.9}$$

where $\sigma(\ell)$ is the entropy at length ℓ. The minimum is acheived when

$$F_{\text{ext}} = -\tau \frac{\partial \sigma}{\partial \ell}. \tag{5.10}$$

Thus, the entropy term $-\tau\sigma(\ell)$ leads to a noticeable force that we can measure. It is called an "entropic force." If we increase the temperature, the entropic force gets stronger. For example, if you heat a rubber band, it contracts.

5.3   Pressure

Consider a system in an energy eigenstate s with energy E_s, and let us suppose that E_s is a function of the volume V of the system. The volume can be changed $V \to V - \Delta V$ by applying an external force. Let us suppose that the volume changes sufficiently slowly that the system remains in an energy eigenstate.[15] The change in energy is

$$E_s(V - \Delta V) = E_s(V) - \frac{dE_s}{dV}\Delta V + \dots . \tag{5.11}$$

Suppose that the system is compressed by applying pressure p to a surface of area A. The force is pA. If the force acts over a distance Δx, then the work imparted to the system is

$$\Delta W = (pA)\Delta x = p\Delta V. \tag{5.12}$$

The work ΔW is equal to the change in energy of the system. Comparing (5.11) and (5.12), we see that the pressure in state s is

$$p_s = -\frac{dE_s}{dV} \tag{5.13}$$

---

[15]This is sometimes called the "adiabatic approximation." Consider a time-dependent Hamiltonian $H(t)$ such that $H(t_0) = H_0$ and $H(t_1) = H_1$. Suppose the system starts out in an eigenstate of H_0, $H_0|E_0\rangle = E_0|E_0\rangle$. What state will it be in at time t_1? If the perturbation happens very slowly, then the eigenstate $|E_0\rangle$ of H_0 will evolve to an eigenstate $|E_1\rangle$ of H_1. However, if the perturbation happens quickly, we typically end in a nontrivial superposition of eigenstates of H_1: $\sum_i c_i|E_i\rangle$. In other words, a slow perturbation takes eigenstates to eigenstates, while a fast perturbation induces transitions between eigenstates. The meaning of slow vs. fast depends on the energy gaps E_{gap} between eigenstates and the amplitude. "Slow" means $\Delta t \gg \hbar/E_{\text{gap}}$.

Averaging over all states in an ensemble, we find

$$\langle p \rangle = \sum_s P(s) \left(-\frac{dE_s}{dV} \right)$$

$$= -\frac{d}{dV} \sum_s P(s) E_s$$

$$= -\left(\frac{\partial \langle E \rangle}{\partial V} \right)_\sigma = -\left(\frac{\partial U}{\partial V} \right)_\sigma. \tag{5.14}$$

The subscript σ means that we take the derivative, keeping the entropy fixed. The entropy is held fixed because the probability distribution on the space of states is fixed during the compression. Specifically, the compression is slow enough that if the system starts out in energy eigenstate state s, it will remain in state s. Thus, the occupation probabilities don't change and the entropy is unchanged.

5.4   Thermodynamic identities

In this section, we derive an alternative expression for pressure in terms of the free energy of a system. This is useful because the free energy is given by $F = -\tau \log Z$, and the partition function Z is the nicest thing to calculate. In our derivation, we will encounter the notion of constrained partial derivatives. Let us explain them in more detail.

5.4.1   Constrained derivatives

A common situation in thermodynamics is that we have three macroscopic measurements x, y, z that depend on each other. Once we specify any two, the third is determined. We can express this with a constraint equation $f(x, y, z) = 0$.

Definition (constrained derivative). A constrained partial derivative $\left(\frac{\partial x}{\partial y} \right)_z$ is the change in x over the change in y, with z held fixed. This notion generalizes to multiple variables: in general if we have a constraint $f(x, y, z, w, \dots) = 0$, then $\left(\frac{\partial x}{\partial y} \right)_{z,w,\dots}$ means that we hold $z, w, \dots$ fixed when we take the derivative.

If x, y, z are related by a constraint $f(x, y, z) = 0$, then their constrained derivatives are related. We can see this by looking at the differential form of the constraint

$$0 = f(x + dx, y + dy, z + dz) - f(x, y, z)$$

$$= \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz. \tag{5.15}$$

To compute the constrained derivative $\left(\frac{\partial x}{\partial y} \right)_z$, we can set $dz = 0$ and solve for $\frac{dx}{dy}$:

$$\left(\frac{\partial x}{\partial y} \right)_z = -\frac{\partial f/\partial y}{\partial f/\partial x}. \tag{5.16}$$

Similarly, we find

$$\left(\frac{\partial y}{\partial z}\right)_x = -\frac{\partial f/\partial z}{\partial f/\partial y},$$
$$\left(\frac{\partial z}{\partial x}\right)_y = -\frac{\partial f/\partial x}{\partial f/\partial z}. \tag{5.17}$$

An immediate consequence of (5.16) and (5.17) is the handy formula

$$\left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_y = -1, \tag{5.18}$$

which we will use in a moment. The nice thing about (5.18) is that it no longer makes reference to the constraint $f(x, y, z)$. This is useful because we don't have to bother writing down $f(x, y, z)$ or choosing conventions for it.

[End of lecture 7]

5.4.2   Pressure, entropy, and free energy

For a homogeneous system (like a gas) with variable volume at finite temperature, the entropy σ, energy U, and volume V are examples of three macroscopic variables that satisfy a constraint. For example, we can compute the entropy $\sigma(U, V)$ as a function of energy and volume. If we consider changes dU, dV, then we have

$$d\sigma = \left(\frac{\partial \sigma}{\partial U}\right)_V dU + \left(\frac{\partial \sigma}{\partial V}\right)_U dV$$
$$= \frac{1}{\tau} dU + \left(\frac{\partial \sigma}{\partial V}\right)_U dV. \tag{5.19}$$

In the second line, we used the definition of temperature. The second term involves a constrained derivative that we don't a-priori know, but it can be related to pressure using the identity (5.18) for U, V, σ:

$$\left(\frac{\partial \sigma}{\partial V}\right)_U \left(\frac{\partial V}{\partial U}\right)_\sigma \left(\frac{\partial U}{\partial \sigma}\right)_V = -1$$
$$\left(\frac{\partial \sigma}{\partial V}\right)_U \left(-\frac{1}{p}\right) \tau = -1$$
$$\left(\frac{\partial \sigma}{\partial V}\right)_U = \frac{p}{\tau} \tag{5.20}$$

In the second line, we used the definition of pressure (5.14) and the definition of temperature.

Substituting this result into (5.19), we find

$$do = \frac{1}{\tau}dU + \frac{p}{\tau}dV$$
$$\tau d\sigma = dU + pdV. \tag{5.21}$$

Let us now compute the change in free energy corresponding to dU, dV. We have

$$dF = dU - \tau d\sigma - \sigma d\tau$$
$$dF = -\sigma d\tau - pdV, \tag{5.22}$$

from which we derive a relationship between pressure and free energy, together with another useful expression for entropy

$$p = -\left(\frac{\partial F}{\partial V}\right)_\tau$$
$$\sigma = -\left(\frac{\partial F}{\partial \tau}\right)_V. \tag{5.23}$$

We will put these to work momentarily.

5.5　Ideal gas

5.5.1　One atom in a box

Our eventual goal is to compute the partition function of a gas of N particles. It is helpful to first compute the partition function Z_1 of a single particle in a box. We can then build the whole partition function from Z_1.

Definition (orbital). An *orbital* is a solution of the single-particle Schrodinger equation.

The term "orbital" is motivated by atomic physics: in that case, orbital refers to a solution of the single-particle Schrodinger equation for an electron in the potential well created by the nucleus. However, we will use the term more generally.

Let us calculate the partition function Z_1 of one particle of mass m in a cubical box of volume $V = L^3$. The orbitals of the free particle Schrodinger equation

$$\left(-\frac{\hbar^2}{2m}\right)\nabla^2\psi = E\psi \tag{5.24}$$

are

$$\psi(x, y, z) = A\sin\frac{n_x\pi x}{L}\sin\frac{n_y\pi y}{L}\sin\frac{n_z\pi z}{L}, \tag{5.25}$$

where n_x, n_y, n_z are any positive integers. The energy eigenvalues are

$$E_n = \frac{\hbar^2}{2m}\left(\frac{\pi}{L}\right)^2(n_x^2 + n_y^2 + n_z^2). \tag{5.26}$$

We neglect the spin and other structure of the particle.

The partition function is the sum

$$Z_1 = \sum_{n_x,n_y,n_z=1}^{\infty} e^{-\frac{\hbar^2\pi^2}{2mL^2\tau}(n_x^2+n_y^2+n_z^2)}$$

$$= \sum_{n_x,n_y,n_z=1}^{\infty} e^{-\alpha^2(n_x^2+n_y^2+n_z^2)}, \tag{5.27}$$

where $\alpha^2 = \frac{\hbar^2\pi^2}{2mL^2\tau}$, and we exclude $n_x = 0$, $n_y = 0$, and $n_z = 0$ terms as the wavefunction vanishes with those values.

If α^2 is sufficiently small, then the quantity inside the sum is slowly varying and we can approximate the sum as an integral,

$$Z_1 \approx \int_0^\infty dn_x \int_0^\infty dn_y \int_0^\infty dn_z e^{-\alpha^2(n_x^2+n_y^2+n_z^2)}$$

$$= \left(\int_0^\infty dn e^{-\alpha^2 n^2} \right)^3$$

$$= \left(\frac{1}{\alpha} \int_0^\infty dx e^{-x^2} \right)^3$$

$$= \left(\frac{\pi^{1/2}}{2\alpha} \right)^3$$

$$= \left(\frac{m\tau}{2\pi\hbar^2} \right)^{3/2} V = n_Q V = \frac{n_Q}{n}, \tag{5.28}$$

where $n = 1/V$ and $n_Q = \left(\frac{m\tau}{2\pi\hbar^2} \right)^{3/2}$ is called the quantum concentration. Note we have set the integration lower bound to 0 with negligible error, as the integrand is slowly varying—the difference between starting from 0 and 1 is small. It is sometimes more common to define the thermal wavelength

$$n_Q = \frac{1}{\lambda^3}, \qquad \lambda = \sqrt{\frac{2\pi\hbar^2}{m\tau}}. \tag{5.29}$$

The thermal wavelength is approximately the de Broglie wavelength of an atom with energy $\langle E \rangle \sim \tau$. Specifically $\lambda \sim \frac{\hbar}{\langle p^2 \rangle^{1/2}} = \frac{\hbar}{(2m\langle E \rangle)^{1/2}} \sim \frac{\hbar}{(m\tau)^{1/2}}$. Note that λ is an intrinsically quantum-mechanical quantity, due to the presence of the $\hbar$. In some formulas, like the ideal gas law derived below, it will drop out. However, in other cases λ will remain, signaling a quantum-mechanical effect that cannot be explained classically. Thus, n_Q is the concentration associated with one atom in a cube of size equal to the thermal average de Broglie wavelength.

The smallness of α is equivalent to the statement

$$L \gg \lambda, \tag{5.30}$$

i.e. the box is big compared to the thermal wavelength. When this condition holds, we say that the gas is in the "classical regime." An ideal gas is a gas of noninteracting particles in the classical regime.

The average energy of the atom in a box is

$$U = \tau^2 \frac{\partial}{\partial \tau} \log Z_1 = \tau^2 \frac{\partial}{\partial \tau} \left(\frac{3}{2} \log \tau \right) = \frac{3}{2} \tau. \tag{5.31}$$

In D spatial dimensions, the answer would be

$$U = \frac{D}{2} \tau. \tag{5.32}$$

This is a special case of a more general result called "equipartition of energy" that we derive in section 8.

5.5.2 N particles and the Gibbs paradox

If we had N distinguishable particles in a box, then we could treat each particle like a separate non-interacting system, and the partition function would be

$$
\begin{aligned}
Z_N^{\text{distinguishable}} &= \sum_{s_1, \ldots, s_N} e^{-(E_{s_1} + \cdots + E_{s_N})/\tau} \\
&= \left(\sum_s e^{-E_s/\tau} \right)^N \\
&= Z_1^N
\end{aligned}
\tag{5.33}
$$

This will turn out to be an incorrect expression for the ideal gas, for reasons that we will explain. However, let us continue with it for now. The energy of the ideal gas follows from the N-particle partition function,

$$
\begin{aligned}
U &= \tau^2 \frac{\partial}{\partial \tau} \log Z_N \\
&= \frac{3}{2} N \tau.
\end{aligned}
\tag{5.34}
$$

The free energy is

$$
\begin{aligned}
F^{\text{distinguishable}} &= -\tau \log Z_N^{\text{distinguishable}} \\
&= -\tau N \log(n_Q V).
\end{aligned}
\tag{5.35}
$$

From here, we can compute the pressure

$$
\begin{aligned}
p &= -\left(\frac{\partial F}{\partial V} \right)_\tau = \frac{N\tau}{V} \\
pV &= N\tau
\end{aligned}
\tag{5.36}
$$

This is called the ideal gas law. Note that the thermal wavelength λ has dropped out of this formula. In more traditional units,

$$pV = N k_B T = N_{\text{moles}} RT. \tag{5.37}$$

We can also go on the compute the entropy

$$
\begin{aligned}
\sigma^{\text{distinguishable}} &= -\left(\frac{\partial F}{\partial \tau}\right)_V \\
&= -\frac{\partial}{\partial \tau}\left(-\tau N \log\left(\left(\frac{m\tau}{2\pi\hbar^2}\right)^{3/2} V\right)\right) \\
&= N \log(n_Q V) + \frac{3}{2}N.
\end{aligned} \tag{5.38}
$$

Here, we used that $\tau \log n_Q \propto \frac{3}{2}\tau \log \tau$. This formula has some surprising properties. Firstly, note that n_Q has *not* dropped out. The notion of counting the states requires quantum mechanics.

Secondly, consider a gas with some fixed concentration of particles $n = N/V$, so that the volume can be written $V = N/n$. The above formula becomes

$$\sigma^{\text{distinguishable}} = N \log\left(\frac{n_Q}{n}\right) + \frac{3}{2}N + N \log N. \tag{5.39}$$

This is mysterious because it is not extensive — i.e. it is not proportional to N. In particular, if this formula were true, then if we took two containers of identical gasses at the same temperature and put them together, then the entropy would increase. We could then bring the containers apart and return to our starting configuration. More directly, this formula just doesn't agree with experiment. This problematic answer is called the "Gibbs paradox."

The fix was understood by Gibbs before the advent of quantum mechanics. However, it is a purely quantum mechanical effect with no classical analog. The key point is that identical fundamental particles are *indistinguishable*. This means that the multi-particle state $|s_1, s_2, \ldots, s_N\rangle$ is the same (up to a sign) under a permutation of the labels,

$$|s_1, s_2, \ldots, s_N\rangle = (-1)^F |s_2, s_1, \ldots, s_N\rangle \tag{5.40}$$

Here, $F = 0$ if the particles are bosons and $F = 1$ if the particles are fermions.

Thus, for indistinguishable particles, the sum (5.33) over-counts the states. To correctly count the states, we must divide by the number of permutations of the labels $s_1, \ldots, s_N$. For example, if we have three bosonic particles in the following states, then the distinguishable partition function overcounts by the following

factors:

$$|1, 1, 1\rangle \to 1$$
$$|1, 1, 2\rangle \to 3$$
$$|1, 2, 3\rangle \to 6$$
$$|123, 14, 999\rangle \to 6. \tag{5.41}$$

If the particles are fermionic, then the first two states above actually vanish, while the last two are again overcounted in $Z^{\text{distinguishable}}$ by a factor of 6.

　　If the number of particles is much smaller than the number of states s_i, then most states have all labels s_i different (i.e. very few orbitals will be multiply occupied), so the distinguishable partition function overcounts by a factor of $N!$. This is a good approximation when we have a dilute gas, where multiple-occupancy is very rare. We will see how to deal correctly with multiple-occupancy later.

　　Let us assume a dilute gas. Taking into account the indistinguishability of fundamental particles, the correct partition function is

$$Z_N^{\text{indistinguishable}} = \frac{1}{N!} Z_1^N \qquad \text{(dilute gas)}. \tag{5.42}$$

The free energy is

$$\begin{aligned}
F &= -\tau \log Z_N \\
&= -\tau N \log \left(n_Q V \right) + \tau \log N! \\
&= -\tau N \log \left(n_Q V \right) + \tau (N \log N - N), \tag{5.43}
\end{aligned}$$

where we have used Stirling's approximation and only kept terms of size N or $N \log N$. The pressure computation is the same as before. However, now the result for the entropy is different

$$\begin{aligned}
\sigma &= - \left(\frac{\partial F}{\partial \tau} \right)_V \\
&= N \log(n_Q V) + \frac{3}{2} N - N \log N + N \\
&= N \left(\log \left(n_Q \frac{V}{N} \right) + \frac{5}{2} \right). \tag{5.44}
\end{aligned}$$

If we fix the concentration N/V, then this formula is now extensive. The extra factor of $1/N!$ has solved the Gibbs paradox. Equation (5.44) called the Sackur-Tetrode equation, and it does agree with experiment.

[End of lecture 8]

6   Blackbody radiation and the Planck distribution

6.1   The Planck distribution

The Planck distribution describes the spectrum of electromagnetic radiation in a cavity in thermal equilibrium. We can think of electromagnetic radiation as a gas of photons which are essentially non-interacting. However there are two main differences that will make the analysis different from the one in the previous section:

- Unlike in the case of the ideal gas, the gas of photons is not dilute. Each orbital can be occupied by multiple photons.

- In addition, photon number is not conserved — states with any number of photons should be considered accessible states.

6.1.1   Energy in a mode

An orbital in this case is an oscillation mode in the cavity. There can be zero or more photons present in each mode. The energy of a single photon in a mode with frequency ω is

$$\hbar\omega, \tag{6.1}$$

and the energy of s photons in this mode is

$$s\hbar\omega. \tag{6.2}$$

Thus, the possible energy levels associated with a mode look just like energy levels of a harmonic oscillator.

The partition function of a single mode is

$$Z_{\mathrm{mode}} = \sum_{s=0}^{\infty} e^{-s\hbar\omega/\tau} = \frac{1}{1 - e^{-\hbar\omega/\tau}} \tag{6.3}$$

where $y = \frac{\hbar\omega}{\tau}$. The average energy in the mode is

$$\begin{aligned}
\langle \mathcal{E} \rangle &= \tau^2 \frac{\partial}{\partial \tau} \log Z_{\mathrm{mode}} \\
&= -\tau^2 \frac{\partial}{\partial \tau} \log(1 - e^{-\hbar\omega/\tau}) \\
&= -\tau^2 \left(-\frac{\hbar\omega}{\tau^2} \right) \frac{e^{-\hbar\omega/\tau}}{1 - e^{-\hbar\omega/\tau}} \\
&= \frac{\hbar\omega}{e^{\hbar\omega/\tau} - 1}.
\end{aligned} \tag{6.4}$$

When $\hbar\omega \ll \tau$, we have simply $\langle \mathcal{E} \rangle \approx \tau$. When $\hbar\omega \gg \tau$, $\langle \mathcal{E} \rangle \approx 0$, and in fact it is exponentially suppressed. This suppression is due to the fact that the minimum energy in the mode is $\hbar\omega$, and the probability of having this energy is suppressed due to a Boltzmann factor.

6.1.2 What are photons?

Let us explain where this picture comes from. In quantum electrodynamics, an oscillating electric field is best understood in terms of its relation to a vector potential **A**:

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}$$
$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{6.5}$$

We can decompose the vector potential into spatial modes

$$\mathbf{A} = \sum_{\mathbf{k}} \mathbf{q_k}(t) \cos(\mathbf{k} \cdot x) + \mathbf{r_k}(t) \sin(\mathbf{k} \cdot x). \tag{6.6}$$

The fields are

$$\mathbf{E} = \sum_{\mathbf{k}} \left(-\dot{\mathbf{q}}_{\mathbf{k}}(t) \cos(\mathbf{k} \cdot x) - \dot{\mathbf{r}}_k(t) \sin(\mathbf{k} \cdot x) \right)$$
$$\mathbf{B} = \sum_{\mathbf{k}} \left(-\mathbf{k} \times \mathbf{q}(t) \sin(\mathbf{k} \cdot x) + \mathbf{k} \times \mathbf{r}(t) \cos(\mathbf{k} \cdot x) \right). \tag{6.7}$$

Plugging these expressions into the energy of an EM field, we have

$$E = \frac{1}{2} \int d^3x (\mathbf{E}^2 + c^2 \mathbf{B}^2)$$
$$\propto \sum_{\mathbf{k}} \frac{1}{2} \left(\dot{\mathbf{q}}_{\mathbf{k}}^2 + c^2 \mathbf{k}^2 \mathbf{q}_{\mathbf{k}}^2 \right) + \sum_{\mathbf{k}} \frac{1}{2} \left(\dot{\mathbf{r}}_{\mathbf{k}}^2 + c^2 \mathbf{k}^2 \mathbf{r}_{\mathbf{k}}^2 \right) \tag{6.8}$$

This is the Hamiltonian for a bunch of oscillators with frequency $\omega = c|\mathbf{k}|$. In quantum field theory, each classical oscillator gets promoted to a quantum oscillator, and the different energy levels of the oscillators are interpreted as photons.

6.1.3 Modes in a cavity

There are an infinite number of modes of the electromagnetic field, and each one can be occupied by any number of photons. Let us solve Maxwell's equations to find the radiation modes in a cubical cavity of size $L \times L \times L$.

Because Maxwell's equations are linear with time-independent coefficients, they can be solved by going to Fourier space in the time variables:

$$\mathbf{E}(\mathbf{x}, t) = \int_{-\infty}^{\infty} d\omega e^{i\omega t} \widetilde{\mathbf{E}}(\mathbf{x}, \omega),$$
$$\mathbf{B}(\mathbf{x}, t) = \int_{-\infty}^{\infty} d\omega e^{i\omega t} \widetilde{\mathbf{B}}(\mathbf{x}, \omega). \tag{6.9}$$

Because $\mathbf{E}$ and $\mathbf{B}$ must be real, we have the condition $\widetilde{\mathbf{E}}(\mathbf{x},\omega)^* = \widetilde{\mathbf{E}}(\mathbf{x},-\omega)$ and similarly for $\widetilde{\mathbf{B}}$.

Maxwell's equations are also translationally-invariant, so the functions $\widetilde{\mathbf{E}}(\mathbf{x},\omega)$ and $\widetilde{\mathbf{B}}(\mathbf{x},\omega)$ can be written as linear combinations of plane waves $e^{i\mathbf{k}\cdot\mathbf{x}}$. However, we must impose boundary conditions, and this will require us to take special linear combinations of $e^{i\mathbf{k}\cdot\mathbf{x}}$ that become sines and cosines. We have two kinds of boundary conditions:

- Firstly $E_\parallel = 0$. This follows from the equation $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$. We integrate this equation over the interior of a very thin rectangular loop with one edge inside the conductor and one edge outside. The right-hand side is zero because the rectangle has zero area. The left-hand side is $(E_\parallel)_{\text{out}} - (E_\parallel)_{\text{in}}$, by Stokes' theorem. Finally, we use the fact that electric fields vanish inside a conductor to conclude that $(E_\parallel)_{\text{out}} = 0$.

- The other boundary condition is $B_\perp = 0$. To derive this, we start from the equation $\nabla \cdot \mathbf{B} = 0$. Integrating this over a small "pill box" with a face inside the conductor and another face outside the conductor, we find $(B_\perp)_{\text{in}} = (B_\perp)_{\text{out}}$. Now the dynamic part of the magnetic field inside a conductor vanishes. The reason is that the electric and magnetic fields are related by

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \tag{6.10}$$

  Thus, a changing magnetic field inside a conductor implies the presence of an electric field, which is not allowed. A conductor can contain a static magnetic field, but we can ignore this in our analysis. All our modes will be oscillating at some nonzero frequency ω and do not include static contributions.

A linear combination of plane waves that satisfies the above boundary conditions is

$$\widetilde{E}_x(\mathbf{x},\omega) = E_{x0} \cos \frac{n_x \pi x}{L} \sin \frac{n_y \pi y}{L} \sin \frac{n_z \pi z}{L}$$
$$\widetilde{E}_y(\mathbf{x},\omega) = E_{y0} \sin \frac{n_x \pi x}{L} \cos \frac{n_y \pi y}{L} \sin \frac{n_z \pi z}{L}$$
$$\widetilde{E}_z(\mathbf{x},\omega) = E_{z0} \sin \frac{n_x \pi x}{L} \sin \frac{n_y \pi y}{L} \cos \frac{n_z \pi z}{L}. \tag{6.11}$$

where $\mathbf{n} = (n_x, n_y, n_z)$ is a vector of positive integers. It is hopefully easy to see that (6.11) obeys the boundary condition $E_\parallel = 0$. To see that it obeys $B_\perp = 0$, we can use (6.10). For example,

$$\dot{B}_x = \partial_y E_z - \partial_z E_y \propto \sin \frac{n_x \pi x}{L} \cos \frac{n_y \pi y}{L} \cos \frac{n_z \pi z}{L}, \tag{6.12}$$

which vanishes at $x = 0$.

The components of the electric field are not independent. We also have to impose

$$\nabla \cdot \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0. \tag{6.13}$$

This gives the condition

$$E_{x0}n_x + E_{y0}n_y + E_{z0}n_z = \mathbf{E}_0 \cdot \mathbf{n} = 0, \tag{6.14}$$

where we have defined $\mathbf{E}_0 = (E_{x0}, E_{y0}, E_{z0})$. For any given $\mathbf{n}$, equation (6.14) says that fluctuations of the electric field must be perpendicular to $\mathbf{n}$. The two perpendicular directions are the two possible polarizations of electromagnetic radiation.

Finally, we also have the wave equation, which says

$$c^2 \nabla^2 \mathbf{E} = \frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{6.15}$$

Plugging in (6.9) with $\widetilde{\mathbf{E}}(\mathbf{x}, \omega)$ given by (6.11), we find a relationship between ω and n:

$$\frac{c^2 \pi^2 n^2}{L^2} = \omega^2. \tag{6.16}$$

6.1.4   Energy of a photon gas and the Planck distribution

Now that we know the modes of the electromagnetic field in a cavity and the partition function (6.3) of a single mode, we can write down the partition function of a photon gas. Because each mode is independent, the partition function is

$$\begin{aligned} Z(\tau) &= \prod_{\mathbf{n}, \text{polarizations}} Z_{\text{mode}}(\mathbf{n}, \tau) \\ &= \prod_{\mathbf{n}} Z_{\text{mode}}(\mathbf{n}, \tau)^2 \\ &= \prod_{\mathbf{n}} \frac{1}{(1 - e^{-\hbar\omega_n/\tau})^2}. \end{aligned} \tag{6.17}$$

Let us take the logarithm, which turns the product into a sum:

$$\log Z = -2 \sum_{\mathbf{n}} \log(1 - e^{-\hbar\omega_n/\tau}). \tag{6.18}$$

We return to this quantity in a moment. For now, let us study the energy U of photons in the cavity. We have

$$\begin{aligned} U &= \tau^2 \frac{\partial}{\partial \tau} \log Z \\ &= 2 \sum_{\mathbf{n}} \frac{\hbar\omega_n}{e^{\hbar\omega_n/\tau} - 1}. \end{aligned} \tag{6.19}$$

We now make a "big box" approximation, so that the quantity $\hbar\omega_n/\tau$ is slowly-varying with n. In more detail, our assumption is $\frac{\hbar\pi c}{L\tau} \ll 1$. The quantity $\frac{\hbar c}{\tau}$ can be thought of as the thermal wavelength of a relativistic particle — it is the typical wavelength at temperature τ. We are making the assumption that $L \gg \frac{\hbar c}{\tau}$, so that the size of the box is large compared to the typical wavelength of radiation in the box. Under our assumption, the sum $\sum_{\mathbf{n}}$ turns into an integral

$$U \approx 2\frac{1}{8} \int_0^\infty 4\pi n^2 dn \frac{\hbar\omega_n}{e^{\hbar\omega_n/\tau} - 1}. \tag{6.20}$$

Let us change variables to the frequency $\omega = \frac{\pi c}{L}n$. We find

$$\frac{U}{V} = \int_0^\infty d\omega \frac{\hbar}{\pi^2 c^3} \frac{\omega^3}{e^{\hbar\omega/\tau} - 1}, \tag{6.21}$$

where $V = L^3$. The quantity in the integrand

$$u_\omega = \frac{\hbar}{\pi^2 c^3} \frac{\omega^3}{e^{\hbar\omega/\tau} - 1} \tag{6.22}$$

has the interpretation as the energy density of electromagnetic radiation per unit frequency range. This is the Planck radiation law.

In the small ω limit, the Planck law becomes

$$u_\omega \approx \frac{\omega^2\tau}{\pi^2 c^3} \qquad \hbar\omega \ll \tau \tag{6.23}$$

This is the Rayleigh-Jeans law. The $\hbar$'s have cancelled and the resulting formula can be understood purely classically. A problem with the Rayleigh-Jeans law is that if we extrapolate it up to high energies, the spectral density continues to grow, and we find that the energy density is divergent. This is the so-called "ultraviolet catastrophe" that was fixed by the advent of quantum mechanics. The catastrophe is avoided in the Planck distribution because high frequency modes are unoccupied due to the fact that they have a minimum energy $\hbar\omega$ (the energy of a single photon in the mode). This leads to exponential suppression by the Boltzmann factor.

Let us finally integrate (6.21) to get the total energy density

$$\frac{U}{V} = \frac{\hbar}{\pi^2 c^3} \left(\frac{\tau}{\hbar}\right)^4 \int_0^\infty dx \frac{x^3}{e^x - 1} \tag{6.24}$$

The integral is

$$\begin{aligned}
\int_0^\infty dx \frac{x^3}{e^x - 1} &= \int_0^\infty dx\, x^3 (e^{-x} + e^{-2x} + e^{-3x} + \dots) \\
&= \Gamma(4) \left(1 + \frac{1}{2^4} + \frac{1}{3^4} + \dots\right) \\
&= 6\zeta(4) \\
&= 6\frac{\pi^4}{90},
\end{aligned} \tag{6.25}$$

where $\zeta(s)$ is the Riemann-ζ function.[16] Altogether, we have

$$
\begin{aligned}
\frac{U}{V} &= \frac{\hbar}{\pi^2 c^3} \left(\frac{\tau}{\hbar}\right)^4 6\frac{\pi^4}{90} \\
&= \frac{\pi^2}{15c^3\hbar^3}\tau^4 \\
&= \alpha\tau^4.
\end{aligned} \tag{6.26}
$$

The fact that the energy density of radiation is proportional to τ^4 is the Stefan-Boltzmann law.

