Statistical Mechanics

And now for a very different, but very modern approach to Statistical Mechanics

You will be thinking about physics and mathematics in a a different way!

We will be developing everything starting from very simple ideas!

DO NOT be misled by simple arguments and simplistic starting points, we will end up with the full theory at the end!

The sections with the Questions/Answers are very important so make sure you read them carefully

Let us begin the fun.

Part 1: Entropy and conservation of information

1.1 Introduction

Statistical mechanics is not really modern physics.

It is pre-modern physics; it is modern physics; and you can be assured that it will be post-modern physics.

The second law of thermodynamics will probably outlast anything that comes up.

Time and again, the second law of thermodynamics has been sort of a guidepost, or our guiding light, to know what we are talking about and to make sure we are making sense.

Statistical mechanics and thermodynamics may not be as sexy as the Higgs boson, but it is at least as deep.

Many would say a lot deeper.

Particle physicists shouldn’t disown me.

It is a lot more general and it covers a lot more ground in explaining the world as we know it.

And in fact without statistical mechanics we probably would not know about the Higgs boson.

So what is statistical mechanics about?

To answer this question, let's go back a step.

The basic laws of physics, Newton's laws, the principles of classical physics, classical mechanics, quantum mechanics, electrodynamics course and so forth are all about perfect predictability.

Now you can object that in quantum mechanics we cannot predict perfectly.

And that is true.

But there are some things we can predict perfectly and those things are the predictables of quantum mechanics.

In all these disciplines, we can make our predictions with maximum precision, or so-called maximal Statistical mechanics is what we use in those cases.

Predictability arises if we know two things:

1. the *starting point* – which is what we call the initial conditions – and

2. the *laws of evolution* of a system.

When working with a *closed system*, that is a system which either comprises everything, or is sufficiently isolated from everything else that external factors don't influence it, if we know the initial conditions exactly or with sufficient precision, and we know the laws of evolution of the system, then we have complete *predictability*.

And that is all there is to say.

In many cases, of course, that complete predictability would be totally useless.

Having a list of the positions and velocities of every particle in the room would not be very useful to us.

The list would be too long, and subject to rather quick changes as a matter of fact.

So we can see that while the basic laws of physics are extremely powerful in their predictability, they can also be in many cases totally useless for actually analyzing what is really going on.

Statistical mechanics is what we use in those cases.

Let's say first of all that the mathematics of statistical mechanics – statistical mechanics as applied to physical systems – is just elementary probability theory.

Some people go so far as saying that statistical mechanics simply is probability theory.

When is statistical mechanics applicable?

It is when we don't know the initial conditions with complete perfection.

It may even be the case if we don't know the laws of motion with infinite precision.

And it is applicable when the system we are investigating is not a closed system, that is when it is interacting with other things on the outside.

In other words, in all those situations where ideal predictability is impossible, we resort to probabilities.

However because, for instance, the number of molecules in the room is so large and probabilities tend to become very precise predictors when the law of large numbers is applicable, statistical mechanics itself can be highly predictable.

But that is not true for everything.

As an illustration, let's say we have a box of gas – it might even be an isolated closed box of gas.

It has some energy in it.

The particles rattle around.

If we know some things about that box of gas, we can predict other things with great precision.

If we know the temperature, we can predict the energy in the box of gas.

We can predict the pressure.

These things are highly predictable. There are other things, as said, we cannot predict.

We cannot predict the position of every molecule.

We cannot predict when there might be a fluctuation.

Fluctuations are phenomena which happen, which don't really violate probability theory, but are due to the tails of the probability distributions.

They are things which are unlikely but not impossible.

Fluctuations happen from time to time.

In a sealed room, every so often a density of molecules bigger than the average will appear in some small region.

Someplace else molecules may be briefly less dense.

Fluctuations like that are hard to predict.

We can calculate the *probability for a fluctuation*.

But we can't predict when and where a fluctuation is going to happen.

It is exactly the same sort of phenomenon which we observe when flipping coins.

Flipping coins is a good example – it is probably a favorite one – for thinking about probabilities.

If I flip a coin a million times, you can bet that approximately half of them will come up heads and half will come up tails, within some margin of error.

But there will also be fluctuations.

Every now and then, if we flip the coin enough times, 1000 heads in a row will come up.

Can we predict when a thousand heads will come up? **No**.

But can we predict how often a thousand heads will come up? **Yes. Not very often!**

So that is what statistical mechanics is for: making statistical or probabilistic predictions about systems which contain elements either too small, or too numerous, or for any reason too difficult to keep track of.

That is when we use statistical mechanics, in other words we use probability theory.

We are going to go through some of the basic statistical mechanics applications.

We will also cover the theory.

We will study the laws of thermodynamics, the laws of statistical mechanics, and then see how they apply to gases, liquids, solids.

Occasionally we will look at quantum mechanical systems.

Another striking feature of statistical mechanics that ought to be mentioned is that all great physicists since the second half of 19th century were masters of statistical mechanics. Why?

First of all because it is useful, but second of all because it is truly beautiful.

Yes, it is a truly beautiful subject of physics and of mathematics.

And it is hard not to get caught up in it, hard not to fall in love with it.

Let's start this class with a brief review of the main concepts of elementary probability theory.

1.2 Elementary probability theory

What is probability? And why does it work?

Everybody has some intuitive feeling about randomness.

Yet to many people it is not far from magic.

And it is easy to get tricked or make mistakes (see discussion at the end of this part).

To say that probability works means that if the probability of some event is say 1/3, and we reproduce many times the experiment producing the event, then it will happen roughly one third of the times.

But as we know, this doesn't always work nicely.

Sometimes the event will happen more often, or less often, than it should.

These are exceptions. And by definition of exceptions they are rare.

Any way we try to pinpoint a good definition of probability, it seems to escape like quicksilver.

Our explanations will always end up involving... probability. But let's give it a try.

In nature there are experiments which, when we replicate them, keep producing exactly the same results.

They are **predictable** or **deterministic**.

They are those we mentioned earlier, appearing for instance in classical mechanics when we can apply Newton's laws.

And there are experiments whose outcomes vary. Those are said to be **random**.

The randomness comes either from some fundamental randomness of nature, like certain phenomena in quantum mechanics, or from our incomplete knowledge of the initial conditions or other things.

Yet the results display some experimental stability in the proportions of occurrences, when the experiment is repeated many times, which we will come back to.

This fundamental distinction established, to work with probability theory we need some primitive ideas about randomness, and then we need to construct a framework.

There are two primitive ideas, which we present with two illustrative examples, Figure 1:

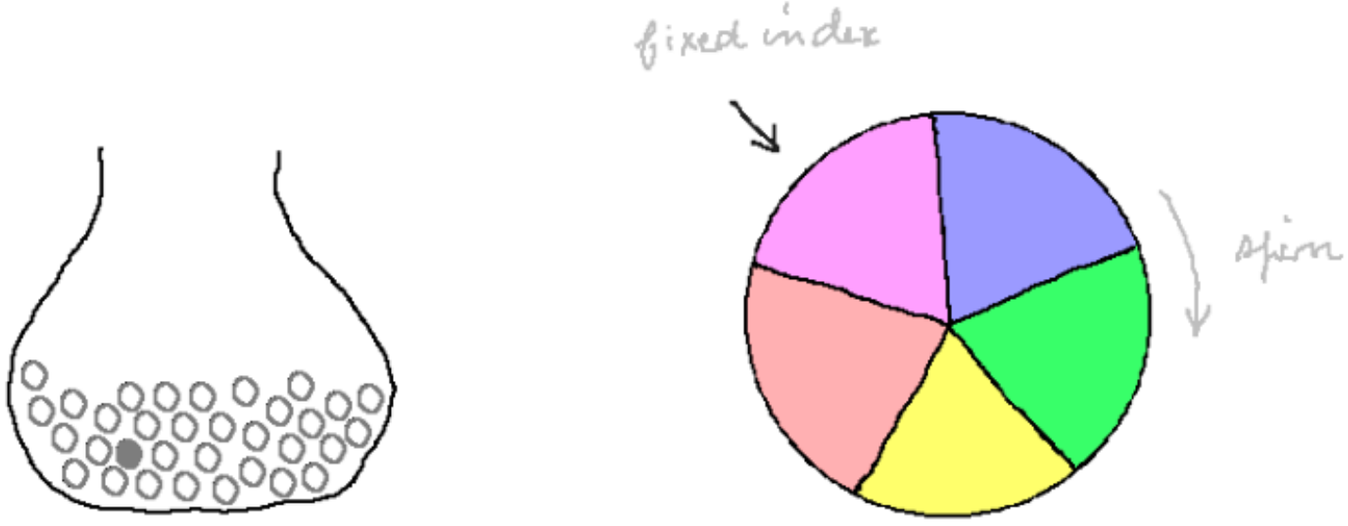


Figure 1: Rare events, and equiprobable events.

1. Rare events:

In an urn, suppose there is a *very large* number of white marbles and one black marble.

If we plunge our hand in the urn without looking, and pick a marble at random, then picking the black marble would be rare, and we can "*safely*" count on picking a white marble .

2. Equiprobable events:

If the possible outcomes of an experiment display some *symmetry*, then each event is considered as likely as any other.

Thus when we spin the wheel on the right part of figure 1, and wait until it has stopped and the fixed index points to a color, the five colors are said to be equiprobable.

Earlier, we mentioned the stability of the proportions of occurrences, in a random experiment reproduced many times.

We intuitively – and correctly – feel that it is linked to the *second idea* listed above.

But we shall see in a moment that it is also actually an instance of the *first idea* about rare events.

Let's now turn to the framework to work with a *random experiment* and probabilities.

The experiment itself is usually denoted ε .

Each time we reproduce ε , at the end of the experiment the world is in some state ω, which may be different from replication to replication.

We are interested in some aspects of ω: we may want to measure some quantity which has a numerical value, or note a color, or whatever when in the state ω .

Sometimes we are interested in only one feature of ω and sometimes in several, like picking someone at random, in a well-specified procedure, and recording both his height and weight, or level of income and academic degree, or, of a more physical flavor, measuring the position and velocity of a particle.

ε

The main source of mistakes when working with probabilities is the ill-definition of the *experiment* .

