8.044: Statistical Physics I

Lecturer: Professor Nikta Fakhri

Notes by: Andrew Lin

Spring 2019

My recitations for this class were taught by **Professor Wolfgang Ketterle.**

1 February 5, 2019

This class's recitation teachers are Professor Jeremy England and Professor Wolfgang Ketterle, and Nicolas Romeo is the graduate TA. We're encouraged to talk to the teaching team about their research – Professor Fakhri and Professor England work in biophysics and nonequilibrium systems, and Professor Ketterle works in experimental atomic and molecular physics.

1.1 Course information

We can read the online syllabus for most of this information. Lectures will be in 6-120 from 11 to 12:30, and a 5-minute break will usually be given after about 50 minutes of class. The class's LMOD website will have lecture notes and problem sets posted – unlike some other classes, **all pset solutions should be uploaded to the website**, because the TAs can grade our homework online. This way, we never lose a pset and don't have to go to the drop boxes.

There are two textbooks for this class: Schroeder's "An Introduction to Thermal Physics" and Jaffe's "The Physics of Energy." We'll have a reading list that explains which sections correspond to each lecture. Exam-wise, there are two midterms on March 12 and April 18, which take place during class and contribute 20 percent each to our grade. There is also a final that is 30 percent of our grade (during finals week). The remaining 30 percent of our grade comes from 11 or 12 problem sets (lowest grade dropped).

Office hours haven't been posted yet; they will also be posted on the website once schedules are sorted out.

1.2 Why be excited about 8.044?

One of the driving principles behind this class is the phrase "More is different." We can check the course website for the reading "More is Different" by P.W. Anderson.

Definition 1

Thermodynamics is a branch of physics that provides phenomenological descriptions for properties of macroscopic systems in thermal equilibrium.

Throughout this class, we'll define each of the words in the definition above, and more generally, we're going to learn about the physics of energy and matter as we experience it at normal, everyday time and length scales. The most important feature is that we're dealing with the physics of **many particles at once** – in fact, we're going to be doing a statistical description of about 10^{24} particles at once. It would be very hard and basically useless to try to use ordinary equations of motion to describe the behavior of each particle.

Fact 2

Because thermodynamics is a study of global properties, like magnetism or hardness, the largeness of our systems will often actually be an advantage in calculations.

The concept of **time asymmetry** will also come up in this class. In Newton's laws, Maxwell's equations, or the Schrodinger equation, there is no real evidence that time needs to travel in a certain direction for the physics to be valid. But the "arrow of time" is dependent on some of the ideas we'll discuss in this class.

Two more ideas that will repeatedly come up are **temperature and entropy**. We'll spend a lot of time precisely understanding those concepts, and we'll understand that it doesn't make sense to talk about the temperature of an individual particle – it only does to define temperature with regards to a larger system. Meanwhile, entropy is possibly the most influential concept coming from statistical mechanics: it was originally understood as a thermodynamic property of heat engines, which is where much of this field originated. But now, entropy is science's fundamental measure of disorder and information, and it can quantify ideas from image compression to the heat death of the Universe.

Here's a list of some of the questions we'll be asking in this class:

- What is the difference between a solid, liquid, and gas?
- What makes a material an insulator or a conductor?
- How do we understand other properties of materials, like magnets, superfluids, superconductors, white dwarfs, neutron stars, stretchiness of rubber, and physics of living systems?

None of these are immediately apparent from the laws of Newton, Maxwell, or Schrodinger. Instead, we're going to need to develop a theoretical framework with two main parts:

- **Thermodynamics:** this is the machinery that describes **macroscopic quantities** such as entropy, temperature, magnetization, and their relationship.
- **Statistical mechanics:** this is the statistical machinery at the microscopic level. What are each of the degrees of freedom doing in our system?

These concepts have been incorporated into different other STEM fields: for example, they come up in Monte-Carlo methods, descriptions of ensembles, understanding phases, nucleation, fluctuations, bioinformatics, and (now the foundation of most of physics) quantum statistical mechanics.

1.3 An example from biology

Many living systems perform processes that are irreversible, and the behavior of these processes can be quantified in terms of how much entropy is produced by them. Statistical physics and information theory help us do this! As a teaser, imagine we have a biological system where movement of particles is influenced by both thermal motion and motor proteins. By watching a video, we can track each individual particle, and looking at the trajectory forward and backward, and we can construct a **relative entropy**

$$\frac{\langle \dot{S} \rangle}{k_B} \equiv D[p_{\text{forward}} || p_{\text{backward}}] = \sum p_f \ln \frac{p_f}{p_b}$$

which compares the probability distributions of forward and backward motion, and the point is that this relates to the entropy production rate of the system! But it'll take us a lot of work to get to that kind of result, so we'll start with some definitions and important concepts.

To summarize this general overview, there's two complementary paths going on here:

$$\text{Thermodynamics} \implies \text{global properties} \implies \text{(temperature, entropy, magnetization, etc.)},$$

and

$$\text{Statistical physics} \implies \text{(microscopic world to macroscopic world)}.$$

We'll also spend time on two "diversions:" quantum mechanic will help us construct the important states that we will end up "counting" in statistical physics, and basic probability theory will give us a statistical description of the properties we're trying to describe (since entropy itself is an information theory metric)! To fully discuss these topics, we're going to need some mathematics, particularly multivariable calculus.

1.4 Definitions

We'll start by talking about the basic concepts of heat, internal energy, thermal energy, and temperature.

Definition 3 (Tentative)

Thermal energy is the collective energy contained in the relative motion of a large number of particles that compose a macroscopic system. **Heat** is the transfer of that thermal energy.

(We'll try to be careful in distinguishing between energy and the transfer of that energy throughout this class.)

Definition 4

Internal energy, often denoted U, is the sum of all contributions to the energy of a system as an isolated whole.

This internal energy U is usually made up of a sum of different contributions:

- Kinetic energy of molecular motion, including translational, vibrational, and rotational motion,
- Potential energy due to interactions between particles in the system, and
- Molecular, atomic, and nuclear binding energies.

Notably, this does not include the energy of an external field or the kinetic and potential energy of the system as a whole, because we care about behavior that is internal to our system of study.

Example 5

Consider a glass of water on a table, and compare it to the same glass at a higher height. This doesn't change the internal energy, even though the glass has gained some overall gravitational potential energy.

Definition 6 (Tentative)

Temperature is what we measure on a thermometer.

As a general rule, if we remove some internal energy from a system, the temperature will decrease. But there are cases where it will plateau as well! For example, if we plot temperature as a function of the internal energy, it is linear for each phase state (solid, liquid, vapor), but plateaus during phase changes, because it takes some energy to transform ice to water to vapor. And now we're ready to make some slightly more precise definitions:

Definition 7

Let U_0 be the energy of a system at tempreature $T = 0$. **Thermal energy** is the part of the internal energy of a system above $U = U_0$.

Notice that with this definition, the binding energy does not contribute to thermal energy (because that's present even at $T = 0, U = U_0$), but the other sources of internal energy (kinetic energy, potential energy) will still contribute.

Definition 8

Heat is the transfer of thermal energy from a system to another system.

This is not a property of the system: instead, it's energy in motion! And heat transfer can occur as heat conduction or radiation, a change in temperature, or other things that occur at the microscopic level.

1.5 States and state variables

The next discussion is a little bit more subtle – we want to know what it means for our system to be in a particular **state**. In classical mechanics, a state is specified by the position and velocity of all objects at time t. So if we're given the two numbers $\{x_i(t), \dot{x}_i(t)\}$ for each i (that is, for every particle in our system), we have specified everything we might want to know. Meanwhile, in quantum mechanics, the state of a system is specified by quantum numbers: for example, $|n_1, \cdots, n_M\rangle$ (for some nonnegative integers n_i) is one way we might describe the system. But we have a completely different definition of "state" now that we're in a macroscopic system:

Definition 9

A system which has settled down is in a **state of thermodynamic equilibrium** or **thermodynamic macrostate**.

Here are some of the characteristics of a system at thermal equilibrium:

- The temperature of the system is uniform throughout the space.
- More generally, at the macroscopic scale, no perceptible changes are occurring, though there are still changes at the microscopic level (like atomic or molecular movement).
- The system is dynamic (meaning that the system continues to evolve and change at the microscopic level).

It's important to note that other properties of the system (that are not temperature) can be non-uniform at equilibrium! For example, if we mix water and oil, there will obviously be some differences in different parts of the system no matter how long we wait for the atoms to mix.

Definition 10

State functions and **state variables** are properties that we can use to characterize an equilibrium state.

Some examples are **pressure** P, **temperature** T, **volume** T, the **number of particles** N, and the **internal energy** U. Note that some quantities are only defined for systems at thermal equilibrium, such as pressure and temperature, while others are defined for more general systems, such as volume.

Another big part of this class is coming up with equations that relate these state functions:

Example 11

The most famous equation of state, the **ideal gas law** ($PV = Nk_BT$), dictates the behavior of an ideal gas.

Definition 12

A **macrostate** is a particular set of values for the state variables of a system. Meanwhile, a **microstate** tells us more at the particle level, specifying the state of each individual particle.

For example, if we have a glass of water, we could (in principle) track each particle, writing down a microstate and describing our system at the microscopic level. But there are many different configurations that give a specific pressure and temperature, so a vast number of microstates can be consistent with a given macrostate.

Definition 13

A macrostate's **multiplicity** is determined by the number of microstates consistent with that macrostate, an **ensemble** is defined as the set of all possible microstates that are consistent with a given macrostate.

This class will develop methods for describing such ensembles corresponding to a specific macrostate, and in particular one important consideration is that each microstate in the ensemble occurs with some **probability**. So we'll be developing some probability theory in the next few lectures, and that will help us approach the physics more precisely.

2 February 6, 2019 (Recitation)

2.1 Introduction

Usually, Professor Ketterle introduces himself in the first class, and this time he notices some people from 8.03. We'll start today by giving a spiel for 8.044, and this class is particularly exciting to Professor Ketterle because he feels a connection to his research! Professor Fakhri does statistical physics in the classical world (in cells in aqueous solution), while Professor Ketterle examines systems at **cold temperatures**, which require quantum mechanics to describe. But it doesn't always matter whether the microscopic picture is quantum or classical! In this class, we're going to learn a framework for systems that have many degrees of freedom, but where we only know the macrostate (such as a liter of water at a certain pressure and temperature).

Fact 14

Professor Ketterle's research can be described as taking temperature $T \to 0$.

His lab has gotten to temperatures of 450 picoKelvin, which is "pretty cold." (In fact, according to Wikipedia, it's the coldest temperature ever achieved.) For comparison, the background temperature in interstellar space is about 2.5 Kelvin, which is more than a billion times warmer than what's been achieved in labs.

Low-temperature developments have opened up fields in physics today, because when we cool gas down to that regime, quantum mechanics becomes more apparent. Basically, when atoms have higher energy (at higher temperatures), they behave like classical particles that collide. But in reality, atoms should be thought of as de Broglie waves – it's just that in the classical situations, the de Broglie wavelength

$$\lambda = \frac{h}{mv}$$

does not play any role in the collisions, because λ is shorter than the size of the nucleus. But when we have an atom, which has low m, and cool it down so that v is small, the de Broglie wavelength can increase to the order of a micron or millimeter, which is large at atomic scales. So if we have a bunch of gas particles, and each particle's wave is localized, new forms of matter can form with completely new properties, and that's what makes low-temperature physics interesting.

Remark 15. *When Professor Ketterle was an undergraduate student, he found statistical mechanics fascinating. He found it attractive that we can predict so much in physics given so few pieces of information: just using statistics and ideas of "randomness" (so that we have* **isotopy** *of velocity), we can make many predictions: find limits on efficiency of engines, derive gas laws, and so on. So we'll get a taste of that through this class!*

Professor Ketterle wants recitations to be interactive, so we should bring questions to the sessions. (He does not want to copy to the blackboard what we can just read from solutions.) Concepts are important: it's important to address all of the important key points so that we can understand derivations. However, to become a physicist, we do need to do problems as well. So in order to compensate for the focus on problems in lecture, usually we will not have material "added." So Professor Ketterle will prepare topics that help us get a deeper understanding or general overview of concepts, but we can also bring questions about homework if we have any.

2.2 Review of lecture material

There were many buzzwords covered during the first lecture. We started thinking about what it means to have a **state** (microstates and macrostates, both classical and quantum), **ensemble**, **energy** (thermal energy Q, internal energy U, heat, and work), and **temperature** (which is particularly interesting in the limit $T \to 0$ or $T \to \infty$). Basically, we have introduced certain words, and it will take some practice to get fluent in those ideas.

Question 16. *How do we explain to a non-physicist why there is an "absolute zero" lowest temperature? (This is a good question to ask a physicist.) In fact, how is temperature actually defined?*

We still don't have a formal definition of temperature, but let's look at an example of a real-life situation in which temperature is very relevant: an ideal gas.

Fact 17

Here's an important related fact: starting on May 20, 2019, the kilogram will be defined in terms of the fixed constant $\hbar$, and along with that, the definition of a Kelvin will also change to depend on the Boltzmann constant k_B rather than the triple point of water.

So the Boltzmann constant will soon specify our units – it's some constant that is approximately $1.38 \cdots \times 10^{-23} J/K$ (and will come up frequently in the rest of this class), and now temperature is related to a measure of energy:

$$k_B T = [J].$$

In an ideal gas (in which we neglect weak interactions between particles), we'll make the definition

$$E = \frac{1}{2}m\overline{v^2} \equiv \frac{3}{2}k_B T,$$

where $\overline{v^2}$ will be defined next lecture. If we take all of this for granted, the lowest possible temperature must occur when there is no kinetic energy: if $\overline{v^2} \to 0$, $T \to 0$. And this also explains why there is no negative temperature in the ideal gas – we can't have less than zero energy. So absolute zero is essentially created by definition: it's a situation in which there is no kinetic energy for any of our particles. For an analogous situation, we can measure the pressure in a vacuum chamber, which is proportional to the density of particles. And thus the lowest pressure is zero, since we can't have negative particles.

On the other hand, what's the highest temperature we can achieve? In principle, there is no upper limit: we can make kinetic energy per particle arbitrarily large. We can add some entertainment to the situation as well: even though the velocity v is upper bounded by c, we do have a divergent expression for **relativistic energy** which is not just $\frac{1}{2}mv^2$:

$$KE = \frac{1}{2}\frac{m_0 c^2}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

In other words, we can keep adding energy to a particle, and it will just get heavier (without going over the speed of light). And in the lab, the highest temperatures we've ever achieved are around 2×10^{12} K in particle accelerators. In general, if we take two nuclei and smash them together when moving near the speed of light, temperature changes happen when during the actual collision. Then energy is converted into particles, and we have a piece of matter in which we are almost at thermal equilibrium.

But what really happens at 10^{12} Kelvin is interesting on its own. As our temperature rises, a solid melts, a liquid evaporates, molecules dissociate, atoms ionize into electrons and ions, and then ions lose more and more electrons until they are bare nuclei. If we go hotter, the nuclei dissociate as well! All of this behavior does actually happen in white dwarfs and neutron stars, but if we go even hotter, the protons and neutrons will dissociate into quarks, and we get a soup of quarks and gluons. Overall, it's interesting that we've actually achieved room temperature times 10^{10}, and we've also achieved room temperature divided by 10^{10}, and there is physical behavior going on at both temperatures.

And now we can talk about what "negative temperature" actually means – we'll have a more rigorous discussion when we study spin systems, but it's good for us to know that there are some magnetic systems that can reach infinite temperatures at finite energies. What's basically happening is that we're traversing $\frac{1}{T}$ from ∞ to 0: when we cross over 0, we can get into negative temperatures, and thus negative temperatures are in some way "even hotter" than infinite temperatures! Expressions of the form $e^{-E/(k_B T)}$ are going to show up **frequently in this class**, so we will actually get $\frac{1}{T}$ reasonably often in our calculations. And we'll understand why this happens as probability theory comes into the picture!

3 February 7, 2019

The first lecture was a glossary of the terms we will see in this class; we'll be slowly building up those concepts over the semester. The first problem set will be posted today, and the deadline will generally be the following Friday at 9pm. Lecture notes have been uploaded, and they will generally be uploaded after each class (as will some recitation materials) under the Materials tab of the LMOD website. For example, starting next Monday in recitation, there will be a review of partial derivatives, and the notes for those are posted already.

Also, we have two graduate TAs this semester! Pearson (from last semester's 8.03 team) will also be available to help. General office hours will also be posted by tomorrow, but in general, we should use email instead of Piazza.

3.1 Overview

We will always begin lecture with a small overview of what will be covered. Statistical physics and thermodynamics are for bringing together the macroscopic and microscopic world, and we're going to start by defining state functions like pressure and temperature using a tractable, simply-modeled system and working from first principles.

Essentially, we will use a **monatomic ideal gas** to define temperature and pressure, and then we will derive the ideal gas law. From there, we'll see how to make empirical corrections to have a more realistic understanding of a system (for example, a van der Waals gas). We'll also briefly talk about the **equipartition theorem**, which lets us connect temperature to energy, as well as the **first law of thermodynamics**, which is basically a restatement of conservation of energy.

3.2 The simplest model

Definition 18

A **monatomic ideal gas** is a system with N molecules, each of which is a single atom with no internal dynamics (such as rotation or vibration). The molecules collide elastically as point particles (and take up no space), and the only energy in the system is kinetic energy.

So putting in our definitions, the kinetic energy, thermal energy, and internal energy are all essentially the same thing, and they are all equal to

$$U = \sum_{i=1}^{N} E_i = \frac{m}{2} \sum_{i=1}^{N} v_{ix}^2 + v_{iy}^2 + v_{iz}^2$$

if all molecules have the same mass. Now assuming that we have an isotropic system, we can assume the three coordinates have equal averages, and we can define an **average (squared) velocity**

$$\overline{v^2} \equiv \langle v_{ix}^2 \rangle = \langle v_{iy}^2 \rangle = \langle v_{iz}^2 \rangle.$$

Plugging this in, the average internal energy is then

$$U_{\text{avg}} = N\langle E \rangle = \frac{3}{2} N m \overline{v}^2.$$

Definition 19

The **temperature** is a measure of the thermal energy of the system, given by

$$m\overline{v^2} \equiv k_B T$$

where k_B is a proportionality constant and T has units of Kelvin (in degrees above absolute zero).

The Boltzmann constant k_B has units of energy per temperature, and it is experimentally about $1.381 \times 10^{-23} J/K$. Plugging this in, we find that the internal energy of an ideal gas is

$$U = \frac{3}{2} N k_B T.$$

Fact 20

Usually chemists write this slightly differently: N is defined to be $N_A n$, where N_A is Avogadro's number $\approx 6.023 \times 10^{23}\text{mol}^{-1}$, and n is the number of moles of the gas. Then the **ideal gas constant** is defined to be $R \equiv N_A k_B \approx 8.314 J/\text{mol K}$, and our equation can be written as

$$U = \frac{3}{2} N k_B T = \frac{3}{2} nRT.$$

The idea is that each of the three dimensions is contributing an equal amount to the energy in the system. We've used the fact that particles have energy to define a temperature, and now we'll similarly use the fact that particles have momentum as well to define pressure. Consider a container shaped like a box with a piston as one of the walls (in the x-direction). We know that by Newton's law, the force can be described as

$$F_x = \frac{dp_x}{dt},$$

where $p = mv_x$ is the momentum in the x-direction for one of the particles. Since the piston will barely move, we can just say that the particle will reverse x-momentum when it bounces off, but has no change in the other two directions:

$$\Delta p_x = 2mv_x, \Delta p_y = \Delta p_z = 0.$$

If we let the cross-sectional area of the piston-wall be A and the length of the box in the x-direction be ℓ, then the time between two collisions with the piston is

$$\Delta t = \frac{2\ell}{v_x}$$

(since it must hit the opposite wall and bounce back). So now plugging in our values, the average force from this one molecule is

$$F_x = \frac{\Delta p_x}{\Delta t} = \frac{mv_x^2}{\ell}.$$

Assuming no internal molecular collisions (since we have an ideal gas), the total force on the piston for this system is then

$$F_x = \sum_{i=1}^{N} F_{xi} = \frac{Nm}{\ell}\overline{v^2} = \frac{N}{\ell} k_b T$$

by our definition of temperature. So now the **pressure** on the piston, defined to be force per unit area, is

$$P = \frac{F_x}{A} = \frac{N k_B T}{\ell A} = \frac{N k_B T}{V}$$

Now we're making some assumptions: if we say that collisions and interactions with the wall don't matter, and that the shape of the container does not matter (an argument using small volume elements), we can rearrange this as

$$\boxed{PV = N k_B T},$$

which is our first equation of state for the class. (As a sidenote, the shape of the container will matter if our system is out of equilibrium, though.) So pressure, volume, and temperature are not independent: knowing two of them defines the third in our ideal gas system, and we're beginning to find a way to relate our state functions!

3.3 An empirical correction

Let's start modifying our equation of state now – we're going to use the chemistry version of the ideal gas law, where n is measured in moles. One key assumption that we have right now is that the particles have no volume: we have to make some corrections if we don't have point particles any more. So we change our volume: letting b be some measure of how much volume is taken up by the particles, we replace

$$V \to V - nb.$$

Also, some particles may have attractive intermolecular forces, and to introduce this, we claim the pressure will change as

$$P = \frac{nRT}{V} - a\left(\frac{n}{V}\right)^2.$$

The constants a and b are empirically measured in a lab, but the point is that these modifications give us the **van der Waals equation**

$$\left(P + a\frac{n^2}{V^2}\right)(V - nb) = nRT.$$

This means the effective volume for a real gas is smaller than an ideal gas, but the pressure can be larger or smaller than an ideal gas because we could have attractive or repulsive molecular interactions.

3.4 The equipartition theorem

If we take another look at the equation for internal energy

$$U = \frac{3}{2}Nk_BT = 3N\left(\frac{1}{2}k_BT\right),$$

notice that our system has $3N$ degrees of freedom: one in each of the x, y, and z coordinates for each of the particles.

Proposition 21 (Classical equipartition theorem)
At thermal equilibrium at temperature T, each **quadratic** degree of freedom contributes $\frac{1}{2}k_BT$ to the total internal energy U of the system.

This is important for being able to consistently define temperature! Unfortunately, this is only true in classical limits at high temperatures. And we should make sure we're precise with our language:

Definition 22
A **degree of freedom** is a quadratic term in a single particle's energy (or **Hamiltonian**). Examples include:

- translational (in each coordinate) about the center of mass: $\frac{1}{2}m\left(v_x^2 + v_y^2 + v_z^2\right)$,
- rotational (in each axis): $\frac{1}{2}\frac{\ell_x^2}{I_x}$ and so on, and
- vibrational: $\frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2$, imagining a molecule with two atoms in simple harmonic oscillation.

Example 23
Let's try writing down the different degrees of freedom for molecules of a diatomic gas.

In such a system, we have **3 translational** degrees of freedom (looking at the center of mass), **2 rotational** degrees of freedom (we don't have the third because there's no moment of inertia about the axis connecting the two atoms), and **2 vibrational** degrees of freedom (coming from the simple harmonic oscillator of the two atoms stretching). Thus, by the equipartition theorem, we already know that we'll have

$$U = 7N\left(\frac{1}{2}k_BT\right),$$

and that's the power of the equipartition theorem: it allows us to have a general method for relating energy and temperature. And as an exercise, we should try to figure out why a **simple crystalline solid** has internal energy $U = 3Nk_BT$, using the same kinds of argument.

3.5 The first law of thermodynamics

We now have a relationship between thermal energy and temperature, and our next step is to think about how the energy of a system can change. We'll start by trying to define a relationship between work and pressure. Let's go back to the piston wall that we used to derive the ideal gas law: the differential work being done on the piston by a particle is

$$dW = F\,dx = PA\,dx = P\,dV.$$

We will use the convention in this class where whenever $dV < 0$, work is being done **on the system**. So the change in the internal energy of the system is the mechanical work done, and $\boxed{dU = -P\,dV}$ (with the negative sign because work is being done **by**, rather than **on**, the system.)

But there are also other ways to transfer energy, particularly through heat (which we denote with the letter Q). We'll use the convention is that heat flow into the system is positive, so $\boxed{dU = dQ}$. So we can write these together to get a total change in internal energy

$$dU = dQ + dW.$$

In other words, the internal energy of the system changes if we add infinitesimal heat to it, or if the system does work externally. We can also add particles to the system to further modify this first law: we'll see later on that we sometimes also get a contribution of the form $dU = \mu\,dN$ where μ is the chemical potential, then

$$dU = dQ + dW + \mu\,dN.$$

Explicitly, the whole point of this kind of statement is to have an energy conservation law, but implicitly, we also have that U, the internal energy of the system, is also a **state function**. But we should keep in mind that work and heat are path-dependent quantities, so the expressions dQ and dW are **inexact differentials**. And in particular, W, Q are not state functions!

Example 24

Consider an evolution of a system as follows. Start in state 1, with N_1 particles, a temperature of T_1, and a volume of V_1. This tells us that we have some internal energy $U(N_1, T_1, V_1)$. We can take this to state 2 by adding some heat, so now we have N_2 particles, a temperature of T_2, and a volume of V_2, giving us a new internal energy. Finally, perform some work on the system to take us back to state 1.

The idea is that U is an **exact differential**, so the internal energy doesn't depend on the path taken. But dW and dQ are both inexact and path-dependent quantities – the example above showed us that $dU = 0$, but the work and heat done depend on what state 2 is. So from here on out, we will use the notation $\text{đ}W$ and $\text{đ}Q$, and now we write the first law as

$$\boxed{dU = \text{đ}Q + \text{đ}W}.$$

So calculus tells us that dU can be obtained from differentiation (of some other function), while $\text{đ}Q$ and $\text{đ}W$ cannot.

In general, state functions can be divided into **generalized displacements** and **generalized forces**, which will be denoted $\{x\}$ and $\{j\}$, respectively. These x_is and j_is are **conjugate variables**, and the idea is to write our differential work as

$$\text{đ}W = \sum_i j_i \, dx_i.$$

Here are some sample state functions that we'll be working with throughout this class

Forces (j_i)	Displacements (x_i)
Pressure P	Volume V
Tension F	Length L
Surface tension σ	Area A
Magnetic field H	Magnetization M
Chemical potential μ	Number of particles N

The quantities under "displacements" are generally **extensive**, while the quantities under "force" are **intensive**. We'll use these words more as the class progresses, but the basic point is that scaling the system up changes our generalized displacements, but not the generalized forces.

4 February 11, 2019 (Recitation)

4.1 Questions and review

We'll begin with a few ideas from the homework. If we want to calculate the work done by moving from one state to another, we can integrate along the path taken and compute

$$W = \oint_{\text{initial}}^{\text{final}} P \, dV.$$

Many such problems are solved using **PV diagrams**, which plot each state based on their pressure (on the y-axis) and volume (on the x-axis), and knowing P and V tells us the temperature T (through the equation of state) as well. And if we have an ideal gas, lines $PV = Nk_BT$ on the PV diagram are where the temperature is constant. For example, if we have an isothermal expansion along one of these lines, we find that

$$\int P \, dV = \int \frac{Nk_BT}{V} dV = Nk_BT(\ln V_f - \ln V_i).$$

There are other kinds of work that can be specified: for instance, isobaric compression is done under constant pressure, so the work is just $P\Delta V$. But we'll come back to all of this later!

Remark 25. *State functions are all the parameters that characterize a system. So for an ideal gas, pressure, volume, and temperature are enough – if we are asked to find the final state, we just need to find the values of those parameters.*

Fact 26

The word **adiabatic** (referring to a process in which a system evolves) has **two different definitions** in statistical physics. In one definition, no heat is exchanged, meaning that $đQ = 0$ throughout the process. (This can happen if our system has insulating walls, so no heat is transferred in or out of the system, or if a process proceeds quickly enough so that heat can't move.) But in the other definition, our process is slow enough so that the system is always in equilibrium. And the main feature of this definition (which isn't true in the other) is that **entropy** is conserved.

Adiabaticity in quantum mechanics means that a particle doesn't deviate from its quantum state, because the process happens very gradually. (If we try to evolve too slowly, though, we will get noise that also interferes with the system.) The bottom line is that adiabaticity forces our system to act "not too slow and not too fast," so that we get the desired constraints.

4.2 Energy

Let's start from the first law of thermodynamics, stated as

$$dU = đW + đQ.$$

As physicists, we don't have access to absolute truth: we do our best to find better and better approximations. So Professor Ketterle doesn't always like it when we call things "laws:" for example, why are we trying to test Coulomb's law to the tenth decimal place, and why do we do it for two electrons that are very close? Regardless, the "law" above is a statement about energy.

Question 27. *What's a form of energy that is internal energy but not thermal energy?*

Two possible answer are "the binding energy of the nucleus" or "the mass energy," but these aren't exactly correct. Thermal energy is supposed to "come and go" as our system heats up, so let's think about a system of water molecules at increasing temperature. At first, our molecules gain kinetic energy, and then after we continue to heat, chemical energy will change through dissociation. So in this system, the binding energy of hydrogen is "reaction enthalpy," which is indeed considered in thermal energy.

And similarly, if we increase temperature so that the kinetic energy is comparable to rest mass energy, we get issues with relativity. If two such particles collide, they can create a particle-antiparticle pair, and in this regime, even the rest mass becomes part of a dynamic process. Therefore, that rest mass can become thermal energy as well. One takeaway we can have is that

$$U = E_{\text{thermal}} + U_0$$

for some constant U_0, and we basically always only care about differences in energy anyway. So this distinction between "thermal energy" and "internal energy" isn't really that important in Professor Ketterle's eyes.

4.3 Scale of systems

Question 28. *Can a single particle have a temperature? That is, can we have an ensemble consisting of one particle?*

Normally, we are given some constant P, V, T, U, and a microstate is specified by a set of positions and momenta $\{x_i, p_i : 1 \le i \le N\}$, where N is a large number of particles. It turns out, though, that even single particles can be thermal ensembles! For example, if we connect a particle in a box to a certain thermal reservoir at temperature T, we can find a "Boltzmann probability distribution" $\propto e^{-E/(k_B T)}$ for being at state E (this is a point which we'll study later). So having an ensemble just means we have many copies of a system that are equally prepared macroscopically, regardless of how many particles this system has, as long as we follow all of the important laws.

And in particular, remember that in an ideal gas, we've assumed the particles are not interacting! So it's perfectly fine to take $N \to 1$ for an ideal gas; rephrased, an ideal gas is just N copies of a single-particle system.

Remark 29. *Schrodinger once said that the Schrodinger equation only describes ensembles when measurements are applied many times. He made the claim that the equation would not apply to just one particle, but recently, single photons, atoms, and ions were observed repeatedly, and it was shown that the quantum mechanical ideas applied there too.*

So we may think it's nonsense that statistics can apply to a single particle, but we can often study a complicated system by simplifying it into multiple copies of a simple one.

4.4 Partial derivatives

Consider the two equations

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx},$$

$$\frac{dz}{dx} = -\frac{dz}{dy}\frac{dy}{dx}.$$

We can ask ourselves "which one is correct?". One general rule to keep in mind is that **in each field of physics, we need to learn some mathematics**. In particular, two tools we'll need to learn for this class are **partial derivatives** and **statistics.** In the handout posted online, we'll see the second statement, but of course we need to ask ourselves why we don't cancel the dys like we have done in ordinary calculus.

The key point is that we use the two equations in different situations. The first equation is valid when z is a function of x, but we have it written as an implicit function y. Rephrased, if we have a function $z(y(x))$, such as $\exp(\sin x)$, then indeed the chain rule indeed tells us that

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}.$$

This is true for a function that depends on a **single independent parameter**. But on the other hand, suppose we have a function where x and y are independent variables: that is, we have $z(x, y)$? (For example, pressure is a function of volume and temperature.) Now z can change by the multivariable chain rule, and we can say that

$$dz = \left.\frac{\partial z}{\partial x}\right|_y dx + \left.\frac{\partial z}{\partial y}\right|_x dy.$$

Often we'll have a situation where we want to keep z constant: for example, we might be keeping the pressure constant as we heat up our system, and we're thinking about pressure as a function of other state variables. In a situation like that, we have

$$0 = \left.\frac{\partial z}{\partial x}\right|_y dx + \left.\frac{\partial z}{\partial y}\right|_x dy,$$

so now x and y must be changed in a certain ratio:

$$\frac{dy}{dx} = -\frac{\left.\frac{\partial z}{\partial x}\right|_y}{\left.\frac{\partial z}{\partial y}\right|_x}$$

And the left hand side of this equation, in more rigorous language, is just $\frac{\partial y}{\partial x}$ as we keep z fixed. So the second equation is now true, and the moral of the story is that we need to be very careful about what variables are being kept constant. We'll do much more study of this in the following weeks!

5 February 12, 2019

We'll start with some housekeeping: our first problem set is due on Friday, and there will be office hours for any questions we have. If we click on the instructor names on the course website, we can see what times office hours are being held. (Hopefully we all have access to the online class materials: if there are any problems, we should send an email to the staff.)

Last lecture, we introduced thermodynamics with a simple and tractable model (the ideal gas). Once we defined pressure and temperature, we derived an equation of state, and we learned that we can empirically modify the ideal gas law to capture real-life situations more accurately. Next, we introduced the first law of thermodynamics, which is essentially conservation of energy. We learned that there are many ways to do work, and we can write the infinitesimal work as the product of a "force" and its corresponding "conjugated variable." This led us to the generalized first law

$$dU = đQ + đW, \quad đW = \sum_i j_i \, dx_i,$$

where each j_i is a generalized force and x_i is its conjugated generalized displacement.

In recitation, we reviewed some material about partial derivatives, and there will be a "zoo" of them in this class. Any macroscopic quantity can be found by taking derivatives of "free energy" (which will be defined later as well). For example, we can take derivatives of energy to find temperature and pressure, which will be helpful since we can use statistical physics to find general quantities like the free energy or entropy. But again, this is all a preview for what will become more rigorous later.

5.1 Experimental properties

Definition 30

A **response functions** are quantities that change when parameters of a system are adjusted, and they are used to characterize the macroscopic properties of that system.

Basically, we introduce a perturbation to a system, and we can then observe the response in our measurements.

Example 31 (Heat capacity)

Suppose we add some heat to a system, and we want to keep track of what happens to the temperature.

We need to be careful, because the system can change while keeping pressure or volume constant? Both are useful quantities, and we'll define the **heat capacities**

$$C_V \equiv \left.\frac{đQ}{dT}\right|_V, C_P \equiv \left.\frac{đQ}{dT}\right|_P.$$

(These can be thought of as variables for a gas on which we perform experiments.)

Example 32 (Force constant)

Suppose we apply a force F on our system, and we want to see the displacement x that results from this external force. (This is a generalization of a spring constant.)

We can therefore define an effective **force constant** via the equation

$$\frac{dx}{dF} \equiv \frac{1}{k}.$$

For example, we can define the **isothermal compressibility** of a gas via

$$K_T \equiv -\frac{1}{V}\left.\frac{\partial V}{\partial P}\right|_T.$$

This is, again, something we can measure experimentally.

Example 33 (Thermal responses)

The **expansivity** of a system is defined as

$$\alpha_P = \frac{1}{V}\left.\frac{\partial V}{\partial T}\right|_P.$$

And finally, if we have an equation of the form

$$dU = đQ + \sum_i j_i \, dx_i,$$

it makes sense to try to write $đQ$ in a similar way as well. And it turns out that if we treat T, the temperature, as a force, we can find a conjugate displacement variable S (called the **entropy**)! So for a reversible process (which we will discuss later), we can write

$$đQ = T\,dS,$$

and now all of our contributions to the internal energy are in the same form.

5.2 Experimental techniques

Now that we've discussed the abstract quantities that we care about measuring, we should ask how we actually measure pressure, volume, and temperature for a given system. We generally deal with **quasi-equilibrium processes**, in which the process is performed sufficiently slowly that the system is always (basically) at equilibrium at any given time. This means that thermodynamic state functions do actually exist throughout our evolution, so we can always calculate well-defined values of P, V, T, and other state functions. And the work done on the system (which is the negative of the work done by the system) is related to changes in thermodynamic quantities, as we wrote above.

Example 34

Let's say we want to measure the potential energy of a rubber band experimentally, and we do this by stretching the rubber band and applying a force.

The idea is that if we are performing the stretching slowly enough, at any point, the force that we apply is basically the same as the internal force experienced by the system. That means that we can indeed take the force we're applying to the rubber band, and we will find that $U = \int F \, d\ell$.

5.3 PV diagrams

Since our state functions are only defined in equilibrium states, all derivatives are also only described in the space of equilibrium states.

Definition 35

In a **PV diagram**, the pressure P is plotted on the y-axis, the volume V is plotted on the x-axis, and every equilibrium state lies somewhere on the graph. The **work** done by or on the system as it transitions from a state I to a state II is defined via

$$W_{by} = -W_{on} = \int_I^{II} P\, dV.$$

The idea here is that (as we discussed with the ideal gas) pressure is force per area, and volume is length times area, so this integral is basically computing $\int F\, dx$. There are ways to go from one state to another without being in equilibrium states along the way as well: for example, if we have sudden free expansion, there is no heat being exchanged and no work done on or by the system, so

$$\Delta Q = \Delta W = 0 \implies U_A = U_B.$$

This tells us that U is a function of only the temperature of the gas (because it doesn't change even when we change our pressure and volume).

Example 36 (Isothermal expansion)

Consider a situation in which a gas moves along an **isotherm** in the PV diagram: for an ideal gas, the equation of this isotherm is just $PV = Nk_BT$.

As the name suggests, we're keeping the temperature of the system constant while we compress the ideal gas. If we start with a volume V_1 and pressure P_1, and we end up at volume V_2 and pressure P_2, then the work done on the system is

$$-\int_{V_1}^{V_2} P\, dV = \int_{V_2}^{V_1} \frac{Nk_BT}{V} dV = Nk_BT(\ln V_2 - \ln V_1) = \boxed{Nk_BT \ln \frac{V_1}{V_2}},$$

since N, k_B, and V are all constants independent of T. If we define $r = \frac{V_1}{V_2}$, and we want to know how much heat is required for this process, we must have

$$0 = dU \implies \boxed{Q = -W = -Nk_BT \ln r},$$

because the internal energy $U(T)$ does not change if T is fixed.

Example 37 (Adiabatic compression)

In an **adiabatic** process, there is no heat added or removed from the system (for example, this happens if we have an isolated container).

Since $\text{đ}Q = 0$, the first law of thermodynamics tells us that

$$dU = \text{đ}W = -P\, dV.$$

Since we have an ideal gas, we also know that

$$PV = Nk_BT, \quad U = \frac{f}{2}Nk_BT,$$

where f is the number of active degrees of freedom for the molecules of the gas. We'll now manipulate these equations a bit: in differential form, we have (using the product rule) that

$$P\,dV + V\,dP = Nk_B dT, dU = \frac{f}{2}Nk_B dT.$$

Combining these equations, we have a relation between changes in internal energy and the state variables:

$$dU = \frac{f}{2}(P\,dV + V\,dP).$$

So now since $dU = -PdV$ from above,

$$-P\,dV = \frac{f}{2}(P\,dV + V\,dP) \implies (f+2)P\,dV + fV\,dP = 0.$$

Definition 38

The **adiabatic index** for a gas with f degrees of freedom is given by

$$\gamma = \frac{f+2}{f}.$$

Now if we rearrange the equation and integrate,

$$0 = \gamma P\,dV + V\,dP \implies \frac{dP}{P} = -\gamma\frac{dV}{V} \implies \frac{P}{P_1} = \left(\frac{V_1}{V}\right)^{\gamma}.$$

This tells us that PV^{γ} is constant, and equivalently that $TV^{\gamma-1}$ and $T^{\gamma}P^{1-\gamma}$ are constant as well. So now the rest is just integration: the work done on the system is

$$W = -\int_{V_1}^{V_2} P\,dV = -P_1V_1^{\gamma}\int_{V_1}^{V_2}\frac{dV}{V^{\gamma}} = \boxed{-P_1V_1^{\gamma}(V_2^{1-\gamma} - V_1^{1-\gamma})}.$$

Plugging in our definition of r,

$$W = \frac{Nk_BT_1}{\gamma - 1}(r^{\gamma-1} - 1).$$

This quantity depends on the number of degrees of freedom in the system, but we can see that in general, the work done for an isothermal process is less than for an adiabatic process! (This is true because the PV curves for $PV^{\gamma} = c$ are "steeper" in a PV diagram than those for $PV = c$, so we are also increasing the temperature, meaning more work is required for us to get the same change in volume.)

Fact 39

It's hard to design an adiabatic experiment, since it's hard to insulate a system completely from its surroundings.

Example 40 (Isometric heating)

In an **isometric** heating, we keep the volume of a system constant while we increase the heat (temperature goes up).

In a PV diagram, this corresponds to moving vertically up. Since there is no change in volume, there is no work done on the system, and thus $dU = \text{đ}Q$: any internal energy change is due to addition of heat. But this is a measurable quantity: we can write

$$\left.\frac{\text{đ}Q}{dT}\right|_V = \left.\frac{\partial U}{\partial T}\right|_V = C_V,$$

and we can measure the change in internal energy to find the specific heat capacity C_V experimentally. Since the energy is dependent only on temperature, **for an ideal gas**,

$$U = \frac{f}{2}Nk_BT \implies \boxed{C_V = \frac{f}{2}Nk_B}.$$

We can also define $\hat{c}_V$ (per molecule) as $\frac{C_V}{N} = \frac{f}{2}k_B$.

Finally, we have one more important process of change along PV diagrams:

Example 41 (Isobaric heating)

This time, we keep the pressure constant and move horizonally in our PV diagram.

Let's differentiate the first law of thermodynamics with respect to T: since $\text{đ}W = -PdV$,

$$\text{đ}Q = dU - \text{đ}W \implies \boxed{\left.\frac{\text{đ}Q}{dT}\right|_P = \left.\frac{\partial U}{\partial T}\right|_P + P\left.\frac{\partial V}{\partial T}\right|_P}.$$

Since pressure is constant, our internal energy U is a function of T and V, and taking differentials,

$$dU = \left.\frac{\partial U}{\partial T}\right|_V dT + \left.\frac{\partial U}{\partial V}\right|_T dV.$$

Dividing through by a temperature differential to make this look more like the equation we had above,

$$\boxed{\left.\frac{\partial U}{\partial T}\right|_P = \left.\frac{\partial U}{\partial T}\right|_V + \left.\frac{\partial U}{\partial V}\right|_T \left.\frac{\partial V}{\partial T}\right|_P}.$$

Combining our equations by substituting the second equation into the first,

$$\left.\frac{\text{đ}Q}{dT}\right|_P = C_V + \left(P + \left.\frac{\partial U}{\partial V}\right|_T\right)\left.\frac{\partial V}{\partial T}\right|_P.$$

But the left side is C_P, so we now have our **general relation between C_P and C_V**

$$\boxed{C_P = C_V + \left(P + \left.\frac{\partial U}{\partial V}\right|_T\right)\left.\frac{\partial V}{\partial T}\right|_P}.$$

Example 42

Let's consider this equation when we have an ideal gas.

Then U is only a function of T, so $\frac{\partial U}{\partial V}$ while keeping T constant is zero. This leaves

$$C_P + C_V + P\left.\frac{\partial V}{\partial T}\right|_P$$

and since we have an ideal gas where $PV = Nk_BT$, $\frac{\partial V}{\partial T}$ at constant P is just $\frac{Nk_B}{P}$, and we have

$$C_P = C_V + Nk_B.$$

Example 43

What if we have an incompressible system, like in solids or liquids?

Then the volume does not change with respect to temperature noticeably, so

$$C_P = C_V + P \left.\frac{\partial V}{\partial T}\right|_P.$$

Defining $\alpha_P = \frac{1}{V} \left.\frac{\partial V}{\partial T}\right|_P$,

$$C_P = C_V + PV\alpha_P.$$

For ideal gasses, $\alpha_P = \frac{1}{T}$, and at room temperature this is about $\frac{1}{300}K^{-1}$. For solids and liquids, the numbers are smaller: $\alpha_P = 10^{-6}K^{-1}$ for quartz and $\alpha_P = 2 \times 10^{-4}K^{-1}$ for water. This essentially means $C_P \approx C_V$ for solids and liquids!

Let's look back again at isometric heating. We found that we had a state function U, which is exactly the amount of heat we added to the system. So in this case, we can directly measure $dU = đQ$. Are there any new state functions such that the change in heat for isobaric heating is the same as the change in that state function? That is, is there some quantity H such that $đQ = dH$? The answer is yes, and we'll discuss this next time! It's enthalpy, and it is $H \equiv U + PV$.

6 February 13, 2019 (Recitation)

Today, we have a few interesting questions, and we'll be using clickers! We can see how we will respond to seemingly simple questions, because professor Ketterle likes to give us twists.

6.1 Questions

Let's start by filling in the details of "adiabatic" processes. In thermodynamics, there are two definitions of "adiabatic" in different contexts.

Fact 44

"dia" in the word means "to pass through," much like "diamagnetic." In fact, in Germany, slide transparencies are called "dia"s as well.

So "adia" means nothing passes through: a system in a perfectly insulated container does not allow transfer of heat. Adiabatic will mean **thermally isolated** in general!

On the other hand, we have adiabatic compression (which we discussed in class), in which we have an equation of the form $PV^\gamma = c$. But what is adiabatic expansion? It sounds like it should just be a decompression: perhaps it is just a reverse of the adiabatic compression process. But this isn't quite right.

In compression, we do the compression slowly: we assume the equation of state for the ideal gas is always valid, so we are always at equilibrium. Indeed, there also exists an adiabatic expansion that is very slow. But in problems like the pset, we can have sudden changes: a sudden, free expansion is not in equilibrium all the time, and it is not reversible!

Fact 45

Adiabatic compression increases the temperature, and slow adiabatic expansion does the opposite. But in sudden free expansion, the internal energy of the system is 0 (as $\text{đ}W = \text{đ}Q = 0$). So the temperature is constant in free expansion.

All three processes have $\text{đ}Q = 0$, since there is no heat transfer. The point is to be careful about whether we have reversible processes, since different textbooks may have different interpretations! We'll talk about entropy later, but the key idea is that the slow adiabatic compression and expansion are **isentropic**: $dS = 0$.

Fact 46

For example, if we change the frequency of our harmonic oscillator in quantum mechanics slowly, so that the energy levels of our system does not jump, that's an adiabatic process in quantum mechanics.

Example 47

Given a PV diagram, what is the graph of sudden, free expansion?

We start and end on the same isotherm, since the temperature is the same throughout the process. But we can't describe the gas as a simple equation of state! In fact, we're not at equilibrium throughout the process, so there is no curve on the PV diagram. After all, the work done $W = \int PdV$ has to be zero. In other words, be careful!

6.2 Clickers

Let's talk about the idea of "degrees of freedom." Molecules can look very different: they can be monatomic, diatomic, or much larger. The degrees of freedom can be broken up into

- center of mass motion
- rotational motion
- vibrational motion.

There are always 3 center of mass degrees of freedom, and let's try to fill in the rest of the table! (We did this using clickers.)

	COM	ROT	VIB	total
atom	3	0	0	3
diatomic	3	2	1	6
CO_2	3	2	**4**	9
H_2O	3	3	**3**	9
polyatomic	3	3	$3N - 6$	$3N$

Some important notes that come out of this:

- There are 2 rotational degrees of freedom for diatomic and straight triatomic molecules: both axes that are not along the line connecting the atoms work. As long as we can distinguish the three directions, though, there are 3 rotational degrees of freedom.

- Here, we count degrees of freedom as normal modes (which is different from 8.223). Recall that in 8.03, we distinguished translational from oscillatory normal modes.
- Water is a triatomic molecule with three modes: the “bending” mode, the “symmetric” stretch, and the “anti-symmetric” stretch.
- Carbon dioxide has 4 modes: the symmetric stretch, the asymmetric stretch, and two bending modes (in both perpendicular axes).

In classical mechanics, if we’re given one particle, we can write 3 differential equations for it: each coordinate gets a Newton’s second law. That’s why we have 3 total degrees of freedom. Similarly, with two particles, we have 6 total degrees of freedom, and the numbers should add up to $3N$ in general. This lets us make sure we don’t forget any vibrational modes!

6.3 Wait a second...

Notice that this definition of “degrees of freedom” is different from what is mentioned in lecture. Thermodynamic degrees of freedom are a whole different story! Now let’s change to using f, the thermodynamic degrees of freedom. Recall that we define $\gamma = \frac{f+2}{f}$.

f measures the number of quadratic terms in the Hamiltonian, and as we will later rigorously derive, if the modes are thermally populated, the energy of each degree of freedom is $\frac{k_B T}{2}$. But the vibrations count twice, since they have both kinetic and potential energy! We’ll also rigorously show this later.