6.1.5   Energy flux

Consider a cubical cavity of radiation, and let us cut a small hole of area A in the cavity. The amount of radiation that escapes in a time Δt depends on the direction the radiation is moving. Specifically, if the angle between the wave vector and the normal to the hole is θ, then the radiation in a column of length

$$
c\Delta t \cos\theta \tag{6.27}
$$

escapes the hole.

   The average volume of radiation that escapes the hole is given by averaging over solid angles. A solid-angle element is $d\phi d\theta \sin\theta$. We integrate only over $\theta \in [0, \frac{\pi}{2}]$ because only right-moving radiation can escape the cavity. We divide by 4π which is the volume of S^2:

$$
V_{\text{escape}} = Ac\Delta t\frac{1}{4\pi}\int_0^{\frac{\pi}{2}} d\theta \int_0^{2\pi} d\phi \sin\theta\cos\theta = \frac{Ac\Delta t}{4}. \tag{6.28}
$$

Thus, the flux of energy per unit area per unit time is

$$
\begin{aligned}
J_U &= \frac{V_{\text{escape}}}{A\Delta t}\frac{U}{V} = \frac{c}{4}\alpha\tau^4 = \frac{\pi^2\tau^4}{60\hbar^3 c^3} = \sigma_B T^4 \\
\sigma_B &= \frac{\pi^2 k_B^4}{60\hbar^3 c^2}.
\end{aligned} \tag{6.29}
$$

[End of lecture 9]

---

[16]If you haven't heard of it, definitely read about the ζ-function, one of the most (in)famous functions in all of mathematics: https://en.wikipedia.org/wiki/Riemann_zeta_function. For a history and some derivations of values of the zeta function at even integers, see https://en.wikipedia.org/wiki/Basel_problem.

6.2   Kirchhoff's Law

Definition (black). An object is defined to be *black* in a given frequency range if all EM radiation incident on it in that frequency range is absorbed.

A small hole in a cavity is black because the radiation will enter the cavity and bounce around many times, being absorbed by the walls long before it escapes back out through the hole.

The radiant energy J_U from a black surface is equal to the energy flux density from a hole in a cavity at the same temperature. To see this, cover the hole with the surface. All the energy radiated from the hole is absorbed by the surface. But to achieve thermal equilibrium, the same amount of energy must be radiated from the surface back into the hole.

Definition (absorptivity). An object has absorptivity $a \in [0,1]$ if it absorbs a fraction a of the radiation incident upon it.

Definition (emissivity). An object has emissivity $e \in [0,1]$ if the energy flux from the object is e times that of a black body.

Realistic objects are not black and have some absorptivity/emissivity less than 1. The above argument about radiation flux from a black body generalizes to show that absorptivity and emissivity are always equal, $a = e$. Consider covering the hole in a cavity with an object with absorptivity a at temperature τ. The object absorbs an energy flux aJ_U^{black}. In order for thermal equilibrium to be maintained, it must emit the same energy flux aJ_U^{black}, implying $a = e$. This is called Kirchhoff's law. The law holds as a function of frequency as well, $a(\omega) = e(\omega)$.

6.3   Planck distribution, temperature, and CMB

The Planck law can be used to compute the temperature of an object. One looks for the peak of the energy density in frequency space. This occurs at $\frac{\hbar\omega}{\tau} \approx 2.82$.

An example is the cosmic microwave background (CMB): the universe is filled with approximately black-body radiation at temperature $2.73K$, see figure 2. This radiation was originally generated when the entire universe was a hot interacting plasma of electrons and protons strongly interacting with the radiation. When the temperature of the radiation cooled to $3000K$, most of the electrons and protons got bound into Hydrogen. This is called "recombination." After recombination, only specific frequencies of EM radiation could interact strongly with matter, so most of the photons decoupled from matter. The gas of photons then underwent isentropic expansion.

Definition (isentropic). We say that a process is *isentropic* if entropy remains constant during the process.
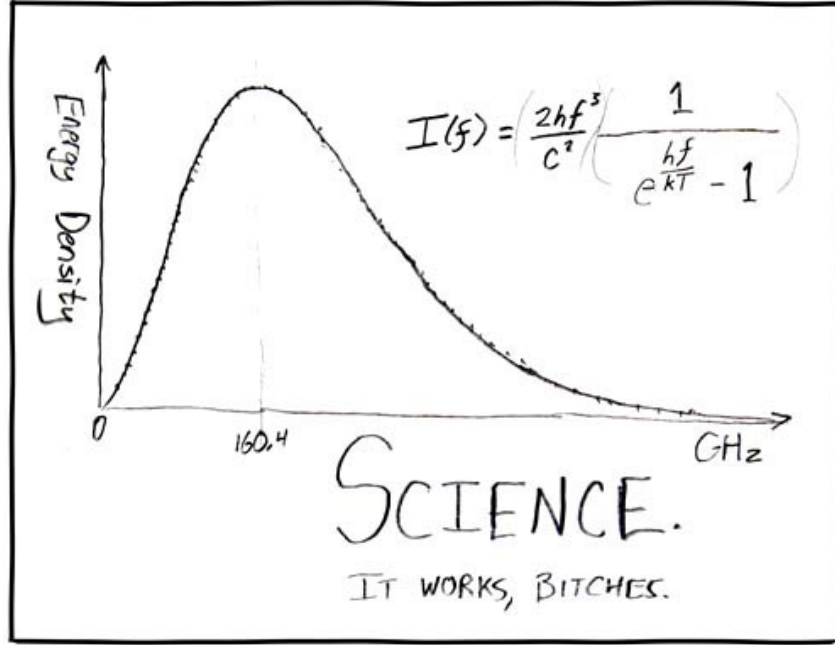
Figure 2: This xkcd comic (https://xkcd.com/54/) shows a plot of the energy density of the CMB as a function of frequency. The dots represent data from the COBE satellite, and the curve is the Planck distribution (with $f = 2\pi\omega$ and $\tau = kT$).

6.4   Partition function, entropy, and pressure of photon gas

Let us return to the log of the partition function (6.18). We can make the big-box approximation and turn it into an integral as before. We will be left with another rather complicated looking integral $\int dx x^2 \log(1 - e^{-x})$. This is doable using the same tricks as before.

However, instead of going through all this, let us simply integrate our expression for $U = \langle E \rangle = \tau^2 \frac{\partial}{\partial \tau} \log Z$:

$$\log Z = \int_0^\tau d\tau' \frac{U(\tau')}{\tau'^2} = \frac{\alpha V \tau^3}{3}. \tag{6.30}$$

The integration constant vanishes because when $\tau = 0$, all modes are in their ground state, and we see explicitly from (6.17) that $Z(\tau = 0) = 1$. The free energy is

$$F = -\tau \log Z = -\frac{\alpha V \tau^4}{3}. \tag{6.31}$$

The entropy is

$$\sigma = -\left(\frac{\partial F}{\partial \tau} \right)_V = \frac{4\alpha V \tau^3}{3}. \tag{6.32}$$

Note that this is consistent with $F = U - \tau\sigma$.

If a gas of photons undergoes isentropic expansion, the temperature scales like $V^{-1/3} = L^{-1}$, where $L = V^{1/3}$. This makes physical sense. Under isentropic expansion, the occupancy of every mode remains the same. The only effect of expansion $L \to \gamma L$ is that the wave vectors rescale by $\mathbf{k} \to \mathbf{k}/\gamma$, and consequently the frequencies and thus the energies in each mode scale like $\omega \to \omega/\gamma$. To keep the Boltzmann factors unchanged, we must have $\tau \to \tau/\gamma$ as well.

The pressure of the photon gas is

$$p = -\left(\frac{\partial F}{\partial V}\right)_\tau = \frac{1}{3}\alpha\tau^4 = \frac{1}{3}\frac{U}{V}. \tag{6.33}$$

Together with the energy density $\frac{U}{V}$, this quantity plays an important role in cosmology because it controls the way that radiation affects the evolution of the geometry of the universe.

The factor of $1/3$ in (6.33) reflects an interesting property of electromagnetic radiation: conformal symmetry. A conformally-symmetric system has a traceless stress-energy tensor. The stress-energy tensor of a stationary perfect fluid is given by

$$T_\mu{}^\nu = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}, \tag{6.34}$$

where $\rho = U/V$ is the energy density. Thus, tracelessness of T implies (6.33).

7   Debye theory

A solid has vibrational modes. A quantum of energy of a vibrational mode of a solid is called a phonon. In this section, we study the thermodynamics of a solid by modeling it as a gas of phonons. Phonons have some similarities to photons, so certain parts of the analysis of phonons will parallel our analysis of a photons. However, there are also some important differences.

One similarity is that in the limit of an infinitely large system, there can be phonons with very low energy. In the case of the electromagnetic field, this followed from the dispersion relation $\omega = c|\mathbf{k}|$, where $\mathbf{k} = \frac{\pi}{L}\mathbf{n}$ is the wavevector. In the limit that L is very large, there are wavevectors $\mathbf{k}$ with very small magnitude (i.e. modes with very long wavelength), and the energy $E = \hbar\omega|\mathbf{k}|$ of quanta in those modes is small.

Phonons also satisfy a dispersion relation of the same form $\omega = v|\mathbf{k}|$, but this time v is the speed of sound in the solid. In particular, note that we have arbitrarily low-energy phonons in an arbitrarily large solid $L \to \infty$, since then the wave number $|\mathbf{k}|$ can get arbitrarily tiny in that case. This is a consequence of translational symmetry. The argument is as follows: As the wavelength of a phonon gets longer, it looks locally more and more like a translation of the lattice. Translation doesn't cost any energy because it is a symmetry. Thus, the energy of a phonon should approach zero as $\mathbf{k} \to 0$. An excitation whose long-wavelength limit is controlled by a symmetry is called a goldstone boson, and a phonon is an example of a goldstone boson for translational symmetry.

An important difference between photons and phonons is as follows. There are an infinite number of electromagnetic modes in any fixed volume. However, the number of phonon modes in a solid is bounded. The reason is that a solid is made up of N atoms. So the number of position coordinates of the atoms is $3N$ and the number of momentum coordinates is $3N$. The normal modes are obtained by writing the Hamiltonian

$$H_{\text{solid}} = \sum_i \frac{p_i^2}{2m} + \frac{1}{2}m\sum_{i,j}\Omega_{ij}x_ix_j \tag{7.1}$$

and diagonalizing the matrix Ω_{ij}. The dimension of the space of modes is unchanged by the diagonalization, and hence the total number of modes is still $3N$.

Example 7.1 (1d lattice). Another way to say this is that wavelengths smaller than the lattice spacing are redundant with other wavelengths. For simplicity, consider a 1-dimensional lattice with lattice spacing a. A mode with number n is equivalent to a mode with number $n + \frac{2mL}{a}$ for $m \in \mathbb{Z}$ because

$$\sin\left(\frac{n\pi x}{L}\right) = \sin\left(\frac{(n + 2mL/a)\pi x}{L}\right), \tag{7.2}$$

for all $x \in a\mathbb{Z}$ in the lattice. (In signal processing, this phenomenon is related to the Nyquist limit.) Note that $L/a = N$, the number of atoms in the lattice, so this redundancy is $n \sim n + 2N$, where "$\sim$" means "is equivalent to." Furthermore, we have $n \sim -n$. Thus, the set of inequivalent mode numbers is $n = 1, \ldots, N$. There are N of them.

A vibrational wave has three polarizations: two transverse polarizations and one longditudinal polarization (think of a slinky). Thus, the sum over modes takes the form

$$\sum_{n,\text{polarizations}} \approx \frac{3}{8} \int 4\pi n^2 dn \tag{7.3}$$

Following Debye, let us suppose that n has a uniform cutoff n_D and determine n_D by solving the equation

$$\frac{3}{8} \int_0^{n_D} 4\pi n^2 dn = \frac{\pi}{2} n_D^3 = 3N$$

$$n_D = \left(\frac{6N}{\pi}\right)^{1/3}. \tag{7.4}$$

The calculation of the energy is the same as for photons. Now we have $\omega = v|\mathbf{k}|$, where v is the speed of sound. The speed of sound can actually be different for transverse and longditudinal polarizations in general, but we will assume it is the same for simplicity. As before $\mathbf{k} = \pi\mathbf{n}/L$ is a wavevector. The energy is

$$U = \sum_{\text{modes}} \frac{\hbar\omega}{e^{\hbar\omega/\tau} - 1}$$

$$= \frac{3}{8} 4\pi \int_0^{n_D} dn\, n^2 \frac{\hbar\omega_n}{e^{\hbar\omega_n/\tau} - 1}. \tag{7.5}$$

We have $\omega = \frac{\pi v}{L} n$, where v is the velocity of sound (which we assume is the same for all polarizations, for simplicity). Thus, the energy can be written

$$= \frac{3\pi^2 \hbar v}{2L} \left(\frac{\tau L}{\pi \hbar v}\right)^4 \int_0^{x_D} dx \frac{x^3}{e^x - 1}, \tag{7.6}$$

where

$$x_D = \frac{\theta}{\tau},$$

$$\theta = \hbar v \left(\frac{6\pi^2 N}{V}\right)^{1/3} \tag{7.7}$$

is the Debye temperature. The integral can be done analytically at low temperatures, where $x_D \to \infty$. The result is

$$U \approx \frac{3\pi^4 N \tau^4}{5\theta^3}. \tag{7.8}$$

The heat capacity is

$$C_V = \left(\frac{\partial U}{\partial \tau}\right)_V = \frac{12\pi^4 N}{5}\left(\frac{\tau}{\theta}\right)^3, \tag{7.9}$$

which matches experiment pretty well. This is called the Debye τ^3 law.

[End of lecture 10]

8   Classical Statistical Mechanics

We have seen that there are some quantities, like the partition function and the entropy from which quantum mechanical effects do not completely decouple. However, it is still possible to approximate these quantities using classical calculations with a small amount of input from quantum mechanics. This stems from a general principle that in the classical limit, numbers of states can be approximated by volumes of phase space. Recall that phase space is the space of positions and momenta of particles. For example, the classical phase space of a single particle in on dimension is specified by $(x, p) \in \mathbb{R}^2$.

Let us study a single quantum mechanical particle moving in 1 dimension, with Hamiltonian

$$\widehat{H} = \frac{\widehat{p}^2}{2m} + V(\widehat{x}). \tag{8.1}$$

I will try to carefully distinguish quantum operators from classical numbers using hats. We have seen that the partition function can be written

$$Z = \mathrm{Tr}(e^{-\beta\widehat{H}}) = \int_{-\infty}^{\infty} dx \langle x|e^{-\beta\widehat{H}}|x\rangle, \tag{8.2}$$

where on the second line, we chose the position basis.

Now, the partition function is complicated because we have the exponential of this operator involving $\widehat{x}$ and $\widehat{p}$, which of course do not commute

$$[\widehat{x}, \widehat{p}] = i\hbar. \tag{8.3}$$

The exponential of a sum of non-commuting matrices can be written

$$e^{\widehat{A}+\widehat{B}} = e^{\widehat{A}}e^{\widehat{B}}e^{-\frac{1}{2}[\widehat{A},\widehat{B}]+\cdots}, \tag{8.4}$$

where ... represents higher nested commutators $[[\widehat{A}, \widehat{B}], \widehat{B}]$, etc.. (This is a form of the Baker-Campbell-Hausdorff formula.)[17]

---

[17]As a sanity check on this formula, let us expand the left-hand side

$$e^{\widehat{A}+\widehat{B}} = 1 + \widehat{A} + \widehat{B} + \frac{1}{2}\left(\widehat{A}^2 + \widehat{A}\widehat{B} + \widehat{B}\widehat{A} + \widehat{B}^2\right) + \ldots$$

$$= 1 + \widehat{A} + \widehat{B} + \left(\frac{1}{2}\widehat{A}^2 + \widehat{A}\widehat{B} + \frac{1}{2}\widehat{B}^2 - \frac{1}{2}[\widehat{A}, \widehat{B}]\right) + \ldots$$

$$= \left(1 + \widehat{A} + \frac{1}{2}\widehat{A}^2 + \ldots\right)\left(1 + \widehat{B} + \frac{1}{2}\widehat{B}^2 + \ldots\right)\left(1 - \frac{1}{2}[\widehat{A}, \widehat{B}] + \ldots\right) \tag{8.5}$$

The claim is that if we include higher-order terms, we can always group the discrepancy between $e^{\widehat{A}+\widehat{B}}$ and $e^{\widehat{A}}e^{\widehat{B}}$ into nested commutators.

However, the classical limit is defined as the limit in which $\hbar \to 0$. (Because $\hbar$ is a dimensionful quantity, this really means that other dimensionful quantities with units of energy$\times$time (i.e. action) are large compared to $\hbar$.) In this limit, we have

$$e^{-\beta \frac{\hat{p}^2}{2m} - \beta V(\hat{x})} \approx e^{-\beta \frac{\hat{p}^2}{2m}} e^{-\beta V(\hat{x})}. \tag{8.6}$$

Plugging this into our trace, we get

$$
\begin{aligned}
Z &\approx \int_{-\infty}^{\infty} dx \langle x | e^{-\beta \frac{\hat{p}^2}{2m}} e^{-\beta V(\hat{x})} | x \rangle \\
&= \int_{-\infty}^{\infty} dx \langle x | e^{-\beta \frac{\hat{p}^2}{2m}} | x \rangle e^{-\beta V(x)} \\
&= \int_{-\infty}^{\infty} dx dp \langle x | e^{-\beta \frac{\hat{p}^2}{2m}} | p \rangle \langle p | x \rangle e^{-\beta V(x)} \\
&= \int_{-\infty}^{\infty} dx dp \langle x | p \rangle \langle p | x \rangle e^{-\beta \frac{p^2}{2m} - \beta V(x)}.
\end{aligned} \tag{8.7}
$$

But we also have

$$\langle x | p \rangle = A e^{\frac{ipx}{\hbar}}. \tag{8.8}$$

This is the wavefunction for a momentum eigenstate. The factor A is determined (up to a phase) by demanding that a momentum eigenstate be properly normalized

$$
\begin{aligned}
\langle p' | p \rangle &= \int_{-\infty}^{\infty} dx \langle p' | x \rangle \langle x | p \rangle \\
&= \int_{-\infty}^{\infty} dx |A|^2 e^{\frac{ix(p-p')}{\hbar}} \\
&= 2\pi\hbar |A|^2 \, \delta(p - p') \qquad \Longrightarrow \qquad |A|^2 = \frac{1}{2\pi\hbar}.
\end{aligned} \tag{8.9}
$$

Thus, finally we have

$$
\begin{aligned}
Z &\approx \frac{1}{2\pi\hbar} \int dx dp\, e^{-\beta \frac{p^2}{2m} - \beta V(x)} \\
&= \frac{1}{2\pi\hbar} \int dx\, dp\, e^{-\beta H(x,p)},
\end{aligned} \tag{8.10}
$$

where $H(x,p)$ is the classical Hamiltonian.

The lesson is that a sum over quantum mechanical states can be approximated by an integral over phase space, where one quantum state has phase-space volume $2\pi\hbar$ for each pair p, x. For example, in $N+N$-dimensional phase space, a quantum state has volume $(2\pi\hbar)^N$,

$$\mathrm{Tr}(e^{-\beta \hat{H}}) \to_{\mathrm{classical}} \int \frac{d^N x d^N p}{(2\pi\hbar)^N} e^{-\beta H(p_1, \ldots, p_N, x_1, \ldots, x_N)}. \tag{8.11}$$

Thus, a little bit of quantum mechanics is still necessary to make sense of a sum over states. However, once we know the density of states in phase space, we can compute various quantities in a classical approximation.

Example 8.1 (Classical ideal gas). As an application, let us recover our formula for the partition function of the ideal gas in the classical regime. In this case, we have $V(x) = 0$, so the single-particle partition function is

$$
\begin{aligned}
Z_1 &= \frac{1}{(2\pi\hbar)^3} \int d^3\vec{x} d^3\vec{p} e^{-\beta \frac{\vec{p}^2}{2m}} \\
&= \frac{V}{(2\pi\hbar)^3} \left(\int_{-\infty}^{\infty} dp\, e^{-\beta \frac{p^2}{2m}} \right)^3 \\
&= \frac{V}{(2\pi\hbar)^3} \left(\frac{2\pi m}{\beta} \right)^{3/2} = V \left(\frac{m\tau}{2\pi\hbar^2} \right)^{3/2}.
\end{aligned} \tag{8.12}
$$

Once we have the single-particle partition function, we can compute the full partition function via

$$
Z_N = \frac{1}{N!} Z_1^N, \tag{8.13}
$$

and this agrees with our previous analysis.

A nice property of the classical phase space integral is that it is simple and flexible: we could have a box with a complicated shape, and we don't have to worry about finding the quantum wavefunctions inside this shape and figuring out the number of states at each energy level. The contribution from the d^3x integral is always V. This means that we have actually *improved* our previous analysis where we had to assume a cubical box. We now see that the thermodynamics of the gas doesn't depend on the shape of the box, in the classical regime.

Example 8.2 (Density of eigenvalues of the Laplacian). We can turn this observation around and deduce an interesting mathematical result. Consider a box with an arbitrary shape. The wavefunctions for energy eigenstates inside the container are solutions of the time-independent Schrodinger equation

$$
-\frac{\hbar^2 \vec{\nabla}^2}{2m} \psi(\mathbf{x}) = E\psi(\mathbf{x}) \tag{8.14}
$$

with the boundary condition $\psi(\mathbf{x})|_{\text{boundary}} = 0$ (called a Dirichlet boundary condition). Thus, the energies are given by $E_i = \hbar^2 \lambda_i / 2m$, where λ_i are the eigenvalues of the Laplacian $-\vec{\nabla}^2$ with Dirichlet boundary conditions. For a general region, the spectrum of $-\vec{\nabla}^2$ can be quite complicated, see e.g. `https://en.wikipedia.org/wiki/Hearing_the_shape_of_a_drum`. However, our result (8.12) encodes a universal answer for the asymptotic density of large eigenvalues.

To see this, let us write the single-particle partition function as an integral over eigenvalues of the Laplacian

$$Z_1 = \sum_i e^{-E_i/\tau} = \sum_i e^{-\hbar^2 \lambda_i/(2m\tau)} = \int d\lambda\, \rho(\lambda) e^{-\hbar^2 \lambda/(2m\tau)}. \qquad (8.15)$$

In the last line, we introduced the density of eigenvalues $\rho(\lambda) = \sum_i \delta(\lambda - \lambda_i)$. In the classical limit $\hbar \to 0$, the exponential decays slowly and the integral is dominated by large λ. Let us suppose that $\rho(\lambda)$ behaves like a power law at large λ,[18]

$$\rho(\lambda) \sim a\lambda^b \qquad (\text{large } \lambda). \qquad (8.16)$$

In the classical limit, the integral becomes

$$Z \approx a\Gamma(b+1) \left(\frac{2m\tau}{\hbar^2} \right)^{b+1} \qquad (\hbar \to 0). \qquad (8.17)$$

Comparing to (8.12), we can solve for a and b to find

$$\rho(\lambda) \sim \frac{V}{4\pi^2} \lambda^{1/2} \qquad (\text{large } \lambda). \qquad (8.18)$$

This is Weyl's law `https://en.wikipedia.org/wiki/Weyl_law` in 3-dimensions. Can you derive the d-dimensional version?

8.1   Equipartition of energy

Equipartition of energy states that in the classical limit, every degree of freedom, defined as a quadratic term in the Hamiltonian, has energy $\frac{\tau}{2}$. This is simple to prove using phase space integrals. Suppose that the Hamiltonian has the form $H = \mathbf{y}^T A \mathbf{y}$, where $\mathbf{y}$ is a vector of phase space variables (x's and p's), and A is some positive-definite matrix. The partition function is

$$Z(\beta) = \int d^N \vec{y}\, e^{-\beta \mathbf{y}^T A \mathbf{y}} = \frac{1}{\beta^{N/2}} \int d^N \vec{y}\, e^{-\mathbf{y}^T A \mathbf{y}} = \frac{\text{const.}}{\beta^{N/2}}. \qquad (8.19)$$

The average energy is

$$U = -\frac{\partial}{\partial \beta} \log Z = -\frac{\partial}{\partial \beta} \left(-\frac{N}{2} \log \beta \right) = \frac{N}{2} \tau. \qquad (8.20)$$

In practice, the Hamiltonian is never exactly quadratic. Equipartition of energy applies when the quadratic approximation to the Hamiltonian is good over the

---

[18]In reality, $\rho(\lambda)$ is a sum of delta functions, so cannot be a power law. However, if we smooth it out a little bit (for example by averaging over a window), then it becomes a power law in the large λ limit.

entire region of phase-space that dominates the partition function integral. As an example, in the monatomic ideal gas, the Hamiltonian is quadratic in momenta to extremely high precision. However, the Hamiltonian is not well-approximated as being quadratic in position — instead the dependence on position is some function that changes sharply at the boundary of the box. Thus, we must separate out the position variables when computing $Z(\beta)$, which is why equipartition only applies to the three momenta in this case.

Furthermore, a quantum harmonic oscillator needs sufficiently high temperature in order to have $\langle E \rangle = \tau$, so equipartition of energy only applies to those degrees of freedom that are in the classical regime, i.e. such that $\tau \gg \hbar\omega$. For temperatures $\tau \ll \hbar\omega$ (the quantum regime), a quantum harmonic oscillator is exponentially likely to be in its ground state, and so has average energy $\langle E \rangle \approx 0$.

Example 8.3 (Monatomic gas). The equipartition theorem gives the energy of a monatomic gas in the classical regime almost instantly. Consider a gas of N particles. The Hamiltonian is $\frac{1}{2m} \sum_{i=1}^{N} \vec{p}_i^2$. This has $3N$ quadratic degrees of freedom, so the average energy is $U = \frac{3}{2}N\tau$ and the heat capacity is $C = \frac{3}{2}N$. This recovers our earlier result (5.34).

Example 8.4. The equipartition theorem also gives a useful qualitative understanding of the heat capacity of a diatomic gas. Diatomic molecules have

- 3 momenta,

- 2 rotation angles,

- 1 longitudinal vibration mode.

To apply equipartition of energy, we must decide which of these degrees of freedom is in the classical regime.

Assuming we have a big box, the momentum degrees of freedom are in the classical regime, and they contribute an energy per molecule of $\frac{3}{2}\tau$.

The rotation degrees of freedom are described by the quantum rigid rotor that you studied on your homework set. The rotor has a characteristic energy scale $E_{\rm rot} = \hbar^2/2I$, where I is the moment of inertia of the molecule. When τ is below this value, the rotational degrees of freedom are in the quantum regime, and when τ is above this value, the rotational degrees of freedom are in the classical regime. In the classical regime, the rotor has two quadratic degrees of freedom, which are the momenta conjugate to the rotation angles. Thus, the rotational degrees of freedom contribute energy $2 \times \frac{\tau}{2} = \tau$ per molecule in the classical regime.

Finally, the longitudinal vibration mode is in the classical regime if $\tau \gg \hbar\omega$, where ω is the characteristic frequency of oscillation. In this regime, there are *two* quadratic degrees of freedom — one position and one momentum variable for the harmonic oscillator. Thus longitudinal vibrations contribute energy $2 \times \frac{\tau}{2} = \tau$ per molecule in the classical regime.

The energy scales associated with spatial momenta, rotations, and longitudinal vibrations are usually ordered as

$$E_{\text{mom}} < E_{\text{rot}} < E_{\text{vib}}. \tag{8.21}$$

Thus, the heat capacity of a diatomic gas as a function of temperature looks like figure 3.
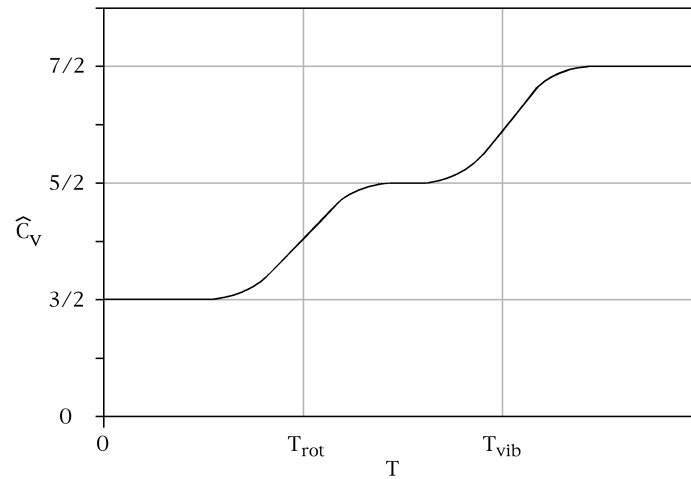
Figure 3: Heat capacity per molecule $\widehat{C} = C/N$ for a diatomic gas, as a function of temperature. At low temperatures, only translational degrees of freedom are in the classical regime and $\widehat{C} = \frac{3}{2}$. At intermediate temperatures, both translations and rotations are classical and $\widehat{C} = \frac{3}{2} + 1 = \frac{5}{2}$, and at high temperatures, translations, rotations, and vibrations are in the classical regime and $\widehat{C} = \frac{3}{2} + 1 + 1 = \frac{7}{2}$. Figure from Wikipedia https://en.wikipedia.org/wiki/Molar_heat_capacity.

