

# A Crash Course in Statistical Mechanics

Noah Miller

December 27, 2018

## Abstract

A friendly introduction to statistical mechanics, geared towards covering the powerful methods physicists have developed for working in the subject.

## Contents

<b>1</b>	<b>Statistical Mechanics</b>	<b>1</b>
1.1	Entropy . . . . .	1
1.2	Temperature and Equilibrium . . . . .	5
1.3	The Partition Function . . . . .	8
1.4	Free energy . . . . .	14
1.5	Phase Transitions . . . . .	15
1.6	Example: Box of Gas . . . . .	17
1.7	Shannon Entropy . . . . .	18
1.8	Quantum Mechanics, Density Matrices . . . . .	19
1.9	Example: Two state system . . . . .	21
1.10	Entropy of Mixed States . . . . .	23
1.11	Classicality from environmental entanglement . . . . .	23
1.12	The Quantum Partition Function . . . . .	27

## 1 Statistical Mechanics

### 1.1 Entropy

Statistical Mechanics is a branch of physics that pervades all other branches. Statistical mechanics is relevant to Newtonian mechanics, relativity, quantum mechanics, and quantum field theory.

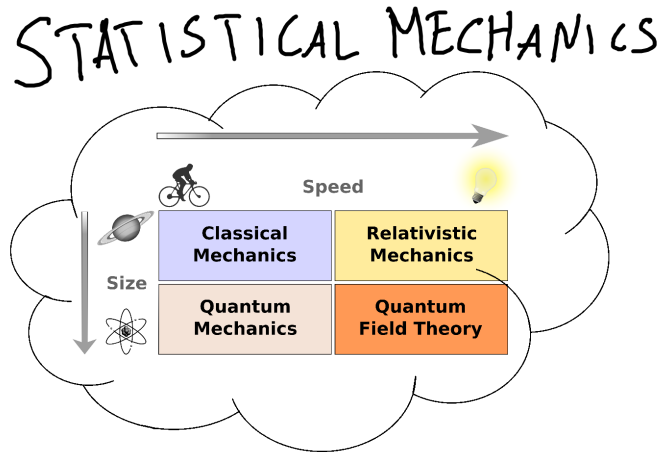


Figure 1: Statistical mechanics applies to all realms of physics.

Its exact incarnation is a little different in each quadrant, but the basic details are identical.

The most important quantity in statistical mechanics is called “entropy,” which we label by  $S$ . People sometimes say that entropy is a measure of the “disorder” of a system, but I don’t think this a good way to think about it. But before we define entropy, we need to discuss two different notions of state: “microstates” and “macrostates.”

In physics, we like to describe the real world as mathematical objects. In classical physics, states are points in a “phase space.” Say for example you had  $N$  particles moving around in 3 dimensions. It would take  $6N$  real numbers to specify the physical state of this system at a given instant: 3 numbers for each particle’s position and 3 numbers for each particle’s momentum. The phase space for this system would therefore just be  $\mathbb{R}^{6N}$ .

$$(x_1, y_1, z_1, p_{x1}, p_{y1}, p_{z1}, \dots, x_N, y_N, z_N, p_{xN}, p_{yN}, p_{zN}) \in \mathbb{R}^{6N}$$

(In quantum mechanics, states are vectors in a Hilbert space  $\mathcal{H}$  instead of points in a phase space. We’ll return to the quantum case a bit later.)

A “microstate” is a state of the above form. It contains absolutely all the physical information that an omniscient observer could know. If you were to know the exact microstate of a system and knew all of the laws of physics, you could in principle deduce what the microstate will be at all future times and what the microstate was at all past times.

However, practically speaking, we can never know the true microstate of a system. For example, you could never know the positions and momenta of every damn particle in a box of gas. The only things we can actually measure are macroscopic variables such as internal energy, volume, and particle number  $(U, V, N)$ . A “macrostate” is just a set of

microstates. For examples, the “macrostate” of a box of gas labelled by  $(U, V, N)$  would be the set of all microstates with energy  $U$ , volume  $V$ , and particle number  $N$ . The idea is that if you know what macrostate your system is in, you know that your system is equally likely to truly be in any of the microstates it contains.

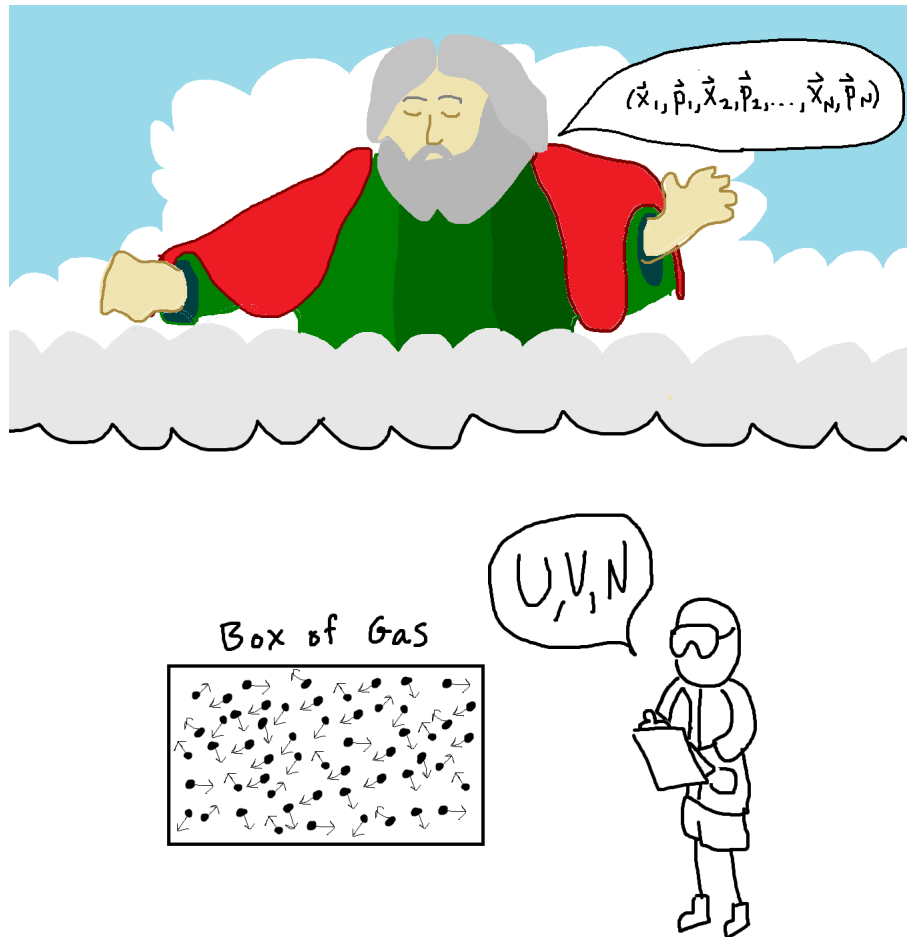


Figure 2: You may know the macrostate, but only God knows the microstate.

I am now ready to define what entropy is. Entropy is a quantity associated with a macrostate. If a macrostate is just a set of  $\Omega$  microstates, then the entropy  $S$  of the system is

$$S \equiv k \log \Omega. \tag{1}$$

Here,  $k$  is Boltzmann’s constant. It is a physical constant with units of energy / temperature.

$$k \equiv 1.38065 \times 10^{-23} \frac{\text{Joules}}{\text{Kelvin}} \tag{2}$$

The only reason that we need  $k$  to define  $S$  is because the human race defined units of temperature before they defined entropy. (We'll see how temperature factors into any of this soon.) Otherwise, we probably would have set  $k = 1$  and temperature would have the same units as energy.