So it’s time to add another column to our table:

	COM	ROT	VIB	total	thermodynamic
atom	3	0	0	3	3
diatomic	3	2	1	6	7
CO_2	3	2	**4**	9	13
H_2O	3	3	**3**	9	12
polyatomic	3	3	$3N-6$	$3N$	$6N-6$

and as we derived in lecture,

$$E = f\frac{k_B T}{2}, C_V = f\frac{k_B}{2}N.$$

However, keep in mind that this concept breaks down when we add too much energy and stop having well-defined molecular structure.

Finally, let’s talk a bit about adiabatic and isothermal compression.

Example 48

Let’s say we do an isothermal compression at T_1, versus doing an adiabatic compression to temperature T_2.

We measure the work it takes to go from an initial volume V_1 to a final volume V_2 under both compressions. The total work is larger for the adiabatic process, since the “area under the curve is larger,” but why is this true intuitively? One way to phrase this is that we press harder, and that means there is more resistance against the work done.

So now let’s prepare a box and do the experiment. We find that there is now no difference: why? (Eliminate the answer of bad isolation.)

- The gas was monatomic with no rotational or vibrational degrees of freedom.

- **Large molecules were put in with many degrees of freedom.**
- The gas of particles had a huge mass.
- This is impossible.

This is because $\gamma = \frac{f+2}{f} \approx 1$ if f is large! Intuitively, the gas is absorbing all the work in its vibrational degrees of freedom instead of actually heating up.

7 February 14, 2019

Remember the pset is due tomorrow night at 9pm. As a reminder, the instructors are only accepting psets on the website: make a pdf file and submit them on LMOD. This minimizes psets getting lost, and it lets TAs and graders make comments directly.

Fact 49

Don't use the pset boxes. I'm not really sure who put them there.

The solutions will become available soon after.

Today, we're starting probability theory. The professor uploaded a file with some relevant information, and delta functions (a mathematical tool) will be covered later on as well.

Also, go to the TA's office hours!

7.1 Review from last lecture

We've been studying thermodynamic systems: we derived an ideal gas law by defining a pressure, temperature, and internal energy of a system. We looked at different processes that allow us to move from one point in phase space (in terms of P, V) to another point.

Thermodynamics came about by combining such motions to form engines, and the question was about efficiency! First of all, let's review the ideas of specific heat for volume and pressure:

- Remember that we discussed an **isometric heating** idea, where the volume stays constant. We could show that $dV = 0 \implies đW = 0 \implies dU = đQ$, which means we can actually get access to the change in internal energy (which we normally cannot do). We also found that

$$\left.\frac{đQ}{dT}\right|_V = \left.\frac{dU}{dT}\right|_V = C_V.$$

- When we have constant pressure (an **isobaric** process), we don't quite have $dU = đQ$, but we wanted to ask the question of whether there exists a quantity H such that $dH = đQ|_P$. The idea is that

$$d(PV) = VdP + PdV, đQ|_P = dU|_P + PdV|_P$$

and this last expression is just $dU|_P + d(PV)_p$ since P is constant. So combining all of these,

$$d(U + PV)|_P = đQ|_P \implies H \equiv U + PV.$$

Definition 50

H is known as the **enthalpy**.

It is useful in the sense that

$$\left.\frac{đQ}{dT}\right|_P = \left.\frac{\partial H}{\partial T}\right|_P = C_P.$$

This is seen a lot in chemistry, since many experiments are done at constant pressure!

We can write a general expression that combines those two:

$$C_P = C_V + \left(P + \left.\frac{\partial U}{\partial V}\right|_T\right)\left.\frac{\partial V}{\partial T}\right|_P$$

and for an ideal gas, this simplifies nicely to $C_P = C_V + Nk_B$.

Last lecture, we also found that

$$\frac{C_P}{C_V} = \frac{f+2}{f} \equiv \gamma,$$

which is the adiabatic index. For a monatomic ideal gas, $f = 3 \implies \gamma = \frac{5}{3}$, and for a diatomic ideal gas, $f = 7 \implies \gamma = \frac{9}{7}$.

7.2 Moving on

Fact 51

If you plot heat capacity C_V per molecule as a function of temperature, low temperatures have $C_V \approx \frac{3}{2}$ (only translational modes), corresponding to $\gamma = \frac{5}{3}$, but this jumps to $C_V \approx \frac{5}{2} \implies \gamma = \frac{7}{5}$ for temperatures between 200 to 1000 Kelvin. Hotter than that, vibrational modes start to come in, and C_V increases while γ approaches 1.

In a **Carnot engine**, we trace out a path along the PV diagram. How can we increase the efficiency?

There are two important principles here: energy is conserved, but entropy is always increasing.

Definition 52 (Unclear)

Define the **entropy** as

$$\Delta S = \frac{Q}{T}.$$

But this doesn't give very much physical intuition of what entropy really is: it's supposed to be some measure of an "ignorance" of our system. Statistical physics is going to help us give an information theoretic definition later:

$$S = -k_B \langle \ln P_i \rangle,$$

which will make sense as we learn about probability theory in the next three or four lectures!

7.3 Why do we need probability?

Almost all laws of thermodynamics are based on observations of macroscopic systems: we're measuring thermal properties like pressure and temperature, but any system is still inherently made up of atoms and molecules, so the motion is described by more fundamental laws, either classical or quantum.

So we care about **likelihoods**: how likely is it that particles will be in a particular microscopic state?

7.4 Fundamentals

Definition 53

A **random variable** x has a set of possible outcomes

$$S = \{x_1, x_2, \cdots, \}.$$

(This set is not necessarily countable, but I think this is clear from the later discussion.)

This random variable can be either **discrete** or **continuous**.

Example 54 (Discrete)

When we toss a coin, there are two possible outcomes: $S_{\text{coin}} = \{H, T\}$. When we throw a die, $S_{\text{die}} = \{1, 2, 3, 4, 5, 6\}$.

Example 55 (Continuous)

We can have some velocity of a particle in a gas dictated by

$$S = \{-\infty < v_x, v_y, v_z < \infty\}.$$

Definition 56

An **event** is a subset of some outcomes, and every event is assigned a **probability**.

Example 57

When we roll a die, here are some probabilities:

$$P_{\text{die}}(\{1\}) = \frac{1}{6}, P_{\text{die}}(\{1, 3\}) = \frac{1}{3}.$$

Probabilities satisfy three important conditions:

- **positivity**: any event has a nonnegative real probability $P(E) \geq 0$.
- **additivity**: Given two events A and B,

  $$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

  where $A \cup B$ means "A or B" and $A \cap B$ means "A and B".
- **normalization**: $P(S) = 1$, where S is the set of all outcomes. In other words, all random variables have some outcome.

There are two ways to find probabilities:

- **Objective approach**: given a random variable, do many trials and measure the result each time. After N of them, we have probabilities N_A for each event A: this is just the number of times A occurs, divided by N. In particular, as we repeat this sufficiently many times,

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}.$$

- **Subjective approach**: We assign probabilities due to our uncertainty of knowledge about the system. For example, with a die, we know all six outcomes are possible, and in the absence of any prior knowledge, they should all be equally probable. Thus, $P(\{1\}) = \frac{1}{6}$.

We'll basically do the latter: we'll start with very little knowledge and add constraints like "knowledge of the internal energy of the system."

7.5 Continuous random variables

Fact 58

We'll mostly be dealing with these from now on, since they're are what we'll mostly encounter in models.

Let's say we have a random variable x which is real-valued: in other words,

$$S_X = \{-\infty < x < \infty\}.$$

Definition 59

The **cumulative probability function** for a random variable X, denoted $F_X(x)$, is defined as the probability that the outcome is less than or equal to x:

$$F_X(x) = \Pr[E \subset [-\infty, x]].$$

Note that $F_X(-\infty) = 0$ and $F_X(\infty) = 1$, since x is always between $-\infty$ and ∞.

Definition 60

The **probability density function** for a random variable X is defined by

$$p_X(x) \equiv \frac{dF_X}{dx}$$

In particular,

$$p_X(x)dx = \Pr\left[E \subset [x, x + dx]\right].$$

It's important to understand that

$$\int_{-\infty}^{\infty} p_X(x)dx = 1,$$

since this is essentially the probability over all x.

Fact 61

The units or dimension of $p_X(X)$ is the reciprocal of the units of X.

Note that there is no upper bound on p_X; it can even be infinity as long as p is still integrable.

Definition 62

Let the **expected value** of any function $f(x)$ of a random variable x be

$$\langle F(x)\rangle = \int_{-\infty}^{\infty} F(x)p(x)dx.$$

As a motivating example, the expected value of a discrete event is just

$$\langle X\rangle = \sum_i p_i x_i,$$

so this integral is just an "infinite sum" in that sent.

7.6 More statistics

Definition 63

Define the **mean** of a random variable x to be

$$\langle x\rangle = \int_{-\infty}^{\infty} xp(x)dx.$$

For example, note that

$$\langle x - \langle x\rangle\rangle = 0.$$

In other words, the difference from the average is 0 on average, which should make sense. But we can make this concept into something useful:

Definition 64

Define the **variance** of a random variable x to be

$$\text{var}(x) = \langle (x - \langle x\rangle)^2\rangle.$$

This tells something about the spread of the variable: basically, how far away from the mean are we? Note that we can expand the variance as

$$\text{var}(x) = \langle x^2 - 2x\langle x\rangle + \langle x\rangle^2\rangle = \langle x^2\rangle - \langle x\rangle^2.$$

Fact 65 (Sidenote)

The reason we square instead of using an absolute value is that the mean is actually the value that minimizes the sum of the squares, while the median minimizes the sum of the absolute values. The absolute value version is called "mean absolute deviation" and is less useful in general.

We're going to use this idea of variance to define other physical quantities like diffusion later!

Definition 66

Define the **standard deviation** as

$$\sigma(x) = \sqrt{\text{var}(x)}.$$

With this, define the **skewness** as a dimensionless metric of asymmetry:

$$\left\langle \frac{(x - \langle x \rangle)^3}{\sigma^3} \right\rangle,$$

and define the **kurtosis** as a dimensionless measure of shape (for a given variance).

$$\left\langle \frac{(x - \langle x \rangle)^4}{\sigma^4} \right\rangle,$$

Let's look at a particle physics experiemnt to get an idea of what's going on:

$$e^+e^- \to \mu^+\mu^-.$$

Due to quantum mechanical effects, there is some probability distribution for θ, the angle of deflection:

$$p(\theta) = c\sin\theta(1+\cos^2\theta), 0 \le \theta \le \pi.$$

To find the constant, we normalize with an integral over the range of θ:

$$1 = c\int_0^\pi \sin\theta(1+\cos^2\theta)d\theta.$$

We will solve this with a u-substitution: letting $x = \cos\theta$,

$$1 = c\int_{-1}^1 (1+x^2)dx \implies 1 = \frac{8}{3}c \implies c = \frac{3}{8}$$

So our probability density function is

$$p(\theta) = \frac{3}{8}\sin\theta(1+\cos^2\theta)$$

Fact 67

This has two peaks and is symmetric around $\frac{\pi}{2}$. Thus, the mean value of θ is $\langle\theta\rangle = \frac{\pi}{2}$, and σ is approximately the distance to the peak.

We can calculate the standard deviation exactly:

$$\langle \boxed{\theta^2} \rangle = \frac{3}{8}\int_0^\pi (1+\cos^2\theta)\sin\theta \cdot \boxed{\theta^2} d\theta = \frac{\pi^2}{4} - \frac{17}{9},$$

and therefore

$$\text{var}(\theta) = \langle\theta^2\rangle - \langle\theta\rangle^2 \approx 0.579 \implies \sigma \approx 0.76.$$

Finally, let's compute the cumulative probability function:

$$F(\theta) = \int_0^\theta \frac{3}{8}\sin\theta(1+\cos^2\theta)d\theta = \frac{1}{8}(4 - 3\cos\theta - \cos^3\theta).$$

This has value 0 at $\theta = 0$, $\frac{1}{2}$ at $\theta = \frac{\pi}{2}$, and 1 at $\theta = \pi$.

Next time, we will talk about discrete examples and start combining discrete and continuous probability. We'll also start seeing Gaussian, Poisson, and binomial distributions!

8 February 19, 2019 (Recitation)

A guest professor is teaching this recitation. We're going to discuss delta functions as a mathematical tool, following the supplementary notes on the website.

There are multiple different ways we can represent delta functions, but let's start by considering the following:

Definition 68

Define

$$\delta_\varepsilon(x) = \frac{1}{\sqrt{2\pi}\varepsilon} \exp\left[-\frac{x^2}{2\varepsilon^2}\right].$$

This is a Gaussian (bell curve) distribution with a peak at $x = 0$ and an inflection point at $\pm\varepsilon$. It also has the important property

$$\int_{-\infty}^{\infty} \delta_\varepsilon(x) = 1,$$

so it is already normalized. This can be shown by using the fact that

$$I = \int_{-\infty}^{\infty} dx e^{-\alpha x^2} \implies I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdy e^{-\alpha(x^2+y^2)}$$

and now switch to polar coordinates: since $dxdy = rdrd\theta$,

$$I^2 = \int_0^\infty drr \int_0^{2\pi} e^{-\alpha r^2} = \frac{\pi}{\alpha}$$

using a u-substitution. So δ_ε is a function with area 1 regardless of the choice of ε. However, ε controls the width of our function! So if ε goes down, the peak at $x = 0$ will get larger and larger: in particular, $\delta_\varepsilon(0) = \frac{1}{\sqrt{2\pi}\varepsilon}$ goes to ∞ as $\varepsilon \to 0$.

So we have a family of such functions, and our real question is now what we can do with integration? What's

$$\int_{-\infty}^{\infty} dx\delta_\varepsilon(x)f(x)?$$

For a specific function and value of ε, this may not be a question you can answer easily. But the point is that if we put an arbitrary function in for f, we don't necessarily know how to do the integration. What can we do?

Well, let's think about taking $\varepsilon \to 0$. Far away from $x = 0$, $\delta_\varepsilon(x)f(x)$ is essentially zero. If we make δ_ε extremely narrow, we get a sharp peak at $x = 0$: zooming in, f is essentially constant on that peak, so we're basically dealing with

$$\int_{-\varepsilon}^{\varepsilon} dxf(0)e^{-\frac{x^2}{2\varepsilon^2}}\frac{1}{\sqrt{2\pi}\varepsilon} = f(0)\cdot 1 = f(0).$$

So the idea is that we start with a particular family of functions and take $\varepsilon \to 0$, and this means that δ_ε is a pretty good first attempt of a "sharp peak."

Definition 69

Let the **Dirac delta function** δ satisfy the conditions

- $\delta(x - x_0) = 0$ for all $x \neq x_0$.
- $\int_{-\infty}^{\infty} dx \delta(x - x_0) = 1$, where the integral can be over any range containing x_0.
- $\int dx \delta(x - x_0) f(x) = f(x_0)$, again as an integral over any range containing x_0.

This seems pretty silly: if we already know $f(x)$, why do we need its evaluation at a specific point by integrating? We're just evaluating the function at $x = 0$. It's not at all clear why this is even useful. Well, the idea is that it's often easier to write down integrals in terms of the delta function, and we'll see examples of how it's useful later on.

For now, let's keep looking at some more complicated applications of the delta function. What if we have something like

$$\int dx \delta(g(x)) f(x)?$$

We know formally what it means to replace $\delta(x)$ with $\delta(x - x_0)$, but if we have a function g with multiple zeros, we could have many peaks: what does that really mean, and how tall are the peaks here? This is useful because we could find the probability that $g(x, y, z) = c$ by integrating

$$\int p(x, y, z) \delta(g(x, y, z) - c) dx dy dz$$

and this answer is not quite obvious yet. So we're going to have to build this up step by step.

Let's start in a simpler case. What's

$$\int dx f(x) \delta(cx)?$$

We can do a change of variables, but let's not rush to that. Note that $\delta_\varepsilon(-x) = \delta_\varepsilon(x)$, and similarly $\delta(x) = \delta(-x)$: this is an even function. So replacing $y = cx$,

$$= \int dy \frac{1}{|c|} f\left(\frac{y}{c}\right) \delta(y) = \frac{1}{|c|} f(0).$$

So we get back $f(0)$, just with some extra constant factor. Be careful with the changing integration limits, both in this example and in general: that's why we have the absolute value in the denominator. In general, the delta function "counts" things, so we have to make sure we don't make bad mistakes with the sign!

Similar to the above, we can deduce that linear functions give nice results of the form

$$\int dx \delta(cx - a) f(x) = \frac{1}{|c|} f\left(\frac{a}{c}\right).$$

But this is all we need! Remember that we only care about δ when the value is very close to 0. So often, we can just make a linear approximation! $f(x)$ looks linear in the vicinity of x_0, and there's a δ peak at x_0. So if we make the Taylor expansion $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$, we have found everything relevant to the function that we need.

Note that by definition, $\delta(g(x)) = 0$ whenever $g(x) \neq 0$. Meanwhile, if $g(x_i) = 0$,

$$g(x) \approx g(x_i) + g'(x_i)(x - x_i) = g'(x_i)(x - x_i).$$

So that means we can treat

$$\delta(g(x)) = \sum_i \delta(g'(x_i)(x - x_i))$$

where we are summing over all zeros of the function!

Fact 70

Remember that δ is an everywhere-positive function, so $\delta(g(x))$ cannot be negative either.

Well, we just figured out how to deal with $\delta(g(x))$ where g is linear! So

$$\int dx f(x)\delta(g(x)) = \sum_i \frac{1}{|g'(x_i)|} f(x_i).$$

So at each point x_i where g is 0, we just take $f(x_i)$ and modify it by a constant. Now this function is starting to look a lot less nontrivial, and we'll use it to do a lot of calculations over the next few weeks.

Example 71

Let's say you want to do a "semi-classical density of states calculation" to find the number of ways to have a particle at a certain energy level. Normally, we'd do a discrete summation, but what if we're lazy?

Then in the classical case, if u is the velocity, we have an expression of the form

$$f(E) = \int p(u)\delta\left(E - \frac{mu^2}{2}\right) du.$$

To evaluate this, note that the δ function is zero at $u = \pm\sqrt{\frac{2E}{m}}$, and the derivative

$$g'(u) = -mu \implies |g'(u_\pm)| = \sqrt{2mE},$$

so the expression is just equal to

$$f(E) = \frac{1}{\sqrt{2mE}}(p(u_+) + p(u_-)).$$

This is currently a one-dimensional problem, so there's only two values of u. In highest dimensions, we might be looking at something like

$$\int d^3\vec{u} p(\vec{u})\delta\left(E - m\frac{u_x^2 + u_y^2 + u_z^2}{2}\right).$$

Now the zeroes lie on a sphere, and now we have to integrate over a whole surface!

By the way, there are different ways to formulate the delta function. There also exists a **Fourier representation**

$$\delta(x) = \int_{-\infty}^{\infty} dk e^{ikx}.$$

It's not obvious why this behaves like the delta function, but remember e^{ikx} is a complex number of unit magnitude. Really, we care about

$$\int \delta(x) f(x) dx,$$

and the point is that for any choice of x other than 0, we just get a spinning unit arrow that gives net zero contribution. But if $x = 0$, e^{ikx} is just 1, so this starts to blow up just like the δ_ε function.

There also exist a Lorentzian representation

$$\delta_\varepsilon = \frac{1}{\pi}\frac{\varepsilon}{x^2 + \varepsilon^2}$$

and an exponential representation

$$\delta_\varepsilon = \frac{1}{2\varepsilon} e^{|x|/\varepsilon}.$$

The point is that there are many different families of functions to capture the intended effect (integrates to 1), but as all of them get sharper, they end up having very similar properties for the important purposes of the delta function.

9 February 20, 2019 (Recitation)

It is a good idea to talk about the concept of an exact differential again, and also to look over delta functions.

9.1 Questions

We often specify a system (like a box) with a temperature, volume, and pressure. We do work on the system when the volume is reduced, so $dW = -PdV$.

When we have a dielectric with separated charges, we can orient the dipoles and get a dipole energy

$$\propto \vec{E} \cdot \vec{p},$$

where $\vec{p}$ is the dipole moment. We now have to be careful if we want to call this potential energy: are we talking about the energy of the whole system, the external field, or something else?

Well, the differential energy can be written as $dW = EdP$: how can we interpret this? Much like with $-PdV$, when the polarization of the material changes, the electrostatic potential energy changes as well.

So now, what's the equation of state for an electrostatic system? Can we find an equation like $PV = nRT$ to have $E(T, P)$? Importantly, note that some analogies to break down: the electric field E is sort of necessary (from the outside) to get a polarization P.

So if we consider an exact differential, and we're given

$$\left.\frac{\partial E}{\partial P}\right|_T = \frac{T}{AT + B}, \left.\frac{\partial E}{\partial T}\right|_P = \frac{BP}{(AT + B)^2},$$

we know everything we could want to know about the system. First, we should show that these do define an equation of state: is

$$dE = \frac{T}{AT + B} dP + \frac{BP}{(AT + B)^2} dT$$

an exact differential? Well, we just check whether

$$\frac{\partial}{\partial T} \frac{T}{AT + B} = \frac{\partial}{\partial P} \frac{BP}{(AT + B)^2}.$$

Once we do this, we can integrate $\frac{\partial E}{\partial P}$ with respect to P to get E up to a function of T, and then differentiate with respect to T to find that unknown function T.

Next, let's talk a bit more about mean and variance. We can either have a set of possible (enumerable, discrete) events $\{p_i\}$, or we could have a probability density $p(x)$. The idea with the density function is that

$$p(x)dx = \Pr[\text{in the range } [x, x + dx]].$$

Remember that probabilities must follow a normalization, which means that

$$\sum p_n = 1 \text{ or } \int p(x)dx = 1.$$

What does it mean to have an average value? In the discrete case, we find the average as

$$\langle n \rangle = \sum_i i p_i,$$

since an outcome of i with probability p_i should be counted p_i of the time. Similarly, the continuous case just uses an integral:

$$\langle x \rangle = \int x p(x) dx.$$

Note that we can replace n and x with any arbitrary functions of n and x. Powers of n and x are called moments, so the mean value is the "first moment." (This is a lot like taking the second "moment of inertia" $\int r^2 \rho(r) dr$.) Then the variance is an average value of

$$\langle (n - \langle n \rangle)^2 \rangle = \langle n^2 - 2n\langle n \rangle + \langle n \rangle^2 \rangle = \langle n^2 \rangle - \langle n \rangle^2.$$

Let's look at another situation where we have a probability density

$$\frac{dP}{dw} = p(w).$$

9.2 Delta functions

Here's a question: what is the derivative

$$\delta'(x)?$$

Well, let's start with some related ideas:

$$\delta(10x) = \frac{1}{10}\delta(x),$$

and this might make us cringe a bit since δ is mostly infinite, but it works for all purposes that we are using it. The important idea is to not always think about infinity: we could consider the delta function to be a rectangle of width ε and height $\frac{1}{\varepsilon}$. This makes it seem more like a real function.

So now, if we take any function $f(x)$, $f(10x)$ is just a function that keeps the maximum constant but shrinks the width by a factor of 10. Well, if we integrate over $f(10x)$, we'll get a factor of $\frac{1}{10}$ less, and that's how we should understand $\delta(10x)$.

Is $\delta'(x)$ defined, then? Let's say we have a triangle function with peak $\frac{1}{\varepsilon}$ and width from $-\varepsilon$ to ε. The derivative of this function is not defined at $x = 0$! So $\delta'(x)$ doesn't necessarily need to be defined. The idea is that we can take $\varepsilon \to 0$ using any representation of a real function, and a derivative would have to be well-defined across all representations: that just doesn't happen here.

Curiously, though, if we used the triangle function, we can actually represent the derivative as a delta function itself, because the derivative is a large number from $-\varepsilon$ to 0 and from 0 to ε:

$$\delta'(x) = \frac{1}{\varepsilon}\delta\left(-\frac{\varepsilon}{2}\right) - \frac{1}{\varepsilon}\delta\left(\frac{\varepsilon}{2}\right),$$

where the $\frac{1}{\varepsilon}$ factor is just for normalization, since the area of the rectangle for the derivative is $\varepsilon \cdot \frac{1}{\varepsilon^2}$. So it seems that

$$\delta'(x) = -\frac{1}{\varepsilon}\left(\delta\left(\frac{\varepsilon}{2}\right) - \delta\left(-\frac{\varepsilon}{2}\right)\right).$$

It's okay, though: delta functions always appear under an integral. (So continuity is important, but not necessarily differentiability.) This means that if we're integrating this with a function $f(x)$,

$$\int f(x)\delta'(x)dx = -\frac{1}{\varepsilon}\int\left(\delta\left(\frac{\varepsilon}{2}\right)-\delta\left(-\frac{\varepsilon}{2}\right)\right)f(x)dx = -\frac{1}{\varepsilon}f\left(\frac{\varepsilon}{2}\right)-f\left(-\frac{\varepsilon}{2}\right) = -f'(0).$$

But the idea is that we want to be faster at manipulating such things. What if we integrated by parts? Then

$$\int f(x)\delta'(x) = -\int f'(x)\delta(x)dx + f(x)\delta(x)|_{-\infty}^{\infty}.$$

The delta function is mathematically zero, so the boundary term disappears (unless we have a bad Lorentzian or other description of the delta function). But now this just gives

$$-\int f'(x)\delta(x)dx = -f'(0).$$

This is maybe how we should use delta functions, but it's still important to have confidence that what we're doing is correct!

Finally, let's ask one more question. We can store energy by pressing air into an underground cave, and we can do that in two ways: adiabatic and isothermal. If we compare the two situations, where does the energy go?

In the isothermal case, the internal energy is the same. So isothermal compression is just transferring the energy as heat to the surrounding ground! Is there a way to retrieve it? (Hint: the process may be reversible if we do it slow enough!)

10 February 21, 2019

Today we're going to continue learning about probability. As a reminder, we're learning probability because there's a bottom-up and top-down approach to statistical physics: thermodynamics gives us state functions that tell us physical properties of the world around us, and we can connect those with microscopic atoms and molecules that actually form the system. Probability allows us to not just follow every particle: we can just think about general distributions instead!

We'll discuss some important distributions today and start our connections between probability distributions and physical quantities. We'll eventually get to the Central Limit Theorem!

10.1 A discrete random variable

Consider a weighted coin such that

$$P(H) = \frac{5}{8}, P(T) = \frac{3}{8}.$$

This is a "biased distribution." Let's say that every time we get a head, we gain \$1, and every time we get a tail, we lose \$1. Letting x be our net money, our discrete probability distribution $P(x)$ satisfies

$$P(1) = \frac{5}{8}, P(-1) = \frac{3}{8}.$$

Here are some important statistics:

- $P(1) + P(-1) = 1$.
- $\langle x\rangle = P(1)\cdot 1 + P(-1)\cdot(-1) = \frac{5}{8} - \frac{3}{8} = \frac{1}{4}$.
- $\langle x^2\rangle = P(1)1^2 + P(-1)1^2 = 1$, so the variance of this distribution is

$$\text{var}(x) = \langle x^2\rangle - \langle x\rangle^2 = \frac{15}{16}.$$

This is a large variance, since most events are far away from the mean!

- To find the probability density function $p(x)$, we can use delta functions:

$$p(x) = \frac{5}{8}\delta(x-1) + \frac{3}{8}\delta(x+1)$$

  Note that if we integrate $p(x)dx$ over any interval containing 1 but not -1, we get a probability of $\frac{5}{8}$, which is what we want.

In general, if we have a probability density function that has both discrete and continuous parts, we can write it as

$$p(x) = f(x) + \sum_{j=1}^{M} p_i \delta(x - x_j).$$

See the Weibull distribution, as well as the Xenon lamp spectrum!

10.2 Important probability distributions

We're going to discuss the Gaussian, Poisson, and Binomial distributions. The idea is that the limit of a Binomial distribution will converge to a Poisson distribution, which will then converge to a Gaussian. We'll see the second part today as part of the Central Limit Theorem!

10.3 Gaussian distribution

Definition 72

The probability density function for a **Gaussian** with standard deviation σ is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right].$$

This distribution has mean a and variance σ^2; let's check for normalization to make sure this is indeed a valid distribution. So

$$\int_{-\infty}^{\infty} p(x)dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right],$$

and now defining

$$y \equiv \frac{x-a}{\sqrt{2}\sigma} \implies dy = \frac{1}{\sqrt{2}\sigma} dx.$$

Substituting in, we end up with

$$\frac{1}{\sqrt{2\pi}\sigma} \cdot \sqrt{2}\sigma \int_{-\infty}^{\infty} e^{-y^2} dy.$$

But the integral is known to be $\sqrt{\pi}$ (by the polar substitution trick), and therefore our integral of $p(x)$ is 1, and indeed we have a normalized distribution.

So the cumulative distribution function is

$$F(x) = \int_{-\infty}^{x} d\xi \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\xi-a)^2}{2\sigma^2}\right].$$

Denoting $\frac{\xi-a}{\sqrt{2}\sigma} = z$ as before, we can substitute again to find

$$F(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{(x-a)/(\sqrt{2}\sigma)} dz e^{-z^2}.$$

Since our probability function is normalized, we can also write this as

$$F(x) = 1 - \frac{1}{\sqrt{\pi}} \int_{(x-a)/(\sqrt{2}\sigma)}^{\infty} dz e^{-z^2}$$

Let erfc, the complementary error function, be defined as

$$\text{erfc}(\eta) = \frac{2}{\sqrt{\pi}} \int_{\eta}^{\infty} dz e^{-z^2}.$$

So that means our cumulative distribution can be written as

$$F(x) = 1 - \frac{1}{2} \,\text{erfc}\left(\frac{x-a}{\sqrt{2}\sigma}\right).$$

It's important to note that the shape of $p(x)$ is the familiar "bell curve" shape. When $a = 0$ (so the curve is centered at $x = 0$),

$$p(0) = \frac{1}{\sqrt{2\pi}\sigma}$$

and we can also compute that $P(\sigma) \approx 0.61P(0), P(2\sigma) \approx 0.135P(0)$.

What exactly is the significance of σ? Well, for any a and σ in the Gaussian,

$$\int_{a-\sigma}^{a+\sigma} p(x)dx \approx 0.68,$$

meaning that about 68% of the time, a random sample from the Gaussian distribution will be within σ of the mean. Similarly,

$$\int_{a-2\sigma}^{a+2\sigma} p(x)dx \approx 0.95,$$

and that means that 95 percent of all measurements are within 2σ of the mean.

Example 73

Consider a measurement of the magnetic moment of a muon

$$m = \frac{ge\hbar}{2M_\mu c}.$$

At first, we expect that $g \approx 2$ theoretically. However, after many measurements, we get a Gaussian distribution for $(g-2)_\mu$:

$$(g-2)_\mu = (116591802 \pm 49) \times 10^{-11},$$

where the first term is the mean and the part after the ± is σ. Theoretical calculations actually end up giving

$$(g-2)_\mu = (116592089 \pm 63) \times 10^{-11},$$

and these distributions are actually significantly different: the discrepancy is still a point of ongoing research! The idea is that this measurement of σ allows us to compare two different distributions.

10.4 Poisson distribution

For random variables X and Y, if they have probability distributions $p(x)$ and $p(y)$, then

$$X, Y \text{ statistically independent} \implies p(x,y) = p(x)p(y).$$

Let's start with an example. Given a random student, the probability that a student is born in May is $\frac{31}{365.25} \approx 0.0849$. Meanwhile, the probability of being born between 9 and 10 in the morning is $\frac{1}{24} \approx 0.0417$. So the probability of being born between 9 and 10 in the morning in May is

$$0.0849 \times 0.0417 = 3.54 \times 10^{-3}.$$

We need this to introduce the idea of a Poisson distribution! This is important for rare events with low probability. Here are two important ideas:

- The probability of an event happening exactly once in the interval $[t, t+dt]$ is proportional to dt as $dt \to 0$: $dp = \lambda dt$ for some λ.
- The probability of events in different events are independent of each other.

If we have these two conditions satisfied, the idea is that we can subdivide a time interval of length T into small intervals of length dt. In each interval, the probability that we observe an event is equal and independent to all the other ones!

Definition 74

Then the probability that we observe a total of exactly n events in an interval time T is given by the **Poisson distribution** $P_n(T)$.

Let's try to compute p_n. We break T into N bins of length dt, so $dt = \frac{T}{N}$, in such a way that the probability of getting two events in the same bin (small time interval) is negligible. Then

$$dP = \lambda dt = \frac{\lambda T}{N} \ll 1.$$

To compute the probability of finding n events, first we find the probability of computing no events, then 1 event, and so on. Note that the probability of observing no event in an interval dt is $1 - \frac{\lambda T}{N}$, so the probability of observing no events overall is

$$\lim_{N\to\infty} \left(1 - \frac{\lambda T}{N}\right)^N$$

since we must have no event observed in all N intervals. By definition of an exponential, this is just

$$P_0(T) = e^{-\lambda T}.$$

Next, find the probability of observing exactly 1 event in time interval T. There are N different places in which this one event can happen, and the probability that it happens is $\frac{\lambda T}{N}$. Then the other $N-1$ intervals must have no event happen, so this is

$$P_1(T) = \lim_{N\to\infty} N \cdot \frac{\lambda T}{N}\left(1 - \frac{\lambda T}{N}\right)^{N-1} = \lambda T \lim_{N\to\infty}\left(1 - \frac{\lambda T}{N}\right)^{N-1}.$$

This gives, again by the definition of an exponential,

$$P_1(T) = \lambda T e^{-\lambda T}.$$

Let's do another example: what about two events? We pick which two intervals are chosen, and then

$$P_k(T) = \lim_{N\to\infty} \binom{N}{2} \cdot \left(\frac{\lambda T}{N}\right)^2 \left(1 - \frac{\lambda T}{N}\right)^{N-2} = \frac{(\lambda T)^2}{2} e^{-\lambda T}.$$

In general, the probability of k events happening is going to be

$$P_k(T) = \lim_{N\to\infty} \binom{N}{k} \cdot \left(\frac{\lambda T}{N}\right)^k \left(1 - \frac{\lambda T}{N}\right)^{N-k} \implies P_k(t) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}.$$

It's important to note that this is a discrete distribution!

Let's check some statistics for our probability distribution function. First of all, is it normalized? Well,

$$\sum_{n=0}^{\infty} p_n(T) = e^{-\lambda T} \sum_{n=0}^{\infty} \frac{(\lambda T)^n}{n!} = e^{-\lambda T} e^{\lambda T} = 1,$$

so the Poisson distribution is indeed normalized.

Next, let's find the mean:

$$\langle n \rangle = \sum_{n=0}^{\infty} n p_n(T) = e^{-\lambda T} \sum_{n=0}^{\infty} \frac{n(\lambda T)^n}{n!}$$

Denoting $Z \equiv \lambda T$, this expression can be written as

$$\langle n \rangle = e^{-Z} \sum_{n=0}^{\infty} \frac{nZ^n}{n!} = Ze^{-Z} \sum_{n=1}^{\infty} \frac{z^{n-1}}{(n-1)!} = e^{-Z} Z \cdot e^Z = Z.$$

So the mean of the Poisson distribution is $\langle n \rangle = \lambda T$, which shouldn't be that surprising: it's saying that if events have a probability $\frac{1}{N}$ of happening, they happen on average once per N.

Finally, let's find the variance: we'll leave this as an exercise, but the idea is to start by computing

$$\langle n(n-1) \rangle = \langle n^2 - n \rangle.$$

It turns out the variance is also λT, and this is an interesting relationship between the mean and variance! We'll introduce a dimensionless quantity

$$\frac{\sigma(n)}{\langle n \rangle}$$

which meausures the width of the distribution. Well, note that as $T \to \infty$ for the Poisson distribution, this goes to 0, so the distribution becomes more and more spiked around λT. It turns out that this approaches a Gaussian distribution! How can we check that?

Taking $T \to \infty$,

$$\lambda T \gg 1 \implies n \gg 1,$$

and we want to find the probability $P_n(\lambda T)$. Denoting $\lambda T \equiv Z$, we expand around the maximum, and we're going to look at the log of the function. By Stirling's approximation, $\ln n! \sim n \ln n - n + \ln(2\pi n)$ as $n \to \infty$, so

$$P_n(Z) = \frac{Z^n}{n!} e^{-Z} \implies \ln P_n(Z) = n \ln Z - Z - \ln n! \approx n \ln Z - Z - n \ln n + n - \frac{1}{2} \ln 2\pi n.$$

The maximum of this function can be found by taking the derivative,

$$\frac{\partial}{\partial n} \ln P_n = \ln Z - \ln n - \frac{1}{2n},$$

and ignoring the $\frac{1}{2n}$ term, we can say that $n_0 = Z$ at the maximum. Doing a Taylor expansion about n_0,

$$\ln P_n(Z) = \ln p_{n_0}(z) + \frac{(n-n_0)^2}{2}\frac{\partial^2}{\partial n^2}P_n(z)\bigg|_{n=n_0},$$

and taking the exponential of both sides, we find that

$$P_n(z) = \frac{1}{\sqrt{2\pi z}}\exp\left[-\frac{(n-z)^2}{2z}\right]$$

which is a Gaussian with standard deviation $\sqrt{z}$ and mean z, as desired! This is our first instance of the Central Limit Theorem, a powerful tool to deal with large numbers.

11 February 25, 2019 (Recitation)

Professor Ketterle was writing a paper on the new definition of the kilogram. Let's spend a few minutes talking about that!

11.1 The kilogram

There's a law that is taking effect in May. Currently, the kilogram is a physical object in France: it's what weighs as much as the "original Paris kilogram." There are copies around the world.

But artifacts like this have discrepancies! There might be diffusion of atoms or fingerprints, so at the microgram level, there are still deviations. This wasn't a problem, but now people have determined Planck's constant with an error of

$$\frac{\Delta h}{h} \approx 10^{-8},$$

and the error is limited from the mass deviation! So this is pretty inconvenient.

Fact 75

So instead, why not define h to be $6.62\cdots \times 10^{-34}$ Joule-seconds?

Now we've turned it around: mass is now determined in terms of h, instead of the other way around!

Question 76. *Why exactly does h determine mass?*

Since $E = h\nu$, and we can consider the frequency of transition in Cesium $= 9.1\cdots \times 10^9$ Hertz (which actually now defines units of time and frequency). With this, now we can measure energy (as either kinetic or rest energy). The idea is that all constants are now defined in terms of c, the speed of light, and h, Planck's constant!

So more precisely, we start with the frequency of Cesium, and we define a second to be $9.1\cdots$ GHz. But this means that if we take our value h, the mass of a photon will be

$$m = \frac{h\nu_{CS}}{c^2}.$$

How do we measure the mass of a photon? Well, a Cesium atom in the upper state is slightly heavier than a Cesium atom in the lower state! That gives us the change in mass Δm, which is on the order of 10^{-40} kilograms.

Fact 77

When a system loses energy, it often loses mass!

It's a tiny effect, but it's important for special relativity. So we can now set up a balance, where 10^{40} Cesium atoms are set up with spins in the "upper state." On the other side, we have the same number of Cesium atoms, but the spins are in the ground state. Then any substance that balances the scale is exactly one kilogram!

11.2 Poisson distribution

Think of the Poisson distribution as modeling a bunch of atoms radiating with some decay rate. If N is the number of particles, λ is the decay rate, and we observe for some rate dt, we have an expectation

$$\langle n \rangle = N\lambda dt$$

for the number of observed events in time dt. We can do combinatorics to find that

$$p_n = \frac{\langle n \rangle^n}{n!} e^{-\langle n \rangle}$$

This is a prettier way to write the Poisson distribution, and it shows that the whole distribution is based entirely on the value of $\langle n \rangle$.

Based on this, let's consider the concept of **shot noise**, which comes from us trying to count something that is random.

When we have a random event like raindrops falling on a roof, you hear a random noise with fluctuations. This is because rain doesn't come as a stream: it's big droplets. So the shot noise comes from the fact that we have a stream of mass (or radioactively decaying particles) that are quantized. So sometimes we have a few more or a few less than expected.

So if we are trying to observe 100 events, the expectation value is 100 ± 10. This is because the variance $\text{var}(n) = \langle n \rangle$, and therefore σ, the standard deviation, is $\sqrt{\langle n \rangle}$. So it's important to remember that $\pm\sqrt{n}$ idea! Basically, the Poisson distribution looks almost normal, and the inflection point occurs around $\sqrt{\langle n \rangle}$ from the mean.

This leads to the next question: as you observe larger expectations, is the shot noise larger or smaller? Well, if we're doing repeated measurements to determine some quantity, our precision goes as

$$\frac{\sigma(n)}{\langle n \rangle} \propto \frac{1}{\sqrt{n}}.$$

So if we measure 10 times longer, we are $\sqrt{10}$ times more accurate. But this is often the best we can do! Much along the same line, we want to do experiments with large counting rate to get higher expectations.

So if we measure the noise in an electric current, the number of electrons that pass through gives shot noise as well. We'll find that

$$i\delta t = n_e e,$$

where n_e is the number of electrons and e is the charge of electron. Well, n_e fluctuates by a factor of $\sqrt{n_e}$, and we can then experimentally measure the charge of an electron! But people did experiments with superconductors, which can also carry current. This has no dissipation, and then the noise for n_e was different. This is because q is now $2e$ (superconductivity happens when electrons combine into pairs). So when the current carriers are Cooper pairs instead of electrons, we only need half the number to get the same current, and this means our fluctuation is larger! This was the first proof that superconductors did not have electron charge carriers.

But here's a real life MIT example for shot noise and how to measure it. Let's say we have 10^6 people on Killian Court, and let's say there are a few exits that people randomly leave through. Through one exit, there's 10^4 people who leave per unit time, and by shot noise, there's a fluctuation of ±100 people.

But now let's say that people leave in groups of 100 instead: now there's 10^2 groups that leave per unit time, so the shot noise is just 10 groups. This means the actual absolute fluctuation is 10^3, which is larger than the original!

So if the carrier of charge, or the unit size of people, is increased by a factor of N, the shot noise is increased by a factor of $\sqrt{N}$.

11.3 Stirling's formula

How do you memorize the formula? If you do

$$n! = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1,$$

we know n^n is a bad estimate, and $\left(\frac{n}{2}\right)^n$ would be a bit closer. And also because we have logarithms, e is a pretty good number to use. So instead

$$n! \approx \left(\frac{n}{e}\right)^n.$$

This gives

$$\log n! \approx n(\log n - \log e) \approx n \ln n - n.$$

There's another term of $\ln \sqrt{2\pi n}$, but when n is large (as it is in this class), that's several orders of magnitude smaller than n! So we can neglect it.

12 February 26, 2019

Remember that we have been introducing probability distributions. We found that the Poisson distribution converges to a Gaussian as the frequency of events becomes larger, and this was an important example of the Central Limit Theorem.

Today, we'll talk about the binomial distribution and connect it to the idea of diffusion. Finally, we'll discuss conditional probability and figure out ideas like "energy given some other knowledge about the system."

12.1 Binomial distribution

Consider a random variable with two possible outcomes A and B, occurring with probability p_A and $p_B = 1 - p_A$. Our goal is to find out how many times A occurs if we repeat the random variable N times. Then this can be calculated as

$$P_N(N_A) = \binom{N}{N_A} p_A^{N_A} (1-p_A)^{N-N_A}$$

The first factor of

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

comes from the number of ways in which we can choose which of the events are A, and the rest is just the probabilities of A and B multiplied the relevant number of times.

Is this normalized? By the Binomial theorem,

$$\sum_{N_A=0}^{N} P_N(N_A) = \sum_{N_A=0}^{N} \binom{N}{N_A} p_A^{N_A} (1-p_A)^{N-N_A} = (p_A + (1-p_A))^N = 1,$$

as desired. We can also find some other statistics by doing mathematical manipulation:

$$\langle N_A \rangle = N \cdot p_A, \text{var}(N_A) = N p_A p_B.$$

Then, the ratio of the standard deviation to the mean of N_A is

$$\frac{\sigma(N_A)}{\langle N_A \rangle} = \frac{\sqrt{N p_A (1 - p_A)}}{N p_A} = \frac{1}{\sqrt{N}} \sqrt{\frac{1 - p_A}{p_A}},$$

so as $N \to \infty$, the distribution becomes narrower and narrower.

One question we should be asking ourselves: how is it possible for physics to have simple equations that explain the complexity of the world? This shows the beauty of statistical mechanics: we can explain the world by just using probability.

12.2 An application

Fact 78

We're going to use the binomial distribution to derive a diffusion equation. A random process is sometimes a random walk, and we can use the diffusion equation to understand that random walk!

Definition 79

A **random walk** is a path of successive steps in random directions in some space.

These describe many physical phenomena, like collisions of particles, shapes of polymers, and so on. For example, DNA is often curled up in a coil, and its shape is usually described by a random variable with mean zero!

There are two kinds of emergent behavior for random walks.

- Given any individual random walk, after a large number of steps, it becomes a fractal (scale invariant). We won't be talking much about this in class though.
- The endpoint of a random walk has a probability distribution that obeys a simple continuum law, which leads to the diffusion equation!

The idea is that these phenomena are global, so they are independent of the microscopic details of the system.

Example 80

Consider a random walk in one dimension: this is also known as **Brownian motion**. Let's say that it moves left or right along a line with step size ℓ, and the probability is $P(+\ell) = P(-\ell) = \frac{1}{2}$.

First of all, we want to find the average displacement after N steps. Well,

$$\langle \Delta x_i \rangle = \frac{1}{2}(\ell) + \frac{1}{2}(-\ell) = 0,$$

so the average is always 0 after N steps. On the other hand, we can consider the mean squared displacement: Then

$$\langle \Delta x_i^2 \rangle = \frac{1}{2}(+\ell)^2 + \frac{1}{2}(-\ell)^2 = \ell^2,$$

and the mean square displacement after N steps is

$$\langle \Delta x^2 \rangle = \sum_{i=1}^{N} \sum_{j=1}^{N} \langle \Delta x_i \Delta x_j \rangle$$

and now all cross terms $\Delta x_i \Delta x_j$ with $i \neq j$ contribute 0 by independence. This means that

$$\langle \Delta x^2 \rangle = \sum_{i=1}^{N} \langle \Delta x_i^2 \rangle = N\ell^2.$$

Fact 81

We could also use the fact that variances of independent variables add! So since each step has variance ℓ^2, the total sum has variance $N\ell^2$.

So if successive jumps happen every δt, the number of jumps in a time t is

$$N = \frac{t}{\delta t} \implies \langle \Delta x^2 \rangle = \frac{\ell^2}{\delta t} t.$$

This is important: the variance scales linearly with time! In comparison, if our random walk has some average velocity

$$\Delta x(t) = vt \implies \langle \Delta x^2 \rangle = \langle v^2 \rangle t^2,$$

which is called **ballistic motion.** In more advanced statistics, this is the setup for the fluctuation-dissipation theorem!

But what's the main physics of what we're working on: where is the randomness of our process coming from?

- The existence of a randomly fluctuating force will push a particle in random directions.
- There is some inertia of the system, as well as a viscous drag.

Our goal is to compute the probability distribution of finding a particle x away from the original position after N steps. If we denote N_L to be the number of steps to the left and N_R the number of steps to the right (so $N = N_L + N_R$), then the net displacement of the walk is $x = \ell(N_R - N_L)$.

Question 82. *How many distinct walks are possible if we give ourselves N steps, N_L of which are to the left and N_R of which are to the right?*

This is just

$$\binom{N}{N_L} = \frac{N!}{N_L! N_R!}.$$

In total, since each move can be to the left or to the right, there are 2^N distinct ways to form a walk of N steps, and the probability of any sequence is $\frac{1}{2^N}$.

Fact 83

It's important to note that **sequences** each have equal probability, but x**, the net distance**, is not uniformly distributed.

So the probability of having a walk of net length x is

$$p(x, N) = \frac{N!}{N_R! N_L!} \left(\frac{1}{2}\right)^N,$$

which is the number of sequences times the probability of any given sequence.

12.3 In the limiting case

We claim that this becomes a Gaussian as N becomes large. Indeed, note that we've define $x = \ell(N_R - N_L)$, so

$$N_L = \frac{N - x/\ell}{2}, N_R = \frac{N + x/\ell}{2}.$$

Substituting these in, we can then use Stirling's approximation $\ln n! \approx n \ln n - n + \frac{1}{2}\ln(2\pi n)$. This yields

$$p(x, N) \propto \sqrt{\frac{2}{\pi N}} \exp\left(-\frac{x^2}{2N\ell^2}\right)$$

This is a Gaussian symmetric about its mean 0, which tells us that we're mostly likely to have a net displacement of $x = 0$.

Fact 84

This explains why in polymers, most of the time there are blobs rather than straighter lines! It's much more probable to be close to the mean.

If we compute the variance,

$$\langle x^2 \rangle = N\ell^2$$

as expected, and if we say that our events are equally spaced by some time δt, the variance is again

$$\langle x^2 \rangle = \frac{\ell^2}{\delta t} t \propto t.$$

Definition 85

Define the **diffusion constant** D such that

$$\langle x^2 \rangle = 2Dt.$$

This has various applications!

12.4 Multiple random variables

Let's say we have two variables x and y.