[End of lecture 11]

9 Chemical potential and the Gibbs Distribution

9.1 Entropy and conserved quantities

We began this course by discussing the notion of *thermal contact*. Two systems are in *thermal contact* if energy is allowed to flow between them, but they are otherwise isolated. Our analysis of systems in thermal contact relied on two ingredients:

- The fundamental assumption of statistical mechanics, which says that the systems were overwhelmingly likely to maximize the entropy

$$\sigma = \sigma_1(U_1) + \sigma_2(U_2), \tag{9.1}$$

- Conservation of energy

$$U = U_1 + U_2. \tag{9.2}$$

From these ingredients, we derived the Boltzmann distribution, defined the partition function, and computed lots of interesting things.

Analogous reasoning applies whenever there is a conserved quantity that the entropy depends on. Such a conserved quantity gives rise to a generalization of the Boltzmann distribution, a generalization of the partition function, and lots more interesting results. In this section, we will develop these methods for an important (approximately) conserved quantity: particle number.

9.2 Chemical potential

Unlike energy, particle number is not a fundamental conserved quantity in the Standard Model of particle physics.[19] However, particle number can be effectively conserved in different physical situations. For example, if we consider physical processes at temperatures small compared to the activation energy of various chemical reactions, then the numbers of each type of atom/molecule are approximately conserved. If we consider processes at temperatures small compared to the binding energies of nuclei, then the numbers of various types of nuclei are approximately conserved.

Thus, let us generalize our tools to take into account conserved particle number. For simplicity, we mostly assume that there is one species of particle, and that the total number N of that species is conserved.

Definition (diffusive contact). In analogy with the definition of thermal contact, we say that two systems $\mathcal{S}_1, \mathcal{S}_2$ are in *diffusive contact* if they can exchange particles.

---

[19]There is a particle-number-like quantity called $B - L$, which measures the difference between baryon number and lepton number, that is very nearly conserved. Experiments have not yet been able to detect changes in $B - L$. However, $B - L$ non-conservation is expected due to quantum gravitational effects. For example, Hawking radiation of black holes violates $B - L$ conservation.

Consider two systems $\mathcal{S}_1, \mathcal{S}_2$ in thermal and diffusive contact. Let the number of particles in system $\mathcal{S}_i$ be N_i, and let the energy of $\mathcal{S}_i$ be U_i. Let the total number of particles be N and the total energy be U. The entropy of each system depends on the number of particles and the energy of that system:

$$\sigma_1(U_1, N_1) \qquad \text{and} \qquad \sigma_2(U_2, N_2). \tag{9.3}$$

By conservation of energy and particle number, the total entropy is

$$\begin{aligned}
\sigma &= \sigma_1(U_1, N_1) + \sigma_2(U_2, N_2) \\
&= \sigma_1(U_1, N_1) + \sigma_2(U - U_1, N - N_1) \tag{9.4}
\end{aligned}$$

The systems will reach thermal and diffusive equilibrium in the most probable configuration, where (9.4) is maximized. This leads to the conditions

$$\begin{aligned}
\left(\frac{\partial \sigma_1}{\partial U_1}\right)_N &= \left(\frac{\partial \sigma_2}{\partial U_2}\right)_N, \\
\left(\frac{\partial \sigma_1}{\partial N_1}\right)_U &= \left(\frac{\partial \sigma_2}{\partial N_2}\right)_U. \tag{9.5}
\end{aligned}$$

The first condition is simply $\tau_1 = \tau_2$. To write the second condition, let us introduce

Definition (chemical potential). The chemical potential μ is defined by

$$\frac{\mu}{\tau} = -\left(\frac{\partial \sigma}{\partial N}\right)_U, \tag{9.6}$$

where τ is the temperature.

The factor of τ is a convention and it means that μ has dimensions of energy. The minus sign is another convention introduced so that chemical potential behaves kind of like a potential energy for particles. To see this, suppose that $\mu_1 > \mu_2$. We see that in order to increase the total entropy (9.4), we must decrease N_1 and increase N_2. Thus, particles move from the region of higher chemical potential to the region of lower chemical potential.

With the above definition, the condition for diffusive equilibrium is

$$\mu_1 = \mu_2 \qquad \text{(diffusive equilibrium)}. \tag{9.7}$$

If several different species of particle are present, each one has its own chemical potential

$$\frac{\mu_j}{\tau} = -\left(\frac{\partial \sigma}{\partial N_j}\right)_{U, N_1, \ldots, N_{j-1}, N_{j+1}, \ldots}. \tag{9.8}$$

(Here, the partial derivative is taken while holding all particle numbers except N_j fixed.) The condition for diffusive equilibrium becomes

$$(\mu_j)_1 = (\mu_j)_2 \qquad \text{for all species } j. \tag{9.9}$$

9.3 Relation to free energy

Consider a change in the energy U and particle number N of a system. The entropy changes by

$$
\begin{aligned}
d\sigma &= \left(\frac{\partial \sigma}{\partial U}\right)_N dU + \left(\frac{\partial \sigma}{\partial N}\right)_U dN \\
&= \frac{1}{\tau}dU - \frac{\mu}{\tau}dN.
\end{aligned}
\tag{9.10}
$$

Now consider the change in free energy

$$
\begin{aligned}
dF &= d(U - \tau\sigma) = dU - \tau d\sigma - \sigma d\tau \\
&= \tau d\sigma + \mu dN - \tau d\sigma - \sigma d\tau \\
&= \mu dN - \sigma d\tau.
\end{aligned}
\tag{9.11}
$$

This leads to another expression for μ in terms of the free energy:

$$
\mu = \left(\frac{\partial F}{\partial N}\right)_\tau.
\tag{9.12}
$$

If the system depends on a volume V, then we must keep that fixed as well when computing the derivative

$$
\mu = \left(\frac{\partial F}{\partial N}\right)_{\tau,V}.
\tag{9.13}
$$

These expressions are useful because F is simple to compute, via its relationship to Z.

Example 9.1 (Ideal gas). We previously computed the free energy of the ideal gas

$$
\begin{aligned}
F &= -\tau \log\left(\frac{1}{N!}(n_Q V)^N\right) \\
&= -\tau\left(N \log n_Q V - N \log N + N\right).
\end{aligned}
\tag{9.14}
$$

The chemical potential is

$$
\mu = -\tau\left(\log n_Q V - \log N\right) = \tau \log(n/n_Q),
\tag{9.15}
$$

where $n = N/V$ is the concentration. Typical gasses are dilute relative to the quantum concentration $n/n_Q \ll 1$, so that μ is negative.

Example 9.2 (Voltage barrier). Consider two systems $\mathcal{S}_1, \mathcal{S}_2$ with chemical potentials μ_1, μ_2. Using (9.12), we have

$$
\frac{\partial F}{\partial N_1} = \mu_1 - \mu_2.
\tag{9.16}
$$

That is, the derivative of the total free energy $F = F_1 + F_2$ with respect to N_1 is the difference of chemical potentials.

Suppose also that the particles are charged with charge q, and imagine turning on a voltage difference between the two systems so that particles in system $\mathcal{S}_i$ have potential energy qV_i.[20] The potential energy will give a new contribution to the total energy and hence the free energy:

$$F' = F + N_1 q V_1 + N_2 q V_2 \qquad (9.17)$$

We now have

$$\frac{\partial F'}{\partial N_2} = \mu_1 - \mu_2 + q(V_1 - V_2) \qquad (9.18)$$

Thus, the new difference of chemical potentials after turning on the voltages is $\mu_1 - \mu_2 + q(V_1 - V_2)$. We can achieve diffusive equilibrium by tuning $q(V_1 - V_2) = -(\mu_1 - \mu_2)$.

Thus, another definition of chemical potential is the electrical potential difference required to maintain diffusive equilibrium. This gives a way of measuring chemical potential for charged particles.

In general, when an external potential is present, the total chemical potential of a system is

$$\mu = \mu_{\text{ext}} + \mu_{\text{int}}, \qquad (9.19)$$

where μ_{ext} is the potential energy per particle in the external potential, and the *internal chemical potential* μ_{int} is the chemical potential that would be present if the external potential were zero.

Example 9.3 (Isothermal model of the atmosphere). We can model the atmosphere as an ideal gas at temperature τ that is in thermal and diffusive equilibrium. The isothermal/equilibrium assumptions are actually not true, but this is a reasonable model. The total chemical potential at height h is

$$\begin{aligned}
\mu &= \mu_{\text{int}} + \mu_{\text{ext}} \\
&= \tau \log(n(h)/n_Q) + Mgh, \qquad (9.20)
\end{aligned}$$

where M is the mass of a gas particle, h is the height, and we have allowed the concentration $n(h)$ to depend on height. In equilibrium, this must be constant, so we have

$$\begin{aligned}
\tau \log(n(h)/n_Q) + Mgh &= \tau \log(n(0)/n_Q) \\
n(h) &= n(0)e^{-Mgh/\tau}. \qquad (9.21)
\end{aligned}$$

_____

[20]Don't confuse V_i with volume in this example.

An alternative fun way to derive this result (that you'll do on the homework) is to demand that the pressure of a slab of gas be enough to hold up the slab above it.

Here's a third slick way that I really like. Consider a single molecule of gas as a subsystem and the whole atmosphere as a reservoir at temperature τ. The density of orbitals at height h is independent of h. Thus, the probability that the molecule will be at height h is proportional to a Boltzmann factor $P(h) \propto e^{-E(h)/\tau} = e^{-Mgh/\tau}$. The concentration $n(h)$ is proportional to this probability.

9.4   Gibbs factor and Gibbs sum

We have seen that N can be treated as a conserved quantity in a way similar to energy U. Let us revisit our discussion of system and reservoir, this time allowing the system and reservoir to be in both diffusive and thermal contact.

Definition (general reservoir). Let us generalize our definition of a reservoir to be a system whose entropy is linear in fluctuations of conserved quantities, to very good approximation.

In the case of energy fluctuations, we argued that this definition is reasonable by considering how derivatives $\frac{\partial^k \sigma}{\partial U^k}$ scaled with the size of a system: the higher derivatives become very small for large systems. The same reasoning holds for any conserved quantity; hence the definition above.

If $\mathcal{R}$ is such a reservoir, then we have

$$\sigma_{\mathcal{R}}(U_0 - E_s, N_0 - N_s) = \sigma_{\mathcal{R}}(U_0, N_0) - \left(\frac{\partial \sigma_{\mathcal{R}}}{\partial N}\right)_U N - \left(\frac{\partial \sigma_{\mathcal{R}}}{\partial U}\right)_N E_s$$

$$= \sigma_{\mathcal{R}}(U_0, N_0) + \mu N_s/\tau - E_s/\tau. \tag{9.22}$$

Let us suppose our system is in a state with energy E_s and particle number N_s. The reservoir has energy $U_0 - E_s$ and particle number $N_0 - N_s$. The probability of the state is proportional to

$$g_{\mathcal{R}}(U_0 - E_s, N_0 - N_s) = e^{\sigma_{\mathcal{R}}(U_0, N_0) + \mu N_s/\tau - E_s/\tau}. \tag{9.23}$$

To compute the constant of proportionality, we can impose that the sum of probabilities is equal to 1. We find

$$P(s) = \frac{e^{\mu N_s/\tau - E_s/\tau}}{\mathcal{Z}(\tau, \mu)} \tag{9.24}$$

$$\mathcal{Z}(\tau, \mu) = \sum_s e^{(N_s \mu - E_s)/\tau} = \sum_{N=0}^{\infty} \sum_{s(N)} e^{(N\mu - E_{s(N)})/\tau}, \tag{9.25}$$

where $s(N)$ indicates states with particle number N.

Definition (grand canonical ensemble). The probability distribution $P(s)$ defined in (9.24) is called the *grand canonical ensemble*.

Definition (grand canonical partition function/Gibbs sum). The object $\mathcal{Z}$ defined in (9.25) is sometimes called the *grand canonical partition function*, although Kittel and Kroemer call it the *Gibbs sum*.

Sometimes it is convenient to define the *activity* $\lambda = e^{\mu/\tau}$. Then the grand canonical partition function can be written

$$\mathcal{Z} = \sum_{N=0}^{\infty} \lambda^N Z_N,$$
$$Z_N = \sum_{s(N)} e^{-E_s/\tau}, \tag{9.26}$$

so it is essentially a generating function for canonical partition functions with different particle number.

We can use the Gibbs sum in the same way that we use the usual partition function. For example, we can compute the expected number of particles

$$\begin{aligned}
\langle N \rangle &= \sum_s N_s P(s) \\
&= \sum_s \frac{N_s e^{(N_s \mu - E_s)/\tau}}{\mathcal{Z}(\tau, \mu)} \\
&= \tau \frac{\partial}{\partial \mu} \log \mathcal{Z}(\tau, \mu). \tag{9.27}
\end{aligned}$$

One tricky point is that because of the way μ is defined, the expression for the average energy is different in the grand canonical ensemble:

$$\begin{aligned}
U = \langle E_s \rangle &= \left(\frac{\mu}{\beta} \frac{\partial}{\partial \mu} - \frac{\partial}{\partial \beta} \right) \log \mathcal{Z} \\
&= \mu \langle N \rangle - \frac{\partial}{\partial \beta} \log \mathcal{Z}. \tag{9.28}
\end{aligned}$$

Example 9.4 (Ionization of impurities in a semiconductor). A semiconductor has energy levels called the "conduction band" which act as an ideal gas of electrons. The semiconductor can also contain impurities, which are atoms whose electrons are in thermal and diffusive equilibrium with the conduction band. In a simple model, an impurity atom has a single valence electron, and that electron can be either absent (zero energy, ionized) or present in either the spin-up or spin-down state (energy $-I$). The Gibbs sum is

$$1 + 2e^{(\mu + I)/\tau}. \tag{9.29}$$

The probability of ionization is

$$\frac{1}{1 + 2e^{(\mu + I)/\tau}}. \tag{9.30}$$

[End of lecture 12]

10   Ideal Gas

10.1   Spin and statistics

10.1.1   Bosons vs. Fermions

Quantum mechanical particles are indistinguishable. This means that probabilities and expectation values of operators are invariant under exchanging the labels associated to a pair of particles. Consider a two-particle state

$$|x_1, x_2\rangle. \tag{10.1}$$

Here, the particles are labeled by their positions. If they also had spin or other internal quantum numbers, then they would carry those labels as well. Under relabeling $x_1 \leftrightarrow x_2$, observables must be unchanged. This means that we must have

$$|x_2, x_1\rangle = e^{i\phi}|x_1, x_2\rangle. \tag{10.2}$$

Note that we don't simply have 1 on the right-hand side because we are free to rotate amplitudes by a phase without changing expectation values and probabilities. However, exchanging the labels twice should bring us back to the same state. Thus, the phase should be a plus or minus sign

$$|x_2, x_1\rangle = \pm|x_1, x_2\rangle. \tag{10.3}$$

Given a general two-particle state $|\Psi\rangle$, the wavefunction is

$$\Psi(x_1, x_2) = \langle x_1, x_2|\Psi\rangle. \tag{10.4}$$

If the particles get a $\pm$ sign under exchange according to (10.3), then the wavefunction satisfies

$$\Psi(x_1, x_2) = \pm\Psi(x_2, x_1). \tag{10.5}$$

We see that particles fall into two types:

- *bosons* have wavefunctions that are symmetric under exchange of labels,

- *fermions* have wavefunctions that are antisymmetric under exchange of labels.

Let us try to build two-particle wavefunctions from a set of single-particle wavefunctions $\psi_a(x)$ ("orbitals"). For example, the single particle wavefunctions could be energy eigenstates of the single-particle Schrödinger equation in some potential. We can define the two-particle wavefunctions

$$\Psi_{ab}(x_1, x_2) = \psi_a(x_1)\psi_b(x_2) \pm \psi_a(x_2)\psi_b(x_1), \tag{10.6}$$

and these will automatically have the correct property under exchange (10.5). However, consider the case where $a = b$:

$$\Psi_{aa}(x_1, x_2) = \psi_a(x_1)\psi_a(x_2) \pm \psi_a(x_2)\psi_a(x_1). \tag{10.7}$$

If we have a $+$ sign (bosons), there is no problem with this wavefunction. However, if we have a minus sign (fermions), then it is simply zero. This is the Pauli exclusion principle: two fermions cannot occupy the same orbital.

Thus, the types of multi-particle states that are available depend on whether we have bosons or fermions. This has dramatic consequences for their statistics, and consequently for experiment.

10.1.2   Spin review

Recall that a rotation by an angle θ acts on a quantum state by

$$R_{\mathbf{n}}(\theta) = e^{i\theta\mathbf{n}\cdot\mathbf{J}}, \tag{10.8}$$

where $\mathbf{n}$ is a unit vector for the axis of rotation, and $\mathbf{J} = (J_x, J_y, J_z)$ are angular momentum generators. By studying the commutation relations of the J's, one can show that the possible eigenvalues for any of the J's are

$$m = -j, -j+1, \ldots, j-1, j, \tag{10.9}$$

where j is either an integer or half-integer. For example, acting on an eigenstate of J_z, and abbreviating $R(\theta) = R_{\mathbf{e}_z}(\theta)$, we have

$$R(\theta)|m\rangle = e^{im\theta}|m\rangle. \tag{10.10}$$

Note that a rotation by 2π gives either ± 1, depending on whether j is integer or half-integer (regardless of which value of m (10.9) we have):

$$R(2\pi)|m\rangle = q|m\rangle, \qquad q = \begin{cases} +1 & j \in \mathbb{Z} \\ -1 & j \in \mathbb{Z} + \frac{1}{2} \end{cases}. \tag{10.11}$$

The generators $\mathbf{J}$ get contributions from both the orbital angular momentum and the intrinsic spin of a particle. In particular, if a spin-s particle is in a position eigenstate at the origin, then $R(2\pi)$ acts on it by ± 1, depending on whether the spin is integer or half-integer.

The fact that $R(2\pi)$ is not necessarily equal to 1 is surprising from the classical point of view. Classically, a 2π rotation brings a system back to itself. However, this is not necessarily true for quantum states. Only probabilities are preserved under a 2π rotation, and this leaves room for the appearance of a phase. You might ask: why can the phase only be ± 1, and not something more general? The answer is related to the topology of the rotation group.

The reason we have the possibility of a nontrivial phase associated with 2π rotations is that, when rotating the system, we can keep track of the path that we take in the rotation group. The final quantum state we get can depend on that path, not just the endpoint of the path.

In 2 dimensions, we can picture the rotation group as a circle, and a rotation by 2π corresponds to moving once around the circle. A rotation by 4π corresponds to moving twice around the circle. The circle allows for arbitrary integer winding numbers, and so we can imagine accumulating many different phases as we go around and around.

In 3 dimensions, the rotation group is a more complicated topological space.[21] It turns out that a rotation by 2π cannot be continuously deformed into a trivial path, but a rotation by 4π can. See for example the demonstration with coffee cup from the lecture, or <https://twitter.com/krzhang/status/1256117297766268936>.

10.1.3   The spin-statistics theorem

A-priori, there is no obvious connection between the sign associated with $R(2\pi)$ and the sign associated with exchange of particle labels. However, it turns out that they are directly related.

Theorem 10.1 (Spin-statistics). All particles with integer spin are bosons and all particles with half-integer spin are fermions.

The spin-statistics theorem is not obvious. It is a consequence of relativistic invariance and the existence of antiparticles. A heuristic picture of why it's true is as follows. Let us imagine drawing the world lines of particles. We draw an arrow to indicate particles vs. anti-particles. These world lines have two important properties:

- They can be created or destroyed in pairs. Thus, we can have loops.

- They keep track of the history of rotations applied to each particle. To model, this we can thicken each line into a "belt."

In the Feynman path integral formalism, we must assign an amplitude to each possible history of particles. *On the board, draw a history associated to an exchange and explain how it can be modified into a history where one particle is rotated by* 2π.

10.2   Statistics of Fermions and Bosons

10.2.1   Fermi-Dirac statistics

Consider an orbital (solution of the single-particle Schrödinger equation) that can be occupied or unoccupied by a fermion, and let the energy of the orbital be E. The

---

[21]Topologically, it is equivalent to $S^3/\mathbb{Z}_2$.

Gibbs sum is

$$\mathcal{Z}_{\text{orbital}} = 1 + \lambda e^{-E/\tau}, \tag{10.12}$$

where $\lambda = e^{\mu/\tau}$. The expected occupancy is

$$f(E) = \langle N_{\text{orbital}} \rangle = \frac{\lambda e^{-E/\tau}}{\mathcal{Z}} = \frac{1}{e^{(E-\mu)/\tau} + 1}. \tag{10.13}$$

This is the Fermi-Dirac distribution. It is equal to 1 when $E \ll \mu$, 0 when $E \gg \mu$, and $1/2$ when $E = \mu$. At low temperatures, it approaches a step function at $E = \mu(\tau = 0) = E_F$ (the Fermi energy).

10.2.2   Bose-Einstein statistics

Let us consider the analogous problem for bosons. In this case, each orbital can be occupied by an arbitrary number of bosons. The Gibbs sum is a geometric series

$$\mathcal{Z}_{\text{orbital}} = \sum_{N=0}^{\infty} \lambda^N e^{-NE/\tau} = \frac{1}{1 - \lambda e^{-E/\tau}}. \tag{10.14}$$

The expected occupancy is

$$f(E) = \lambda \frac{\partial}{\partial \lambda} \log \mathcal{Z}_{\text{orbital}} = \frac{\lambda e^{-E/\tau}}{1 - \lambda e^{-E/\tau}}$$
$$= \frac{1}{e^{(E-\mu)/\tau} - 1}. \tag{10.15}$$

This is the Bose-Einstein distribution. The only difference with the Fermi-Dirac distribution is the minus sign in the denominator, but that difference is dramatic. Note that the occupancy blows up when $E \ll \mu$ — bosons love to live together in low energy orbitals.

10.3   The classical limit

An ideal gas is a system of free non-interacting particles in the classical regime. Here, classical regime means that the occupancy of each orbital is much smaller than 1. In this case, we have

$$e^{(E-\mu)/\tau} \gg 1, \tag{10.16}$$

so that

$$f_{FD}(E) \approx f_{BE}(E) \approx e^{(\mu-E)/\tau} = \lambda e^{-E_s/\tau}. \tag{10.17}$$

This is called the "classical distribution function."

You might recall that in our initial analysis of the ideal gas, we used the low-occupancy assumption to justify dividing by $N!$ in the multi-particle partition function. This ignored the combinatorics associated with multiple occupancy. Using the technology of Gibbs sums and chemical potentials, we will be able to treat the low-occupancy approximation in a different way.[22] We will study an ideal gas of particles in the classical limit, where the difference between FD and BE statistics is unimportant. Later, we will study gasses in the quantum regime, where the different statistics of fermions and bosons will play a major role.

Let us determine the chemical potential of a gas of N atoms in the classical regime. We must have

$$N = \langle N \rangle = \sum_s f(E_s)$$

$$= \lambda \sum_s e^{-E_s/\tau}$$

$$= \lambda Z_1, \tag{10.18}$$

where Z_1 is simply the single-particle partition function. We have already evaluated this: $Z_1 = n_Q V$, where $n_Q = \left(\frac{M\tau}{2\pi\hbar^2}\right)^{3/2}$ is the quantum concentration. Thus, we find

$$N = \lambda n_Q V \quad \Longrightarrow \quad \lambda = n/n_Q,$$
$$\Longrightarrow \quad \mu = \tau \log(n/n_Q). \tag{10.19}$$

where $n = N/V$ is the conentration. This recovers our result from before.

From the chemical potential $\mu = \left(\frac{\partial F}{\partial N}\right)_{\tau,V}$, we can recover the free energy

$$F = \int_0^N dN \mu$$

$$= \int_0^N \tau \left(\log(N) + \log(1/(n_Q V))\right)$$

$$= \tau \left((N \log N - N) + N \log(1/(n_Q V))\right)$$

$$= \tau N (\log(n/n_Q) - 1). \tag{10.20}$$

The pressure is

$$p = -\left(\frac{\partial F}{\partial V}\right)_{\tau,N} = N\tau/V, \tag{10.21}$$

as before. The entropy is

$$\sigma = -\left(\frac{\partial F}{\partial \tau}\right)_{V,N} = -N \left(\log(n/n_Q) - 1 - \frac{3}{2}\right)$$

$$= N \left(\log(n_Q/n) + \frac{5}{2}\right), \tag{10.22}$$

---

[22]A way where the errors are more easily quantified.

which is the Sackur-Tetrode equation again.

11   Fermi Gases

In the previous section, we considered gases in the classical regime $n \ll n_Q$. The behavior of gases in the quantum regime $n \gg n_Q$ is rich and interesting, and in particular depends in a crucial way on whether the gas is composed of fermions or bosons. In this section, we will study a Fermi gas (gas of fermions). Some examples of Fermi gases are

- Electrons in the conduction band of a metal,

- Electrons in a white dwarf star,

- Liquid ^{3}He,

- Nuclear matter.

11.1   Gibbs sum and energy

For concreteness, we will consider a gas of nonrelativistic spin-$\frac{1}{2}$ particles (e.g. electrons). Such particles have two spin states, $s = \pm\frac{1}{2}$. Suppose the particles are constrained to a cube of side length L. The orbitals are labeled by mode numbers which are a triplet of nonnegative integers $\mathbf{n} = (n_x, n_y, n_z) \in \mathbb{Z}^3_{\geq 0}$, and also a spin $s = \pm\frac{1}{2}$. Their energies are a function of $n = |\mathbf{n}|$ given by

$$E_n = \frac{\hbar^2}{2m}\left(\frac{\pi n}{L}\right)^2. \tag{11.1}$$

Each orbital can be occupied or unoccupied, so the partition function of a single orbital is

$$\mathcal{Z}_{\mathbf{n}} = 1 + e^{\beta(\mu - E_n)}, \tag{11.2}$$

where $\beta = 1/\tau$.

   The Gibbs sum of the entire system is a product of the Gibbs sums for each orbital

$$\mathcal{Z} = \prod_{\substack{\mathbf{n} \in \mathbb{Z}^3_{\geq 0} \\ s = \pm\frac{1}{2}}} \mathcal{Z}_{\mathbf{n}} = \prod_{\mathbf{n} \in \mathbb{Z}^3_{\geq 0}} \mathcal{Z}_{\mathbf{n}}^2. \tag{11.3}$$

Note that we have a square because of the possibility of having two different spins.

The expected number of particles is

$$\langle N \rangle = \frac{\sum_s N_s e^{\beta(\mu N_s - E_s)}}{\mathcal{Z}}$$

$$= \frac{1}{\beta}\frac{\partial}{\partial\mu}\log\mathcal{Z}$$

$$= \frac{1}{\beta}\frac{\partial}{\partial\mu}2\sum_{\mathbf{n}}\log(1 + e^{\beta(\mu - E_n)})$$

$$= 2\sum_{\mathbf{n}}\frac{e^{\beta(\mu - E_n)}}{1 + e^{\beta(\mu - E_n)}} = 2\sum_{\mathbf{n}}f_{\mathrm{FD}}(E_n, \tau, \mu), \tag{11.4}$$

where we have abbreviated $\sum_{\mathbf{n}} = \sum_{\mathbf{n}\in\mathbb{Z}^3_{\geq 0}}$. This is an intuitive formula: the Fermi-Dirac distribution gives the expected occupancy of each mode number, so the expected value of N is the sum of the expected occupancies, times 2 to account for spin. We abbreviate $f_{\mathrm{FD}}(E, \tau, \mu)$ as $f(E, \tau, \mu)$ or sometimes $f(E)$, throughout this section.

[End of lecture 13]

Let us make the "big box" assumption that the thermal wavelength is much smaller than the size of the box. The sum over modes $\sum_{\mathbf{n}}$ can then be approximated as a phase space integral. For any function f, we have

$$\sum_{\mathbf{n}} f(E_n) \approx \int \frac{d^3\mathbf{x}\, d^3\mathbf{p}}{(2\pi\hbar)^3} f\left(\frac{\mathbf{p}^2}{2m}\right)$$

$$= \frac{4\pi V}{(2\pi\hbar)^3}\int_0^\infty dp\, p^2 f\left(\frac{p^2}{2m}\right)$$

$$= \frac{4\pi V}{(2\pi\hbar)^3}\int_0^\infty d(\sqrt{2mE})\,(2mE)f(E)$$

$$= \frac{V}{4\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\int_0^\infty dE\, E^{1/2} f(E)$$

$$= \int_0^\infty dE\, \mathcal{D}_0(E) f(E), \tag{11.5}$$

where we have defined the "density of modes"

$$\mathcal{D}_0(E) = \frac{V}{4\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} E^{1/2}. \tag{11.6}$$

You may recognize this as just Weyl's law (8.18). Remember that the density of eigenvalues of the Laplacian $-\vec{\nabla}^2$ is given by

$$\rho(\lambda)d\lambda = \frac{V}{4\pi^2}\lambda^{1/2}d\lambda. \tag{11.7}$$

To get $\mathcal{D}_0(E)dE$, we can just plug in $\lambda = \frac{2mE}{\hbar^2}$.

Because we have particles with spin $s = 1/2$, the "density of orbitals" is

$$\mathcal{D}(E) = (2s + 1)\mathcal{D}_0(E) = 2\mathcal{D}_0(E). \tag{11.8}$$

With this definition, we can write

$$\langle N \rangle = \int_0^\infty dE \mathcal{D}(E) f_{\mathrm{FD}}(E, \tau, \mu). \tag{11.9}$$

We can interpret $\mathcal{D}(E) f_{\mathrm{FD}}(E, \tau, \mu)$ as the density of occupied orbitals.

To compute the expectation value of the energy, we apply (9.28):

$$\left(\frac{\mu}{\beta} \frac{\partial}{\partial \mu} - \frac{\partial}{\partial \beta} \right) \log \mathcal{Z}_{\mathbf{n}} = \left(\frac{\mu}{\beta} \frac{\partial}{\partial \mu} - \frac{\partial}{\partial \beta} \right) \log(1 + e^{\beta(\mu - E_n)})$$

$$= \frac{E_n e^{\beta(\mu - E_n)}}{1 + e^{\beta(\mu - E_n)}} = E_n f_{\mathrm{FD}}(E_n, \tau, \mu). \tag{11.10}$$

Thus, we have

$$U = 2 \sum_{\mathbf{n}} E_n f(E_n, \tau, \mu)$$

$$\approx \int_0^\infty dE \mathcal{D}(E) E f_{\mathrm{FD}}(E, \tau, \mu), \tag{11.11}$$

where in the second line, we have made the big-box approximation.