You might be wondering how we actually count  $\Omega$ . As you probably noticed, the phase space  $\mathbb{R}^{6N}$  is not discrete. In that situation, we integrate over a phase space volume with the measure

$$d^3x_1 d^3p_1 \dots d^3x_N d^3p_N.$$

However, this isn't completely satisfactory because position and momentum are dimensionful quantities while  $\Omega$  should be a dimensionless number. We should therefore divide by a constant with units of position times momentum. Notice, however, that because  $S$  only depends on  $\log \Omega$ , any constant rescaling of  $\Omega$  will only alter  $S$  by a constant and will therefore never affect the change in entropy  $\Delta S$  of some process. So while we have to divide by a constant, whichever constant we divide by doesn't affect the physics.

Anyway, even though we are free to choose whatever dimensionful constant we want, the "best" is actually Planck's constant  $h$ ! Therefore, for a classical macrostate that occupies a phase space volume  $\text{Vol}$ ,

$$\Omega = \frac{1}{N!} \frac{1}{h^{3N}} \int_{\text{Vol}} \prod_{i=1}^N d^3x_i d^3p_i. \quad (3)$$

(The prefactor  $1/N!$  is necessary if all  $N$  particles are indistinguishable. It is the cause of some philosophical consternation but I don't want to get into any of that.)

Let me now explain why I think saying entropy is "disorder" is not such a good idea. Different observers might describe reality with different macrostates. For example, say your room is very messy and disorganized. This isn't a problem for you, because you spend a lot of time in there and know where everything is. Therefore, the macrostate you use to describe your room contains very few microstates and has a small entropy. However, according to your mother who has not studied your room very carefully, the entropy of your room is very large. The point is that while everyone might agree your room is messy, the entropy of your room really depends on how little you know about it.

## 1.2 Temperature and Equilibrium

Let's say we label our macrostates by their total internal energy  $U$  and some other macroscopic variables like  $V$  and  $N$ . (Obviously, these other macroscopic variables  $V$  and  $N$  can be replaced by different quantities in different situations, but let's just stick with this for now.) Our entropy  $S$  depends on all of these variables.

$$S = S(U, V, N) \quad (4)$$

The temperature  $T$  of the  $(U, V, N)$  macrostate is then be defined to be

$$\frac{1}{T} \equiv \left. \frac{\partial S}{\partial U} \right|_{V, N}. \quad (5)$$

The partial derivative above means that we just differentiate  $S(U, V, N)$  with respect to  $U$  while keeping  $V$  and  $N$  fixed.

If your system has a high temperature and you add a bit of energy  $dU$  to it, then the entropy  $S$  will not change much. If your system has a small temperature and you add a bit of energy, the entropy will increase a lot.

Next, say you have two systems  $A$  and  $B$  which are free to trade energy back and forth.

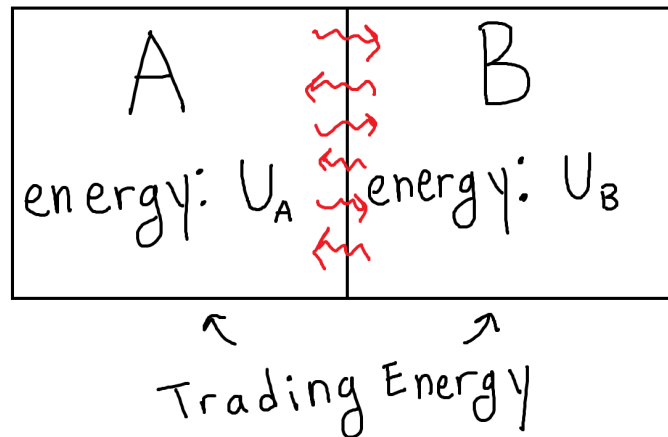


Figure 3: Two systems  $A$  and  $B$  trading energy.  $U_A + U_B$  is fixed.

Say system  $A$  could be in one of  $\Omega_A$  possible microstates and system  $B$  could be in  $\Omega_B$  possible microstates. Therefore, the total  $AB$  system could be in  $\Omega_A \Omega_B$  possible microstates. Therefore, the entropy  $S_{AB}$  of both systems combined is just the sum of entropies of both sub-systems.

$$S_{AB} = k \log(\Omega_A \Omega_B) = k \log \Omega_A + k \log \Omega_B = S_A + S_B \quad (6)$$

The crucial realization of statistical mechanics is that, all else being equal, a system is most likely to find itself in a macrostate corresponding to the largest number of microstates. This is the so-called “Second law of thermodynamics”: for all practical intents and purposes, the entropy of a closed system always increases over time. It is not really a physical “law” in the regular sense, it is more like a profound realization.

Therefore, the entropy  $S_{AB}$  of our joint  $AB$  system will increase as time goes on until it reaches its maximum possible value. In other words,  $A$  and  $B$  trade energy in a seemingly random fashion that increases  $S_{AB}$  on average. When  $S_{AB}$  is finally maximized, we say that our systems are in “thermal equilibrium.”

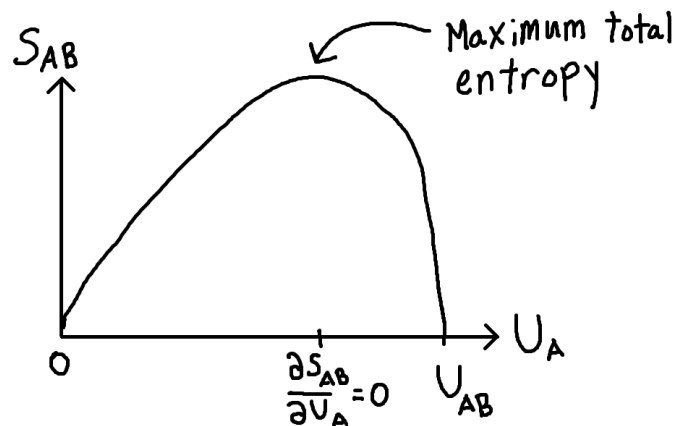


Figure 4:  $S_{AB}$  is maximized when  $U_A$  has some particular value. (It should be noted that there will actually be tiny random "thermal" fluctuations around this maximum.)

Let’s say that the internal energy of system  $A$  is  $U_A$  and the internal energy of system  $B$  is  $U_B$ . Crucially, note that the total energy of combined system

$$U_{AB} = U_A + U_B$$

is constant over time! This is because energy of the total system is conserved. Therefore,

$$dU_A = -dU_B.$$

Now, the combined system will maximize its entropy when  $U_A$  and  $U_B$  have some particular values. Knowing the value of  $U_A$  is enough though, because  $U_B = U_{AB} - U_A$ . Therefore, entropy is maximized when

$$0 = \frac{\partial S_{AB}}{\partial U_A}. \quad (7)$$

However, we can rewrite this as

$$\begin{aligned}
 0 &= \frac{\partial S_{AB}}{\partial U_A} \\
 &= \frac{\partial S_A}{\partial U_A} + \frac{\partial S_B}{\partial U_A} \\
 &= \frac{\partial S_A}{\partial U_A} - \frac{\partial S_B}{\partial U_B} \\
 &= \frac{1}{T_A} - \frac{1}{T_B}.
 \end{aligned}$$

Therefore, our two systems are in equilibrium if they have the same temperature!

$$T_A = T_B \tag{8}$$

If there are other macroscopic variables we are using to define our macrostates, like volume  $V$  or particle number  $N$ , then there will be other quantities that must be equal in equilibrium, assuming our two systems compete for volume or trade particles back and forth. In these cases, we define the quantities  $P$  and  $\mu$  to be

$$\frac{P}{T} \equiv \left. \frac{\partial S}{\partial V} \right|_{U,N} \qquad \frac{\mu}{T} \equiv - \left. \frac{\partial S}{\partial N} \right|_{U,V}. \tag{9}$$

$P$  is called “pressure” and  $\mu$  is called “chemical potential.” In equilibrium, we would also have

$$P_A = P_B \qquad \mu_A = \mu_B. \tag{10}$$

(You might object that pressure has another definition, namely force divided by area. It would be incumbent on us to check that this definition matches that definition in the relevant situation where both definitions have meaning. Thankfully it does.)