Definition 86

The **joint probability distribution function**

$$p(x_0, y_0) = \left.\frac{d^2 F}{dx dy}\right|_{x=x_0, y=y_0} \implies d^2 F = p(x, y) dx dy$$

is the probability of x occurring between x_0 and $x_0 + dx$ and y between y_0 and $y_0 + dy$.

Then

$$F(x, y) = \int_{-\infty}^{x} d\xi \int_{-\infty}^{y} d\eta p(\xi, \eta).$$

We can also define some other quantities like we did in the one-dimensional case:

Definition 87

The **expectation value** of a function $f(x, y)$ is

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy p(x, y) f(x, y).$$

We can also integrate one variable out of the equation to find the probability distribution for the other variable: for example,

$$p(x) = \int_{-\infty}^{\infty} dy p(x, y).$$

With this in mind, let's try to relate our variables. Can we answer questions like

Question 88. *What is the probability that X lies between x and $x + dx$ given that Y is certain, denoted $P(X|Y)$?*

Note that $p(x|y)$ should be proportional to $p(x, y)$: we'll talk more about this next time!

13 February 27, 2019 (Recitation)

Some of Professor Ketterle's colleagues said that photons don't have mass, which is true in the basic sense. But there's a relativistic mass-energy equivalence, so writing the equation for a photon

$$E = m_0 c^2$$

does actually make sense and has a nonzero m_0 for photons. But there's a question beyond semantics of "real mass" versus "relativistic mass" here:

Question 89. *If we take a cavity and put many photons inside bouncing back and forth, does the mass increase? Does it have more inertia?*

The answer is yes, since we have the photons "basically at a standstill!" But the whole point is to be careful what we mean by "mass."

13.1 Small aside for surface tension

Given a water droplet on a surface, there's three different surface tensions, corresponding to the three pairs out of {surface, air, water}. The concept of surface tension is that surfaces want to shrink, and this creates net forces at interfaces.

Water droplets stop existing when the forces cannot be balanced anymore, so an equilibrium state cannot exist. This point is called **critical wetting**, and anything beyond that point results in water coating the whole surface!

13.2 Delta functions and probability distributions

Example 90

Let's say we have a harmonic oscillator in a dark room, and there is a lightbulb on the oscillator. If we take a long-exposure photo, what does the light distribution look like?

If we let $x = \sin \omega t$, all phases ϕ are equally probable, since ϕ is proportional to time. So we want to go from a probability distribution of ϕ to one of x: how could we do that?

Well, we know that the velocity is slower at the ends, so we expect more values of x on the ends than the middle. (This is explained more rigorously in the problem set.) The punchline is that the probability distribution is going to be proportional to $\frac{1}{v(x)}$, and now we can proceed with mathematics.

So in this case, we have time as our random variable, and we need

$$1 = \int_0^T p(t)dt$$

for a period of length T. So our starting distribution $p(t) = \frac{1}{T}$, and now we want to turn this into a distribution $p(x)$. Well,

$$p(\overline{x}) = \int_0^T p(t)\delta(\overline{x} - x(t))dt.$$

since **we ask for the moments in time where** $x(t) = x$**.** In this case, $p(t)$ is constant, so this is

$$\frac{1}{T}\sum_i \frac{1}{|f'(t_i)|} = \frac{1}{T}\frac{1}{|v(t)|}$$

where $f(t) = \overline{x} - x(t)$ and t_i are the roots. (Notice that this gives us our probability normalization for free!) So now $\frac{dx}{dt} = \omega \cos \omega t = \omega\sqrt{1 - \sin^2 \omega t}$, which we can write in terms of x as

$$\left|\frac{dx}{dt}\right| = \omega\sqrt{1 - x^2} \implies p(\overline{x}) = \frac{1}{\omega T}\frac{1}{\sqrt{1 - x^2}}.$$

But wait! We haven't been careful enough, because there's two different points where $x(t) = \overline{x}$. The slopes are negative of each other, so even though $\left|\frac{dx}{dt}\right|$ are the same, we need to count the two roots separately. Thus the actual number we want is

$$\frac{2}{\omega T}\frac{1}{\sqrt{1 - x^2}} = \frac{1}{\pi\sqrt{1 - x^2}}.$$

The idea in general is that if we have a probability distribution $p(x)$, and $x = f(y) \implies y = g(x)$ is some function, we can find the probability distribution $p(y)$ by

$$p(\overline{y}) = \int p(x)\delta(\overline{y} - g(x))dx.$$

Basically, we want to take all values where $g(x) = \overline{y}$. But here's another way to see it: the probability differential $p(x)dx$ should correspond to another probability differential $p(y)dy$, so

$$p(y) = p(x)\frac{dx}{dy}.$$

We'd just have to be careful about multiple roots, which the delta function does a good job of.

13.3 Poisson radioactive decay

When sampling decay, scientists often take a small time interval Δt such that $\lambda \Delta t = \langle n \rangle$ is very small. This is the limit $\langle n \rangle \ll 1$: the distribution is also correct in this limit.

In general, the probability to get one count is $\langle n \rangle = p$: our question is to find the probability of two counts in that small interval. Is it p^2?

Example 91

Consider a die with N faces. We throw the dice twice ($n = 2$): what is the probability we get two 1s? It's $\frac{1}{N^2}$. On the other hand, what's the probability we get exactly one 1? It'll be $\frac{2N-2}{N^2}$; as $N \to \infty$, this goes to $\frac{2}{N}$.

So notice that the probability of two 1s is actually not the square of the probability of one 1! In fact, it's $\frac{p^2}{4}$.

But back to the radioactive decay case. Does the same argument work here? Well, the Poisson distribution is

$$p_n = \frac{\langle n \rangle^n}{n!} e^{-\langle n \rangle}.$$

Taking this to the limit where $\langle n \rangle \ll 1$, we can neglect the exponential term, and

$$p_n = \frac{p^n}{n!}.$$

This isn't the same as the $\frac{p^2}{4}$, because throwing a die is not Poisson - it's binomial! To modify the distribution into one that's more Poisson, we have to make N, the number of sides, go to infinity, but we also need to take n, the number of throws, to infinity. We'll do this more formally next time, but if we take $n, N \to \infty$ while keeping the expected number of events the same, then $\frac{n}{N}$ should be constant. This will indeed get us the desired Poisson distribution!

Question 92. *Let's say we have a count of N for radioactive decay: what is σ?*

This is shot noise: it's just $\sqrt{N}$.

Question 93. *Let's say we do a coin toss and flip it some number of times, getting N heads. What's σ for the number of heads that appear?*

Binomial variance works differently: since $\sigma^2 = np(1-p)$, $\sigma = \sqrt{N}/2$!

Question 94. *What if the probability for a head is 0.999?*

In this case, σ^2 is much less than N, and we'll have basically no fluctuation relative to $\sqrt{N}$. So binomial distributions work in the opposite direction! On the other hand, taking probability of a head to be 0.001 will give basically $\sqrt{N}$. So that's the idea of taking the binomial distribution

$$a^n(1-a)^{N-n}\binom{N}{n}$$

with mean Na and variance $Na(1-a)$. If $1-a$ is very small, this yields similar statistics to the Poisson distribution!

14 February 28, 2019

There is an exam on March 12th, so here is some information on it! It will be at 11-12:30 (during regular class hours), and there will be four questions on the exam. This year, Professor Fakhri will post the past five years' worth of exams, and three of the four problems will be from the past five years. They will be posted in the next few days, and we'll have about 10 days to work through those. (One question will be new.)

The material will cover things up to next Thursday. The next two lectures will talk about Shannon entropy, and those are the last lectures that will be on the exam.

There will also be an optional review session held by the TAs next Thursday.

Next week, the American Physical Society meeting will be taking place, so if we want extra office hours, we should send over an email!

14.1 Overview

We started talking about conditional probability last time, which will be helpful in talking about canonical ensemble properties. We thought about the probability distribution of a sum of random variables, particularly thinking about doing repeated measurements of some quantity. The idea is that we'll get closer and closer on average to the actual quantity.

The idea was that we started with a two-variable probability distribution, and we wanted to find the probability that X lies between x and $x + dx$ given a fixed value of y. This is denoted $p(X|Y)$.

14.2 Conditional probability

Claim 95. *$p(x, y)$ is proportional to $p(x|y)$.*

We know that

$$\int p(x|y)dx = 1,$$

since with y held fixed, we expect to find x somewhere (we're just limiting ourselves to a one-dimensional probability distribution). In addition,

$$\int p(x, y)dx = p(y),$$

since this is "all possible values of x" for a given y. Thus, we can see that (essentially removing the integrals),

$$p(x|y) = \frac{p(x, y)}{p(y)}.$$

This is the Bayesian conditional probability formula!

Fact 96

We plot $p(x|y)$ using "contour plots."

Example 97

Let's say that the probability of an event happening is uniform inside a circle of radius 1 and 0 everywhere else.

We can write this mathematically using the Heaviside step function:

Definition 98

Define the **Heaviside step function**

$$\theta(x) = \int_{-\infty}^{0} \delta(s)ds.$$

This can be written as

$$\theta(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$$

and it is unclear what $\theta(0)$ is.

Then the example above has a probability distribution of

$$p(x, y) = \frac{1}{\pi}\theta(1 - x^2 - y^2),$$

since we only count the points with $1 - x^2 - y^2 \geq 0$ and the normalization factor comes from the area of the circle (π). Then if we want to find the probabilities $p(x), p(y), p(y|x)$, we can do some integration:

- To find $p(x)$, integrate out the ys:

  $$p(x) = \int dy \frac{1}{\pi}\theta(1 - x^2 - y^2) = \frac{1}{\pi}\int_{\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi}\sqrt{1 - x^2}$$

  since we take the limits to be the zeros of the argument of θ. This holds for all $|x| < 1$ (the probability is 0 otherwise).

- Similarly, we find $p(y) = \frac{2}{\pi}\sqrt{1 - y^2}$ for $|y| < 1$.
- Finally, to find the conditional probability,

  $$p(y|x) = \frac{p(x, y)}{p(x)} = \frac{\frac{1}{\pi}\theta(1 - x^2 - y^2)}{\frac{2}{\pi}\sqrt{1 - x^2}} = \frac{\theta(1 - x^2 - y^2)}{2\sqrt{1 - x^2}},$$

  which is $\frac{1}{2\sqrt{1-x^2}}$ for $|y| < \sqrt{1 - x^2}$ and 0 everywhere else. The idea is that we were initially choosing points randomly in the circle, so the distribution for a given x should also be uniform in y.

Definition 99

Given two random variables X and Y, define them to be **statistically independent** if and only if

$$p(x, y) = p(x)p(y)$$

This means knowing y tells us nothing about x and vice versa: in other words, it's a corollary that

$$p(x|y) = p(x), p(y|x) = p(y).$$

Data analysis uses Bayes' Theorem often, so we should read up on it if we're curious! Also, see Greytak's probability notes page 27 to 34 on jointly Gaussian random variables.

14.3 Functions of random variables

Suppose x is a random variable with probability distribution $\frac{dF_x}{dx} = p(x)$ [recall F is the cumulative distribution function]. Let $y = f(x)$ be a function of a random variable x: what is the probability distribution $\frac{dF_y}{dx} = p(y)$?

Example 100

Consider a one-dimensional velocity distribution

$$p_V(v) = \frac{dF_V}{dv} = ce^{-mv^2/(2k_BT)}.$$

Given that $E = \frac{mv^2}{2}$, is there any way we can find the probability distribution $p(E)$?

The naive approach is to use the chain rule: just say

$$\frac{dF_E}{dE} = \frac{dF_E}{dv}\frac{dv}{dE}.$$

But if we compute this,

$$\frac{dE}{dv} = mv = \sqrt{2mE},$$

and we can plug this in to find

$$p(E) = ce^{-E/kt}\frac{1}{\sqrt{2mE}}.$$

Unfortunately, this is not normalized: instead, let's use delta functions to try to get to the right answer! We can write

$$\frac{dF_x}{dx} = p(x) = \int_{-\infty}^{\infty} d\xi p(\xi)\delta(x-\xi)$$

where the delta function only plucks out the term where $x = \xi$. So

$$\frac{dF_y}{dy} = \int_{-\infty}^{\infty} d\xi p(\xi)\delta(y-f(\xi))$$

(basically, we only select the values of ξ where $f(\xi) = y$). Using the important property of delta functions, this is

$$\frac{dF_y}{dy} = \sum_{\xi_i} \frac{p(\xi_i)}{\left|\frac{df}{d\xi}\right|_{\xi_i}}.$$

So now if we have

$$\frac{dF_v}{dv} = p(v) = ce^{-mv^2/(2k_BT)}$$

as before, we can write

$$\frac{dF_E}{dE} = \int_{-\infty}^{\infty} p(u)\delta\left(E - \frac{mu^2}{2}\right) du$$

since we want all values of E equal to $\frac{mu^2}{2}$. Notice this happens at $u_i = \pm\sqrt{\frac{2E}{m}}$: there's two roots, so that results in

$$\frac{dF_E}{dE} = \frac{2}{\sqrt{2mE}} ce^{-E/k_BT}$$

which (can be checked) is normalized in the same way. The chain rule method misses the multiple roots!

14.4 Sums of random variables

Let's say we have $x_1, \cdots, x_n$ random variables with probability density functions $p_j(x_j)$ for all $1 \le j \le n$.

Assume x_js are all statistically independent: then

$$p(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} p_j(x_j).$$

For simplicity, let's say that all probability distributions p_j are the same function $p(x)$. Then we can also write this as

$$p(x_1, \cdots, x_n) = \prod_{i=1}^{n} dy_j p(y_j)\delta(x_j - y_j)$$

(the delta function notation will make our lives easier later on). Our goal is to find the mean, average, variance, and other statistics for $x_1 + \cdots + x_n$.

Fact 101

This is applicable for experiments that repeatedly try to measure a quantity x. The average measured value $\overline{x}$ should have a probability distribution that grows narrower and narrower!

Using the notation

$$S_n = \sum_{j=1}^{n} x_j, \overline{x_n} = \frac{S_n}{n},$$

our goal is to find the uncertain in our measurement after n trials. More specifically, we want to find the probability distribution of $\overline{x}$.

Proposition 102

The variance of the average is proportional to $\frac{1}{n}$, and as $n \to \infty$, this becomes a Gaussian.

Why is this? The probability distribution for $\overline{x_n}$, much like the examples above, is

$$p(\overline{x_n}) = \prod_{j=1}^{n} \int p(y_j) \delta\left(\overline{x}_n - \frac{1}{n}\sum_{k=1}^{n} y_k\right) dy_j$$

and the mean of $\overline{x}_n$ is

$$\int p(\overline{x_n})\overline{x_n} d\overline{x_n}$$

$$= \prod_{i=1}^{n} \int dy_j p(y_j) \frac{1}{n} \sum_{k=1}^{n} y_k.$$

Switching the sum and product, this is

$$\frac{1}{n}\sum_{k=1}^{n}\prod_{j=1}^{n}\int p(y_j) y_k dy_j = \frac{1}{n} n\langle x\rangle = \langle x\rangle,$$

and this is just a convoluted way of saying that the average expected measurement is just the average of x.

Next, let's find the variance of the averages: we can first compute

$$\langle \overline{x_n}^2 \rangle = \int \overline{x_n}^2 p(\overline{x_n}) d\overline{x_n},$$

which expands out to

$$\int d\overline{x_n}\overline{x_n}^2 \prod_{j=1}^{n} \int dy_j p(y_j) \delta\left(\overline{x_n} - \frac{1}{n}\sum_{k=1}^{n} y_k\right)$$

and simplifying this by evaluating the delta functions, this becomes

$$\frac{1}{n^2}\prod_{j=1}^{n}\int p(y_j) dy_j \left(\sum_{k=1}^{n} y_k\right)^2 = \frac{1}{n^2}\prod_{j=1}^{n}\int p(y_j) dy_j \left(\sum_{k=1}^{n} y_k^2 + 2\sum_{k>\ell} y_k y_\ell\right)$$

So this yields

$$\frac{1}{n^2}\left(n\langle x^2\rangle + n(n-1)\langle x\rangle^2\right),$$

so since we are trying to find the variance,

$$\text{var}(\overline{x_n}) = \langle \overline{x_n}^2\rangle - \langle \overline{x_n}\rangle^2 = \frac{1}{n}\langle x^2\rangle - \frac{1}{n}\langle x\rangle^2 = \frac{\text{var}(x)}{n},$$

as desired. So

$$\sigma(\overline{x_n}) = \frac{1}{\sqrt{n}}\sigma(x),$$

and this means that the standard deviation of the average gets smaller relative to the standard deviation of x as we make more measurements!

Proposition 103

So the distribution of the sum of random variables is

$$\frac{dF_{S_n}}{dS_n} = p_{S_n}(S_n) = \prod_{j=1}^{n} \int dy_j p(y_j) \delta \left(S_n - \sum_{k=1}^{n} y_k \right).$$

For example, if we have two random variables,

$$p(S_2) = \iint dx dy p_1(x) p_2(y) \delta(s_2 - x - y) = \int dx p_1(x) p_2(s - x)$$

which is the **convolution** of p_1 and p_2, denoted $P_1 \otimes P_2$.

Fact 104

The sum of Gaussian random variables is another Gaussian with mean and variance equal to the sums of the means and variances of the two original Gaussians. Similarly, the sum of Poissons

$$\frac{(\lambda T)^n}{n!} e^{-\lambda T}$$

is a Poisson distribution with $\langle S_N \rangle = N \langle n \rangle$ and variance $\operatorname{var}(S_N) = N \operatorname{var} n = N\lambda T$.

15 March 4, 2019 (Recitation)

We'll go through some ideas from the problem set.

15.1 Probability distributions and brightness

We can basically think of this problem as a sphere with fireflies emitting light. What do we see? Well, we know that the light intensity depends on ρ, where $\rho = \sqrt{x^2 + y^2}$. So we basically want to integrate out the z-direction.

The idea is that we should go from $p(x, y, z)$ to different coordinates. Any probability distribution always integrates to 1, so we are interested in only picking out the values

$$p(\rho) = \int p(x, y, z) \delta(\rho - \sqrt{x^2 + y^2}) dx dy dz$$

where $\rho^2 = x^2 + y^2$.

Question 105. *Is $p(\rho)$ the brightness?*

Not quite! $p(\rho)d\rho$ gives the "probability" or number of stars in a narrow strip from ρ to $\rho + d\rho$. So to find the brightness, we need to divide through by $2\pi\rho$, since we need to divide by the area. This means

$$B(\rho) = \frac{p(\rho) d\rho}{2\pi\rho d\rho} = \frac{p(\rho)}{2\pi\rho}.$$

When we calculate this and plot it, $p(\rho)$ looks linear for small ρ, so $B(\rho)$ starts off approximately constant close to the center! This makes sense: the "thickness" at each point is about the same near the center.

15.2 Change of random variables

Let's say we have some function (think potential energy) $E = f(x)$ with inverse function $x = g(E)$. Let's say we're given some spatial distribution $p(x)$, but there is an energy E at each point: how can we find the probability density function $p(E)$?

We can use the cumulative function $F(x)$, defined by $\frac{d}{dE}F(E) = p(E)$. By the chain rule, we can regard $E = f(x)$, so

$$\frac{dF}{dx} = \frac{dF}{df(x)}\frac{df(x)}{dx} \implies p(x) = p(E)\frac{df}{dx}.$$

But it is possible that our function is multi-valued: for example, what if $E = f(x) = x^2$ is quadratic? Then the function is not one-to-one: in this case, our cumulative distribution should not look like

$$F(E) = \int_{-\infty}^{g(E)=\sqrt{E}} p(x)dx,$$

but rather

$$F(E) = \int_{-\sqrt{E}}^{\sqrt{E}} p(x)dx$$

so that we get all values x such that $x^2 \le E$.

Another example of this is that if we have $E = x^3$, we just have

$$F(E) = \int_{-\infty}^{\sqrt[3]{E}} p(x)dx,$$

since again we want all values of x that make $x^3 \le E$. The idea is that being "cumulative" in x doesn't necessarily mean the corresponding values are any kind of "cumulative" function in $f(x)$, so we need to be careful! This is part of the reason why we like to use delta functions instead: the central idea is that we need to look at all the roots of $E = p(x)$.

So the rest is now a differentiation:

$$p(E) = \frac{d}{dE}F(E) = \frac{d}{dE}\int_{-\infty}^{g(E)} p(x)dx,$$

where g is a root of the equation $f(x) = E \implies x = f^{-1}(E)$ (there could be multiple such roots, which means we'd have to split into multiple integrals). So by the first fundamental theorem of calculus, this is

$$p(E) = p(g(E))\frac{dg}{dE} \implies p(E) = p(x)\frac{1}{df/dx}$$

as before! In general, if our expression looks like

$$p(E) = \frac{d}{dE}\left(\int_{x_i}^{x_{i+1}} p(x)dx + \int_{x_{i+2}}^{x_{i+3}} p(x)dx + \cdots\right)$$

and this gives the sum

$$\sum_i p(x_i)\left|\frac{dg}{dE}\right|_{f(x_i)=E},$$

where the absolute value comes from the fact that the lower limits have a different sign of slope from the upper limits, which cancels out with the negative sign from the first fundamental theorem.

Here's one more method that we will all like! Add a theta function $\theta(E - f(x))$: we want to integrate over all values $f(x) \le E$, where θ is the step function, and this is just

$$F(E) = \int p(x) dx \theta(E - f(x))$$

(think of this as "including" the parts that have a small enough value of $f(x)$). But now taking the derivative,

$$p(E) = \frac{d}{dE} F(E) = \int p(x) \theta'(E - f(x)) dx$$

and the derivative of the theta function is the delta function

$$p(E) = \int p(x) \delta(E - f(x)) dx.$$

We've done this before! We just care about all roots where $f(x) = E$, and this is

$$\sum_i p(x_i) \frac{1}{|f'(x_i)|},$$

and indeed this is the same result as we've gotten through other methods.

Fact 106

Key takeaway: sum over roots and correct with a factor of the derivative!

Now if we have a function E as a function of multiple variables $f(x, y, z)$, we can just pick out the "correct values" via

$$p(E) = \iiint p(x, y, z) \delta(E - f(x, y, z)) dx dy dz.$$

How do we evaluate the derivatives here? It's possible that δ in general could be a product of three delta functions: for example, think of a point charge in electrostatics. But in this case, we're dealing with a one-dimensional delta function. We need to solve the equation $E = f(x_i, y, z) = 0$: we may have two roots x_1, x_2, so now we have

$$p(E) = \iint \sum_{i=1}^{2} p(x_i, y, z) \cdot \frac{1}{\left|\frac{\partial P}{\partial x}(x, y, z)\right|_i} dy dz$$

The point is that the delta function eliminates one variable, so integrate one variable at a time! Alternatively, there may be some condition on y and z (for example, if $E = x^2 + y^2 + z^2$, then $y^2 + z^2$ are forced to be within some range), and that just means we have to add an additional condition: either an integration or a theta function.

16 March 5, 2019

Fact 107

The optical trap, a focused beam of light, can trap particles of size approximately 1 micron. People can play Tetris with those particles! Once the optical traps are turned off, though, particles begin to diffuse, and this can be explained by statistical physics.

The American Physical Society meeting was held this week: 90 percent of the awards this year were given to statistical physics applied to different fields. The material covered was a bit more challenging than areas we've explored so far, but there are many applications of what we've discussed!

16.1 Entropy and information: an overview

Let's start by reviewing why we've been talking about all the topics of this class so far. We used the first law

$$dU = đQ + đW,$$

where $đW$ is a product of an intensive and extensive quantity such as $-PdV$ (generalized force and displacement). We found that $đQ$ is not a state function: it does depend on the path we take to get to a specific state. **Entropy** is an old idea, and it comes from thermodynamics as a way to keep track of heat flow! It turns out that this S is indeed a state function, and it can characterize a macroscopic state.

We're going to use the ideas from probability theory to write down an "entropy of a probability distribution" and measure how much "space" is needed to write down a description of the states. One definition we'll see later is the probabilistic

$$S = -\sum p_j \log p_j.$$

Consider the following thought experiment: let's say we have some particles of gas in a box, and we have the experimental tools to measure all of those positions and velocities. We write them down in a file, and we want to **compress the files** to the smallest possible length.

If we compress efficiently, the length of the file tells us something about the entropy of the distribution. For example, if all of the particles behave "simply," it will be easy to efficiently compress the data, but if the system looks more "random" or "variable," the compression will be less effective.

Fact 108

In general, if we heat up the room by 10 degrees and repeat this process, the compressed file will have a longer length. Generally, we want to see the **change in length of the file** per temperature or heat added!

This is a connection between two different ideas: an abstract length of a computer program and a tangible heat.

16.2 Shannon entropy and probability

Let's say we have a set $S = \{s_1, \cdots, s_N\}$ whose outcomes have probability $p_1, \cdots, p_N$. An example of a "well-peaked" distribution is

$$p_1 = 1, p_j = 0 \ \forall j \neq 1.$$

If we see an event from this probability, we are "not surprised," since we knew everything about the system from the beginning. On the other hand, if we have all p_j approximately equal,

$$p_j = \frac{1}{N} \forall j.$$

The amount of surprise for any particular event is about as high as it can be in this case! So we're really looking at the amount of randomness we have. Claude Shannon published a paper that was basically the birth of information theory:

Proposition 109 (Shannon, 1948)

What is the minimum number of binary bits σ needed on average to reproduce the precise value of a symbol from the given bits? This turns out to be

$$\sigma = -\sum_{j=1}^{N} p_j \log_2 p_j.$$

Example 110

Let's say we have a fair coin that can come up heads or tails with $P_H = P_T = \frac{1}{2}$. If we have a string of events like $HTHHTHTTHTTT$, we can represent this in a binary string by $H \to 1, T \to 0$.

Clearly, we do need 1 bit to represent each coin flip. Our "symbol" here is an individual "head" or "tail" event, and the minimum number of bits needed is

$$-\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = 1.$$

So our "coding scheme" sending heads and tails to 0 and 1 is "maximally efficient."

Example 111

Let's say we have four symbols A, B, C, D, all equally likely to come up with probability $\frac{1}{4}$. We can represent this via

$$A \to 00, B \to 01, C \to 10, D \to 11.$$

The Shannon entropy of this system is indeed

$$-\sum_{1}^{4} \frac{1}{4}\log_2\frac{1}{4} = 2,$$

so we need at least 2 bits to encode each symbol.

Example 112

Let's say we have three symbols with probability

$$P_A = \frac{1}{2}, P_B = \frac{1}{4}, P_C = \frac{1}{4}.$$

Naively, we can represent $A = 00, B = 01, C = 10$, so we need 2 bits per symbol. But there is a better code? Yes, because the Shannon entropy

$$-\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{4}\log_2\frac{1}{4}\right) = \frac{3}{2},$$

so there should be a way to code each symbol in 1.5 bits on average! Here's a better coding scheme: use $A = 0, B = 10, C = 11$, which gives an average of

$$\frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = \frac{3}{2}$$

Now given a string

$$00101000111101001010,$$

we can reconstruct the original symbol: if we see a 0, then pull it out as an A, if we see a 1, pull it and the next number out to form a B or C, and rinse and repeat until we reach the end!

In general, the idea is to group symbols to form **composites**, and associate **high probability with shortest bit strings.** In the case above, we had a high chance of having A, so we made sure it didn't require too many bits whenever it appeared.

Example 113

If we have a biased coin with probability $\frac{3}{4}$ of heads (A) and $\frac{1}{4}$ of tails (B), the Shannon entropy is

$$-\left(\frac{1}{4}\log_2\frac{1}{4}+\frac{3}{4}\log_2\frac{3}{4}\right)\approx 0.811.$$

So there should be a way to represent the heads-tails method in less than 1 character per flip!

We can group symbols into composites with probabilities

$$\begin{cases} AA: & \frac{9}{16} \\ AB: & \frac{3}{16} \\ BA: & \frac{3}{16} \\ BB: & \frac{1}{16} \end{cases}.$$

These are all fairly close to powers of 2, so let's represent A as 0, AB as 10, BA as 110, and BB as 111 (this is not perfect, but it works pretty well). Then on average, we need

$$\frac{9}{16}\cdot 1+\frac{3}{16}\cdot 2+\frac{3}{16}\cdot 3+\frac{1}{16}\cdot 3\approx 1.688$$

bits to represent two symbols, for an average of less than 1 bit per symbol! This is better than the version where we just use 1 for heads and 0 for tails.

Fact 114

If we instead group 3 symbols, we may be able to get an even better coding scheme! We do have to make sure we can umambiguously decode, though.

By the way, for our purposes from now on, we'll be using natural log instead of base 2 log, since we have continuous systems instead. Note that

$$\log_2 X=\frac{\ln X}{\ln 2},$$

so the Shannon entropy is

$$\sigma=-\frac{1}{\ln 2}\sum_n p_n\log p_n,$$

and if we instead have a continuous probability distribution, we can integrate instead:

$$\sigma=-\int p(n)\log_2 p(n)dn,$$

where we normalize such that $\int p(n)dn=1$.

16.3 Entropy of a physical system

Now that we have some intuition for "representing" a system, let's shift to some different examples.

Consider the physical quantity

$$S = -k_B \sum_i p_i \ln p_i.$$

Note that all terms here are nonnegative, so the minimum possible value is $S = 0$: this occurs when there is only one event with probability 1 and no other possibility. This is called a **delta function distribution**. On the other hand, the maximum possible value occurs with a **uniform distribution**: where all p_is are the same. If there are M events each with probability $\frac{1}{M}$, this evaluates to

$$-k_B \sum_i \frac{1}{M} \ln M = k_B \ln M.$$

(By the way, the k_B is a way of converting from Joules to Kelvin for our measure of temperature.)

Proposition 115

This means S is a measure of "dispersion" or "disorder" of the distribution! So this gives an estimate of our probability distribution, or at least its general shape.

For example, if we have no information about our system, we expect it to be uniform. This yields the maximum possible value of S (or entropy), and this is the best unbiased estimate of our distribution. Once we obtain additional information, our unbiased estimate is obtained by maximizing the entropy given our new constraints.

Fact 116

This is done using Lagrange multipliers!

If we have some new information $\langle F(x) \rangle = f$ (we measure the value of some function $F(x)$), we want to maximize

$$S(\alpha, \beta, \{p_j\}) = -\sum_i p_i \ln p_i - \alpha(\sum(p_i) - 1) - \beta(\sum(p_i F(x_i)) - f)$$

Our constraints are that our distribution must be normalized and that we want $\langle F(x) \rangle - f$ to be close to 0 as well. It turns out this gives a **Boltzmann distribution**

$$p_i = \alpha \exp(-\beta F(x_i)).$$

Here β is fixed by our constraints, and α is our normalization factor! For example, we could find β by knowing the average energy of particles. We'll see this a bit later on in the course.

16.4 Entropy in statistical physics

Recall that we specify a system by stating a thermodynamic equilibrium macrostate: for example, we give the internal energy, pressure, temperature, and volume of a system. This is **specifying an ensemble.**

On the other hand, we can look at the microstates of our system: they can be specified in quantum systems by numbers $\{n_j, 1 \le j \le N\}$ or in the classical systems by positions and velocities $\{x_i, v_i, 1 \le i \le N\}$.

We can set up a distinction here between information theory and statistical mechanics. In information theory, our ensembles look very simple: we have usually a small number of possible outcomes, but the probability distributions can look very complicated. On the other hand, ensembles in statistical mechanics are often much more complicated (lots of different possible microstates), but our probability distributions are much more simple. The idea is that S, our entropy, will be the maximum value of

$$S = -k_B \sum p_i \ln p_i$$

across all probability distributions $\{p_i\}$.

But what are the distributions given our constraints? That's what we'll be looking at in the rest of this class!

17 March 6, 2019 (Recitation)

Let's start with a concrete example of the discussion from last recitation. Let's say we have a probability distribution that is uniform inside a circle of radius R: $p(x, y) = \frac{1}{\pi R^2}$ for $x^2 + y^2 \le R^2$ and 0 outside. We're going to find the probability distribution $p(r)$ in three different ways.

17.1 The messy way

First of all, if we use Cartesian coordinates, we can directly write this in terms of a delta function

$$\boxed{p(r) = \int_{-R}^{R} \int_{-\sqrt{R^2-y^2}}^{\sqrt{R^2-y^2}} p(x, y)\delta(r - \sqrt{x^2 + y^2})dxdy}$$

Let's take care of the delta function as a function of x. We know the delta function $\delta(f(x))$ is $\delta(x)$, divided by $f'(x)$ at a zero of the function, so let's compute the roots!

$$f(x) = r - \sqrt{x^2 + y^2} \implies x_\pm = \pm\sqrt{r^2 - y^2}.$$

The absolute value of the derivative is equal at both roots:

$$f'(x) = -\frac{x}{\sqrt{x^2 + y^2}} \implies |f'(x_\pm)| = \frac{\sqrt{r^2 - y^2}}{r} = \sqrt{1 - \left(\frac{y}{r}\right)^2}.$$

So now, we can evaluate our boxed expression above. The delta function is integrated out (except that we gain a factor of $|f'|$ in the denominator, and we replace x with the root x_i wherever it appears. But here the probability distribution is uniform (does not depend on x explicitly), and the two roots have equal $|f'(x_i)|$, so we get a factor of 2. This simplifies to

$$= \sum_{x_i \text{ roots}} \int_{-R}^{R} \boxed{\int_{-\sqrt{R^2-y^2}}^{\sqrt{R^2-y^2}}} \frac{\boxed{\delta(x - x_i)}}{\pi R^2} \cdot \frac{1}{\sqrt{1 - \left(\frac{y}{r}\right)^2}} \boxed{dx}\, dy = \int_{-R}^{R} \frac{1}{\pi R^2} \cdot \frac{2}{\sqrt{1 - \left(\frac{y}{r}\right)^2}} dy (?)$$

where the boxed terms integrate out to 1, since $x_i = \pm\sqrt{r^2 - y^2}$ is always in the range $\left[-\sqrt{R^2 - y^2}, \sqrt{R^2 - y^2}\right]$.

But we must be careful: if $|y| > |r|$, or if $|r| > R$, we don't actually have these two roots! So we put in some constraints in the form of theta (step) functions: they force $R > r$ and $r^2 > y^2$:

$$= \boxed{\theta(R - r)\frac{2}{\pi R^2} \int_{-R}^{R} \frac{\theta(r^2 - y^2)}{\sqrt{1 - \left(\frac{y}{r}\right)^2}} dy}$$

(where we have r^2 and y^2 in the second theta function to deal with potentially negative values of y). What does that

θ function mean? The inner one just means we integrate across a different range of y:

$$\theta(R-r)\frac{2}{\pi R^2}\int_{-r}^{r}\frac{1}{\sqrt{1-\left(\frac{y}{r}\right)^2}}dy$$

and now we can integrate this: substituting $u = \frac{y}{r}$, this is

$$\theta(R-r)\frac{2r}{\pi R^2}\int_{-1}^{1}\frac{1}{\sqrt{1-u^2}}du$$

and the integral is $\sin^{-1}(u)|_{-1}^{1} = \pi$, resulting in a final answer of

$$p(r) = \theta(R-r)\frac{2r}{R^2} = \boxed{\begin{cases}\frac{2r}{R^2} & r < R\\ 0 & r \geq R\end{cases}}$$

It's stupid to use Cartesian coordinates here, but this shows many of the steps needed!

17.2 Polar coordinates

Here's a faster way: the probability distribution

$$x = \rho\cos\theta, y = \rho\sin\theta$$

becomes $p(x,y) = \frac{1}{\pi R^2}$ for $\rho < R$. So now we can write our boxed double integral above in our new coordinates:

$$\int_0^R\int_0^{2\pi}p(x,y)\delta(r-\rho)\rho d\theta d\rho.$$

The integration over $d\theta$ gives a factor of 2π, and p is uniform, which simplifies this to

$$\frac{2\pi}{\pi R^2}\int_0^R\delta(r-\rho)\rho d\rho.$$

The delta function has only the root $\rho = r$: since we're integrating over $[0,R]$, this is

$$p(r) = \frac{2\pi}{\pi R^2}r\cdot\theta(R-r) = \frac{2r}{R^2}\theta(R-r),$$

which is identical to what we had before.

17.3 Without delta functions

We can use cumulative density distributions instead! What is the cumulative probability

$$F(r) = \int_0^r\int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}}p(x,y)dxdy?$$

This is the probability over all $x^2+y^2 \leq r^2$. This integral is just $p(x,y)$ (which is constant) times the area of a circle with radius r, which is

$$\frac{\pi r^2}{\pi R^2} = \frac{r^2}{R^2}$$

as long as $r < R$. So $p(r)$ is just the derivative of $F(r)$:

$$p(r) = \frac{dF(r)}{dr} = \frac{2r}{R^2},$$

and we're done! We could have fixed up the edge case of $r > R$ by adding a theta function $\theta(r - \rho)$ inside the original integrand. Then the derivative of the theta function is the delta function, which gives the same delta function as in our first method.

To summarize, we can generally avoid delta functions with cumulative densities.

17.4 Parity violation

We once assumed that if we flip our coordinates $x \to -x, y \to -y$, and so on, there is no difference in our laws. Basically, everything in a mirror would also obey the laws of physics. But the Wu experiment proved this to be false!

This is called parity (P) violation. But there's something more interesting: people managed to include charge (C) conjugation, changing matter and antimatter, and then CP conservation seemed to be true. But it was found that even this is violated!

Fact 117

So if you want to tell your friends in the alien world what the right side is, you can say to run a current through a coil of wire. Put in Cobalt-60, and the magnetic field from the coil will have more electrons coming out from the top than the bottom if our coil is counterclockwise. This is a **right-handed current**!

But if we want to explain that our heart is on the left-hand side, we can put ourselves in a solenoid. Run a current through the solenoid, and you can say that the Cobalt-60 emission goes up. Then the current now flows across our chest from the right to the left!

But they'll be made of antimatter, so they might hold out the wrong hand when they shake hands.

18 March 7, 2019

Just a reminder: there is an exam on Tuesday in this classroom. It will go from 11 to 12:30 (usual class time). The past exams have been uploaded: note that three of the four problems on this year's exam will be reused from previous exams posted on Stellar. If we are able to do the problem sets and past exams, we have a good mastery of what's going on.

By the way, the concept of **density of states** has been moved: it will come later, and it will not come up on the exam.

There will be an optional review session from 5 to 7 today, and Professor Fakhri will also hold additional office hours. Material is everything from class until today, though entropy will be more about concepts than specific examples.

The next problem set will not be due on Friday.

18.1 Quick review and overview

We've been learning about macroscopic quantities and connecting them to microscopic systems, and this led us to the idea of entropy. This was a concept that Boltzmann introduced before information theory: basically, we care about how "confused" we are about a system.

Today, we're going to expand on the concept of thermodynamic entropy and introduce the second law of thermodynamics, which claims that entropy is nondecreasing with time. This is an **emergent** phenomenon!

Remember that for a thermodynamic ensemble, we defined our entropy to be

$$S = k_B \sum_j p_j \ln p_j,$$

where we're summing over all possible microstates j that occur with a probability of p_j. Note that this is also

$$\equiv (k_B \ln 2)\sigma,$$

where σ is the Shannon entropy.

18.2 Looking more at entropy

Proposition 118

Thermodynamic equilibrium occurs at the (unbiased) probability distribution which maximizes entropy:

$$S = \max_{p_j} -k_b \sum_j p_j \ln p_j.$$

Recall from last time that an example of such an unbiased probability distribution is the **uniform** distribution: all states occur with equal probability. If we have no prior information, this is the best "guess" we can have for what our system looks like. In this case, the distribution looks like

$$p_j = \frac{1}{\Gamma},$$

where Γ is the total number of consistent microstates. (Γ is known as the **multiplicity**.) Plugging this in,

$$S = -k_b \sum_{j=1}^{\Gamma} \frac{1}{\Gamma} \ln \frac{1}{\Gamma} = \boxed{k_b \ln \Gamma}.$$

Fact 119

On Boltzmann's tombstone in Vienna, $S = k \log W$ is written. This equation is kind of the foundation of statistical physics!

Note that S is a measure of the macroworld, while W or Γ is a measure of the microworld, so this is a good relationship between the two.

Here's some additional facts about our entropy S.

- S is a state function of P and V. In other words, it is independent of the path we took, so we can compute S from other macrostates. For example, we can write the entropy S of a gas in a box in terms of the volume, number of molecules, and internal energy

  $$S = S(U, V, N).$$

  On the other hand, if we have a magnet, the state function depends on the magnetization $\vec{M}$.

- The proportionality constant of k_B arises because we chose to use units of temperature. In particular, we could have units of temperature in Joules if we just let $k_B = 1$.

18.3 The second law

Recall the first law of thermodynamics, which tells us about conservation of energy:

$$dU = \text{đ}Q - PdV.$$

Proposition 120 (Second Law of Thermodynamics)

Entropy of an isolated system cannot decrease.

From an information theory perspective, this is saying that our ignorance of a system only increases with time. Let's look at an example by **time-evolving** a system!

Proposition 121

In both classical and quantum systems, the time-evolution of a microstate is both **causal** and **time-reversal invariant**.

What do those words mean? Causality says that each microstate at some time t_1 evolves into a unique, specific microstate at time $t_2 > t_1$. So causality says that we can't have two different microstates at t_2 that both originated from t_1: if we had 100 microstates at time t_1, we can't have more than that at a later time t_2.

Meanwhile, the concept of time-reversal invariance is that both laws of classical and quantum physics are reversible if we switch $t \to -t$. For instance, any wavefunction $|\psi(t)\rangle$ or classical $\vec{x}(t), \vec{v}(t)$ that is a valid also gives a valid $|\psi(-t)\rangle$ and $\vec{x}(-t), \vec{v}(-t)$.

So if we think about this, it means we cannot have two microstates at time t_1 that converge into one at a later time t_2 either. So our ignorance about the system cannot decrease!

But can the entropy increase?

18.4 A thought experiment

Example 122

Consider a box with a partition, and one half of the box is filled with a gass with a known U, V, N at some time $t < 0$. (The other part is filled with a vacuum.) At time $t = 0$, the partition is removed.

Now the gas fills a volume of $2V$, and U and N are still the same, so there are many more microstates that are possible. This increases our entropy! The kinetic energy of the particles has not changed, but our ignorance of the system has increased. There are many more possible values for the initial position and momentum of every particle.

What's the change in the number of microstates Γ? If we assign a binary variable to each particle, which tells us whether the particle is on the left or right side of the box, after $t > 0$, we now need an extra binary bit to tell us about the system. Thus, with N particles, our change in Shannon entropy is $\Delta\sigma = N$. Thus

$$\Delta S = k_b \ln 2 \Delta\sigma \implies \Delta S = N k_b \ln 2$$

and since $S \sim \log \Gamma$, we get a factor of 2^N more possible microstates!

Fact 123

This doesn't break causality or time-reversal. The idea is that every microstate before $t < 0$ goes to exactly one microstate at $t > 0$, but we don't know which one it is: the probability distribution is still uniform, just with a larger range of possibilities.

Notice that there is some time from our initial state (U, V, N) to our final state $(U, 2V, N)$ to reach equilibrium again (so that we can define our state functions). We can think of this as "mixing" states and making the probability distribution more uniform! There is a whole different field called **ergodicity.**

Fact 124

In any (real) macroscopic system, regardless of the initial configuration, over a long time, the system will uniformly sample over all microstates.

Basically, over a long time, the sampling of a probability distribution will yield all microstates with equal probability. For example, instead of preparing many initial configurations, we can prepare one particle and sample it many times.

Fact 125 (Ergodic hypothesis)

We can compute the properties of an equilibrium macrostate by averaging over the ensemble.

If there are microstates S_i that occur with probability p_i, and we have some function $f(S_i)$ of microstates, we can compute a property

$$\langle f \rangle = \sum_{s_i} f(s_i)p(s_i).$$

But instead, we can sample our system and average:

$$\langle f \rangle = \frac{1}{T}\int_0^T f(t)dt.$$

This time T may be large though!

Fact 126 (Systems out of equilibrium)

There are some systems that have a slow relaxation time, so they'll never reach equilibrium within a reasonable amount of time! An example is the universe.

In the rest of this class, we'll come up with ensembles, and find unbiased probability distributions consistent with a macrostate. We'll try to see what conditions we can impose to define thermodynamic quantities!

18.5 Moving on

We'll be talking about different kinds of ensembles (collections of microstates) in this class. A **microcanonical ensemble** is mechanically and adiabatically isolated, so its volume V and number of particles N is constant. In such a system, we can define a temperature! After that, we will discuss the **canonical ensemble**, which trades fixed U for fixed T. We can then look at **grand canonical ensembles**, which are systems at fixed chemical potential.

Recall that $S = S(U, V, N)$ is a state function on equilibrium states, and $\Delta S > 0$ for isolated systems. We also know that it is an **extensive** quantity (it is additive) like N, V, and U: it turns out the conjugated quantity (generalized force) here is temperature T.

How can we show additivity of entropy?

Lemma 127
Given two independent non-interacting systems A and B, the entropy $S_{AB} = S_A + S_B$.

Proof. A has N_A possible microstates with probability $P_{\alpha,A}$, so

$$S_A = -k_B \sum_{\alpha} P_{\alpha,A} \ln P_{\alpha,A}$$

and similar for B. By statistical independence,

$$P_{A,B} = P_A \cdot P_B,$$

so (here $P_{\alpha,\beta}$ refers to the probability $P_{\alpha,A}P_{\beta,B}$ for brevity)

$$S_{AB} = -k_B \sum_{\alpha,\beta} P_{\alpha,\beta} \ln P_{\alpha,\beta} = -k_B \sum_{\alpha,\beta} P_{\alpha,A}P_{\beta,B} \ln(P_{\alpha,A}P_{\beta,B}) = -k_B \sum_{\alpha,\beta} P_{\alpha,A}P_{\beta,B} \left(\ln P_{\alpha} + \ln P_{\beta}\right))$$

which can be written as

$$S_{AB} - k_B \boxed{\sum_{\beta} P_{\beta,B}} \sum_{\alpha} P_{\alpha,A} \ln P_{\alpha,A} - k_B \boxed{\sum_{\alpha} p_{\alpha,A}} \sum_{\beta} P_{\beta,B} \ln P_{\beta,B}$$

and as the boxed terms are 1, this is just $S_A + S_B$ as desired. ☐

19 March 11, 2019 (Recitation)

We'll cover some short questions and then relate Poisson, Binomial, and Gaussian distributions to each other.

As a quick refresher, if we have a probability distribution $p(x, y)$ and we want to find it in terms of another variable $z = f(x, y)$, then

$$p(z) = \int p(x, y)\delta(z - f(x, y))dxdy$$

will pick out the correct values of z. The rest is mathematics: find the roots, magnitudes of derivatives, and make the relevant substitutions.

19.1 Different probability distributions

Question 128. *What is the counting distribution for radioactive decay? Basically, measure the number of particles that decay / do something else in some interval T: if we do this multiple times, what's the distribution going to look like?*

Remember that we've discussed three kinds of probability distributions here: binomial, Poisson, and Gaussian. We're always looking for a count rate: can we distinguish anything between these three kinds?

Well, in a binomial distribution, we have some finite number of trials N, so the possible range of n, our count, is always between 0 and N. But for the Poisson distribution, n is any nonnegative integer, and the Gaussian can be any real.

The idea is that if our binomial distribution's tail is sufficiently flat on the positive end, because our probability p is small or our number of trials N is large enough, then we can extend it to ∞ and treat it similar to a Poisson distribution. But on the other hand, if our tail is sufficiently flat on the negative end, we can also extend it to $-\infty$ and treat it like a Gaussian distribution!

So the answer to the question is "yes, Poisson is correct," but not quite! There is indeed a maximum count rate: N, the number of total atoms in our radioactive material. So this is sort of binomial, but those events are so unlikely that we can neglect them completely.

How do we rigorize this? Remember that our binomial distribution for N trials of an event of probability a is

$$p(n, N) = a^n(1-a)^{N-n}\binom{N}{n}.$$

If we let $N \to \infty$, but we keep our mean $Na = \langle n \rangle$ constant, then $a = \frac{\langle n \rangle}{N}$, and our distribution becomes

$$p(n, N) = \frac{\langle n \rangle^n}{N^n}\left(1 - \frac{\langle n \rangle^n}{N^n}\right)\binom{N}{n}.$$

We can neglect the $-n$ in the second exponent, since $n \ll N$, and this middle term now approaches $e^{-\langle n \rangle}$. What's more, $\binom{N}{n} = \frac{N(N-1)\cdots(N-n+1)}{n!}$ is essentially $\frac{N^n}{n!}$, and now we're left with

$$p(n, N) \approx \frac{\langle n \rangle^n}{N^n}e^{-\langle n \rangle}\frac{N^n}{n!} = \frac{\langle n \rangle^n}{n!}e^{-\langle n \rangle}$$

which is the Poisson distribution as we wanted!