For instance, when we simply say "let's pick a person at random", it is ill-defined.

The experiment is not sufficiently specified.

Do we mean, in the room, or out of a subway entrance in New York, and in that case in which district, at what time of day, or in the United States, or on Earth?

Another example, of impossibility this time, is when we say "let's pick, at random, a point distributed uniformly over the entire line".

It is actually impossible.

There is no such thing as a density of probability with the same value from $-\infty$ to $+\infty$.

So we have to be careful.

So, once the experiment ε has been specified, we consider that each time it is performed, the world comes out in some state ω.

The set of the ω's is denoted Ω and is called the space of states, or set of states.

Probabilists call it the universe of possible states attached to performing ε .

We are interested in measuring some features, numerical data or non numerical data, about ω.

Suppose we are interested in a feature denoted X.

It depends on ω.

In other words it is a function of ω.

It is called a *random variable* and the measurement made on a given ω is denoted

$$X(\omega) \tag{1}$$

It is the result, after having performed the experiment ε once, of the measurement of X on the state ω that ε produced.

$X(\omega)$ takes its values in the set of possible outcomes of X.

Let's denote it A.

Mathematicians write that X is a function from Ω to A as follows

$$X : \Omega \to A \tag{2}$$

The set A can be finite, infinite and countable, or infinite and continuous.

For instance if the experiment ε is the throw of a die, X is the number shown on top after a throw, and the set A is

$$A = \{1, 2, 3, 4, 5, 6\} \tag{3}$$

In this case the abstract set Ω can just be taken to be the set of results of throwing the die, i.e. $\Omega = A$ itself.

Then the ω's are simply the six possible results, and we don't even have to think about X.

It is simply the identity.

Although this simplification is often appropriate, it is better – particularly when there are several random variables that we want to measure on the outcome of ε – to clearly distinguish the set of states Ω that is once and for all attached to ε and the various sets of outcomes of random variables.

Let's stress again that A doesn't have to be numerical.

Our die, for instance, could have faces painted with different colors rather than bearing numbers.

Once ε, Ω, X and A have been defined, the last fundamental concept to introduce in the framework is a probability P.

Mathematicians technically talk about a measure of probability P on Ω that is σ-additive, etc.

And they introduce it before any random variable and they brings in unnecessary complications.

In this review of elementary probability we don't need to go into such extra complications.

Furthermore in an elementary approach, it may obscure the link with probabilities as we intuitively know them.

P is defined such that any subset of Ω – called an event – has a probability.

In the case of the die, it is particularly simple.

Each ω is itself an interesting event.

There are six of them.

And if the die is well balanced, then they are equiprobable.

Thus we write

$$P\{X = 5\} = \frac{1}{6} \tag{4}$$

meaning that the probability of getting a 5, when throwing the die, is 1/6.

Events are any subsets of Ω, not only the ω's themselves.

We can also write

$$P\{X \leq 2\} = \frac{1}{3} \tag{5}$$

The principle that *symmetry*, or equivalence somehow, between the possible states ω implies equiprobability – in other words, whatever makes them different doesn't affect their propensity to occur – is often invoked to figure out the distribution of probability *P* attached to an experiment.

Another possible way to figure out *P* is through a large number of replications of the experiment.

We will explain it in a moment.

Finally there may be various probabilistic calculations which we can also make to relate the distribution of a random variable *X* to those of other random variables we already know.

This pertains to the calculus of probability.

And we shall work out many examples in this class.

Our framework is now complete.

It consists of a random experiment ε, a big set Ω of possible states of the world after the performance of ε, and a probability *P*.

$$\text{framework} = [\mathcal{E}, \Omega, P] \tag{6}$$

And we are interested in measuring various random variables *X*, *Y* , *Z*, etc., after having performed ε.

Let's now simplify a bit the setting and the notations.

For the time being, the set of possible states will be

$$\Omega = \{\omega_1, \omega_2, \omega_3, \cdots, \omega_n\} \tag{7}$$

that is a finite set of outcomes of ε.

The states are indexed by i running from 1 to n.

For example, when flipping a coin once, ω_1 = heads, and ω_2 = tails.

Later on, we will extend this to an infinite countable, and then even an infinite continuous set Ω

In this last case, probabilities will be need to be replaced by *densities of probability*.

Instead of considering $P\{X = x\}$, which will usually be equal to 0, we will always consider

$$P\{X \in [x, x + dx]\} = p(x)dx$$

And, following custom, we will often still just write it as *P(x)dx*, keeping in mind that each random variable has its own density.

Staying with a finite set Ω, the probabilities of the single states ω_i's will simply be denoted

$$P(i) \tag{8}$$

They must satisfy

$$P(i) \geq 0 \tag{9}$$

$$\sum_i P(i) = 1$$

Indeed, probabilities are positive numbers.

And the total probability, when we add up the probabilities of all possibilities, should be equal to one.

When performing ε we certainly should get some result - **something happens!!**

Probabilities have all sorts of interesting, beautiful and sometimes surprising properties.

The most useful one for us in this class is the *law of large numbers*.

Here is what it says.

Suppose that we either make many replicas of the same system, or do the same experiment ε over and over a large number N of times, and we count how many times we get the i-th outcome ω_i.

That is some count that we denote N_i.

Then the law says that

The frequentist view

$$\lim_{N\to\infty} \frac{N_i}{N} = P(i) \tag{10}$$

This is a statement about probabilities, which can be stated more precisely and rigorously within the framework [ε , Ω, P].

But let's approach it at an intuitive level.

It says that when we replicate ε a larger and larger number of times, and measure the experimental frequency of occurrence of the *i*-th outcome, this experimental frequency gets closer and closer to the actual probability *P(i)*.

For instance, if we toss a coin a thousand times, the frequency of heads will be close to 1/2.

If we throw it 10000 times, the frequency will be even closer to 1/2.

In each case, it is only a probabilistic statement.

There can be – and in fact most of the times there will be – a discrepancy.

That discrepancy is itself a random variable.

But it will have a distribution more and more concentrated, relatively to its range, around 0.

We said earlier that the convergence of experimental frequencies toward their corresponding theoretical probabilities is actually an instance of the rare event idea, see figure 1 and its comments.

The law of large numbers is neither magic, nor some kind of eerie principle of nature.

It stems from the fact that in an urn with one black ball and very many white balls, if we pick one at random, we can assume safely that we will pick a white one.

It is not always the case, but it will be extremely rare to pick a black ball.

And for all practical purpose it can be neglected.

Let's see why, in the case of tossing coins, the law is a *simple result in numbering*.

Consider the experiment $\mathcal{F}$ which consists in tossing the coin 1000 times, i.e. repeating $\mathcal{E}$ 1000 times.

The space $\Omega_{\mathcal{F}}$ attached to $\mathcal{F}$ has 2^{1000} elements – that is a huge number.

Each are equiprobable.

When we perform $\mathcal{F}$ once, i.e. when we repeat $\mathcal{E}$ a thousand times, we pick one element in $\Omega_{\mathcal{F}}$.

It turns out – and it is not hard to show, although we won't do it – that most elements in $\Omega_{\mathcal{F}}$ contain about as many heads as tails — all "near" the most probable.

View them as the white balls in the urn if you like.

And the black balls would be very few.

So when we pick one, we pick a white one.

To try to shed even more light on the phenomenon, rather than do some combinatorics, consider the 16 possible results, displayed below, of throwing the coin four times.

You can check that there is only one result with zero head.

Four results with 1 head.

Six results with 2 heads.

Four results with 3 heads, and one result with 4 heads.

So there is a kind of concentration around an equal number of heads and tails.

The counts of the number of heads actually correspond to the so-called **Pascal triangle**.

They are also the coefficients in the development of the polynomial $(a+b)^N$.

$T,\ T,\ T,\ T$
$T,\ T,\ T,\ H$
$T,\ T,\ H,\ T$
$T,\ T,\ H,\ H$
$T,\ H,\ T,\ T$
$T,\ H,\ T,\ H$
$T,\ H,\ H,\ T$
$T,\ H,\ H,\ H$
$H,\ T,\ T,\ T$
$H,\ T,\ T,\ H$
$H,\ T,\ H,\ T$
$H,\ T,\ H,\ H$
$H,\ H,\ T,\ T$
$H,\ H,\ T,\ H$
$H,\ H,\ H,\ T$
$H,\ H,\ H,\ H$

Exponent	Pascal's Triangle	Binomial Expansion
0	1	$(a+b)^0 = 1$
1	1 1	$(a+b)^1 = 1a + 1b$
2	1 2 1	$(a+b)^2 = 1a^2 + 2ab + 1b^2$
3	1 3 3 1	$(a+b)^3 = 1a^3 + 3a^2b + 3ab^2 + 1b^3$
4	1 4 6 4 1	$(a+b)^4 = 1a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 1b^4$
5	1 5 10 10 5 1	$(a+b)^5 = 1a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + 1b^5$
6		

The concentration about half and half is more marked, of course, when *N* is larger than 4, and it grows more and more marked as *N* increases.

That is what the law of large numbers is about.

In formal probability theory, it is stated more rigorously than we have done here.

It is proved via an intermediate result called Chebyshev inequality. $P(|X - m_x| \geq a) \leq \frac{D_x}{a^2}$

It is not particularly hard, and is rather elegant.

But it is outside what we want to do in this review

The law of large numbers, expressed by equation (10), says that the ratio N_i/N converges toward $P(i)$ when N gets very large.

In other words,

> *when we have repeated an experiment a large number of times, we can use the experimental frequency N_i/N of occurrence of the i-th outcome as an estimate of $P(i)$.*

We use this result all the time.

Now let's go back to a random variable we want to measure, which is not the outcome of ε itself.

To get closer to physics concepts and notations, let's call the random variable F.

So let's suppose that there is a quantity, denoted $F(i)$, that is associated with the i-th state ω_i.

Recall expression (7) defining the set of states. $\Omega = \{\omega_1, \omega_2, \omega_3, \cdots, \omega_n\}$

F can be some meaningful physical quantity.

We can also make it up.

For example if our system is heads and tails, and nothing but heads and tails, we could assign

$$\begin{aligned} F(\text{heads}) &= +1 \\ F(\text{tails}) &= -1 \end{aligned} \tag{11}$$

If our system has many more states, we may want to assign a larger number of possible values taken by F – not necessarily the same number as the number of elements in Ω though.