11.2 Zero temperature

Fermi gases have a characteristic energy scale called the Fermi energy (defined below). We would often like to understand the physics of Fermi gases at temperatures much lower than the Fermi energy. To do this, we can perform perturbation theory around zero temperature. The starting point is to understand $\tau = 0$.

In the limit of zero temperature, the Fermi-Dirac distribution becomes a step function

$$\lim_{\tau \to 0} f(E, \tau, \mu) = \theta(\mu - E) = \begin{cases} 1 & \text{if } E < \mu \\ 0 & \text{if } E > \mu \end{cases}, \tag{11.12}$$

where $\theta(x)$ is the Heaviside step function (https://en.wikipedia.org/wiki/Heaviside_step_function).

Definition (Fermi energy). The *Fermi energy* is defined as the chemical potential at zero temperature: $E_F = \mu(\tau = 0)$.

Thus, at zero temperature, we have

$$N = \int_0^{E_F} dE\, \mathcal{D}(E) = \frac{2}{3} E_F \mathcal{D}(E_F) = \frac{V}{3\pi^2}\left(\frac{2mE_F}{\hbar^2}\right)^{3/2}$$

$$E_F = \frac{\hbar^2}{2m}\left(3\pi^2 n\right)^{2/3}, \tag{11.13}$$

where we have used that $\mathcal{D}(E) \propto E^{1/2}$. This gives a connection between the Fermi energy and the concentration.

The total energy at zero temperature is

$$U = \int_0^{E_F} dE\, E\mathcal{D}(E) = \frac{2}{5} E_F^2 \mathcal{D}(E_F) = \frac{3}{5} E_F N, \tag{11.14}$$

where we have used that $E\mathcal{D}(E) \propto E^{3/2}$. This is a striking result. The energy at zero temperature is enormous when N is large or the concentration n is large, coming from the fact that the Pauli exclusion principle forces orbitals with high energy to be occupied.

Because $E_F \propto n^{2/3} \sim V^{-2/3}$, the energy decreases with decreasing volume. This leads to an outward pressure sometimes called *degeneracy pressure*. The pressure at zero temperature is given by

$$\begin{aligned} p &= -\left(\frac{\partial F}{\partial V}\right)_{\tau,N}\Bigg|_{\tau=0} = -\left(\frac{\partial (U - \tau\sigma)}{\partial V}\right)_{\tau,N}\Bigg|_{\tau=0} = -\left(\frac{\partial U}{\partial V}\right)_{\tau,N} \\ &= -\left(-\frac{2}{3}\right)\frac{U}{V} \\ &= \frac{2}{5} E_F n. \end{aligned} \tag{11.15}$$

In metals, the degeneracy pressure is cancelled in metals by Coulomb attraction between electrons and ions, and in white dwarf stars by gravitational attraction.

11.3   Heat capacity at low temperature

For electrons in the conduction band of a metal, typical Fermi energies are $E_F/k_B = 50000K$, so the regime of low temperatures $\tau \ll E_F$ is of practical interest. For concreteness, consider the energy U as a function of temperature. When we turn on the temperature, the Fermi-Dirac distribution $f(E, \tau, \mu)$ is no longer a step function. However, when τ is small, $f(E, \tau, \mu)$ only changes significantly near the Fermi energy $E \sim E_F$. Thus, there will be some reshuffling of particles between energy levels near E_F, but not elsewhere.

We have to be a bit careful because we must let μ depend on temperature in order to keep N constant. Let us study the condition that N is constant

$$N = \int_0^\infty dE \mathcal{D}(E) f(E, \tau, \mu)$$
$$0 = \int_0^\infty dE \mathcal{D}(E) \left(\frac{\partial f}{\partial \tau} + \frac{\partial \mu}{\partial \tau} \frac{\partial f}{\partial \mu} \right). \tag{11.16}$$

Here, we encounter two important functions:

$$\frac{\partial f}{\partial \mu} = \frac{1}{\tau} \frac{e^{(E-\mu)/\tau}}{(e^{(E-\mu)/\tau} + 1)^2} = \frac{1}{\tau} \frac{1}{4 \cosh^2 \left(\frac{E-\mu}{2\tau} \right)},$$
$$\frac{\partial f}{\partial \tau} = \frac{E - \mu}{\tau^2} \frac{e^{(E-\mu)/\tau}}{(e^{(E-\mu)/\tau} + 1)^2} = \frac{E - \mu}{\tau^2} \frac{1}{4 \cosh^2 \left(\frac{E-\mu}{2\tau} \right)}. \tag{11.17}$$

Note that μ just shifts over the Fermi-dirac distribution, so $\frac{\partial f}{\partial \mu}$ looks like the derivative of a step function, which is a delta function $\delta(E - \mu)$. Meanwhile, $\frac{\partial f}{\partial \tau}$ looks like two closeby bumps: one negative just below μ and one positive just above μ. This should remind you of the derivative of a delta function $\delta'(E - \mu)$.[23]

Let us turn this into a quantitative approximation. Let us guess

$$\frac{\partial f}{\partial \mu} \approx A \, \delta(E - \mu), \qquad (\tau \ll \mu, E),$$
$$\frac{\partial f}{\partial \tau} \approx B \, \delta'(E - \mu), \qquad (\tau \ll \mu, E), \tag{11.21}$$

We can determine the coefficients A, B by integrating both sides against 1 and $E - \mu$. Note that we can integrate from $-\infty$ to ∞ instead of from 0 to ∞ because

---

[23]The dirac delta function is defined by

$$\int_{-\infty}^\infty \delta(x - a) f(x) = f(a). \tag{11.18}$$

Technically speaking, the delta function is not really a function, but a *distribution* — something that can be integrated against a function to get a number. Derivatives of delta functions are similar: to define them we must explain how to integrate them against a function. Specifically, they are defined by integrating by parts

$$\int_{-\infty}^\infty \delta'(x - a) f(x) = - \int_{-\infty}^\infty \delta(x - a) \frac{\partial}{\partial x} f(x) = -f'(a). \tag{11.19}$$

Similarly, we can compute an integral against the n-th derivative $\delta^{(n)}(x - a)$ by integrating by parts n times:

$$\int_{-\infty}^\infty \delta^{(n)}(x - a) f(x) = (-1)^n f^{(n)}(a). \tag{11.20}$$

the functions are very small when $E \ll \mu$, and we will assume that μ is large. For example,

$$\int_{-\infty}^{\infty} dE \frac{\partial f}{\partial \mu} = \int_{-\infty}^{\infty} dE \left(-\frac{\partial f}{\partial E} \right) = f(-\infty) - f(\infty) = 1. \qquad (11.22)$$

(In this case, we didn't actually have to do an integral.) Comparing to $\int dE\, \delta(E - \mu) = 1$, we conclude

$$\frac{\partial f}{\partial \mu} \approx \delta(E - \mu), \qquad (\tau \ll \mu, E). \qquad (11.23)$$

In the case of $\frac{\partial f}{\partial \tau}$, to compute the coefficient B, we should integrate against $E - \mu$:[24]

$$\int_{-\infty}^{\infty} dE(E - \mu) \frac{\partial f}{\partial \tau} = \int_{-\infty}^{\infty} dE \frac{(E - \mu)^2}{\tau^2} \frac{e^{(E-\mu)/\tau}}{(e^{(E-\mu)/\tau} + 1)^2}$$

$$= \tau \int_{-\infty}^{\infty} dx \frac{x^2 e^x}{(e^x + 1)^2} = \frac{\pi^2 \tau}{3}. \qquad (11.25)$$

comparing to

$$\int dE(E - \mu)\delta'(E - \mu) = -\int dE \frac{\partial(E - \mu)}{\partial E} \delta(E) = -1, \qquad (11.26)$$

we conclude

$$\frac{\partial f}{\partial \tau} \approx -\frac{\pi^2 \tau}{3} \delta'(E - \mu) \qquad (\tau \ll \mu, E). \qquad (11.27)$$

These expansions in terms of δ-functions and their derivatives are extremely useful. They are the first terms in more general series expansions at small τ, e.g.

$$\frac{\partial f}{\partial \mu} = \delta(E - \mu) + a_2 \tau^2 \delta''(E - \mu) + \dots$$

$$\frac{\partial f}{\partial \tau} = -\frac{\pi^2 \tau}{3} \delta'(E - \mu) + b_3 \tau^3 \delta'''(E - \mu) + \dots. \qquad (11.28)$$

---

[24]To evaluate the integral, we first compute

$$I(a) = \int_0^{\infty} dx \frac{x}{e^{ax} + 1} = \int_0^{\infty} dx\, x \left(e^{-ax} - e^{-2ax} + e^{-3ax} - e^{-4ax} + \dots \right)$$

$$= \int_0^{\infty} dx\, x \left(e^{-ax} + e^{-2ax} + \dots \right) - 2\int_0^{\infty} dx\, x \left(e^{-2ax} + e^{-4ax} + \dots \right)$$

$$= \frac{1}{a^2} \zeta(2) - \frac{2}{(2a)^2} \zeta(2) = \frac{\pi^2}{12a^2}. \qquad (11.24)$$

The desired answer is $-2I'(1)$.

The coefficients $a_1, a_2, \ldots$ and $b_3, \ldots$ can be computed by integrating both sides against $(E - \mu)^n$ for various n. The additional terms would be necessary to compute further corrections in τ to the formulas that we will derive next.

Note that this kind of expansion is only good if $\frac{\partial f}{\partial \tau}$ is integrated against another function that is smooth at $E = \mu$. This is the case for $\mathcal{D}(E)$ above. However, we sometimes encounter integrals like $\int dE g(E) \frac{\partial f}{\partial \tau}$, where $g(E)$ is not smooth at $E = \mu$. One example is graphene (which you may encounter on your problem set). In those cases, we cannot separately expand $\frac{\partial f}{\partial \tau}$ — we must consider the complete product $g(E) \frac{\partial f}{\partial \tau}$.

We can now return to (11.16) and plug in our approximations for $\frac{\partial f}{\partial \tau}$ and $\frac{\partial f}{\partial \mu}$. Because we are working near $\tau = 0$, we can also substitute $\mu = E_F$ in the δ functions. We find

$$
\begin{aligned}
0 &= \int_0^\infty dE \mathcal{D}(E) \left(-\frac{\pi^2 \tau}{3} \delta'(E - E_F) + \frac{\partial \mu}{\partial \tau} \delta(E - E_F) \right) \\
&= \frac{\pi^2 \tau}{3} \mathcal{D}'(E_F) + \frac{\partial \mu}{\partial \tau} \mathcal{D}(E_F).
\end{aligned}
\tag{11.29}
$$

This determines the leading correction to μ at small τ:

$$
\frac{\partial \mu}{\partial \tau} = -\frac{\pi^2 \tau}{3} \frac{\mathcal{D}'(E_F)}{\mathcal{D}(E_F)} = -\frac{\pi^2 \tau}{6 E_F}
$$

$$
\mu(\tau) = E_F - \frac{\pi^2 \tau^2}{12 E_F} \qquad (\tau \ll E_F).
\tag{11.30}
$$

Let us now study the energy at small τ. We have

$$
\begin{aligned}
C_V = \frac{\partial U}{\partial \tau} &= \int_0^\infty dE\, E \mathcal{D}(E) \left(\frac{\partial f}{\partial \tau} + \frac{\partial \mu}{\partial \tau} \frac{\partial f}{\partial \mu} \right) \\
&= \int_0^\infty dE\, E \mathcal{D}(E) \left(-\frac{\pi^2 \tau}{3} \delta'(E - E_F) + \frac{\partial \mu}{\partial \tau} \delta(E - E_F) \right) \\
&= \frac{\pi^2 \tau}{3} \left(\mathcal{D}(E_F) + E_F \mathcal{D}'(E_F) \right) + \frac{\partial \mu}{\partial \tau} E_F \mathcal{D}(E_F) \\
&= \frac{\pi^2 \tau}{3} \mathcal{D}(E_F) = \frac{\pi^2}{2} \frac{N \tau}{E_F}.
\end{aligned}
\tag{11.31}
$$

where we cancelled the term involving $\mathcal{D}'(E_F)$ using the first line of (11.30).

Note that the heat capacity is proportional to τ: it is small at small temperature, which is another striking prediction. (Compare to an ideal gas in the classical regime whose heat capacity is independent of τ.) The physical interpretation is that when we increase τ, particles in orbitals with energy $E_F - \tau$ get lifted to have energy $E_F + \tau$. The change in energy of each particle is proportional to τ, but the number of particles whose orbitals are changing is also proportional to τ. Thus, the total energy changes quadratically with τ, and the heat capacity is linear in τ.

Example 11.1 (Metals). The heat capacity of many metals can be written as a sum of an electronic contribution from a free electron gas and a contribution from phonons. The former goes like τ, and the latter goes like τ^3 according to Debye theory. Overall, we have

$$C_V^{\text{metal}} = \gamma\tau + A\tau^3 \tag{11.32}$$

which fits the data extremely well. In more detail, we expect the form

$$C_V^{\text{metal}} = \frac{\pi^2 N}{2E_F}\tau + \frac{12\pi^4}{5}\frac{N}{\theta^3}\tau^3 \tag{11.33}$$

At low temperatures, the free electron term will be dominant. The crossover temperature occurs at

$$\tau = \left(\frac{5\theta^3}{24\pi^2 E_F}\right)^{1/2}, \tag{11.34}$$

which in metals is about $1\,\text{K}$.

In practice, the heat capacity in metals is linear at low temperatures. However, the slope γ does not precisely agree with the prediction (11.33). For example, in potassium, the observed slope γ satisfies $\gamma/\gamma_0 \approx 1.23$, where $\gamma_0 = \frac{\pi^2 N}{2E_F}$ is the ideal gas prediction. The reason is that electrons in a metal are not really free — they experience Coulomb interactions with each other and with the lattice of ions.

In fact, it is initially surprising that the linear slope prediction works at all. The reason is that the important excitations in a metal aren't really individual electrons, but "quasiparticles" — an electron together with accompanying deformations and vibrations of the lattice of ions (called "dressing"). The dressing changes the way electrons interact with the lattice and with each other. Firstly, it tends to screen Coulomb interactions, so that the dressed electrons interact more weakly with each other and the lattice than undressed electrons. Secondly, it changes the effective mass of the excitations $m \to m_*$, so that the Fermi energy of dressed electrons E_{F*} is different.

This is an example of the phenomenon of "renormalization." We are given a microscopic theory of particles with some masses and interactions. Interactions dress the particles, so that the quasiparticles relevant for long-distance dynamics have different properties from their microscopic constituents. Thus, effective masses and interactions change with distance scale. This phenomenon is ubiquitous in particle physics — in fact, many of the so-called "fundamental" particles might be quasiparticles of some underlying more fundamental theory. Even the fine-structure constant α, which controls the strength of electromagnetic interactions, changes with distance scale.

[End of lecture 14]

Example 11.2 (White dwarf stars). White dwarf stars have masses approximately equal to the mass of the sun, but radii 100 times smaller. The density of the sun is approximately 1g/cm^3. The density of a white dwarf star is 10^4 to 10^7g/cm^3. All the atoms in a white dwarf star are ionized and the electrons form an approximately free electron gas.

Typical interatomic spacings are 1 Bohr radius, or 10^{-8}cm. When the atoms are 100 times closer together, their typical spacing is 10^{-10}cm, so the concentration of electrons is approximately $n = 10^{30}$ electrons/cm^3. The corresponding Fermi energy is[25]

$$\frac{\hbar^2}{2m}(3\pi^2 n)^{2/3} \approx 3 \times 10^5 \,\text{eV} = 0.36\,\text{MeV}. \tag{11.35}$$

The Fermi temperature is 10^9K. The actual temperature inside a white dwarf is expected to be on the order of 10^7K, thus we are in the same highly quantum regime that we just analyzed.

One important point is that the Fermi energy is comparable to the rest mass of an electron, which is $m_e c^2 = 0.5\,\text{MeV}$. Thus, relativistic effects are nontrivial, but don't completely invalidate our analysis.

What determines the size of a white dwarf star? Very roughly, the energy due to gravitational potential energy is

$$U_{\text{grav}} = -\frac{GM^2}{r}. \tag{11.36}$$

The kinetic energy of the Fermi gas is

$$U_{\text{fermi}} = \frac{3}{5}E_F N \approx N\frac{\hbar^2}{m_e}\left(\frac{N}{r^3}\right)^{2/3} = \frac{\hbar^2}{m_e}\frac{N^{5/3}}{r^2}. \tag{11.37}$$

We have $N = M/m_p$, where m_p is the proton mass. Minimizing the energy, we find

$$0 = \frac{\partial}{\partial r}(U_{\text{grav}} + U_{\text{fermi}}) = \frac{GM^2}{r^2} - \frac{\hbar^2}{m_e}\frac{M^{5/3}}{m_p^{5/3}r^3}, \tag{11.38}$$

which gives

$$r \approx \frac{\hbar^2}{Gm_e m_p^{5/3} M^{1/3}}. \tag{11.39}$$

So more massive white dwarf stars are actually smaller. Plugging in a solar mass $M_\odot \approx 10^{33}$g, we find

$$r(M_\odot) \approx 10^4 \,\text{km}, \tag{11.40}$$

roughly the size of the earth.

---

[25]I often use google for these types of calculations: https://www.google.com/search?q=hbar%5E2%2F(2*electron+mass)*(3+pi%5E2+*+10%5E30%2Fcm%5E3)%5E(2%2F3)+in+MeV

Example 11.3 (Relativistic white dwarf stars and the Chandrasekhar limit). When a white dwarf star more massive, the Fermi energy can get larger than the electron rest mass and relativistic effects become more important. When the star gets very massive, we can approximate the electron mass as a relativistic degenerate Fermi gas. Recall that the energy of a relativistic particle is $E = pc = \frac{\hbar c \pi n}{L}$. To determine the Fermi energy, we have

$$N = 2\frac{1}{8} \int_0^{n_F} 4\pi n^2 dn$$

$$= \pi \left(\frac{L}{\hbar c \pi}\right)^3 \int_0^{E_F} dE E^2 = \frac{\pi}{3(\hbar c \pi)^3} V E_F^3 \tag{11.41}$$

The kinetic energy of the gas is

$$U = \pi \left(\frac{L}{\hbar c \pi}\right)^3 \int_0^{E_F} dE E^3 = \frac{3N}{4} E_F = (\hbar c \pi)\frac{3N}{4}\left(\frac{3N}{\pi V}\right)^{1/3} \sim \frac{\hbar c N^{4/3}}{V^{1/3}} \sim \frac{\hbar c N^{4/3}}{r} \tag{11.42}$$

This depends on $1/r$ in the same way that gravitational potential energy does. Thus, if the relativistic dwarf star is heavy enough, gravitational potential energy will dominate kinetic energy and the star will collapse. The condition for collapse is

$$GM^2 \approx \hbar c N^{4/3} = \hbar c (M/m_p)^{4/3}$$

$$M_{\text{Chandrasekhar}} \approx \frac{M_{\text{Pl}}^3}{m_p^2}, \tag{11.43}$$

where $M_{\text{Pl}} = \left(\frac{\hbar c}{G}\right)^{1/2}$ is the Planck mass (approximately 2×10^{-5}g). Plugging in numbers, this gives about 10^{30}kg, or about 1 solar mass $M_\odot$. The current accepted value, based on a more careful analysis, for $M_{\text{Chandrasekhar}}$ is $1.4 M_\odot$.

Example 11.4 (Neutron star). When a relativistic white dwarf star collapses, what does it turn into? A way for the star to decrease its energy is to reduce the number of electrons, which it can do via the "electron capture" reaction

$$p + e \to n + \nu_e, \tag{11.44}$$

where ν_e is a neutrino. Under normal conditions, this reaction is (thankfully) not energetically favorable because $m_p + m_e < m_n + m_{\nu_e}$. However, the high temperature of a collapsing dwarf star is enough to overcome the activation energy and the reaction occurs, creating a neutron star. In a neutron star, the gravitational energy is balanced by the kinetic energy of a degenerate neutron gas. The Fermi energy of a neutron gas is set by $m_n \approx 2 \times 10^3 m_e$, so from (11.39) the neutron star is about 10^3 times smaller than a white dwarf, or a few kilometers. The flux of neutrinos from the

core of a newly-forming neutron star blasts away a shell of other collapsing matter, creating a supernova. Most of the energy is released in the form of neutrinos.[26]

The Schwarzchild radius of a black hole is

$$r_S = \frac{2GM}{c^2}, \tag{11.45}$$

which for $M = M_\odot$ is about 3 kilometers. Thus, a solar-mass neutron star is close to being a black hole. Indeed, if the neutron star has mass about $3M_\odot$, it will collapse into a black hole.

---

[26]The Wikipedia article on neutron stars (https://en.wikipedia.org/wiki/Neutron_star) is pretty mindblowing. One sentence I like is: "A neutron star is so dense that one teaspoon (5 milliliters) of its material would have a mass over 5.5×10^{12} kg, about 900 times the mass of the Great Pyramid of Giza."

12   Bose Gases and Bose-Einstein Condensation

Bose gases behave very differently from Fermi gases in the degenerate regime. Recall that it is possible for multiple bosons to occupy the same orbital. In fact, a spectacular thing happens in Bose gases at low temperatures: many particles accumulate in the lowest-energy orbital (the "ground orbital").[27] This phenomenon is called *Bose-Einstein condensation*. It is not surprising that this might happen at temperatures that are extremely low compared to the energy spacing ΔE between the lowest orbital and the first excited orbital. However, the surprising fact is that Bose-Einstein condensation happens even for temperatures much larger than ΔE.

As an example, the spacing between the lowest and first excited orbital for atoms in a cubical box with side length L is

$$
\begin{aligned}
\Delta E &= \frac{\hbar^2}{2m}\left(\frac{\pi}{L}\right)^2\left(\mathbf{n}_{\text{excited}}^2 - \mathbf{n}_{\text{ground}}^2\right) \\
&= \frac{\hbar^2}{2m}\left(\frac{\pi}{L}\right)^2\left(4+1+1-(1+1+1)\right) \\
&= 3\frac{\hbar^2}{2m}\left(\frac{\pi}{L}\right)^2.
\end{aligned}
\tag{12.1}
$$

The mass of ^{4}He is $4\,\text{GeV}/c^2$ (4 times the mass of a nucleon). For a box of size $L = 1$cm, we have

$$
\Delta E \approx 10^{-18}\,\text{eV},
\tag{12.2}
$$

which is a temperature of about $\Delta E/k_B \approx 10^{-14}\,\text{K}$. However, the observed temperature of Bose-Einstein condensation in such a system is about $2\,\text{K}$.

12.1   Chemical potential near $\tau = 0$

Naively, the Boltzmann factors $e^{-E/\tau}$ for the lowest several energy levels would be nearly the same at a temperature of $\tau = 2\,\text{K}$. The key to Bose-Einstein condensation is that the chemical potential μ can be extremely small compared to ΔE, even when τ is not small. The smallness of μ makes the physics of the ground orbital very different from the physics of the excited orbitals.

Consider a Bose gas of N particles. For simplicity, we will consider spin-0 bosons. Recall that the expected occupancy of an orbital with energy E is

$$
f(E,\mu,\tau) = \frac{1}{e^{(E-\mu)/\tau}-1},
\tag{12.3}
$$

---

[27]People usually say that particles accumulate in the ground state, where here "state" means a solution to the single-particle Schrodinger equation, i.e. what we have been calling an "orbital."

where in this section f stands for the Bose-Einstein distribution. The occupancy of the ground orbital $E = 0$ is

$$f(0, \mu, \tau) = \frac{1}{e^{-\mu/\tau} - 1}. \tag{12.4}$$

Let us begin by calculating the chemical potential at very small $\tau \ll \Delta E$. As we emphasized above, $\tau \ll \Delta E$ is *not necessary* for Bose-Einstein condensation. However, it will help us understand how small μ is. When $\tau \ll \Delta E$, all N particles are in the ground orbital, and we have

$$\lim_{\tau \to 0} f(0, \mu, \tau) \approx N = \frac{1}{e^{-\mu/\tau} - 1} \qquad (\tau \ll \Delta E).$$

$$\mu = -\tau \log \left(1 + \frac{1}{N} \right) \approx -\frac{\tau}{N} \qquad (\tau \ll \Delta E, 1 \ll N). \tag{12.5}$$

For example, if we have 10^{22} particles, then the chemical potential when $\tau \ll \Delta E$ is $-\tau/10^{22}$.

12.2   Orbital occupancy versus temperature

The fact that the chemical potential is so much closer to the energy of the lowest orbital means that the physics of that orbital is very different from the physics of the other orbitals. Let us study the occupancy of various orbitals as a function of temperature τ.

Our central claim is that when τ is larger than ΔE (but not too large) we can still have an $O(1)$ fraction of the particles in the ground orbital. This is because μ remains extremely small even as we increase τ. To see this, let us assume that an $O(1)$ fraction of the particles are in the ground orbital. Using this assumption, we will solve for μ and verify that our assumption is consistent for some range of τ.

Let us write

$$N = N_0(\tau) + N_e(\tau), \tag{12.6}$$

where $N_0 = f(0, \mu, \tau)$ is the number of particles in the ground orbital (which we assume to be very large). As before, we can solve for μ

$$N_0(\tau) = \frac{1}{e^{-\mu/\tau} - 1} \implies \mu = -\frac{\tau}{N_0(\tau)}, \tag{12.7}$$

where we used our assumption that $N_0(\tau)$ is large. If an $O(1)$ fraction of particles are in the ground orbital, μ is still ridiculously small. For example, we will later find that at temperature $\tau = k_B \cdot 1.5K = 10^{-3}$eV, half of the particles in a gas of ^{4}He are in the ground orbital. In this case, we have $\mu = -10^{-25}$eV, which is much smaller than the spacing $\Delta E = 10^{-18}$ eV we computed earlier.

In general, for a box of size L, if we increase L while holding the concentration fixed, then the number of particles scales like L^3, so that $\mu \sim 1/L^3$. Meanwhile, the energy spacing ΔE scales like $1/L^2$. Thus, in a sufficiently large box, we can ensure $\mu \ll \Delta E$.

The number of particles in excited orbitals is

$$N_e(\tau) = \sum_{\mathbf{n} \neq 0} f(E_n, \mu, \tau) = \sum_{\mathbf{n} \neq 0} \frac{1}{e^{(E_n - \mu)/\tau} - 1}, \tag{12.8}$$

where

$$E_n = \frac{\hbar^2}{2m} \left(\frac{\pi n}{L} \right)^2 - E_{\text{ground}}. \tag{12.9}$$

Here, we have shifted the energy by $E_{\text{ground}} = 3 \frac{\hbar^2}{2m} \frac{\pi^2}{L^2}$ so that the ground orbital has energy 0. For each excited orbital, we have $E_n \gg \mu$, so we can safely replace the above sum with

$$N_e(\tau) = \sum_{\mathbf{n} \neq 0} f(E_n, 0, \tau) = \sum_{\mathbf{n} \neq 0} \frac{1}{e^{E_n/\tau} - 1} \tag{12.10}$$

We can now analyze this sum using our usual tools. When $L \gg \frac{\hbar}{\sqrt{m\tau}}$, we can approximate the sum as an integral. The density of orbitals is half of what we computed before, because now we have spin-0 particles

$$\mathcal{D}_B(E) \equiv \frac{V}{4\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} E^{1/2}. \tag{12.11}$$

We use the subscript "B" to indicate spin-0 bosons.

The combination $\mathcal{D}_B(E) f(E, \mu, \tau)$ behaves like $E^{-1/2}$ at small E. Even though this function has a singularity at $E = 0$, it is an *integrable* singularity near $E = 0$ — the area under the curve $E^{-1/2}$ is finite. This guarantees that the sum $N_e(\tau)$ can be approximated by an integral

$$N_e(\tau) \approx \int_0^\infty \mathcal{D}_B(E) f(E, \mu, \tau). \tag{12.12}$$

There are some confusing things about this integral: it is supposed to represent only the contribution of excited states, and yet the integral goes from 0 to ∞. Doesn't this include the ground state? And if so, aren't we double-counting the ground state?

Example 12.1. The answer is no, we are not double-counting. To see why let us study the following toy model of the sum over energy levels:

$$S_{0,\epsilon}(a) = \frac{1}{a} \sum_{n=0}^{\infty} \frac{(n/a + \epsilon)^{1/2}}{e^{n/a+\epsilon} - 1}, \tag{12.13}$$

in the limit $\epsilon \ll 1/a \ll 1$. The important feature of this toy model is that the small quantity ϵ is only important in the first term $n = 0$. Thus, we will have to treat this term differently from the others. We will see that the other terms can be approximated as an integral from $n = 0$ to ∞, and this integral *does not double-count the $n = 0$ term.*

To begin, consider the following simplified version of the sum over the "excited" levels from $n = 1$ to ∞,

$$S_1(a) = \frac{1}{a} \sum_{n=1}^{\infty} \frac{(n/a)^{1/2}}{e^{n/a} - 1} \tag{12.14}$$

The sum is the area under the red curve in the following picture (here, we have $a = 10$)
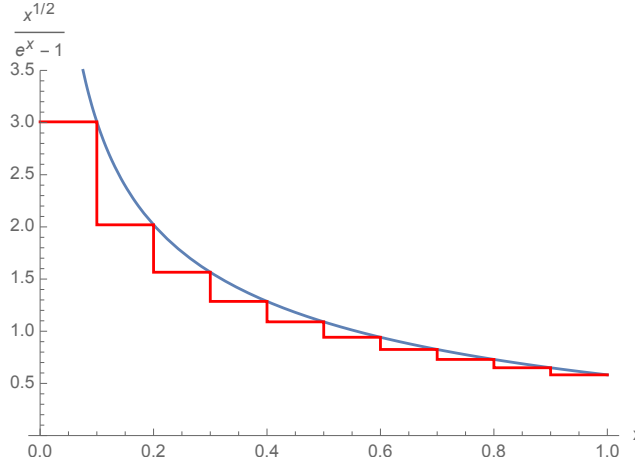
Meanwhile, the area under the blue curve is

$$I = \int_0^{\infty} \frac{dn}{a} \frac{(n/a)^{1/2}}{e^{n/a} - 1} = \int_0^{\infty} dx \frac{x^{1/2}}{e^x - 1}. \tag{12.15}$$

Because I is finite, it is hopefully clear that $S_1(a) \approx I$, with the approximation getting better for larger a. Note that the integral goes from $x = 0$ to $x = \infty$, but it approximates the terms in the sum with $n = 1, 2, \ldots$.