### 1.3 The Partition Function



Figure 5: If you want to do statistical mechanics, you really should know about the partition function.

Explicitly calculating  $\Omega$  for a given macrostate is usually very hard. Practically speaking, it can only be done for simple systems you understand very well. However, physicists have developed an extremely powerful way of doing statistical mechanics even for complicated systems. It turns out that there is a function of temperature called the “partition function” that contains all the information you’d care to know about your macrostate when you are working in the “thermodynamic limit.” This function is denoted  $Z(T)$ . Once you compute  $Z(T)$  (which is usually much easier than computing  $\Omega$ ) it is a simple matter to extract the relevant physics.

Before defining the partition function, I would like to talk a bit about heat baths. Say you have some system  $\mathcal{S}$  in a very large environment  $\mathcal{E}$ . Say you can measure the macroscopic variables of  $\mathcal{S}$ , including its energy  $E$  at any given moment. (We use  $E$  here to denote energy instead of  $U$  when talking about the partition function.) The question I ask is: if the total system has a temperature  $T$ , what’s the probability that  $\mathcal{S}$  has some particular energy  $E$ ?



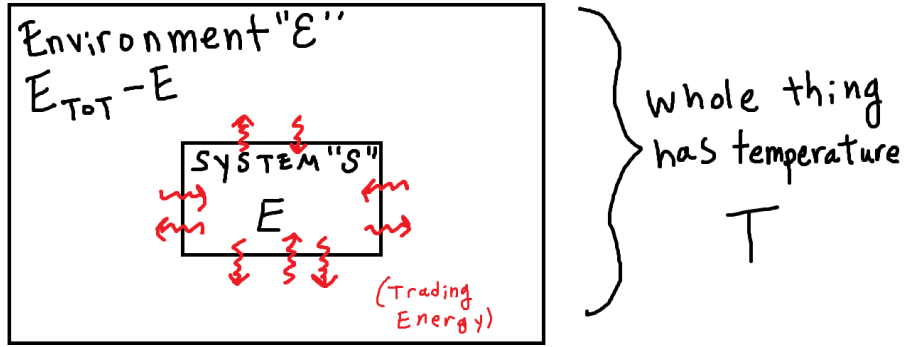


Figure 6: A large environment  $\mathcal{E}$  and system  $\mathcal{S}$  have a fixed total energy  $E_{\text{tot}}$ .  $\mathcal{E}$  is called a “heat bath” because it is very big. The combined system has a temperature  $T$ .

We should be picturing that  $\mathcal{S}$  and  $\mathcal{E}$  are evolving in some complicated way we can’t understand. However, their total energy

$$E_{\text{tot}} = E + E_{\mathcal{E}} \quad (11)$$

is conserved. We now define

$$\begin{aligned} \Omega_{\mathcal{S}}(E) &\equiv \text{num. microstates of } \mathcal{S} \text{ with energy } E \\ \Omega_{\mathcal{E}}(E_{\mathcal{E}}) &\equiv \text{num. microstates of } \mathcal{E} \text{ with energy } E_{\mathcal{E}}. \end{aligned} \quad (12)$$

Therefore, the probability that  $\mathcal{S}$  has some energy  $E$  is proportional to the number of microstates where  $\mathcal{S}$  has energy  $E$  and  $\mathcal{E}$  has energy  $E_{\text{tot}} - E$ .

$$\text{Prob}(E) \propto \Omega_{\mathcal{S}}(E)\Omega_{\mathcal{E}}(E_{\text{tot}} - E) \quad (13)$$

Here is the important part. Say that our heat bath has a lot of energy:  $E_{\text{tot}} \gg E$ . As far as the heat bath is concerned,  $E$  is a very small amount of energy. Therefore,

$$\begin{aligned} \Omega_{\mathcal{E}}(E_{\text{tot}} - E) &= \exp\left(\frac{1}{k}S_{\mathcal{E}}(E_{\text{tot}} - E)\right) \\ &\approx \exp\left(\frac{1}{k}S_{\mathcal{E}}(E_{\text{tot}}) - \frac{E}{kT}\right) \end{aligned}$$

by Taylor expanding  $S_{\mathcal{E}}$  in  $E$  and using the definition of temperature. We now have

$$\text{Prob}(E) \propto \Omega_{\mathcal{S}}(E) \exp\left(-\frac{E}{kT}\right).$$

$\Omega_{\mathcal{S}}(E)$  is sometimes called the “degeneracy” of  $E$ . In any case, we can easily see what the ratio of  $\text{Prob}(E_1)$  and  $\text{Prob}(E_2)$  must be.

$$\frac{\text{Prob}(E_1)}{\text{Prob}(E_2)} = \frac{\Omega_{\mathcal{S}}(E_1)e^{-E_1/kT}}{\Omega_{\mathcal{S}}(E_2)e^{-E_2/kT}}$$

Furthermore, we can use the fact that all probabilities must sum to 1 in order to calculate the absolute probability. We define

$$\begin{aligned} Z(T) &\equiv \sum_E \Omega_{\mathcal{S}}(E)e^{-E/kT} \\ &= \sum_s e^{-E_s/kT} \end{aligned} \tag{14}$$

where  $\sum_s$  is a sum over all states of  $\mathcal{S}$ . Finally, we have

$$\text{Prob}(E) = \frac{\Omega_{\mathcal{S}}(E)e^{-E/kT}}{Z(T)} \tag{15}$$

However, more than being a mere proportionality factor,  $Z(T)$  takes on a life of its own, so it is given the special name of the “partition function.” Interestingly,  $Z(T)$  is a function that depends on  $T$  and not  $E$ . It is not a function that has anything to do with a particular macrostate. Rather, it is a function that has to do with *every* microstate at some temperature. Oftentimes, we also define

$$\beta \equiv \frac{1}{kT}$$

and write

$$Z(\beta) = \sum_s e^{-\beta E_s}. \tag{16}$$

The partition function  $Z(\beta)$  has many amazing properties. For one, it can be used to write an endless number of clever identities. Here is one. Say you want to compute the expected energy  $\langle E \rangle$  your system has at temperature  $T$ .

$$\begin{aligned} \langle E \rangle &= \sum_s E_s \text{Prob}(E_s) \\ &= \frac{\sum_s E_s e^{-\beta E_s}}{Z(\beta)} \\ &= -\frac{1}{Z} \frac{\partial}{\partial \beta} Z \\ &= -\frac{\partial}{\partial \beta} \log Z \end{aligned}$$

This expresses the expected energy  $\langle E \rangle$  as a function of temperature. (We could also calculate  $\langle E^n \rangle$  for any  $n$  if we wanted to.)

Where the partition function really shines is in the “thermodynamic limit.” Usually, people define the thermodynamic limit as

$$N \rightarrow \infty \quad (\text{thermodynamic limit}) \quad (17)$$

where  $N$  is the number of particles. However, sometimes you might be interested in more abstract systems like a spin chain (the so-called “Ising model”) or something else. There are no “particles” in such a system, however there is still something you would justifiably call the thermodynamic limit. This would be when the number of sites in your spin chain becomes very large. So  $N$  should really just be thought of as the number of variables you need to specify a microstate. When someone is “working in the thermodynamic limit,” it just means that they are considering very “big” systems.