F is simply some function of the states.

We already mentioned this in expression (2), let's write it again

$$F : \Omega \to A \tag{12}$$

The random variable F acts on the set of states Ω and takes its value in the set A.

In the case of the coin, $\Omega = \{H, T\}$, and the set A in which F takes its values is $\{+1, -1\}$.

Thus, we have made up a numerical random variable – or *measurement* – attached to flipping a coin.

As said, *F(i)* can also be some meaningful physical quantity.

It could be the energy of the *i*-th state ω_i.

Given the state in which is some system, it has an energy.

Its measurement would perhaps be called in that case

$$E(i) \tag{13}$$

Or it could be the momentum of the state.

We would have to choose a good notation not interfering with probabilities.

Or it could be something else.

It could be whatever we happen to like to measure on our system.

In any case, then an important quantity is the *average* of *F(i)*.

We will use the quantum mechanical notation for it, even though we are not doing quantum mechanics.

It is a nice notation.

Physicists tend to use it all over the place.

Mathematicians hate it, so physicists want to use it even more!

We just put a pair of brackets around F to mean its average.

It is defined as follows

$$\langle F \rangle = \sum_{i}^{n} F(i)P(i) \qquad \text{standard definition of the average} \tag{14}$$

In words, it is the average of the values *F(i)* weighted by their respective probabilities *P(i)*.

Notice that the average of *F(i) does not* have to be any of the possible values that *F* can take.

For example, in the case of the coin, where $F(H) = +1$, and $F(T) = -1$, and we flip it a million times, and the probability is 1/2 for heads and 1/2 for tails, the average of *F* will be 0.

It is **not** one of the possible outcomes of *F* in the experiment.

Yet it is its average.

There is **no rul**e that the average of a measure should be one of its possible experimental values.

Thanks to the law of large numbers, we can write equation (14) another way.

$$\langle F \rangle \approx \sum_{i}^{n} F(i) \frac{N_i}{N} \tag{15}$$

This approximate equality becomes a true equality in the limit when there is a large number of measurements.

That is it for our mathematical preliminary.

We need to know what a random experiment is, what a probability is, what a random variable or random measurement is, and what is an average, because we will use them over and over.

Now we shall go deeper into the link between probability and symmetry.

And we shall introduce time.

1.3 Probability, symmetry and systems evolving with time

Let's start with coin flipping again.

For the usual coin, the probability for heads is usually deemed to be 1/2, and the probability for tails is usually also deemed to be 1/2.

But why do we think that?

Why is it 1/2 and 1/2?

What is the logic there?

In this case it is **symmetry**.

Of course no coin is perfectly symmetric.

Making a little mark on it, however tiny, to distinguish heads and tails biasses it a little.

But apart from that tiny bias, for example a small scratch, the coin is symmetric.

Heads and tails are **symmetric** with respect to each other.

Therefore there is no rationale, when we flip a coin, for it to turn up heads more often than tails.

So it is symmetry quite often – we might even say always, in some deeper sense, but at least in many cases – which dictates probabilities.

Probabilities are usually taken to be equal for configurations which are related to each other by some symmetry.

Symmetry means if you act with a symmetry, if you reflect everything, if you turned everything over, that the system behaves the same way.

Another example besides coin tossing would be die throwing.

Now the space Ω, instead of having two states, has six states.

When we throw the die, and it has finished rolling, the state it is in is the face showing up.

To stress that states don't have to be numerical values, let's consider a die with colored faces, figure 2.

So here we don't keep track of numbers.

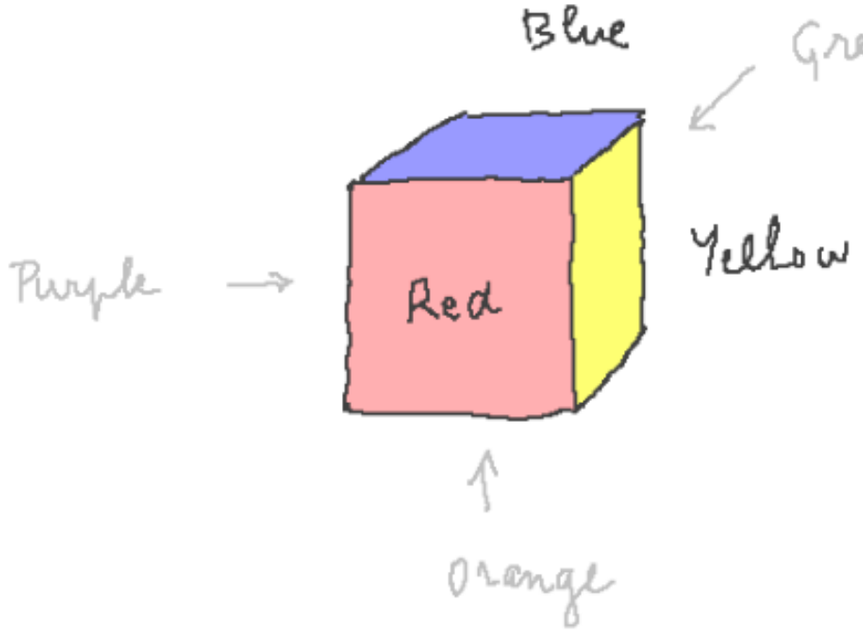


Figure 2: Die with colored faces.

We keep track of colors.

The space of states is

$$\Omega = \{\text{red, yellow, blue, orange, green, purple}\} \tag{16}$$

In other words, ω_1 = red, ω_2 = yellow, ω_3 = blue, etc.

What is the probability that after a throw the die turns up, for instance, blue like in figure 2? It is 1/6.

Why? Because there are six possibilities.

They are all symmetric with respect to each other.

We use the principle of symmetry that tell us that the *P(i)*'s are all equal, therefore

$$P(i) = \frac{1}{6} \qquad \text{for all } i \tag{17}$$

But what if the die is not symmetric?

For example, if it is weighted in some unfair way.

Or if it is been cut with faces that are not nice parallel squares, thus not making it an honest cube.

Then symmetry won't tell us anything.

There are cases where we may be able to use some deeper underlying theory in which symmetry plays a role.

But, in the absence of such deeper information, there is no theoretical way to figure out the probabilities.

In that case we resort to experiment.

Do the experiment ε a billion times; keep track of the numbers; assume that things have converged.

And that way we measure the probabilities and thereafter we can use them.

So either we have some symmetry, or some underlying theory like in quantum mechanics, that tell us what the probabilities are, or we do experiments to get them.

Whatever the case, we end up with probabilities.

Statistical mechanics tends to rely mostly on symmetry as we will see.

But if there is no symmetry to guide us in our determination of probabilities then it is experiment.

Now there is another possible answer.

This answer is often frequently invoked and it is a correct answer under the circumstances.

It can have to do with the evolution of a system, the way a system changes with time.

Let's see some examples.

Let's take our six-sided polyhedron, and assume that it is not a nice symmetric cube.

Furthermore suppose this object has the following funny habit: when it is in one state, it then jumps to another definite state, then to another definite state, etc.

It is called the law of motion of the system.

The *law of motion of a system* is that wherever it is – i.e. in whatever state, also called a *configuration* – at one instant, at the next instant it will be in some other state according to a definite rule.

The time spent in each state is some seconds or some microseconds or whatever.

And they are all equal.

And for simplicity the jumps are instantaneous

So we imagine a discrete sequence.

And let's suppose there is a genuine law that tells us how this funny cube moves around.

A law of motion is a rule telling us what the next configuration will be, given where we are.

It is a rule of updating configurations.

Here is an example.

The letters are abbreviations for the colors in expression (16).

$$
\begin{aligned}
\mathrm{R} &\longrightarrow \mathrm{B} \\
\mathrm{B} &\longrightarrow \mathrm{G} \\
\mathrm{G} &\longrightarrow \mathrm{Y} \\
\mathrm{Y} &\longrightarrow \mathrm{O} \\
\mathrm{O} &\longrightarrow \mathrm{P} \\
\mathrm{P} &\longrightarrow \mathrm{R}
\end{aligned}
$$

Given the configuration, that is the state, in which the system is at any time, we know what it will be next.

And we know what it will continue to do.

Of course we may not actually know the law.

Maybe all we know is that there is a law of the type above.

Anyway, let's draw the one above as a diagram, Figure 3.

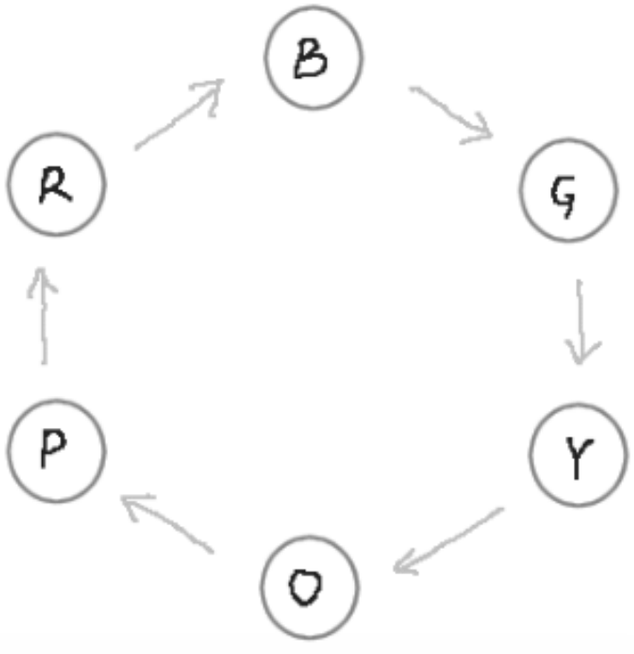


Figure 3: Law of motion.

And we make the assumption that the same amount of time is spent in each state.

We are not assuming that the cube has any symmetry anymore.

It may not be symmetric at all.

Imagine that the faces are not parallel, they are quadrilaterals but not squares.

Some are big, some are small.

They may even be not quite flat, etc.

But figure 3 is the rule to go from one configuration to another.

And each step takes, let's say, a microsecond.

We might have no idea when and in which state the system began.