Fact 129

So letting N go to infinity but adjusting the probability accordingly, we get a Poisson distribution. On the other hand (or as a subsequent step), if we make $\langle n \rangle$ larger and larger, this gives us a Gaussian distribution by using Stirling's approximation and using a Taylor expansion. This will yield

$$C \exp\left[-\frac{(n - \langle n \rangle)^2}{2\langle n \rangle}\right].$$

The idea is that all that matters is the values of n and $\langle n \rangle$. But here's an alternative way to go from binomial to Gaussian: keep a constant, and let N get larger. This now yields

$$C' \exp\left[-\frac{(n - \langle n \rangle)^2}{2\langle n \rangle}\right] \cdot (1-a).$$

So in this case, the variance is not $\langle n \rangle = Na$ but $Na(1-a) = \langle n \rangle(1-a)$. The reason this is different is because when we went to the Poisson as an intermediate step, we forced a to be small, which meant we could neglect the $1-a$ term!

So now let's look at $a = \frac{1}{2}$, which is the case of a random walk. So our variance is $\frac{\langle n \rangle}{2}$, and let's say the step size of our random walk is 1 (so we move to the right or to the left by 1 unit each time).

Fact 130

Notice that, for example, if we have 10 steps, we expect to move to the right 5 times. But if we move to the right 6 times instead, our net walk is $6 - 4 = 2$: in general, if our number of right moves is k more than the mean, the net walk is $2k$.

So if we substitute in for a step size of our random walk x, $n - \langle n \rangle = \frac{x}{2}$, and $\langle n \rangle = \frac{N}{2}$. Rewriting our Gaussian, we will just get Brownian motion

$$p(x) \propto e^{-x^2/2N}.$$

In general, measuring N atoms at a rate of λ for time t just yields

$$\langle n \rangle = N\lambda t.$$

But the concept behind all the counting is to track n, the number of counts, relative to $\langle n \rangle$, the average number of counts.

Fact 131

If we have a probability distribution p_n, the values of $\langle n \rangle$ and $\text{var}(n)$ are telling us data about one trial or sample from the distribution. But if we want to calculate $\overline{n}$, which is the average of N measurements, the variance changes by a factor of $\frac{1}{N}$ (since variances add, so the variance of our sum is $N\,\text{var}(n)$, and then we divide our sum by N, which divides our variance by N^2).

By the way, the formula for a normalized Gaussian distribution is on the equation sheet, so we don't have to worry about it too much.

Let's think about the stars problem from a past exam: we have stars distributed with density ρ stars per light-year cubed. What is the probability that there are no stars within r of a given star?

Basically, we take a volume V, and we want the probability no other star is in that given volume. We can think of this as taking small pieces of volume, where each one is independent, and where there is a finite, consistent value of $\langle n \rangle$: average number of stars in each piece of volume. This is a Poisson distribution! So our expectation value is

$$\langle n \rangle = \rho \cdot V = \frac{4\pi\rho}{3} r^3,$$

and

$$p(n, N) = \frac{\langle n \rangle^n}{n!} e^{-\langle n \rangle}.$$

But why do we take $n = 0$ instead of $n = 1$? The first star just gives us a volume to look at, so we can completely ignore it.

Fact 132 (Clearly false)

If we go on an airplane, we should bring an explosive to be safe, because the probability of there being two explosives on the plane is almost zero.

The central idea here is independence! If we guarantee that we have one star, the other stars don't care that the first star is there.

20 March 13, 2019 (Recitation)

20.1 Geometric versus binomial distribution

Let's quickly look at the tunneling problem from the quiz: we can think of having an alpha particle inside a nucleus that eventually escapes.

We know the probability that it does exactly n bounces is

$$p_n = a^n(1 - a).$$

This is a geometric distribution!

On a related note, if we have $n+1$ atoms, what is the probability that exactly 1 of them decays? Well, this looks very similar to what we have above, but we get an extra combinatorial factor (because any of the atoms can decay):

$$P_1 = p(1-p)^n \binom{n+1}{1},$$

and this turns the geometric distribution into a binomial one!

Fact 133

Here's another example of a geometric distribution: let's say a patient needs a kidney transplant, but we need to screen donors to see if there is a match. Given a random blood test, the probability of a match is p: then the number of people we need to screen is

$$p_n = p(1-p)^{n-1}.$$

and we can replace $a = 1-p$ to get something similar to the tunneling problem.

So the distinction is that you try again and again until success in a geometric distribution, but there's a finite number of trials in a binomial one.

20.2 Entropy of a probability distribution

Let's say we have an average count rate of $\langle n \rangle$ in a Poisson distribution, meaning the variance is $\langle n \rangle$ as well. (Think of the standard deviation as being about at 0.6 times the maximum value in a distribution that is about Gaussian.)

So what does it mean to have an **entropy** of a distribution?

Example 134

Consider a coin toss or letters in a book: can we measure the entropy of that somehow?

The idea is to have N random events pulled from our probability distribution: what is the number of bits needed to represent that information on average? It's kind of like image compression: using much less space to display the same data, but we don't want any loss of resolution.

In a coin toss, we need a 1 or 0 for each toss, since all events are randomly likely. So N random events must come from N bits, and indeed, the Shannon entropy for one event

$$S = -\sum p_i \log_2 p_i = 1,$$

Let's go to the extreme: let's assume we have a coin which comes up heads 99 percent of the time. How many bits of information do we need to communicate the random series $0000\cdots010\cdots01$?

Fact 135

We can just send a number that counts the number of 0s between 1s! So instead of needing about 100 bits to represent each group between 1s, we can use on average $\approx \log_2(100)$ bits.

More rigorously, let's say the probability of having a 1 (corresponding to a tail) is small: $\varepsilon \approx \frac{1}{100}$. If we have N coin tosses, we will need an expected $N \cdot \varepsilon$ differences between the 1s. Each difference is about $\frac{1}{\varepsilon}$, and we need $\log_2 \frac{1}{\varepsilon}$

bits for each one. So the expected entropy here is

$$N\varepsilon \log \frac{1}{\varepsilon} = -N\varepsilon \log \varepsilon,$$

while the theoretical Shannon entropy yields

$$S = -N\left(\varepsilon \log \varepsilon + (1-\varepsilon)\log(1-\varepsilon)\right)$$

and as $\varepsilon \to 0$, the second term dies out to first order! This confirms the result we had before.

Fact 136

One way to think of this is that the entropy gets smaller (we need less bits) as our sequence becomes less random and more predictable.

20.3 Entropy of a Poisson distribution

If we think in terms of information, we are looking at a random sequence: we want to think of coding the resulting data. Well, what's the data that we're trying to code?

If we set up our system and repeatedly count, we'll get a series of numbers pulled from the distribution. So our question is **how we can code this**? How concisely can we represent the stream of random numbers?

Example 137

First of all, let's say we have a uniform distribution from 91 to 110, so there are 20 equally likely outcomes. What's the entropy of this system?

Well, we need $\log_2(20)$ bits on average to represent the data! As a sanity check, if we go to the formula

$$S = -\sum p_i \log p_i$$

and we have W equally probable options, then this Shannon entropy is just

$$S = -W\frac{1}{W}\log\frac{1}{W} = \log W.$$

Fact 138

Ludwig Boltzmann has $S = k_B \ln W$ written on his tombstone - notice that this is just $k_B \ln 2$ times the quantity we've been thinking about! The multiplicative factor is just a convention to connect different fields and stick to historical reasons.

So back to the probability distributions. We can perform N measurements, each of which can be one of W possibilities (for example, lottery numbers or pieces of colored paper). How many bits do we need to convey the random series that comes up? Again, we need $\log_2 W$ bits for each number.

But how would we encode these large numbers? In general, we want to index them: if there are 20 possible numbers, we should send them out as $0, 1, \cdots, 19$, not as their actual values.

So looking at a Poisson distribution, we care much more about the events that occur more frequently. Looking at the $p \log p$ term, as $p \to 0$, this number approaches 0. So the improbable wings of the Poisson distribution are not important: we really care about those values within a few standard deviations!

So the number of numbers that dominate the sequence is around $\sqrt{\langle n \rangle}$ for the Poisson distribution, and thus we estimate the entropy to be

$$\log\left(c\sqrt{\langle n \rangle}\right) \sim \log_2 c + \frac{1}{2}\log_2(\langle n \rangle)$$

where c is some order 1 number (from 1 to 10). Converting this to statistical mechanics, this gives us a $\frac{1}{2\ln 2}$ term in leading order! Indeed, if we actually plug in

$$S = -\sum p_i \log p_i$$

for the Poisson distribution (as we do in our homework), and in fact the actual number looks like (with $x = \langle n \rangle$)

$$S(x) = \frac{1}{2\ln 2}(1 + \ln 2\pi x) = \frac{1}{2\ln 2}\ln x + \frac{1}{2\ln 2}(1 + \ln 2\pi).$$

So our handwaving gave us the correct result asymptotically by replacing the Poisson with a Gaussian. That second factor is approximately 4.1, so indeed we have our order 1 number.

21 March 14, 2019

Happy Pi Day! Read the paper posted on the course website.

We did well on the exam; some of them haven't been graded and will be done by the end of today. They will be brought to class on Tuesday; we can also email Professor Fakhri. If we feel like our score doesn't reflect our understanding, also send an email.

There is a problem set due on Monday (instead of Friday) at 7pm. There will be another pset for the Friday before break, but it is short (only 2 problems). Finally, there will be an interesting problem in the problem set after that about statistics of the Supreme Court.

Last time, we introduced Shannon and thermodynamic entropy. Near the end of class, we found that the total entropy of a system is the sum of its independent parts, so entropy is an extensive quantity. Now we'll relax this assumption and allow heat and work to be exchanged as well, and we'll see what the new entropy becomes! This allows us to define temperature, and then we can connect those microscopic pictures to the macroscopic world. After that, we'll talk about reversibility and quasi-equilibrium and how we can compute changes in entropy based on the initial and final state.

21.1 Entropy and thermal equilibrium

Recall from last time that if we have independent, non-interacting systems A and B, then $S_{AB} = S_A + S_B$.

This time, let's say that we still have A and B isolated from the outside, but there is a partition between A and B. This means that N_A, N_B are fixed, and so are V_A, V_B, but heat can flow between the two systems. Our goal is to somehow define a temperature based on S_{AB}.

We'll let the system go to equilibrium: at that point, our entropy is maximized. Then if A has some entropy S_A and internal energy U_A, and B has some entropy S_B and internal energy U_B, we claim that

$$S_{AB} = S_A(U_A) + S_B(U_B)$$

for functions S_A and S_B that depend on U_A and U_B, respectively. This is a good assumption even if we have small fluctuations.

Now let's say some infinitesimal heat $đQ$ passes from A to B. Then $dU_A = -đQ$ and $dU_B = đQ$. But the change in S_{AB} should be zero, since we have a maximum entropy at this point! Expanding out the differential,

$$dS_{AB} = \left.\frac{\partial S_A}{\partial U_A}\right|_{V_A,N_A} dU_A + \left.\frac{\partial S_B}{\partial U_B}\right|_{V_B,N_B} dU_B.$$

Now plugging in $dS_{AB} = 0$ and using the fact that $dU_A = -dU_B$, we have **at thermal equilibrium** that

$$\left.\frac{\partial S_A}{\partial U_A}\right|_{V_A,N_A} = \left.\frac{\partial S_B}{\partial U_B}\right|_{V_B,N_B}.$$

So we want to define a state function that is equal at these two different points (in two systems in thermal equilibrium)! When we bring two systems together, the temperatures should become equal, which motivates the following definition:

Definition 139

Define the **temperature** T of a system via

$$\left.\frac{\partial S}{\partial U}\right|_{V,N} = \frac{1}{T}.$$

There's no constant of proportionality here, because we used k_B in our definition of entropy.

Here are some important facts:

- $\frac{\partial S}{\partial U}\big|_{V,N}$ only applies to systems at thermal equilibrium, but it's a "hot" area of research to think about non-equilibrium states as well.
- Temperature is **transitive**: if A and B are at thermal equilibrium, and so are B and C, then A and C are at thermal equilibrium. This is the **zeroth law of thermodynamics**.
- There are other notions of equilibrium (e.g. mechanical) as well, but for now we're ignoring ideas like partitions being able to move due to pressure.

21.2 Particles in binary states

Let's put together everything we've learned so far with an example!

Example 140

Let's say we have N particles, each of which can have 2 states (for example, a bit taking on values 0 or 1, or a particle with spins in a magnetic field). One way to represent this is by placing them along a number line and representing each one with an up or down arrow. Spin up gives an energy of ε_0 and spin down gives an energy of ε_1: let's say n_0 is the number of particles in the lower energy state ε_0; without loss of generality we let $\varepsilon_0 = 0$. Similarly, define n_1 to be the number of particles in the upper energy state $\varepsilon_1 = \varepsilon$.

Note that $n_0 + n_1 = N$, and we can write this as a frequency:

$$n_1 = fN \implies n_0 = (1-f)N.$$

The total internal energy of this system is

$$\varepsilon_0 \cdot n_0 + \varepsilon_1 \cdot n_1 = f\varepsilon N.$$

Let's compute the entropy of this system. If we can count the number of microstates Γ (also called the multiplicity), all such states should be equally probable, and then we can compute the entropy from there:

$$S = k_B \ln \Gamma.$$

If we have fN particles in the upper energy state, and we have $(1-f)N$ particles in the lower energy state,

$$\Gamma = \frac{N!}{n_0!n_1!}.$$

Now by the Stirling approximation (since N is large),

$$S = k_b \ln \Gamma = k_b(N \ln N - N - (n_0 \ln n_0 - n_0) - (n_1 \ln n_1 - n_1))$$

and since $N = n_0 + n_1$, this simplifies to

$$S = k_B(N \ln n - n_0 \ln n_0 - n_1 \ln n_1) = k_B(n_0(\ln n - \ln n_0) + n_1(\ln n - \ln n_1)) = -Nk_B(f \ln f + (1-f)\ln(1-f)).$$

Notice that S is indeed extensive: it depends on N, the number of particles that we have. We can also find our temperature:

$$\frac{1}{T} = \left.\frac{\partial S}{\partial U}\right|_N = \frac{\partial S}{\partial f}\frac{\partial f}{\partial U} = (-N \ln f + N \ln(1-f))\frac{k_B}{\varepsilon N}$$

(here volume is not well-defined and not relevant). Moving terms around,

$$\frac{\varepsilon}{k_B T} = \ln(1-f) - \ln f$$

and this can be represented as comparing two kinds of energy: the difference in energy ε between states and $k_B T$, the thermal energy. That's the kind of comparison we'll be doing a lot in this class, since the ratio tells us a lot about the macroscopic quantities of the system! So now defining $\beta = \frac{1}{k_B T}$, we can write

$$\frac{1-f}{f} = e^{\beta\varepsilon}.$$

Then

$$f = \frac{n_1}{N} = \frac{1}{1+e^{\beta\varepsilon}}, 1-f = \frac{n_0}{N} = \frac{1}{1+e^{-\beta\varepsilon}},$$

which also means we can rewrite

$$n_1 = \frac{Ne^{-\beta\varepsilon}}{1+e^{-\beta\varepsilon}}, n_0 = \frac{N}{1+e^{-\beta\varepsilon}}.$$

So now we can compute our internal energy:

$$U = f\varepsilon N = \frac{N\varepsilon}{1+e^{\beta\varepsilon}}.$$

If we plot $\frac{n_0}{N}$ and $\frac{n_1}{N}$ as functions of $k_B T = \frac{1}{\beta}$, the number of excited particles in the upper state increases to an asymptotic limit. Similarly, $\frac{U}{\varepsilon N}$ approaches $\frac{1}{2}$ as $k_B T \to \infty$. Finally, let's plot the heat capacity $C = \left.\frac{\partial U}{\partial T}\right|_N$:

$$C = \frac{N\varepsilon^2}{k_B T^2}\frac{e^{\varepsilon/(k_B T)}}{\left(1+e^{\varepsilon/(k_B T)}\right)^2}.$$

We'll often look at these when $T \to 0, T \to \infty$: as $T \to 0, C \to \frac{1}{T^2}e^{-\varepsilon/(k_B T)}$, and as $T \to \infty, C \to 0$. There will be physical explanations for each of these behaviors as well!

Finally, remember that we computed the entropy: we're going to change our variables:

$$S = S(U,N) = S(T,N) \implies S = -Nk_B(f \ln f + (1-f)\ln f)$$

Using the fact that $f = \frac{1}{1+e^{\varepsilon\beta}}$, we now have

$$S(N,T) = Nk_b \ln(1+e^{-\beta\varepsilon}) + \frac{N\varepsilon}{T}\left(\frac{1}{1+e^{\beta\varepsilon}}\right)$$

and notice that we can split this up in a familiar way:

$$= \frac{U(N,T)}{T} + Nk_b \ln(1+e^{-\beta\varepsilon}).$$

As $T \to 0$, $S \to 0$, and $T \to \infty$, $S \to Nk_b \ln 2$. This gives us a good interpretation of information entropy: high temperature gives high uncertainty!

21.3 Back to macroscopic systems

Let's look back at our equation

$$\frac{1}{T} = \left.\frac{\partial S}{\partial U}\right|_V .$$

This is important because S is computed from the number of microstates in our system! So let's go back to our system with two parts and an immovable thermal partition (so again, we have fixed volume). Let a small amount of heat be transferred from one part to the other.

No work is done, so $đQ = dU$, which means that $dS = \left.\frac{\partial S}{\partial U}\right|_{V,N} dU = \frac{1}{T} dU$.

Here dU is our thermal energy, and since we add a small amount of heat, our systems remain close to thermal equilibrium: thus U and S remain well-defined. Finally, it's important to note that this can be reversed without any effect on our system, since dU is infinitesimal.

Fact 141

Those three conditions are what dictate a **reversible heat transfer**: $dU = đQ$, so $dS = \frac{dQ_{\text{rev}}}{T}$. However, this is only an equality when we have reversible quantities: in general, we have the inequality

$$dS \geq \frac{đQ}{T}.$$

This is the first time we write our heat as the product of an extensive and intensive quantity. It's important that S is a state function, so we can compute the change dS by looking at any path!

Example 142

Consider a system with a partition: the left half has volume V, temperature T_A, and pressure P_A, and the right half is empty. At time $t = 0$, we remove the partition, and now we have a new volume $2V$, T_B, P_B.

Since the system is isolated, $T_B = T_A$ (as no heat is transferred, and an ideal gas's energy only depends on temperature). By the ideal gas law, then, $P_B = \frac{P_A}{2}$. This is an irreducible process, so we can't follow a specific path on the PV diagram. But S is a state function, so we can pick any path!

Let's say we go from A to C (reversibly add heat at constant pressure P until volume $2V$), and then from C to B (reversibly remove heat at constant volume until the pressure drops to $\frac{P}{2}$). Then dQ_{rev} along the first part of our path is $C_P(T)dT$, and dQ_{rev} along the second part is $C_V(T)dT$, so our total change in entropy is

$$S_B = S_A = \int_{T_A}^{T_C} \frac{dT'}{T'} C_P(T') - \int_{T}^{T_C} \frac{dT}{T'} C_V(T')$$

and these integrals can be combined to

$$\int_T^{T_C} \frac{dT'}{T'}(C_P(T') - C_V(T')) = Nk_B \ln \frac{T_C}{T}$$

Since we have an ideal gas, $T_C = 2T$, and therefore

$$\Delta S = S_B - S_A = Nk_B \ln 2,$$

as we expect!

22 March 18, 2019 (Recitation)

22.1 Questions

Let's discuss the entropy of a Poisson ensemble from the problem set. We've been given a Poisson distribution with mean $\langle n \rangle$: then

$$p_n = \frac{\langle n \rangle^n}{n!} e^{-\langle n \rangle}.$$

By definition, the Shannon entropy is

$$-\sum_n p_n \log_2 p_n.$$

Mathematically, this is pretty straightforward, and we discussed last Wednesday what an entropy of a distribution actually means. For example, if $\langle n \rangle \approx 100$, the entropy lets us know how many bits we generally need to represent the random samples.

In particular, if we measure $100, 90, 105, 98$, and so on, then we can encode this by taking differences from the mean. We expect $\sigma \approx \sqrt{n}$, so we can find a coding that only uses about $\log_2(c\sqrt{n})$ bits! Entropy, then, is the number of bits needed for an optimal encoding.

Fact 143

In most cases, the entropy is basically $\log_2$ of a few times the standard deviation. This means that we expect

$$S = \frac{1}{2}\log_2 n + \log_2 c$$

for some constant c.

Next, let's take a look again at a biased coin: the information-theory entropy is

$$S = -p \log_2 p - (1-p)\log_2(1-p).$$

This makes sense to be maximized at $\frac{1}{2}$: near 0 and 1, it's easy to expect or predict the outcome of the coin flip. So the information-theoretic limit is the Shannon entropy.

How do we construct a better coding scheme than 1 bit each? We can block our flips into larger groups and encode them one by one. We'll find that often, we do better than 1 bit each, but not better than the Shannon theoretical limit!

22.2 Entropy

We know that the entropy of a macrostate is

$$S = k_b \ln \Gamma,$$

where Γ is the number of microstates. It's a powerful concept, but it can also be confusing.

The key concept is that entropy must always get larger. We can only lose information by not carefully controlling our experiment! The **second law of thermodynamics** can be formulated in many ways: $\frac{dS}{dt} > 0$ is one example, but we need to be careful with that.

If we have a particle in a box, the particle can be anywhere, and we can describe the entropy of the particle. (We'll look at this more in the future.) If the particle is in the left part of the box only, though, its entropy will decrease. Since the number of microstates is a factor of $\frac{1}{2}$ smaller due to entropy, we lose a factor of $k_b \ln 2$.

What if we have a delocalized particle that is then measured? We can "catch" the particle in a box of smaller volume. But this is done through a measurement, and through the process, the observer increases the entropy! So the second law applied to isolated systems, not subsystems.

Let's say we have a reservoir that is connected to a system, and we cool down the system (for example, if we put an object in a refrigerator). The entropy flows in the same direction as the heat: the object that cools down will lose entropy as well.

Essentially, our question is "what happens to a system when we add energy?" It will have more microstates, unless we have a pathological system. Intuitively, a system (like a harmonic oscillator) has more energy for larger velocities, because we have a larger sphere of possibilities. This has to do with density of states!

Fact 144

So a system that is cold and heats up will gain entropy: entropy flows in the same direction as heat.

But maybe there are 10^{10} microstates in the reservoir and only 20 microstates in the small system. Let's say we heat up the small system so it now has 100 microstates, and at the same time, the reservoir reduces its microstates to 5×10^9. Is this possible?

Let's look at the total entropy of the reservoir plus the system! Remember that the second law of thermodynamics applies to complete systems: we should consider $S = \log_2 \Gamma$, where Γ is the total number of microstates.

Fact 145

So what we really care about is the **product of the number of microstates**, because we have multiplicity!

Since $20 \cdot 10^{10} < 100 \cdot 5 \times 10^9$, this process is indeed allowed in nature.

Entropy extends beyond energy conservation, though: it also tells us when energetically allowed processes will not happen in nature. For example, is it possible for us to take a reservoir of heat, extract work, with only the consequence that the temperature cools down? Also, is it possible for two objects at the same temperature T to develop a temperature differential?

No, because these things **violate entropy!** Reversibility is an important concept here: free expansion of a gas greatly increases the number of microstates. On the other hand, we know the time-reversibility of the system means the actual number of microstates must be constant, so what does it mean for entropy to increase? How can we go from a small number of microstates to a large number?

People have done many studies of various systems and how they behave in time. These are called **billiard problems**: have particles bouncing off a wall, and if we have a spherical system, it's possible that we may never fill the full region of microstates! On the other hand, more irregular systems may allow much more randomness in phase space. So the important assumption in entropy is that **all microstates are equally likely.** We don't have particles kept on certain trajectories, mostly due to our inability to measure perfectly. Loss of knowledge when going from highly deterministic systems to complete randomness is the central idea of entropy increase. It is in principle possible for small deviations at the microscopic level to happen (fluctuation), but it's often immeasurably small.

23 March 19, 2019

Class is very quiet - maybe everyone is ready for spring break. Our exams are at the front of the room; the problems and solutions will be uploaded tomorrow, because there are two people still taking the exam.

There is a problem set (2 problems) due on Friday.

23.1 Overview and review

As a quick review, we started by reviewing Shannon entropy and thinking of entropy as a thermodynamic quantity connecting the microscopic world with macroscopic problems. Last time, we looked at a two-level system, going from counting states to computing entropy, temperature, and heat capacity. It's interesting, because it's the first system where quantum effects manifest themselves in the macroscopic world. One thing to think about is the unimodal heat capacity - there are important quantum effects there!

We'll define some terms that are useful for describing systems, and we're going to keep thinking about changes in entropy in the PV plane. We have all the tools necessary into understanding fundamental systems! Next lecture, we'll also look at some more quantum systems.

Fact 146

Recall that a **two-level** system has N particles, each of which can be in a lower or higher energy level. If we plot $\frac{n}{N}$ versus k_BT, which is the number of particles in the higher energy state versus energy, the number of particles in the higher energy state increases to $\frac{1}{2}$. If we plot $U/(\varepsilon N)$, where ε is the energy level of the higher energy state, this also saturates at $\frac{1}{2}$. Finally, heat capacity $c/(Nk_B)$ increases and then decreases.

Remember that we found the entropy of this system: as $T \to 0, S \to 0$ [this will be the third law of thermodynamics], and as $T \to \infty$, $S \to Nk_B \ln 2$. Let's try to justify what we see and physically explain each of these graphs!

- At lower temperatures / energies, all energy is in the lower level. But if we add energy to our system, we have particles evenly distributed among all states.
- This is also why internal energy saturates at $\frac{1}{2}\varepsilon N$: half of the particles will be in the ε energy state. (We take groud state to be 0.)
- The heat capacity is harder to explain. At high temperatures, we expect entropy to read its maximum value of $Nk_B \ln 2$, and we have evenly distributed particles. So changing the temperature a little bit does not change the configuration very much, so the heat capacity is very low. This is a **saturation phenomenon**! Meanwhile, at very low temperatures, we have to overcome the gap of energy ε to actually change the state of the system. This **gapped behavior** is also an important characteristic: we should have a vanishing of order $\exp\left[-\frac{\varepsilon}{k_BT}\right]$.

- The maximum value of C (heat capacity) is coming from the **scale temperature**

$$T_\varepsilon = \frac{\varepsilon}{k_B}.$$

This is where we start exciting more and more particles. This is where we have the highest disorder.

By the way, the saturation effect means that we can't have a system where most particles are at the higher energy level. However, we can prepare our system in a way that gives more particles than we normally have at the high energy level! It's called a **metastable state**, and it's interesting for "negative temperature."

23.2 A closer look at the First Law

Remember the First Law of Thermodynamics, which relates an exact differential to inexact differentials:

$$dU = đQ + đW.$$

We know that if work is done in a quasi-equilibrium process (so that we're pretty close to being in an equilibrium state), then pressure is defined throughout the process, so we can write $đW = -PdV$. Meanwhile, if heat is added in a quasi-equilibrum process, temperature is defined throughout, so we can write $đQ = TdS$.

This means we can write the First Law in a bit more detail:

$$dU = TdS - PdV + \mu dN + \sigma dA + \cdots .$$

Recall that we defined (for a system at thermal equilibrium)

$$\left.\frac{\partial S}{\partial U}\right|_{V,N} = \frac{1}{T}.$$

This also tells us

$$\left.\frac{\partial U}{\partial V}\right|_{S,N} = -P$$

(if we plug into $dU = TdS - PdV$, the TdS term goes away), and similarly

$$\left.\frac{\partial S}{\partial V}\right|_{U,N} = \frac{P}{T}$$

which is "volume equilibrium." Remember that we had a fixed partition that only allowed heat transfer between two parts of a system: now, let's see what happens with a movable partition!

23.3 Deriving some important cases

Let's say we have a system with A and B separated by a partition. Now, let's say that A and B can exchange volume V, but they cannot exchange N, the number of particles, or U, the internal energy.

If A and B are at thermal equilibrium, then the entropy is maximized. What changes when we move the partition a bit? volume is being exchanged, so

$$dV = dV_A = -dV_B.$$

Writing out changes in entropy in terms of partial derivatives like we did last time, since we're at a maximum entropy,

$$\left.\frac{\partial S_A}{\partial V_A}\right|_{U_A,N_A} dV_A + \left.\frac{\partial S_B}{\partial V_B}\right|_{U_B,N_B} dV_B = 0,$$

and this means we want $\frac{\partial S_A}{\partial V_A}\Big|_{U_A,N_A} = \frac{\partial S_B}{\partial V_B}\Big|_{U_B,N_B}$. Plugging in, this means $\frac{P_A}{T_A} = \frac{P_B}{T_B} \implies P_A = P_B$ (since we're at thermal equilibrium). So this means that **pressures are equal at equilibrium**! Similarly, if we have a partition where we can exchange only particles but not internal energy or volume, we find that

$$\frac{\partial S}{\partial N}\bigg|_{U,V} = -\frac{\mu}{T}$$

is constant.

23.4 Entropy's role in thermodynamic processes

Remember that an **adiabatic** process has no heat transfer (because it moves too quickly, for example): $dQ = 0$.

Definition 147

An **isoentropic** process has no change in entropy: $\Delta S = 0$.

These two are not interchangeable! It's possible to have zero change in heat but a nonzero change in entropy. For example, free expansion is adiabatic if it happens fast enough and is isolated, but the entropy does increase. Let's also be a bit more specific about some other words we've discussed:

Definition 148

A **quasi-equilibrium** process is one where the state is always near thermodynamic equilibrium, so that state variables are defined throughout the process, meaning we can use the first law. A **reversible** process is a quasi-equilibrium process in which the direction of heat flow and work can be changed by infinitesimal changes in external parameters.

It's not necessarily true that quasi-equilibrium processes are reversible though. For example, processes with friction have some dissipation of energy, which means that we can't do them in reverse and get back to our original state. However, if we do them sufficiently slowly, state functions can be consistently defined.

23.5 Looking at change in entropy in the PV plane

Example 149

Consider the isometric heating of a gas: in other words, we hold the volume constant and $dV = 0$.

This means $dU = \text{đ}Q = T\,dS \implies dS = \frac{dU}{T}$. Therefore,

$$\frac{\partial U}{\partial T}\bigg|_V = T\,\frac{\partial S}{\partial T}\bigg|_V \implies \frac{\partial S}{\partial T}\bigg|_V = \frac{C_V(T)}{T} \implies S(T,V) - S(T_0,V_0) = \int_{T_0}^{T} \frac{dT'}{T'} C_V(T')$$

where $C_V(T)$ is some function of T. Looking at a simple system like the monatomic ideal gas, we have $C_V = \frac{3}{2}Nk_B$, so

$$\boxed{\Delta S = \frac{3}{2}Nk_B \ln\frac{T}{T_0}}.$$

Example 150

Now let's look at the isobaric heating of a gas: keep pressure constant and $dP = 0$.

Remember from our problem set and other discussions that we can define the **enthalpy**

$$H = U + PV,$$

and in our system here, $dH = dU + d(PV) = TdS - PdV + PdV + VdP = TdS + VdP$. Now, because we have an isobaric process,

$$dH = TdS \implies \left.\frac{\partial H}{\partial T}\right|_P = T\left.\frac{\partial S}{\partial T}\right|_P.$$

Rearranging this, we have that $C_P(T) = T\left.\frac{\partial S}{\partial T}\right|_P$, so

$$S(T, P) - S(T_0, P) = \int_{T_0}^{T} \frac{dT'}{T'} C_P(T'),$$

so again for a monatomic ideal gas, $C_P = \frac{5}{2}Nk_B$, and we have

$$\Delta S = \frac{5}{2}Nk_B \ln \frac{T}{T_0}.$$

Example 151

Finally, let's think about isothermal expansion (from a volume V_0 to a final volume V): $dT = 0$.

Let's rewrite $dU = TdS - PdV$ as

$$dS = \frac{1}{T}(dU + PdV) = \frac{1}{T}\left(\left.\frac{\partial U}{\partial T}\right|_V dT + \left.\frac{\partial U}{\partial V}\right|_T dV + PdV\right)$$

(since U is a function of V and T), and this simplifies to

$$dS = \frac{1}{T}\left(C_V dT + \left(\left.\frac{\partial U}{\partial V}\right|_T + P\right) dV\right).$$

Since we have an isothermal expansion, the dT term goes away, and thus

$$dS = \frac{1}{T}\left(P + \left.\frac{\partial U}{\partial V}\right|_T\right) dV.$$

To compute this more easily, we'll expand by volume first:

$$S(T, V) - S(T, V_0) = \int_{V_0}^{V} \frac{dV'}{T}\left(P + \left.\frac{\partial U}{\partial V}\right|_T\right)$$

For an ideal gas, U is only a function of T, so $\left.\frac{\partial U}{\partial V}\right|_T = 0$, and pressure $P = \frac{Nk_BT}{V}$, so

$$S(T, V) - S(T, V_0) = Nk_B \ln \frac{V}{V_0}.$$

Fact 152

Draw a cycle using all three of these processes:

- Start at pressure P_0 and volume V_0, and heat it isometrically to pressure P.
- Isothermally expand the gas from volume V_0 to V.
- Finally, do an isobaric compression from volume V back to volume V_0.

Notice that $\frac{V}{V_0} = \frac{T}{T_0}$ in an ideal gas, and in this case, the change in entropy over the whole cycle is zero!

So as long as we have a well-defined pressure, temperature, and internal energy at all points in our process, we are at thermal equilibrium, and our entropy is a state function. In general, this means we can use any path to calculate our entropy!

24 March 20, 2019 (Recitation)

A lot of interesting new material was covered in class.

24.1 Reviewing the two-level system

Let's look at the two-level system, where we have N particles that can either be at low energy (0) or high energy (ε). The lowest possible energy of the system is 0, and the highest possible is $N\varepsilon$.

We've picked three special states: 0, $\frac{N\varepsilon}{2}$, $N\varepsilon$. Notice that 0 and $N\varepsilon$ have minimum entropy, since there is only 1 possible microstate. In general, the number of microstates for having n_0 objects in the lower state and n_1 in the higher state is

$$\Gamma(n_0, n_1) = \frac{n!}{n_1!n_0!}.$$

This is $\binom{N}{n_1}$: we're choosing n_1 of the N states. This is consistent with the edge cases of 0 and $N\varepsilon$.

Well, what's the entropy when the energy is half of the maximum? We know that if we have N fair coins, we have entropy of N bits, which corresponds to $Nk_b \ln 2$. So we don't need any mathematics to understand the entropy of the system!

What about the temperature of the system? We can argue that it's usually zero at the ground state, because we've "taken out all of the energy." We claim that the temperature at the middle state is ∞ and that the energy at the high state is negative zero!

How do we argue this? We can plot entropy versus energy: it peaks at the middle and starts and ends at 0. So now

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

so the temperature is infinite at $\frac{N\varepsilon}{2}$, because the slope of the S versus E graph is 0. It turns out we have the Boltzmann factor

$$\frac{n_1}{n_0} = e^{-(E_1-E_0)/(k_bT)};$$

the idea is that if two groups are equally populated, and the energy states are different, then the ratio on the left hand side is 1, the exponent must be 0, and a finite energy different means T must be infinite. The idea is that at infinite temperature, energy doesn't matter, because it is free!

Well, notice that the graph has positive slope for $E < \frac{N\varepsilon}{2}$ and negative slope for $E > \frac{N\varepsilon}{2}$. So the temperature is positive at first but then becomes negative!

So let's start plotting other quantities versus temperature. Let's plot only in the range from $T = 0$ to $T = \infty$: this means we only care about energies between 0 and $\frac{N\varepsilon}{2}$. If we plot E versus T, we start at 0 and then asymptotically approach $\frac{N\varepsilon}{2}$.

Fact 153

By the way, note that T, U, E are all parametrized together, so $\frac{\partial S}{\partial U}$ keeps N constant, but there's no other dependencies to worry about.

So what's the specific heat of the system? Specific heat measures how much energy we need to change the temperature, and as $T \to \infty$, the energy is not changing anymore: it's saturated to $\frac{N\varepsilon}{2}$. So at high temperatures, the energy **saturates**, and $C \to 0$.

Fact 154

Here, we have a finite number of possible states and a bounded energy system. This is very different from a gas that can have infinite kinetic energy! We'll also find later that there are usually more energy states for ordinary systems: the spin system is an exception.

What if T is very small? The system wants to be in the ground state, and for very low (infinitesimal) temperature, we only care about the ground state or the first excited state: there's only a possibility of one of the N particles increasing its energy by ε. In the Boltzmann distribution, we have a factor

$$e^{-E/(k_B T)},$$

so the probability of the first energy distribution is proportional to $e^{-\varepsilon/(k_B T)}$. This is something that characterizes any system with a ground state and an excited state with an energy gap of ε!

Then we just use the physics of combinatorics: we'll discuss this as partition functions later, but this exponential factor is universal across all systems like this. So systems at low temperature show exponential behavior, and if we plot energy versus temperature, we'll start off with an exponential behavior. This is what we call **gapped behavior**.

Example 155

What's an example of a system without gapped behavior? In other words, when can the excited state have arbitrarily small energy?

We want to say a classical gas, where the kinetic energy $\frac{1}{2}mv^2$: the velocity v can be arbitrarily small. But not really: remember that a particle in a box has a smallest energy $\frac{\hbar^2}{L^2}$, so we do need to take the classical limit or use a very big box. Then we'll see that the heat capacity does not have that exponential curve near $T \to 0$.

24.2 Negative temperature?

Let's look back at T being negative for high energy (in our system, where $E > \frac{N\varepsilon}{2}$. We've always heard that we can't cool down and get to negative temperature: there's an absolute zero. The idea is that there's a lowest energy state, and there's no way to get less energy that that lowest state.

So negative temperature doesn't mean we can keep cooling to negative Kelvin. In fact, this system tells us that negative temperatures appear in a different way! We increase the energy of the system, and the temperature increases to ∞ and then goes negative. So somehow ∞ temperature and $-\infty$ temperature are very close to each other! This is because we inherently defined our temperature as a slope being equal to $\frac{1}{T}$, and often, Boltzmann factors give $\frac{1}{T}$ as well.

So if we get confused, think about $\frac{1}{T}$ instead! Plotting energy versus $-\frac{1}{T}$ (so that we start off with positive temperature on the left), we have our energy increasing from 0 to $\frac{N\varepsilon}{2}$ at $-\frac{1}{T} = 0$, and then it further increases from $\frac{N\varepsilon}{2}$ to $N\varepsilon$! This is a continuous curve, and everything is smooth between positive and negative temperatures.

In other words, the connection between positive and negative temperatures does not happen at 0: it happens at ∞. This is only possible because we have a maximum energy state! If we were to do this with a system like a classical gas, the energy would go to ∞ as $T \to \infty$. So there wouldn't be a connection to negative temperature there.

25 March 21, 2019

Today's physics colloquium is about turbulence, which is a very challenging problem. Part of Professor Fakhri's group is looking at biological signals, and it turns out the dynamics of the signaling molecules in the cell follow a class of turbulence phenomena that is connected to quantum superfluid turbulence.

Remember there's a problem set due tomorrow, and there will be another one posted soon.

25.1 Overview

Today, we're going to start discussing some general systems where we can do **counting of states.** The idea is to do thermodynamics and statistical physics in systems like a two-level system! We start by counting states to find multiplicity, and from there we can compute entropy, temperature, and heat capacity.

In particular, we'll look at a particle in a box, as well as the quantum harmonic oscillator. We'll introduce an idea called the **density of states**, which will tell us important information about the internal energy of the system! This is not a lecture on quantum mechanics, but we'll get some descriptions that are related.

By the way, there is a way to count a density of states in a classical system too, even though the states are continuous! We'll treat them in a semi-classical way.

25.2 Quantum mechanical systems: the qubit

Recall that in classical mechanics, the number of states is infinite: for example, a system of N particles depends on the position and momentum $(\vec{x}_i, \vec{p}_i)$, and these quantities (which define states) are continuous variables.

However, in quantum mechanical systems, states are quantized and finite, so they are countable! We'll learn how to label and count some simple examples: the **qubit**, **quantum harmonic oscillator**, and the **quantum particle in a box**.

A **qubit** is a system with only two states: $|0\rangle = |+\rangle = |\uparrow\rangle$ and $|1\rangle = |-\rangle = |\downarrow\rangle$. Here, $|\rangle$ refers to a set of states. The higher energy state $|+\rangle$ has energy E^+, and the lower energy state $|-\rangle$ has energy E^-.

The simplest example of a qubit is a spin $\frac{1}{2}$ particle: for example, an electron in a magnetic field. We know (or are being told) that the magnetic moment

$$\vec{\mu} = -g\left(\frac{e}{2mc}\right)\vec{s},$$

where $g \approx 2$ for an electron, e is the charge of the electron, m is the mass of the electron, and c is the speed of light. Then when a particle is placed in a magnetic field $\vec{H}$, the energy is

$$E = -\vec{\mu} \cdot \vec{H} = \frac{e}{mc} \vec{s} \cdot \vec{H}.$$

In particular, the different energy states are

$$E^+ = \frac{e\hbar}{2mc} H, E^- = \frac{-e\hbar}{2mc} H,$$

where $\hbar$ is Planck's constant $\frac{h}{2\pi} \approx 1.054 \times 10^{-34} J \cdot s$. There are many systems that behave like this at low energy!

Fact 156

These are interesting derivations, but for those of us who haven't taken 8.04, we shouldn't worry too much about it.

In general, if we have a quantum particle with spin j in a magnetic field, the different possible energy states are

$$E_m = \frac{e\hbar g}{2Mc} mH,$$

where $m = j, j-1, \cdots, -j$. This has to do with systems exhibiting paramagnetic properties!

25.3 Quantum harmonic oscillator

Let's start by looking at a classical example. We have a harmonic oscillator with spring k and mass m, where the potential energy $V(x) = \frac{1}{2}kx^2 = \frac{1}{2}m\omega^2 x^2$ (if the natural frequency is $\omega = \sqrt{\frac{k}{m}}$.

This is used to explain vibrations of solids/liquids/gases, as well as other material properties. The energy (or Hamiltonian) of this system can be written classically as

$$H = \frac{p^2}{2m} + \frac{1}{2}kx^2,$$

and we find that $x(t) = A\cos(\omega t + \phi)$, and the energy of this system is a continuous quantity $m\omega^2 A^2$.

On the other hand, in the quantum version, we can label our quantum states as $|n\rangle$, for $n = 0, 1, 2, \cdots$, and there's only a few allowed energies. They are equally spaced in units of $\hbar\omega$:

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega,$$

where the lower allowed energy $\frac{1}{2}\hbar\omega$ is the **zero point energy**. We can also have a bunch of non-interacting particles in the same harmonic oscillator potential: then we just add the individual energies.

Fact 157

The energy of a set of non-interacting particles moving in a harmonic oscillator potential is (for example)

$$E(n_1, n_2, n_3) = \hbar\omega\left(n_1 + n_2 + n_3 + \frac{3}{2}\right).$$

The states here are denoted as $|n_1, n_2, n_3\rangle$, where each $n_j = 0, 1, 2, \cdots$. It's pretty easy to see how this generalizes.

Well, there's a one-to-one correspondence between a one-dimensional system with N particles and an N-dimensional harmonic oscillator for one particle: the energy states look the same!

$$E_N(n_1, \cdots, n_N) = \hbar\omega \left(n_1 + \cdots + n_N + \frac{N}{2} \right).$$

25.4 Particle in a box

This system has applications to an ideal gas in a box, as well as to black-body radiation and Fermi and Bose gases. Basically, we care about the limit of a large box.

Consider a particle with mass M in a three-dimensional cubic box of length L. Quantum energy states can be labeled $|n_1, n_2, n_3\rangle$ by three integers! We can show that including boundary conditions,

$$E(n_1, n_2, n_3) = \frac{\pi^2\hbar^2}{2ML^2}(n_1^2 + n_2^2 + n_3^2).$$

These energy levels are very closely spaced! Let's compute this for an oxygen molecule. If the box measures 1 mm on a side,

$$E \approx \frac{10 \times 10^{-68}}{2 \times (32 \times 1.67 \times 10^{-27}) \cdot 10^{-6}} \approx 10^{-36} J \approx 6 \times 10^{-16} eV.$$

These are very small numbers: compare this to $k_B T$ at room temperature

$$k_B T = \frac{1}{40} eV.$$

So most of the time, we can assume the energy levels are almost smooth (since we often compare energy to $k_B T$ in our calculations), which makes our work a lot easier!

25.5 Counting states and finding density

Let's look at our particle in a box.

Fact 158

The math will be more complicated from this point on. We should review theta and delta functions!

Let's say we want to find the number of energy states $N(E)$ in a box of volume V with energy less than E. Once we find $N(E)$, the number of states, we can differentiate it to find $\frac{dN}{dE}$, which will give the number of states between E and $E + dE$ (this is similar to how we differentiated a cumulative distribution to get a probability distribution function!).

Note that in our cumulative distribution $N(E)$ will be stepwise (since we have some finite number of states for each energy level), and we will normalize it by dividing by $\frac{\pi^2\hbar^2}{2ML^2}$. But as we go to higher temperatures, the steps are very small compared to $k_B T$, so we can do a smooth interpolation! The idea is that $\frac{dN}{dE}$ is a sum of delta functions (because N is a bunch of step functions), but we can approximate those with a smooth curve as well. This is called a **delta comb**, by the way.

So now

$$N(E) = \sum_{n_1, n_2, n_3} \theta \left(E - \frac{\pi^2\hbar^2}{2ML^2}(n_1^2 + n_2^2 + n_3^2) \right),$$

since we only count a state if the total energy is at most E. Doing our interpolation, the sum becomes an integral: since n_1, n_2, n_3 can take on positive values,

$$N(E) = \int_0^\infty dn_1 \int_0^\infty dn_2 \int_0^\infty dn_3 \theta\left(c^2 - n_1^2 - n_2^2 - n_3^2\right),$$

where $c^2 = \frac{2ML^2E}{\pi^2\hbar^2}$, just to make the calculations a bit easier. To evaluate this integral, let's let $n_j = cy_j$:

$$N(E) = c^3 \int_0^\infty dy_1 \int_0^\infty dy_2 \int_0^\infty dy_3 \theta(1 - y_1^2 - y_2^2 - y_3^2).$$

Notice that this is one-eighth of the volume of a unit sphere! This is because we want y_1, y_2, y_3 to be positive and $y_1^2 + y_2^2 + y_3^2 \le 1$. So the volume is just $\frac{\pi}{6}$, and

$$N(E) = \frac{\pi}{6}c^3 = \frac{\pi}{6}\left(\frac{2ML^2E}{\pi^2\hbar^2}\right)^{3/2}$$

and since $L^3 = V$, we get a factor of V out as well, and we can simplify further.

For any quantum system, if $E = E(n_1, \cdots, n_N)$, then

$$N(E) = \sum_{\{n_j\}} \theta(E - E(\{n_j\})).$$

Then we can differentiate

$$\frac{dN}{dE} = \sum_{n_j} \delta\left(E - E(\{n_j\})\right).$$

So in this case, since

$$N(E) = \frac{\pi}{6}\left(\frac{2M}{\pi^2\hbar^2}\right)^{3/2} VE^{3/2},$$

we have

$$\frac{dN}{dE} = \frac{\pi}{4}\left(\frac{2M}{\pi^2\hbar^2}\right)^{3/2} VE^{1/2}.$$

Fact 159

In general, if E_k occurs some g_k number of times, this is called a **degeneracy factor.** Since we want to count all the states, we get a g_k factor in front of our theta function! This is better explained with examples, though.

Example 160

Let's count the density of states in a quantum harmonic oscillator.

Then $E_n = \left(n + \frac{1}{2}\right)\hbar\omega$, and the number of states $N(E)$ is a step function that increases by 1 at $\frac{1}{2}\hbar\omega$, $\frac{3}{2}\hbar\omega$, and so on. So

$$N(E) = \sum_{n=0}^{\infty} \theta\left(E - \hbar\omega\left(n + \frac{1}{2}\right)\right)$$

If $E \gg \hbar\omega$, we can again make this into an integral:

$$N(E) = \int_0^\infty \theta\left(\frac{E}{\hbar\omega} - n - \frac{1}{2}\right) dn \approx \frac{E}{\hbar\omega},$$

so we have $\frac{dN}{dE} = \frac{1}{\hbar\omega}$.

Next time, we'll talk about how classical systems can still be counted! This is the idea of a semi-classical description.

26 April 1, 2019 (Recitation)

Today's recitation is being taught by Professor England.

Let's go back to discussing the ideas of entropy! Remember that we started off by discussing information entropy, but now we want to start developing entropy more as a thermodynamic and statistical quantity.