Of course, in real life  $N$  is never infinite. However, I think we can all agree that  $10^{23}$  is close enough to infinity for all practical purposes. Whenever an equation is true “in the thermodynamic limit,” you can imagine that there are extra terms of order  $\frac{1}{N}$  unwritten in your equation and laugh at them.

What is special about the thermodynamic limit is that  $\Omega_{\mathcal{S}}$  becomes, like, really big...

$$\Omega_{\mathcal{S}} = (\text{something})^N$$

Furthermore, the entropy and energy will scale with  $N$

$$S_{\mathcal{S}} = NS_1 \quad E = NE_1$$

In the above equation,  $S_1$  and  $E_1$  can be thought of as the average amount of entropy per particle.

Therefore, we can rewrite

$$\begin{aligned} \text{Prob}(E) &\propto \Omega_{\mathcal{S}}(E) \exp\left(-\frac{1}{kT}E\right) \\ &= \exp\left(\frac{1}{k}S_{\mathcal{S}} - \frac{1}{kT}E\right) \\ &= \exp\left(N\left(\frac{1}{k}S_1 - \frac{1}{kT}E_1\right)\right). \end{aligned}$$

The thing to really gawk at in the above equation is that the probability that  $\mathcal{S}$  has some energy  $E$  is given by

$$\text{Prob}(E) \propto e^{N(\dots)}.$$

I want you to appreciate how insanely big  $e^{N(\dots)}$  is in the thermodynamic limit. Furthermore, if there is even a miniscule change in  $(\dots)$ ,

$\text{Prob}(E)$  will change radically. Therefore,  $\text{Prob}(E)$  will be extremely concentrated at some particular energy, and deviating slightly from that maximum will cause  $\text{Prob}(E)$  to plummet.

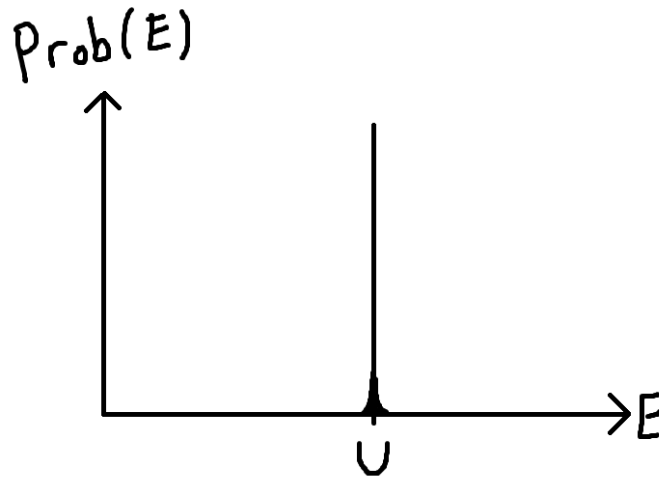


Figure 7: In the thermodynamic limit, the system  $\mathcal{S}$  will have a well defined energy.

We can therefore see that if the energy  $U$  maximizes  $\text{Prob}(E)$ , we will essentially have

$$\text{Prob}(E) \approx \begin{cases} 1 & \text{if } E = U \\ 0 & \text{if } E \neq U \end{cases}.$$

Let's now think back to our previously derived equation

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

Recall that  $\langle E \rangle$  is the expected energy of  $\mathcal{S}$  when it is coupled to a heat bath at some temperature. The beauty is that in the thermodynamic limit where our system  $\mathcal{S}$  becomes very large, we don't even have to think about the heat bath anymore! Our system  $\mathcal{S}$  is basically just in the macrostate where all microstates with energy  $U$  are equally likely. Therefore,

$$\langle E \rangle = U \quad (\text{thermodynamic limit})$$

and

$$U = -\frac{\partial}{\partial \beta} \log Z(\beta) \tag{18}$$

is an *exact* equation in the thermodynamic limit.

Let's just appreciate this for a second. Our original definition of  $S(U)$  was

$$S(U) = k \log(\Omega(U))$$

and our original definition of temperature was

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

In other words,  $T$  is a function of  $U$ . However, we totally reversed logic when we coupled our system to a larger environment. We no longer knew what the exact energy of our system was. I am now telling you that instead of calculating  $T$  as a function of  $U$ , when  $N$  is large we are actually able to calculate  $U$  as a function of  $T$ ! Therefore, instead of having to calculate  $\Omega(U)$ , we can just calculate  $Z(T)$  instead.

I should stress, however, that  $Z(T)$  is still a perfectly worthwhile thing to calculate even when your system  $\mathcal{S}$  isn't "big." It will still give you the exact average energy  $\langle E \rangle$  when your system is in equilibrium with a bigger environment at some temperature. What's special about the thermodynamic limit is that you no longer have to imagine the heat bath is there in order to interpret your results, because any "average quantity" will basically just be an actual, sharply defined, "quantity." In short,

$$Z(\beta) = \Omega(U)e^{-\beta U} \quad (\text{thermodynamic limit}) \quad (19)$$

It's worth mentioning that the other contributions to  $Z(\beta)$  will also be absolute huge; they just won't be as stupendously huge as the term due to  $U$ .

Okay, enough adulation for the partition function. Let's do something with it again. Using the above equation there is a very easy way to figure out what  $S_{\mathcal{S}}(U)$  is in terms of  $Z(\beta)$ .

$$\begin{aligned} S_{\mathcal{S}}(U) &= k \log \Omega_{\mathcal{S}}(U) \\ &= k \log(Ze^{\beta U}) \quad (\text{thermodynamic limit}) \\ &= k \log Z + k\beta U \\ &= k\left(1 - \beta \frac{\partial}{\partial \beta}\right) \log Z \end{aligned}$$

(Gah. Another amazing identity, all thanks to the partition function.)

This game that we played, coupling our system  $\mathcal{S}$  to a heat bath so we could calculate  $U$  as a function of  $T$  instead of  $T$  as a function of  $U$ , can be replicated with other quantities like the chemical potential  $\mu$  (defined in Eq. 10). We could now imagine that  $\mathcal{S}$  is trading particles

with a larger environment. Our partition function would then be a function of  $\mu$  in addition to  $T$ .

$$Z = Z(\mu, T)$$

In the thermodynamic limit, we could once again use our old tricks to find  $N$  in terms of  $\mu$  and  $T$ .

## 1.4 Free energy

Now that we're on an unstoppable victory march of introductory statistical mechanics, I think I should define a quantity closely related to the partition function: the "free energy"  $F$ .

$$F \equiv U - TS \tag{20}$$

(This is also called the "Helmholtz Free Energy.")  $F$  is defined for any system with some well defined internal energy  $U$  and entropy  $S$  when present in a larger environment which has temperature  $T$ . Crucially, *the system does not need to be in thermal equilibrium with the environment*. In other words, free energy is a quantity associated with some system which may or may not be in equilibrium with an environment at temperature  $T$ .

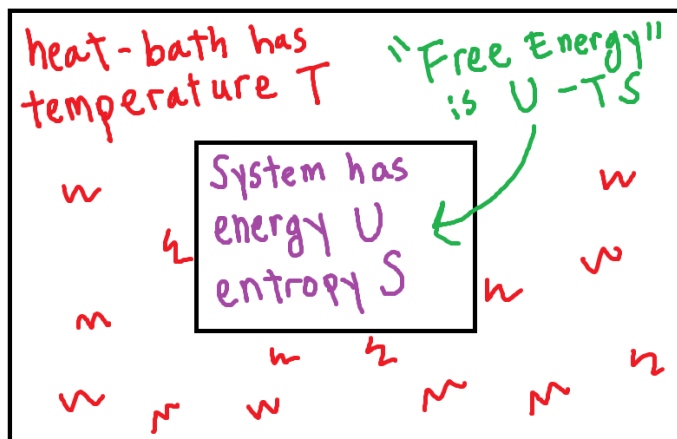


Figure 8: A system with internal energy  $U$  and entropy  $S$  in a heat bath at temperature  $T$  has free energy  $F = U - TS$ .