But suppose our job is to catch it at a particular instant and ask what the color is.

Even though we don't know the starting point, we can still say that the probability for each state is 1/6.

It is not related to any symmetry of the die, and as we said the die may not be symmetric at all.

But it is a consequence of the symmetry of the time structure.

As it passes through the sequence of states, the die spends one sixth of its time red, one sixth of its time blue, one sixth of its time green, etc.

If we don't know where it starts, and we take a flash snapshot, the probability will be 1/6 for each state.

It doesn't really depend on knowing the detailed law.

For example the law could have been different. Figure 4 shows a different example

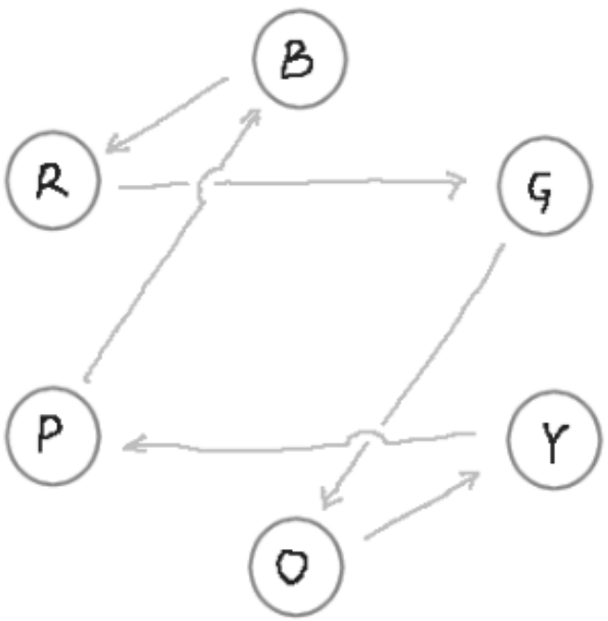


Figure 4: Another law of motion.

This shares with the previous law that there is a closed cycle of events in which we pass through each color once before we cycle around.

We may not know which law of nature is for the system, figure 3 or figure 4 or even another law which makes a complete cycle through all the colors, but we can say again that the probability will be 1/6 for each one of them.

So this prediction of 1/6 doesn't depend on knowing the starting point, and doesn't depend on knowing the law of physics.

It is just important to know that there was a particular kind of law.

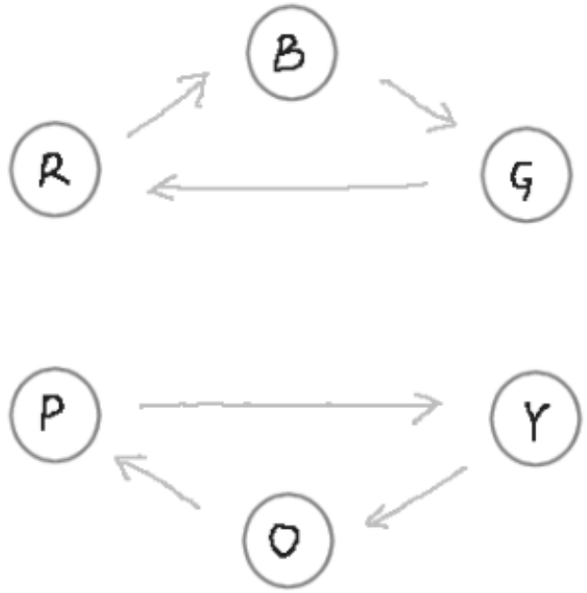


Figure 5: Law of motion with two cycles.

Are there possible laws for the system which will not give us 1/6? **Yes.**

Let's write another law, Figure 5

This rule says that if we start with red, we go to blue.

If we are in blue, we go to green and if we get to green, we go back to red.

Or if we start with purple, we go to yellow.

Yellow goes to orange.

And orange goes back to purple.

Notice in this case, when we are on one of the two cycles, we stay on it forever.

Suppose we knew we were on the upper cycle.

It doesn't matter knowing in which state we started, but if we knew that we started on the upper cycle somewhere, then we would know that there is a 1/3 probability to be in the red state when we flash the system, 1/3 probability to be in blue, 1/3 probability to be in green.

And of course there is 0 probability to be in yellow, orange or purple, since we are not on their cycle.

The probabilities 1/3, 1/3 and 1/3 to be in red, blue or green are also called the *conditional probabilities*, given that we know we are on the first cycle

On the other hand, we could have started on the second cycle.

We could have started with purple.

Or we might not know exactly where we started, just that we started on the lower cycle.

Then the probabilities are: purple 1/3, yellow 1/3 and orange 1/3.

And again, of course, 0 for the other colors.

Now what about a more general case?

It might be that we know with some probability, that we started on the upper cycle and with some other probability on the lower cycle.

In fact, let's give the cycles labels.

The upper cycle we call $+1$. And the lower triangle we call -1.

It is just attaching to them a number, a numerical value.

If we are on the upper cycle, something or other is called $+1$, and below it is -1.

Now we have to append something that tells us how we get from some place else.

It doesn't follow from symmetry.

And it doesn't follow from cycling through the system.

It is the probability that we are on cycle $+1$, or its complement to one that we are on cycle -1.

Where might that come from?

Well, flipping somebody else's coin might decide which of these two cycles we shall be on.

And the coin doesn't have to be balanced.

So we have two more probabilities to consider:

$$P\{\text{ cycle} = +1\}$$
$$P\{\text{ cycle} = -1\} \tag{18}$$

And they must add up to 1.

These two probabilities are not probabilities for individual colors.

They are probabilities for individual cycles.

Now what is the probability for blue?

From elementary probability theory, the product of the probability that we are on cycle +1 times the conditional probability that we get blue, conditioned on knowing that we are on cycle +1.

$$P\{\text{blue}\} = P\{+1\}P\{\text{ blue} \mid +1\} \tag{19}$$

where $P\{\text{blue}\}$ is the overall probability to get blue, $P\{+1\}$ is an abbreviation for $P\{$ cycle = +1 $\}$, and $P\{$ blue |+1$\}$ is the standard notation for the conditional probability to get blue given that we already know that we are on the first cycle.

Since the three conditional probabilities, once we are on cycle +1, are 1/3, formula (19) becomes

$$P\{\text{blue}\} = P\{+1\}\frac{1}{3} \tag{20}$$

And we have similar formulas for the other colors.

Of course, in the case where $P\{+1\} = 1/2$, we get back to $P\{\text{blue}\} = 1/6$.

So in the case where there are several different possible cycles, we would need to supply another probability that we must get from somewhere else.

The example we just went through, with several cycles, is what we call having a *conservation law*.

In our example, the conservation law would be just the conservation of the number assigned to the cycle.

For red, blue and green we have assigned the value +1.

That +1 could be the energy, or it could be something else, whatever.

Let's just think of it as the energy to keep things familiar.

The energies of the three configurations in the upper cycle might all be +1.

And the energies of the three configurations in the lower cycle might be −1.

The point is that, because the rule of motion keeps us always on the same cycle, the quantity – energy or whatever – is conserved.

It doesn't change. That is what a conservation law is.

In summary, a conservation law says that the configuration space divides up into cycles, or trajectories if they don't loop, and that in each of them a certain characteristic of the configuration doesn't change.

While the system jumps from configuration to configuration as time passes, it stays on the same cycle or trajectory.

As a consequence the quantity, whose value is determined only by the cycle or trajectory, is conserved.

The cycles don't have to have equal size, figure 6.

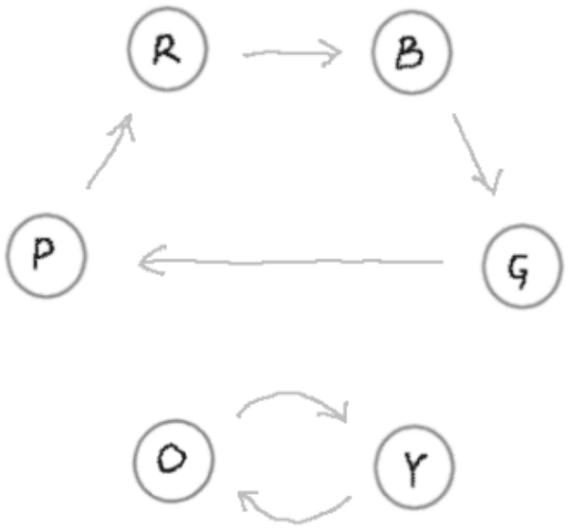


Figure 6: Example of cycles with different size.

We could have a conservation law in this space of configuration, if there is a quantity with one value for the upper cycle, and another value for the lower cycle.

The number of states don't have to be the same in each cycle.

But still there can be a conservation law.

And again somebody would have to supply for us some idea of the relative probabilities of the two cycles.

Where that comes from is part of the study of statistical mechanics.

That is what determines probabilities in statistical mechanics: some probabilities, coming from somewhere, which tell us the probabilities for different conserved quantities, and then going from configuration to configuration through the configuration space.

1.4 Questions/answers session - *to make us think about what we have said so far*

Question:

So, we assume that once we are on a given cycle, the probabilities of all the configurations on that cycle are the same?

Answer:

Right. That comes from the fact that the time spent in each state is the same.

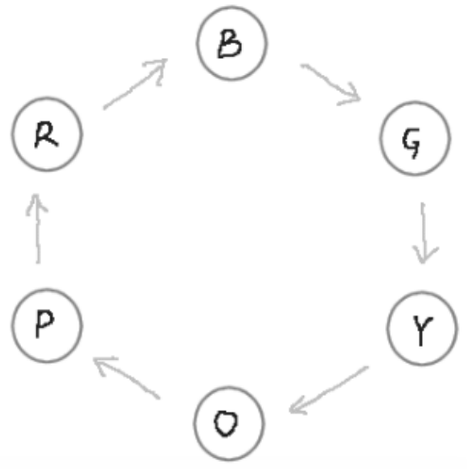
We are in classical physics.

Everything is deterministic.

The randomness that we introduce is only a consequence of our imperfect knowledge of the initial conditions of the system, or of the exact time at which we observe it.

Because we don’t have precise enough instruments, or because we are lazy, we don’t know exactly in which state the system was at time zero, or we don’t know exactly the time at which we take the measurement of the color.