Now, if $\epsilon \ll 1/a$, then modifying the sum over "excited" levels to include ϵ doesn't change the answer very much

$$S_{1,\epsilon}(a) = \frac{1}{a} \sum_{n=1}^{\infty} \frac{(n/a + \epsilon)^{1/2}}{e^{n/a+\epsilon} - 1} \approx S_1(a) \approx I, \qquad (\epsilon \ll 1/a \ll 1) \tag{12.16}$$

The reason is that $n/a \gg \epsilon$ for all $n = 1, 2, \ldots$, by our assumption that $\epsilon \ll 1/a$. Thus every individual term in the sum is well-approximated by replacing $n/a+\epsilon \to \epsilon$.

Now consider the sum including the "ground" level $n = 0$. We can evaluate it by separating out the $n = 0$ term:

$$
\begin{aligned}
S_{0,\epsilon}(a) &= \frac{1}{a} \sum_{n=0}^{\infty} \frac{(n/a + \epsilon)^{1/2}}{e^{n/a+\epsilon} - 1} \\
&= \frac{1}{a} \frac{\epsilon^{1/2}}{e^{\epsilon} - 1} + S_{1,\epsilon}(a) \\
&\approx \frac{\epsilon^{-1/2}}{a} + I \qquad (\epsilon \ll 1/a \ll 1)
\end{aligned}
\tag{12.17}
$$

The sum $S_{0,\epsilon}(a)$ is analogous to the one we must do: because the ground energy is so close to a singularity of the Bose-Einstein distribution, we must treat it separately. The remaining energy levels are not too close to the singularity and they can be approximated in terms of an integral.

Let us now return to the sum (12.8). Approximating it as an integral, we have

$$
\begin{aligned}
N_e(\tau) &= \int_0^{\infty} dE \, \mathcal{D}_B(E) \frac{1}{e^{E/\tau} - 1} \\
&= \frac{V}{4\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \int_0^{\infty} dE \, \frac{E^{1/2}}{e^{E/\tau} - 1} \\
&= \frac{V}{4\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \tau^{3/2} \int_0^{\infty} dx \, \frac{x^{1/2}}{e^x - 1} \\
&= \frac{V}{4\pi^2} \left(\frac{2m\tau}{\hbar^2} \right)^{3/2} \Gamma(\tfrac{3}{2}) \zeta(\tfrac{3}{2}) \\
&= \zeta(\tfrac{3}{2}) n_Q V,
\end{aligned}
\tag{12.18}
$$

where $n_Q = \left(\frac{m\tau}{2\pi\hbar^2} \right)^{3/2}$ is the quantum concentration. Above, we have used $\Gamma(\tfrac{3}{2}) = \sqrt{\pi}/2$. Numerically, $\zeta(\tfrac{3}{2}) \approx 2.61238$. Note that the shift $E_n \to E_n - E_{\text{ground}}$ is not important in the integral approximation.

[End of lecture 15]

Thus, the fraction of particles in excited orbitals is

$$
\frac{n_e}{n} = \zeta(\tfrac{3}{2}) \frac{n_Q}{n},
\tag{12.19}
$$

where $n = N/V$. Because particle number is conserved, the number of particles in the ground orbital is

$$
\frac{n_0}{n} = 1 - \zeta(\tfrac{3}{2}) \frac{n_Q}{n} = -\frac{\tau}{N\mu}.
\tag{12.20}
$$

This finally determines the temperature dependence of μ (for $\tau < \tau_E$ — see below).

The Einstein condensation temperature τ_E is the temperature above which all particles are in excited orbitals. It is given by

$$\zeta(\tfrac{3}{2})\frac{n_Q(\tau_E)}{n} = 1 \qquad \implies \qquad \tau_E = \frac{2\pi\hbar^2}{m}\left(\frac{n}{\zeta(\tfrac{3}{2})}\right)^{2/3}. \qquad (12.21)$$

Plugging this back in, we can write

$$\frac{n_e}{n} = \left(\frac{\tau}{\tau_E}\right)^{3/2} \qquad (12.22)$$

Above this temperature, we can no longer use (12.20) to determine μ. Instead, we must solve the equation $N = N_e(\tau)$, taking into account μ-dependent corrections to $N_e(\tau)$ (which we ignored above because μ was so small).

The interpretation of the calculation that we just did is as follows. At a given temperature τ, the excited orbitals can only hold a certain fraction of the particles in the gas. Any particle that doesn't go in an excited orbital goes into the ground state.

12.3   Liquid ^{4}He and superfluidity

Plugging in numbers for ^{4}He, we find that for a concentration of $2 \times 10^{22}/\text{cm}^3$, the Einstein condensation temperature calculated using (12.21) is $\tau_E = 3\text{K}$. Nothing actually happens at this temperature — our discussion above is modified strongly by interactions.[28]

Instead, when $T \approx 2.17\text{K}$ (at low pressure), one finds a new phase with negligible viscosity, meaning it flows almost without resistance. Below this temperature, liquid Helium is in a two-fluid state: a superfluid, which is essentially a Bose condensate consisting of particles in the ground orbital, and a normal fluid consisting of particles in excited orbitals.

Note that ^{3}He also shows superfluidity, but at a much lower temperature $\tau \approx 10^{-3}\text{K}$. This might be surprising because ^{3}He atoms are fermions. What happens is that interactions between the fermions cause them to form "Cooper" pairs which are bosons. These Cooper pairs can then condense. The very low temperatures are necessary so that the Cooper pairs are sufficiently tightly bound.

Superfluidity is not implied by Bose-Einstein condensation, but it is related. Its appearance depends on interactions, so it goes beyond what we can analyze in this lecture. However, we can give some idea of how superfluidity works.

Consider a heavy object with mass M moving through a liquid. The object may excite the liquid, and if it does, that will manifest as resistance to the object's motion. However, it can happen that conservation of energy and momentum don't

---

[28]Later, we will discuss some examples of true Bose Einstein condensation that can be described using the framework we developed above.

allow the object to excite the liquid. If this occurs, the object won't experience resistance, and the liquid will appear to have zero viscosity.

Suppose the excitations of the liquid have momentum $\mathbf{p}$ and energy $E_{\mathbf{p}}$. Then during a scattering process with the object, conservation of energy and momentum imply

$$\frac{1}{2}MV^2 = \frac{1}{2}MV'^2 + E_{\mathbf{p}}$$
$$M\mathbf{V} = M\mathbf{V}' + \mathbf{p}, \tag{12.23}$$

where $\mathbf{V}$ and $\mathbf{V}'$ are the initial and final velocity of the object. Subtracting $\mathbf{p}$ from both sides of the second equation and taking $1/2M$ of its square, we get

$$\frac{1}{2}MV^2 - \mathbf{V} \cdot \mathbf{p} + \frac{p^2}{2M} = \frac{1}{2}MV'^2. \tag{12.24}$$

Subtracting from the first equation, we find

$$\mathbf{V} \cdot \mathbf{p} - \frac{p^2}{2M} = E_{\mathbf{p}}. \tag{12.25}$$

Suppose that M is very large, so that we have

$$\mathbf{V} \cdot \mathbf{p} = E_{\mathbf{p}}. \tag{12.26}$$

Depending on the functional form of $E_{\mathbf{p}}$, it may not be possible to satisfy this equation. The dot product $\mathbf{V} \cdot \mathbf{p}$ is maximized when the velocity and momentum are aligned. In that case, we find

$$V = \frac{E_p}{p}. \tag{12.27}$$

The lowest possible V for which this equation can be satisfied (the "critical velocity") is obtained by plotting E_p vs p and finding the lowest slope of a line through the origin tangent to the curve.

If the excitations of the liquid have a free-particle dispersion relation $E_p = \frac{p^2}{2m}$, then this equation becomes $V = p/(2m)$, which always has a solution for any value of V. The curve E_p vs p is a quadratic, and the slope of the line through the origin tangent to the curve is zero.

However, the excitations of ^{4}He do not have a free-particle dispersion relation — instead E_p is a somewhat complicated function of p. At small p, the excitations of ^{4}He are longditudinal sound waves with a dispersion relation

$$E_p = v_s p, \tag{12.28}$$

where $v_s \approx 237m/s$ is the speed of sound. The actual value of the critical velocity in liquid ^{4}He is much lower, about $50m/s$. This is because the curve of E_p vs. p dips back down at larger p, corresponding to a wavelength of about 10^{-8} cm.

12.4   Dilute gas Bose-Einstein condensates

In the mid-90's, two competing groups at UC Boulder (Cornell and Weimann) and at MIT (Ketterle) created Bose-Einstein condensates (BECs) in dilute atomic gases that are well-described by the theory we have developed. Because the gases are dilute, interatomic collisions are rare, and thermodynamic properties are not strongly modified by interactions as in the case of superfluid helium.

The experiments were done with ^{87}Rb (Boulder) and ^{23}Na (MIT). These are alkali atoms which have a single valence electron. A benefit of alkali atoms is that they have nonzero magnetic moments, and thus can be held in a magnetic trap. Magnetic trapping was essential for achieving Bose condensation at very low temperatures, since otherwise the atoms would interact strongly with whatever container they were placed in.

The trapped atoms have a density of about 10^{14}cm^{-3}, for which the Einstein temperature is $\tau_E \approx 1\mu K$. The main challenge in achieving a dilute gas BEC was achieving such low temperatures, while maintaining a suitable concentration of atoms.

12.4.1   Magnetic traps

A magnetic field $\mathbf{B}$ produces a splitting of energy levels called the Zeeman effect. The energy of a magnetic dipole in the field $\mathbf{B}$ is

$$E = -\mu \cdot \mathbf{B} \tag{12.29}$$

Alkali atoms have a single valence electron with zero orbital angular momentum $\mathbf{L} = 0$, so their magnetic moment is entirely due to the spin of the valence electron

$$\mu = -g_s \mu_B \mathbf{S} \tag{12.30}$$

where $g_s \approx 2$ is the electron spin g-factor, $\mathbf{S}$ is the electron spin angular momentum, and $\mu_B = e\hbar/2m_e$ is the Bohr magneton.

An atom whose magnetic moment is aligned with $\mathbf{B}$ experiences a lower potential at high values of $|\mathbf{B}|$, and hence seeks regions of large $|\mathbf{B}|$. An atom whose magnetic moment is anti-aligned will seek regions of small $|\mathbf{B}|$. Earnshaw's theorem says that it is not possible to have a local maximum of $|\mathbf{B}|$. Thus, magnetic traps work by setting up a local minimum of $|\mathbf{B}|$. If the field is sufficiently strong, the magnetic moments of atoms moving around in the trap will adiabatically follow the orientation of $\mathbf{B}$ — i.e. high-field-seekers will remain high-field-seekers and low-field-seekers will remain low-field-seekers. The low-field-seekers will be attracted to the local minimum of $|\mathbf{B}|$, and the high-field-seekers will be expelled from the trap.

The magnetic traps used in the first BEC experiments were quadrupole traps, whose local magnetic field looks like

$$\mathbf{B} \propto x\mathbf{e}_x + y\mathbf{e}_y - 2z\mathbf{e}_z. \tag{12.31}$$

(Note that this indeed satisfies $\nabla \cdot \mathbf{B} = 0$.) This can be achieved locally by placing two finite-length solenoids end-to-end with opposite fields. Unfortunately, quadrupole traps have a zero of $\mathbf{B}$ in the middle. Atoms can experience spin-flips at this location, resulting in them being lost from the trap. The MIT team dealt with this problem by "plugging" the hole at $(x, y, z) = 0$ using a laser with frequency tuned to repel the atoms.[29] A nicer choice is the Ioffe-Pritchard trap, which has a nonzero magnetic field at its center and gives rise to a harmonic (though not isotropic) potential.

12.4.2   Laser cooling

Magnetic traps are not very strongly confining, so atoms must be cooled significantly to even remain in the trap. The technique used for this initial cooling is laser cooling or "optical molasses," which is a very cool idea.

The idea is to shine a laser through the gas of atoms at a frequency $\nu = \omega - \delta$ slightly lower than a resonant frequency ω of the atoms. That is, ω is such that $\hbar\omega = E_1 - E_2$ is a difference between two atomic energy levels, and δ is called the "detuning." Atoms moving toward the laser with velocity $\mathbf{v}$ will experience a blue-shifted frequency $\nu \to \nu - \mathbf{k} \cdot \mathbf{v} = \omega - (\delta + \mathbf{k} \cdot \mathbf{v})$, where $\mathbf{k}$ is the wave vector of the laser, satisfying $c|\mathbf{k}| = \nu$. This blue-shift will make the laser light closer to resonance, and therefore make the atom more likely to absorb a photon, receiving a momentum kick $\Delta\mathbf{p} = m\Delta\mathbf{v} = -\hbar\mathbf{k}$ opposite to its direction of motion. Once the photon is absorbed, the atom radiates it away isotropically, resulting in zero momentum kick on average. Thus, the atom is more likely to slow down than speed up during its interaction with the laser.

More precisely, the rate of absorption of photons with detuning δ is proportional to

$$R_{\text{abs}} \propto \frac{A}{\gamma^2 + (\nu - \omega)^2} = \frac{A}{\gamma^2 + \delta^2}, \tag{12.32}$$

where ω is the resonant frequency of the atom, ν is the applied frequency of light, γ is the rate of spontaneous emission of photons, and A depends on the squared amplitude of the applied light. You can think of the atom as being like a harmonic oscillator with frequency ω and damping rate 2γ. Then (12.32) is the expression for the squared amplitude of a damped-driven oscillator with driving frequency ν.

---

[29]The laser is tuned to a frequency $\omega + \delta$, where ω is a characteristic oscillation frequency of the atom and δ is a large positive "detuning." The laser creates an oscillating electric field that drives oscillations of the electric dipole moment of the atoms. The energy of an electric dipole in a field E is $-\mathbf{E} \cdot \mathbf{d}$. When δ is positive, oscillations of $\mathbf{d}$ occur π out of phase with the driving electric field, causing the atom to have higher energy in higher electric oscillating electric fields.

Thus, the force due to absorption of photons is proportional to

$$F_{\mathbf{k}} = \hbar \mathbf{k} R_{\mathrm{abs}} = \hbar \mathbf{k} \frac{A}{\gamma^2 + (\delta + \mathbf{k} \cdot \mathbf{v})^2}$$

$$\approx \frac{A \hbar \mathbf{k}}{\gamma^2 + \delta^2} \left(1 - \frac{2\delta \mathbf{k} \cdot \mathbf{v}}{\gamma^2 + \delta^2} \right). \tag{12.33}$$

where the shift of the detuning $\delta \to \delta + \mathbf{k} \cdot \mathbf{v}$ in the denominator is due to the doppler shift. In the second line, we expanded in small $v = |\mathbf{v}|$. The first term gives a universal kick in the direction of the laser light, while the second term gives the velocity-dependent cooling force. We do not need to consider the force due to spontaneous emission because it averages to zero.

A nifty trick for canceling the unwanted first term is to use two counterpropagating laser beams — one with wavevector $\mathbf{k}$ and the other with wavevector $-\mathbf{k}$. The sum of the resulting forces is

$$F_{\mathbf{k}} + F_{-\mathbf{k}} = -\frac{4A\hbar\delta\mathbf{k}(\mathbf{k} \cdot \mathbf{v})}{(\gamma^2 + \delta^2)^2}. \tag{12.34}$$

For example, if the lasers and velocity are oriented in the z-direction, the force is

$$F_z \propto -\frac{4A\hbar\delta k^2}{(\gamma^2 + \delta^2)^2} v_z, \tag{12.35}$$

which looks like a viscous drag force. For this reason, the counterpropagating lasers are called "optical molasses." The setup we just described is sufficient to slow atoms in one dimension. To slow them in three dimensions, we must apply three pairs of counterpropagating beams.

In the MIT experiment, they start with sodium atoms effusing from an oven at a temperature of 600K and a concentration of $10^{14}/\mathrm{cm}^3$. Using laser cooling, the atoms are eventually cooled to about 100μK.

12.4.3   Evaporative cooling

The next step is evaporative cooling. Conceptually, the height of the trap is lowered gradually, so atoms near the top of the potential well (i.e. those with the most energy) escape, while those deeper in the potential well do not. Atomic collisions occur often enough to keep the remaining atoms close to thermal equilibrium with each other, so that the temperature steadily decreases. In this way, the temperature is further reduced from 100μK to $T \sim T_E \sim 1\mu$K, with approximately 10^7 atoms remaining in the BEC.

The way the walls of the trap are "lowered" is really interesting. Recall that the atoms are trapped by an inhomogeneous magnetic field. They experience an effective potential

$$V_{\mathrm{eff}} = \pm \frac{1}{2} g_s \mu_B |\mathbf{B}|. \tag{12.36}$$

The source of this effective potential is the Zeeman splitting between spin states

$$\Delta E = g_s \mu_B |\mathbf{B}| \equiv \hbar \omega_{\text{Zeeman}}. \tag{12.37}$$

Atoms whose magnetic moments are anti-aligned with the magnetic field experience a plus sign in (12.36), and thus seek low values of $|\mathbf{B}|$ (and remain in the trap), while atoms whose magnetic moments are aligned seek high values of $|\mathbf{B}|$ (and get ejected).

One can induce oscillations between energy levels $\Delta E = \hbar \omega$ by applying an oscillating electric field with frequency $\nu \approx \omega$. Thus, by applying such a field with $\nu = \omega_{\text{Zeeman}}$, one can turn low-field-seeking atoms into high-field-seeking atoms and kick them out of the trap. The key trick is to match the oscillation frequency ν to the value of ω_{Zeeman} for atoms at the top of the trap. In this way, only atoms at the top of the trap get ejected.

12.4.4   BEC in a harmonic trap

The fact that the atoms are in a harmonic trap changes slightly the analysis of Bose Einstein condensation. For simplicity, let us assume that the trap is isotropic,[30] so that the energy levels for a single atom are the energy levels of three independent harmonic oscillators

$$E_{\mathbf{n}} = \hbar \omega_0 (n_x + n_y + n_z). \tag{12.38}$$

The number of particles in excited orbitals is

$$\begin{aligned} N_e &= \sum_{\mathbf{n} \neq 0} f(E_n, \mu, \tau) \\ &\approx \int_0^\infty dn_x dn_y dn_z f(\hbar \omega_0 (n_x + n_y + n_z), \mu, \tau). \end{aligned} \tag{12.39}$$

Last time, the energy depended on $n_x^2 + n_y^2 + n_z^2$, so it was useful to introduce a variable $n = \sqrt{n_x^2 + n_y^2 + n_z^2}$ representing the radial direction in mode number space. We can use the same trick here. This time, our "radial" variable is $s = n_x + n_y + n_z$. By dimensional analysis, the measure for the radial direction will be $as^2 ds$ for some constant a. We can fix the constant by requiring that $\int_0^1 as^2 ds = \frac{1}{6}$, the volume of a unit simplex. This gives $a = \frac{1}{2}$.

---

[30] For typical BEC's in the laboratory, this is not the case — the trap is typically elongated in some direction.

Overall, we get

$$
\begin{aligned}
N_e &= \frac{1}{2} \int_0^\infty s^2 ds \frac{1}{e^{(\hbar\omega_0 s - \mu)/\tau} - 1} \\
&\approx \frac{1}{2} \frac{\tau^3}{(\hbar\omega_0)^3} \int_0^\infty dx \frac{x^2}{e^x - 1} \\
&= \frac{\zeta(3)\tau^3}{(\hbar\omega_0)^3}
\end{aligned}
\tag{12.40}
$$

The Einstein condensation temperature is the temperature where $N_e = N$, i.e.

$$
\tau_E = \hbar\omega_0 \left(\frac{N}{\zeta(3)} \right)^{1/3}.
\tag{12.41}
$$

The mean squared position of an atom in an excited state in a harmonic trap is

$$
\langle x^2 \rangle_\tau = \frac{2}{m\omega_0^2} \left\langle \frac{1}{2} m\omega_0^2 x^2 \right\rangle = \frac{\tau}{m\omega_0^2},
\tag{12.42}
$$

by equipartition of energy. However, the squared width of the ground state is

$$
\langle 0 | x^2 | 0 \rangle = \frac{\hbar}{2m\omega_0}.
\tag{12.43}
$$

Thus, at temperatures comparable to τ_E, we have

$$
\frac{\langle 0 | x^2 | 0 \rangle^{1/2}}{\langle x^2 \rangle_\tau^{1/2}} \approx \left(\frac{\hbar\omega_0}{2\tau_E} \right)^{1/2} \approx N^{-1/6}.
\tag{12.44}
$$

Consequently, the BEC appears as a dense "pit" inside a cloud.

[End of lecture 16]

13   Heat and Work

13.1   Reversibility, heat, and work

Definition (reversible process). A reversible process is a process where the combined entropy of a system plus reservoir remains constant. Because the total entropy remains constant, it is possible to run the process in reverse without violating the laws of thermodynamics (in this case, without leaving the "most probable configuration" for the combined system).

Note that the entropy $\sigma_\mathcal{S}$ of the system *can* change in a reversible process — it is only the combined entropy $\sigma_\text{tot} = \sigma_\mathcal{S} + \sigma_\mathcal{R}$ that cannot change. Henceforth, we denote $\sigma_\mathcal{S}$ by σ. Achieving reversibility in practice requires e.g. removing friction, and ensuring that two objects with different temperatures never come into direct contact. If heat is only exchanged between objects whose temperatures are infinitesimally close to each other, then the heat flow can be reversed by an infinitesimal change in one of the temperatures. In this way, we can imagine running the process backwards. However, if heat flows from a hotter object to a cooler object, then we know from our discussion of thermal contact that this involves an increase in total entropy that cannot be undone.

In a reversible process, there are two types of energy exchange that we can distinguish:

Definition (heat). An exchange of energy that accompanies a change in entropy of the system $\sigma \to \sigma + d\sigma$ is called *heat* and we denote it by $đQ$. In a reversible process, the change in entropy of the reservoir must be $d\sigma_\mathcal{R} = -d\sigma$. By conservation of energy, the change in energy of the reservoir is $dU_\mathcal{R} = -đQ$. Finally, from the definition of temperature $\frac{1}{\tau} = \frac{\partial \sigma_\mathcal{R}}{\partial U_\mathcal{R}}$, we have $đQ = \tau d\sigma$.

Definition (work). An exchange of energy that comes from a change in external parameters (such as the position of a piston) is called *work* and is denoted $đW$. Work does not involve a change in entropy of the system.

The $đ$-notation $đQ$ and $đW$ is to emphasize that there do not exist functions $Q(\sigma, V)$ and $W(\sigma, V)$ of which $đQ$ and $đW$ are differentials. That is, heat and work are not functions of the state of a system — we cannot say that my car has heat 1000J or that an iceberg has work 1000J. Instead, heat and work characterize changes. By contrast σ, U, and V are functions of the state of a system — one can observe the system without knowing its history and (in principle) write down σ, U, and V. We write their differentials as $d\sigma$, dU, and dV. Don't worry if this distinction is confusing — we will return to it later.[31]

---

[31]For the mathematically inclined, the distinction is this: differentials like $d\sigma$ and dV are exact differential forms obtained by applying the exterior derivative operator d to some function of the state of a system. By contrast, "differentials" with a bar $đW$ and $đQ$ are not exact or even closed.

It is often useful to think of $\mathcal{S}$ as coupled to two different auxiliary systems: a reservoir $\mathcal{R}$ with which it can exchange heat (i.e. energy with an accompanying exchange of entropy), and another system $\mathcal{S}'$ on which it can do work but not exchange entropy. In practice, there may be no physical distinction between $\mathcal{R}$ and $\mathcal{S}'$.

By conservation of energy, we have

$$
\begin{aligned}
dU &= đW + đQ \\
&= đW + \tau d\sigma,
\end{aligned}
\tag{13.1}
$$

which says that the total change in energy is caused partly by heat exchanged with the reservoir and partly by work done on the system.

13.2   Heat engines and Carnot efficiency

All forms of work can be completely converted into each other — for example, mechanical energy can be used to generate an electrical current, and an electrical current can be converted back into mechanical energy. Meanwhile, work can be completely converted into heat (e.g. by friction), but not vice-versa. This makes work a valuable quantity. It is possible to convert some heat to work, but there are fundamental limits on how efficient this process can be.

Definition (heat engine). A heat engine is a cyclic device that converts heat into work by exploiting a difference in temperature between two reservoirs.

Here, "cyclic" means that the device periodically returns to its initial state.

Let us model such a device as a system $\mathcal{S}$ coupled to three auxiliary systems

- A reservoir $\mathcal{R}_h$ at high temperature τ_h,

- A reservoir $\mathcal{R}_l$ at low temperature τ_l,

- An auxiliary system $\mathcal{S}'$ on which $\mathcal{S}$ can do work.

Over the course of a cycle, the device takes in heat Q_h at high temperature τ_h, does some work W on $\mathcal{S}'$, and ejects heat Q_l at low temperature τ_l. By conservation of energy, we have

$$
Q_h = W + Q_l.
\tag{13.2}
$$

Let us assume the device operates reversibly. This means that the change in entropy of the system must be zero over the course of a cycle. The entropy increase accompanying the intake of Q_h is $\sigma_h = Q_h/\tau_h$, and the entropy decrease accompanying the output of Q_l is $\sigma_l = Q_l/\tau_l$. Overall, we must have

$$
\begin{aligned}
\sigma_h &= \sigma_l \\
Q_h/\tau_h &= Q_l/\tau_l.
\end{aligned}
\tag{13.3}
$$

They are more property thought of as 1-forms on configuration space.

From here, we can solve for W in terms of Q_h:

$$W = Q_h(1 - \tau_l/\tau_h) = \frac{\tau_h - \tau_l}{\tau_h}Q_h = \eta_C Q_h. \qquad (13.4)$$

The quantity η_C is called the *Carnot efficiency*. It gives the fraction of heat that can be converted into work by a heat engine operating reversibly.

Suppose now that the heat engine operates irreversibly, so that there is some entropy production inside the engine. For example, at some point the engine could allow gas to expand suddenly into a larger chamber, or friction inside the engine could directly turn work into heat. Because entropy must increase, in this case we have

$$\sigma_h \leq \sigma_l$$
$$Q_h/\tau_h \leq Q_l/\tau_l \qquad \text{(irreversible engine)} \qquad (13.5)$$

Solving for W in this case, we have

$$W = Q_h - Q_l \leq Q_h(1 - \tau_l/\tau_h) = \eta_C Q_h \qquad \text{(irreversible engine)}. \qquad (13.6)$$

Thus, heat engines operating irreversibly have a lower energy conversion efficiency $W/Q_h \leq \eta_C$.

13.3 Refrigerators

A refrigerator is a heat engine run in reverse. The refrigerator takes input heat Q_l from a reservoir at low temperature τ_l, together with input work, and outputs heat Q_h to a reservoir at temperature τ_h. In a reversible refrigerator, we again have $Q_h/\tau_h = Q_l/\tau_l$ so that total entropy is unchanged. Conservation of energy implies $W = Q_h - Q_l$. The efficiency of a refrigerator is measured by

$$\gamma = \frac{Q_l}{W} = \frac{Q_l}{Q_h - Q_l} \qquad (13.7)$$

In the case of reversible operation, this is

$$\gamma_C = \frac{\tau_l}{\tau_h - \tau_l}. \qquad (13.8)$$

For an irreversible refrigerator, we have $Q_h/\tau_h \geq Q_l/\tau_l$, so that the coefficient γ is smaller than γ_C.

An air conditioner is a refrigerator that cools the inside of a building. Note that the efficiency of an air conditioner is highest when the inside temperature τ_l is close to the outside temperature τ_h. A heat pump is an air conditioner with the input and output switched, so it uses work to heat up a building. The efficiency of a reversible heat pump is

$$\frac{Q_h}{W} = \frac{Q_h}{Q_h - Q_l} = \frac{\tau_h}{\tau_h - \tau_l} > 1, \qquad (13.9)$$

which is better than directly converting W into heat and injecting it into the room (as with a heater).

13.4   Different types of expansion

Let us give some different examples of processes involving a classical ideal gas and their relationship to reversibility.

Example 13.1 (Reversible isothermal expansion). Isothermal means the temperature is held constant. The gas is in contact with a reservoir and can exchange energy with the reservoir during expansion. As the gas expands isothermally, it does work on the walls of its container. The work done is

$$W = \int_{V_i}^{V_f} p dV = \int_{V_i}^{V_f} \frac{N\tau}{V} dV = N\tau \log(V_f/V_i). \tag{13.10}$$

The entropy is given by (10.22). The quantum concentration n_Q is constant during isothermal expansion, so the entropy of the gas increases by

$$\Delta\sigma = N \log(V_f/V_i). \tag{13.11}$$

Note that $W = \tau \Delta\sigma$. This is the integral of the thermodynamic relation

$$dU = \tau d\sigma - p dV$$
$$\Delta U = \tau \Delta\sigma - W. \tag{13.12}$$

For an ideal gas, τ is unchanged, so since $U = \frac{3}{2}N\tau$, the energy U is unchanged as well. This implies $W = \tau \Delta\sigma$.

Note that there are two sources of energy flow into/out of the gas. 1) The gas does work on its environment through a piston, meaning energy W leaves the gas. 2) Energy in the form of heat enters the gas as it expands to maintain constant temperature.

Example 13.2 (Reversible isentropic expansion). Isentropic means the entropy is held fixed during the expansion. This can be achieved by insulating the gas from a reservoir so that no heat can flow. The entropy of a monatomic gas depends on the volume and temperature as

$$\sigma(\tau, V) = N \left(\log \tau^{3/2} + \log V + \dots \right). \tag{13.13}$$

To keep the entropy constant, we must have $V\tau^{3/2} = $ const. So as the system expands, the temperature drops. The energy is proportional to the temperature, so the energy drops as well.