26.1 Rewriting the first law

Earlier in the class, we found that we could describe our energy dU as a product of the generalized force and generalized displacement:

$$dU = \sum_i \left(\frac{\partial U}{\partial x_i} \right) dx_i.$$

Here, we referred to TdS as "heat" dQ, and PdV, as well as other terms, as "work" dW. But our $dS = \frac{đQ}{T}$ could be interpreted differently: if we define Ω as a function of state variables $U, V, N, \cdots$ as the number of different microstates consistent with our data, we could also define

$$S = k_B \ln \Omega.$$

There's a lot of plausibility arguments that can support our theory working this way! For example, two objects in contact have maximum entropy when they have the same temperature, and indeed we want entropy to increase as time evolves.

So let's run with this, and let's see if we can rewrite the first law in a more sensible way. Since S is proportional to $\ln \Omega$, we have $dS = k_B d(\ln \Omega)$: plugging this into the first law,

$$d(\ln \Omega) = \frac{dU}{k_B T} + \frac{PdV}{k_B T} - \frac{\mu dN}{k_B T} \cdots$$

This gives us another way to think of how to "count" our states! We can ask questions like "how does a flow of heat contribute to our entropy?" or "how does entropy change if I increase the volume?" So the right hand side is a bunch of levers that we can pull: now we can see how our statistical quantity changes when I adjust my other terms!

26.2 Applications to other calculations

We can go back to our ideal gas model, and we'll see that the calculations fall out pretty easily!

Let's say we have N particles in a box of volume V, and we want to think about the number of microstates we can have.

Fact 161

Scaling is very useful; we don't have to be too exact with our calculations.

Microstates are dictated by two variables: the position and the momentum of the individual particles. It's true that if we tried to count the number of discrete "float numbers" that would work, we'd have infinitely many possibilities for $\vec{x}$. But there's still a sense in which having twice as much volume gives twice as many possible positions, so the number of states here is basically proportional to V! Similarly, momentum, which is independently assigned to particles of position, has some function $a(U)$ in terms of the internal energy. Since we have N particles, the number of microstates here is proportional to

$$\Omega = (cV \cdot a(U))^N.$$

Fact 162

One other way to patch this up is with the uncertainty principle: $dpdx \sim \hbar$: we do need discrete states in that case.

To get to entropy from here, we take a logarithm and multiply by k_B (which is just a units conversion factor):

$$k_B \ln \Omega = N k_B (\ln V + \ln a(U) + c).$$

Notice that the constant term c comes from the fact that we only care about differentials and differences in the first law! For example, pressure is really just

$$\frac{P}{k_B T} = \left(\frac{\partial(\ln \Omega)}{\partial V}\right)_{U,N,\cdots}.$$

Let's try plugging that into our expression for $\ln \Omega$: this yields

$$\frac{P}{k_B T} = \frac{N}{V},$$

which rearranges to the ideal gas as desired!

Fact 163

We can always take a derivative (in principle) fixing all the other variables. It might be experimentally difficult to do this, though.

So another way to say this is that the ideal gas is just related to the scaling of states with volume.

But there's something wrong with the entropy we're using: we wanted the change in entropy to be an extensive quantity. U is extensive: putting two copies of the system next to each other doubles the energy. Since U and $\ln \Omega$ play similar roles in the first law, we want them to both be extensive. But $\ln \Omega = N \ln V$, but doubling the size of the system doubles both N and V, and this doesn't give exact extensivity. What's going on here?

Fact 164

This is called Gibbs' paradox! The idea is that we treated our particles as different: if particles 1 and 2 are on the boundary and 3 is in the middle, this is different from 1 and 3 on the boundary and 2 in the middle. But there's really no way for us to be able to distinguish our particles!

So it seems that there are $N!$ permutations, and we want to divide through by $N!$. But this isn't actually an exact answer! Remember that we have quantized states, so we're not overcounting particle states when many particles are close to each other. So we'd run into problems where dividing by $N!$ is artificially penalizing particles nearby - we'll return to those ideas for now.

But we'll deal with dilute gases, since we're doing ideal gas calculations anyway, and then it is exact to say that

$$\Omega = \frac{(V \cdot a(U))^N}{N!}.$$

With Stirling's approximation, this gives

$$\ln \Omega = N \ln V - N \ln N + N = N \ln \frac{V}{N} + N.$$

Now this is an extensive quantity! $\frac{V}{N}$ is an intensive quantity: it's related to the density, and now scaling the system by a factor of 2 does scale $\ln \Omega$ by a factor of 2 as well.

Example 165

If we have two independent identical systems, they have the same number of possible microstates Ω, so the total system has Ω^2 possible states: this does yield twice the entropy. On the other hand, if we put the two systems to together, what happens to the entropy? Doesn't it increase again?

This is again an issue of Gibbs' paradox! The particles in the two systems are indistinguishable, so we should really think a bit more about those $N!$-type terms. Keep this in mind when we start thinking about quantum systems (particularly Bose-Einstein condensates), though.

26.3 Relating this to temperature

Let's think a bit more about our ideal gas system. The internal energy of this system is

$$U = \sum_{i=1}^{3N} \frac{p_i^2}{2m},$$

since each particle has 3 degrees of freedom. So if we have some given energy U, how many possible arrangements $\Omega(U, V, N)$ are there? Basically, how does Ω scale with U?

Let's do the cheap thing: we have a bunch of independent coordinates p_i whose squares add to some constant, which means that the momentum coordinates in phase space are on a hypersphere of radius $\sqrt{2mU}$. Really, we care about scaling: how does the size of a sphereical shell change with volume? Because the volume of the whole sphere goes as r^{3N}, the boundary goes as r^{3N-1}, but we can ignore the 1 if N is really big. So that means

$$\Omega(U, V, N) \sim \frac{V^N}{N!}\sqrt{2mU}^{3N} \sim \frac{V^N}{N!}U^{3N/2}.$$

Going back to our first law equation, let's look at the part

$$d(\ln \Omega) = \frac{dU}{k_B T} + \cdots$$

and differentiate with respect to U. This yields

$$\left(\frac{\partial(\ln \Omega)}{\partial U}\right)_{V,N} = \frac{1}{k_B T} = \left(\frac{\partial(\ln U)}{\partial U}\right)\frac{3N}{2} \implies U = \frac{3Nk_BT}{2}.$$

This gives us the temperature definition out of our calculation of microstates as well!

26.4 Increasing entropy

Finally, let's do another look at the second law of thermodynamics. This law is initially encountered empirically: given a process that we design and take a system through, we can (for example) draw a loop in our PV diagram. In these cases, the total entropy will always increase: ΔS of the environment plus ΔS of our system cannot get smaller.

So this is a very empirical law, but our new description gives us a way to think about this more clearly! (By the way, it's important to note that this is always an average law.) So suppose we have a surface of states in U, V, N space, and we follow our system by doing dynamics with Hamilton's equations. What does it really mean for entropy to increase? There's two notions here: one is to count the number of states on our surface and take the logarithm, but that doesn't really tell us anything in this case.

Instead, we can think of some quantity R: maybe this is "the number of particles on the left side" or "some measure of correlation with a 3-dimensional figure." The point is that we can assign a value of R to all microstates, and now we can group our microstates together by values of R: this is called "coarse-graining."

So what's $\Omega(R)$? Much like with Ω, it's the number of microstates consistent with some value of R, and we can define an "entropy" $k_B \ln \Omega(R)$ along with it. The concept is that R will eventually stabilize to some quantity - for example, particles are not likely to unfill a region.

That means $S(R)$ will increase generally - it may fluctuate, just as R does, but S is an average quantity. One way to think of this is that as we move in our state space, we'll generally get stuck in the biggest volumes of R: it's much more likely that we'll end up there on average! Analogously, two systems at thermal equilibrium can go out of equilibrium, but they're likely to return back into equilibrium (because it is thermodynamically favorable to do so). So entropy can decrease, but that is usually nothing more than a statistical fluke.

27 April 2, 2019

The energy level of the class is higher after the break! There's a problem set due this Friday; we should make sure we understand the material from the required reading.

There will be an exam on April 17th, which is similar to last exam. We'll start hearing announcements about it soon.

Recall that we started by introducing information and thermodynamic entropy. We found that counting the number of microstates consistent with a macrostate gives us a measure of entropy as well, and then we could calculate the temperature, heat capacity, and many equations of state. This then led us to introducing a few different systems, where we could actually count states.

We discussed a two-level qubit, particle in a box, and quantum harmonic oscillator, because these all have discrete states. Today, we'll talk about how to extend this to classical systems and give them this kind of treatment as well. The idea is to count states to find the density of states as well, which gives us a probability distribution.

27.1 Returning to the particle in a box

As we calculated last time, the density of states depends on the energy E of the system:

$$\frac{dN}{dE} = \frac{1}{4\pi^2\hbar^3}(2M^{)}3/2VE^{1/2}.$$

Our goal is to find a connection here to a classical system: let's say we have some volume dV that is infinitesimal on the classical scale but large enough to contain many quantum states. Then for our box,

$$dV = dxdydz = d^3x$$

(for simplicity). If we rewrite E in terms of momentum as $\frac{p^2}{2m}$, then $dE = \frac{pdp}{m}$; substitute this into our equation above, and we can pull out a factor of 4π:

$$dN = \frac{4\pi}{8\pi\hbar^3}p^2dpd^3x$$

Now notice that the $4\pi p^2dp$ looks a lot like spherical coordinates: $d^3p = p^2dp\sin\theta d\theta d\phi$, and integrating out θ and ϕ gives the result we want. So that means (because $\hbar = \frac{h}{2\pi}$),

$$dN = \frac{1}{h^3}d^3xd^3p.$$

This means that dx and dp, which are classical degrees of freedom, yield something about our states dN, just with a normalization factor $\frac{1}{h^3}$! So looking back at

$$\frac{dN}{dE} = \frac{1}{4\pi^2\hbar^2}(2M)^{3/2}VE^{1/2},$$

we can think of $\frac{1}{h^3}$ as a "volume" that contains one state. This means that we can go from classical systems to semi-classical descriptions: our "volume" in phase space is

$$h^3 = \Delta x \Delta p_x \Delta y \Delta p_y \Delta z \Delta p_z.$$

In other words, we can think of our volume h^3 as contain 1 state in our phase space.

Example 166

Let's go back to the density of states calculation for a quantum harmonic oscillator, but we'll start with a 1-dimensional classical harmonic oscillator and do this trick - let's see if we end up with the same result.

The number of states with an energy less than E is

$$N(E) = \int \frac{dxdp}{2\pi\hbar}\theta\left(E - \frac{p^2}{2m} - \frac{1}{2}kx^2\right)$$

(where $h = 2\pi\hbar$ is the normalization factor as before). Writing $\frac{1}{2}kx^2 = \xi^2$, $\frac{p^2}{2m} = \eta^2$, our number of states can be written as

$$N(E) = \frac{1}{\pi\hbar\omega}\int d\eta d\xi \theta(E - \eta^2 - \xi^2)$$

and since the integral is the area of a circle with radius $\sqrt{E}$, this is just

$$N(E) = \frac{\pi E}{\pi\hbar\omega} = \frac{E}{\hbar\omega},$$

which is the same result that we got with the quantum harmonic oscillator! Then we can find $\frac{dN}{dE}$, which will give the number of states with energy between E and $E + dE$. We can think of our $\frac{dxdp}{2\pi\hbar}$ idea as a "volume normalization."

27.2 A useful application

Now, we're going to use this counting-states argument to look at thermodynamics of an ideal gas. Remember that we could count our states in our two-level system if we were given the energy of our overall system, so let's try to think about a **microcanonical ensemble**: systems that are mechanically and adiabatically isolated with a fixed energy. We'll discuss Sackur-Tetrode entropy along the way as well.

First of all, our goal is to find the multiplicity: how many states $\Gamma(U, V, N)$ are consistent with a given U, V, N? Once we know this, we'll write our entropy $S(U) = k_B \ln \Gamma$ (at thermal equilibrium), and then we can calculate our temperature with the familiar equation

$$\left.\frac{\partial S}{\partial U}\right|_{N,V} = \frac{1}{T}$$

to find our energy U in terms of V, T, N (and by extension the heat capacity). Finally, we'll find the equation of state by taking some derivatives.

Let's use the model where we have N molecules of a monatomic gas in a box of side length L. Then the energy of our system (as a particle in a box) is

$$E(n_x, n_y, n_z) = \frac{\pi^2\hbar^2}{2ML^2}(n_x^2 + n_y^2 + n_z^2).$$

To count the number of states, let's assign a vector $\vec{n}_j$ to each particle, where $1 \le j \le N$. Every particle contributes to the state, but since the particles are indistinguishable, we overcount by a factor of $N!$. (We'll return to this idea later.) So the multiplicity is

$$\Gamma(U, V, N) = \frac{1}{N!}\sum_{\vec{n}_1}\sum_{\vec{n}_2}\cdots\sum_{\vec{n}_N}\delta\left(\frac{2ML^2U}{\pi^2\hbar^2} - \sum \vec{n}_j^{\,2}\right)$$

since we want to only pick out the states with a certain energy. We'll find that we have very closely spaced states, so we can approximate this sum as an integral from 0 to ∞. Let's then replace this as half of the integral from $-\infty$ to ∞ by symmetry: since each of N particles has 3 degrees of freedom, and there are 3 integrals, this means we now have an integral over all degrees of freedom

$$\Gamma(U, V, N) = \frac{1}{2^{3N}N!}\int d^3n_1 \cdots d^3n_N \delta(R^2 - \sum_{j=1}^{N}\vec{n}_j^{\,2}),$$

where $R^2 = \frac{2ML^2U}{\pi^2\hbar^2}$. So we now have $3N$ variables (basically, we want the surface area of a $3N$-dimensional sphere of radius R): let's replace our $\vec{n}_i$s with the $3N$ variables $\xi_1, \xi_2, \cdots, \xi_{3N}$. Our integral is now

$$\Gamma = \frac{1}{2^{3N}N!}\int d^{3N}\xi\delta(R^2 - \vec{\xi}^2)$$

and normalizing by letting $\xi_j = Rz_i$, this becomes

$$\Gamma = \frac{1}{2^{3N}N!}R^{3N-2}\int d^{3N}z\delta(\vec{z}^2 - 1).$$

(The $3N - 2$ comes from $\delta(cx) = \frac{1}{c}\delta(x)$.) Since N is very large, we can approximate $3N - 2$ as $3N$, and so now we want to deal with the integral

$$\int d^{3N}z\delta(\vec{z}^2 - 1) = \frac{1}{2}\int d^{3N}\delta(|\vec{z}| - 1).$$

The integral is now the surface area of a $3N$-dimensional sphere

$$S_{3N} = \frac{3N}{2}\frac{\pi^{3N/2}}{\left(\frac{3N}{2}\right)!}$$

so we can plug that in: this yields

$$\Gamma = \frac{R^{3N}}{2^{3N-1}N!}S_{3N}$$

and we're done with our first step!

27.3 Calculating the entropy of the system

Now, let's start calculating entropy and our other relevant quantities. Since $S = k_B \ln \Gamma$, we can plug in the values we have: we'll approximate 2^{3N-1} as 2^{3N}, and by Stirling's approximation,

$$S = k_B\left[3N\ln\frac{R}{2} - N\ln N + N + \frac{3N}{2}\ln\pi - \frac{3N}{2}\ln\frac{3N}{2} + \frac{3N}{2}\right].$$

We'll factor out an N and combine some other terms as well:

$$S = Nk_B\left[3\ln\frac{R}{2} - \ln N + \frac{3}{2}\ln\pi - \frac{3}{2}\ln\frac{3N}{2} + \frac{5}{2}\right].$$

We'll now substitute in our value of $R = \frac{\sqrt{2mL^2U}}{\pi\hbar}$:

$$S = Nk_B\left(\frac{3}{2}\ln U + 3\ln L + \frac{3}{2}\ln\frac{M}{2\pi^2\hbar^2} + \frac{3}{2}\ln\pi - \frac{5}{2}\ln N - \frac{3}{2}\ln\frac{3}{2} + \frac{5}{2}\right).$$

We'll now combine the terms with a $\frac{3}{2}$ in front: since $3\ln L = \ln L^3 = \ln V$,

$$\boxed{S = k_BN\left(\frac{3}{2}\ln\frac{U}{N} + \ln\frac{V}{N} + \frac{3}{2}\ln\frac{M}{3\pi\hbar^2} + \frac{5}{2}\right)}$$

and now we've found our entropy in terms of the variables we care about! This is called the **Sackur-Tetrode entropy**.

27.4 A better analysis: looking at temperature

So let's try to think about the consequences of having all of the different parts here. If we didn't have the $\hbar$ term in our entropy, imagine we take $\hbar$ to 0. Then our entropy S goes to infinity, and we know that this isn't supposed to happen. The idea then is that there is some length scale where quantum effects become obvious!

Also, entropy is an extensive quantity: we have the N in front of our other terms, and that's why the $N!$ correction term was important for us to include. Notice that if we double our volume, $\Delta S = Nk_B\ln 2$ (exercise), so this is consistent with what we want!

So now let's calculate temperature:

$$\left.\frac{\partial S}{\partial U}\right|_{V,N} = \frac{1}{T} = \frac{3Nk_B}{2U} \implies U = \frac{3}{2}Nk_BT$$

which is the same result we had before: this is consistent with the equipartition theorem! Finally, since $TdS = PdV + c$ (where c is other terms that are not relevant to S and V),

$$\frac{P}{T} = \left.\frac{\partial S}{\partial V}\right|_{U,N} = \frac{Nk_B}{V} \implies PV = Nk_BT,$$

which is the equation of state for an ideal gas! Similarly, we know that $-\mu dN = TdS + c'$ (where c' is other terms that aren't relevant to N and S),

$$-\frac{\mu}{T} = \frac{S}{N} - \frac{5}{2}k_B \implies \frac{5}{2}Nk_BT = TS + \mu N,$$

and this can be rewritten as

$$\frac{3}{2}Nk_BT + Nk_BT = \boxed{U + PV = TS + \mu N},$$

which is known as **Euler's equation**.

27.5 Microcanonical ensembles

Definition 167

A **microcanonical ensemble** is a mechanically and adiabatically isolated system, so the total energy of the system is specified.

This means that in phase space, all possible microstates - that is, all members of the ensemble - must be located on the surface $H(\mu) = E$. At thermal equilibrium, all states are equally likely, so the probability of any given state is

$$P_E(\mu) = \begin{cases} \frac{1}{\Gamma(E)} & H(\mu) = E \\ 0 & \text{otherwise.} \end{cases}$$

Next class, we'll comment a bit more about uncertainty in this surface (to create volumes)!

28 April 3, 2019 (Recitation)

28.1 The diffusion equation

Let's start with a concept from the problem set. We've always been discussing thermodynamic entropy as

$$S = -k_B \sum p_i \ln p_i = k_b \ln \Gamma$$

(where Γ is the number of microstates) if all states are equally likely.

But now when we look at the diffusion equation, we're given a slightly different equation for a gas in one dimension:

$$S = -k_B \int \rho(x,t) \ln \rho(x,t) dx.$$

What's the connection between the two? If we take some small volume ΔV, then the probability that we find a molecule in that volume at position x is

$$p(x) = \rho(x) \Delta V.$$

So this equation is just the probability distribution entropy, ignoring the normalization constant!

Next question: let's look at the diffusion equation more carefully:

$$\frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial x^2}.$$

Can we argue why this always increases the diffusion of the gas from general principles? If we spread our probability over more possibilities, then entropy goes up: the larger Γ is, the more microstates we have, which means all individual probabilities are small.

Well, diffusion makes our peaks in ρ go down! So the probability distribution is getting wider and flatter, and this is a lot like making all of our probabilities go to $\frac{1}{n}$: we're moving toward "equilibrium."

28.2 Black holes

There's a lot of physics going on here! We have a relation between relativistic mass and energy from special relativity: $E = Mc^2$, and we have the Schwarzchild radius from general relativity; $R_s = \frac{2GM}{c^2}$. Black holes are a good place to do research, because they are a system where we can make new discoveries: there are big open questions about combining quantum physics with general relativity.

Here's one interesting idea: if there's nothing to say about a black hole other than its angular momentum and mass, then matter with entropy sucked into that black hole has disappeared: doesn't this mean entropy has decreased? We can never combine two microstates into one, so what's going on here?

Turns out black holes aren't completely black! They are actually at a "temperature"

$$T_h = \frac{\hbar c^3}{8\pi G k_B M}.$$

This is called "Hawking radiation," and this is a way for black holes to communicate with the outside world (using electromagnetic radiation). This means that we can actually have an entropy for black holes, and the contradiction is gone!

Fact 168 (Sort of handwavy)

This is a bit past standard explanations, but the idea is that the Schwarzchild radius is an "infinite potential:" no particles can get past the event horizon. But adding quantum fluctuation (due to quantum field theory "spontaneously" producing particle-antiparticle pairs), we can create virtual photons at the boundary: one with positive energy can escape, and another with negative energy gets sucked into the black hole.

So we can think of this as having a black-body radiation spectrum, where there is a peak at the temperature

$$\hbar\omega = k_B T_{\text{BB}}.$$

So now you can integrate to find the entropy of a black hole: we can find that it is related to the surface area of a sphere with Schwarzchild radius! This is relevant to understanding complicated materials: for example, the entropy of most systems (like an ideal gas) is dependent on volume. Black holes tell us a different story: entropy doesn't necessarily scale with volume. After all, $\frac{1}{M}$, so the derivative $\frac{\partial S}{\partial E}$ is proportional to E. So S is proportional to E^2, which is proportional to R_s^2!

So the fact is that we have a complicated system for which only the surface matters: this scales much slower than volume. Thus, describing superconductors and other materials is motivated by this discussion of a black hole! This is called the "holographic principle."

28.3 Negative temperatures

Let's start by talking about heat transfer: it's about energy conservation and increase in entropy. Usually, we have systems coupled to a reservoir: those reservoirs have a lot of energy, but we don't want to go against what nature wants! So then a "deal" is made: if the reservoir gives an energy $\Delta E = T\Delta S$, then the entropy of the reservoir decreases by $\Delta S_R = \frac{\Delta E}{T}$. So the system just needs to make sure that the entropy created by the system is larger: $|\Delta S_S| \geq |\Delta S_R|$, and then all the energy in the world can be transferred. This has to do with free energy, which we'll talk about later!

So let's say that we transfer the energy from our reservoir R to our system S, and this just heats up our system S. Then the increase in entropy of the system is

$$\Delta S_s = \frac{\Delta E}{T_S}:$$

then if we want total entropy to increase, clearly this only happens if $T_S \leq T_R$. So it's only thermodynamically favorable to transfer heat ΔQ from higher temperature to lower temperature. But what happens when negative temperatures are involved?

Remember that when we drew our graph of S versus E for the two-level system: the temperature is $\frac{1}{\text{slope}}$. The slope goes from positive to 0 and then 0 to some negative value, which avoids the singularity of ∞ temperature! So now we want $\frac{1}{T_S} > \frac{1}{T_R}$: regardless of whether we have two systems at positive temperature or two systems at negative temperature, they will want to move "towards" each other. But if we look at the total entropy, the equation we care about is that heat is transferred from R to S if and only if

$$\frac{1}{T_R} < \frac{1}{T_S}.$$

But now the question: if we bring a temperature with negative temperature in contact with one of positive temperature, what will happen? $\frac{1}{T}$ is continuous in this case, and two sysetms always want to equilibriate! So what is going on here is that systems at negative temperature gain entropy when you lose energy! The reservoir pays for both the energy and the entropy, so it will willingly transfer energy ΔE: then the entropy of both the reservoir and the system both go up.

So what happens is that if a system with negative temperature is in contact with an ideal gas with some positive temperature, then the gas will always gain more energy!

29 April 4, 2019

This is a reminder that the problem set is due tomorrow!

29.1 Brief overview

Last class, we used our knowledge of quantum systems (namely the particle in a box) to determine thermodynamics of an ideal gas. The idea is that the internal energy of the system is fixed, so we can count the number of possible states: then because at thermal equilibrium, all microstates are equally likely, we can just use the formula $S = k_B \ln \Gamma$ to find the entropy. From there, we could find the temperature of the system, as well as the general equation of state. This then led us to the definition of a microcanonical ensemble.

Today, we're going to go back to how thermodynamic started: using heat engines! We'll see how to extract work from those engines, and we'll find a bound on the efficiency on such engines. From there, we'll start looking beyond having "isolated systems:" we'll see how to do statistical physics with just parts of systems, which will lead us to the idea of free energy.

By the way, the question of $N!$ in our calculations from last week is an important idea: this will be a problem on our next problem set.

29.2 The third law of thermodynamics

We've talked quite a bit about entropy as information and also in terms of heat transfer, but let's start looking beyond just those ideas!

Theorem 169 (Third law of thermodynamics)

The following are equivalent statements:

- As temperature $T \to 0$, the entropy of a perfect crystal goes to 0.
- As temperature $T \to 0$, the heat capacity $C(T) \to 0$.
- It is impossible to cool a system to $T = 0$ in any finite number of steps.

Let's go through each of these concepts one by one. The idea with the first statement is that a perfect crystal should always be at the ground state (in quantum mechanics) when all thermal energy is taken away ($T = 0$), so there should only be $\Gamma = 1$ microstate! Thus, $S = k_b \ln \Gamma = 0$. However, systems do have imperfections: for example, some atoms in a piece of glass may not be in this perfect crystal shape. So it will take a very long time for the system to "relax:" this is a kind of residual entropy, and there are interesting ways to deal with this.

Thinking about the second statement, let's write entropy in terms of the heat capacity.

Lemma 170

Entropy for a given pressure and temperature is

$$S(T, P) = S_0 + \int_0^T \frac{dT}{T} C_P(T).$$

where S_0 is some "residual entropy" mentioned above.

Proof. We start from the first law and formula for enthalpy:

$$dU = TdS - PdV \implies dH = TdS + VdP.$$

At constant pressure,

$$\left.\frac{\partial H}{\partial T}\right|_P = T \left.\frac{\partial S}{\partial T}\right|_P = C_P(T),$$

where the third term disappears due to $dP = 0$, and now integrating both sides of $\left.\frac{\partial S}{\partial T}\right|_P = \frac{C_P(T)}{T}$ yields the result above. $\square$

But now if we take $T \to 0$, the integral will diverge unless $C_P(T)$ goes to zero! So we must have $C_P(T) \to 0$.

Finally, let's think about the third statement: how can we think about cooling a system in this way?

29.3 Engines and efficiency

Definition 171

A **heat engine** is a machine that executes a closed path in the (phase) space of thermodynamic states by absorbing heat and doing work.

Since internal energy is a state function, a closed loop will not change the internal energy of our system! We can think of this alternatively as assigning a U to each point in our phase space. This can be written as

$$0 = \oint dU = \oint TdS - \oint PdV.$$

Note that $\oint PdV > 0 \implies$ our engine does work, and $\oint TdS > 0 \implies$ heat is added to the system. If we draw a clockwise loop in PV-space, the engine will do work on the surroundings, and we can make an analogous statement about TS-space.

So our goal is to construct such a closed loop and find the efficiency: how much work do we get for the amount of heat we are putting in?

Definition 172

A **Carnot engine** is a special engine where all heat is absorbed at a fixed temperature T^+ and expelled at a fixed T^-.

This is an ideal engine: in other words, we are assuming that this is a reversible process and that all heat exchange only takes place between a source T^+ and sink T^-. We can think of this as having our system between a source and sink, absorbing some heat Q^+, expelling some heat Q^-, and doing some work W.

By the first law, since $\Delta U = 0$,

$$Q^+ = Q^- + W.$$

If we have a reversible transfer of energy between the system and the source, then the change in entropy

$$S^+ = \frac{Q^+}{T^+}.$$

(If our system is not ideal, we may have additional entropy gain, since we lose additional knowledge about our system. So in general, we'd have $S^+ \ge \frac{Q^+}{T^+}$.) Similarly, we also have a reversible transfer of energy between the source and sink, so the entropy expelled to the environment is

$$\frac{S^-}{=}\frac{Q^-}{T^-}.$$

(In an irreversible process, some of the entropy remains as residual entropy inside our engine, so generally we have $S^- \le \frac{Q^-}{T^-}$.

Definition 173

Define the **efficiency** of an engine to be

$$\eta = \frac{W}{Q^+} = \frac{Q^+ - Q^-}{Q^+} = 1 - \frac{Q^-}{Q^+}$$

to be the ratio of work we can do with the engine to the heat that we put in.

We know that in our cycle, the entropy in and out of our system should also be a state function. So $\Delta S = 0$, and

$$S^+ = S^- \implies \frac{Q^+}{T^+} \le \frac{Q^-}{T^-} \implies \frac{Q^-}{Q^+} \ge \frac{T^-}{T^+}.$$

Plugging this in to our definition, we have the following:

Proposition 174 (Carnot's bound on efficiency of a heat engine)

The maximum efficiency of a Carnot engine between a source of temperature T^+ and sink of temperature T^- is

$$\eta = 1 - \frac{T^-}{T^+}.$$

Example 175

The maximum efficiency of an engine between 327° Celsius and 27° Celsius is $\frac{600-300}{300} = \frac{1}{2}$ (convert to Kelvin).

This Carnot efficiency is an absolute thermodynamic bound, and it's saturated when all heat is added at some temperature T^+ and expelled at some temperature T^-.

If we plot this process in the TS plane, we trace out a rectangle and have to do the following steps:

- Isothermally compress the gas at temperature T^-: this requires expelling Q^- heat and doing an equal amount of work.
- Put in work to adiabatically compress our gas from T^- to T^+.
- Isothermally expand the gas at temperature T^+: this requires putting in Q^+ heat and expelling that amount of work.
- Let the gas adiabatically expand back to temperature T^- (and get work out of it).

In the PV plane, the first and third steps (for an ideal gas) follow the isotherms $PV = c$, and the second and fourth steps follow $PV^\gamma = c$. The total work done here is the area enclosed by our curve in the PV plane! For an ideal gas, we have the following: If we say that our steps lead us between states $1, 2, 3, 4$, then we can make the following table:

Step	Process	Work by	Heat
$1 \to 2$	isothermal compression	$-nRT \ln \frac{V_1}{V_2}$	$-nRT^- \ln \frac{V_1}{V_2}$
$2 \to 3$	adiabatic compression	not needed	0
$3 \to 4$	isothermal expansion	$nRT^+ \ln \frac{V_4}{V_3}$	$nRT^+ \ln \frac{V_4}{V_3}$
$4 \to 1$	adiabatic expansion	not needed	0

where the "not needed" steps will cancel out.

So the total heat added is

$$Q^+ = nRT^+ \ln \frac{V_4}{V_3},$$

while the work done is

$$Q^+ - Q^- = nRT^+ \ln \frac{V_4}{V_3} - nRT^- \ln \frac{V_1}{V_2}.$$

We can show that $\frac{V_4}{V_3} = \frac{V_1}{V_2}$ by using the fact that $PV^\gamma = c$ for adiabatic compression and expansion, and therefore we indeed have indeed

$$\eta = \frac{W}{Q^+} = \frac{T^+ - T^-}{T^+}$$

saturates the Carnot efficiency!

Does our system change when we deal with small numbers and more fluctuation? The question of "whether we can beat the Carnot efficiency" is still a good question of research today at small scales!

Next session, we will introduce an interesting engine called Stirling's engine, whose work per cycle is better than the Carnot theoretical efficiency!

30 April 8, 2019 (Recitation)

This recitation was taught by Professor England.

We often consider a situation where a piston is pushing on a gas. One assumption that is usually made is that we have a "quasi-equilibrium" state: the system is compressed slowly, so that we're always close to an equilibrium state (for example on our PV diagram).

In this scenario, we do actually use exact differentials in a way that can be measured. Then if we apply some force F over an infinitesimal distance dx, we have our first law

$$dU = \text{đ}W + \text{đ}Q = -PdV + TdS.$$

But in real life, we usually push our piston down quickly, and the compression is done rapidly. Now our question is: what's the amount of work we do if our gas isn't in equilibrium? (Basically, our system is not given enough time to relax.)

Example 176

Let's say our gas is placed in a bath at temperature T.

When we compress an ideal gas very slowly, $dU = 0$ (since the temperature stays constant). This means

$$\text{đ}Q = \text{đ}W \implies PdV = TdS.$$

On the other hand, what's the change in entropy for this process? The entropy change over the universe is

$$\Delta S_{\text{universe}} = \Delta S_{\text{system}} + \Delta S_{\text{bath}}.$$

In a quasi-static process, $\Delta S_{\text{universe}} = 0$: we have a reversible process, because the change of entropy at a constant temperature is

$$\Delta S = \frac{\text{đ}Q}{T};$$

since all heat transferred out of the system is transferred into the bath at the same rate! So the total change in a quasi-static process is

$$\Delta S = \frac{\text{đ}Q}{T} - \frac{\text{đ}Q}{T} = 0.$$

Fact 177

Also, we can think about the fact that the entropy is proportional to $-N \ln \rho$, where ρ is the density of our ideal gas.

Indeed, this process is reversible: if we push the piston back out slowly enough, the heat will flow back into the system, and we'll again have equal and opposite changes in entropy.

But now let's say we push the piston the same amount, but we do it fast enough that heat doesn't transfer as smoothly: for example, we can imagine doing an adiabatic process, and then (once that's finished) let the heat transfer with the bath happen. This is significantly different: in an isothermal process, PV is constant, while in an adiabatic constant, PV^γ is constant: these yield different amounts of work for the same compression ΔV, because (mathematically) the integral of PdV is different or (physically) adiabatic compression means the particles will push back more, so the work we have to do is larger! So

$$W > \int_{\text{isothermal}} PdV;$$

and this argument works even when the process isn't adiabatic. Ultimately, we are putting in some extra energy, and that energy will eventually exit into the heat bath, since $U = \frac{3}{2}Nk_BT$ only depends on the temperature of our gas. Ultimately, then we know that the change in entropy of our system (ideal gas) doesn't depend on our final state: it will be

$$\Delta S_{\text{system}} = -k_B \Delta \left(N \ln \rho \right),$$

which is the same as what we had before in the slow case. But the difference comes from the change ΔS in the environment: because the fast work we did is larger than the isothermal work, some heat is pushed into the heat bath! So $\frac{\Delta Q}{T}$, entropy change in the surrounding bath, is greater than the original $\int \frac{\text{đ}Q}{T}$ in the isothermal case, which means our process is now irreversible: $\Delta S > 0$.

Fact 178

So in summary, a fast process means we do work that is larger than the slow work we needed to do, which increases heat flow and therefore generates entropy. That's where the second law of thermodynamics is coming from:

$$\Delta S_{\text{fast}} = 0.$$

But it would also be nice to understand this in a microscopic scale as well. Let's say we have a bunch of particles in our ideal gas that are flying around in this case: there will be times where (by chance) there's some vacuum near the piston, so we can get lucky and compress the gas without doing any work (as fast as we want)! So we've then decreased the entropy of the gas without doing any work at all, and that seems to violate the second law of thermodynamics: how do we deal with this?

This gets us into a more current topic: let's say we have our system and we follow some path from the initial to final state (for example, we have some fast compression $h(t)$ which dictates the height of our piston). Now we can define a state function

$$F = U - TS,$$

the Helmholtz free energy. We know that $dU = -PdV + \mu dN + \cdots + TdS$, so taking differentials of the free energy (by the product rule), the TdS cancels out:

$$dF = -SdT + \text{đ}W$$

So at constant temperature, $dF = \text{đ}W$. That means that holding the temperature constant will give a total change in free energy equal to the work that we've done! So it's important to build here because a first law that holds for quasi-static processes is deeply connected to it.

To deal with the second law violation, we have to start thinking about the statistical properties of the microstates. A slow process always averages over all possibilities: a bunch of states are getting visited. But a really fast piston will visit the system in some microstate, which explains why our work is statistically varying. So we get some distribution $p(W)$, and this will lead us to a fact:

Proposition 179 (Jarzynski, 1997)

We have

$$\int dW p(W) e^{-W/(k_BT)} = e^{-\Delta F/(k_BT)}$$

We'll be able to prove this later on! What does this tell us? Inherently, the right hand side is a change in a state function, and the left hand side measures our statistical fluctuations: this is a limitation on the shape of our distribution. For example, $p(W) = \delta(W - \Delta F)$ in a quasi-static process, and indeed this checks out with our integral.

This can alternatively be written as a statement about expectations:

$$\left\langle e^{-W/(k_B T)} \right\rangle = e^{-\Delta F/(k_B T)}.$$

So now we can define a quantity

$$W_d = W - \Delta F$$

which tells us how much work is dissipated. A quasi-static process has $\Delta F = W$, but in general we can have some difference based on statistical variance. But dividing our statement above by $e^{-\Delta F/(k_B T)}$, we have

$$\left\langle e^{-(W-\Delta F)/(k_B T)} \right\rangle = \left\langle e^{-W_d/(k_B T)} \right\rangle = 1.$$

Now recall that the second law is about doing extra work that we didn't need to! Dividing our ΔF definition through by T,

$$\frac{\Delta F}{T} = \frac{\Delta U}{T} - \Delta S$$

we can think of the right side as $-\Delta S_{\text{env}} - \Delta S_{\text{sys}}$, which is $-\Delta S_{\text{total}}$. But by convexity (specifically, Jensen's inequality),

$$\langle e^x \rangle \geq e^{\langle x \rangle}$$

So plugging that in, $\langle W_d \rangle \geq 0$, so we always expect to do a nonnegative amount of dissipated work!

31 April 9, 2019

Some announcements about the exam: it will take place next Thursday, April 18. It'll cover material after quiz 1 - it won't have any problems on probability theory, but we will need to understand those concepts to do (for example) problems on a microcanonical ensemble!

Again, previous exams will be posted online, and 3 of the 4 problems will be from previous years' exams. There will be an optional review session as well.

Notice that there won't be any classes next week: Tuesday has no classes, and Thursday will be the exam.

31.1 A quick overview

Last time, we talked about extracting work from heat engines, and we found ways to combine different processes in the PV plane to make a cycle and extract work. We found a theoretical bound, called the Carnot efficiency, and we constructed an example that saturates that efficiency.

Fact 180

Note that it's not possible to actually achieve this efficiency in real life because of dissipation and other processes!

Now we're going to try to introduce some other cycles as well, such as the Stirling engine. We'll also discuss statistical physics for systems that are actually connected to the outside world using free energy!

31.2 A new heat engine

Recall that in a Carnot cycle, we trace a clockwise cycle in our PV plane bounded by curves of the form $PV = c$ and $PV^\gamma = C$. We derived that the efficiency here is the theoretical bound

$$\eta = \frac{W}{Q^+} = \frac{T^+ - T^-}{T^+}.$$

But if we replace our adiabatic processes with isometric properties, this allows us to do more work in a single cycle! This is called a **Stirling engine**. In the TS diagram, we now no longer trace out a rectangle, since we no longer have adiabatic steps that keep entropy constant. Instead, for a monoatomic ideal gas, the entropy change is

$$S = Nk_B \left(\frac{3}{2} \ln T + \cdots \right).$$

So in the TS-plane, our process is now bounded by horizontal lines (at constant temperature) and exponential curves ($T \propto \exp(\frac{2}{3}\frac{S}{Nk_B})$). So our question now is **how we construct such an engine?** Again, let's say our process takes us between states $1, 2, 3, 4$.

Step	Process	Work by engine	Heat added
$1 \to 2$	Isothermal compression	$-nRT^- \ln \frac{V_1}{V_2}$	$-nRT^- \ln \frac{V_1}{V_2}$
$2 \to 3$	Isometric heating	0	$C_V(T^+ - T^-)$
$3 \to 4$	Isothermal expansion	$nRT^+ \ln \frac{V_1}{V_2}$	$nRT^+ \ln \frac{V_1}{V_2}$
$4 \to 1$	Isometric cooling	0	$-C_V(T^+ - T^-)$

So the efficiency here is

$$\eta = \frac{W}{Q^+} = \frac{nR(T^+ - T^- \ln(V_1/V_2)}{nRT^+ \ln(V_1/V_2) + C_V(T^+ - T^-)}.$$

So the idea here is that the $C_V(T^+ - T^-)$ is what limits us from reaching the Carnot efficiency - how do we get around this?

Proposition 181

Consider a chamber with two pistons: the left piston is for "expansion" and the right piston is for "compression." Keep the portion of the system to the left of the chamber at temperature T^+, and keep the portion of the system to the right at temperature T^-. In between, we have an ideal gas (equivalently a fluid), but put a material with high heat capacity in the middle as well (called the **regenerator**).

So that amount of heat can be reused: it is stored and then reused in the next cycle! So now this effectively eliminates the $C_V(T^+ - T^-)$ term, and we do indeed reach the Carnot efficiency $\frac{T^+ - T^-}{T^+}$ as desired.

How much work do we get for each cycle? We can find that

$$\frac{W_{\text{Stirling}}}{W_{\text{Carnot}}} = \frac{1}{1 - \frac{\ln(T^+/T^-)}{(\gamma - 1)\ln r}}$$

where r is the ratio of volumes. Plugging in $\frac{T^+}{T^-} = 2, r = 10, \gamma = 1.4$, we find that a Stirling engine gives 4 times as much work as a Carnot engine for the same ratio of volumes and temperatures!

31.3 Brief aside about the third law

Remember that one of our statements of the third law was that **it is impossible to cool any system to absolute zero in a finite amount of time**. One example of this is by switching between two different pressures: low to high with an adiabatic process and high to low with an isothermal process. This can be done using nuclear magnetization!

But both P curves end at $(0, 0)$: that's the main point of the third law. If we try to do any process, following some curve in the TS-plane, we will never actually be able to get to the origin. However, we've gotten pretty good at getting close to absolute zero: ask professor Ketterle about this!

31.4 Free energy

So as a basic summary of what we've been talking about: we introduced S, our entropy, as a function of the number of microstates. We've made some statements about general properties of S, but we're going to start looking at smaller parts of the system (and treat the rest as being a heat bath with constant temperature). So now our system exchanges energy, volume, and particles with an outside world in thermal equilibrium, so we have our temperature fixed (rather than the total internal energy).

There's many different notions of free energy: they will be used in different constraints. Remember that in an isolated system, we need $\Delta S \geq 0$ to have a spontaneous change of an isolated system (though this isn't necessary sufficient). Let's now imagine that our system is in contact with a heat bath at a fixed T: now the total entropy change is (at constant temperature)

$$\Delta S_{\text{total}} = \Delta S_{\text{system}} + \Delta S_{\text{bath}}.$$

If our system does no work (because it is at a fixed volume),

$$\Delta S_{\text{bath}} = \frac{\Delta Q}{T} = -\frac{\Delta U_{\text{system}}}{T},$$

and we can plug this in to find

$$\Delta S_{\text{total}} = \Delta S_{\text{system}} - \frac{\Delta U_{\text{system}}}{T}$$

This means that we have a new necessary condition:

Fact 182

We must have

$$\Delta S_{\text{sys}} - \frac{\Delta U_{\text{system}}}{T} \geq 0$$

for a spontaneous change of our system in contact with a heat bath.

This motivates us to make a new definition:

Definition 183

Define the **Helmholtz free energy** F of a system to be the state function $F = U - TS$.

We now require $\Delta F \leq 0$ to make a spontaneous change in a system at thermal equilibrium with fixed V and T. It seems that because S is a function of U, V, N in our system, F should also be a function of the four variables T, V, N, U.

But the free energy is not actually a function of the internal energy! Indeed,

$$F = U - T_b S_{\text{system}}(U_S) \implies \frac{\partial F}{\partial U_S} = 1 - T_b \frac{\partial S_S}{\partial U_S} = 1 - \frac{T_b}{T_s},$$

and at thermal equilibrium (which is the only situation where we use F to represent the system), $T_b = T_s$. So the derivative is 0, so U_S is not an independent variable here!

Fact 184

This is an example of a **Legendre transformation**, which helps us change from one type of energy or free energy to another by using a different set of independent variables. For example, we go from (U, V, N) to (T, V, N).

As another example, we can change our variables in another way: considering a system at constant temperature and pressure (this happens a lot in biology and chemistry). Then recall that

$$dU = dQ - PdV \implies dH = dQ + VdP \implies \Delta Q = \Delta H$$

at constant pressure, and now we can play the same trick! The heat emitted into the environment at constant pressure is $-\Delta H$, so the total change in entropy

$$\Delta S_{\text{total}} = \Delta S_{\text{system}} + \Delta S_{\text{bath}} = \Delta S_{\text{system}} - \frac{\Delta H}{T},$$

and by the second law, we must have this quantity be at least zero.

Definition 185

This motivates the definition of the **Gibbs free energy**

$$\Delta G = H - TS$$

and for a spontaneous change in our system, we must have $\Delta G \leq 0$: free energy needs to decrease!

It's important to note that this wasn't all just to introduce a new state function: here's a preview of what's coming. In some systems, we have a fixed energy and can calculate the temperature of that system: in a microcanonical ensemble, we consider all possible microstates that are consistent with it. Then we can take some subset of our system, where the temperature T is still fixed (though the energy is not), and that's what a canonical ensemble deals with!

31.5 Natural variables and Maxwell relations

The idea we've been developing in this class so far is to go from various constraints on our system to some mathematical relationship between thermodynamic variables. For example, the first law tells us that if we define our internal energy U in terms of S and V, then

$$dU = TdS - PdV \implies \left.\frac{\partial U}{\partial S}\right|_V = T, \left.\frac{\partial U}{\partial V}\right|_S = -P.$$

Because we have a state function U, the mixed second partial derivatives should be the same: this means

$$\frac{\partial}{\partial V}\left(\left.\frac{\partial U}{\partial S}\right|_V\right)_S = \frac{\partial}{\partial S}\left(\left.\frac{\partial U}{\partial V}\right|_S\right)_V$$

and that gives us the relation

$$\left.\frac{\partial T}{\partial V}\right|_S = -\left.\frac{\partial P}{\partial S}\right|_V.$$

This applies to other state functions we have too - we can apply the same logic with our free energy definitions as well! The idea is that free energies are often easier to measure experimentally under certain conditions. For example, since $H = U + PV$, $dH = TdS + VdP$, so writing H as a function of S and P,

$$\left.\frac{\partial H}{\partial S}\right|_P = T, \left.\frac{\partial H}{\partial P}\right|_S = V,$$

and doing the mixed partial derivatives yields

$$\left.\frac{\partial T}{\partial P}\right|_S = \left.\frac{\partial V}{\partial S}\right|_P.$$

Finally, since $F = U - TS, dF = dU - TdS - SdT$, which is also $dF = -PdV - SdT$. So this time, it's natural to write F as a function of V and T. Now

$$\left.\frac{\partial F}{\partial T}\right|_V = -S, \left.\frac{\partial F}{\partial V}\right|_T = -P \implies \left.\frac{\partial S}{\partial V}\right|_T = \left.\frac{\partial P}{\partial T}\right|_V.$$

These partial derivative equations often relate a quantity that is easy to measure with something that is generally less experimental in nature! We'll talk more about these concepts next time.

32 April 10, 2019 (Recitation)

32.1 Exchanging energy

Let's start with the Helmholtz free energy

$$F = E - TS.$$

It's easy to say what energy or momentum "does," but what about free energy? One way to think about this is the "balance" between energy and entropy. Any system that interacts with a reservoir at temperature T (for example, room temperature in the world around us) cares about "free energy" to consider favorability of a reaction.

The main idea is basically that the second law must hold! If we create an entropy ΔS, then we can gain an energy of $E = T\Delta S$ from the reservoir. Another way to think about this is connected to the problem set: if we have a system with (non-normalized) occupation numbers

$$n_a \propto e^{-\beta E_a},$$

notice that all Boltzmann factors are equal if we have infinite temperature, and that means all states are equally populated! On the other hand, with zero temperature, only the ground state is occupied, because any higher state has exponentially smaller occupation numbers.

32.2 Looking at the Sackur-Tetrode entropy

Recall the equation that was derived by these two researchers from Germany back in 1913:

$$S(U, T, N) = k_B N \left(\ln \frac{U}{N} + \ln \frac{V}{N} + \frac{3}{2} \ln \frac{4\pi m}{3h^2} + \frac{5}{2} \right)$$

Note that the constants after the first two terms are just prefactors. Interestingly, though, Sackur got $\frac{3}{2}$ instead of $\frac{5}{2}$: this comes from $\ln n! = n(\ln n - 1)$, and he didn't include the -1 in his approximation.

But does this really matter? In almost all calculations, we ignore the entropy and only consider ΔS. But what would we actually get wrong?

Notice that in the correct form of Sackur-Tetrode entropy, the entropy S goes to 0 as the temperature T goes to

0. It turns out we can measure absolute entropy via

$$S(T) = \int_0^T \frac{dQ}{T} = \int_0^T \frac{c_P(T)dT}{T}.$$

(You also have to add latent heat to go from solid to liquid state, and so on. This can be written as a delta function, but the details aren't that important.) Well, some researchers followed mercury across different temperatures: knowing C_P, they integrated from 0, and eventually mercury became an ideal gas at high enough temperature! So through measurements, they found agreement with the theoretical answer. (They did assume $\frac{5}{2}$.)

Fact 186

This allowed them to find an experimental value of h, Planck's constant, and they did this to 1 percent precision! It's pretty amazing that we can determine this by measuring heat - this is really adding "entropy."