Okay. So why did we define this quantity  $F$ ? The hint is in the name "free energy." Over time, the system will equilibrate with the environment in order to maximize the entropy of the whole world. While doing so, the energy  $U$  of the system will change. So if we cleverly leave our system in a larger environment, under the right circumstances we

can let the second law of thermodynamics to do all the hard work, transferring energy into our system at no cost to us! I should warn you that  $\Delta F$  is actually not equal to the change in internal energy  $\Delta U$  that occurs during this equilibration. This is apparent just from its definition. (Although it does turn out that  $F$  is equal to the “useful work” you can extract from such a system.)

The reason I’m telling you about  $F$  is because it is a useful quantity for determining what will happen to a system at temperature  $T$ . Namely, in the thermodynamic limit, the system will *minimize*  $F$  by equilibrating with the environment.

Recall Eq. 19 (reproduced below).

$$Z(\beta) = \Omega(U)e^{-\beta U} \quad (\text{thermodynamic limit})$$

If our system  $\mathcal{S}$  is in equilibrium with the heat bath, then

$$\begin{aligned} Z(\beta) &= \exp\left(\frac{1}{k}S - \beta U\right) \quad (\text{at equilibrium in thermodynamic limit}) \\ &= \exp(-\beta F). \end{aligned}$$

First off, we just derived another amazing identity of the partition function. More importantly, recall that  $U$ , as written in Eq. 19, is defined to be the energy that maximizes  $\Omega(U)e^{-\beta U}$ , A.K.A. the energy that maximizes the entropy of the world. Because we know that the entropy of the world always wants to be maximized, we can clearly see that  $F$  wants to be minimized, as claimed.

Therefore,  $F$  is a very useful quantity! It always wants to be minimized at equilibrium. It can therefore be used to detect interesting phenomena, such as phase transitions.

## 1.5 Phase Transitions

Let’s back up a bit and think about a picture we drew, Fig. 7. It’s a very suggestive picture that begs a very interesting question. What if, at some critical temperature  $T_c$ , a new peak grows and overtakes our first peak?

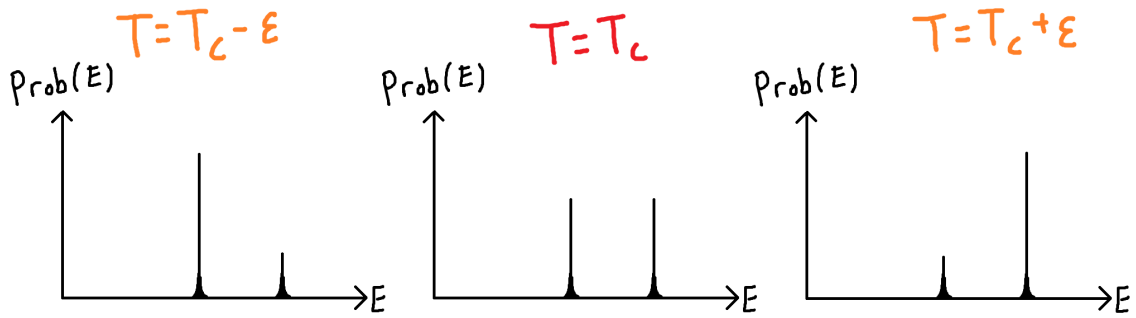


Figure 9: A phase transition, right below the critical temperature  $T_c$ , at  $T_c$ , and right above  $T_c$ .

This can indeed happen, and is in fact what a physicist would call a “first order phase transition.” We can see that there will be a discontinuity in the first derivative of  $Z(T)$  at  $T_c$ . You might be wondering how this is possible, given the fact that from its definition,  $Z$  is clearly an analytic function as it is a sum of analytic functions. The thing to remember is that we are using the thermodynamic limit, and the sum of an infinite number of analytic functions may not be analytic.

Because there is a discontinuity in the first derivative of  $Z(\beta)$ , there will be a discontinuity in  $E = -\frac{\partial}{\partial \beta} \log Z$ . This is just the “latent heat” you learned about in high school. In real life systems, it takes some time for enough energy to be transferred into a system to overcome the latent heat energy barrier. This is why it takes so long for a pot of water to boil or a block of ice to melt. Furthermore, during these lengthy phase transitions, the pot of water or block of ice will actually be at a constant temperature, the “critical temperature” (100°C and 0°C respectively). Once the phase transition is complete, the temperature can start changing again.

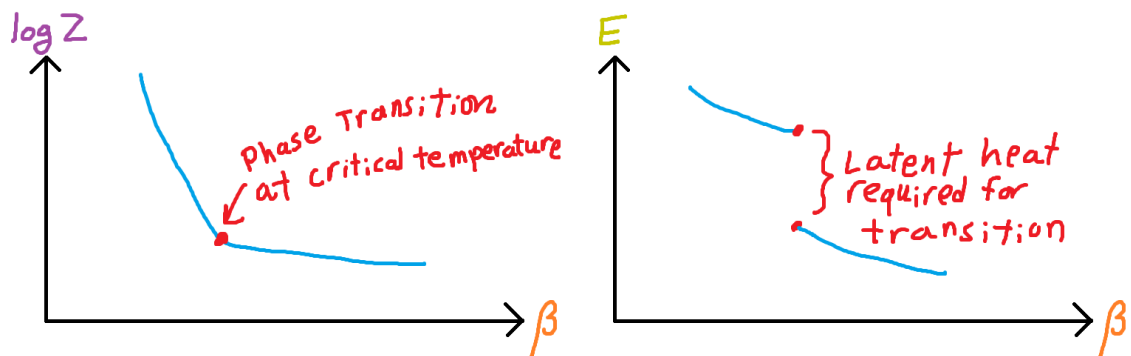


Figure 10: A discontinuity in the first derivative of  $Z$  corresponds to a first order phase transition. This means that you must put a finite amount of energy into the system called “latent heat” at the phase transition before the temperature of the system will rise again.



## 1.6 Example: Box of Gas

For concreteness, I will compute the partition function for an ideal gas. By ideal, I mean that the particles do not interact with each other.

Let  $N$  be the number of particles in the box and  $m$  be the mass of each particle. Suppose the particles exist in a box of volume  $V$ . The positions and momenta of the particles at  $\vec{x}_i$  and  $\vec{p}_i$  for  $i = 1 \dots N$ . The energy is given by the sum of kinetic energies of all particles.

$$E = \sum_{i=1}^N \frac{\vec{p}_i^2}{2m}. \quad (21)$$

Therefore,

$$\begin{aligned} Z(\beta) &= \sum_s e^{-\beta E_s} \\ &= \frac{1}{N!} \frac{1}{h^{3N}} \int \prod_{i=1}^N d^3x_i d^3p_i \exp\left(-\beta \sum_{i=1}^N \frac{\vec{p}_i^2}{2m}\right) \\ &= \frac{1}{N!} \frac{V^N}{h^{3N}} \prod_{i=1}^N \int d^3p_i \exp\left(-\beta \frac{\vec{p}_i^2}{2m}\right) \\ &= \frac{1}{N!} \frac{V^N}{h^{3N}} \left(\frac{2m\pi}{\beta}\right)^{3N/2} \end{aligned}$$

If  $N$  is large, the thermodynamic limit is satisfied. Therefore,

$$\begin{aligned} U &= -\frac{\partial}{\partial \beta} \log Z \\ &= -\frac{3}{2} N \frac{\partial}{\partial \beta} \log \left( N!^{\frac{-2}{3N}} \left(\frac{V}{h^3}\right)^{\frac{2}{3}} \frac{2m\pi}{\beta} \right) \\ &= \frac{3N}{2\beta} \\ &= \frac{3}{2} NkT. \end{aligned}$$

You could add interactions between the particles by adding some potential energy between  $V$  each pair of particles (unrelated to the volume  $V$ ).

$$E = \sum_{i=1}^N \frac{\vec{p}_i^2}{2m} + \frac{1}{2} \sum_{i,j} V(|\vec{x}_i - \vec{x}_j|) \quad (22)$$

The form of  $V(r)$  might look something like this.