Let's focus on figure 3 where there is only one overall cycle.



Suppose we take a measure at time t.

Suppose also we know that the system started in state blue, and it stays one microsecond in each state, and *t = 2 minutes 17 seconds 754 milliseconds* and *123 microseconds*.

Then it is easy to calculate that the system should be in orange.

But if the precision of our timing is only one millisecond, taking a flash of the system amounts to a random experiment where the probability of each of the six states is 1/6.

We have assumed that the flash is instantaneous and there is no blur between several colors.

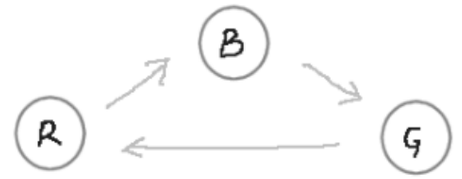
That's the kind of circumstance that we are talking about

If they are both closed systems there will be two conserved quantities, the energy of the left system and the energy of the right system.

Question:

If we take two pictures, presumably we will be able to check that we stayed in the same cycle?

Answer:



Oh yes. If there are several distinct cycles, like for instance in figure 5, and we take two pictures one after the other, the two colors we shall obtain will be on the same cycle.



In fact, once we have made a first picture, we know in which cycle we are, and there are no more probabilities for cycles to heed.

Now the probabilities are uniform over the configurations of the cycle we are on.

In other words, once you determine the value of some conserved quantity, then you know it.

And then you can reset the probabilities for the various possible configurations of the system.

Let's talk about energy for a minute.



Figure 7: Closed system.

Suppose we have a closed system.

To represent the close system we just draw a box

Now closed means that it is not in interaction with anything else and therefore can be thought of as a whole universe unto itself.

It has an energy.

The energy is some function of the state of the system, whatever determines the state of the system.

Now let's suppose we have another closed system which is built out of two, identical or not identical, versions of the same thing.



Figure 8: Two closed systems.

They will both be separately conserved. Why?

Because they don't talk to each other.

They don't interact with each other.

The two energies are conserved.

And the probability frameworks for each system will just be independent experiments and analyses.

But now supposing they are connected by a little tube which allows energy to flow back and forth, figure 9.

Then there is only one conserved quantity, the total energy, which is sort of split between the two systems.



Figure 9: Two systems connected.

We can ask: given a total amount of energy, what is the probability that the energy of one subsystem is one thing and the energy of the other subsystem is the other?

If the two boxes are equal, we would expect that on the average they have equal energy.

But we can still ask what is the probability that energy in the left box has such and such value, given some overall piece of information.

So, if we consider only the system on the left in figure 9, that is a circumstance where the probability for which cycle we are on may be determined by thinking about the system as part of a bigger system.

And we are going to do that.

That is important.

So in general, you need some other ingredients besides just cycling around through the system, represented in diagram 5 or diagram 6, to tell you the relative probabilities of conserved quantities.

So we are off and flying in our statistical mechanics type of thinking!

Now let's turn to some fundamental laws, and in particular to some forbidden.

1.5 Bad laws of motion and the –1st law of physics

By bad laws, we don't mean in the sense of DOMA or any of those kind of laws, but in the sense that the principles of physics don't allow them.

If you know anything about classical mechanics, you know that a bad law of motion is one that violates the conservation of information – the most primitive and basic principle of physics.

Conservation of information is not a standard conservation law that says that there are certain quantities which are conserved when a system evolves over time through a cycle or any other kind of trajectory.

Conservation of information is the principle which says that we can keep track of our trajectory both going forward and going backward.

Let us discuss it briefly.

Figure 10 is an example of a bad law.

There are six states, or configurations, for the system, but the rule says that wherever we are, the next state is red.

Even if we are at red, we go to red.

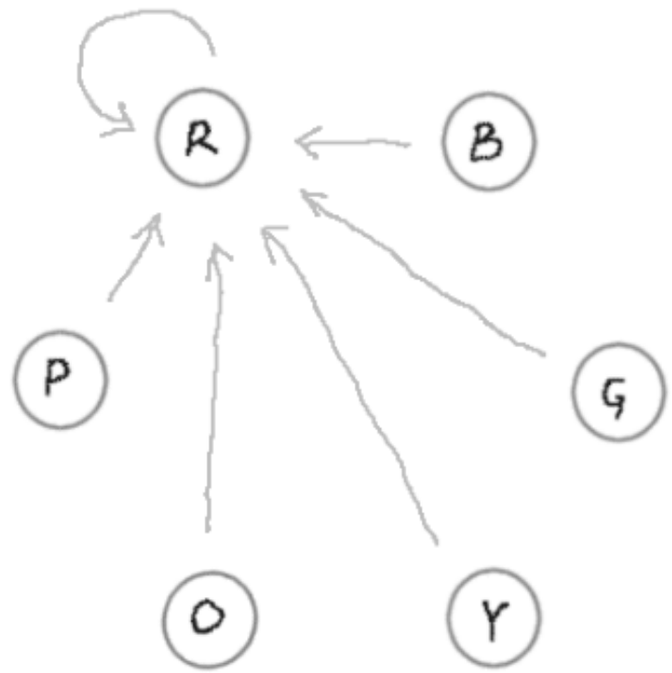


Figure 10: Example of bad law. It keeps track of where we go, but not of where we come from.

We readily notice that it is **not** reversible.

We can follow a trajectory when the time passes forward.

It is very simple: we always end up at red and then stay there.

But if we look at the trajectory in reverse, with the time passing backwards, it doesn't correspond to a well defined law

When we are at red, where would the law say we should go?

It wouldn't say, because there are several arrows pointing at red in figure 10, and therefore when we reverse time several arrows would leave red for different states.

That would not be a law at all, because in classical physics a law of evolution must be deterministic.

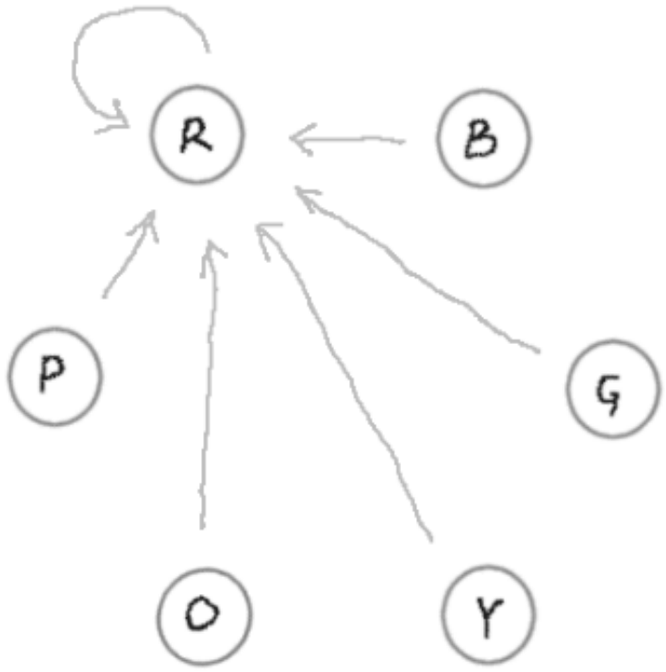
In other words, the law in figure 10 is bad because it loses track of where we started and therefore is not reversible.

We could make it more complicated, and such that it still loses track of the past, if we wanted to.

But the story would be the same.

A law must be deterministic, and it must be reversible.

This second characteristic is equivalent to saying that we don't lose track of the past.



We can express the principle simply:

> *A good law of motion is one such that for each state, one and only one arrow arrives at it, and one and only one arrow leaves it.*

The laws in figures 3, 4, 5 and 6, were all good laws according to this principle.

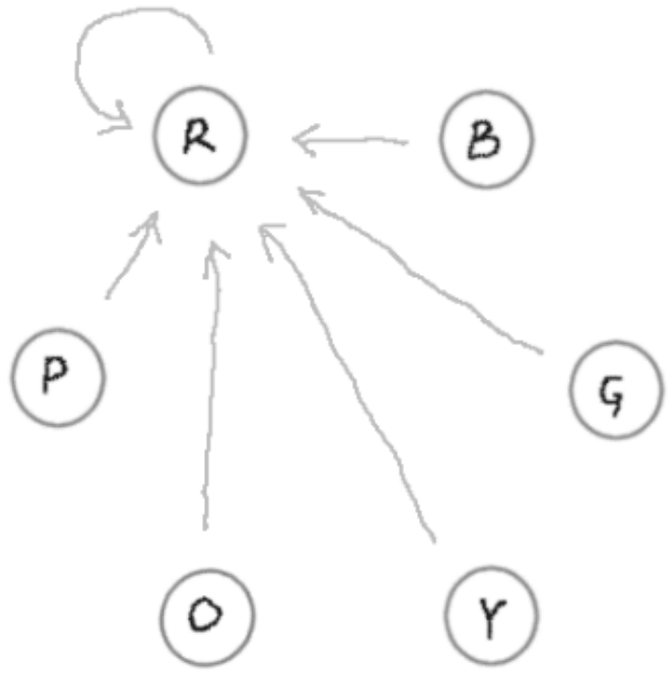
In any of those examples, if we start from any state S_1 and go, say, 56 steps forward and end up in state S_2, then starting from S_2 and going backward 56 steps we will end up in state S_1.

They don't lose information.

The law in figure 10 loses information.

That is exactly the kind of thing that classical physics doesn't allow.

Quantum physics also doesn't allow laws of motions that are not deterministic in the future, or not deterministic in the past, i.e. not reversible.



There is no name for this law because it is so primitive that everybody always forgets about it.

But I and some others call it the **1st law of physics**.

I wish that it would catch on, that people would start using it, because it is the most basic law of physics, that information is never lost.

The 1st law says that distinctions or differences between states propagate with time.

Nature never loses track of the difference between two states.

In principle, if we have the capacity to follow the system, from wherever we are we will always know exactly where we go, as well as where we come from.

Because we don't have the appropriate instruments, or we are too lazy, we may lose track, but nature doesn't have this problem and keeps distinctions.

And in principle we could too because nature – by definition – doesn't hide anything from us.

To a physicist or a natural philosopher it doesn't make sense to say that some things are, or have been in the past, but we cannot one way or another be aware of them.