But if we replace $\frac{5}{2}$ with $\frac{3}{2}$ and try to get that experimental result for h, we actually change it by a factor of $e^{-1/3} \approx 0.72$. So that would give you an incorrect experimental result by 30 percent! This is why absolute entropy does occasionally matter.

32.3 Multiple ground states

In quantum mechanics, we have multiple energy levels: it's possible that (for example at zero magnetic field) we have two low-energy states. Then we have a special symmetry in our system!

For example, if we have N distinguishable particles and each can be in 2 degenerate states, we have an extra contribution S_0 to the entropy. This was actually mentioned in Sackur-Tetrode: mercury has multiple isotopes, so we have to be careful. There's other ways quasi-degeneracy could come up as well.

32.4 Fluctuations in a microcanonical ensemble

Let's try to think about a system with two subsystems, but instead of exchanging energy, let's think about exchanging particles. If the system is divided into two symmetric parts A and B, and we have $N_A + N_B = N$ particles, we expect there to generally be $\frac{N}{2}$ particles in both halves.

The total number of microstates Γ_{AB} is then multiplicative, because we essentially pick a microstate from both A and B. But then taking logarithms, the entropy S_{AB} is now additive!

But let's think a bit more about the number fluctuation. Is it true that the number of microstates for the whole system, is

$$\Gamma_{N,A+B} = \Gamma_{\frac{N}{2},A} \cdot \Gamma_{\frac{N}{2},B}?$$

Not quite! Some arrangements of the system $A + B$ don't actually have the number of particles or the energy split exactly: there are fluctuations! So we basically have to add over all possible number of particles in system A:

$$\Gamma_{N,A+B} = \sum_k \Gamma_{k,A}\Gamma_{N-k,B}.$$

But in principle, the extreme values of k are not likely - they contribute very little to the entropy. Specifically, the distribution of probabilities N_A is centered around $\frac{N}{2}$, and now we have a sharp distribution that's basically Gaussian with standard deviation $\propto \sqrt{N}$.

That means that allowing fluctuations, $\Gamma_{N,A+B} = \Gamma_{\frac{N}{2},A} \cdot \Gamma_{\frac{N}{2},B}$ has an approximate "width" of $\sqrt{N}$: this gives us an approximate answer of

$$\Gamma_{N,A+B} = \sqrt{N}\Gamma_{\frac{N}{2},A} \cdot \Gamma_{\frac{N}{2},B}.$$

But this kind of fluctuation is very small: if we have, for example, 10^{26} particles, we have a precision of 10^{-13}: this is unmeasurably small! So we wouldn't be able to just count them. Now note that our counting of microstates is not quite multiplicative with this approximation: we have an entropy

$$S \propto S_{\frac{N}{2},A} + S_{\frac{N}{2},B} + c \ln N.$$

So we should always be careful! The first two terms are extensive quantities, while the last term is not. Luckily, that is almost negligible compared to the other terms.

33 April 11, 2019

33.1 Overview and review

This class is about studying the connection between microscopic and macroscopic descriptions of a system. What we've been doing recently is imposing various constraints on our system: for example, setting a fixed energy U gives us a microcanonical ensemble, and we'll find that setting a fixed temperature T will give us what is called a canonical ensemble.

Last time, we started discussing more relationships between our thermodynamic variables. If we have a first law condition

$$dU = TdS - PdV,$$

then we can define our energy U in terms of S and V: then because U is a state function, the mixed partial derivatives with respect to S and V gives us

$$\left.\frac{\partial U}{\partial S}\right|_V = T, \left.\frac{\partial U}{\partial V}\right|_S = -P \implies \left.\frac{\partial T}{\partial V}\right|_S = -\left.\frac{\partial P}{\partial S}\right|_V.$$

With the same kind of argument, we can also derive values from the enthalpy $H = U + PV$: this yields

$$\left.\frac{\partial H}{\partial S}\right|_P = T, \left.\frac{\partial H}{\partial P}\right|_S = V \implies \left.\frac{\partial T}{\partial P}\right|_S = \left.\frac{\partial V}{\partial S}\right|_P.$$

With the Helmholtz free energy $F = U - TS$,

$$\left.\frac{\partial F}{\partial T}\right|_V = -S, \left.\frac{\partial F}{\partial V}\right|_T = -P \implies \left.\frac{\partial S}{\partial V}\right|_T = \left.\frac{\partial P}{\partial T}\right|_V,$$

and finally with the Gibbs free energy $G = F + PV$,

$$\left.\frac{\partial G}{\partial T}\right|_P = -S, \left.\frac{\partial G}{\partial P}\right|_T = V \implies \left.\frac{\partial S}{\partial P}\right|_T = -\left.\frac{\partial V}{\partial T}\right|_P.$$

These four relations between our state variables are known as **Maxwell's equations**: we'll soon see why they're important. For example, if we want to see change in entropy per volume at constant temperature, there's really no way to measure that quantity directly, since it's really hard to measure entropy. On the other hand, we can measure $\left.\frac{\partial P}{\partial T}\right|_V$ a lot more easily: just change our temperature and measure the pressure inside some fixed volume!

Fact 187

Don't memorize these equations: we can always derive these from the first law directly.

More generally, we can make a table between intensive variables X and extensive conjugate variables Y:

X	Y
$-P$	V
σ	A
H	M
F	L
E	P
μ	N

and in general, we can always write down our internal energy U as a function of S and Y, H as a function of S and X, F as a function of T and Y, and G as a function of T and X! We've mostly only been focusing on a three-dimensional gas, which is why we've been using P and V, but we could replace this with other pairs of variables.

33.2 An application of Maxwell's relations

Let's try to compute heat capacity of an arbitrary material: in general, the formula

$$C_V = \left.\frac{\partial U}{\partial T}\right|_V$$

measures how much internal energy of a system depends on its temperature. For an ideal gas, we know that internal energy only depends on temperature, but we may want to measure other heat capacities as well.

Example 188

What is a good way to find

$$\left.\frac{\partial U}{\partial V}\right|_T?$$

Let's do this systematically so that it's easy to understand how to do related problems! Start with the first law,

$$dU = TdS - PdV.$$

This means that $\left.\frac{\partial U}{\partial V}\right|_S = -P$. Let's write U as a function of T and V, so

$$dU = \left.\frac{\partial U}{\partial T}\right|_V dT + \left.\frac{\partial U}{\partial V}\right|_T dV.$$

This includes the term we want! Going back to the first law, writing S as a function of T and V as well,

$$\left.\frac{\partial U}{\partial T}\right|_V dT + \left.\frac{\partial U}{\partial V}\right|_T dV = dU = TdS - PdV = T\left(\left.\frac{\partial S}{\partial T}\right|_V dT + \left.\frac{\partial S}{\partial V}\right|_T dV\right) - PdV$$

Now all differentials are in terms of dV and dT, so we can gather terms of dT and dV and divide to find

$$\left.\frac{\partial U}{\partial V}\right|_T = T\left.\frac{\partial S}{\partial V}\right|_T - P.$$

If we want to make this a measurable quantity, though, we should replace the $\frac{\partial S}{\partial V}\big|_T$ term: this can be done with Maxwell's relations! Since $\frac{\partial S}{\partial V}\big|_T = \frac{\partial P}{\partial T}\big|_V$, we just have

$$\boxed{\left.\frac{\partial U}{\partial V}\right|_T = T\left.\frac{\partial P}{\partial T}\right|_V - P}.$$

So if we just have the equation of state, we have enough to find the quantity we desired!

Example 189

What is the V-dependence of C_V; that is, what is

$$\left.\frac{\partial C_V}{\partial V}\right|_T ?$$

Recall that C_V is itself a partial derivative! Specifically,

$$\frac{\partial}{\partial T}\left(\left.\frac{\partial U}{\partial V}\right|_T = \frac{\partial}{\partial V}\left(\left.\frac{\partial U}{\partial T}\right|_T = \left.\frac{\partial C_V}{\partial V}\right|_T .$$

Since we just found the leftmost quantity, we can plug that in:

$$\left.\frac{\partial C_V}{\partial V}\right|_T = \frac{\partial}{\partial T}\left(T\left.\frac{\partial P}{\partial T}\right|_V - P\right) = \boxed{T\left.\frac{\partial^2 P}{\partial T^2}\right|_V}.$$

Example 190

Can we find a relationship in general between C_P and C_V?

Remember that we defined a quantity $H = U + PV$ to help with this kind of problem:

$$\left.\frac{\partial H}{\partial T}\right|_P = C_P = \left.\frac{\partial U}{\partial T}\right|_P + P\left.\frac{\partial V}{\partial T}\right|_P .$$

(This came from expanding out the derivatives on both sides.) Expanding out the first term on the right hand side,

$$\left.\frac{\partial U}{\partial T}\right|_P = \left.\frac{\partial U}{\partial T}\right|_V + \left.\frac{\partial U}{\partial V}\right|_T\left.\frac{\partial V}{\partial T}\right|_P$$

Note that $\frac{\partial U}{\partial T}\big|_V = C_V$, and we also computed earlier that $\frac{\partial U}{\partial V}\big|_T = T\frac{\partial P}{\partial T}\big|_V - P$, so plugging everything in,

$$C_P = C_V + \left(\left.\frac{\partial U}{\partial V}\right|_T + P\right)\left.\frac{\partial V}{\partial T}\right|_P \implies \boxed{C_P = C_V + T\left.\frac{\partial P}{\partial T}\right|_V\left.\frac{\partial V}{\partial T}\right|_P}.$$

We defined response functions earlier in the class: we had variables like expansivity that are known for certain materials! So if we're doing a problem in that realm, we can rewrite our equation in terms of coefficients like thermal expansion:

$$\alpha = \frac{1}{V}\left.\frac{\partial V}{\partial T}\right|_P \implies C_P = C_V + TV\alpha\left.\frac{\partial P}{\partial T}\right|_V .$$

So now can we say anything about our function $\frac{\partial P}{\partial T}\big|_V$? Note that we have the partial derivative identity

$$\left.\frac{\partial P}{\partial T}\right|_V\left.\frac{\partial T}{\partial V}\right|_P\left.\frac{\partial V}{\partial P}\right|_T = -1 \implies \left.\frac{\partial P}{\partial T}\right|_V = -\frac{1}{\frac{\partial T}{\partial V}\big|_P\frac{\partial V}{\partial P}\big|_T} = \frac{-V\alpha}{\frac{\partial V}{\partial P}\big|_T}$$

where $\beta = -\frac{1}{V} \frac{\partial V}{\partial P}\big|_T$ is the coefficient of isothermal compressibility. This lets us just write down $\frac{\partial P}{\partial T}\big|_V = \frac{\alpha}{\beta}$: plugging this in, we get the relation

$$\boxed{C_P = C_V + TV\frac{\alpha^2}{\beta}}.$$

This is basically a game of thinking like an experimental physicist: we know what's easy to measure, so we just try to write our derivatives in terms of those quantities!

33.3 The Joule-Thomson effect

This phenomenon is also known as the throttling process!

Example 191

What is the temperature change of a real gas when it is forced through a valve, if the container is isolated from the environment (so no heat exchange)?

In a real gas, there is actually interaction between the particles, so the temperature will actually change. Specifically, we then have a potential energy associated with our system as well!

To describe our system, imagine having two containers A and B connected with a small valve in the middle: let's say there is a pressure P_0 on the left container A and a pressure P_1 in the right container 2. We force some volume V_0 through the valve from P_0 to P_1: then the work done by the piston for container A is

$$W_A = P_0V_0,$$

and the work done by the piston for container B is

$$W_B = -P_1V_1.$$

where V_1 is the volume the gas takes up in the new container. Since there is no heat exchange, by the first law,

$$U_1 - U_0 = W_A + W_B = P_0V_0 - P_1V_1 \implies U_0 + P_0V_0 = U_1 + P_1V_1.$$

This means that this is a **constant enthalpy process**! So does the gas that is pushed through get cooler or warmer? We can define the **Joule-Thomson coefficient**

$$\mu_{JT} = \frac{\partial T}{\partial P}\bigg|_H.$$

Since we have expansion, $\partial P < 0$. This means that $\mu_{JT} < 0 \implies dT > 0$, so the gas warms up, and $\mu_{JT} > 0 \implies dT < 0$, so the gas cools. This actually has to do with liquifying gases at room temperature!

So how can we say things about μ_{JT}? Starting from the enthalpy equation, if H is written as a function of T and P, since $dH = 0$,

$$dH = \frac{\partial H}{\partial T}\bigg|_P dT + \frac{\partial H}{\partial P}\bigg|_T dP \implies \mu_{JT} = -\frac{\frac{\partial H}{\partial P}\big|_T}{\frac{\partial H}{\partial T}\big|_P} = -\frac{1}{C_P}\frac{\partial H}{\partial P}\bigg|_T.$$

This $\frac{\partial H}{\partial P}\big|_T$ term is analogous to the $\frac{\partial U}{\partial V}\big|_T$ term from earlier! Specifically, we can go through the analogous derivations to find

$$\mu_{JT} = -\frac{1}{C_P}\left(V - T\frac{\partial V}{\partial T}\bigg|_P\right).$$

For an ideal gas, $PV = nRT \implies \frac{\partial V}{\partial T}\big|_P = \frac{V}{T}$, so $\mu_{JT} = 0$: we can't see any effect on temperature. But for a van der

Waals gas, we have constants a and b so that

$$(V - nb)\left(P + a\frac{n^2}{V^2}\right) = nRT;$$

if we expand out the left side, since a and b are generally very small, the $-nb\left(\frac{an^2}{V^2}\right)$ term is negligible, and thus

$$PV - Pnb + \frac{an^2}{V} = nRT;$$

taking differentials,

$$PdV - \frac{an^2}{V^2}dV = nRdT,$$

and substituting in for P and doing the relevant calculations,

$$\boxed{\mu_{JT} \approx \frac{n}{C_P}\left(\frac{2a}{RT} - b\right)}.$$

So the temperature where μ_{JT} changes sign is the **inversion temperature**

$$T_{\text{inversion}} = \frac{2a}{Rb}.$$

If our temperature is larger than the inversion temperature, μ_{JH} is negative, so changing the pressure will increase our temperature. Otherwise, if T is smaller, then temperature will decrease!

34 April 17, 2019 (Recitation)

There is a quiz tomorrow!

34.1 Types of free energy

Last time, we mentioned the Helmholtz free energy and Gibbs free energy

$$F = U - TS$$

$$G = U + PV - TS,$$

where both of these are a kind of balance between energy and entropy. We can think of this by looking at the Boltzmann distribution as $T \to 0$ and $T \to \infty$.

Question 192. *What does free energy mean?*

Let's say we have our system S connected to a reservoir R fixed at temperature T. If we want to see if a certain reaction can happen, we must have the total $\Delta S_{\text{total}} > 0$ to have a favorable reaction (by the second law). Information-theoretically, we can go from more precise knowledge to less precise knowledge.

What is the change in entropy here? The change in internal energy of the reservoir is $\Delta U = \Delta Q$ (since no work is being done), which is $-T\Delta S_{\text{sys}}$ in a reversible process. This means that the change in entropy

$$\Delta S_{\text{total}} = \Delta S_{\text{sys}} + \Delta S_{\text{res}} = \Delta S_{\text{sys}} - \frac{\Delta U_{\text{sys}}}{T} = -\frac{1}{T}(\Delta U - T\Delta S_{\text{sys}}) \geq 0.$$

Defining $F = U - TS$, any favorable reaction has to go to lower free energy!

Fact 193

Think of a reservoir as a swimming pool and our system as a piece of cold chalk being thrown into it. The change in temperature of the swimming pool can be assumed to be almost zero, and if not in a specific case, we can always scale the system. This is the limit we take in a canonical ensemble!

On the other hand, when we have a system at constant pressure, we often use the Gibbs free energy $G = F + PV$, much like how we use enthalpy for some calculations instead of internal energy.

But why is it called free energy? We can think of this as "energy that can be converted." Looking again at a system with T and P constant (so we want to look at G), let's say that we are part of the system, and we want to transform some internal energy U into something useful. For example, can we turn the energy gained from

$$2H_2 + O_2 \to 2H_2O$$

into something useful by the laws of thermodynamics? Is the energy actually available?

Well, let's say 1 mole of a large molecule dissociates into 5 moles of smaller constituents: then there are more molecules, so to conserve the pressure P, we must increase the volume V, which does work

$$\Delta U = -P\Delta V.$$

This works in the water reaction as well! Since we now have less molecules, we have to reduce our volume, which does work on the system. This is where the PV term comes from in our equation, and we can finally account for the $+T\Delta S$ from the change in entropy in our system (since that needs to be transferred to our reservoir).

So we change the internal energy by some ΔU, and this now gives us an associated work $P\Delta V$ and heat transfer $-T\Delta S$. This is indeed the $G = U + PV - TS$ that we desire!

Fact 194

In summary, we do "boundary condition" corrections for the variables we are fixing constant: in this case P and T.

34.2 Ice packs

Some of these work without needing a frozen material: break something in a plastic pack, and it gets cold. How does this happen?

It turns out that it requires energy $\Delta U > 0$ to dissolve ammonium nitrate in water. So if this is in contact with something warmer (like an arm), it will grab energy and cool down the environment. Now think of the system as the plastic pack and our body as the environment! This is allowed because it creates some entropy $\Delta S > 0$. So now

$$F = U - TS < 0$$

is indeed true because the temperature T is sufficiently large (so the entropy overcompensates for the change in internal energy).

But what happens if the ice pack were not in contact with anything? Then the water and ammonium nitrate would be completely isolated, and now this cannot dissolve!

34.3 Heat engines

We have a theoretical bound on the efficiency of a heat engine

$$\eta_C = \frac{T^+ - T^-}{T^+}.$$

Let's say someone managed to create some $\tilde{\eta} > \eta_C$: what would this look like?

We know that a reservoir T^+ provides heat Q^+, and the sink T^- gains heat $Q^+ - W$. Well, let's imagine doing this in reverse: pull heat $Q^+ - W$ from the T^- sink and output heat Q^+ to the T^+ reservoir. (For example, this is a refrigerator!)

We can scale this in such a way that the Q^+ here is the same as the Q^+ in the theoretical other engine! Then putting the two engines together, the work we've created is $\tilde{W} - W$, and the heat we've extracted is $-(\tilde{W} - W)$: we've removed the upper reservoir from the problem when we run the Carnot engine in reverse.

But then we'd create work by extracting energy from a single temperature reservoir: this doesn't make sense thermodynamically! We can't get work and do a refrigeration at the same time, or at least it hasn't been observed.

Thus, the Carnot efficiency is actually the best possible bound on our heat engine! What's more, **any reversible engine has to work at exactly the Carnot efficiency**, or else we could run it in reverse and combine it with the Carnot cycle forward, and we'd get the exact same contradiction.

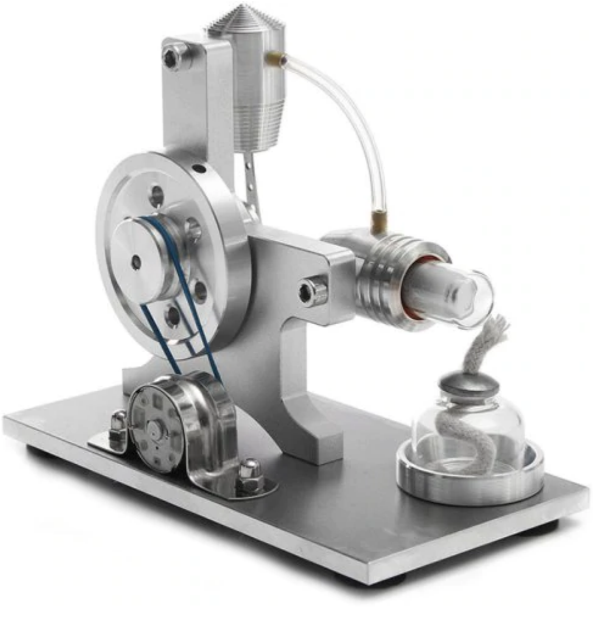
35 April 22, 2019 (Recitation)

35.1 Engines

Since we've been talking about engines, Professor Ketterle wanted to show us a Stirling engine!

Most engines we have in real life have some combustible material, so we have an open system with fuel and some exhaust material (and this turns out to be more efficient in general). In principle, though, we should be able to operate an engine by just having one heat source (like in our Carnot cycle).

The example at the front of class is just an alcohol burner, which acts as a hot reservoir, together with a cold reservoir at room temperature. Then we can run a motor in reverse, which is an electrical generator!



How does a heat engine work, generally? In principle, we can just say that we have a gas that compresses and contracts with a piston. But in this case, when we heat up the gas in the hot reservoir, a piston system actually moves the gas to the cold reservoir, where it cools down! We have pistons 90 degrees out of phase, and that lets our motor run.

Recall that last time, we showed that if we ever had an engine with efficiency $\eta > \eta_{\text{rev}}$, we could couple it together with a Carnot engine (running one in reverse) to create work for free by extracting heat from a single reservoir. This would be very convenient, but it violates the second law of thermodynamics.

Fact 195

This also told us that all reversible engines between reservoirs at T_1 and T_2 must have the same η efficiency. This is the maximum efficiency! The Carnot engine is just special because it's an example of an easily described engine with that efficiency

$$\eta = \frac{T_2 - T_1}{T_2}.$$

35.2 Free energy

Recall that $F = U - TS$, the free energy, tells us something about whether a system's entropy is going up or down:

$$\Delta F < 0 \implies \Delta S_{\text{total}} = \Delta S_{\text{sys}} + \Delta S_{\text{res}} > 0$$

at constant temperature. Basically, defining new "energies" like F, G, H includes information about the reservoir as well: for example, G describes our environment on earth (at constant pressure and constant temperature), which is why we use

$$G = U - TS + PV.$$

Basically, TS and PV tell us that we need to do some work or transfer some heat to satisfy the environment conditions! So F and G tell us what is "left," and that's why their sign tells us about whether processes can happen spontaneously.

Fact 196

We can think of "spontaneous" processes as those that can give us work, and that's how to reconcile thoughts like "spontaneous combustion."

35.3 Looking at free expansion of a gas again

Let's say we have a gas that expands from a volume V into a volume of $2V$, while keeping the internal energy U and temperature T the same. (For example, consider an ideal gas). Then $F = U - TS$ goes down, and free energy going down means that we can extract work out of this! That means we "missed our opportunity:" we've increased our entropy without getting the work out of it.

But we can also just keep that whole system at constant temperature T: if we're being formal about it, with a quasi-static process, we should actually have a piston that is isothermally expanding at temperature T. We'll then find that the work done does satisfy

$$|\Delta W| = |\Delta F|.$$

If the internal energy and temperature stay the same, where is that work coming from? Well, the gas is losing some kinetic energy when the piston is moving back! It is the heat added by the environment at constant T that keeps the internal energy constant, and that's basically coming from the TS term in the free energy.

So we create energy by reducing our internal energy, but the PV and TS terms are "taxed" by our reservoir and conditions. In our case here, we actually get a bonus from the environment to get work!

Proposition 197

Use F when we have constant temperature, and use G when we also additionally have a requirement that P is constant.

35.4 Partial derivatives

We have a bunch of thermodynamic variables:

$$P, V, S, T, H, U, F, G.$$

This is a system with only PV work. Well, that already gives us eight variables: we can now play in eight-dimensional space and write expressions of the form

$$\left.\frac{\partial a}{\partial b}\right|_c.$$

But how many of these eight variables are independent? For example, in 3-dimensional space, we can use r, θ, ϕ versus x, y, z and do lots of coordinate transformations, but no matter what, we have three independent variables. How many do we have in this situation.

Well, let's think about an ideal gas, where N is constant. Defining volume and temperature gives us pressure (by the ideal gas law), entropy (by the Sackur-Tetrode equation), and then we can get all types of energy. This means that we **always have two independent variables!** Often, we students will write things like

$$\left.\frac{\partial S}{\partial P}\right|_{V,T},$$

which doesn't actually make any sense, since fixing V and T tells us exactly what our other variables are! It's important to not get lost in the jungle of variables: we can only keep one constant at a time (of course, unless one of the two is N).

So Maxwell's relations come about because P, V, S, T can be written as first derivatives of the thermodynamic potentials H, U, F, G. Specifically, for a function X (which is one of U, H, F, G), we can think of

$$\frac{\partial^2 x}{\partial a \partial b} = \frac{\partial^2 x}{\partial b \partial a}.$$

For example, in problem 4 of last week's exam, we were told to compare

$$\frac{\partial V}{\partial T}, \frac{\partial S}{\partial P}.$$

The relevant differentials here are

$$dU = TdS - PdV, dG = -SdT + VdP.$$

This means U is written most naturally as a function of V and S, but G is written as a function of T and P. We can always convert between our variables, but this is the way we generally want to do it!

We are supposed to say what we keep constant in the two partial derivatives above, but we should ask the question: should we use the Maxwell's relation for U or for G?

We do have V and S in the numerators, and we have T and P in the denominators, so either would work at this point. But note that we were told to actually deal with

$$\left.\frac{\partial V}{\partial T}\right|_S, \left.\frac{\partial S}{\partial P}\right|_V.$$

So now looking at the V and S, it's most natural to work with U! But G gives the answer as well, and this is what Professor Ketterle did initially! Let's work out some of the steps:

$$dG = -SdT + VdP \implies -S = \left.\frac{\partial G}{\partial T}\right|_P, V = \left.\frac{\partial G}{\partial P}\right|_T.$$

So now

$$\frac{\partial^2 G}{\partial T \partial P} = -\left.\frac{\partial S}{\partial P}\right|_T = \left.\frac{\partial V}{\partial T}\right|_P.$$

This isn't actually what we want, though: we have the wrong variables held constant! There is a way to get around this though: just like we can transform C_P and C_V into each other, let's look at how to relate $\frac{\partial S}{\partial P}\big|_T$ to $\frac{\partial S}{\partial P}\big|_V$. We're going to have to label our variables carefully: writing S as a function of P and V (since this is what we want in the end), where V is actually a function of P and T. This means we're representing our entropy as

$$S(P,T) = S(P, V(P,T)) \implies \left.\frac{\partial S}{\partial P}\right|_T = \left.\frac{\partial S}{\partial P}\right|_V + \left.\frac{\partial S}{\partial V}\right|_P \left.\frac{\partial V}{\partial P}\right|_T$$

by the chain rule. So it's possible to work from there, but it is very messy! Alternatively, we could go back and do the smarter thing: looking at U as the thermodynamic potential instead, we end up with

$$dU = TdS - PdV \implies \left.\frac{\partial T}{\partial V}\right|_S = -\left.\frac{\partial P}{\partial S}\right|_V.$$

This is exactly the reciprocal of what we wanted: we end up with

$$\left.\frac{\partial V}{\partial T}\right|_S = -\left.\frac{\partial S}{\partial P}\right|_V.$$

36 April 23, 2019

All exam grades will be posted by the end of today. We can always email the professor to schedule a meeting if we want to discuss! Also, drop date is very soon.

There is no pset due this week, because we just started some new material. Instead, we can focus on other homework assignments we might have!

36.1 Overview

It's time to start a new section in our class - we're going to talk about a new kind of ensemble. We found that for a system that is well-defined in a certain way (mechanically and adiabatically isolated) can be described by a microcanonical ensemble, which relates to the probability distribution of the individual microstates. The key idea there was that the **energy** was held constant, and in these cases, all microstates are equally probable. The temperature T is then a natural consequence!

We often deal with systems that aren't adiabatically isolated, though. Then the microstates may have different probabilities: that's what we'll discuss today, and we'll use the **canonical ensemble** as a tool! We'll find that there really isn't much of a difference between the properties of a microcanonical and canonical ensemble in some aspects, but we don't have to compute the actual number of microstates anymore.

36.2 The canonical ensemble

Let's start by making a distinction:

Fact 198

In a microcanonical ensemble, we specify the internal energy U, and at thermal equilibrium we can deduce the temperature T. On the other hand, in a canonical ensemble, we specify the temperature T and use this to deduce the internal energy U.

Then our macrostates M are specified by our temperature T, as well as (potentially) other variables. In this scenario, we're allowing heat to be inputted, but no external work can be done on or by the system.

Proposition 199

We can have our system maintained at a fixed temperature T if it is in contact with a **reservoir** or **heat bath** at temperature T.

We make the assumption that the heat bath is sufficiently large that it has basically negligible change in temperature! For example, a glass of boiling water in a large room will cool down to room temperature, while the room's temperature is basically fixed.

Question 200. *How do we find the probability of any given microstate*

$$P_T(\mu)?$$

Let's say our reservoir has some microstates μ_R and energies associated to them $H_R(\mu_R)$. Similarly, our system has some microstates μ_S and energies $H_S(\mu_s)$. We can think of the reservoir and system as one larger system $R \oplus S$: this is now mechanically and adiabatically isolated, so we can think of this as a microcanonical ensemble!

Our total energy here is E_{total}, and we know that (because our system is much smaller than the reservoir), we have $E_{\text{sys}} \ll E_{\text{total}}$. Then the probability of some microstate $(\mu_s \oplus \mu_R)$ is

$$P(\mu_S \oplus \mu_R) = \frac{1}{\Gamma_{S\oplus R}(E_{\text{total}})}$$

if our total energy $H_S(\mu_S) + H_R(\mu_R) = E_{\text{total}}$), and 0 otherwise. This is a joint probability distribution, so if we want a specific μ_S for our system, we "integrate" or "sum out" all values of R:

$$P(\mu_S) = \sum_{\{\mu_R\}} P(\mu_S \oplus \mu_R).$$

Also recall that we can write the conditional probability

$$P(S|R) = \frac{P(S,R)}{P(R)},$$

so (because probability is $\frac{1}{\Gamma}$, the multiplicity), this expression can also be written as

$$P(\mu_S) = \frac{\Gamma_R(E_{\text{total}} - H_S(\mu_S))}{\Gamma_{S\oplus R}(E_{\text{total}})}$$

since we sum over the multiplicity of R given a specific energy. Notice that the denominator here is independent of μ_S: it's a constant, so we can rewrite this in terms of the entropy of our reservoir: given that $S = k_b \ln \Gamma$,

$$P(\mu_s) \propto \exp\left[\frac{1}{k_B} S_R(E_{\text{total}} - H_S(\mu_S))\right].$$

In the limit where the energy of the system is significantly smaller than the total energy of the system and reservoir, we can do a Taylor expansion of this entropy to simplify this further! Then (treating $H_S(\mu_S)$ as our variable), because $E_R = 1 - H_S(\mu_S)$,

$$S_R(E_{\text{total}} - H_S(\mu_S)) \approx S_R(E_{\text{total}}) - H_S(\mu_S)\frac{\partial S_R}{\partial E_R}.$$

Dropping the S subscripts for simplicity and plugging into the probability distribution, since $\beta = \frac{1}{k_B T}$, and $\exp[S_R]$ is some constant that doesn't depend on our microstate (to first order), we have the following fact:

Theorem 201 (Canonical ensemble)

For a fixed temperature, the probability of our microstate is

$$\boxed{P(\mu) = \frac{e^{-\beta H(\mu)}}{Z}}.$$

This is a probability distribution, which means Z is our normalization factor! Z is taken from the German word "zustandssumme," which means "sum over states." That explains why it's our normalization factor here: if we add up $P(\mu)$ for all states, we'll get $\frac{Z}{Z} = 1$.

Definition 202

Z here is called a **partition function**. In this case,

$$Z = \sum_\mu e^{-\beta H(\mu)}.$$

36.3 Looking more at the partition function

Why are partition functions useful? Basically, many macroscopic quantities can be described in terms of our function Z. Remember that the multiplicity Γ was important in our microcanonical ensemble for finding S: now Z takes its place, and it's a lot easier to write this down!

Fact 203

The exponential term here is called a **Boltzmann factor**.

So now notice that the energy is no longer fixed: it's some random variable $\mathcal{H}$, and we want to find its mean, variance, and other information. (Spoiler: the fluctuations are small, so in the thermodynamic limit, this will look a lot like the microcanonical ensemble!)

First of all, we can write the probability of a given energy $\overline{H}$ as a sum

$$P(\overline{H}) = \sum_\mu P(\mu)\delta(H(\mu) - \overline{H}),$$

since we want to add up the states with our given energy. Since all probabilities in this sum are equal, this evaluates to

$$\frac{e^{-\beta H}}{Z}\sum_{\mu}\delta(H(\mu)-\overline{H})$$

The sum is the number of microstates with a given energy H', which is just our multiplicity $\Gamma(\overline{H})$. Thus, this can be rewritten as

$$P(\overline{H})=\frac{\Gamma(\overline{H})e^{-\beta\overline{H}}}{Z}.$$

We'll talk more about this later, but we can write Γ in terms of our entropy: this will yield an expression of the form

$$\frac{1}{Z}\exp\left[\frac{S(\overline{H})}{k_B}-\frac{\overline{H}}{k_BT}\right],$$

and since this is a constant times $\overline{H}-TS$, this is related to our Helmholtz free energy!

36.4 Evaluating the statistics

Note that as $\overline{H}$ increases, Γ rapidly increases, while $e^{-\beta\overline{H}}$ rapidly decreases. Therefore, if we plot this, we'll find that the distribution is sharply peaked around some energy U. What's the mean and variance of this distribution?

Well, the average here is

$$\langle H\rangle=\sum_{\mu}H(\mu)\frac{e^{-\beta H(\mu)}}{Z}$$

which can be written as a derivative

$$-\frac{1}{Z}\frac{\partial}{\partial\beta}\sum_{\mu}e^{-\beta H(\mu)}.$$

But the sum here is just Z, so

$$\langle H\rangle=-\frac{1}{Z}\frac{\partial Z}{\partial\beta}=\boxed{-\frac{\partial(\ln Z)}{\partial\beta}}.$$

How sharp is this distribution - that is, how narrow is it? We can compute the variance $\text{var}(H)$ by first considering

$$-\frac{\partial Z}{\partial\beta}=\sum_{\mu}He^{-\beta H}.$$

Taking another derivative with respect to β,

$$-\frac{\partial^2 Z}{\partial\beta^2}=\sum_{\mu}H^2e^{-\beta H},$$

but this looks a lot like the average value of H^2! We then find

$$\langle H^2\rangle=\frac{1}{Z}\frac{\partial^2 Z}{\partial\beta^2}.$$

Therefore,

$$\text{var}(H)=\langle H^2\rangle-\langle H\rangle^2=\boxed{\frac{1}{Z}\frac{\partial^2 Z}{\partial\beta^2}-\left(\frac{1}{Z}\frac{\partial Z}{\partial\beta}\right)^2}.$$

This can be rewritten as

$$\frac{1}{Z}\frac{\partial}{\partial\beta}\left(\frac{\partial Z}{\partial\beta}\right)+\frac{\partial}{\partial\beta}\left(\frac{1}{Z}\right)\frac{\partial Z}{\partial\beta}$$

which is actually just

$$\frac{\partial}{\partial\beta}\left(\frac{1}{Z}\frac{\partial Z}{\partial\beta}\right)$$

by the product rule! If we look back, this tells us that

$$\text{var}(H) = -\frac{\partial}{\partial\beta}\langle H\rangle,$$

where $\langle H\rangle = U$ is our "average energy." Since $\beta = \frac{1}{k_B T}$, this means we can rewrite this as

$$\text{var}(H) = k_B T^2 \frac{\partial U}{\partial T}.$$

Since C_V, our heat capacity, is defined to be $\frac{\partial U}{\partial T}\big|_{V,N}$, if we define $\hat{C}_V = \frac{C_V}{N}$ (the heat capacity per particle),

$$\boxed{\text{var}(H) = N k_B T^2 \hat{C}_V}.$$

If we think of fractional fluctuations, we want to look at the ratio of our standard deviation to the mean. The mean is proportional to N, the number of particles, since it is an extensive quantity, but the standard deviation of H is proportional to $\sqrt{N}$. This means

$$\frac{\sqrt{\text{var}(H)}}{\text{mean}(H)} \propto \frac{1}{\sqrt{N}}$$

which is very small on the thermodynamic scale, and thus our variable is highly concentrated! This means that we can basically think of the system almost as being a microcanonical ensemble with fixed energy $\langle H\rangle$: it was just easier to get to this point, since all we need to do is compute our partition function Z.

36.5 Writing macroscopic quantities in terms of the partition function

We found earlier that

$$\langle H\rangle \equiv U = -\frac{\partial}{\partial\beta}\ln Z.$$

Since $\beta = \frac{1}{k_B T}$,

$$\frac{\partial}{\partial\beta} = \frac{\partial}{\partial T}\left(\frac{\partial T}{\partial\beta}\right) \implies \frac{\partial}{\partial\beta} = -k_B T^2\frac{\partial}{\partial T} \implies U = k_B T^2 \frac{\partial}{\partial T}\ln Z.$$

Let's next look at the entropy: by definition,

$$S = -k_B \sum_j p_j \ln p_j.$$

Plugging in our probability distribution,

$$S = -k_B \sum \frac{e^{-\beta E_j}}{Z}(-\beta E_j - \ln Z).$$

Rewriting this in terms of the internal energy U,

$$S = k_B\beta U + k_B \ln Z = \boxed{\frac{U}{T} + k_B \ln Z}.$$

From this, we can see that

$$-k_B T \ln Z = U - TS = F,$$

which is the Helmholtz free energy! Recall that we have (by the differential $dF = -SdT - PdV$,

$$\left.\frac{\partial F}{\partial T}\right|_V = -S, \left.\frac{\partial F}{\partial V}\right|_T = -P,$$

so we can then get our other quantities by taking partial derivatives of F. This can always yield our equation of state!

37 April 24, 2019 (Recitation)

Let's start by discussing a topic from last problem set.

37.1 Mixing entropy

Consider two gases (red and white) that are mixed in a box with volume V. How can we compare this situation to one where the two are in separated boxes (each with volume V)?

Specifically, we have three scenarios: in A, red and white are in the same box of volume V, in B, red and white are in two different boxes each with volume V, and in C, red and white are in two different boxes each with volume $\frac{V}{2}$.

What can we say about the entropy? First of all, any scenario in C can occur in scenario A, so the entropy $S_A > S_C$. But let's go ahead and do the mixing derivation again! Recall the Sackur-Tetrode equation

$$S = Nk_B\left(\cdots + \ln\frac{V}{N}\right).$$

(Everything else - masses, temperatures, and so on - are constant.) Usually, A and C are compared in textbooks: there is a partition in a box that is then removed. If this partition breaks our volume V into two parts with $N_1 = c_1 N$ and $N_2 = c_2 N$ of the total (for example, 80 percent Nitrogen on one side and 20 percent on the other), we can apply the Sackur-Tetrode equation: V and N are proportional if our system is at equal pressure on both sides:

$$S_C = k_B\left[c_1 N \ln\frac{c_1 V}{c_1 N}\right] + k_B\left[c_2 N \ln\frac{c_2 V}{c_2 N}\right] = Nk_B \ln\frac{N}{V}$$

(before we mix). After we mix, though, we sum instead

$$S_A = k_B\left[c_1 N \ln\frac{V}{c_1 N}\right] + k_B\left[c_2 N \ln\frac{V}{c_2 N}\right]$$

because each component now has the full volume instead of only a fraction! Note that we can write this as

$$\begin{aligned} S_A &= k_B\left[c_1 N \ln\frac{V}{N} - \ln c_1\right] + k_B\left[c_2 N \ln\frac{V}{N} - \ln c_2\right] \\ &= S_C - \boxed{k_B N(c_1 \ln c_1 + c_2 \ln c_2)}, \end{aligned}$$

and this last term is called the **mixing entropy**.

Fact 204

This generalizes to more than two components as well.

Here's a second derivation that is even simpler! We can ignore the color of the gas particles at first and go through the derivation for the Sackur-Tetrode equation in both cases. But then we have $N = N_1 + N_2$ particles, and we need to label N_1 of them red and N_2 of them white. In situation C, we must label all of the ones on the left red and all of the ones on the right white, but we're free to do this arbitrarily in situation A! So we get an extra multiplicity of $\binom{N}{N_1}$, and that will contribute an extra

$$\Delta S = k_B \ln \Gamma = k_B \frac{N!}{N_1!N_2!} = -k_B N \left(\frac{N_1}{N} \ln \frac{N_1}{N} + \frac{N_2}{N} \ln \frac{N_2}{N} \right),$$

as expected.

But notice that situations A and B are now essentially the same! When doing the Sackur-Tetrode equation, we made the explicit assumption that we could treat the red and white gases separately. So that actually tells us that

$$S_A = S_B.$$

Proposition 205

So is there a way for us to separate the two gases from situation A into situation B without doing any work?

We can just construct a (one-directional) special membrane that only allows one of them to pass through! For example, if one of them is large and one is small, we could have small pores - we don't violate any real laws of physics this way.

Fact 206

Specifically, think of a red and white membrane that allow only red and white particles to pass through, respectively. Then we can enclose our red and yellow particles in boxes and "translate" the red box until it is side-by-side with the white box!

So to get from situation A to situation B, we can place two boxes next to each other (each with volume V), separated by a membrane. No work is done this way - the translation doesn't do any work or allow for any transfer of heat!

To get from situation B to situation C, then, we can compress the containers of situation B: if we do this isothermally, the work that we do creates an equivalent amount of heat (since $\Delta U = \Delta Q + \Delta W = 0$), and the heat ΔQ is actually just $T\Delta S$, the change in entropy of mixing.

37.2 Looking at the canonical ensemble

Recall the setup: we have a system connected to a reservoir at constant temperature T. What's the probability of our system being at an energy E?

Basically, we think of the reservoir plus the system as a microcanonical ensemble: then the probability is just proportional to the number of microstates of the system at energy E, which is the same as the reservoir being at energy $U - E$ (where the total energy of reservoir plus system is U). Then writing this in terms of entropy,

$$p(E) = ce^{S_R(U-E)/k_B} \propto \Gamma_R(U - E).$$

But now assuming $E \ll U$, we can do the Taylor expansion

$$= ce^{S_R(U)/k_B} e^{\frac{\partial S_R}{\partial U} \cdot -E/k_B}$$

and the first two terms here can be encapsulated as one constant: remembering that $\frac{\partial S}{\partial U} = \frac{1}{T}$,

$$= \frac{e^{-E/(k_B T)}}{Z}$$

for some constant Z. As a normalization factor, we now have our "partition function"

$$Z = \sum_j e^{-E_j/(k_B T)}.$$

Fact 207

Probably the most important idea in this derivation is thinking of the reservoir and the system as one big system, since it allows us to think about energy!

As a question: Taylor expansions are first order, so what's the range of validity for the assumptions that we made? Specifically, remember that the second derivative, which is related to the "curvature," tells us how much the first derivative (which is related to **temperature**) is changing when we vary our parameter. So we neglect the change in temperature as we change our energy E!

So this error term only comes about when the reservoir is too small: this means that with small systems (or theoretical "canonical ensembles," really), our derivation gives an exact result.

Recall that we did some derivations related here earlier on: on an earlier problem set, we had n_a particles in states of energy ε_a, and we had specific limits on $\sum n_a = N$ and $\sum n_a \varepsilon_a = U$. We found then that

$$n_a = N \frac{e^{-\beta \varepsilon_a}}{\sum_j e^{-\beta \varepsilon_j}};$$

this was found by looking at the combinatorics of distributing particles over energy states and incorporating Lagrange parameters for the constraints on N and U. Notice that we can now translate this: we have N systems, each at some energy level, and now $\frac{n_a}{N}$ tells us the probability that a random system has energy ε_a!

How do we make this into the language of a "canonical ensemble?" Any given system is coupled to the combined reservoir of $N-1$ other systems: as long as $1 \ll N$, and we are at thermal equilibrium, the temperature of the reservoir will not change very much, and we can use our canonical ensemble formula.

38 April 25, 2019

We'll discuss the canonical ensemble some more today! We'll also start looking at a few more examples - the main idea is that the canonical ensemble will result in significantly simpler calculations, since microcanonical and canonical ensembles predict the same results in the thermodynamic limit.

38.1 Review and remarks on the canonical ensemble

Here's a helpful table: "macrostate" tells us which macrostates are fixed or given to us as constraints.

Ensemble	Macrostate	$p(\mu)$	Normalization factor
Microcanonical	(U, x)	$\frac{\delta(H(\mu)-U)}{\Gamma}$	$S(U, x) = k_B \ln \Gamma$
Canonical	(T, x)	$\frac{\exp(-\beta H(\mu))}{Z}$	$F(T, x) = -k_B T \ln Z$

In other words, the free energy in the canonical ensemble and the entropy in the microcanonical ensemble play similar roles.

Note that the logarithm of the partition function also gives other quantities. For example,

$$C_V = N\hat{c}_V = \frac{1}{k_B T^2}\sigma_x^2$$

where σ_x^2 is the variance of the energy of our system at thermal equilibrium. This shows up in more sophisticated discussions of statistical physics! Since our heat capacity is a response function

$$C_V = \left.\frac{\partial Q}{\partial T}\right|_V ,$$

it describes how much heat is transferred given some perturbation of our temperature. But instead of having to do that perturbation, this equations tells us that observing fluctuations at thermal equilibrium works as well!

Fact 208

This is a special case of something called the **fluctuation-dissipation theorem**.

As another comment: we've mostly been summing over states to calculate the partition function, but we're often instead given the energies as opposed to states. For example, if we're given that our energies are $\{E_i | 1 \le i \le M\}$, we can write our partition function as

$$Z(T, N, V) = \sum_i g_i e^{-\beta E_i},$$

where g_i is the **degeneracy** of the energy level (letting us know how many microstates are at that given energy E_i). In the continuous limit, we can approximate this instead as

$$Z(T, N, V) = \int dE \frac{dN}{dE} e^{-\beta E},$$

where $\frac{dN}{dE}$ is the density of states we discussed earlier in class.

One more comment: since $F = -k_B T \ln Z$, we know that $Z = \exp(-\beta F)$. This looks a lot like the Boltzmann factor $\exp(-\beta E_i)$ for a particular microstate! Thus we can think of $\exp(-\beta E_i)$ as being a volume or weight in phase space for a particular state: Z is then the weight contributed by all states for a given N and V. (For a canonical ensemble, since $F = U - TS$ is the "free energy," it can be thought of as the amount of energy available for the system to do useful work.)

38.2 Looking at the two-state system again

Example 209

Let's go back and say that we have a system where a particle can be in two states: $E_\uparrow = \frac{\varepsilon}{2}$ and $E_\downarrow = -\frac{\varepsilon}{2}$.

The partition function here is then

$$Z = e^{\beta\varepsilon/2} + e^{-\beta\varepsilon/2} = 2\cosh\left(\frac{\beta\varepsilon}{2}\right),$$

and now the internal energy of our system

$$U = -\frac{d}{d\beta}\ln Z = -\frac{\frac{\partial Z}{\partial \beta}}{Z} = -\frac{\varepsilon}{2}\tanh\left(\frac{\beta\varepsilon}{2}\right)$$

goes from $-\frac{\varepsilon}{2}$ to 0 as our temperature T gets larger (which looks a lot like our microcanonical ensemble!). Then the probability of being in each of our two states is

$$P_\uparrow = \frac{e^{-\varepsilon/(2k_BT)}}{2\cosh\frac{\varepsilon}{2k_BT}+1} = \frac{1}{1+e^{\varepsilon/(k_BT)}}, P_\downarrow = 1 - P_\uparrow.$$

Indeed, $P_\uparrow$ goes from 0 to $\frac{1}{2}$, and $P_\downarrow$ goes from 1 to $\frac{1}{2}$ as our temperature T gets larger.

Next, let's calculate the entropy: from our free energy F, we find that

$$S = \frac{U}{T} + k_B \ln Z = -\frac{\varepsilon}{2T}\tanh\frac{\varepsilon}{2k_BT} + k_B \ln\left(2\cosh\frac{\varepsilon}{2k_BT}\right).$$

Looking at the limits, this goes to $-\frac{\varepsilon}{2T} + k_B \ln\exp\frac{\varepsilon}{2k_BT} \to 0$ as $T \to 0$. This is consistent with what we already know: there's only one possible ground state! On the other hand, as $T \to \infty$, this goes to $k_B \ln 2$, which is consistent with the microcanonical ensemble and the intuition of our system.

We can also calculate our heat capacity:

$$C = \frac{\partial U}{\partial T} = -\frac{\varepsilon}{2}\frac{\partial}{\partial T}\tanh\frac{\varepsilon}{2k_BT} = k_B\left(\frac{\varepsilon}{2k_BT}\right)^2\frac{1}{\cosh^2\frac{\varepsilon}{2k_BT}}.$$

Plotting this as a function of thermal energy k_BT, the distribution is (just like in the microcanonical ensemble) unimodal. Think of this in terms of fluctuations! We can calculate the standard deviation of our energy at a given temperature:

$$\sigma(U) = k_BT\sqrt{\frac{C_V}{k_B}} = \frac{\varepsilon/2}{\cosh\frac{\varepsilon}{k_BT}},$$

and the behavior models the microcanonical ensemble as $T \to 0$ and as $T \to \infty$: it is

38.3 Systems with a large number of particles

Let's now think about how to compute the partition function in larger systems! Specifically, how can we do this for distinguishable versus indistinguishable particles - more specifically, how many microstates are there with a given number distribution $(n_1, \cdots, n_M)$?