Figure 11: An example for an interaction potential  $V$  between particles as a function of distance  $r$ .

The calculation of  $Z(\beta)$  then becomes more difficult, although you could approximate it pretty well using something called the “cluster decomposition.” This partition function would then exhibit a phase transition at a critical temperature between a gas phase and a liquid phase. It is an interesting exercise to try to pin down for yourself where all the new states are coming from at the critical temperature which make  $Z(\beta)$  discontinuous. (Hint: condensation.)

Obviously, the attractions real life particles experience cannot be written in terms of such a simple central potential  $V(r)$ . It’s just a simplified model. For example, there should be some angular dependence to the potential energy as well which is responsible for the chemical structures we see in nature. If we wanted to model the liquid-to-solid transition, we’d have to take that into account.

## 1.7 Shannon Entropy

So far, we have been imagining that that all microstates in a macrostate are equally likely to be the “true” microstate. However, what if you assign a different probability  $p_s$  to each microstate  $s$ ? What is the entropy then?

There is a more general notion of entropy in computer science call “Shannon entropy.” It is given by

$$S = - \sum_s p_s \log p_s. \quad (23)$$

It turns out that entropy is maximized when all the probabilities  $p_s$  are equal to each other. Say there are  $\Omega$  states and each  $p_s = \Omega^{-1}$ . Then

$$S = \log \Omega \quad (24)$$

matching the physicist’s definition (up to the Boltzmann constant).

One tiny technicality when dealing with the Shannon entropy is interpreting the value of

$$0 \log 0.$$

It is a bit troublesome because  $\log 0 = -\infty$ . However, it turns out that the correct value to assign the above quantity is

$$0 \log 0 \equiv 0.$$

This isn't too crazy though, because

$$\lim_{x \rightarrow 0} x \log x = 0.$$

## 1.8 Quantum Mechanics, Density Matrices

So far I have only told you about statistical mechanics in the context of classical mechanics. Now it's time to talk about quantum mechanics.

There is something very interesting about quantum mechanics: states can be in superpositions. Because of this, even if you know the exact quantum state your system is in, you can still only predict the probabilities that any observable (such as energy) will have a particular value when measured. Therefore, there are *two* notions of uncertainty in quantum statistical mechanics:

1. Fundamental quantum uncertainty
2. Uncertainty due to the fact that you may not know the exact quantum state your system is in anyway. (This is sometimes called "classical uncertainty.")

It would be nice if we could capture these two different notions of uncertainty in one unified mathematical object. This object is called the "density matrix."

Say the quantum states for your system live in a Hilbert space  $\mathcal{H}$ . A density matrix  $\rho$  is an operator

$$\rho : \mathcal{H} \rightarrow \mathcal{H}. \tag{25}$$

Each density matrix is meant to represent a so-called "classical superposition" of quantum states.

For example, say that you are a physics PhD student working in a lab and studying some quantum system. Say your lab mate has prepared the system in one of two states  $|\psi_1\rangle$  or  $|\psi_2\rangle$ , but unprofessionally forgot

which one it is in. This would be an example of a “classical superposition” of quantum states. Usually, we think of classical superpositions as having a thermodynamical nature, but that doesn’t have to be the case.

Anyway, say that your lab mate thinks there’s a 50% chance the system could be in either state. The density matrix corresponding to this classical superposition would be

$$\rho = \frac{1}{2} |\psi_1\rangle \langle\psi_1| + \frac{1}{2} |\psi_2\rangle \langle\psi_2|.$$

More generally, if you have a set of  $N$  quantum states  $|\psi_i\rangle$  each with a classical probability  $p_i$ , then the corresponding density matrix would be

$$\rho = \sum_{i=1}^N p_i |\psi_i\rangle \langle\psi_i|. \quad (26)$$

This is useful to define because it allows us to extract expectation values of observables  $\hat{O}$  in a classical superposition. But before I prove that, I’ll have to explain a very important operation: “tracing.”

Say you have quantum state  $|\psi\rangle$  and you want to calculate the expectation value of  $\hat{O}$ . This is just equal to

$$\langle\hat{O}\rangle = \langle\psi|\hat{O}|\psi\rangle. \quad (27)$$

Now, say we have an orthonormal basis  $|\phi_s\rangle \in \mathcal{H}$ . We then have

$$1 = \sum_s |\phi_s\rangle \langle\phi_s|. \quad (28)$$

Therefore, inserting the identity, we have

$$\begin{aligned} \langle\hat{O}\rangle &= \langle\psi|\hat{O}|\psi\rangle \\ &= \sum_s \langle\psi|\hat{O}|\phi_s\rangle \langle\phi_s|\psi\rangle \\ &= \sum_s \langle\phi_s|\psi\rangle \langle\psi|\hat{O}|\phi_s\rangle. \end{aligned}$$

This motivates us to define something called the “trace operation” for any operator  $\mathcal{H} \rightarrow \mathcal{H}$ . While we are using an orthonormal basis of  $\mathcal{H}$  to define it, it is actually independent of which basis you choose.

$$\text{Tr}(\dots) \equiv \sum_s \langle\phi_s|\dots|\phi_s\rangle \quad (29)$$

We can therefore see that for our state  $|\psi\rangle$ ,

$$\langle \hat{\mathcal{O}} \rangle = \text{Tr}\left(|\psi\rangle \langle \psi| \hat{\mathcal{O}}\right). \quad (30)$$

Returning to our classical superposition and density matrix  $\rho$ , we are now ready to see how to compute the expectation values.

$$\begin{aligned} \langle \hat{\mathcal{O}} \rangle &= \sum_i p_i \langle \psi_i | \hat{\mathcal{O}} | \psi_i \rangle \\ &= \sum_i p_i \text{Tr}\left(|\psi_i\rangle \langle \psi_i| \hat{\mathcal{O}}\right) \\ &= \text{Tr}\left(\rho \hat{\mathcal{O}}\right) \end{aligned}$$

So I have now proved my claim that we can use density matrices to extract expectation values of observables.

Now that I have told you about these density matrices, I should introduce some terminology. A density matrix that is of the form

$$\rho = |\psi\rangle \langle \psi|$$

for some  $|\psi\rangle$  is said to represent a “pure state,” because you know with 100% certainty which quantum state your system is in. Note that for a pure state,

$$\rho^2 = \rho \quad (\text{for pure state}).$$

It turns out that the above condition is a necessary and sufficient condition for determining if a density matrix represents a pure state.

If a density matrix is instead a combination of different states in a classical superposition, it is said to represent a “mixed state.” This is sort of bad terminology, because a mixed state is not a “state” in the Hilbert space  $\hat{\mathcal{H}}$ , but whatever.