Because no heat can flow, we have $\Delta U = -W$. Consider a small amount of expansion $V \to V + dV$. The change in energy due to work is

$$dU = -p dV = -\frac{N\tau}{V} dV = -\frac{N}{V} \frac{2}{3} \frac{U}{N} dV = -\frac{2U}{3} \frac{dV}{V}$$
$$\frac{dU}{U} = -\frac{2}{3} \frac{dV}{V}$$
$$U = CV^{-2/3} \tag{13.14}$$

Thus, we have $\tau V^{2/3} \propto U V^{2/3} = $ const., which agrees with our direct analysis using the formula for the entropy.

Example 13.3 (Irreversible expansion into vacuum). If a gas suddenly expands, without exchanging energy with its environment, the process is irreversible. The total energy doesn't change, and so the temperature doesn't either. The entropy thus changes by $N \log(V_2/V_1)$. For example, if we double the size of the box, the entropy changes by $N \log 2$. We can think of this as adding 1 bit of information per particle, since now to specify the state of the system, we must say whether it is in the left- or right-half of the box.

13.5   The Carnot cycle

We have argued that any heat engine has an upper bound on its efficiency $\eta \leq \eta_C = \frac{\tau_h - \tau_l}{\tau_h}$ coming from the principle that entropy does not decrease. The Carnot cycle is an example of a heat engine that achieves this maximum efficiency.

In the Carnot cycle, a gas is in a box adjacent to two reservoirs at temperatures τ_l and τ_h. At any given time, we can remove an insulating barrier to couple the gas to a reservoir, or add a barrier to decouple the gas. Furthermore, the gas can do work against a piston by changing its volume. By adding/removing barriers and doing work on or extracting work from the gas, we change the macroscopic configuration of the gas in a cycle so that it comes back to its starting point.

In order to have a reversible engine, we must only couple the gas to a reservoir when the gas already has the temperature of that reservoir. This tells us that some parts of the cycle should be isothermal at temperatures τ_l and τ_h. Furthermore, when the gas is not coupled to a reservoir, we must ensure that its entropy does not change. This tells us that the other parts of the cycle should be isentropic.

Essentially the only solution is a four-stage process. Isothermal expansion, isentropic expansion, isothermal compression, isentropic compression. In the τ, σ plane, the gas traverses a rectangle between the points

$$p_1 = (\tau_h, \sigma_L), \quad p_2 = (\tau_h, \sigma_H), \quad p_3 = (\tau_l, \sigma_H), \quad p_4 = (\tau_l, \sigma_L). \tag{13.15}$$

Recall that the entropy of a gas is given by

$$\sigma = N(\log \tau^{3/2} + \log V + \text{const.}). \tag{13.16}$$

Thus, we see that $\sigma_H - \sigma_L = N \log(V_2/V_1) = N \log(V_3/V_4)$, so that we must have $V_2/V_1 = V_3/V_4$. In more detail, the process is[32]

---

[32]Following Kittel and Kroemer, we use conventions where we always associate a positive number and a direction (i.e. from the reservoir to the system) to heat and work, as opposed to using negative numbers to indicate direction. Thus, when combining different sources of heat and work, we will have to keep in mind the direction of flow to know what signs to use.

- $(1 \to 2)$ The gas starts with temperature τ_h and entropy σ_L. After being coupled to the reservoir at temperature τ_h, it undergoes isothermal expansion, doing work on the piston and also increasing its entropy to σ_H. During this stage, heat flows from the reservoir:

$$Q_h = \tau_h(\sigma_H - \sigma_L) = N\tau_h \log(V_2/V_1). \tag{13.17}$$

Recall that the internal energy of a monatomic ideal gas is given by $U = \frac{3}{2}N\tau$. Since the temperature isn't changing during isothermal expansion, the internal energy of the gas isn't changing either. Thus, all the heat Q_h goes into work done on the piston

$$W_{12} = Q_h. \tag{13.18}$$

- $(2 \to 3)$ The gas now has temperature τ_h and entropy σ_H. The reservoir is decoupled from the gas, and the gas undergoes isentropic expansion. During this expansion, the gas does more work on the piston, but its temperature drops to τ_l. The work is given by the change in internal energy of the gas, which is given by

$$W_{23} = U(\tau_h) - U(\tau_l) = \frac{3}{2}N(\tau_h - \tau_l). \tag{13.19}$$

During isentropic expansion or compression, $V\tau^{3/2}$ is constant. Thus, we have

$$V_2\tau_h^{3/2} = V_3\tau_l^{3/2}. \tag{13.20}$$

- $(3 \to 4)$ The gas now has temperature τ_l and entropy σ_H. It is coupled to the low-temperature reservoir with temperature τ_l, and then undergoes isothermal compression. During the compression, its entropy changes from σ_H back to σ_L, so that it dumps heat $Q_l = \tau_l(\sigma_H - \sigma_L)$ into the low-temperature reservoir. Because the internal energy of the gas doesn't change, this heat must come from work done by the piston.

$$W_{34} = Q_l = \tau_l(\sigma_H - \sigma_L) = N\tau_l \log(V_3/V_4) = N\tau_l \log(V_2/V_1) \tag{13.21}$$

- $(4 \to 1)$ The gas has temperature τ_l and entropy σ_L. It is decoupled from the reservoir and undergoes isentropic compression. During the compression, the temperature increases from τ_l to τ_h. To achieve the compression, the piston must do work

$$W_{41} = U(\tau_h) - U(\tau_l) = \frac{3}{2}N(\tau_h - \tau_l) \tag{13.22}$$

The total work done by the heat engine is

$$
\begin{aligned}
W_{12} + W_{23} - W_{34} - W_{41} &= W_{12} - W_{34} \\
&= (\tau_h - \tau_l)(\sigma_H - \sigma_L).
\end{aligned}
\tag{13.23}
$$

This is simply the area of the rectangle in the τ-σ plane. This simple result is a consequence of the fact that the integral of dU over a closed cycle in configuration space must vanish

$$
\oint dU = 0.
\tag{13.24}
$$

On the other hand, we have

$$
dU = đQ + đW = \tau d\sigma - p dV
\tag{13.25}
$$

Thus, we find

$$
\oint p dV = \oint \tau d\sigma = \int_{\text{rectangle}} d\tau d\sigma,
\tag{13.26}
$$

where in the last line we used Stokes theorem.

Meanwhile, the heat taken in from the high temperature reservoir is $\tau_h(\sigma_H - \sigma_L)$. Thus, the Carnot cycle exactly achieves the Carnot efficiency $\frac{W}{Q_h} = \frac{\tau_h - \tau_l}{\tau_h} = \eta_C$.

13.5.1 Path dependence of heat and work

In the Carnot cycle, we see an example of the path-dependence of heat and work. The integrals

$$
\oint đQ = \oint \tau d\sigma, \quad \oint đW = -\oint p dV
\tag{13.27}
$$

are both nonzero. This shows that Q cannot be a function of the state of the system — i.e. there cannot be a function $Q(\sigma, V)$ such that $đQ$ is the differential of that function.[33] If such a function existed, then we'd have $\oint dQ = Q(\sigma_f, V_f) - Q(\sigma_i, V_i) = 0$ for any closed cycle. The object $đQ$ is something that must be integrated along a path. The resulting heat transfer Q depends on that path — not just the initial and final points. This is why we use the funny $đ$. The regular d is reserved for something whose integral along a closed path is zero, as in (13.24).

---

[33]The variables we use to parametrize the system are not important as long as they completely characterize the macrostate. For example, there can't be a function $Q(\tau, V)$ either.

13.5.2   Universality of reversible engines

Suppose that we had two reversible engines with different efficiencies

$$\frac{W_1}{Q_{h1}} = \eta_1 > \eta_2 = \frac{W_2}{Q_{h2}}. \tag{13.28}$$

We can hook them up to each other with the lower-efficiency engine operating as a refrigerator. Specifically, we set

$$Q_{h1} = Q_{h2} = Q_h, \tag{13.29}$$

so that $W_1 = W_2 + \Delta W$. The refrigerator takes in heat

$$Q_{l2} = Q_{h2} - W_2 = Q_h - W_1 + \Delta W, \tag{13.30}$$

from the low temperature reservoir and requires work $W_1 - \Delta W$. The heat engine deposits heat

$$Q_{l1} = Q_{h1} - W_1 \tag{13.31}$$

into the low-temperature reservoir and does work W_1.

Overall, the net effect is to take heat $Q_{l2} - Q_{l1} = \Delta W$ from the low temperature reservoir and do work $W_1 - (W_1 - \Delta W) = \Delta W$. That is, we have converted heat into work, thus violating the second law of thermodynamics. Of course, the contradiction is that a reversible heat engine or refrigerator must not produce entropy, and we already showed that this fixes their efficiency to be the Carnot efficiency. Still this argument is a way of seeing that the efficiency of a reversible engine is universal without invoking entropy explicitly. Now we can exhibit the Carnot cycle, which is a reversible cycle, and deduce that all reversible engines have efficiency η_C.

14   Gibbs free energy

14.1   Heat and work at constant temperature

In a process at constant temperature, the work done on a system is

$$dW = dU - dQ = dU - \tau d\sigma = dU - d(\tau\sigma) = dF, \tag{14.1}$$

where $F = U - \tau\sigma$ is the free energy. Recall also that the condition for thermodynamic equilibrium for a system at constant temperature τ is that F is minimized. This is because the number of accessible states of a system+reservoir in thermal contact is

$$g = g_S g_R = e^{\sigma(U)} e^{\sigma_R(U_{\text{tot}} - U)} = e^{\sigma(U)} e^{\sigma_R(U_{\text{tot}}) - U/\tau} = e^{\sigma_R(U_{\text{tot}})} e^{-F/\tau}, \tag{14.2}$$

Thus, a minimum of F is a maximum of g, i.e. the "most probable configuration."

For these reasons, free energy plays a role for isothermal processes at nonzero temperature analogous to the role that energy plays for processes at zero temperature. The extra term $-\tau\sigma$ in the free energy accounts for heat exchange with the reservoir. If heat is provided by the reservoir, then we get it for free — no external work is required to make it happen. If heat is given to the reservoir, then that energy cannot be used to do work.

14.2   Heat and work at constant pressure

Many processes, particularly those open to the atmosphere, take place at constant pressure. An example is the boiling of a liquid against a piston with a constant applied force. Such a system is in thermal and *mechanical* contact with a reservoir $\mathcal{R}$ at temperature τ and pressure p.

Definition (mechanical contact). Two systems are in mechanical contact if they can exchange volume.

Recall that one of our expressions for pressure was

$$\frac{p}{\tau} = \left(\frac{\partial \sigma}{\partial V} \right)_U. \tag{14.3}$$

Thus, for the reservoir $\mathcal{R}$, we have

$$\sigma_R(U_{\text{tot}} - U, V_{\text{tot}} - V) = \sigma_R(U_{\text{tot}}, V_{\text{tot}}) - \frac{p}{\tau}V - \frac{1}{\tau}U + \dots . \tag{14.4}$$

The definition of a reservoir is that the "..." terms can be dropped (we should imagine that they are higher order in $1/N$, where N is some extensive quantity), so that the entropy of the reservoir is linear in V and U (to very good approximation).

   Mechanical contact means that in order for a system to change volume $V \to V + \delta V$, it must *steal* volume from the reservoir $V_{\mathcal{R}} \to V_{\mathcal{R}} - \delta V$. That is, volume becomes a conserved quantity in the context of mechanical contact. Consequently, the number of accessible states of the combined system and reservoir is

$$
e^{\sigma(U,V)} e^{\sigma_{\mathcal{R}}(U_{\text{tot}} - U, V_{\text{tot}} - V)} = e^{\sigma_{\mathcal{R}}(U_{\text{tot}}, V_{\text{tot}})} e^{\sigma(U,V) - \frac{p}{\tau}V - \frac{1}{\tau}U}
$$

$$
= e^{\sigma_{\mathcal{R}}(U_{\text{tot}}, V_{\text{tot}})} e^{-G/\tau} \tag{14.5}
$$

Here, we have defined a quantity called the Gibbs free energy

$$
G = F + pV = U + pV - \tau\sigma \tag{14.6}
$$

Thus, the minimum of G is the "most probable configuration" for a system and reservoir in thermal and mechanical contact.

   One of the equilibrium conditions for thermal and mechanical contact is

$$
0 = \left(\frac{\partial \sigma_{\mathcal{S}}}{\partial V}\right)_U - \left(\frac{\partial \sigma_{\mathcal{R}}}{\partial V}\right)_U, \tag{14.7}
$$

which says that the pressure of the system and reservoir must be equal.

   Recall that the free energy $F = U - \tau\sigma$ has an interpretation as the energy, where we subtract off the "useless" contribution from heat $\tau\sigma$. The Gibbs free energy has a similar interpretation. When a system is in mechanical contact with a reservoir, the an amount of energy $-pV$ goes into changing the volume of the reservoir. This energy is "useless" in the same form as heat and it is useful to subtract it off. Thus G represents the "useful" work that can be extracted from a system at constant temperature and pressure.

14.3   Thermodynamic identities for Gibbs free energy

Consider a system with variable energy U, volume V, and number of particles N. Under changes dU, dV, dN, the entropy changes by

$$
d\sigma = \left(\frac{\partial \sigma}{\partial U}\right)_{V,N} dU + \left(\frac{\partial \sigma}{\partial V}\right)_{U,V} dV + \left(\frac{\partial \sigma}{\partial N}\right) dN
$$

$$
= \frac{1}{\tau} dU + \frac{p}{\tau} dV - \frac{\mu}{\tau} dN. \tag{14.8}
$$

The corresponding change in G is

$$
dG = dU + pdV + V\,dp - \tau d\sigma - \sigma d\tau
$$

$$
= (\tau d\sigma - pdV + \mu dN) + pdV + V\,dp - \tau d\sigma - \sigma d\tau
$$

$$
= V\,dp - \sigma d\tau + \mu dN. \tag{14.9}
$$

As a result, we get the following thermodynamic identities

$$\left(\frac{\partial G}{\partial N}\right)_{\tau,p} = \mu$$

$$\left(\frac{\partial G}{\partial \tau}\right)_{N,p} = -\sigma$$

$$\left(\frac{\partial G}{\partial p}\right)_{\tau,N} = V. \tag{14.10}$$

14.4   Gibbs free energy and chemical potential

Note that p and τ are intensive quantities: they do not change when two identical systems with the same p and τ are put together. By contrast, U, V, σ, F, G, N are extensive quantities: they are additive under putting two systems together.

If we think of G as a function of τ, p, N, then N and G are the only extensive quantities, so they must be proportional:

$$G = N\phi(\tau, p), \tag{14.11}$$

where ϕ is some function of the intensive quantities that is independent of N. However, we also saw that $\left(\frac{\partial G}{\partial N}\right)_{\tau,p} = \mu$, which implies $\phi(\tau, p) = \mu(\tau, p)$, so that we can write

$$G = N\mu(\tau, p). \tag{14.12}$$

Example 14.1 (Gibbs free energy of ideal gas). Recall that the entropy of an ideal gas is

$$\sigma = N\left(\log \frac{n_Q}{n} + \frac{5}{2}\right) = N\left(\log \frac{V n_Q}{N} + \frac{5}{2}\right). \tag{14.13}$$

The free energy is

$$F(\tau, V, N) = U - \tau\sigma = \frac{3}{2}N\tau - N\left(\log \frac{V n_Q}{N} + \frac{5}{2}\right)$$

$$= N\tau\left(\log \frac{N}{V n_Q} - 1\right). \tag{14.14}$$

We previously found the formula

$$\mu = \left(\frac{\partial F}{\partial N}\right)_{\tau,V} = \tau\log(N/V n_Q). \tag{14.15}$$

We can alternatively get this from the Gibbs free energy

$$G(\tau, p, N) = F + pV = N\tau \left(\log \frac{N}{V n_Q} - 1 \right) + N\tau$$

$$= N\tau \log \frac{p}{\tau n_Q}$$

$$= N\mu(\tau, p). \tag{14.16}$$

where we have used the ideal gas law $pV = N\tau$.

Consider now a system made up of multiple chemical species with particle numbers N_j. Because G is extensive, we must have

$$G(\lambda N_1, \lambda N_2, \ldots, \lambda N_k) = \lambda G(N_1, N_2, \ldots, N_k). \tag{14.17}$$

Taking a derivative with respect to λ and evaluating at $\lambda = 1$, we find

$$\sum_j N_j \frac{\partial G}{\partial N_j} = \sum_j N_j \mu_j(\tau, p) = G, \tag{14.18}$$

where we used a generalization of (14.10) to compute the μ_j from derivatives of G. Here, we made a physical assumption that μ_j is only a function of τ and p, and doesn't depend for instance on the intensive ratios N_i/N_j. This assumption is true in the limit that the different species are weakly interacting, so the presence of one type of particle does not affect others. The thermodynamic identity is

$$\tau d\sigma = dU + pdV - \sum_j \mu_j dN_j \tag{14.19}$$

so that we have

$$dG = \sum_j \mu_j dN_j - \sigma d\tau + V dp. \tag{14.20}$$

14.5   Chemical processes

The equation for a chemical reaction can be written

$$\sum_j \nu_j A_j = 0, \tag{14.21}$$

where A_j represent chemical species, and ν_j are coefficients. For example, for the reaction $H_2 + Cl_2 = 2HCL$, we have

$$A_1 = H_2, \quad A_2 = Cl_2, \quad A_3 = HCL$$

$$\nu_1 = 1, \quad \nu_2 = 1, \quad \nu_3 = -2. \tag{14.22}$$

Consider this reaction happening at constant temperature τ and pressure p. The equation (14.20) becomes

$$dG = \sum_j \mu_j dN_j \qquad (14.23)$$

On the other hand, we can write

$$dN_j = \nu_j d\widehat{N}, \qquad (14.24)$$

where $d\widehat{N}$ is the change in the number of times the reaction has taken place. Plugging this into the expression for dG, we have

$$dG = \left(\sum_j \mu_j \nu_j \right) d\widehat{N}. \qquad (14.25)$$

In equilibrium, we have $dG = 0$, so that

$$\sum_j \mu_j \nu_j = 0. \qquad (14.26)$$

14.6 Equilibrium for ideal gases

Consider the case where each constituent is an ideal gas. We will have to generalize our discussion of the ideal gas slightly to make this interesting. Recall that the partition function of a single monatomic atom was

$$Z_1^{\mathrm{monatomic}} = \sum_{\mathbf{n}} e^{-E_{\mathbf{n}}/\tau} = n_Q V. \qquad (14.27)$$

Here, the energies appearing in the sum come from the center-of-mass momentum of the atoms, and we have written the result after approximating the sum over modes as an integral in the usual way. If the gas is not monatomic, but the particles have some internal structure, then the partition function becomes

$$Z_1 = \sum_{\mathbf{n},\mathrm{internal}} e^{-(E_{\mathbf{n}}+E_{\mathrm{int}})/\tau} = \sum_{\mathbf{n}} e^{-E_{\mathbf{n}}/\tau} \sum_{\mathrm{internal}} e^{-E_{\mathrm{int}}/\tau}$$

$$Z_1 = n_Q V Z_{\mathrm{int}}(\tau), \qquad (14.28)$$

where $Z_{\mathrm{int}}(\tau)$ is the partition function associated with the internal degrees of freedom — for instance, rotational or vibrational motion.

As an example, for rotational degrees of freedom contribute

$$Z_{\mathrm{rot}}(\tau) = \sum_{j=0}^{\infty} (2j+1) e^{-\frac{\hbar^2}{2I\tau} j(j+1)}, \qquad (14.29)$$

as you computed on problem set 4. Vibrational degrees of freedom contribute

$$Z_{\text{vib}}(\tau) = \sum_{n=0}^{\infty} e^{-\hbar n\omega/\tau} = \frac{1}{1 - e^{-\hbar\omega/\tau}}. \tag{14.30}$$

The total internal partition function for a diatomic molecule is a product of these factors

$$Z_{\text{int}}(\tau) = Z_{\text{rot}}(\tau) Z_{\text{vib}}(\tau). \tag{14.31}$$

The Gibbs sum is

$$\mathcal{Z} = \sum_{n=0}^{N} \lambda^N Z_1^N = \frac{1}{1 - \lambda Z_1} \approx 1 + \lambda Z_1, \tag{14.32}$$

where $\lambda = e^{\mu/\tau}$, and we have assumed $\lambda \ll 1$ so that we are in the classical regime. The expectation value of N is

$$N = \langle N \rangle = \lambda \frac{\partial}{\partial \lambda} \log \mathcal{Z} \approx \lambda \frac{\partial}{\partial \lambda} (\lambda Z_1) = \lambda Z_1. \tag{14.33}$$

Thus, the chemical potential is given by

$$\mu = \tau \log \frac{N}{Z_1} = \tau(\log n - \log c), \tag{14.34}$$

where $c = n_Q Z_{\text{int}}$. This is a generalization of our previous results for monatomic gases.

Suppose that we have multiple species, so that

$$\mu_j = \tau(\log n_j - \log c_j), \tag{14.35}$$

where $c_j = n_{Qi} Z_{\text{int},i}(\tau)$. The equilibrium condition (14.26) can be written

$$\log \prod_j n_j^{\nu_j} = \sum_j \nu_j \log n_j = \sum_j \nu_j \log c_j = \log K(\tau),$$
$$K(\tau) \equiv \prod_j c_j^{\nu_j}. \tag{14.36}$$

Here, K is called the *equilibrium constant*, which is a function only of temperature. It can also be written

$$K(\tau) = \prod_j n_{Qj}^{\nu_j} e^{-\nu_j F_{\text{int},j}/\tau}, \tag{14.37}$$

where $F_{\text{int},j} = -\tau \log Z_{\text{int},j}$ is the internal free energy. Equation (14.36) is called the *law of mass action*.

Example 14.2 (Atomic and molecular hydrogen). As an example, consider the reaction $H_2 - 2H = 0$ for the dissociation of molecular hydrogen into atomic hydrogen. The concentrations satisfy

$$[H_2][H]^{-2} = K(\tau). \tag{14.38}$$

Here, the notation $[A] = n_A$ is shorthand for the concentration of species A. Since H is monatomic, it has $Z_{\text{int},H} = 1$. What about H_2? One important point is that there is a nontrivial binding energy $E_B = 4.476\text{eV}$ for two H's inside an H_2 molecule. When computing Z_{int,H_2}, we must use the same conventions for the zero of energy as we used for H. Thus, we have

$$Z_{\text{int},H_2} = e^{-(-E_B)/\tau} Z_{\text{rot},H_2} Z_{\text{vib},H_2}, \tag{14.39}$$

where Z_{rot,H_2} and Z_{vib,H_2} are the partition functions associated with rotational motion and vibrations. The equilibrium constant is

$$K(\tau) = \frac{n_{Q,H_2} Z_{\text{int},H_2}}{n_{Q,H}^2} = 2^{3/2} e^{E_B/\tau} \frac{Z_{\text{rot},H_2} Z_{\text{vib},H_2}}{n_{Q,H}}. \tag{14.40}$$

Here, we used $\frac{n_{Q,H_2}}{n_{Q,H}} \propto \frac{m_{H_2}^{3/2}}{m_H^{3/2}} = 2^{3/2}$. At low temperatures, the vibrational and rotational partition functions become 1, and the dominant term above is the binding energy term $e^{E_B/\tau}$. We can write our result as

$$\frac{[H_2]}{[H]} = 2^{3/2} e^{E_B/\tau} Z_{\text{rot},H_2} Z_{\text{vib},H_2} \frac{[H]}{n_{Q,H}} \tag{14.41}$$

At low concentrations $[H] \ll n_{Q,H}$, the concentration of H_2 is suppressed. This is a consequence of entropy: even though H_2 has lower energy due to binding, dissociated molecules have higher entropy. Entropy competes against the binding energy through the term $e^{E_B/\tau}$. As an example, in intergalactic space, the hydrogen has very low concentration, so most of it is unbound.

14.6.1 Heat capacity

The heat capacity at constant volume is defined by

$$C_V = \left(\frac{\partial U}{\partial \tau}\right)_V = \left(\frac{\partial \sigma}{\partial \tau}\right)_V \left(\frac{\partial \sigma}{\partial U}\right)_V^{-1} = \tau \left(\frac{\partial \sigma}{\partial \tau}\right)_V. \tag{14.42}$$

For the ideal gas, it is

$$\tau \frac{\partial}{\partial \tau} \left(\frac{3}{2} N \log \tau + \dots\right) = \frac{3N}{2}. \tag{14.43}$$

The heat capacity at constant pressure is defined by

$$C_p = \tau \frac{\partial \sigma}{\partial \tau} = \left(\frac{\partial U}{\partial \tau} \right)_p + p \left(\frac{\partial V}{\partial \tau} \right)_p. \tag{14.44}$$

The energy of an ideal gas is $U = \frac{3}{2} N \tau$, which is independent of p. Thus, this term is the same as before. We get a new contribution from

$$p \left(\frac{\partial V}{\partial \tau} \right)_p = p \frac{N}{p} = N. \tag{14.45}$$

Thus, $C_p = C_V + N = \frac{5}{2} N$ for an ideal gas. The physical interpretation is that in order to raise the temperature when the system is at constant pressure, we need to pour in more energy. Some of the energy goes to expanding the system, doing work against the pressure.

15   Phase transitions

You will notice that these lecture notes are lacking in pictures, especially in this section. For the pictures, you can (a) copy them down in class, (b) consult the textbook (which we're following relatively closely), (c) check out some of the other notes linked from the website.

Sources: K&K chapter 10, David Tong's notes on statistical physics

A phase transition is a discontinuity in thermodynamic quantities or their derivatives. For example, the heat capacity and the compressibility of water jump between the liquid phase and the gas phase. A small change in temperature or pressure leads to a big change in the equilibrium configuration.

15.0.2   Nonanalyticities

Discontinuities in thermodynamic quantities can only occur in infinite-size systems. To see this, consider a system in the canonical ensemble at temperature τ. The partition function is

$$Z(\tau) = \sum_s e^{-E_s/\tau}. \tag{15.1}$$

Quantities like, e.g., the heat capacity can be computed from derivatives of Z, $C_V \propto \beta^2 \frac{\partial^2}{\partial \beta^2} \log Z$. If the sum over states s is finite, then C_V can never be discontinuous because $Z(\tau)$ will be a finite sum of smooth curves $E^{-E_s/\tau}$ as a function of τ, and hence itself a smooth curve. However, as the number of states goes to infinity, $Z(\tau)$ can potentially develop non-smooth features, called "nonanalyticities."

In fact, it's not sufficient simply for the number of states to be infinite — for example, a single harmonic oscillator will not exhibit a phase transition. In practice, the number of *degrees of freedom* must be infinite as well. This means that phase transitions only actually occur in the infinite volume limit. In practice, the rounding out of thermodynamic quantities due to finite system size is unobservable for macroscopic systems.

To see how nonanalyticities can emerge, consider a density of states of the form

$$f(E) = \frac{1}{\Gamma(N\alpha)} E^{\alpha N - 1} + \frac{1}{\Gamma(N\beta)} E^{\beta N - 1}, \tag{15.2}$$

where $\beta > \alpha$. (The important thing is the two different power laws — the coefficients are chosen for convenience.) The partition function is

$$Z(\tau) = \int_0^\infty e^{-E/\tau} f(E) = \tau^{N\alpha} + \tau^{N\beta}. \tag{15.3}$$

The free energy is

$$F = -\tau \log Z = -\tau \log \left(\tau^{N\alpha} + \tau^{N\beta} \right). \tag{15.4}$$

Consider the limit $N \to \infty$. Note that the free energy grows like $F \sim N$, so this is like the limit of a large number of degrees of freedom. If $\tau < 1$, then the first term in parentheses dominates and we have $F \approx -\alpha N \tau \log \tau$. If $\tau > 1$, the second term dominates, and we have $F \approx -\beta N \tau \log \tau$. Thus, in the large-$N$ limit, we have

$$F = \begin{cases} -\alpha N \tau \log \tau & (\tau < 1), \\ -\beta N \tau \log \tau & (\tau > 1) \end{cases} \qquad (N \gg 1).$$

The derivative of the free energy becomes discontinuous at $\tau = 1$. The free energy itself is not discontinuous — it is equal to zero at $\tau = 1$ in both expressions.

This type of discontinuity in the derivative of F is characteristic of a *first-order* phase transition. First-order transitions are typically cases where two distinct contributions to Z are switching places in importance, as is happening in our toy model. Physically, one contribution $\tau^{N\alpha}$ comes from states in one kind of configuration, and the other contribution $\tau^{N\beta}$ comes from states in a different type of configuration. When their contributions switch places, it means that thermodynamic quantities go from being dominated by one type of configuration to the other.

15.0.3 Phase diagram of water

As a physical example of a system with phase transitions, let us look at the phase diagram of water as a function of temperature and pressure. There are three (naively) distinct phases: solid, liquid, and gas.[34] The phases are separated by *coexistence curves* in the τ-p plane. For example, the upper-left curve separates solid and liquid. It is almost vertical, with a huge downward slope. The downward slope reflects the fact that ice, when put under pressure, melts.

The top-right curve separates the liquid and gas phases — boiling occurs here. Interestingly, the curve ends at a point $(\tau_c, p_c) = (647\mathrm{K} = 374°\mathrm{C}, 218\,\mathrm{atm})$ called the *critical point*. Here, the distinction between liquid and gas disappears, and we might call the system simply a "fluid." By going around the critical point, you can actually turn liquid into gas without ever encountering a discontinuity. Thus, in some sense we cannot say they are different phases until we go sufficiently close to the coexistence curve and observe a discontinuity in thermodynamic quantities.

At sufficiently low pressure, the liquid phase disappears and only solid and gas are allowed. Ice sublimates directly to vapor as the temperature is increased. The three phase boundaries meet at a *triple point* $(273.16\mathrm{K} = 0.01°\mathrm{C}, 0.006\,\mathrm{atm})$,[35] where solid, liquid, and gas simultaneously coexist.[36]

---

[34]Actually, the full phase diagram includes other exotic phases, like different types of ice. Here, we are considering the part of the diagram involving the three most familiar phases.