Assume our particles are distinguishable. If we have N particles, there are $\binom{N}{n_1}$ ways to pick n_1 of them to go in the first energy level, then $\binom{N-n_1}{n_2}$ ways for the next energy level, and so on: this gives the multinomial coefficient

$$\frac{N!}{n_1!n_2!\cdots n_m!}.$$

This is the degeneracy factor for a given energy that we want! Let's look at us having N distinct harmonic oscillators, where there are M distinct energy levels $\varepsilon_1, \cdots, \varepsilon_M$. Then n_k denotes the number of distinct harmonic oscillators that are excited at our energy level ε_k. Now we can write our partition function

$$Z = \sum_{n_1+\cdots+n_m=N}\frac{N!}{\prod_{j=1}^m n_j!}\exp\left[-\beta\sum_{k=1}^M \varepsilon_k n_k\right].$$

where the first term is the degeneracy factor, and the $\sum_{k=1}^M \varepsilon_k n_k$ is the total energy of our state. Let's see if we can simplify this: first of all, we can expand the sum inside the exponential as a product:

$$Z = \sum_{n_1+\cdots+n_m=N}\frac{N!}{\prod_{j=1}^M n_j!}\prod_{k=1}^M \exp[-\beta\varepsilon_k]^{n_k}.$$

Now remember the binomial theorem:

$$(a_1 + a_2)^N = \sum_{n_1+n_2=N} \frac{N!}{n_1!n_2!} a_1^{n_1} a_2^{n_2}$$

which generalizes to the multinomial theorem

$$(a_1 + \cdots + a_M)^N = \sum_{n_1+\cdots+n_M=N} \frac{N!}{n_1!\cdots n_M!} a_1^{n_1} \cdots a_M^{n_M}.$$

This is exactly what we want here! By the multinomial theorem, we can rewrite our Z as

$$Z = (\exp[-\beta\varepsilon_1] + \exp[-\beta\varepsilon_2] + \cdots + \exp[-\beta\varepsilon_M])^N,$$

and since the sum inside the parentheses is actually the partition function for an individual harmonic oscillator Z_1, we can write

$$Z = Z_1^N.$$

Proposition 210

This means that the **partition function for identical distinguishable systems is multiplicative!**

It's important here that the individual oscillators here were distinguishable to get the multinomial coefficient. But if we want to do the same exercise with indistinguishable particles, the degeneracy factor becomes just 1.

Then calculations in general are a lot uglier, but we can look at a special case: let's say we have high temperatures, which is basically looking at our harmonic oscillators in the classical limit! Then our states are basically uniquely defined by just the list of energy levels (and we rarely need to deal with overcounting issues), so we just end up with

$$Z \to \frac{1}{N!} Z_1^N.$$

Talk to Professor Ketterle when we're not looking at the classical limit, though!

Example 211

Let's explicitly compute the partition function for N distinct harmonic oscillators and work with it!

Then we have (using facts from quantum physics)

$$Z_1 = \sum_{n=0}^{\infty} \exp\left(-\beta\hbar\omega\left(n + \frac{1}{2}\right)\right) = e^{-\beta\hbar\omega/2} \sum_{n=0}^{\infty} \left(e^{-\beta h\bar{\omega}}\right)^n = \frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}}$$

by the geometric series formula, and Z is just Z_1^N.

Now that we have our partition function, we can compute our other macroscopic variables:

$$U = -\frac{\partial}{\partial B} \ln Z = -N\frac{\partial}{\partial\beta} \ln Z_1$$

which evaluates out to

$$U = N\hbar\omega\left(\frac{1}{e^{-\beta\hbar\omega} - 1} + \frac{1}{2}\right).$$

As we take our temperature $T \to 0$, we only occupy the ground energy level, which is $\frac{1}{2}N\hbar\omega$. On the other hand, if our energy is large, $k_B T \gg$

$hbar\omega \implies T \gg \frac{\hbar\omega}{k_B}$, which is the high-temperature classical limit: think about what happens this way! It's also good to think about how to calculate the heat capacity and entropy of this system from here.

One comment: we'll deal with a lot of series in these kinds of calculations, so we should become comfortable with the relevant techniques!

39 April 29, 2019 (Recitation)

Today's recitation is being taught by the TA.

Recently we've been discussing the canonical ensemble - as a first question, when do we use it?

Basically, whenever we have a system, we have some different states: for example, we can consider them by energy level. In a microcanonical ensemble, we consider a system that is cut off from all of its surroundings: no mass or heat transfer, so we have conservation of energy.

But in the real world, most systems are connected to the environment, even if we have a fixed number of particles. A canonical ensemble is just the most basic example of this!The main idea is that the probability of finding the system in an energy ε_i is

$$p_i = \frac{e^{-\beta\varepsilon_i}}{Z},$$

where Z is the partition function. How do we derive this? In general, our system needs to maximize entropy S

$$\frac{S(\{p_i\})}{k_B} = -\sum p_i \ln p_i.$$

We have the constraints $\sum p_i = 1$, and also we can say that the "average" energy of our system is constant: writing that in terms of our probabilities, we have $\sum p_i\varepsilon_i = U$.

Fact 212

An idea here is that thermodyanmics and statistical mechanics are connected by the idea of "averaging."

Now that we have our constraints, we use Lagrange multipliers! Basically, given a function f that we want to maximize and a bunch of constraints of the form $\{g_i(x) = 0\}$, it's equivalent to minimizing over x and λ (the Lagrange multipliers)

$$F(x) = f(x) - \sum \lambda_i g_i(x).$$

To minimize a vector function, consider the gradient: we want $\frac{\partial F}{\partial x_i} = \frac{\partial F}{\partial \lambda_j} = 0$ for all i, j. (The latter, by the way, is just saying that each $g_i(x) = 0$.)

So let's apply this to our problem! We have

$$G(\{p_i\}) = -\sum_i p_i \ln p_i - \lambda(\sum_i p_i - 1) - \beta(\sum p_i\varepsilon_i - U)$$

Taking the derivative with respect to p_i,

$$\frac{\partial G}{\partial p_i} = -\ln p_i - 1 - \lambda - \beta E_i = 0 \implies p_i = e^{-1-\lambda-\beta E_i}.$$

Meanwhile, taking the derivative with respect to λ just yields $\sum p_i = 1$: this means we have to normalize our probability distribution: this yields

$$p_i = \frac{e^{-\beta\varepsilon_i}}{\sum_a e^{-\beta\varepsilon_a}},$$

as before!

So we can actually define temperature T directly in terms of our Lagranage multiplier: $\beta = \frac{1}{k_B T}$. Our partition function looks important, but can we do anything with it? First of all, we can write it in two different ways:

$$Z = \sum_{\text{states } i} e^{-\beta\varepsilon_i} = \sum_{\text{energies } \{\varepsilon_n\}} g_n e^{-\beta\varepsilon_n}.$$

(The g_n degeneracy factor is pretty important here!) Let's try to write this in terms of some quantities that we already know about: since our energy

$$U = \langle E\rangle = \sum_i p_i\varepsilon_i, \sum_i \frac{\varepsilon_i e^{-\beta\varepsilon_i}}{Z},$$

note that

$$-\frac{1}{Z}\frac{\partial Z}{\partial\beta} = \frac{1}{Z}\sum_i \varepsilon_i e^{-\beta\varepsilon_i} = -\frac{\partial \ln Z}{\partial B} = U.$$

Now that we have energy, we can think about this in certain kinds of systems. A real magnet is a bunch of small atoms that are magnetic dipoles: microscopically, those dipoles are oriented in magnetic domains, and this order breaks down if the magnetic heats up! (Basically, there's no clear direction for the dipoles to point.) On the other hand, magnets at very low temperature have a defined order: there's a quantity "magnetization" which is the average $\frac{|\mu|}{N}$, and it decreases to 0 at some Curie temperature T_C. We might see later on that the decay is actually square-root up until that point!

Example 213

So if we have a system with a bunch of magnetic spins, we have a total energy (in a two-state system)

$$E = -(\sum \mu_i)B.$$

Knowing B and knowing the temperature T, we want to think about the average magnetization

$$m = \frac{\langle\sum \mu_i\rangle}{N}.$$

We can think about this in another way: consider

$$\sum_{\{\mu_i\}} \frac{(\sum \mu_j)e^{\beta\sum_i \mu_i)B}}{Z} = \frac{1}{\beta}\frac{\partial \ln Z}{\partial B} = Nm,$$

and now we have what we want: we've found m in terms of the partition function!

So looking at probability distribution, we want the fluctuation in U

$$(\Delta U)^2 = \langle E^2\rangle - \langle E\rangle^2.$$

The second term here is already known: it is $\left(-\frac{\partial \ln Z}{|partial\beta}\right)^2$, and then how can we find the first term? Take more derivatives! Note that

$$\frac{\partial^2 \ln Z}{\partial\beta^2} = \frac{\partial}{\partial\beta}\left(-\frac{1}{Z}\sum \varepsilon_i e^{-\beta\varepsilon_i}\right),$$

and now by the product rule,

$$= \frac{1}{Z}\left(\sum \varepsilon_i^2 e^{-\beta\varepsilon_i}\right) - \frac{1}{Z^2}\left(\frac{\partial Z}{\partial\beta}\right)^2.$$

The first term is now the expected value of $\langle E^2\rangle$, and therefore the fluctuations $(\Delta U)^2$ are connected to partial derivatives $\frac{\partial^2 \ln Z}{\partial\beta^2}$!

We can think of this in terms of other situations as well: remember that when we have our heat capacity $C = \frac{\partial U}{\partial T}$, we can also think about the relationship between $\frac{\partial}{\partial T}$ and $\frac{\partial}{\partial \beta} = \frac{\partial}{\partial T}\frac{\partial T}{\partial B}$, and now $\frac{\partial B}{\partial T} = -\frac{1}{k_B T^2}$ by direct substitution.

Fact 214

It will be important to switch between β and T in the future!

So then

$$C = \frac{\partial U}{\partial T} = -\frac{1}{k_B T^2}\frac{\partial U}{\partial \beta} = \frac{1}{k_B T^2}\frac{\partial^2 \ln Z}{\partial \beta^2} \implies \boxed{(\Delta U)^2 = k_B T^2 C}.$$

C is a response function: we can generally do something to our system and see how this changes the properties. The equation, then, connects theoretical quantities to measurable experimental quantities!

As an exercise, we can prove that if $\chi = \frac{\partial \langle M \rangle}{\partial B}$, we have a similar relation: then $(\Delta M)^2 = k_B T^2 \chi$.

Finally, is there a way for us to compute the entropy from our partition function? We have

$$S = -k_B \sum p_i \ln p_i = -k_B \sum p_i(-\ln Z - \beta \varepsilon_i).$$

This can be rewritten as

$$= -k_B \beta \sum p_i \varepsilon_i + k_B \ln Z \sum p_i.$$

The sum of the p_is is 1, and the sum of the $p_i \varepsilon_i$s is our energy $\langle E \rangle$, so we have

$$S = \frac{1}{T}\langle E \rangle + k_B \ln Z.$$

This should look familiar: since the Helmholtz free energy $F = U - TS$, we can rewrite

$$F = k_B T \ln Z = \frac{1}{\beta} \ln Z.$$

But now knowing F and U, we are able to find S, and then we can get whatever else we want! In fact, we can also use this to prove the first and second law.

Example 215

Let's go back to our two-state system.

We can say that our energy is

$$E = -B \sum \mu_i = -(\sum \sigma_i)\mu B,$$

where each σ_i is ± 1. Intuitively, we should expect independence, so Z should be multiplicative. Indeed, our collection of spins $\{\sigma_i\} = (\pm 1, \cdots, \pm 1)$: this is 2^N configurations, and we can then write this as

$$\sum_{\{\sigma_i\}} e^{-\beta \mu B (\sum \sigma_i)}$$

and now we can expand this out as (since $\sum a_i b_j = \sum a_i \sum b_j$)

$$\sum_{\{\sigma_i\}} e^{\beta \mu B \sigma_1} e^{\beta \mu B \sigma_2} \cdots = \prod_i \left(\sum_{\sigma_i = \pm 1} e^{\beta \mu B \sigma_i} \right)$$

But we always have the same B, μ, β, σ, so each sum is equal: this means we have

$$Z = \left(e^{-\beta \mu B} + e^{\beta \mu B}\right)^N, = 2^N \cosh(\beta \mu B)^N$$

and we've found Z explicitly.

Finally, looking at the magnetization one more time,

$$\langle M\rangle = \frac{1}{\beta}\frac{\partial \ln Z}{\partial B} = N\mu\tanh(\beta\mu B).$$

So m, the average magnetization, goes from 1 to -1: a strong enough magnetic field will align all the spins in the same direction! We also can expand the curve around $B = 0$ to figure out the linear relationship between m and B at small B.

40 April 30, 2019

There is a problem set due this week on Friday - it covers a lot of concepts about the canonical ensemble, and each one is a good way to put what we have learned into practice! Today, we'll look at more examples of canonical ensembles and see the applications to black-body radiation and other topics. (This is because we can think of black-body radiation as a bunch of oscillators, and we can learn about their equation of state to deduce further information.)

40.1 Back to the two-level system

Let's again imagine that we have N particles, each with two energy levels. This can be thought of as assigning an up or down arrow to each of N spots on a lattice - the particles are then distinguishable, because we can identify them by their position.

Let's say the up and down spin have energy levels of $\frac{\varepsilon}{2}$ and $-\frac{\varepsilon}{2}$, respectively. We found last time that the partition function for one particle was

$$Z_1 = e^{-\varepsilon\beta/2} + e^{\varepsilon\beta/2} = 2\cosh\frac{\beta\varepsilon}{2},$$

and then the partition function for all N particles, by independence, is just

$$Z = Z_1^N = \left(2\cosh\frac{\beta\varepsilon}{2}\right)^N.$$

We can now calculate the thermodynamic properties of the system: first of all,

$$\ln Z = N\ln\left(2\cosh\frac{\beta\varepsilon}{2}\right)$$

This actually gives us the other quantities that we want: we can calculate

$$U = -\frac{N\varepsilon}{2}\tanh\frac{\varepsilon}{2k_BT},$$

$$S = N\left(-\frac{\varepsilon}{2k_BT}\tanh\frac{\varepsilon}{2k_BT} + k_B\ln\left(2\cosh\frac{\varepsilon}{2k_BT}\right)\right).$$

We'll also find that the heat capacity C_V is just N times the heat capacity of a single particle, and the fluctuation standard deviation of U is

$$\sigma(U) = k_BT\sqrt{\frac{C_V}{k_B}} = \frac{\sqrt{N}\varepsilon/2}{\cosh\frac{\varepsilon}{2k_BT}}.$$

Since $U \propto N$ and $\sigma \propto N$, we have concentration of the energy!

40.2 Another look at the harmonic oscillators

Remember that for a similar system we examined (the quantum harmonic oscillator), we have that

$$Z_1 = \sum_{n=0}^{\infty} \exp(-\beta\hbar\omega(n+\tfrac{1}{2})),$$

and then again by independence,

$$Z = Z_1^N \implies U = -\frac{\partial(\ln Z)}{\partial\beta} = N\hbar\omega\left(\frac{1}{e^{\beta\hbar\omega}-1}+\frac{1}{2}\right).$$

A good question here: what do we know about temperature in the limits? As $T \to 0$, $\beta \to \infty$, and that means $U = \frac{N\hbar\omega}{2}$. This is the ground state: all harmonic oscillators are in their ground state of $\frac{\hbar\omega}{2}$.

Meanwhile, when $T \gg \frac{\hbar\omega}{k_B}$, so the temperature is large enough for $\beta\hbar\omega$ to go to 0, U becomes large. We can define $\beta\hbar\omega = x$: now doing a Laurent expansion because our quantity blows up at 0,

$$\frac{1}{e^x-1} = \frac{1}{x+\frac{x^2}{2}+\frac{x^3}{6}+\cdots} = \frac{1}{x}\cdot\frac{1}{1+\frac{x}{2}+\frac{x^2}{2}+\cdots}.$$

Now we can neglect higher order terms and this simplifies to

$$\frac{1}{x}\left(1-\frac{x}{2}+O(x^2)\right) = \frac{1}{x}-\frac{1}{2}+O(x).$$

Plugging this back in, as $T \to \infty$, we have

$$U \to N\hbar\omega\left(\frac{1}{\beta\hbar\omega}-\frac{1}{2}+\frac{1}{2}\right) = Nk_BT,$$

and this is the famous result from the equipartition theorem! (Harmonic oscillators contribute a potential and kinetic quadratic term to the Hamiltonian.)

So now, how can we determine the probability distribution for a single oscillator? We have

$$p(E) = \frac{1}{Z_1}e^{-E/(k_BT)}$$

(as a property of the canonical ensemble in general), and plugging in our specific value of Z_1,

$$p(E) = (1-e^{-\beta\hbar\omega})e^{\beta\hbar\omega/2}e^{-\beta(n+\frac{1}{2})\hbar\omega},$$

and this is actually a probability distribution over our ns:

$$p(E) = p(n) = (1-e^{-\beta\hbar\omega})e^{-n\beta\hbar\omega}.$$

This is actually a **geometric distribution**: if we take $a = e^{-\beta\hbar\omega}$, the probability

$$p(n) = (1-a)a^n.$$

This is indeed normalized: $\sum_n p(n) = 1$, and the average value of n, called the average **occupation number** of a harmonic oscillator, is

$$\langle n\rangle = \frac{1}{e^{\hbar\omega/(k_BT)}-1}.$$

Finally, let's compute the heat capacity:

$$C - \frac{\partial U}{\partial T} = N\hbar\omega \frac{\partial}{\partial T}\left(\frac{1}{e^{\hbar\omega/(k_B T)} - 1} + \frac{1}{2}\right)$$

which simplifies to

$$= Nk_B \left(\frac{\hbar\omega}{k_B T}\right)^2 \frac{e^{\hbar\omega/(k_B T)}}{(e^{\hbar\omega/(k_B T)} - 1)^2}.$$

Again, let's look at the limits: when $T \to \infty$, all $\frac{\hbar\omega}{k_B T}$ terms go to 0 (define that fraction to be y). Specifically,

$$(e^y - 1)^2 \sim (y + O(y^2))^2 \sim y^2, e^y \sim 1 + \cdots \sim 1,$$

and thus as $T \to \infty$, we have

$$C \to Nk_B.$$

This is a familiar result as well - at high temperatures, we're expecting $U = Nk_B T$, so it makes sense for $C = Nk_B$. Remember also that because we have a gap between the ground state and next lowest energy level, we have **gapped behavior** where C becomes exponentially small at low temperature T.

The distinguishing factor here from a two-level system is that there are always higher and higher energy levels!

40.3 Deriving the ideal gas law again

Let's try to do this derivation without needing to count states!

We have our N indistinguishable particles, so we found last time that the partition function (by overcounting arguments) is

$$Z = \frac{1}{N!}(Z_1)^N.$$

To calculate Z_1, remember that we need to do our density of states argument: by the semi-classical density of states argument,

$$Z_1 = \sum_j e^{E_j/(k_B T)} \to \int \frac{d^3x d^3p}{(2\pi\hbar)^3} e^{-p^2/(2mk_B T)}.$$

Evaluating this integral, integrating out the d^3x gives us a volume, and writing $d^3p = 4\pi p^2 dp$ by spherical symmetry,

$$Z_1 = V\frac{4\pi}{(2\pi\hbar)^3}\int_0^\infty dp p^2 e^{-p^2/(2mk_B T)}.$$

Using the change of variables $y^2 = \frac{p^2}{2mk_B T}$, the integral now becomes

$$Z_1 = V4\pi\left(\frac{2mk_B T}{4\pi^2\hbar^2}\right)^{3/2}\int_0^\infty dy y^2 e^{-y^2}$$

and the known integral has value $\frac{\sqrt{\pi}}{4}$. This gives us a final partition function for one particle of

$$Z_1 = \left(\frac{mk_B T}{2\pi^2\hbar^2}\right)^{3/2} V.$$

Notice that the units of the partition function are dimensionless! That means that the $\frac{mk_B T}{2\pi\hbar^2}$ term should have units of $\frac{1}{\text{length}^2}$, and this helps us define a new length scale

$$\lambda_D = \sqrt{\frac{2\pi\hbar^2}{mk_B T}} = \frac{2\pi\hbar}{\sqrt{2\pi m k_B T}} = \frac{h}{\sqrt{2mE_{\text{thermal}}}}.$$

This is called the thermal de Broglie wavelength: how do we compare it to others? Thinking of our ideal gas as noninteracting point particles, we have an average inter-particle distance

$$\left(\frac{V}{N}\right)^{1/3}.$$

The main point is this: if we have $\left(\frac{V}{N}\right)^{1/3} \gg \lambda_D$, we have a classical system, and we can use Maxwell-Boltzmann statistics to evaluate our system. But if $\left(\frac{V}{N}\right)^{1/3} \ll \lambda_D$ (for example, when we start lowering our temperature), quantum mechanical effects are dominant, and we describe the system either with Bose-Einstein statistics or Fermi-Dirac statistics, based on whether we have distinguishability.

Example 216

Consider an electron at room temperature. Then we have

$$\lambda_D = \frac{h}{\sqrt{2\pi m_e k_B T}} \approx 4.5\text{nm},$$

which is pretty small.

This means that in our ideal gas situation, we should be using Maxwell-Boltzmann statistics! So now we have

$$Z_1 = \frac{V}{\lambda_D^3} \implies Z = \frac{Z_1^N}{N!},$$

and now let's try to calculate our thermodynamic quantities. By Stirling's approximation,

$$U = -\frac{\partial}{\partial \beta} \ln Z = -\frac{\partial}{\partial \beta}(N \ln Z_1 - N \ln N + N),$$

and if we work this out, we'll derive the well-known

$$U = \frac{3}{2}\frac{N}{\beta} = \frac{3}{2} N k_B T.$$

Calculating the entropy for this system,

$$S = \frac{U}{T} + k_B \ln Z = \frac{3}{2} N k_B + k_B \ln\left(\frac{Z_1^N}{N!}\right).$$

If we use Stirling's approximation again, we'll find that

$$S = N k_B \left(\frac{3}{2} + \ln Z_1 - \ln N + 1\right),$$

and we can substitute in to find

$$S = N k_B \left(\frac{3}{2} \ln\left(\frac{m k_B T}{2\pi\hbar^2}\right) + \ln\frac{V}{N} + \frac{5}{2}\right)$$

which is the familiar Sackur-Tetrode equation!

40.4 The Maxwell-Boltzmann equation

Finally, we want to look at the probability distribution for the kinetic energy of a specific molecule in an ideal gas. Using the semi-classical limit argument again, now that we know the partition function Z_1,

$$p(E) = \frac{1}{Z_1} \int \frac{d^3x d^3p}{(2\pi\hbar)^3} e^{-p^2/(2mk_BT)} \delta\left(E - \frac{p^2}{2m}\right),$$

since we want to pick out those energies that are equal to $\frac{p^2}{2m}$ specifically. Making the same simplifications and plugging in Z_1,

$$= \left(\frac{2\pi\hbar^2}{mk_BT}\right)^{3/2} \frac{1}{V}\frac{1}{(2\pi\hbar)^3}V \cdot 4\pi \int_0^\infty dp p^2 e^{-p^2/(2mk_BT)}\delta\left(E - \frac{p^2}{2m}\right)$$
$$= \left(\frac{2\pi}{mk_BT}\right)^{3/2} \frac{1}{2\pi^2} 2mEe^{-E/k_BT}\left(\frac{m}{|p|}\right).$$

Now since we can write $p = \sqrt{\frac{m}{2E}}$ in terms of kinetic energy, this all simplifies to

$$\boxed{p(E) = 2\pi\left(\frac{1}{\pi k_BT}\right)^{3/2}\sqrt{E}e^{-E/(k_BT)}}.$$

The $\sqrt{E}$ factor essentially tells us about the momentum: the density of states is proportional to $\sqrt{E}$. This is known as the **Maxwell-Boltzmann distribution** for the kinetic energy of a molecule in an ideal gas situation, and it works whenever we have $\left(\frac{V}{N}\right)^{1/3} \gg \lambda_D$.

We can find the average energy here:

$$\langle E\rangle = \int Ep(E)dE = \frac{3}{2}k_BT = \langle\frac{1}{2}mv^2\rangle,$$

which gives us the root-mean-square of the velocity:

$$v_{\text{rms}} = \sqrt{\langle v^2\rangle} = \sqrt{\frac{3k_BT}{m}}.$$

This also gives us the velocity distribution

$$p(v) = \int_0^\infty dEp(E)\delta\left(v - \sqrt{\frac{2E}{m}}\right)$$

and through some routine algebra, we get the Maxwell-Boltzmann velocity

$$\boxed{p(v) = 4\pi\left(\frac{m}{2\pi k_BT}\right)^{3/2} v^2e^{-mv^2/(2k_BT)}}.$$

Next time, we'll start talking about black-body radiation. Look at the problem set to examine some of these systems more carefully!

41 May 1, 2019 (Recitation)

41.1 General comments

Let's start with some general remarks about the partition function. We're in a situation where we know about states, entropy, and the general behavior of nature: now, we'll try to tie that in to the canonical ensemble using our function

$$Z = \sum_i e^{-\beta E_i}.$$

This was derived in class: in particular, the probability of a state i is $\frac{e^{-\beta E_i}}{Z}$, and then the average value of E is

$$\langle E \rangle = \sum_i E_i \frac{e^{-\beta E_i}}{Z}$$

which is connected to a β-derivative. We also know that $S = k_B \sum p_i \ln p_i$ can be written in terms of the derivatives, as can free energy and all our other variables! So it basically has all the knowledge we need in our system.

So in relation to the problem set, we always "sum over states," even if some of them may look different from others!

What makes the partition function easy? The idea is that having N indistinguishable particles gives a partition function

$$Z = \frac{1}{N!} Z_1^N \implies \ln Z = N \ln Z_1 - \ln N!,$$

and it's often much easier to calculate Z_1 for one particle. This wasn't true in the microcanonical ensemble: since we had a fixed total U, we didn't have independence between the energy states of our particles there, so our calculations were more complicated! We don't have to make the ugly approximation for the $3N$-dimensional sphere's surface area, as we did when deriving Sackur-Tetrode with the microcanonical ensemble.

41.2 Particles in a potential

There's a lot of different ways we can work with a particle with a potential and kinetic energy. The potential energy can be constant (a box), linear (for example, due to gravity on Earth), or quadratic (a harmonic oscillator) in terms of x, and the kinetic energy can be quadratic or linear in terms of p. Finally, we can pick how many degrees of freedom d we have for our particle.

The idea is that many problems are just some combinations of these parameters! We're going to generalize so that we can see the big picture.

First of all, if our energy $H(\vec{x}, \vec{p})$ can be written in terms of these phase variables, our equation for the partition function Z_1 (for one particle) is just (semi-classically)

$$Z = \int \frac{d^3x d^3p}{h^3} e^{-\beta H(\vec{p}, \vec{x})}.$$

The first term's 3 can be replaced with the number of degrees of freedom we have. So now we just insert our expressions for kinetic and potential energy into H, but let's not integrate yet. What are we trying to work with? For example, if we're trying to find the internal energy, we don't need Z: instead, we want

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

So all the normalization factors in Z can be ignored if all we want is U: all of the Gaussian integrals give multiplicative factors, and $\ln Z$ turns those into constants, which become zero under β. Given this, let's rederive the relevant part of this calculation for the ideal gas. The d^3x integration gives a volume, and we don't care about the h^3 either, so we're left with

$$Z_1 \propto \int_{-\infty}^{\infty} d^3p e^{-\beta(p_x^2+p_y^2+p_z^2)/(2m)}.$$

This is essentially just three independent integrals! To work with this, let's define a new variable $\xi^2 = \frac{\beta p_x^2}{2m}$: then $d\xi = \beta dp_x \cdot C$ - here we only care about the β-dependence, since we take the derivative with respect to β later on.

This gives us a scale factor of $\beta^{-1/2}$ in each direction! So we're left with

$$Z_1 = \left[\beta^{-1/2} \int d\xi e^{-\xi^2}\right]^3 = \beta^{-3/2} \cdot C,$$

where C has the volume, the h^3 term, and so on - those aren't relevant right now! Now $\ln Z$ is some constant plus $-\frac{3}{2} \ln \beta$, and now

$$U = -\frac{\partial}{\partial \beta}\left(-\frac{3}{2} \ln \beta\right) = \frac{3}{2\beta} = \frac{3}{2} k_B T,$$

as we expect! (This also gives us things like $C_V = \frac{3}{2} k_B$ for free.)

Fact 217

If we need something like the entropy, then we do need the constants we've tossed out along the way, but the internal energy doesn't depend on them!

So we've now looked at one example of a potential: let's now think about the relativistic gas, where the kinetic energy is $c|p|$. Most of our derivation stays the same, except that now

$$Z_1 = \int_{-\infty}^{\infty} d^3 p e^{-\beta c|p|/2m}.$$

Defining $\vec{\xi} = -\beta c \vec{p}$, this converts us to an integral proportional to

$$\beta^{-3} \int d^3 \xi e^{-|\xi|}$$

(with the point being that we want to convert to generic, normalized integrals which just give us fixed constants!). Then almost everything stays the same: we just have $U = 3k_B T$ and $C_V = 3k_B$ now.

In both of these cases, we've found that our partition function

$$Z_1 = (c)\beta^{-x},$$

where $x = \frac{d}{2}$ is half the number of degrees of freedom for the quadratic kinetic energy and $x = d$ is the number of degrees of freedom for the linear kinetic energy.

But we've only been dealing with cases where our potential is constant: what if we have a linear or quadratic potential? Well, the situation is exactly symmetric! We can make exactly the same arguments and substitutions, so in general, we'll actually have

$$Z_1 = (c)\beta^{-x+y},$$

where $y = \frac{1}{2} d$ if we have a harmonic oscillator (quadratic) potential and d if we have a linear potential!

Fact 218

The whole point is that this is why statistics is useful: we can just look at how everything scales with β. This gives us (at least in the semi-classical limit) the internal energy and many other quantities.

So now if we want the specific heat for a relativistic gas in two dimensions which lives in a potential

$$V(\vec{x}) = \alpha |x|^{10},$$

we know that the relativistic gas integration gives us (since $d = 2$),

$$x + y = 1 \cdot 2 + \frac{1}{10} \cdot 2 = \frac{11}{5}.$$

So the partition function is just going to be proportional to $\beta^{-11/5}$. That gives us a specific heat of $\frac{11}{5} k_B T$ with very few calculations!

Fact 219

This is a generalization of the equipartition theorem, which tells us that each quadratic term gives a $\beta^{-1/2}$ in the partition theorem, which gives a specific heat of $\frac{1}{2} k_B$.

42 May 2, 2019

In terms of exam regrades, everything will be updated later today. The problem set is due tomorrow!

We're going to discuss black-body radiation today: the main idea is that being able to absorb light at all frequencies leads to interesting behavior in terms of the spectrum of colors. In particular, we'll start understanding ideas like why we humans can see in the visible light range! (In particular, things would be very different if the sun was at a different temperature.)

Looking ahead, we'll look at photons as particles of a gas, and we'll consider why the classical model of this gas isn't good enough to understand that spectrum, particularly in the UV range! This was part of the birth of quantum mechanics, and it'll let us actually plot the energy density with respect to our frequency ω correctly.

42.1 Black-body radiation

As a reminder, quantum mechanical particles can be described by a wavefunction, and the energy eigenstates (for a particle in a box) are given by

$$\psi_k = \frac{1}{\sqrt{V}} e^{i\vec{k}\vec{x}}.$$

In three dimensions, we have $\vec{k} = (k_1, k_2, k_3)$, where each $k_i = \frac{2\pi n_i}{L}$ is in terms of the **mode numbers** $n_i \in \mathbb{Z}$. It turns out this can also describe the quanta of the electromagnetic field! Each mode can be thought of associated with a quantum harmonic oscillator, and we need all plane waves to travel at the speed of light, so we have $\omega_k = c|k|$. (Looking at the polarization, there are two modes per wavevector $\vec{k}$.)

So we'll treat our black-body system as a gas of photons, and let's see what happens as our temperature T increases! We'll start by essentially counting states: how many states are there available to a single photon whose energy is between E and $E + dE$? We'll calculate this in terms of frequency:

$$\frac{dN}{d\omega} d\omega = (4\pi k^2) \left(\frac{dk}{d\omega} d\omega \right) \left(\frac{2V}{(2\pi)^3} \right),$$

where $k^2 = \frac{\omega^2}{c^2}$ and $\frac{dk}{d\omega} = \frac{1}{c}$. What do each of these terms mean? The first term $4\pi k^2$ is the surface area of a sphere of radius k, the second term is the thickness of a spherical shell for $d\omega$, and the last term is the density of a single particle plane-wave eigenstate wavevector (where the factor of 2 comes from the fact that there are 2 directions of polarization). So now simplifying,

$$\frac{dN}{d\omega} d\omega = \frac{V\omega^2}{\pi^2 c^2} d\omega,$$

We want to calculate the partition function, so let's look at a particular frequency ω for our photons and then integrate later. This is essentially the partition function for a harmonic oscillator, but we'll disregard the zero point energy: it's a constant, and in practice, we only care about energy differences (unless we're talking about dark energy and so on). This means that our partition function simplifies to

$$Z_\omega = 1 + e^{-\beta\hbar\omega} + e^{-2\beta\hbar\omega} + \cdots = \frac{1}{1 - e^{-\beta\hbar\omega}}.$$

We want to sum this over all possible frequencies - theoretically, we don't have any upper limit (though in solids, we do have some physical limit due to the properties). So now we can integrate out

$$\ln Z = \int_0^\infty d\omega \frac{dN}{d\omega} \ln Z_\omega,$$

which simplifies to

$$\ln Z = -\frac{V}{\pi^2 c^3} \int_0^\infty d\omega \omega^2 \ln(1 - e^{-\beta\hbar\omega}).$$

We'll leave the partition function like this, but we can already calculate some quantities. First of all, the average internal energy can be found via

$$\langle E \rangle = U = -\frac{\partial}{\partial\beta} \ln Z = \frac{V\hbar^2}{\pi^2 c^3} \int_0^\infty \frac{\omega^3}{e^{\beta\hbar\omega} - 1} d\omega,$$

(since we can swap the integral and derivative). In differential form, this says that

$$E(\omega) d\omega = \frac{V\hbar^2}{\pi^2 c^3} \frac{\omega^3}{e^{\beta\hbar\omega} - 1} d\omega,$$

which is the **amount of energy carried by photons with frequency between ω and $\omega + d\omega$.** This is actually **Planck's distribution**! If we plot the integrand with respect to ω, this has a single peak: at low frequency, the distribution is consistent with the equipartition theorem, but then the energy gaps at higher frequencies bring I back to 0 instead of going off to infinity.

Fact 220

Calculating $\frac{dE(\omega)}{d\omega}$ gives Wein's radiation law, which is a good exercise.

By the way, when the temperature is around 6000 Kelvin, which is the temperature of the Sun, the peak is mostly in the visible light range, which makes sense!

Fact 221

This is known as a "radiation gas," which has different properties from the normal gasses that we've been dealing with.

42.2 Doing some derivations

Looking again at our internal energy integral, we have

$$U = \int_0^\infty d\omega \left(\frac{\omega^2 V}{\pi^2 c^3}\right) \hbar\omega \left(\frac{1}{e^{\beta\hbar\omega} - 1}\right).$$

Here, the first term $\left(\frac{\omega^2 V}{\pi^2 c^3}\right)$ is the density of modes per frequency. The $\hbar\omega$ is a kind of energy unit excitation, and combined with the last term, this gives the mean thermal excitation energy per mode. Substituting in $x = \beta\hbar\omega$, we actually want to calculate

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

That last integral is actually known because there is no upper limit on the frequency - otherwise, it's only possible to numerically approximate it! It turns out that this is

$$= \Gamma(4)\zeta(4) = 6 \cdot \frac{\pi^4}{90} = \frac{\pi^4}{15},$$

and therefore we can think about the **energy density of our gas of photons**

$$\frac{U}{V} = \xi = \frac{\pi^2 k_B^4}{15\hbar^3 c^3} T^4.$$

In other words, we can just write down this law as

$$\boxed{\frac{U}{V} \propto T^4}.$$

Example 222

So we can imagine having a box of photons with a hole: how can we find the energy flux?

Flux is the rate of energy transfer, and it will just be $\frac{\xi c}{4}$, where c is the speed of light: the $\frac{1}{4}$ factor comes from us not having a point source, and where the size of the hole is larger than the wavelength of the photons. This can then be written as $\boxed{\sigma T^4}$, and this proportionality is called the **Stefan-Boltzmann law.** Here,

$$\sigma = \frac{\pi^2 k_B^4}{60\hbar^3 c^2}.$$

Example 223

Let's see if we can find the pressure of this gas. (It will be interesting, because the pressure only depends on the temperature!)

First of all, let's calculate the Helmholtz free energy

$$F = -k_B T \ln Z = \frac{V k_B T}{\pi^2 c^3} \int_0^\infty d\omega \omega^2 \ln(1 - e^{-\beta\hbar\omega}).$$

Again letting $x = \beta\hbar\omega$ and doing similar simplifications,

$$F = \frac{V (k_B T)^4}{\pi c^3 \hbar^3} \int_0^\infty dx x^2 \ln(1 - e^{-x}),$$

and now by integration by parts, the integral simplifies to

$$\int_0^\infty dx \left(\frac{1}{3}\frac{d}{dx} x^3\right) \ln(1 - e^{-x}) = \frac{1}{3} x^3 \ln(1 - e^{-x}\Big|_0^\infty - \frac{1}{3}\int_0^\infty dx \frac{x^3}{e^x - 1}.$$

We've seen the second term before: it evaluates to $\frac{\pi^4}{45}$, and therefore

$$F = -\frac{\pi^2 (k_B T)^4 V}{45\hbar^3 c^3} = -\frac{1}{3} U.$$

Now we can find our pressure:

$$P = -\left.\frac{\partial F}{\partial V}\right|_T = -\frac{U}{3V} = \frac{4\sigma}{3c}T^4.$$

This equation of state tells us that interestingly, this gas's pressure only depends on the temperature! In particular, the pressure P is $\frac{1}{3}$ of the energy density, and this is an interesting fact in cosmology.

42.3 Continuing on

Let's try calculating some other thermodynamic quantities: the entropy of our gas is

$$S = -\left.\frac{\partial F}{\partial T}\right|_V = \frac{16V\sigma}{3c}T^3,$$

the specific heat

$$C_V = \left.\frac{\partial U}{\partial T}\right|_V = \frac{16V\sigma}{c}T^3,$$

and to find the number of photons,

$$\frac{dN}{d\omega dV} = \frac{\omega^2}{\pi^2 c^3}\frac{1}{e^{\beta\hbar\omega}-1},$$

and integrating out V and ω, we find that

$$N = \frac{V}{\pi^2 c^3}\int_0^\infty d\omega \frac{\omega^2}{e^{\beta\hbar\omega}-1} \approx 1.48\frac{\sigma T^3 V}{k_B c}.$$

We can also now calculate the fluctuations in energy: normalizing the standard deviation with respect to U,

$$\frac{\sigma(E)}{U} = \frac{k_B T}{\frac{4}{c}\sigma T^4 V}\sqrt{\frac{16\sigma T^3 V}{k_B c}} = \sqrt{\frac{k_B c}{6T^3 V}} = \sqrt{\frac{1.48}{N}}.$$

These fluctuations are very small as N grows large! So we can basically treat this from the microcanonical point of view, and the thermodynamic behavior would look basically the same.

Question 224. *So what does all of this mean?*

First of all, this shows that classical mechanics isn't enough to describe our systems. If we take $\hbar \to 0$ - that is, we don't assume that we have quantization - then

$$\frac{d^2U}{d\omega dV} = \frac{\hbar\omega^3}{\pi^2 c^3}\frac{1}{e^{\hbar\omega\beta - 1}} \to \frac{\omega^2}{\pi^2 c^3}k_B T$$

as $\hbar \to 0$. This quadratic growth is consistent for $\hbar\omega \ll k_B T$ with the equipartition theorem, but then quantum mechanical effects give an exponential decay eventually, which classical mechanics can't predict. Also, this gas is very interesting, because local properties only depend on temperature. $\frac{U}{V}$, $\frac{N}{V}$, $\frac{S}{V}$, and P are all dependent on T and not V!

Finally, this was a gas of photons, but we can also think about vibrations of solids in terms of phonons instead. The same derivations can be made, but we just have some different quantities: c is now the speed of sound (since we care about sound modes), and also, we have a bounded integral instead of integrating from 0 to ∞. Basically, the wavelength can't be smaller than the inter-atom distance! That makes it interesting to think about how C_V of a solid changes as we decrease the temperature - this will lead us to Einstein's law.

An exercise: what is the chemical potential μ for this radiation gas? It will turn out to be zero, and we'll see a connection when we discuss bosons later on!

43 May 6, 2019 (Recitation)

43.1 Homework questions

In one of the problems of the problem set, we have particles oriented with a dipole moment $\vec{m}u$ which gives us an energy

$$-\vec{E} \cdot \vec{\mu} = -|E||\mu| \cos \theta.$$

We want to find the partition function, which is an integral over states:

$$= \int \frac{\text{"} dxdp \text{"}}{h} e^{-\beta H},$$

where H, the Hamiltonian, is the kinetic plus potential energy that we have if we use the semi-classical approximation. The main differences here are that we have two-dimensional space (and therefore two-dimensional momentum space as well), so we really want

$$Z_1 = \int \frac{d^2x d^2p}{h^2} e^{-\beta H}.$$

What's special is that we are integrating over angles θ and π, and we also have the canonical momenta p_θ and p_ϕ as a result.

But now we can integrate our $\cos \theta$ out through our polar coordinates, and everything works out! (We will get a $\cos \theta$ in the exponent.)

In a different problem, when we give a gravitational potential energy $-mgh$ to each particle, that just adds a $e^{-\beta mgh}$ to each term. This is generally solved more easily with the grand canonical ensemble (where we are actually allowed to exchange particles), but if we are to solve the problem in the way that we know, the $e^{-\beta mgh}$ factor comes out of the z-integral. Now we just get an additive constant in F (because we take $\ln Z$) and the rest of the problem looks like an ideal gas!

Finally, considering the polymer chain, the main concept is to treat each individual monomer separately! Since we have a classical system, we can just compute Z_1 for one monomer and take the Nth power. We basically have a bunch of two-level systems, for which the partition function is easy to calculate!

Now all the degeneracies of the form $\binom{N}{n}$ are accounted for in the binomial expansion of the two-level system $1 + e^{-\beta \Delta}$, raised to the Nth power. We don't even need to do the combinatorics from the microcanonical ensemble!

43.2 Clicker questions

Question 225. *Canonical versus microcanonical ensembles: do both ensembles have the same properties?*

The answer is yes! We'll unfold more details about this soon.

Question 226. *Do both ensembles have the same number of microstates?*

No! In the microcanonical ensemble, we're given a total energy E for which all our states must reside, but in the canonical ensemble, we're given more freedom. Here's a way to think about it: let's say we divide a system into two halves with $\frac{N}{2}$ particles. Then the total number of microstates is

$$\Gamma_{\text{total}} = \Gamma_{\text{Left}} \Gamma_{\text{Right}},$$

but this isn't quite correct in general because we might not have exactly $\frac{N}{2}$ on each side! (Think of "fixing the number of particles on the left side" as "fixing the energy of our system" as in the microcanonical ensemble.) It turns out in general that the right order of magnitude is to deal with the Gaussian fluctuation

$$\Gamma_{2V} = (\Gamma_V)^2 \sqrt{N}.$$

In this case, it is easy to count microstates, but it may be more difficult in other versions! Well, the reason we don't worry about it too much is because taking log often makes the $\sqrt{N}$ negligible. We really only care about those microstates with non-negligible probability, so the Binomial distribution with mean $\frac{N}{2}$ and standard deviation on the order of $\sqrt{N}$ can be viewed as basically Gaussian, and we only care about those values within a few standard deviations.

So what is the number of microstates in a microcanonical distribution? We often have a density of states $\frac{\partial N(E)}{\partial E}$, but that's a classical description and therefore can't be used directly for counting. So we need to have some energy width δE, and we often didn't specify what that was!

Fact 227

Usually we don't need to because the δE cancels out later, or because it only adds a negligible logarithmic term.

There was one example where it didn't need to be specified: given a harmonic oscillator, there are only specific states at width $\hbar\omega$ apart, so perfect, identical harmonic oscillators do actually have some "exact number of states." So in our equation

$$\Gamma(E) = \frac{\partial N(E)}{\partial E} \delta E,$$

we can think of "$\delta E = \hbar\omega$" as an effective width between our energies.

Question 228. *So what's the energy gap that we've assumed in the canonical ensemble?*

We have a Boltzmann factor $e^{-\beta E}$ that exponentially decreases with E, and we have a density of states that increases very rapidly. Putting these together, we get a sharp peak at the average energy $\langle E \rangle$, which is why we say that the microcanonical and canonical ensembles are equivalent for lots of purposes. So how broad is the distribution?

Well, we calculated this in class: we can find $\langle E \rangle$ and $\langle E^2 \rangle$, which allow us to find

$$\sigma_E = k_B T \sqrt{\frac{C}{k_B}}$$

where C is the specific heat. Thinking about this for an ideal gas: $C \propto N k_B$, so this is proportional to $\sqrt{N}$. This is the same factor that's been popping up in all the other situations, and that's because that's the right factor for fluctuations! Now it's the same idea: $\sqrt{N}$ is small compared to N, so it can be neglected.

44 May 7, 2019

Some preliminary announcements: homework will be graded soon, and we should check the website to make sure everything is accurate. The last graded problem set is due on Friday! The next one (on photons, phonons, gases) is not graded but is good practice for the final. Finally, some past finals will be posted soon - there will be more problems on the final because it is longer (3 hours), but some fraction will be taken from past exams.

44.1 Overview

Last time, we went over black-body radiation by thinking of photons as quanta of the electromagnetic field. We find that raising the temperature of a black-body gives a spectrum of radiation, and we can figure out the thermodynamic quantities of this body using the canonical ensemble!

We're essentially going to go over the physics of solids today, examining quantities like the heat capacity with different models. We'll also consider what happens when we look at the extremes of temperature: for example, what if the de Broglie wavelength is comparable in size to the inter-atom distance? That's where the grand canonical ensemble comes in!

44.2 Phonons

Our goal, ultimately, is to consider the heat capacity C of a solid. A phonon is essentially a quanta of vibrational mechanical energy!

Fact 229

We can discretize sound waves in solids using these quasi-particles. The energy of a phonon is

$$E = \hbar\omega = \hbar k c_s,$$

where c_s is the speed of sound.

We'll be working with a perfect crystal of atoms as our system for convenience. The first thing we want to do is to consider the density of states of our phonons

$$\frac{dN}{d\omega} d\omega = \frac{3V}{2\pi^2 c_s^2} \omega^2 d\omega.$$

where the 3 in the numerator comes from having a multiplicity of polarization.

Fact 230

The factor of 3 is really from having one longitudinal mode (compression) and two transverse modes (shear). Since we have N atoms in our crystal, this is $3N$ different normal modes, meaning that there are $3N$ different types of phonons with frequency $\omega_1, \cdots, \omega_{3N}$.