Thus we see that the 1st law is not a stringent requirement.

It is at the foundation of what knowledge means, as opposed to fictitious beliefs.

Let's see an apparent counterexample from mechanics.

Friction seems to be such a case.

If we consider an object sliding with friction, it eventually comes to rest.

It is sort of equivalent to always ending up at red in figure 10.

That seems like a violation of the law that tells us that the distinctions have to be preserved.

But of course it is not really true.

What is really going on is that as the object slides, decelerates and stops, it is heating up the surface on which it slid.

If we could keep track of every molecule we would find out that the distinctions between trajectories is recorded.

In other words, two different starting states of the universe would lead, when taking into account everything, to two different ending states of the universe, even if the object ends up at the same place.

Let's imagine now that there was a fundamental law of physics for a collection of particles.

The particles are labeled by an index n, they each have a position x_n, possibly multi-dimensional, depending on the time t.

And they satisfy the equation (Newton's second law F = ma)

$$\frac{d^2 x_n}{dt^2} = -\gamma \frac{dx_n}{dt} \tag{21}$$

The left-hand side is acceleration.

And the right-hand side corresponds to a frictional force opposing motion.

The factor γ is the viscous drag.

We could put masses on the left-hand side multiplying the second derivative, but that would not change anything for our purpose, so let's take them all equal to 1.

Again this is friction.

Equation (21) has the property that if we start with a moving particle it will very quickly come almost to rest.

It will exponentially come to rest, and that will happen fast if there is a significant drag.

Imagine that all the particles, in a gas for example, satisfied this law of physics.

It is perfectly deterministic, back and forth.

It tells you what happens next, and it is reversible.

But it has the unfortunate consequence that every particle almost comes to rest.

That sounds odd.

It sounds like no matter at what temperature we start the room, it will quickly come to zero degree.

It does not happen that way.

Equation (21) is a perfectly good differential equation.

But there is something wrong with it from the point of view of conservation of energy, and from the point of view of thermodynamics.

If a closed system starts with a lot of kinetic energy – we usually call it *temperature* – it doesn't evolve to zero temperature.

That is not what happens.

And it is not only a violation of energy conservation, it also looks like a violation of the second law of thermodynamics.

Equation (21) describes a process where things get simpler.

We start with a random bunch of particles moving in random directions, and we let it run, and they all come to rest.

What we end up with is simpler and requires less information to describe than what we started with.

It is very much like, in figure 10, everything going to red.

The second law of thermodynamics generally says things get worse.

Things get more complicated not less complicated.

So even though it is mathematically sound, deterministic in the future and in the past, equation (21) is not a good law of physics in view of the law of conservation of information.

We saw that one way to express the 1st law, or law of conservation of information, is to say that every state must have one arrow coming in and one arrow going out.

There is another important way to formulate it.

1.6 Conservation law and probability

Let's consider a system and the collection of states or configurations, indexed by i, it can be in.

And we assign probabilities, $P(1)$, $P(2)$, $P(3)$, etc. that is, probability, when we look at it, that the system be in state 1 – earlier called ω_1 –, probability that the system be in state 2, probability that it be in state 3,... for some subset of the complete set of possible states.

All the others, we say that they have probability zero.

For example, we can take our colored die in figure 2 and assign red, yellow and blue, each probability 1/3.

And the other three colors have probability zero.

Where did we get that information from?

It doesn't matter.

We got it from somewhere.

Somebody secretly told us in our ear:

"It is either red, yellow or blue, but I'm not going to tell you which".

The upper face of the die actually shows one color, that is the die is in one specific state, but for some reason we don't know it.

We only know the partial information given by the whispering chap.

We could also imagine the die having three faces with three shades of green, and the others with three shades of purple.

And, because we did not take a good enough look, we only got the general color of the state, not the precise shade.

And to avoid being in a Bayesian setting where probabilities don't correspond to any reproducible experiment – and therefore, to many of us, would be meaningless – let's imagine that at least theoretically we could play the game many times and those three colors would turn up randomly with three equal probabilities, because the die is loaded or some other reason.

Now we follow the system as it evolves in time.

The system evolves according to whatever law of physics, as long as it is an allowable law of physics.

After a while, suppose we are capable of following it in detail.

We are no longer constrained by our laziness or inaccurate instruments in this case.

Then what are the probabilities at a later time?

Well, if we don't know what the laws of physics are, we can't say of course.

But we can say one thing: we can say there are three states with probability 1/3, and three states with probability zero.

They may get reshuffled, which ones are probable and which ones are improbable.

But after a certain time, there will continue to be three states which have each probability 1/3 and the rest probability zero.

Let's be very precise, because when talking about probabilities it is easy to slip into confusion.

If we reproduce the experiment many times here is what will happen.

And we suppose we have omnipotent observational powers of colors and time.

At first the three colors red, yellow and blue are the possible initial colors.

So they will each appear roughly one third of the time.

And after a specific time t, that is a specific numbers of steps, the three corresponding states after time t will also appear – necessarily – roughly one third of the time.

There is no loss of information, even though the initial conditions are random between three states from our point of view.

So in general we could characterize these information conserving series as follows.

We have a total number of states N, and a number of possible states M.

$$\begin{aligned} N &= \text{total number of states} \\ M &= \text{number of possible states} \\ M &< N \\ P &= \frac{1}{M} \end{aligned} \tag{22}$$

The number of possible states stays the same over time – even though they may get reshuffled among the N states as time passes.

And of course, at a given time, the probability of any of the possible states at that time is $P = 1/M$.

That is a different characterization of the information conserving law.

It introduces probability, but it specifies that the randomness over initial conditions is carried along with time, with no further loss.

Then the number of states which have non-zero probability will remain constant and their probabilities will remain equal to $1/M$.

As we have described, these possible states may reshuffle as time passes, but the number with non-zero probability will remain fixed.

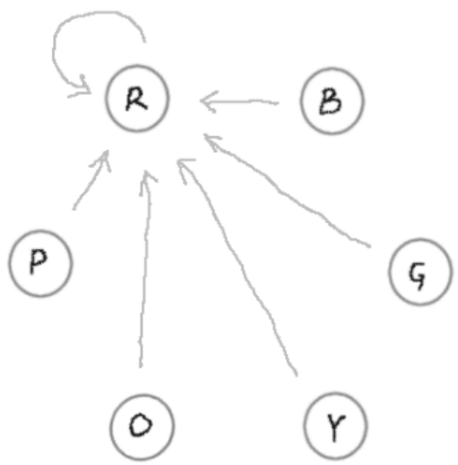
For the information non-conserving laws every configuration goes to red.

In figure 10, we may start with a probability distribution that is for instance 1/3 for red, green and purple and 0 for the others.

And then a little bit later there is only one state that has a non-zero probability. And that is red.

So we now have another way to describe information conservation.

And we can quantify that.



Let M be the number of states which are possible, and let's assume they all have equal probability.

These states have a name.

They are called the **occupied** states – states which have non-zero probability, with equal probability.

Again these may shift around as time passes.

But M doesn't change.

What is M characterizing?

M is characterizing our **ignorance**.

The bigger M is, the **greater** is our ignorance.

If M is equal to N, that means equal probability for every configuration, or **maximal** ignorance.

If M is equal to $N/2$, that means we know that the system is in one out of half the states.

We are still pretty ignorant, but we are not that ignorant.

We are less ignorant.

What is the minimum amount of ignorance we could have?

That is when we know precisely, at any time, in which state the system is.

In that case $M = 1$.

That is when we have no ignorance resulting from one cause or another.

You could say that we are then the omnipotent observer that we talked about in explaining the probabilities attached to the occupied states.

So M , in relation to N , is a measure of our ignorance.

And, associated with it, is the concept of **entropy**.

1.7 Entropy

We have come to the concept of entropy.

Notice that entropy is coming before anything else.

Entropy is coming before temperature.

It is even coming before energy.

Entropy is more fundamental in a certain sense than any of them.

But it is also different from temperature or energy in the sense that it is not an absolute characteristic of the system – irrespective of what the observer is or does – but is a characteristic of the system plus the observer.

This will become clear in a moment

The entropy is denoted by the letter S, and it is **defined** as the logarithm of M:

$$S = \log M \tag{23}$$

It is the logarithm of the number of states that have an appreciable probability, more or less all equal, in the specific circumstances that we described.

When there is no loss of information, the entropy S is conserved.

All that happens is that the occupied states reshuffle.

But there will always be M of them with probability $1/M$.

That is the law of conservation of entropy – if we can follow the system in detail.

There is a subtle point here.

Think, for a moment, that the initial randomness is not due to our imperfect observations but to some other cause.

Randomness in the initial conditions means that if we reproduce the experiment with the system, they will be different.

But they belong to a set of M initial states.

Then, if we can observe things precisely, this randomness propagates with no loss of information as time goes on.

Thus, if there is no loss of information, entropy is conserved.

Now of course in reality we may be again lazy, lose track of the system.

We might have, after a point, lost track of the equations, lost track of our timing device, and so forth and so on.

Thus we may have started with a lot of knowledge and wind up with very little knowledge.

That is not because the equations cause information to be lost, but because we just weren't careful.

Perhaps it is impossible to be perfectly careful – to be an omnipotent observer – perhaps there are too many degrees of freedom to keep track of.

Indeed the laws in physics are deterministic.

And the equations used to describe them are deterministic too.

But the lack of precision in our observations can generate the equivalent of randomness in initial conditions, or in measurements at time t.

And the lack of precise information might worsen.

So when that happens, the entropy increases.

But it simply increases because our ignorance has gone up, **not** because anything has really happened in the system.

If the initial system randomness is given and we could follow it, we would find that the entropy is **conserved**.

That is the concept of entropy in a nutshell.

We are going to expand on it a lot.

We are going to redefine it with a more careful definition.

What does entropy measure at this point?

It measures approximately the number of states that have a non-zero probability.

The bigger it is, the less we know.

What is the maximum value of S? It is $\log N$.

Now of course N could be infinite.

We might have an infinite number of states, and if we do, then there is no upper bound to the amount of our ignorance.

But when studying a closed system with only N states our ignorance is bounded.

So the notion of maximum entropy is also a measure of how many states there are altogether.

We said that entropy is deep and fundamental – and so it is.