[35]The Kelvin scale is defined such that the triple point of water is 273.16K.

[36]There are some awesome videos of triple points online, e.g. https://www.youtube.com/watch?v=r3zP9Rj7lnc.

15.1   Coexistence curves

At a coexistence curve, the two phases, say liquid and gas, are in thermal, mechanical, and diffusive equilibrium. The conditions for this are

$$\tau_l = \tau_g, \quad \mu_l = \mu_g, \quad p_l = p_g. \tag{15.5}$$

In particular, the chemical potentials must be equal, and this determines a curve in the τ-p plane

$$\mu_g(p, \tau) = \mu_l(p, \tau). \tag{15.6}$$

Let us relate this curve to some more easily measurable quantities.

We start by determining its slope. Consider changes $\tau \to d\tau$ and $p \to dp$, moving along the coexistence curve. In addition to (15.6), we have

$$\mu_g(p + dp, \tau + d\tau) = \mu_l(p + dp, \tau + d\tau)$$
$$\left(\frac{\partial \mu_g}{\partial p}\right)_\tau dp + \left(\frac{\partial \mu_g}{\partial \tau}\right)_p d\tau = \left(\frac{\partial \mu_l}{\partial p}\right)_\tau dp + \left(\frac{\partial \mu_l}{\partial \tau}\right)_p d\tau, \tag{15.7}$$

which can be rearranged to give

$$\frac{dp}{d\tau} = \frac{\left(\frac{\partial \mu_l}{\partial \tau}\right)_p - \left(\frac{\partial \mu_g}{\partial \tau}\right)_p}{\left(\frac{\partial \mu_g}{\partial p}\right)_\tau - \left(\frac{\partial \mu_l}{\partial p}\right)_\tau} \tag{15.8}$$

Recall that we have the relations

$$G = N\mu(p, \tau), \quad \left(\frac{\partial G}{\partial p}\right)_{N,\tau} = V, \quad \left(\frac{\partial G}{\partial \tau}\right)_{N,p} = -\sigma. \tag{15.9}$$

Let us define

$$v = V/N, \quad s = \sigma/N \tag{15.10}$$

the volume and entropy per particle. We then have

$$\left(\frac{\partial \mu}{\partial p}\right)_\tau = \frac{1}{N}\left(\frac{\partial G}{\partial p}\right)_{N,\tau} = \frac{V}{N} = v$$
$$\left(\frac{\partial \mu}{\partial \tau}\right)_p = \frac{1}{N}\left(\frac{\partial G}{\partial \tau}\right)_{N,p} = -\frac{\sigma}{N} = -s. \tag{15.11}$$

Thus, we can write our equation for the slope of the coexistence curve as

$$\frac{dp}{d\tau} = \frac{\Delta s}{\Delta v}, \quad \Delta s = s_g - s_l, \quad \Delta v = v_g - v_l \tag{15.12}$$

The quantity in the numerator is the change in entropy per particle in going from liquid to gas, and the quantity in the denominator is the change in volume per particle in going from liquid to gas.

The quantity Δs is related to the amount of heat needed to boil a molecule — i.e. to transfer a molecule reversibly from the liquid to the gas while keeping the temperature constant. The heat during the transfer is

$$\dj Q = \tau \Delta s \equiv L, \tag{15.13}$$

where L is called the *latent heat of vaporization*. With this definition, we have

$$\frac{dp}{d\tau} = \frac{L}{\tau \Delta v}, \tag{15.14}$$

which is called the *Clausius-Clapeyron equation* or *vapor pressure equation*.

15.1.1 Approximate solution to the vapor pressure equation

We can obtain a useful approximation to (15.14) with two assumptions. First that the volume per molecule in the gas phase is much larger than the volume per molecule in the liquid phase $v_g \gg v_l$, which implies $\Delta v \approx v_g$. Second, that the gas phase is well-described by the ideal gas law $pv = \tau$, so that $\Delta v \approx v_g = \tau/p$. We find

$$\frac{dp}{d\tau} = \frac{L}{\tau^2}p. \tag{15.15}$$

Let us furthermore assume that $L \approx L_0$ is approximately constant as a function of τ. This is a good approximation in water and actually ice as well for a large range of temperatures from $220K$ to $650K$ (the critical point). With this approximation, we can integrate the above differential equation to obtain

$$\log p(\tau) = -\frac{L_0}{\tau} + C$$
$$p(\tau) = p_0 e^{-L_0/\tau} \tag{15.16}$$

15.1.2 Physics on the coexistence curve

Suppose τ and p lie on a liquid-gas coexistence curve. Because the chemical potentials $\mu_l(p, \tau)$ and $\mu_g(p, \tau)$ are equal, we can move molecules from liquid to gas in a reversible manner. Specifically,

$$G = \mu_g N_g + \mu_l N_l = \mu(N_g + N_l) = \mu N, \tag{15.17}$$

which is independent of N_g. However, the volume is not independent of N_g. We have

$$V = N_g v_g + N_l v_l = N_g v_g + (N - N_g)v_l, \tag{15.18}$$

and $v_g \neq v_l$. Thus, we can solve

$$N_g = \frac{V - Nv_l}{v_g - v_l}, \quad N_l = \frac{Nv_g - V}{v_g - v_l}. \tag{15.19}$$

There exists a range of V such that both of these quantities are nonzero, and hence the system will be in an inhomogeneous mixture of two phases. In this range, the volume can be changed reversibly while the Gibbs free energy, temperature, and pressure remain constant.

If we change V into a region where $V < Nv_l$ or $V > Nv_g$, we naively find a contradiction because N_g or N_l appears to become negative. In practice, this means that if we try to move V into these regions, the temperature and or pressure will change.

15.2   The Van der Waals equation of state

We will now study a toy model of an interacting non-ideal gas that illustrates some of the physics we've discussed. The Van der Waals equation of state is a modification of the ideal gas $p = \frac{N\tau}{V}$ law that takes into account a simple model of interactions between molecules. It is

$$\begin{aligned}
p_{\mathrm{VdW}} &= \frac{N\tau}{V - Nb} - a\left(\frac{N}{V}\right)^2 \\
&= \frac{\tau}{v - b} - \frac{a}{v^2}
\end{aligned} \tag{15.20}$$

where a and b are constants. The replacement $V \to V - Nb$ models the fact that particles have short-range repulsive forces when they are brought together — each molecule essentially has some volume b, which means the total volume available for the particles to move is reduced by Nb.

The term $p \to -a\frac{N^2}{V^2}$ takes into account long-range pairwise attraction between molecules. The shift in energy due to pairwise interactions is

$$\delta U = \sum_{\mathrm{pairs}} V_{\mathrm{pair}} = \frac{1}{2} \int d^3\mathbf{x}\, d^3\mathbf{y}\, n(\mathbf{x}) n(\mathbf{y}) \phi(\mathbf{x} - \mathbf{y}), \tag{15.21}$$

where ϕ is the intermolecular potential and the $\frac{1}{2}$ is to avoid overcounting pairs. Let us make the *mean field* approximation that the density of atoms is constant $n(\mathbf{x}) = n$. This is of course not correct — for example, the density of atoms $n(\mathbf{x})$ is a sum of delta functions in the classical limit. However, the mean field approximation should be ok when there are not large spatial fluctuations in the density $n(\mathbf{x})$. It is also important that the gas is dilute — if the atoms were close and could feel large variations in each others' potentials, then $n(\mathbf{x}) = \mathrm{const}$ would be a very poor approximation. And finally, it is also important that the interactions $\phi(\mathbf{x})$ are weak — otherwise it would not make sense to start with the ideal gas law $p = \frac{N\tau}{V}$ and

make small corrections to it. We will talk more about the validity of mean field approximations later. Plugging in $n(\mathbf{x}) = n$, we have

$$\delta U = \frac{n^2}{2} \int d^3\mathbf{x} d^3\mathbf{y} \phi(\mathbf{x} - \mathbf{y}) = \frac{n^2}{2} V \int d^3\mathbf{x} \phi(\mathbf{x})$$
$$= \frac{n^2}{2} V(-2a) = -a\frac{N^2}{V}. \tag{15.22}$$

Where we simply gave the integral $\int d\mathbf{x}\phi(\mathbf{x})$ a name $-2a$. The negative sign is because the potential should be attractive. This gives a correction to the pressure via

$$\delta p \approx -\frac{\partial \delta U}{\partial V} = -a\frac{N^2}{V^2}. \tag{15.23}$$

This derivation was a reasonable approximation in the limit of low density and weak interactions. However, we are going to now take the Van der Waals equation of state and use it outside of its regime of validity because it gives a very useful toy model in which to understand some properties of phase transitions. The quantitative results will be incorrect, but some of the qualitative features will be correct.

15.2.1   Isotherms

Let us fix τ, and plot p vs V in the Van der Waals model. Such curves are called *isotherms*. If τ is large, the term $-a/v^2$ is unimportant, and the curves closely follow the ideal gas law $p \approx \tau/v$. However, as τ gets smaller, the curve $p(v)$ develops a wiggle. The value $\tau = \tau_c$ below which the wiggles occur is the one where the curve has an inflection point

$$\frac{\partial p}{\partial v} = \frac{\partial^2 p}{\partial v^2} = 0. \tag{15.24}$$

A nice way to find the inflection point is to write the VdW equation as

$$pv^3 - (pb + \tau)v^2 + av - ab. \tag{15.25}$$

This is a cubic in v with three solutions for $\tau < \tau_c$. When $\tau = \tau_c$ and $p = p_c$, the three solutions coincide, so that the equation has the form

$$C(v - v_c)^3 = p_c v^3 - (p_c b + \tau_c)v^2 + av - ab, \tag{15.26}$$

for some constant C. Expanding out and equating coefficients of powers of v, we find $C = p_c$ and

$$\tau_c = \frac{8a}{27b}, \quad v_c = 3b, \quad p_c = \frac{a}{27b^2}. \tag{15.27}$$

Let us consider $\tau < \tau_c$. There is a range of p for which there are three possible solutions v_1, v_2, v_3 for the equation of state $p(v)$. What is happening in this range? Note that the middle solution $v = v_2$ has the feature

$$\frac{\partial p}{\partial v} > 0. \tag{15.28}$$

This means that as the system grows in volume, the pressure grows. As the system shrinks in volume, the pressure decreases. The result is an instability — the system will evolve away from this configuration toward some other state.

Example 15.1 (Negative heat capacity). Incidentally, a similar instability condition applies to heat transfer. Recall that the heat capacity is defined as

$$C_V = \frac{\partial U}{\partial \tau}. \tag{15.29}$$

If the heat capacity were negative, then increasing the energy would decrease the temperature. This would cause more energy to flow into the system, further decreasing the temperature. Meanwhile, decreasing the energy would increase the temperature, causing more energy to flow out, causing the temperature to increase further. A famous example of a system with negative heat capacity is a black hole. The Hawking temperature is $\tau \propto \frac{1}{M}$, which means it decreases as the mass increases. The instability results in the black Hawking evaporating to nothing.

Back to our liquid-gas system. What configuration replaces the unstable one? Well, the Van der Waals equation of state was derived assuming the system has a fixed density n/V. We can move off this curve by making the density inhomogeneous — i.e. by turning the system into an inhomogeneous mixture of coexisting gas and liquid. As we argued before, in the inhomogeneous phase, the pressure and temperature are constant as a function of volume. Thus, the physically correct isotherm does not follow the curve $p_{\mathrm{VdW}}(\tau)$ between v_1 and v_3, but rather is a straight line in that region.

15.2.2 The Maxwell construction

To find out where the straight line occurs, we should set the chemical potentials of the liquid and gas phases equal

$$\mu(p, v_1(p)) = \mu(p, v_3(p)). \tag{15.30}$$

This equation can be solved for p once we know μ for the Van der Waals gas. Remember that we have

$$dG = \mu dN - \sigma d\tau + V dp = d(N\mu) = \mu dN + N d\mu$$
$$N d\mu = -\sigma d\tau + V dp. \tag{15.31}$$

We can integrate this equation along any path in configuration space to compute the chemical potential μ. A convenient choice is to consider paths of constant N and τ, in which case we have

$$N d\mu = V dp \qquad (d\tau = 0)$$
$$\mu_g = \mu_l + \frac{1}{N} \oint_{p_l}^{p_g} V dp \qquad (15.32)$$

To find two points with equal μ, we should find two points such that the integral of $V dp$ between the points vanishes. This says that the difference between the area above the straight line and below the straight line should be equal, which tells us where to draw the isotherm in the coexistence region. This is called the *Maxwell construction*.

Actually, there is something problematic about the Maxwell construction: it requires us to integrate $d\mu$ through the unstable region, where the physics described by the Van der Waals curve is completely wrong! A better derivation would be to choose a path between the gas point and liquid point that does not pass through the unstable region. Ultimately this would be more mathematically complicated, and will give the same answer. Anyway, the Maxwell construction is not really very important — it is a trick to help us visualize where the condition (15.30) holds.

Let us consider changing the pressure a little bit away from its value at the coexistence curve $p_{\text{coexistence}}$ at some temperature τ. If p is slightly above $p_{\text{coexistence}}$, then we follow the liquid part of the curve $\left(\frac{\partial \mu}{\partial p}\right)_\tau = v_l$. If p is slightly below $p_{\text{coexistence}}$, then we follow the gas part of the curve $\left(\frac{\partial \mu}{\partial p}\right)_\tau = v_g$. Thus, there is a discontinuity in the *first derivative* of the chemical potential (and also the Gibbs free energy) as we move across the coexistence pressure at fixed τ. Note that there is no discontinuity in the Gibbs free energy itself, since the chemical potentials have to be equal for coexistence. A discontinuity in the first derivative of the free energy is characteristic of a *first order transition*. This is the same thing that happened in our toy model of the partition function $Z(\tau) = \tau^{N\alpha} + \tau^{N\beta}$.

15.2.3   Metastable states

For each isotherm, we can determine the pressure at which the liquid and gas states are in equilibrium. The coexistence region lies inside a downward-facing parabola-like curve whose top is at the critical point. Note that there are regions inside this curve where there is no instability, i.e. $\frac{\partial p}{\partial V} < 0$ as it should be. These are metastable phases — i.e. supercooled gas or superheated liquid. We can coax the system into these states by, e.g. slowly cooling the gas. A small perturbation will cause the supercooled gas to liquify. (By contrast, in the unstable region, no small perturbation is needed.)

15.3   The critical point

Let us look at how various quantities behave near the critical point in the Van der Waals model. For example, let us ask how the difference in volume $v_g - v_l$ between the liquid and gas phases changes with temperature near the critical point.

Plugging v_g and v_l into the Van der Waals equation of state, we find

$$p = \frac{\tau}{v_g - b} - \frac{a}{v_g^2} = \frac{\tau}{v_l - b} - \frac{a}{v_l^2} \tag{15.33}$$

We can use this to solve for τ,

$$\tau = \frac{a(v_l + v_g)(v_l - b)(v_g - b)}{v_l^2 v_g^2}. \tag{15.34}$$

When we're near the critical point, we can write $v_l = v_c - \epsilon/2$ and $v_g = v_c + \epsilon/2$, where ϵ is small. Plugging this in, we find

$$\tau \approx \frac{2av_c((v_c - b)^2 - (\epsilon/2)^2)}{v_c^4 - (\epsilon v_c)^2 + O(\epsilon^3)}$$

$$= \frac{2av_c(v_c - b)^2}{v_c^4}\left(1 - \frac{\epsilon^2}{4(v_c - b)^2} - \frac{\epsilon^2}{v_c^2}\right) + O(\epsilon^3) \tag{15.35}$$

where $O(\epsilon^3)$ means terms of order ϵ^3 and higher. When $\epsilon = 0$, we must have $\tau = \tau_c$, so we can rewrite this as

$$\tau_c - \tau \sim (v_g - v_l)^2$$

$$(\tau_c - \tau)^{1/2} \sim v_g - v_l. \tag{15.36}$$

This is our first example of a *critical exponent* — the exponent in a power law relationship between quantities as one approaches a critical point.

Recall that $\left(\frac{\partial \mu}{\partial p}\right)_{\text{gas}} = v_g$ and $\left(\frac{\partial \mu}{\partial p}\right)_{\text{liquid}} = v_l$. Below the critical temperature, we noticed that these quantities jumped as we dialed the pressure past its coexistence value. Thus, the Gibbs free energy $(G = N\mu)$ had a discontinuous first derivative. However, at the critical temperature, the difference between v_l and v_g vanishes. The Gibbs free energy no longer has a discontinuous first derivative as a function of pressure. Instead, we have $\frac{\partial^2 \Delta \mu}{\partial p^2} = \frac{\partial (v_g - v_l)}{\partial p} \sim (p - p_c)^{-2/3}$, which diverges. This is characteristic of a *second-order phase transition*.

Let us compute a few more critical exponents in the Van der Waals model. We can ask how the volume changes with pressure as we move along the critical isotherm. Remember that the critical point was an inflection point $\frac{\partial p}{\partial v} = \frac{\partial^2 p}{\partial v^2} = 0$. Thus, the Taylor expansion starts with a cubic term

$$p - p_c \sim (v - v_c)^3 \qquad (\tau = \tau_c). \tag{15.37}$$

Finally, let us consider the compressibility, defined as

$$\kappa = -\frac{1}{v}\left(\frac{\partial v}{\partial p}\right)_{\tau}. \tag{15.38}$$

We would like to see how κ changes as we approach $\tau \to \tau_c$ from above. We already know that at the critical point $\left(\frac{\partial p}{\partial v}\right)_{\tau=\tau_c} = 0$. Thus, expanding for temperatures close to τ_c, we have

$$\left(\frac{\partial p}{\partial v}\right)_{\tau, v=v_c} \sim \tau - \tau_c, \tag{15.39}$$

so that

$$\kappa \sim (\tau - \tau_c)^{-1}. \tag{15.40}$$

The compressibility diverges near the critical point.

How well do these results agree with experiment? The actual answers for the behavior of water near its critical point are

$$
\begin{aligned}
v_g - v_l &\sim (\tau_c - \tau)^{\beta}, & \beta &= 0.326419(3) \\
p - p_c &\sim (v - v_c)^{\delta}, & \delta &= 4.78984(1) \\
\kappa &\sim (\tau - \tau_c)^{-\gamma}, & \gamma &= 1.23708(7).
\end{aligned} \tag{15.41}
$$

So the Van der Waals model does not actually get the critical exponents right. This is perhaps not surprising, since it is a very crude model and doesn't incorporate many of the properties of actual water. However, a very surprising fact is that the above values of the critical exponents don't actually depend on special properties of water at all. If you measure critical exponents in any other liquid-vapor transition, you find the same crazy numbers. The above critical exponents are actually some kinds of universal "constants of nature" like e and π. The failure of the Van der Waals model has nothing to do with the failure to build in details of water molecules — rather it is missing some essential feature of critical points that leads to the emergence of these universal numbers.

I have to show you so many decimal digits for β, δ and γ because the current most precise values for these quantities are actually due to me — your instructor — together with some collaborators. See, for example `https://arxiv.org/abs/1603.04436`, which computes precise values for the numbers $\Delta_\sigma = 0.5181489(10), \Delta_\epsilon = 1.412625(10)$ which are related to the critical exponents above by $\beta = \frac{\Delta_\sigma}{3-\Delta_\epsilon}, \delta = \frac{3-\Delta_\sigma}{\Delta_\sigma}, \gamma = \frac{3-2\Delta_\sigma}{3-\Delta_\epsilon}$. These numbers are computed using a method called the "conformal bootstrap," which is based on techniques from quantum field theory. In last few lectures of the term, I hope to explain a little bit about why critical exponents are universal and why quantum field theory is a good tool for describing them.

15.4   The Ising Model

Other important examples of phase transitions and critical points occurs in magnets. In fact, the phases of a ferromagnet are very closely related to those of a liquid-vapor transition in a way that we will make precise shortly.

The Ising model is a simple toy model of a ferromagnet that still displays very nontrivial behavior. It was inveted by Wilhelm Lenz in 1920 and given to his student Ernst Ising, who solved the 1-dimensional case for his Ph.D. thesis in 1924. It consists of N sites in a d-dimensional lattice. On each lattice site i lives a spin that can be either up $(+1)$ or down (-1): $s_i \in \{\pm 1\}$. The energy of the system of spins is

$$E = -J \sum_{\langle ij \rangle} s_i s_j - B \sum_i s_i. \tag{15.42}$$

The interpretation of the second term is that the spins sit in an applied external magnetic field B, causing it to be energetically favorable to point in the direction of B. The energy of a magnetic moment ms_i in a magnetic field is $-Bms_i$, and we have defined $Bm \to B$ for brevity.

In the first term, the notation $\sum_{\langle ij \rangle}$ means that the sum runs over nearest-neighbor pairs of spins. Its precise meaning depends on the dimension d and the type of lattice. We write q for the number of nearest neighbors of a given lattice site. For a square lattice in d-dimensions, each spin has $q = 2d$ nearest neighbors, corresponding to the sites reached by moving by ± 1 along each of the d axes. The Ising model can be defined on any type of lattice, but for concreteness we will mostly consider a square lattice. The cases $d = 1, 2, 3$ are relevant for real-world materials — they describe magnetic strings, surfaces, and bulk materials.

The term $-J \sum_{\langle ij \rangle} s_i s_j$ models interactions between neighboring spins and the interaction strength J is a parameter. When $J > 0$, this term makes it energetically favorable for the spins to align with each other. This case is called a *ferromagnet*. The case $J < 0$ is called an *antiferromagnet*. For the purposes of the present discussion, the distinction is not so important, but we will imagine that we have a ferromagnet.

Let us work in the canonical ensemble and introduce the partition function

$$\begin{aligned} Z &= \sum_{\{s_i\}} e^{-\beta E[s_i]} \\ &= \sum_{s_1 = \pm 1} \sum_{s_2 = \pm 1} \cdots \sum_{s_N \pm 1} e^{-\beta E(s_1, s_2, \ldots, s_N)}, \end{aligned} \tag{15.43}$$

where $\beta = 1/\tau$. The notation $\sum_{\{s_i\}}$ means we sum over all configurations of spins $s_i \in \pm 1$, and we have written it out more explicitly on the second line. An important observable is average spin or *magnetization*, which can be written in terms of a

derivative of the partition function with respect to B,

$$m = \frac{1}{N} \sum_i \langle s_i \rangle = \frac{1}{N\beta} \frac{\partial \log Z}{\partial B}. \tag{15.44}$$

15.4.1 The Ising model as a Lattice Gas

The Ising model has another interpretation as a lattice model of a gas. Consider a d-dimensional lattice, but now with particles hopping between lattice sites. The particles have hard cores, so that we do not allow multiple particles on the same site. Thus, the state of the lattice is described by occupation numbers $n_i \in \{0, 1\}$. The particles have an attractive force that makes them want to sit on neighboring sites, and we have some chemical potential μ. A Gibbs factor is $e^{-\beta(E-\mu N)}$, where

$$E - \mu N = -K \sum_{\langle ij \rangle} n_i n_j - \mu \sum_i n_i. \tag{15.45}$$

However, this system is related to the Ising model by a simple change of variable:

$$s_i = 2n_i - 1. \tag{15.46}$$

As an exercise, you can determine how J, B are related to K, μ. Up to some shifts, the chemical potential μ plays the role of an applied magnetic field, and the attraction term K plays the role of the nearest-neighbor spin-spin interaction.

For this reason, you might optimistically expect the Ising model to have something to do with liquid-vapor transitions as well. This might seem over-optimistic because this model of a gas is so much cruder than the Van der Waals model we have discussed before. However, it turns out to be right!

15.5 Mean Field Theory

For general lattices in general dimensions d, the partition function (15.43) cannot be computed analytically. An exact solution exists in $d = 1$ (which we will derive shortly). The case $B = 0$ and $d = 2$ was solved by Onsager in 1944 in a tour de force calculation. The case $d = 3$ is particularly interesting for real world materials, but has been an open problem for almost 100 years.

Mean field theory is an approximate method to evaluate Z that makes manifest some of the physics of the Ising model. We write the interactions between neighboring spins in terms of their deviation from the average spin m,

$$\begin{aligned} s_i s_j &= [(s_i - m) + m][(s_j - m) + m] \\ &= (s_i - m)(s_j - m) + m(s_i + s_j) - m^2 \end{aligned} \tag{15.47}$$

In the mean field approximation, we assume that fluctuations of spins away from the average are small, so that we can neglect the first term. The energy is then

$$E \approx -J \sum_{\langle ij \rangle} [m(s_i + s_j) - m^2] - B \sum_i s_i$$

$$= \frac{1}{2} JNqm^2 - (Jqm + B) \sum_i s_i, \qquad (15.48)$$

where q is the number of nearest neighbors for a given lattice site. The factor of $1/2$ in the first term is to avoid overcounting pairs of lattice sites.

The mean field approximation has removed the interactions. We now have a bunch of non-interacting spins in the effective magnetic field

$$B_{\text{eff}} = B + Jqm, \qquad (15.49)$$

which includes the external applied field and also a field generated by nearest neighbors. We analyzed a system of noninteracting spins a long time ago in section **??**. The partition function is

$$Z = \sum_{\{s_i\}} e^{-\beta \left(\frac{1}{2} JNqm^2 - B_{\text{eff}} \sum_i s_i \right)}$$

$$= e^{-\frac{1}{2}\beta JNqm^2} (e^{\beta B_{\text{eff}}} + e^{-\beta B_{\text{eff}}})^N$$

$$= e^{-\frac{1}{2}\beta JNqm^2} 2^N \cosh^N(\beta B_{\text{eff}}). \qquad (15.50)$$

This expression for Z depends separately on m and B_{eff}. However, we can determine m from (15.44), and this gives us a consistency equation

$$m = \tanh(\beta B_{\text{eff}}) = \tanh(\beta B + \beta Jqm). \qquad (15.51)$$

This is a transcendental equation that we can solve for m for various values of β, J, B. We can visualize the solution by plotting the curves m and $\tanh(\beta B + \beta Jqm)$ and asking that they intersect. Let us consider some different cases.

- $B = 0$, $\beta Jq < 1$ (zero applied magnetic field, high temperature/low spin-spin interaction). When $B = 0$, the curve $\tanh(\beta B + \beta Jqm)$ is symmetric around $m = 0$. When $\beta Jq < 1$, there is only one solution to (15.51) at $m = 0$. The interpretation is that at high temperatures and zero applied field, the spins fluctuate a lot and average out to zero.

- $B = 0$, $\beta Jq > 1$ (zero applied magnetic field, low temperature/high spin-spin interaction). There are now three solutions to equation (15.51), $m = -m_0, 0, m_0$. The middle solution is unstable. To see this, we can expand the free energy near $m = 0$ when $B = 0$,

$$F = -\frac{1}{\beta} \log Z \approx -\frac{N \log 2}{\beta} - \frac{1}{2}(JNq(Jq\beta - 1))m^2 + O(m^4) \qquad (B = 0). \qquad (15.52)$$

133

We see that $m = 0$ is a local maximum of the free energy and thus does not dominate the canonical ensemble. Thus, the system settles into the states $m = \pm m$. The interpretation is that the spin-spin interactions overwhelm the effects of thermal fluctuations and the spins overwhelmingly want to point in the same direction.

Precisely which direction they chooses depends on the history of the system (hysteresis). If more spins start out up than down, then the system is most likely to evolve to a state where most of the spins are up (an vice-versa). This is an example of a phenomenon called *spontaneous symmetry breaking*. When $B = 0$, the system has a $\mathbb{Z}_2$ symmetry under flipping all the spins $s_i \to -s_i$, which acts as $m \to -m$. However, the actual state of the system at low temperature is not invariant under $m \to -m$. (When the temperature is high, the symmetry is unbroken.)

- $B = 0$, $\beta Jq = 1$ (zero applied magnetic field, critical temperature). This is the point where the distinction between the high and low temperature phases disappears. The critical temperature is $\tau_c = Jq$.

- $B \neq 0$. In this case, there is always a solution for m with the same sign as B. At sufficiently high values of βJq, there are two additional solutions. The middle one is always unstable. The other one, with opposite sign to B is metastable — it is a local but not global minimum of the free energy. Thus, for the stable solution, the magnetization always points in the direction of B. The applied magnetic field breaks the $\mathbb{Z}_2$ symmetry. If we slowly change the applied magnetic field from positive to negative, the system will move into the metastable state and then at some point decay into the stable state where most spins are flipped.

To summarize, if we fix $\tau < \tau_c$ and vary B, there is a phase transition at $B = 0$ where the sign of the magnetization m jumps. Because m is a first derivative of the free energy, this is a first-order transition. As we increase τ, the first-order transition disappears at $\tau = \tau_c$. There, we have a second-order transition. Near the critical point, the phase diagram in the B, τ plane looks like a rotated version of the liquid-vapor phase diagram we discussed last time.

15.5.1   Critical exponents in Mean Field Theory

Let us compute the critical exponents of the Ising model in the mean field approximation. We first consider $\beta = 0$ and ask how the difference $2m_0$ in magnetization between the two phases depends on τ. Just below $\tau = \tau_c$, m is small and we can Taylor expand (15.51) to obtain

$$m \approx \beta Jqm - \frac{1}{3}(\beta Jqm)^3. \tag{15.53}$$

Thus, the magnetization goes like

$$m_0 \sim \pm(\tau_c - \tau)^{1/2}. \tag{15.54}$$

We see that we get the same exponent as in our computation of $n_g - n_l$ in the Van der Waals model.

We can also set $\tau = \tau_c$ and ask how the magnetization changes as we approach $B = 0$. We then have $\beta J q = 1$ and our equation is

$$m = \tanh(B/Jq + m) \tag{15.55}$$

Expanding in small B gives

$$m \approx \frac{B}{Jq} + m - \frac{1}{3}\left(\frac{B}{Jq} + m\right)^3 \approx \frac{B}{Jq} + m - \frac{1}{3}m^3 + O(B^2), \tag{15.56}$$

so that

$$m \sim B^{1/3}. \tag{15.57}$$

This is analogous to our result $v_g - v_l \sim (p - p_c)^{1/3}$ in the Van der Waals model.