As a result of this, we have another difference between phonons and photons: the frequency spectrum for light waves is unbounded (it can go arbitrarily high), but the sound waves have a minimum wavelength

$$\lambda = \frac{2\pi c_s}{\omega},$$

where λ is the spacing between atoms (since something needs to be able to propagate the wave). This means we have a **maximum frequency** corresponding to our $\lambda_D \sim \sqrt[3]{\frac{V}{N}}$, meaning that our maximum frequency

$$\omega_D \sim \left(\frac{N}{V}\right)^{1/3} c_s.$$

To find the proportionality constant in front, note that we can count single-phonon states as

$$\int_0^{\omega_D} d\omega \frac{dN}{d\omega} = \frac{V\omega_D^3}{2\pi^2 c_s^2}$$

(by direct integration). One way to deal with this (with solid-state physics) is to deal with a "primitive cell," but instead we'll argue that this number is just $3N$, the number of degrees of freedom. Thus we can find our maximum allowed frequency:

$$3N = \frac{V\omega_D^3}{2\pi^2 c_s^3} \implies \boxed{\omega_d = \left(\frac{6\pi^2 N}{V}\right)^{1/3} c_s}.$$

We'll associate ω_D with a new temperature scale now: define the **Debye temperature**

$$T_D = \frac{\hbar\omega}{k_B}.$$

44.3 Using the canonical ensemble

We'll now calculate the partition function for our system: for a fixed frequency,

$$Z_\omega = 1 + e^{-\beta\hbar\omega} + e^{-2\beta\hbar\omega} + \cdots = \frac{1}{1 - e^{-\beta\hbar\omega}}.$$

Assuming the frequencies are independent of each other, the partition function is just the product over all ω

$$Z = \prod_\omega Z_\omega,$$

and now taking the log, we can approximate the sum as an integral

$$Z = \int_0^{\omega_D} d\omega \frac{dN}{d\omega} \ln Z_\omega.$$

(Note that we now have an upper frequency ω_D instead of ∞.) Now the energy of our system is

$$E = \int_0^\infty d\omega \frac{dN}{d\omega} \frac{\hbar\omega}{e^{\beta\hbar\omega - 1}},$$

which simplifies to

$$= \frac{3V\hbar}{2\pi c_s^3} \int_0^{\omega_D} d\omega \frac{\omega^3}{e^{\beta\hbar\omega} - 1}.$$

There isn't an analytic expression for this anymore, so we'll instead just look at the low- and high-temperature limits. Letting $x = \beta\hbar\omega$, we can rewrite our energy as

$$E = \frac{3V}{2\pi^2(\hbar c_s)^3}(k_B T)^4 \int_0^{T_D/T} dx \frac{x^3}{e^x - 1}.$$

Example 231

One extreme is where $T \ll T_D$, so our upper limit goes to ∞. Then we've seen our integral before: it evaluates to $\frac{\pi^4}{15}$.

So in this case, we can calculate

$$C_V = \frac{\partial E}{\partial T} = \frac{2\pi^2 V k_B^4}{5\hbar^3 c_s^3} T^3 = N k_B \frac{12\pi^4}{5}\left(\frac{T}{T_D}\right)^3.$$

In other words, at low temperatures, our heat capacity is proportional to T^3. This does line up with experimental observations! In particular, $C_V \to 0$ as $T \to 0$., which is consistent with our third law of thermodynamics.

Fact 232

Einstein has a different description of this behavior with a different kind of calculation - Debye just extended that model by allowing many different frequencies.

Example 233

On the other hand, can we recover the results we expect at high (classical) temperatures $T \gg T_D$?

Then we integrate over very small values of x, so we can do a Taylor expansion. In these cases,

$$\int_0^{T_D/T} dx \frac{x^3}{e^x - 1} = \int_0^{T_D/T} dx(x^2 + O(x^3)) = \frac{1}{3}\left(\frac{T_D}{T}\right)^3 + \cdots,$$

and now we find that our heat capacity reaches a constant

$$C_V = \frac{V k_B^4 T_D^3}{2\pi^2 \hbar^3 c_s^3} = 3Nk_B,$$

which is consistent with the Dulong-Petit law! The data doesn't perfectly fit with this, and that's because we have a flaw in the calculation: our dispersion relation $\omega = kc_s$ isn't quite accurate, but that can be left for a physics of solids class.

44.4 Back to the monatomic ideal gas

Remember that when we talked about the ideal gas for a canonical ensemble, we treated the particles quantum mechanically in a box: we put the system at fixed temperature, and we used the partition function to pull out the thermodyanmics of our system. There, we introduced the de Broglie wavelength (or length scale), and we mentioned that we should relate that to the inter-particle distance to see whether or not quantum effects are important.

Well, when we're close to the de Broglie wavelength, we can't use the $Z = \frac{1}{N!} Z_1^N$ formula for our partition function anymore: we need a better way to describe our system. That's where the **grand canonical ensemble** comes in: we relax our constraint on having a fixed number of particles, but we **fix the chemical potential**.

Remember that this length scale

$$\lambda = \sqrt{\frac{2\pi\hbar^2}{mk_B T}}$$

increases as our temperature decreases. Eventually, this becomes comparable to our interparticle spacing $\propto \left(\frac{V}{N}\right)^{1/3}$, which is where quantum effects start to have an effect.

Remember that in a microcanonical ensemble, we fix V, U, N, and in a canonical ensemble, we fix V, T, N. In both of these cases, we can calculate the energy, and the macroscopic properties are basically the same here because the fluctuations are so small.

So in a grand canonical ensemble, we fix V, T, μ instead: it's a pressure-like term that adds particles. Specifically, our internal energy in differential form can be written as

$$dU = TdS - PdV + \mu dN \implies \left.\frac{\partial U}{\partial N}\right|_{S,V} = \mu.$$

It's usually difficult to keep S and V constant, so we instead work with the Gibbs free energy: there, we can instead calculate

$$\left.\frac{\partial G}{\partial N}\right|_{T,P} = \mu,$$

which is generally easier to work with experimentally! We can show that in a system at thermal equilibrium,

$$\left.\frac{\partial S}{\partial N}\right|_{U,V} = -\frac{\mu}{T},$$

so two systems brought together at thermal equilibrium will also have the same chemical potential.

Proposition 234

This means we can think of our system as being connected to a large reservoir, so that both are at some fixed T, μ. Now the total system and reservoir are a microcanonical ensemble!

So we want the probability distribution of a given state s_j for our system: then

$$p(S_j) = \frac{\Gamma_R}{\Gamma_{R\oplus S}},$$

and similar to our canonical ensemble derivation, we can write the Γs in terms of our entropy:

$$\Gamma_R(s_j) = \exp(S_R(S_j)/k_B) \implies p_j(s_j) = \frac{1}{\mathcal{Z}} \exp\left[S_R(U - E_j, M - N_j)/k_B\right]$$

where $\mathcal{Z}$ is our normalization factor, analogous to the partition function. Now

$$\frac{\partial p_j}{\partial E_j} = -\frac{1}{k_B} \left.\frac{\partial S_R}{\partial U}\right|_{U-E_j} p_j = -\frac{1}{k_B T} p_j,$$

and

$$\frac{\partial p_j}{\partial N} = -\frac{1}{k_B} \left.\frac{\partial S}{\partial M}\right|_{M-N_j} p_j = \frac{\mu}{k_B T} p_j.$$

Thus, we can write our probability of any state

$$p_j(S_j) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{E_j}{k_B T} + \frac{\mu N_j}{k_B T}\right) = \frac{1}{\mathcal{Z}} \exp\left[(\mu N_j - E_j)/(k_B T)\right].$$

We can now write our expression for our grand partition function

$$\boxed{\mathcal{Z} = \sum_j \exp\left[(\mu N_j - E_j)/(k_B T)\right]}.$$

44.5 Thermodynamics of the grand canonical ensemble

How can we do physics with this new quantity? Let's look at a similar quantity as in the canonical ensemble

$$\frac{\partial}{\partial \beta} \ln \mathcal{Z} = \frac{1}{\mathcal{Z}} \sum_j (\mu N_j - E_j) e^{(\mu N_j - E_j)\beta},$$

which can be rewritten as

$$= \sum_j p_j(s_j)(\mu N_j - E_j).$$

Thus, we actually have the average value of $\mu N_j - E_j$, which is

$$\frac{\partial}{\partial \beta} \ln \mathcal{Z} = \mu\langle N\rangle - \langle E\rangle = \mu N - U.$$

We can also consider

$$\frac{\partial}{\partial \mu} \ln \mathcal{Z} = \frac{1}{\mathcal{Z}} \sum_j \beta N_j e^{(\mu N_j - E_j)\beta};$$

again, we can extract out our probability term to get

$$= \beta \langle N \rangle = \beta N.$$

This means we can find the expected number of particles N, and in the thermodynamic limit, fluctuations are much smaller (in order of magnitude) than N, so this is pretty accurate almost all the time!

Is it possible for us to find the entropy of our system? We know that

$$S = -k_B \sum p_j \ln p_j = -k_B \sum_j \frac{e^{\beta(\mu N_j - E_j)}}{\mathcal{Z}} \cdot \left(\beta(\mu N_j - E_j) - \ln \mathcal{Z}\right),$$

which can be arranged as

$$S = -\frac{1}{T}(\mu N - U) + k_B \ln \mathcal{Z} \implies \boxed{U - TS - \mu N = k_B T \ln \mathcal{Z}}.$$

So the central theme is that $\ln \mathcal{Z}$ basically gives us an energy term! This lets us create a new free energy

$$\boxed{\Omega = U - TS - \mu N} = -k_B T \ln \mathcal{Z},$$

which is the **grand potential**: this is essentially a sum over all the different states!

In differential form, the grand potential can be written out as

$$d\Omega = dU - TdS - SdT - \mu dN - Nd\mu,$$

and writing out $dU = TdS - PdV + \mu dN$, this simplifies as

$$d\Omega = -SdT - PdV - Nd\mu.$$

This means we can take derivatives again: we can find our entropy, pressure, and number of particles via

$$\left.\frac{\partial \Omega}{\partial T}\right|_{V,\mu} = -S, \left.\frac{\partial \Omega}{\partial V}\right|_{T,\mu} = -P, \left.\frac{\partial \Omega}{\partial \mu}\right|_{T,V} = -N.$$

Next time, we'll look at bosons and fermions, and we'll try to recover the Boltzmann statistics at high temperatures!

45 May 8, 2019 (Recitation)

45.1 More on density of states

Let's start by summarizing some concepts from last recitation. We have a density of states $\frac{dN}{dE}$ (which usually goes as E^{N_A} if we have a mole of atoms), which is very steep. On the other hand, we have the Boltzmann factor $e^{-E/(k_B T)}$, which usually goes as e^{-N}. Multiplying together these rapidly growing and decaying distributions, we get a very sharp peak!

Almost all of the action occurs around that sharp peak, because the probability of being found far away from $\langle E \rangle$ is very small. More quantitatively, note that the distribution is a delta function at $\langle E \rangle$ for a microcanonical ensemble, since we fix the energy. Really, the only way to do this is to pick a bunch of harmonic oscillators with quantized energy states, but there are still some perturbations there - in other words, our discrete energy states are broadened a bit.

Fact 235

To account for this, we can think of the microcanonical ensemble as having an energy width of $\hbar\omega$ if our energy states of the oscillators differ by $\hbar\omega$.

So when we derived the Sackur-Tetrode equation, remember that we used quantized box numbers n_i such that the sum of the squares of those quantities is fixed: this ended up being the surface area of a sphere, which helped us find the multiplicity of a given energy. Specifically, if we say that

$$\sum_i n_{x_i}^2 + n_{y_i}^2 + n_{z_i}^2$$

is constant, we can integrate to find the number of states: since we have a microcanonical ensemble, we pick out a specific energy U to find that

$$\Gamma \propto \int d^{3N} n_i \cdot \delta\left(\sum n_i^2 - U\right).$$

But there's one thing we did not do: we never specified that we use a spherical shell with some energy width δE. Surface areas are not volumes, so we have a density of states instead of a number of states - how can we introduce an energy uncertainty?

Well, remember that we replaced r^{3N-1} with r^{3N} due to $N \gg 1$: this means that we're essentially allowing for the whole volume instead of the surface area, so our energy width is actually from 0 to our energy E. This also tells us that in an N-dimensional sphere, almost all of the volume is concentrated towards the surface area: in fact, half the volume is concentrated within an $O\left(\frac{r}{N}\right)$ distance from the surface area! That's why it's okay for us to also count the volume of states inside: it's negligible compared to that part near the surface.

Fact 236

So going back to the canonical ensemble, remember that we can find the uncertainty

$$\sigma_E = k_B T \sqrt{\frac{C_V}{k_B}} \propto \sqrt{N} \propto \frac{E}{\sqrt{N}}.$$

Thus, the width of the energy peak is proportional to $\sqrt{N}$, which is precise (proportionally) up to experimental accuracy! But when we made the approximation from r^{3N-1} to r^{3N}, we actually get a fluctuation on the order of $\frac{E}{N}$.

So it's often okay to make what looks like a crude approximation!

45.2 Quick note about the different distributions

Fact 237

As a result, we almost never use the microcanonical ensemble. An awkward situation is that if we divide a system in two, a microcanonical ensemble doesn't stay microcanonical (because we can exchange energy between the two halves)! On the other hand, the partition function is simple: we can just multiply $Z = Z_1 \cdot Z_2$, and that makes many calculations much easier.

On the other hand, allowing particle exchange means we have to use the grand canonical distribution. This helps us with quantum mechanical descriptions of particles - in particular, because of quantum indistinguishability, we no longer have independent particles, but we do have independent states, and we can take our partition function and factor it

$$Z = \prod Z_i$$

over states! We'll probably go over this a bit more in class soon, but it's interesting that non-interacting particles still affect independence for each other. (For example, we have the Pauli exclusion principle for fermions!)

We're going to find that some derivations are easier for the grand canonical ensemble than in the microcanonical or canonical ensemble as well. Le' L

45.3 One last conceptual question

We know that we have a Boltzmann factor

$$e^{-E/(k_B T)}$$

in our probability distribution, so any energy state, no matter how huge, has some positive probability of happening. Is that an artifact of the approximations we have made, or is it physical?

One assumption we make is that our reservoir (that our system is connected with) is large enough to sustain the fluctuations of energy! Since we treat the reservoir and system as a microcanonical ensemble with some total energy U, we can't actually have energies $E > U$.

But other than that, this is indeed how nature works! Let's assume that there is indeed some point where the energy can't get any larger. and then we have some small probability p of getting to an energy above that. Then the entropy and energy change are approximately

$$\Delta = -k_B p \ln p, \Delta E = pE \implies \Delta F = \Delta E - T\Delta S = p(E - Tk_B \ln p)$$

So transferring some probability to a higher state changes the free energy F: $\Delta F < 0$ if $p < e^{-E/(k_B T)}$, and that means that it is always favorable to allow that energy state! This means the canonical ensemble probability distribution is indeed a stable equilibrium, and any collision between atoms (or other perturbation) of the system moves us towards that exact distribution.

46 May 9, 2019

46.1 Overview

Today, we're going to continue talking about the grand canonical ensemble, which is a useful framework for trying to understand quantum effects in different kinds of gases! As a quick reminder, this is a system where instead of fixing energy or temperature (as in the microcanonical and canonical ensembles) along with the number of particles, we have to make a more careful argument. At the quantum level, we can no longer assume that all of our particles are likely to be in different states, so we can't have a simple overcounting term like $\frac{1}{N!}$. In addition, we now allow our system to be open.

The number of particles can now fluctuate, but we'll find that in the thermodynamic limit again, the fluctuations are small. So this will also give similar results to the other formulations!

We'll see some systems where we can use this model today.

46.2 Types of quantum particles

The main idea is that in relativistic quantum field theory, there are two types of particles.

Definition 238

Bosons are particles that have an intrinsic spin of $0, 1, 2, \cdots$ (in multiples of $\hbar\omega$).

The main concept here is that we have indistinguishable particles: the quantum states are **symmetric** until particle exchange, which can be written as

$$\psi(\vec{r}_1, \vec{r}_2) = \psi(\vec{r}_2, \vec{r}_1).$$

Examples of bosons include photons, gluons, Helium-4, as well as quasi-particles like phonons and plasmons.

Definition 239

On the other hand, **fermions** are particles that have an intrinsic spin of $\frac{1}{2}, \frac{3}{2}, \cdots$ (in multiples of $\hbar\omega$).

Examples here include electrons in metals at low temperature, liquid Helium-3, white dwarfs, and neutron stars. This time, quantum states are **antisymmetric**: we have

$$\psi(\vec{r}_1, \vec{r}_2) = -\psi(\vec{r}_2, \vec{r}_1).$$

Fact 240

Bosons can have infinitely many particles per state, but by the **Pauli exclusion principle**, there can only be 0 or 1 fermion in each quantum state.

We're going to label our states by occupation numbers $\{n_r\}$, often as $|n_1, n_2, \cdots, n_r\rangle$. At high temperatures, they will be the same as in our Boltzmann statistics.

Fact 241

By the way, these are three-dimensional particles: 2-D systems are completely different!

Remember that we have our canonical ensemble partition function

$$Z = \sum_{\{n_r\}} e^{-\beta n_r \varepsilon_r},$$

where we sum over all ways of partitioning N particles into sets $\{n_r\}$, with the constraint $\sum n_r = N$. But we have indistinguishable particles here, and it's hard to account for the fact that we can only have 1 particle in each energy state for fermions.

46.3 Using the grand canonical ensemble

So our first step is to look at an open system: let's fix our chemical potential μ and let N fluctuate! (We'll also put our system in contact with a heat bath at temperature T.)

Then for a given state, we have the variables

$$N_j = \sum n_r, E_j = \sum_r n_r \varepsilon_r$$

which are allowed to vary, and now our grand partition function

$$\mathcal{Z} = \sum_j \exp\left[(\mu N_j - E_j)\beta\right] = \sum_{\{n_r\}} \exp\left[\left(\mu \sum_r n_r - \sum_r n_r \varepsilon_r\right)\beta\right].$$

We can rewrite the sum of the exponents as a product:

$$= \sum_{\{n_r\}} \prod_r \exp\left[(\mu - \varepsilon_r)n_r\beta\right] = \boxed{\prod_r \sum_{\{n_r\}} \exp\left[(\mu - \varepsilon_r)n_r\beta\right]}$$

by swapping the sum and product. **Now our path diverges for bosons and fermions:** in one case, the n_rs can be anything, and in the other, the n_rs must be 0 or 1.

Proposition 242

For fermions, since n_rs are all 0 or 1, we find that

$$\mathcal{Z}^{FD} = \prod_r (1 - e^{(\mu - \varepsilon_r)\beta}).$$

This is referred to as **Fermi-Dirac**.

Proposition 243

Meanwhile, for bosons, since n_rs can b, anything, we have an infinite sum

$$\mathcal{Z}^{BE} = \prod_r \left(\sum_{n=0}^{\infty} e^{(mu - \varepsilon_r)\beta}\right)^n = \prod_r \frac{1}{1 - e^{(\mu - \varepsilon_r)\beta}}.$$

This is referred to as **Bose-Einstein**.

In both cases, we often want to deal with the logarithm of the partition function: then the product becomes a sum, and we have the following expressions:

$$\ln \mathcal{Z}^{FD} = \sum_r \ln(1 + e^{(\mu - \varepsilon_r)\beta}),$$

$$\ln \mathcal{Z}^{BE} = \sum_r - \ln(1 - e^{(\mu - \varepsilon_r)\beta}),$$

It's important to note here that $\sum_r$ is the sum over a single state! So if we want the **ensemble average occupation number**, we take (as was derived previously)

$$N = k_B T \frac{\partial}{\partial \mu} \ln \mathcal{Z}.$$

For fermions,

$$N = \sum_r \frac{e^{(\mu - \varepsilon_r)\beta}}{1 + e^{(\mu - \varepsilon_r)\beta}} = \sum_r \frac{1}{1 + e^{(\varepsilon_r - \mu)\beta}}$$

Each term here can be thought of as $\langle n_r \rangle$, the average occupation number of state r! That means that for a given state,

$$\boxed{\langle n_r \rangle = \frac{1}{e^{(\varepsilon_r - \mu)\beta}}},$$

and this is known as the **Fermi-Dirac occupation number.** In particular, we have the Fermi function

$$f(t_r) = \frac{1}{1 + e^{\beta(t_r - \mu)}}.$$

Note that μ is allowed to be both positive and negative here. If we take low temperature T, $\beta \to \infty$, and then chemical potential μ essentially separates our filled and empty states! The probability of having a state $\varepsilon_r > \mu$ is almost zero, and the probability of having a state for $\varepsilon_r < \mu$ is almost 1. As we increase temperature T, the jump from probability 0 to 1 becomes less steep, and we'll see how that works in a minute.

Meanwhile, **for bosons,**

$$N = \sum_r -\frac{e^{(\mu - \varepsilon_r)\beta}}{1 - e^{(\mu - \varepsilon_r)\beta}} = \sum_r \frac{1}{e^{(\varepsilon_r - \mu)\beta} - 1}.$$

Analogously, this can be thought os as a sum $\sum_r \langle n_r \rangle$, so we know that the average occupation

$$\boxed{\langle n_r \rangle = \frac{1}{e^{\beta(t_r - \mu) - 1}}}.$$

Notice that the main difference here from the Fermi-Dirac occupation number is the -1 instead of the $+1$! (By the way, remember that this is exactly the expression we saw with photons if we set $\mu = 0$.)

One important idea with this system, though, is that $\langle n_r \rangle \geq 0$ for non-interacting particles. If $\mu \to \varepsilon_r$, the denominator goes to 0, and thus $\langle n_r \rangle$ goes to infinity. In fact, $\mu > \varepsilon_r$ means our occupation number becomes negative! This isn't allowed, so that actually sets a bound on our allowed μ in relation to our energy states ε_r.

Proposition 244

The allowed values of μ for a boson system are

$$\mu \leq \varepsilon_{r_{\text{min}}},$$

where $\varepsilon_{r_{\text{min}}}$ is the lowest energy state for a single particle.

By the way, Professor Ketterle may talk about Bose-Einstein condensates next Thursday!

46.4 Looking more carefully at the occupation numbers

Let's take $\langle n_r \rangle$ to high temperature in both cases, so $\varepsilon_r - \mu \gg k_B T$. Then we should expect that our results agree with the canonical ensemble answer of

$$\langle n_r \rangle = e^{-\beta(\varepsilon_r - \mu)}$$

for classical, identical particles. Indeed, this is what happens in the limiting case for both FD and BE statistics, since the exponential terms dominate the 1 in the denominator! So our new model is really mostly useful at lower temperatures.

This means we really care about having our inter-atom distance on the order of of our length scale meaning that

$$\frac{V}{N} \sim \left(\frac{h}{\sqrt{2\pi m_e k_B T}} \right)^3.$$

Example 245

Electrons at room temperature have a length scale of around 4.5 nanometers, and copper has a density of 9×10^3 kg$/m^3$, which means there are 8.5×10^{28} atoms per cubic meter. So if we have one "conduction electron" per atom, the volume we have to work with is about

$$\frac{1}{8.5 \times 10^{28}} m^3 \text{ per atom.}$$

This is about $(0.23nm)^3$, and that means the inter-atom distance is small enough for us to want to use Fermi-Dirac statistics! But looking at copper atoms instead of the electrons themselves, the relevant length scale is around 0.012 nm. In other words, copper is a bunch of non-degenerate atoms immersed in a degenerate electron gas.

46.5 Fermions at low temperature

We should think of "degenerate" as having density large enough for quantum effects to be important. This is because two particles being too close gives overlapping wavefunctions.

So for the remainder of this class, we'll be looking at degenerate Fermi systems (where we use Fermi-Dirac statistics). Examples include electrons in a conduction band of metals, white dwarf stars, neutron stars, and heavy metals.

Here our occupation numbers follow

$$n_j = \frac{1}{1 + e^{(\varepsilon_j - \mu)\beta}}.$$

Note that when $\mu \gg k_B T$, levels with $\varepsilon < \mu$ have $n_j \to 1$, and levels with $\varepsilon > \mu$ have $n_j \to 0$.

We get **complete degeneracy** at $T = 0$, where we have essentially a sharp change from filled to empty states. Let's think a little more about such a degenerate Fermi gas: now we essentially have

$$n_j = \frac{1}{1 + e^{(\varepsilon_j - \mu)\infty}} = \begin{cases} 1 & \varepsilon_j < \mu \\ 0 & \varepsilon_j > \mu \end{cases}.$$

This is essentially a step function! All states with low energy are filled, and all others are empty. If we increase our temperature a little, the ∞ in the exponent becomes a large number, and we get a little bit of wiggle room. Then in some interval $\Delta\varepsilon \sim k_B T$, we can have partially filled states.

Fact 246

The filled levels are known as the "Fermi sea," and the set of states with $\varepsilon = \mu$ is known as the "Fermi surface." We can then define a "Fermi momentum"

$$p_F = \sqrt{2m\varepsilon_F}$$

where $\varepsilon_F = \mu$.

In three dimensions, we can calculate the density of states of our Fermi gas:

$$dN = g\frac{d^3x d^3p}{(2\pi\hbar)^3},$$

where g is our degeneracy factor (2 for an electron). This can then be written as

$$N = g \int \frac{d^3x d^3p}{(2\pi\hbar)^3} \frac{1}{e^{(p^2/2m - \mu)\beta} + 1},$$

and we can do similar tricks with the spherical integration as before: this evaluates to

$$= \frac{4\pi V g}{(2\pi\hbar)^3}\int_0^\infty \frac{p^2\,dp}{e^{(p^2/2m-\mu)\beta}+1}.$$

We'll finish this and calculate some of the thermodynamic properties here next time!

47 May 13, 2019 (Recitation)

Professor Ketterle showed us another Stirling engine today! Basically, we have a displacer which displaces gas between a hot reservoir (hot water) and cold reservoir (room temperature air). Whenever the piston moves down towards the hot reservoir, the air above gets colder, so there is a temperature differential. This causes a pressure modulation, which allows another piston to drive the wheel forward.

47.1 Entropic force

Often, we think of energy as the cause of force. But in statistics, we can think of entropy causing some kind of force as well! Recall the differential formula

$$dU = TdS - PdV;$$

we can use this to find the pressure at constant S or U:

$$P = T\left.\frac{\partial S}{\partial V}\right|_U = -\left.\frac{\partial U}{\partial V}\right|_S.$$

On the other hand, we also know that we have the free energy

$$F = U - TS \implies dF = -SdT - PdV,$$

which leads us to P in terms of isothermal derivatives:

$$P = -\left.\frac{\partial F}{\partial V}\right|_T = \left.\frac{\partial (U-TS)}{\partial V}\right|_T = \boxed{-\left.\frac{\partial U}{\partial V}\right|_T + T\left.\frac{\partial S}{\partial V}\right|_T}.$$

But the pressure is always the same, regardless of what process we're using to get to this point (because we have a state function)!

So does pressure come from internal energy or from entropy? There's actually contributions from both an internal energy change and an entropy change! Applying this to the ideal gas, because $\left.\frac{\partial U}{\partial V}\right|_T = 0$, we actually have

$$P = T\left.\frac{\partial S}{\partial V}\right|_T,$$

and **pressure is determined by entropy alone!**

Fact 247

It's important to think about isothermal versus adiabatic compression here: in the former case, compression leads to no change in internal energy (which is only dependent on T), so we get an output of heat. On the other hand, adiabatic compression leads to an increase in the kinetic energy of the particles, which means the temperature does increase.

47.2 Applying this to the chain model

Let's go back to the question where we have N chains of length ℓ, each of which is either flipped to the left or to the right in a one-dimensional system. Here, assume that the system doesn't have any internal energy, so all states are equally probable.

There are 2^N possible configurations of our chain (since each chain can be flipped to the left or right independently), and this is essentially a random walk! We can describe such a system by a binomial distribution, which can be approximated as a Gaussian for large N: since the variance of each individual chain is ℓ^2,

$$p(x) = p_0 \exp\left(-\frac{1}{2}\frac{x^2}{\ell^2 N}\right).$$

From this, we know that our multiplicity

$$\Gamma(x) = \Gamma_0 \exp\left(-\frac{1}{2}\frac{x^2}{\ell^2 N}\right),$$

so the entropy

$$S = k_B \ln \Gamma = k_B \left(c - \frac{1}{2}\frac{x^2}{\ell^2 N}\right):$$

(since the Gaussian exponential nicely cancels out with the logarithm!)

But now, we can calculate the force exerted by the chain: since P and V were conjugate variables in our boxed equation above, we can also replace them with F and x. Notably, if $\frac{\partial U}{\partial x}\big|_T = 0$ here (notice that $-PdV$ is work, and so is $+Fdx$, so we gain a negative sign),

$$F = -T \left.\frac{\partial S}{\partial x}\right|_T \implies \boxed{F = k_B T \frac{x}{\ell^2 N}}.$$

So force is proportional to temperature, and it's also a Hooke's law relation! This is a "pure entropy" situation, where we just needed to count the number of microstates to find dependence of S on the length x of the chain.

In other words, we put energy into the chain when we stretch it, but there's no way to store internal energy in this chain! So the energy of pulling will be transferred as heat, as that's the only way we can move the chain to a **lower entropy state** by the second law.

Question 248. *What causes the restoration here?*

Let's imagine that our chain is now vertical, and there is a mass hanging from the end. At a given temperature, this gives us the length x of the chain where we have equilibrium.

But if we increase the temperature of our reservoir, what happens? Because our weight is constant, the force in the above equation is constant, so Tx is constant. In other words, when temperature goes up, the chain gets shorter! This is a lot like if we put a weight on a piston sitting on top of an ideal gas: increasing the temperature of the gas increases the pressure, which makes the piston move higher up.

How exactly can such a massless chain with no internal energy even hold up a mass? Remember that the chain is always connected to a reservoir at temperature T! For entropic reasons, this nudges some of the chains upward. There's always a process to transfer force and energy between any reservoirs that exist and the system at hand.

47.3 Blackbody radiation and the Debye model

The former deals with a "gas" of photons, and the latter (which deals with specific heat of a solid) deals with a "gas" of phonons, the quantized excitations of sound.

In both cases, we can say that we have a bunch of energy states: then what's the expectation value for the average occupation number of a harmonic oscillator? It's given by the expression

$$\frac{1}{e^{\hbar\omega/(k_BT)}-1}.$$

If we want to know the total energy of a bunch of harmonic oscillator quanta like this, we multiply by the energy $\hbar\omega$ (for each harmonic oscillator), and then we need a density of states to know how many harmonic oscillators we have per interval. So

$$U = \int d\omega \hbar\omega \frac{1}{e^{\hbar\omega/(k_BT)}-1}\frac{dN_\omega}{d\omega},$$

and this is essentially just a dimensional analysis argument! Integrating this out from 0 frequency to some maximum ω_{max}, we have found our internal energy.

But how are the two different in their descriptions? The main difference is that we have dispersion relations $\omega = c_s k$ and $\omega = ck$, and in both cases, we find that $\frac{dN}{d\omega}$ is quadratic in ω^2. The difference is just the factor of the speed of photons versus phonons! In addition, there is also a factor of $\frac{3}{2}$ from the polarization factors. But most importantly, **the maximum frequency is upper bounded for the Debye model** because of the limitations of the solid, while the maximum frequency is unbounded for blackbody radiation. Each of these concepts is important to understand!

48 May 14, 2019

Today's lecture is being given by Professor Ketterle.

40 minutes ago, Professor Ketterle sent a press release about the kilogram changing next Monday! It will no longer be defined by the Paris artifact but by Planck's constant. We should celebrate this, because Planck's constant is much more beautiful than a chunk of metal. But he wants to mention that explaining physics should be simple: it should not just be for "the physicists and the nerds."

Today's class will go in three parts: an introduction with the essence of quantum statistics, using the Bose-Einstein distribution to describe phase transition, and how to create the first Bose-Einstein condensate in a gas.

48.1 Bose-Einstein condensate: an introduction

Bose-Einstein condensate allows us to create a matter wave by having many particles in one quantum state. This is based on quantum statistics, and we can use situations from class to understand it!

Example 249

Let's say we have 3 particles with some total energy 3.

This is a microcanonical ensemble: what's the classical description of this? There's 1 way to have all three particles in energy level 1, $3! = 6$ ways to have 3 particles in energy levels $2, 1, 0$, and 3 ways to have one in energy 3 and the others in energy 0.

But if we have indistinguishable particles, our distribution is different. Now we don't have the multiplicity of 6 and 3; in particular, the distribution of particles in energy levels for these **bosons** is different from the classical model. This means bosons actually have a tendency to clump together!

Finally, there's one more problem: if we have **fermions**, we're forced into a specific case, because you can't have two different particles in the same state!

Fact 250

Notice that the difference for occupancy level iis just a -1 or $+1$ in the denominator of $\frac{1}{e^{(\varepsilon-\mu)/(k_B T)}}$.

In particular, photons are bosons, but they can be freely created! This means there is no "chemical potential," and we can set $\mu = 0$ here. This gives us the Planck blackbody spectrum

$$n(\varepsilon) = \frac{1}{e^{\varepsilon/(k_B T)} - 1}.$$

Notably, Bose rederived Planck's formula with a new method, and Einstein used that method to add in the μ chemical potential term! That became the "Bose-Einstein distribution."

Question 251. *How do we describe* **Bose-Einstein condensation***, then?*

The idea is that the classical distribution just shifts to be skinnier and taller when we adjust our temperature T. But then at a certain critical temperature $T < T_C$, corresponding to a critical wavelength, the population per energy state goes to infinity. This specific singularity is described in Einstein's paper.

But back in 1924, Einstein said (after describing the theory) "The theory is pretty, but is there also some truth to it?" The idea is that "mathematical singularities" may not need to be part of actual physics, so there was lots of skepticism. It was only in 1938 that Fritz London realized that Bose-Einstein condensation is indeed physical and observable!

48.2 Why use the formalism that we do?

We're going to look more at the equations and understand the singularity mathematically. First of all, note that we can formulate things in many ways, but smarter choices (for example, in terms of coordinates) make our job easier.

We know that for atoms and molecules, energy is conserved, and so is the number of particles. But there is a problem here: distributing energies is not independent, and this means one of the particles needs to "pay the price." We like to assume in classical physics that each particle is an independent entity, and fluctuations shouldn't affect each other! That's why we use the canonical ensemble: we can then say that N particles have a partition function

$$Z_n = Z_1^N \text{ or } \frac{1}{N!} Z_1^N.$$

But then Einstein's 1924 paper seemed to cause some problems: the particles turned out to not be independent anymore under the canonical distribution. But indeed, descriptions under quantum physics no longer have particle independence (for example, the Pauli exclusion principle)!

Instead, our independence shifts to the **quantum states** themselves. By allowing each quantum state to run through each of its possible occupation numbers, we also allow for our total number of particles N to fluctuate, and now we use the grand canonical ensemble

$$\mathcal{Z} = \prod_i \mathcal{Z}_i.$$

That's the beauty of the grand canonical ensemble - we get our independence back again!

48.3 Mathematical derivation

We've derived in class before that the occupation number under the Bose-Einstein distribution

$$n_j = \frac{1}{e^{(\varepsilon_j-\mu)/(k_B T)} - 1}.$$

Question 252. *What is the population n_0 (corresponding to zero momentum and lowest energy)?*

We use the fact that as $T \to 0$, the chemical potential is negative but approaches 0^-. Then the occupation of the zero energy state is

$$n_0 \to \frac{k_B T}{-\mu} \to \infty.$$

Notably, this means the chemical potential can't be positive, or we'd have a singularity at some non-ground state!

Question 253. *What is the number of particles in an excited state (all $j \neq 0$)?*

This is just the sum over all nonzero states

$$N = \sum_{\varepsilon_j > \delta} \frac{1}{e^{\beta(\varepsilon_j - \mu)} - 1}$$

where we'll make $\delta \to 0$. If we only care about having an upper bound, we can ignore the small negative value of chemical potential:

$$N \le \sum_{\varepsilon_j > \delta} \frac{1}{e^{\beta \varepsilon_j} - 1}.$$

Remember that when we sum over all states, we can do this in a semi-classical manner instead:

$$= \int \frac{d^3x d^3p}{h^3} \frac{1}{e^{\beta\varepsilon_j} - 1}.$$

The position integral just becomes V, the volume of our system, and then we can use the spherical integration trick again:

$$= \frac{V 4\pi}{h^3} \int dp p^2 \frac{1}{e^{\beta\varepsilon_j} - 1}.$$

We can now replace $p^2 dp$ with $\sqrt{\varepsilon} d\varepsilon$ with a constant factor; $C\sqrt{\varepsilon}$ here is our density of states! This ends up giving us (replacing h with $\hbar$ for simplicity) an upper bound

$$N_{\text{max}} = \int_{\delta}^{\infty} \frac{N(\varepsilon)}{e^{\beta\varepsilon - 1}}, N(\varepsilon) = \frac{V \left(\frac{2m}{\hbar}\right)^{3/2} \sqrt{\varepsilon}}{4\pi^2}.$$

We'll remove most of the ugly constants by introducing a thermal de Broglie wavelength and dimensionless variable

$$\lambda = \sqrt{\frac{2\pi\hbar^2}{mk_B T}}, x = \beta\varepsilon.$$

Remember that our integral is still counting the number of particles in non-ground states: this gives us

$$= \frac{V}{\lambda^3} \frac{2}{\sqrt{\pi}} \int_{\delta\beta}^{\infty} \frac{\sqrt{x} dx}{e^x - 1}.$$

But now we can actually replace $\delta\beta$ with 0, since the integral converges, and we get an estimate of

$$\boxed{N_{\text{max}} \approx 2.612 \frac{V}{\lambda^3}}.$$

Notice that this is a fixed number dependent on T! So if we put more particles into our system at a fixed temperature, they don't go into excited states! So we have some absolute limit on the number of excited particles in terms of V and T: all others go into the ground state.

Proposition 254

In other words, eventually we have a **saturation of the quantum gas**: when we have too many particles, the whole system must condense! This is similar to the way in which eventually water vapor at 100 degrees Celsius must start forming water droplets when the pressure is too large.

So if we lower the temperature enough, and N_{max} reaches a point comparable to N, every subsequent increase in the number of particles or decrease in temperature will create a Bose-Einstein condensate.

Fact 255

Notably, if we set the boxed $N_{\text{max}} = N$, we can find our critical temperature

$$k_B T_C = \left(\frac{N}{2.612}\right)^{2/3} \frac{\hbar^2 2\pi}{L^2 m},$$

where $V = L^3$ if we assume we have a box.

Notably, the $\frac{N^{2/3}}{L^2}$ is proportional to $n^{2/3}$, where n is the density of our gas!

There's two things we should discuss here. First of all, the density we can get to realistically (when we work with individual atoms) is $n = 10^{14}$, which is a factor of 10^5 smaller than room temperature gases. Then we find that the T_C here is 100 nanoKelvin: that means verifying this experimentally requires us to get to very cold temperatures!

Question 256. *If we want almost all atoms in the ground state, what's the temperature scale that we are allowed to have?*

Remember that there is a first excited state: we have **gapped behavior here**. In our classical model, an energy lower than that must put all the particles in the ground state! On the other hand, though, the Bose-Einstein model has an extra factor of $N^{2/3}$ in it. Because bosons are indistinguishable, we get this extra factor that actually helps us: everything goes to the ground state much faster!

48.4 How do we do this experimentally?

Fact 257

In a paper that Schrodinger wrote in 1952, he expressed the opinion that van der Waals corrections (attractions and repulsions between molecules) and effects of Bose-Einstein condensation are almost impossible to separate. He thinks that most systems will become liquids before any quantum effects can be seen.

But it turns out that there exist cooling methods that can get us to very cold atomic clouds! Atoms cannot "keep photons," so if we shine a laser beam at an atom, the photons that hit the atom must be emitted out by fluorescence.

But now we can blue-shift the emitted radiation, every time an atom absorbs and emits a photon, it radiates away some of its energy! This is hard to implement in the laboratory, but this is the way **laser cooling** works.

Fact 258

Atoms are a special system: they are almost completely immune to black-body radiation, so we don't need to do the same shielding from cameras and beams and other materials as in more complicated systems.

Well, this gets us to the microKelvin level: what can we do to get closer to the 100 nanoKelvin level that we want?

It turns out **evaporative cooling** is very easy to understand. In a thermos, steam molecules (which have the highest energy) escape, and the lower-energy molecules stay behind. In our atomic system, we have a container made by magnetic or electric fields, and then use "radio frequency spin flips" to select the particles with highest energy (This is the same as "blowing on a cup of coffee").

In other words, if we remove our electric/magnetic field container, the gas will expand at a thermal velocity. Since the kinetic energy $\frac{mv^2}{2} = \frac{k_B T}{2}$ by equipartition, this is a way for us to figure out the temperature without needing to explicitly put something like a thermometer in contact with it!

Fact 259

By the way, right now the temperature of 1 Kelvin is $\frac{1}{273.15}$ times the triple point of water. Soon it'll be defined in terms of the Boltzmann constant instead!

Basically, when we "blow" with our radio waves, we should expect smaller clouds, and indeed, this is what happens! There's an elliptical object in the middle that stays put - that's the Bose-Einstein condensate that we can observe.

Remember that a thermal gas is isotropic: the shadow should always be perfectly circular. So why is the condensate elliptical? It's because the ground state of an elliptical container is elliptical!

Fact 260

Finally, how do we prove that the atoms act as one single wave?

The key idea is **interference**! Two waves that collide form a standing wave, and it turns out that we can do the same kind of interference by taking a Bose-Einstein condensate and cutting it in half. The interesting thing here is that we have now accomplished interference of matter waves, since the positive and negative part of the wavefunction add up to zero!

49 May 15, 2019 (Recitation)

49.1 Drawing more connections

Let's start by trying to understand the last concept we introduced in this class. Exchanging energy in the canonical ensemble is very similar to exchanging particles in a grand canonical ensemble! In the former, we're connected to a temperature reservoir, and in the second, we're connected to some chemical potential reservoir.

What are the transitions here? We fix the internal energy U in our microcanonical ensemble, but in a canonical ensemble, we essentially tweak our energy as a function of T, our temperature. In particular, we don't clamp down the energy: one thing we can do is to maximize entropy as a function of β (using Lagrange multipliers). The mathematics here is that if we want to maximize $F(x, y, z)$ subject to some constraint $E(x, y, z) = \varepsilon_0$, we can introduce a new parameter β which also varies independently:

$$F(x, y, z) - \beta E = 0$$

and then all derivatives with respect to x, y, z, β must be zero. So the **microcanonical condition is enforced with the Lagrange parameter**, and ultimately that parameter is adjusted (aka, the temperature is our independent variable) to reach our energy E.

Fact 261

So temperature plays a dual role: temperature affects internal energy, but we can also use it as a Lagrange parameter to ask questions like "how to maximize entropy." E and β depend each other.

Well, this same dependence happens with N and μ! If we have a reservoir with some chemical potential, μ controls the number of particles N, just like the temperature T controls our energy E. We now want to maximize the free energy F, and we use a similar Lagrange parameter: this time, it is μ. So now μ can also be our "knob" that controls N.

Example 262

Consider a semiconductor piece of metal. A battery is then a source of chemical potential: increasing the voltage charges up the system and introduces more electrons!

So the battery becomes the "energetic cost" of delivering another particle. Our grand partition function

$$\mathcal{Z} = \sum_j e^{-\beta(\varepsilon_j - n_j \mu}$$

now takes into account that each particle being present changes our energy somehow. Remember where this all comes from: we treat our system as being connected to a reservoir, and we think of this whole system-reservoir as being a microcanonical ensemble. Then the probability of any specific microstate μ_j is

$$\Pr(\mu_j) \propto \Gamma_{\text{Res}}(E - \varepsilon_j, N - n_j),$$

since E and N are fixed across the whole system-reservoir entity. This is then proportional to $e^{S_{\text{Res}}}$, and expanding out S to first order in our Taylor expansion yields the result that we want:

$$= \frac{\partial S}{\partial N}(n_j) + \frac{\partial S}{\partial E}(-\varepsilon_j) = \beta \mu n_j - \beta \varepsilon_j$$

and that yields the $-\beta(\varepsilon_j - \mu n_j)$ that we want! This Taylor expansion basically tells us "how much energy it costs" to give away a particle or give away some energy.

49.2 Going to the lab

Most particles in nature are fermions (quarks, electrons, and so on), rather than bosons.

Fact 263

By the way, the main difference is the spin (inherent angular momentum) of the particle, and whether it's an integer or a half-integer. Two indistinguishable particles must satisfy $\psi(x_1, x_2) = \pm\psi(x_2, x_1)$, where the ± comes from us only observing the square of the wave function. Turns out $-$ corresponds to fermions, and $+$ corresponds to bosons: now the Pauli exclusion principle comes from the "spin statistic theorem!"

50 May 16, 2019

We're going to finish talking about Fermi gas thermodynamics, and we'll finish by talking about what lies beyond this class!

Fact 264

By the way, the grade cutoffs for A/B, B/C, and C/D last year were 85, 65, and 55. This fluctuates year to year, though.

We should check this weekend that everything is graded! The deadline for all grading-related things is tentatively **Sunday**, but everything will definitely be done by Wednesday morning. Exam 1 and 2 solutions will be posted soon, and past exams have been posted for reviewing for the final as well.

50.1 Back to Fermi systems: a gas of electrons

Remember that fermions are a gas of spin $\frac{1}{2}$. We found that because of the Pauli exclusion principle, we have an interesting result for the occupation number at zero temperature for a **degenerate Fermi gas**:

$$n_j(\varepsilon) = \frac{1}{1 + e^{(\varepsilon - \mu)\beta}}$$

this is basically a step function at $\varepsilon = \mu$. (If we increase the temperature to some T, we get partially filled states with a width of around $\Delta\varepsilon \approx k_B T$.)

At temperature $T = 0$, define $\mu = E_f$ to be the **Fermi energy**: we can then also define a Fermi momentum $p_F = \sqrt{2mE_f}$, which has applications in condensed matter. The terms **Fermi sea** and **Fermi surface** then refer to the filled states where $E < E_f$ and the set of states where $E = E_f$, respectively.

We're going to try to derive the thermodynamics of a 3D Fermi gas now! We start with a density of states calculation: we have

$$dN = g\frac{d^3x d^3p}{(2\pi\hbar)^3},$$

where g, the degeneracy factor, is $2s + 1$ (where s is the spin of our particle). Since electrons have spin $\frac{1}{2}$, $g = 2$ in this case! Integrating out,

$$N = g\int \frac{d^3x d^3p}{(2\pi\hbar)^3}\frac{1}{e^{(p^2/2m - \mu)\beta} + 1}.$$

We'll do the same tricks we keep doing: integrating out d^3x and using spherical coordinates, this simplifies to a single integral

$$N = \frac{4\pi V g}{(2\pi\hbar)^3}\int_0^\infty \frac{p^2 dp}{e^{(p^2/2m - \mu)} + 1}.$$

As we take temperature to 0, we have some upper limit equal to our Fermi momentum (since the denominator is 1 for $p < p_F$ and ∞ for $p > p_F$), so this simplifies to (combining some constants)

$$N = \frac{gV}{2\pi^2\hbar^3}\int_0^{p_F} p^2 dp = \frac{gVp_F^3}{6\pi^2\hbar^3};$$

substituting back for our Fermi energy yields

$$E_f = \frac{\hbar^2}{2m}\left(\frac{6\pi^2}{g}\frac{N}{V}\right)^{2/3}.$$

The idea is that this is the energy of the last filled state, because fermions that enter our system successively fill energy levels from lowest to highest! This gives us a **Fermi wavenumber**

$$\frac{p_F}{\hbar} = k_f \implies N = \frac{g}{6\pi^2}Vk_F^3.$$

We can also define a **Fermi temperature**

$$T_f = \frac{E_f}{k_B};$$

surprisingly, for electrons in metal, this temperature is around 10^4 Kelvin, and for electrons in a white dwarf, this is around 10^7 Kelvin! Since these numbers are so large, the description of Fermi gases is very good for deriving material properties of solids, as well as other fermion systems.

So now if we want to calculate our energy, we integrate

$$U = g \int \frac{d^3x d^3p}{(2\pi\hbar)^3} \frac{p^2/2m}{e^{(p^2/2m-\mu)\beta}+1};$$

If we take our temperature $T \to 0$, the same behavior with the denominator happens, and we're left with

$$U = \frac{gV}{2\pi^2\hbar^3} \int_0^{p_F} dp p^2 \frac{p^2}{2m} = \frac{gVp_F^5}{20m\pi^2\hbar^3}.$$

Again, we can rearrange to write U in terms of N, our number of particles, and E_f, our Fermi energy: this yields

$$U = \frac{3}{5} N E_F.$$

Finally, how do we derive the equation of state? If U is written as a function of $N, V, T = 0$ (in our limiting case), then

$$P = -\left.\frac{\partial U}{\partial V}\right|_{N,T=0}.$$

We can rewrite our expression above for U: it turns out that we have

$$U = \text{constant} \cdot \frac{N^{5/3}}{V^{2/3}},$$

so taking the derivative,

$$P = \frac{2}{3}\frac{U}{V} = \frac{gp_F^5}{30\pi^2 m\hbar^3},$$

which can be rewritten in terms of Fermi energy as

$$\boxed{PV = \frac{2}{5} N E_f}.$$

So even at temperature 0, there is some residual pressure! This is known as "degenerate pressure," and it occurs because of the Pauli exclusion principle - this has stabilizing effects in certain systems.

The only problem is that we've assumed our electrons are free particles, but we know that this isn't true - they're bound to nuclei! So we need to start adding correction terms to account for interactions like this.

50.2 What's next?

There's a lot of cool and exciting areas that we can study after this class! Here's some of them:

- Phase transitions. These can be observed in real life, and they also have applications in biophysics and other areas like particle physics!
- Non-equilibrium physics: how does a system out of equilibrium relax into equilibrium? There's something called "linear response theory" here, as well as a notion of "non-equilibrium steady states."
- Dynamical processes.

- Thermodynamics of small systems - we've been using large N to simplify a lot of our calculations, but there's an exciting area of theoretical development where we start caring more about fluctuations! There's notions of "work-fluctuation" and other strange phenomena.

We should all continue with statistical physics after this point! This class is a basis for doing other, more exciting things.