But there is also an aspect to it which makes it in a certain sense less fundamental.

It is not just a property of the system.

This is a very important point worth stressing:

*Entropy is a property of the system **and** our state of knowledge of the system.*

It depends on two things.

It depends on characteristics of the system, and it **also** depends on our state of knowledge of the system.

You should **keep** that in mind.

Now let's talk about **continuous** mechanics.

1.8 Continuous mechanics, phase space, and Liouville's theorem

We want to talk about the mechanics of particles moving around with continuous positions, continuous velocities.

How do we describe that?

How do we describe the space of states of a mechanical system?

We describe real mechanical systems, particles and so forth, as points in a **phase space.**

The phase space consists of positions and momenta.

In simple contexts, momentum is mass times velocity.

So we can also say, roughly speaking, that the phase space is the space of positions and velocities.

Note on terminology: sometimes the term configuration space refers only to positions and sometimes it refers to positions and momenta, in which case it is another name for the phase space.

Anyway, phase space is the usual and unambiguous name for the space of positions and momenta.

That is sufficient when considering a system consisting of particles moving about, with positions x_i and velocities $v_i = \dot{x}_i$ ($p_i = m_i \dot{x}_i$), like for instance a gas.

Figure 11 shows the phase space for a system made of a large number of particles(entire plane).

The *i*-th particle has position x_i and momentum p_i.

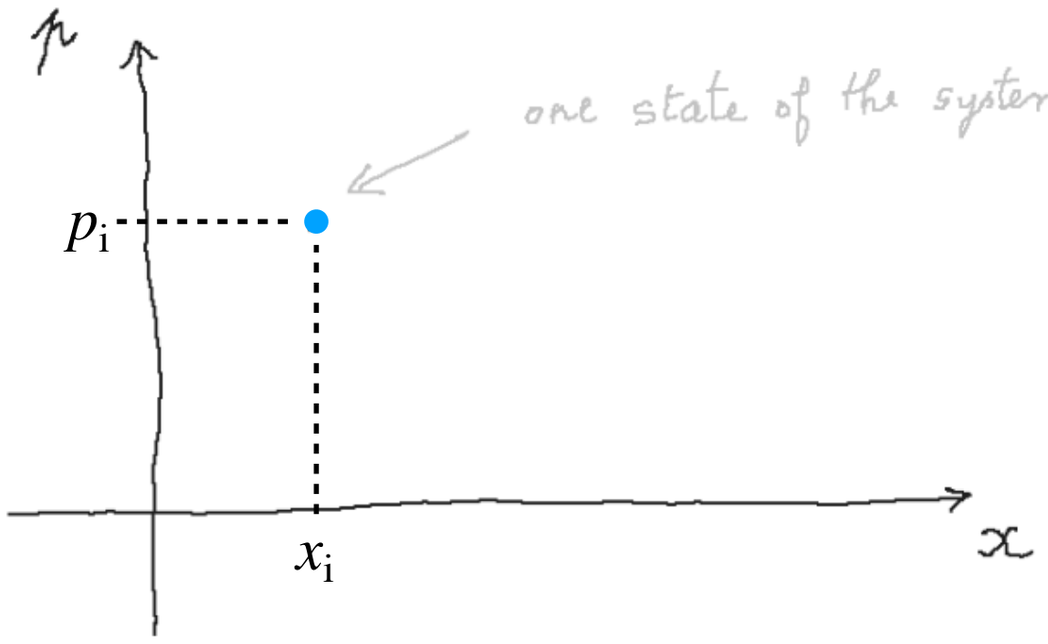


Figure 11: Phase space for a system of particles.

The vertical axis is for the momenta p_i's.

And as usual this axis is a stand-in for all of the momentum degrees of freedom, that is all the p_i's.

If there are 10^{23} particles there are 10^{23} p's.

And each of them is actually a vector of three values.

But we can't draw more than one of them.

Well, we could draw two of them but then we wouldn't have any room for the q's, that is for the x's.

Horizontally we record the positions q_i's of the particles.

But here we simply denote them x_i's.

And again the horizontal axis is a stand-in for the large number of dimensions corresponding to all the particles.

So we just write x to mean all the particles positions.

Let's start with the analog of a probability distribution, which is constant, that is equiprobable, on some subset of the total set of states of the system, and is zero on the remaining states.

We can represent that by drawing a sub-region in the phase space, figure 12.

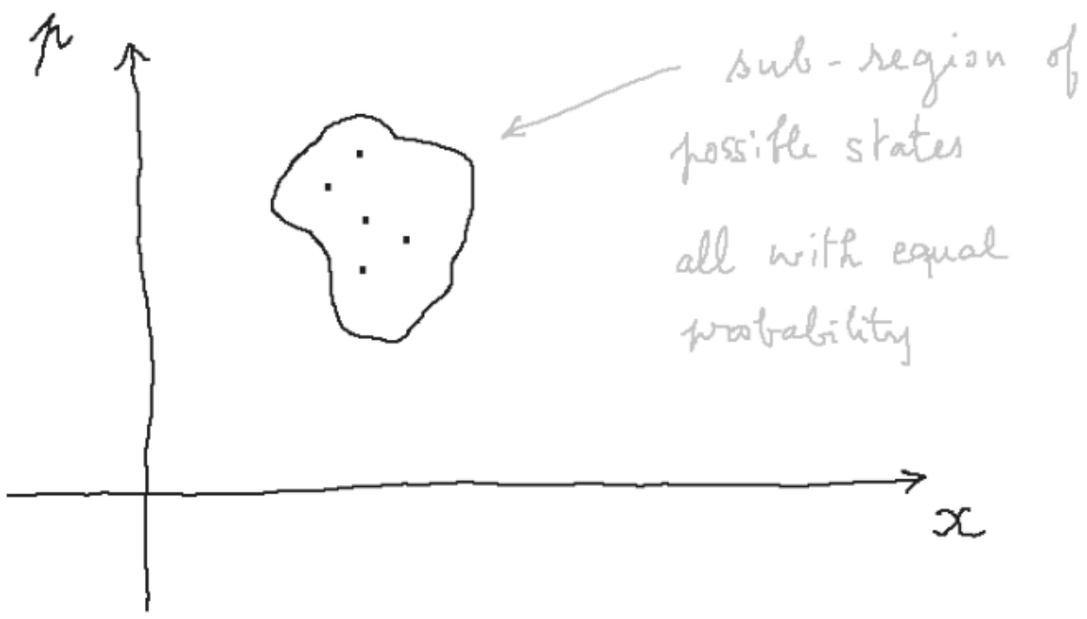


Figure 12: Phase space. Sub-region of possible states. Photograph at time t.

In the sub-region shown in figure 12, all the states have equal probability.

The time is fixed for the moment.

That is, at that time t, the system can be in any of the states in the patch, equally probably for each of them.

And the probability that the system be outside is zero.

That is the sort of situation where we may have some information about the particles, but we don't know exactly where each of them is and with which velocity.

For example we know that all the particles in the room are... in the room.

In fact we know quite a bit more than that.

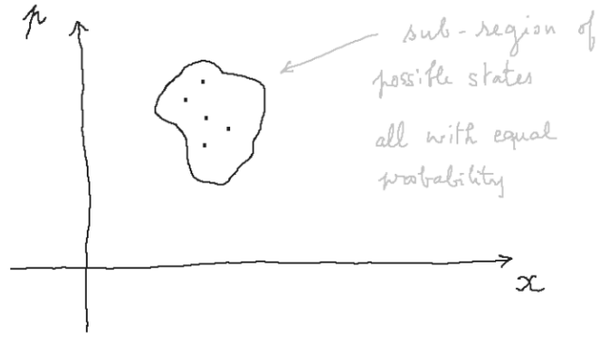
We know that they are more or less evenly spread out in the volume of the room.

We will formalize it, but let's go one step at a time.

Let's note, to start with, that knowing that the particles are in the room puts some boundaries on where x is in figure 12.

Remember that we now call x the entire collection of positions of the particles.

x is a unique multi-dimensional variable attached to the state of the system.

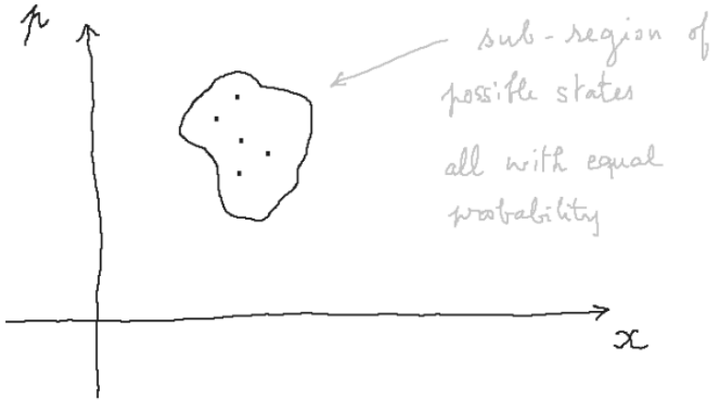


And for convenience we represent it as a simple number on the horizontal axis.

We may also know that all of the particles have momenta which are within some range.

That also confines them between values on the vertical axis.

So a typical bit of knowledge about the particles in the room might be represented, at least approximately, by saying that there is a zero probability that the system be outside the region shown in figure 12, and a uniform probability over all the states within the region.



Now the system evolves over time.

As the system evolves x and p change.

Over a period Δt, each point in the subregion in figure 12 will go to another point elsewhere in the figure, not necessarily in the initial region.

Points nearby will go to points nearby.

The whole sub-region – also called the occupied patch – will flow to another region with a possibly different shape.

The motion of the system with time is almost like a fluid flowing in the phase space.

We think of the points of the phase space as fluid points moving with time.

So the occupied patch flows to another occupied patch, figure 13.