Finally, let us consider the magnetic susceptibility

$$\chi = N\left(\frac{\partial m}{\partial B}\right)_\tau. \tag{15.58}$$

This is the analog of compressibility for a gas. Let us ask how χ changes as we approach $\tau \to \tau_c$ from above at $B = 0$. Differentiating (15.51) with respect to B, we find

$$\chi = \frac{N\beta}{\cosh^2(\beta Jmq)}\left(1 + \frac{Jq}{N}\chi\right). \tag{15.59}$$

Setting $m = 0, B = 0$, we find

$$\chi = \frac{N\beta}{1 - Jq\beta} \sim (\tau - \tau_c)^{-1}. \tag{15.60}$$

We again get the same critical exponent as in the Van der Waals model.

15.6 Critical phenomena and universality

We have seen that the mean field approximation gives the same answers for critical exponents in both the Van der Waals gas and the Ising model. A rough analogy between the two cases is

gas	magnet
τ	τ
p	B
τ_c	τ_c
$p = p_c$	$B = 0$
n	m
κ	χ
$n = n_g$	$m = m_0$
$n = n_l$	$m = -m_0$

On the other hand, there are some differences — the Ising model has an inbuilt $\mathbb{Z}_2$ symmetry under $B, m \to -B, -m$, whereas there is no obvious symmetry relating the liquid and gas phases.

Just like with liquid-vapor transitions, the critical exponents of the Ising model turn out to be different from those predicted by mean field theory. However, amazingly **they are the same critical exponents as in liquid vapor transitions**. The very same β, γ, δ that we discussed before appear not only in myriad liquid-vapor transitions, but at critical points of ferromagnets, and many other settings. We say that all these systems are in the same *universality class* — their critical points are controlled by the same critical exponents. The particular universality class we have been discussing is called the 3d Ising universality class. However, this is only because the 3d Ising model is the simplest model in the same universality class. Other than being simple, there is nothing special about it — it is just one of many models with the same emergent critical behavior.

What does mean field theory get wrong in all of these systems, and what is shared between different critical points? A defining feature of critical points is that fluctuations are important. In fact, the structure of fluctuations at critical points is extremely special. As an example, in a ferromagnet with $\tau < \tau_c$ and $h = 0$, there are two phases $m = \pm m_0$. In practice, these phases will coexist, so that we will have large pockets of up spins and pockets of down spins. As we approach the critical point, the distinction between $m \pm m_0$ starts to disappear and thermal fluctuations become more important. Inside the pockets of up spins, there are smaller pockets of down spins. Inside those pockets are even smaller pockets of up spins, etc.. At the critical point, this nested structure goes on ad infinitum, and we get (in the limit of an infinite lattice) infinitely nested pockets of up and down spins at all distance scales.

In liquid-vapor transitions, something similar happens. When $\tau < \tau_c$ the liquid can boil. Boiling happens by a bubble of gas appearing inside the liquid. As we approach the critical temperature, fluctuations in density become more important and we end up with droplets inside the bubbles, and bubbles inside the droplets, etc. until there are bubbles and droplets at all distance scales.

This infinitely nested structure of different phases at a critical point is reflective of an emergent symmetry: scale invariance. The system looks the same at all

distance scales, like a fractal. There is strong evidence that scale-invariant theories are extremely tightly mathematically constrained, and this is the reason why the same scale-invariant theories show up in so many different physical systems.

As an example, the tight mathematical constraints on scale-invariant theories in 2-dimensions have led to an exact solution for the critical exponents of the 2d Ising model

$$m_0 \sim (\tau_c - \tau)^\beta, \quad \beta = \frac{1}{8}$$
$$m \sim B^{1/\delta}, \quad \delta = 15$$
$$\chi \sim (\tau - \tau_c)^{-\gamma}, \quad \gamma = \frac{7}{4}. \tag{15.61}$$

The critical exponents in 3 dimensions are not known exactly, but they can be computed numerically using conformal bootstrap techniques to give the values in section **??**.

Let us now develop a bit more technology to understand the significance of fluctuations and scaling symmetry in the Ising model. Along the way, we will solve the 1d Ising model (though unfortunately we won't earn a Ph.D. for doing so like Ising did).

15.7   Solving the 1d Ising model

In this section, we will solve the 1d Ising model the *right way*. This is not necessarily the easiest way, but it is the best way in that it can be vastly generalized to many different types of lattice models and beyond.

15.8   The 1d Ising model and the transfer matrix

Let us start with the Ising lattice model in 1-dimension. For concreteness, we will study the theory on a periodic lattice with length M, so the spins s_i are labeled by $i \in \mathbb{Z}_M$. The partition function is given by

$$Z_{1\mathrm{d}} = \sum_{\{s_i = \pm 1\}} e^{-S[s]}$$
$$S[s] = -K \sum_{i=1}^{M} s_i s_{i+1} - h \sum_{i=1}^{M} s_i, \tag{15.62}$$

where $K = \beta J$ and $h = \beta B$. I will sometimes refer to $S[s]$ as the "action," even though it is equal to βE, where E is the classical energy.

We're going to solve this theory using an analogy with quantum mechanics. The partition sum can be thought of as a discrete version of a path-integral. This path integral can be computed by relating it to a quantum-mechanical theory. This is an example of the notion of *quantization*.

The key idea is to build up the partition sum by moving along the lattice site-by-site. Forget about periodicity for the moment, and consider the contribution to the partition function from spins $j < i$ for some fixed i,

$$Z_{\text{partial}}(i, s_i) = \sum_{\{s_j : j < i\}} e^{K \sum_{j<i} s_j s_{j+1} + h \sum_{j<i} s_j}. \tag{15.63}$$

Because of the interaction term $s_{i-1} s_i$, we cannot do the sum over $\{s_1, \ldots, s_{i-1}\}$ without specifying the spin s_i. Thus, we have a function of s_i. In short, $Z_{\text{partial}}(i, s)$ is the partition function of the theory on the lattice $1 \ldots i$, with fixed boundary condition s at site i.[37]

Note that $Z_{\text{partial}}(i + 1, s_{i+1})$ can be related to $Z_{\text{partial}}(i, s_i)$ by inserting the remaining Boltzmann weights that depend on s_i and performing the sum over $s_i = \pm 1$,

$$Z_{\text{partial}}(i + 1, s_{i+1}) = \sum_{s_i = \pm 1} T(s_{i+1}, s_i) Z_{\text{partial}}(i, s_i), \tag{15.64}$$

where

$$T(s_{i+1}, s_i) \equiv e^{K s_i s_{i+1} + h s_i}. \tag{15.65}$$

The key step is to recognize (15.64) as a discrete version of the Schrodinger equation in a 2-dimensional Hilbert space $\mathcal{H}$. This Hilbert space has basis $|s\rangle = |\pm 1\rangle$. The $T(s', s)$'s are elements of a 2×2 matrix $\widehat{T}$ acting on $\mathcal{H}$

$$T(s', s) = \langle s' | \widehat{T} | s \rangle, \qquad \widehat{T} = \begin{pmatrix} e^{K+h} & e^{-K-h} \\ e^{-K+h} & e^{K-h} \end{pmatrix}, \tag{15.66}$$

and $Z_{\text{partial}}(i, s)$ are the components of a vector $|\Psi_i\rangle \in \mathcal{H}$,

$$Z_{\text{partial}}(i, s) = \langle s | \Psi_i \rangle. \tag{15.67}$$

In this notation, (15.64) becomes

$$|\Psi_{i+1}\rangle = \widehat{T} |\Psi_i\rangle. \tag{15.68}$$

The matrix $\widehat{T}$ is called the "transfer matrix", and it plays the role of a discrete time-translation operator. Here, i should be thought of as a discrete imaginary time coordinate.

To be explicit, the (integrated) Schrodinger equation in a quantum theory in imaginary time is

$$|\Psi(t_E + \Delta t_E)\rangle = e^{-\Delta t_E \widehat{H}} |\Psi(t_E)\rangle, \tag{15.69}$$

---

[37]We must also impose some boundary condition at site 1. The precise choice is not important for this discussion, so we have left it implicit.

where t_E is the imaginary time coordinate, Δt_E is some imaginary time-step, and $\widehat{H}$ is the quantum Hamiltonian. Thus, the 1-dimensional Ising lattice model is equivalent to a 2-state quantum theory with Hamiltonian

$$\widehat{H} = -\frac{1}{\Delta t_E} \log \widehat{T}. \tag{15.70}$$

When the lattice is periodic with length M, the partition function is related to the transfer matrix by

$$\begin{aligned} Z &= \sum_{\{s_i\}} \langle s_M|\widehat{T}|s_{M-1}\rangle\langle s_{M-1}|\widehat{T}|s_{M-2}\rangle \cdots \langle s_1|\widehat{T}|s_M\rangle \\ &= \mathrm{Tr}(\widehat{T}^M). \end{aligned} \tag{15.71}$$

This is easy to evaluate by diagonalizing $\widehat{T}$,

$$\mathrm{Tr}(\widehat{T}^M) = \lambda_+^M + \lambda_-^M, \tag{15.72}$$

where

$$\begin{aligned} \lambda_\pm &= e^K \cosh h \pm \sqrt{e^{2K} \sinh^2 h + e^{-2K}} \\ &\rightarrow \begin{cases} 2\cosh K \\ 2\sinh K \end{cases} \quad (\text{when } h = 0). \end{aligned} \tag{15.73}$$

In the thermodynamic limit $M \to \infty$, the partition function is dominated by the larger eigenvalue

$$Z_1 = \lambda_+^M \left(1 + \left(\frac{\lambda_-}{\lambda_+}\right)^M\right) \approx \lambda_+^M. \tag{15.74}$$

This is an interesting result: the 1d Ising model has a partition function that is completely smooth as a function of the parameters K, h. Thus, *there is no phase transition in the 1d Ising model.* The interpretation is that the spin-spin interactions aren't strong enough to create different high and low temperature phases. To see a phase-transition, we will have to look in higher dimensions.

There is a nice interpretation of (15.74) in terms of our quantum mechanics analogy: the state with the largest eigenvalue of $\widehat{T}$ has the smallest eigenvalue of $\widehat{H}$ — i.e. it is the ground state, and we should call it $|0\rangle$. We have shown that the ground state dominates the thermodynamic limit. Contributions from the excited state are exponentially suppressed in the size of the system

$$\left(\frac{\lambda_-}{\lambda_+}\right)^M = e^{-M/\xi},$$

$$\frac{1}{\xi} \equiv -\log(\lambda_-/\lambda_+). \tag{15.75}$$

The constant of proportionality is $1/\xi$, where ξ is called the "correlation length". The correlation length characterizes the rate at which correlations fall off with distance.

We can also use the transfer matrix to compute correlation functions. For example, consider the two-point function $\langle s_{i_1} s_{i_2} \rangle_{\mathbb{Z}_M}$ where the subscript M indicates that we are on a periodic lattice with length M. Suppose $i_1 > i_2$. We have

$$
\begin{aligned}
\langle s_{i_1} s_{i_2} \rangle_{\mathbb{Z}_M} &= \frac{1}{Z_1} \sum_{\{s_i\}} \langle s_M | \widehat{T} | s_{M-1} \rangle \cdots \langle s_{i_1+1} | \widehat{T} | s_{i_1} \rangle s_{i_1} \langle s_{i_1} | \widehat{T} | s_{i_1-1} \rangle \cdots \\
&\quad \times \cdots \langle s_{i_2+1} | \widehat{T} | s_{i_2} \rangle s_{i_2} \langle s_{i_2} | \widehat{T} | s_{i_2-1} \rangle \cdots \langle s_1 | \widehat{T} | s_M \rangle \\
&= \frac{1}{Z_1} \mathrm{Tr}(\widehat{T}^{M-i_1} \widehat{\sigma}^z \widehat{T}^{i_1-i_2} \sigma^z \widehat{T}^{i_2}) \qquad (i_1 > i_2).
\end{aligned}
\tag{15.76}
$$

Here, we introduced the Pauli spin operator σ^z that measures the spin of a state

$$
\sigma^z |s\rangle = s |s\rangle.
\tag{15.77}
$$

It is easy to compute the correlation function (15.76) by expressing σ^z in the eigenbasis of $\widehat{T}$. One can show that in the limit of large M and large "distance" $i_1 - i_2$, the correlator factorizes into a product of expectation values $\langle 0 | \sigma^z | 0 \rangle = \langle s_i \rangle$, plus exponential corrections from the excited state

$$
\langle s_{i_1} s_{i_2} \rangle = \langle s_i \rangle \langle s_j \rangle + O(e^{-|i_1 - i_2|/\xi}, e^{-(M-|i_1-i_2|)/\xi}).
\tag{15.78}
$$

Thus, the "correlation length" indeed controls the rate at which correlations between different spins die off.

15.9   Quantization in quantum mechanics

In problem set 5, you studied a quantum mechanical particle moving in 1 dimension. The system had hamiltonian

$$
\widehat{H} = \frac{\widehat{p}^2}{2m} + V(\widehat{x}), \qquad \text{where} \qquad \widehat{p} |x\rangle = -i\hbar \frac{\partial}{\partial x} |x\rangle.
\tag{15.79}
$$

You showed that the partition function could be written

$$
\begin{aligned}
&\mathrm{Tr}\left[\left(1 - \frac{T\widehat{H}}{\hbar N} + O\left(\frac{1}{N^2}\right) \right)^N \right] \\
&= \int_{-\infty}^{\infty} A_{\frac{T}{N}} dx_0 \ldots \int_{-\infty}^{\infty} A_{\frac{T}{N}} dx_{N-1} \exp\left(-\frac{1}{\hbar} \int_0^T dt L(x(t), \dot{x}(t)) \right),
\end{aligned}
\tag{15.80}
$$

where $x(t)$ is a piecewise linear path between the x_k at times kT/N and $L(x, \dot{x}) = \frac{1}{2} m \dot{x}^2 - V(x)$ is the "Lagrangian." Taking the limit $N \to \infty$ of (15.80), we get

$$
\mathrm{Tr}(e^{-\frac{T\widehat{H}}{\hbar}}) = \int_{x(0)=x(T)} \mathcal{D}x(t) \exp\left(-\frac{1}{\hbar} \int_0^T dt L(x(t), \dot{x}(t)) \right),
\tag{15.81}
$$

where here the integral is over periodic paths. Note that the left-hand side is the partition function at inverse temperature $\beta = 1/\tau = T/\hbar$. Thus, the partition function is a path integral over paths that are periodic in imaginary time with periodicity $T = \hbar\beta$, or $\Delta t = -i\hbar\beta$.

This transformation is analogous to the one we just did for the 1d Ising model, but backwards. The analogy is as follows. Firstly, the sum over a spin is analogous to the integral over x

$$\sum_{s_i=\pm 1} \Longleftrightarrow \int_{-\infty}^{\infty} dx_i \tag{15.82}$$

The energy functional $S[s]$ in the Ising model is analogous to the action

$$S[s] \Longleftrightarrow \frac{1}{\hbar} \int_0^T dt L(x(t), \dot{x}(t)) \tag{15.83}$$

The partition function of the Ising model is analogous to the path integral

$$\sum_{\{s_i\}} e^{-S[s]} \Longleftrightarrow \int_{-\infty}^{\infty} A_{\frac{T}{N}} dx_0 \ldots \int_{-\infty}^{\infty} A_{\frac{T}{N}} dx_{N-1} \exp\left(-\frac{1}{\hbar} \int_0^T dt L(x(t), \dot{x}(t))\right) \tag{15.84}$$

The state $|s\rangle$ is analogous to $|x\rangle$

$$|s\rangle \Longleftrightarrow |x\rangle, \tag{15.85}$$

and the transfer matrix is analogous to the operator that evolves with the Hamiltonian from one time to the next

$$\widehat{T} \Longleftrightarrow e^{-T\widehat{H}/\hbar N} = \left(1 - \frac{T\widehat{H}}{\hbar N} + O\left(\frac{1}{N^2}\right)\right) \tag{15.86}$$

The lesson is: a statistical system on a 1d lattice can be transformed into a quantum mechanical system, where the direction along the lattice gets interpreted as imaginary time. This begs the question: can we do something similar for higher-dimensional lattices, and what do we get?

15.10   The 2d Ising model

Let us now consider a slightly more complicated case: the 2d Ising model. For simplicity, we set $h = 0$. We consider the partition function on the doubly-periodic lattice $\mathbb{Z}_M \times \mathbb{Z}_N$ and label spins $s_{i,j}$ by a pair $(i,j) \in \mathbb{Z}_M \times \mathbb{Z}_N$.

The action is given by

$$S[s] = -K \sum_{i,j} (s_{i,j} s_{i+1,j} + s_{i,j} s_{i,j+1})$$

$$= K \sum_{i,j} \left(\frac{1}{2} (s_{i,j+1} - s_{i,j})^2 - 1 \right) - K \sum_{i,j} s_{i,j} s_{i+1,j}$$

$$= \text{const.} + \sum_{j=1}^{N} L(\mathbf{s}_{j+1}, \mathbf{s}_j). \tag{15.87}$$

In the last line, we split the action into contributions from pairs of neighboring rows. The notation $\mathbf{s}_j$ represents the configuration of spins in the j-th row,

$$(\mathbf{s}_j)_i = s_{i,j}. \tag{15.88}$$

The action associated with a pair of neighboring rows is given by

$$L(\mathbf{s}', \mathbf{s}) = \frac{1}{2} K \sum_{i=1}^{M} (\mathbf{s}'_i - \mathbf{s}_i)^2 - \frac{1}{2} K \sum_{i=1}^{M} (\mathbf{s}_{i+1} \mathbf{s}_i + \mathbf{s}'_{i+1} \mathbf{s}'_i). \tag{15.89}$$

(The constant in (15.87) gives an unimportant multiplicative constant that we will ignore.)

To quantize the theory, we can think of the j direction as time, so $\mathbf{s}_j$ is interpreted as a configuration of spins on a fixed time-slice. The Hilbert space has an orthonormal basis vector for each such configuration,

$$\mathcal{H}_M = \text{Span} \{ | \pm 1, \pm 1, \cdots, \pm 1 \rangle \}$$

$$= \bigotimes_{i=1}^{M} \mathcal{H}_i, \tag{15.90}$$

where $\mathcal{H}_i$ is a 1-qubit Hilbert space for each site i. $\mathcal{H}_M$ is the quantum Hilbert space of M qubits, and is 2^M-dimensional. This is sometimes called a (quantum-mechanical) spin-chain.

The transfer matrix between successive time slices is a $2^M \times 2^M$ matrix with entries

$$\langle \mathbf{s}' | \widehat{T} | \mathbf{s} \rangle = e^{-L(\mathbf{s}', \mathbf{s})}. \tag{15.91}$$

The partition function on $\mathbb{Z}_M \times \mathbb{Z}_N$ is then

$$Z(\mathbb{Z}_M \times \mathbb{Z}_N) = \text{Tr}_{\mathcal{H}_M} (\widehat{T}^N). \tag{15.92}$$

To compute correlation functions, we need an operator that measures the spin at site i. This is simply the Pauli spin matrix σ_i^z associated with the i-th site

$$\sigma_i^z | s_1, \ldots, s_j, \ldots, s_M \rangle = s_i | s_1, \ldots, s_i, \ldots, s_M \rangle. \tag{15.93}$$

Correlation functions become traces of time-ordered products, e.g.

$$\langle s_{i_1,j_1} s_{i_2,j_2} \rangle = \mathrm{Tr}_{\mathcal{H}_M}(\widehat{T}^{N+j_2-j_1} \sigma_{i_1}^z \widehat{T}_2^{j_1-j_2} \sigma_{i_2}^z). \tag{15.94}$$

Let us write $\widehat{T}$ in a more familiar way as an operator on a spin chain. First split L into contributions from horizontal and vertical bonds

$$L(\mathbf{s}', \mathbf{s}) = L_h(\mathbf{s}') + L_h(\mathbf{s}) + L_v(\mathbf{s}', \mathbf{s}).$$

$$L_h(\mathbf{s}) = -\frac{1}{2}K \sum_i \mathbf{s}_{i+1}\mathbf{s}_i,$$

$$L_v(\mathbf{s}', \mathbf{s}) = \sum_i \frac{1}{2}K(\mathbf{s}_i' - \mathbf{s}_i)^2. \tag{15.95}$$

Note that

$$e^{\frac{1}{2}K \sum_i \sigma_i^z \sigma_{i+1}^z} |\mathbf{s}\rangle = e^{-L_h(\mathbf{s})} |\mathbf{s}\rangle. \tag{15.96}$$

Meanwhile, L_v only involves spins at a single site, so let us imagine that we have only one site. Note that

$$\langle s'|(1 + e^{-2K}\sigma^x)|s\rangle = e^{-\frac{1}{2}K(s'-s)^2}. \tag{15.97}$$

We also have

$$1 + e^{-2K}\sigma^x = e^{A+K'\sigma^x}, \qquad \text{where}$$

$$\tanh K' = e^{-2K},$$

$$e^A = \sqrt{1 - e^{-4K}}, \tag{15.98}$$

which follows by expanding out the Taylor series for $e^{A+K'\sigma^x}$ and matching the coefficients of $1, \sigma^x$. Thus,

$$e^{AM} \langle \mathbf{s}'|e^{K' \sum_i \sigma_i^x}|\mathbf{s}\rangle = e^{-L_v(\mathbf{s}',\mathbf{s})}. \tag{15.99}$$

The constant e^{AM} will cancel in correlation functions, so we will ignore it. Putting everything together, we find

$$\widehat{T} \propto \exp\left(\frac{1}{2}K \sum_i \sigma_i^z \sigma_{i+1}^z\right) \exp\left(K' \sum_i \sigma_i^x\right) \exp\left(\frac{1}{2}K \sum_i \sigma_i^z \sigma_{i+1}^z\right). \tag{15.100}$$

In a quantum-mechanical interpretation, we would write $\widehat{T} = e^{-\Delta\tau\widehat{H}}$, but the resulting $\widehat{H}$ would be very complicated.

Onsager's 1944 solution of the 2d Ising model consisted of diagonalizing the above matrix. The solution has an interesting and important feature: when the parameter

K is dialed to a very special value, the correlation length ξ (which is the inverse of the difference between the two lowest eigenvalues of $\widehat{T}$) diverges — i.e. the low-lying eigenvalues of $\widehat{T}$ collapse toward zero. This is the critical point of the theory, and it results in lots of interesting and rich physics. In fact, the right definition of a critical point/second-order phase transition is a point where the correlation length of a system diverges. A divergent correlation length means the system is in some sense coupled to itself on all distance scales, and can display interesting emergent phenomena.

We have started with a 2d lattice model, and in the end obtained a quantum mechanical system that has a spatial direction (in addition to time). In the continuum limit, where we zoom out to distance scales much larger than the lattice spacing, this system becomes a quantum field theory. In general, a d-dimensional lattice model in the continuum limit can be rewritten as a quantum field theory with $d-1$ spatial directions and 1 time direction. In fact, one of our best definitions of the strong nuclear force is as the continuum limit of a 4d lattice model.

A critical point is described by a quantum field theory with an infinite correlation length. The fact that the correlation length is infinite means that the theory doesn't have any intrinsic length scales — it becomes invariant under rescaling.

16   Maxwell's demon

Maxwell's demon is a deceptively simple paradox about entropy that was proposed in 1867 and resolved more than 100 years later.

Imagine a box of gas with a divider separating it into two halves. Both halves initially have the same temperature, volume, and number of particles. On the divider, there is a small door operated by a machine that we call the "demon." The demon looks at each gas molecule approaching the door. By selectively choosing to open or close the door, it allows only fast-moving molecules to enter the right half, and only slow-moving molecules to enter the left half. In this way, the temperature on the left decreases and the temperature on the right increases. The demon has apparently violated the second law of thermodynamics, since the configuration where both sides have different temperatures has lower entropy than the configuration where they have the same temperature (i.e. thermal equilibrium).

The demon does not need to be conscious — it could be a computer with a laser sensor, or even something as simple as a spring-loaded door that only opens to the right and can only be knocked open by a sufficiently fast molecule. There are also versions of the paradox that violate the second law in other ways. For example, starting with a mixture of two gases A and B, the demon could selectively let A molecules pass to the left and B molecules pass to the right, separating the mixtures. The separated gases have lower entropy than the mixed gas.

No matter what, the demon has to perform some kind of computation: it has to measure a molecule, make a decision, and act accordingly. The resolution to the paradox comes from understanding some thermodynamic properties of computation: specifically, which kinds of computations are reversible and which ones are not.

Let us start with a very concrete model of a bit of information: a ball in a double-well potential $V(x) = (x^2 - a^2)^2$. We say that the ball in the left well represents 0, and the ball in the right well represents 1. Given a ball in the 1 state, we can move it to the 0 state in a completely reversible manner without doing any work or dissipating any heat. For example, we can hook the ball up via a pulley to another ball in an exactly opposite potential $-V(x)$. Now an infinitesimally-tiny nudge will cause the ball to move from state 1 to state 0.

Suppose, however, that we do not know the initial state of the ball, and we wish to move it to the 0 state. Importantly, we want an operation that works no matter what the initial state is. Let us call this operation "SetToZero." Almost by definition, SetToZero cannot be reversible. The problem is that it maps two different initial states (0 and 1) to a single final state 0. The laws of physics can always be run backwards, so how can we achieve an irreversible operation? We must use the environment.

As a concrete example, suppose that the ball experiences some friction. Then our SetToZero operation can be: swing a hammer at the right well. If the ball is already on the left, nothing will happen. If the ball is on the right, it will get kicked

over the barrier, and settle on the other side due to friction. Friction is important for this, since otherwise the ball will just come back over the barrier. More abstractly, in this example, information about the initial state has moved via heat dissipation into the environment, where it is forgotten. In general, this must happen any time we want to erase information, since erasing is an irreversible operation. This is

Definition (Landauer's principle). Erasing information requires energy to be dissipated as heat.

The heat that's dissipated is at least $\tau \Delta \sigma$, where $\Delta \sigma$ is the information entropy associated with the erased bits.

We can now see a problem with Maxwell's demon. During operation, the demon must measure a molecule. This measurement requires storing information, like "the molecule is moving fast" in some kind of memory bank. The demon makes a decision based on this information. Then before the next molecule arrives, the demon must erase its memory bank. Erasure requires dissipating heat, which means the entropy of that information is put into the environment.

As an example, consider the spring-loaded door. In order to function correctly, the door must settle back down after each molecule comes through. This is a form of erasing information: the door is more excited if a fast molecule hits it and less excited if a slow molecule hits it, but either way it must settle down to the same state. However, settling down requires sending away energy, and hence entropy, into the environment.

Thus, the demon is ultimately taking entropy from the gas and sending it into the environment. The total entropy of system plus environment is non-decreasing, and the paradox is resolved.

One could ask: why must the demon erase its memory? Couldn't it just leave the information in memory? To understand the answer, let us imagine a memory bank given by a ticker tape that can store a string of 0's and 1's. To avoid erasing its memory, the demon must start filling up the ticker tape. After it has separated a large number of molecules, it has created a long ticker tape with an essentially random sequence of 0's and 1's. The entropy of the gas has turned into information entropy of the ticker tape. An important insight is that we should count the information entropy of the tape as physical entropy. After all, the tape is a physical system. A tape with length N has at least 2^N different microstates, and these should be treated on the same footing as the microstates of any other physical system. Thus, in this case, the demon has transferred entropy from the gas to the tape, but it has not decreased the total entropy of the combined gas-tape system.

The insight that information entropy should be counted as thermodynamic entropy, and that this gives the resolution to the Maxwell's demon paradox, was largely due to Bennett in 1982.

A   Review: trace

Let us review the trace and some of its properties. Consider a linear operator A. Its matrix elements A_{ji} in an n-dimensional basis $\vec{e}_i$ are defined by[38]

$$A\vec{e}_i = \sum_{j=1}^{n} A_{ji}\vec{e}_j. \tag{A.1}$$

The trace of A is defined by

$$\text{Tr}(A) = \sum_{i=1}^{n} A_{ii}. \tag{A.2}$$

The trace has two important properties:

- **Cyclicity.** The trace is invariant under cyclic permutations of a product of matrices

$$\text{Tr}(A_1 \cdots A_{n-1}A_n) = \text{Tr}(A_n A_1 \cdots A_{n-1}). \tag{A.3}$$

  Actually, this follows from cyclicity for a product of two matrices (exercise!)

$$\text{Tr}(AB) = \text{Tr}(BA). \tag{A.4}$$

  We can prove this as follows:

$$\text{Tr}(AB) = \sum_i (AB)_{ii} = \sum_i \sum_j A_{ij}B_{ji} = \sum_i \sum_j B_{ji}A_{ij} = \sum_j (BA)_{jj} = \text{Tr}(BA) \tag{A.5}$$

- **Basis independence.** Under a change of basis implemented by a matrix U, A changes by $A \to UAU^{-1}$. We can prove invariance of the trace under this transformation using cyclicity

$$\text{Tr}(UAU^{-1}) = \text{Tr}(U^{-1}UA) = \text{Tr}(A). \tag{A.6}$$

Suppose A is diagonalizable, i.e. there exists a change-of-basis matrix U such that

$$A = U\text{diag}(\lambda_1, \ldots, \lambda_n)U^{-1}, \tag{A.7}$$

---

[38]If $\vec{e}_i$ is an orthonormal basis, we have $A_{ji} = \vec{e}_j \cdot (A\vec{e}_i)$, or in Dirac notation $A_{ji} = \langle j|A|i \rangle$. However, what we say in this section holds in any basis.

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. Then the trace is given by the sum of the eigenvalues:

$$\text{Tr}(A) = \text{Tr}(\text{diag}(\lambda_1, \ldots, \lambda_n)) = \sum_i \lambda_i. \tag{A.8}$$

Finally, for a diagonalizable matrix, a function of that matrix is defined by applying the function to each eigenvalue

$$f(A) = U\text{diag}(f(\lambda_1), \ldots, f(\lambda_n))U^{-1}. \tag{A.9}$$

In particular,

$$\text{Tr}(f(A)) = \sum_i f(\lambda_i). \tag{A.10}$$

This is the equation used in writing (3.38), in the case $f(H) = e^{-H/\tau}$.