## 1.9 Example: Two state system

Consider the simplest Hilbert space, representing a two state system.

$$\mathcal{H} = \mathbb{C}^2$$

Let us investigate the difference between a quantum superposition and a classical super position. An orthonormal basis for this Hilbert space is given by

$$|0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad |1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Say you have a classical superposition of these two states where you have a 50% probability that your state is in either state. Then

$$\begin{aligned}\rho_{\text{Mixed}} &= \frac{1}{2} |0\rangle \langle 0| + \frac{1}{2} |1\rangle \langle 1| \\ &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.\end{aligned}$$

Let's compare this to the pure state of the quantum super position

$$|\psi\rangle = \frac{1}{\sqrt{2}} |0\rangle + \frac{1}{\sqrt{2}} |1\rangle.$$

The density matrix would be

$$\begin{aligned}\rho_{\text{Pure}} &= \left( \frac{1}{\sqrt{2}} |0\rangle + \frac{1}{\sqrt{2}} |1\rangle \right) \left( \frac{1}{\sqrt{2}} \langle 0| + \frac{1}{\sqrt{2}} \langle 1| \right) \\ &= \frac{1}{2} (|0\rangle \langle 0| + |1\rangle \langle 1| + |0\rangle \langle 1| + |1\rangle \langle 0|) \\ &= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}\end{aligned}$$

The pure state density matrix is different from the mixed state because of the non-zero off diagonal terms. These are sometimes called “interference terms.” The reason is that states in a quantum superposition can “interfere” with each other, while states in a classical superposition can't.

Let's now look at the expectation value of the following operators for both density matrices.

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

They are given by

$$\begin{aligned}\langle \sigma_z \rangle_{\text{Mixed}} &= \text{Tr} \left( \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right) = 0 \\ \langle \sigma_z \rangle_{\text{Pure}} &= \text{Tr} \left( \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right) = 0 \\ \langle \sigma_x \rangle_{\text{Mixed}} &= \text{Tr} \left( \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right) = 0 \\ \langle \sigma_x \rangle_{\text{Pure}} &= \text{Tr} \left( \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right) = 1\end{aligned}$$

So we can see that a measurement given by  $\sigma_z$  cannot distinguish between  $\rho_{\text{Mixed}}$  and  $\rho_{\text{Pure}}$ , while a measurement given by  $\sigma_x$  can distinguish between them! There really is a difference between classical superpositions and quantum superpositions, but you can only see this difference if you exploit the off-diagonal terms!

## 1.10 Entropy of Mixed States

In quantum mechanics, pure states are microstates and mixed states are the macrostates. We can define the entropy of a mixed state drawing inspiration from the definition of Shannon entropy.

$$S = -k \text{Tr}(\rho \log \rho) \quad (31)$$

This is called the Von Neumann Entropy. If  $\rho$  represents a classical superposition of orthonormal states  $|\psi_i\rangle$ , each with some probability  $p_i$ , then the above definition exactly matches the definition of Shannon entropy.

One thing should be explained, though. How do you take the logarithm of a matrix? This is actually pretty easy. Just diagonalize the matrix and take the log of the diagonal entries. Thankfully, density matrices can always be diagonalized (they are manifestly self-adjoint and therefore diagonalizable by the spectral theorem) so you don't have to do anything more complicated.

## 1.11 Classicity from environmental entanglement

Say you have two quantum systems  $A$  and  $B$  with Hilbert spaces  $\mathcal{H}_A$  and  $\mathcal{H}_B$ . If you combine the two systems, states will live in the Hilbert space

$$\mathcal{H}_A \otimes \mathcal{H}_B.$$

Say that  $|\phi_i\rangle_A \in \mathcal{H}_A$  comprise a basis for the state space of  $\mathcal{H}_A$  and  $|\phi_j\rangle_B \in \mathcal{H}_B$  comprise a basis for the state space  $\mathcal{H}_B$ . All states in  $\mathcal{H}_A \otimes \mathcal{H}_B$  will be of the form

$$|\Psi\rangle = \sum_{i,j} c_{ij} |\phi_i\rangle_A |\phi_j\rangle_B$$

for some  $c_{ij} \in \mathbb{C}$ .

States are said to be “entangled” if they can **not** be written as

$$|\psi\rangle_A |\psi\rangle_B$$

for some  $|\psi\rangle_A \in \mathcal{H}_A$  and  $|\psi\rangle_B \in \mathcal{H}_B$ .

So, for example, if  $\mathcal{H}_A = \mathbb{C}^2$  and  $\mathcal{H}_B = \mathbb{C}^2$ , then the state

$$|0\rangle \left( \frac{1}{\sqrt{2}} |0\rangle - \frac{i}{\sqrt{2}} |1\rangle \right)$$

**would not** be entangled, while the state

$$\frac{1}{\sqrt{2}} \left( |0\rangle |0\rangle + |1\rangle |1\rangle \right)$$

**would** be entangled.

Let's say a state starts out unentangled. How would it then become entangled over time? Well, say the two systems  $A$  and  $B$  have Hamiltonians  $\hat{H}_A$  and  $\hat{H}_B$ . If we want the systems to interact weakly, i.e. "trade energy," we'll also need to add an interaction term to the Hamiltonian.

$$\hat{H} = \hat{H}_A \otimes \hat{H}_B + \hat{H}_{\text{int}}.$$

It doesn't actually matter what the interaction term is or if it is very small. All that matters is that if we really want them to interact, it's important that the interaction term is there at all. Once we add an interaction term, we will generically see that states which start out unentangled become heavily entangled over time as  $A$  and  $B$  interact.

Say for example you had a system  $\mathcal{S}$  described by a Hilbert space  $\mathcal{H}_{\mathcal{S}}$  coupled to a large environment  $\mathcal{E}$  described by a Hilbert space  $\mathcal{H}_{\mathcal{E}}$ . Now, maybe you are an experimentalist and you are really interested in studying the quantum dynamics of  $\mathcal{S}$ . You then face a very big problem:  $\mathcal{E}$ . Air molecules in your laboratory will be constantly bumping up against your system, for example. This is just intuitively what I mean by having some non-zero  $\hat{H}_{\text{int}}$ . The issue is that, if you really want to study  $\mathcal{S}$ , you desperately *don't* want it to entangle with the environment, because you have no control over the environment! This is why people who study quantum systems are always building these big complicated vacuum chambers and cooling their system down to fractions of a degree above absolute zero: they want to prevent entanglement with the environment so they can study  $\mathcal{S}$  in peace!



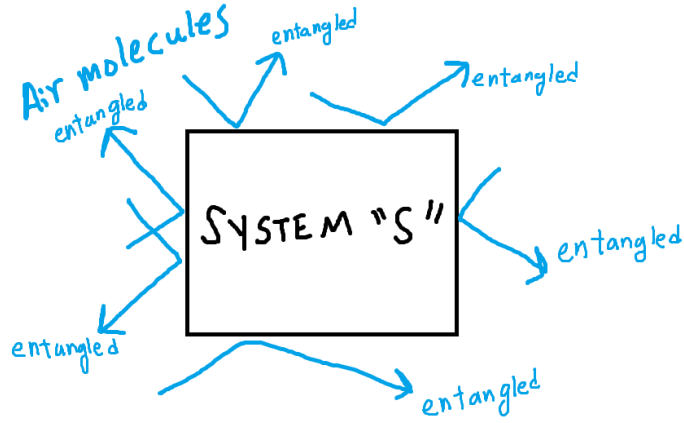


Figure 12: Air molecules bumping up against a quantum system  $\mathcal{S}$  will entangle with it.

Notice that the experimentalist will not have access to the observables in the environment. Associated with  $\mathcal{H}_{\mathcal{S}}$  is a set of observables  $\hat{\mathcal{O}}_{\mathcal{S}}$ . If you tensor these observables together with the identity,

$$\hat{\mathcal{O}}_{\mathcal{S}} \otimes 1_{\mathcal{E}}$$

you now have an observable which only measures quantities in the  $\mathcal{H}_{\mathcal{S}}$  subsector of the full Hilbert space. The thing is that entanglement within the environment gets in the way of measuring  $\hat{\mathcal{O}}_{\mathcal{S}} \otimes 1_{\mathcal{E}}$  in the way the experimenter would like.

Say, for example,  $\mathcal{H}_{\mathcal{S}} = \mathbb{C}^2$  and  $\mathcal{H}_{\mathcal{E}} = \mathbb{C}^N$  for some very big  $N$ . Any state in  $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{E}}$  will be of the form

$$c_0 |0\rangle |\psi_0\rangle + c_1 |1\rangle |\psi_1\rangle \quad (32)$$

for some  $c_0, c_1 \in \mathbb{C}$  and  $|\psi_0\rangle, |\psi_1\rangle \in \mathcal{H}_{\mathcal{E}}$ . The expectation value for our observable is

$$\langle \hat{\mathcal{O}}_{\mathcal{S}} \otimes 1_{\mathcal{E}} \rangle = \left( c_0^* \langle 0| \langle \psi_0| + c_1^* \langle 1| \langle \psi_1| \right) \hat{\mathcal{O}}_{\mathcal{S}} \otimes 1_{\mathcal{E}} \left( c_0 |0\rangle |\psi_0\rangle + c_1 |1\rangle |\psi_1\rangle \right)$$

$$= |c_0|^2 \langle 0| \hat{\mathcal{O}}_{\mathcal{S}} |0\rangle + |c_1|^2 \langle 1| \hat{\mathcal{O}}_{\mathcal{S}} |1\rangle + 2 \operatorname{Re} \left( c_0^* c_1 \langle 0| \hat{\mathcal{O}}_{\mathcal{S}} |1\rangle \langle \psi_0 | \psi_1 \rangle \right)$$

The thing is that, if the environment  $\mathcal{E}$  is very big, then any two random given vectors  $|\psi_0\rangle, |\psi_1\rangle \in \mathcal{H}_{\mathcal{E}}$  will generically have almost no overlap.

$$\langle \psi_0 | \psi_1 \rangle \approx e^{-N}$$

(This is just a fact about random vectors in high dimensional vector spaces.) Therefore, the expectation value of this observable will be

$$\langle \hat{\mathcal{O}}_{\mathcal{S}} \otimes 1_{\mathcal{E}} \rangle \approx |c_0|^2 \langle 0| \hat{\mathcal{O}}_{\mathcal{S}} |0\rangle + |c_1|^2 \langle 1| \hat{\mathcal{O}}_{\mathcal{S}} |1\rangle.$$

Because there is no cross term between  $|0\rangle$  and  $|1\rangle$ , we can see that when we measure our observable, our system  $\mathcal{S}$  seems to be in a classical superposition, A.K.A a mixed state!

This can be formalized by what is called a “partial trace.” Say that  $|\phi_i\rangle_{\mathcal{E}}$  comprises an orthonormal basis of  $\mathcal{H}_{\mathcal{E}}$ . Say we have some density matrix  $\rho$  representing a state in the full Hilbert space. We can “trace over the  $\mathcal{E}$  degrees of freedom” to receive a density matrix in the  $\mathcal{S}$  Hilbert space.

$$\rho_{\mathcal{S}} \equiv \text{Tr}_{\mathcal{E}}(\rho) \equiv \sum_i \langle \phi_i | \rho | \phi_i \rangle_{\mathcal{E}}. \quad (33)$$

You be wondering why anyone would want to take this partial trace. Well, I would say that if you can't perform the  $\mathcal{E}$  degrees of freedom, why are you describing them? It turns out that the partially traced density matrix gives us the expectation values for any observables in  $\mathcal{S}$ . Once we compute  $\rho_{\mathcal{S}}$ , by tracing over  $\mathcal{E}$ , we can then calculate the expectation value of any observable  $\hat{O}_{\mathcal{S}}$  by just calculating the trace over  $\mathcal{S}$  of  $\rho_{\mathcal{S}}\hat{O}_{\mathcal{S}}$ :

$$\text{Tr}\left(\rho\hat{O}_{\mathcal{S}} \otimes 1_{\mathcal{E}}\right) = \text{Tr}_{\mathcal{S}}(\rho_{\mathcal{S}}\hat{O}_{\mathcal{S}}).$$

Even though the whole world is in some particular state in  $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{E}}$ , when you only perform measurements on one part of it, that part might as well only be in a mixed state for all you know! Entanglement looks like a mixed state when you only look at one part of a Hilbert space. Furthermore, when the environment is very large, the off diagonal “interference terms” in the density matrix are usually very close to zero, meaning the state looks very mixed.

This is the idea of “entanglement entropy.” If you have an entangled state, then trace out over the states in one part of the Hilbert space, you will receive a mixed density matrix. That density matrix will have some Von Neumann entropy, and in this context we would call it “entanglement entropy.” The more entanglement entropy your state has, the more entangled it is! And, as we can see, when you can only look at one tiny part of a state when it is heavily entangled, it appears to be in a classical superposition instead of a quantum superposition!

The process by which quantum states in real life become entangled with the surrounding environment is called “decoherence.” It is one of the most viciously efficient processes in all of physics, and is the reason why it took the human race so long to discover quantum mechanics. It's very ironic that entanglement, a quintessentially quantum phenomenon, when taken to dramatic extremes, hides quantum mechanics from view

entirely!

I would like to point out an important difference between a classical macrostate and a quantum mixed state. In classical mechanics, the subtle perturbing effects of the environment on the system make it difficult to keep track of the exact microstate a system is in. However, in principle you can always just re-measure your system very precisely and figure out what the microstate is all over again. This isn't the case in quantum mechanics when your system becomes entangled with the environment. The problem is that once your system entangles with the environment, that entanglement is almost certainly never going to undo itself. In fact, it's just going to spread from the air molecules in your laboratory to the surrounding building, then the whole university, then the state, the country, the planet, the solar system, the galaxy, and then the universe! And unless you "undo" all of that entanglement, the show's over! You'd just have to start from scratch and prepare your system in a pure state all over again.

## 1.12 The Quantum Partition Function

The quantum analog of the partition function is very straightforward. The partition function is defined to be

$$\begin{aligned} Z(T) &\equiv \text{Tr} \exp\left(-\hat{H}/kT\right) \\ &= \sum_s e^{-\beta E_s}. \end{aligned} \tag{34}$$

Obviously, this is just the same  $Z(T)$  that we saw in classical mechanics! They are really not different at all. However, there is something very interesting in the above expression. The operator

$$\exp\left(-\hat{H}/kT\right)$$

looks an awful lot like the time evolution operator

$$\exp\left(-i\hat{H}t/\hbar\right)$$

if we just replace

$$-\frac{i}{\hbar}t \longrightarrow -\beta.$$

It seems as though  $\beta$  is, in some sense, an "imaginary time." Rotating the time variable into the imaginary direction is called a "Wick Rotation,"

and is one of the most simple, mysterious, and powerful tricks in the working physicist's toolbelt. There's a whole beautiful story here with the path integral, but I won't get into it.

Anyway, a mixed state is said to be "thermal" if it is of the form

$$\begin{aligned}\rho_{\text{Thermal}} &= \frac{1}{Z(T)} \sum_s e^{-E_s/kT} |E_s\rangle \langle E_s| \\ &= \frac{1}{Z(\beta)} e^{-\beta \hat{H}}\end{aligned}\tag{35}$$

for some temperature  $T$  where  $|E_s\rangle$  are the energy eigenstates with eigenvalues  $E_s$ . If you let your system equilibrate with an environment at some temperature  $T$ , and then trace out by the environmental degrees of freedom, you will find your system in the thermal mixed state.