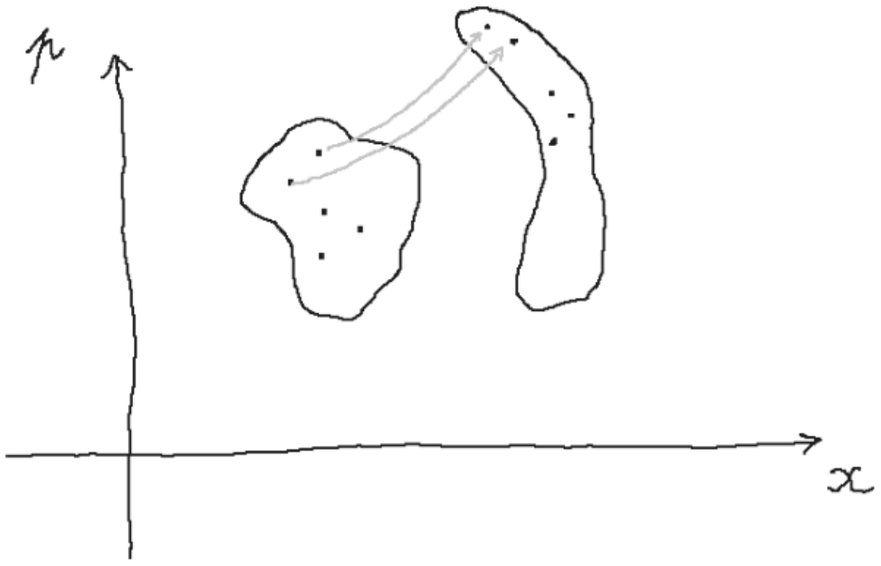


Figure 13: Flowing of the occupied patch in the phase space.

And just like there was equal probability for the system to be at any point in the initial patch in figure 12, after Δt there is equal probability for the system to be at any point in the new patch, wherever it is and whatever its new shape is now.

In truth we are in a continuous situation, and probabilities should be replaced by densities of probabilities multiplied by small volumes, but the reasoning is the same.

There is a theorem that goes with this flowing.

It is called **Liouville's** theorem.

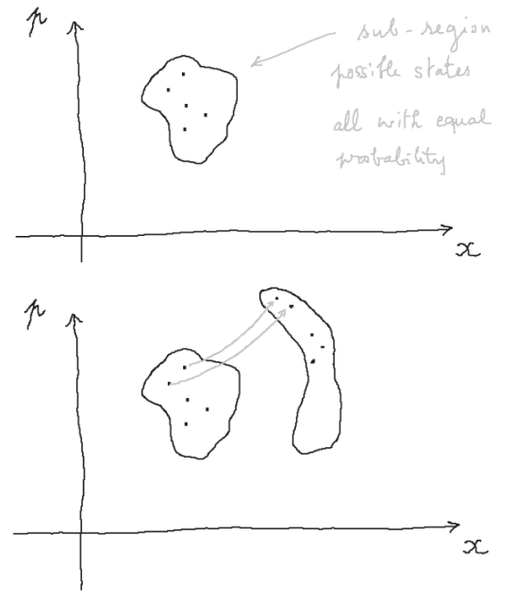
It says that the volume of the occupied patch is conserved, that is, remains constant over time.

In figures 12 and 13 the patches are surfaces and their volumes are actually areas.

But remember that the figures are simplified representations.

The x axis can be high-dimensional, and so is the p axis with the same dimension

So the volume of the patch is actually a doubly high-dimensional integral.



Liouville's theorem tells us is that whatever the initial patch evolves into, it is into something of the same volume, in other words figuratively speaking the same number of states.

It is the immediate continuous analog of the discrete situation where, if we start with M states and we follow the system according to the equations of motion, we will occupy the same number of states afterwards, as we started with.

They will be different states, but the motion will preserve the number of them and the probabilities will remain equal.

Let's stress that Liouville's theorem not only says that the volume of the occupied region will stay the same as it flows, but also, a little bit better, that if we start with a uniform probability distribution, it will **remain** uniform.

So there is a very close analogy between the discrete case and the continuous case.

And Liouville's theorem is what prevents a system from evolving according to an equation like equation (21) where everything comes to rest.

That can't happen.

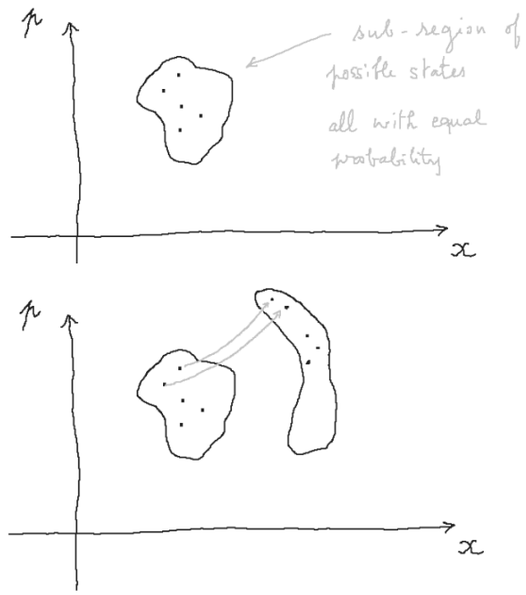
Let's see more precisely why it can't happen.

In figure 13, imagine that no matter where we started, we ended up with $p = 0$.

That would mean every point of the occupied patch got mapped onto the x axis.

Its area would go from some positive value to zero.

But Liouville's theorem prevents that.



What Liouville's result says in fact is if the blob squeezes in one direction, it must expand in another direction, like a drop of water trapped beneath a plastic film coating a surface, which we can move around but not remove.

Let's think of Liouville's theorem in the context of an eraser sliding on a table top, slowing down and coming to rest.

So we consider that *we don't know exactly* the state of the system, and therefore it is in a patch like in figures 12 and 13.

If, in the phase space, the components of the eraser get shrunk because it stops, it means the components of some other elements of the system [eraser + table] in the phase space must expand.

It is the p's and x's of the molecules that are in the table.

For the case of the eraser sliding on the table, there is really a very high dimensional phase space.

And as the eraser may come to rest, or almost rest, so that the phase space squeezes one way, it spreads out in another direction having to do with the other hidden microscopic degrees of freedom.

Our possible partial ignorance about the exact position and velocity of the eraser cannot vanish as it stops.

At the least, it is transferred as an increased partial ignorance about the exact positions and velocities of the molecules of the table top.

So if we can – as an omnipotent observer – keep track of the system the entropy is conserved.

We will see that it can also go up.

That will be the subject of the second law of thermodynamics.

We now know how the -1^{st} law of physics, about conservation of information, is expressed in the context of statistical mechanics and imperfect knowledge of all the degrees of freedom describing a system.

To follow the order of the laws of thermodynamics, let's talk briefly about the zeroth law.

Then we will talk about the first law, and then about the famous second law of thermodynamics.

1.9 Zeroth law of thermodynamics

The zeroth or 0-th law of thermodynamics has to do thermal equilibrium.

We haven't explained what a thermal equilibrium is, but we can already give an intuitive idea, and give a **preliminary** statement of the zeroth law.

A gas in a vase is in thermal equilibrium if all its molecules moves around in such a way that the global distribution of their positions and velocities somehow doesn't change over time.

And the concept can be applied to any system.

Similarly, we say that two systems A and B are in thermal equilibrium with each other if, when put in contact so that they can exchange energy, the global distributions of the positions and velocities of the molecules in each of them don't change.

The zeroth law of thermodynamics says that if we have several systems, and system A is in thermal equilibrium with B, and B is in thermal equilibrium with C, then A is in thermal equilibrium with C.

We will come back to that.

Just keep it in the back of our mind for the time being, because we haven't described what thermal equilibrium precisely is.

But having gone through the -1^{st} and the zeroth laws, we can now jump to the first law.

1.10 First law of thermodynamics

The first law is simply energy conservation, nothing more.

It is really simple to write down.

But its simplicity belies its power.

It is the statement that, first of all, there is a conserved quantity.

And the fact that we call that conserved quantity the energy will play for the moment not such a big role.

But let's just say there is energy conservation.

How is that expressed?

As we can guess, whatever the energy E is, the equation is written

$$\frac{\Delta E}{\Delta t} = 0 \rightarrow \lim_{\Delta t \rightarrow 0} \frac{\Delta E}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{E(t + \Delta t) - E(t)}{\Delta t} = \frac{dE}{dt} = 0 \tag{24}$$

Now this is the law of energy conservation for a closed system.

If a system consists of several parts in interaction with each other, then of course any one of the parts can have a changing energy.

But the sum total of all of the parts will conserve energy.

For instance, in the system represented in figure 14 we have

$$\frac{dE_A}{dt} = -\frac{dE_B}{dt} \tag{25}$$



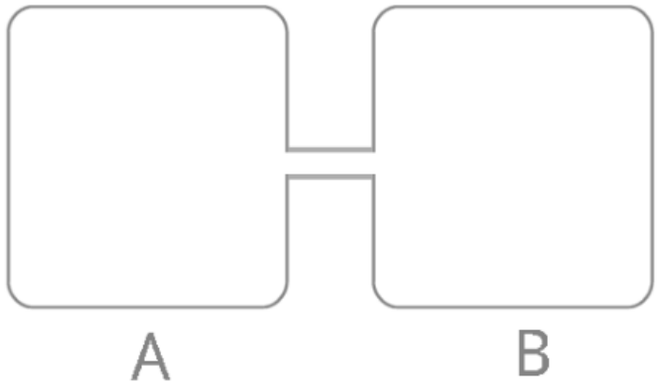
Figure 14: Two systems forming altogether one closed system.

We could have written that the sum of the two derivatives is equal to zero, but equation (25) emphasizes that what energy we lose on one side we gain on the other.

That is the first law of thermodynamics.

That is all it says.

The total energy of a closed system is conserved.



Now in the context of figure 14 there is a slightly hidden assumption.

We have assumed that, if the system is composed of two parts, its total energy is the sum of the energies of each parts.

That is really not generally true.

If you have two systems and they interact with each other, there maybe for example forces between the two parts.

So there might be a potential energy that is a function of both of the coordinates.

For example, let's look at a closed system made of two particles orbiting around each other, or, say, a simplified Solar System made of the Sun and the Earth.

The energy consists of the kinetic energy of one body plus the kinetic energy of the other body plus a term which doesn't belong to either body.

It belongs to both of them in a sense.

It is the potential energy of interaction(gives the force) between the Sun and the Earth.

In that context we really can't say that the energy is the sum of the energy of one thing plus the energy of the other thing.

Energy conservation is still true, but we can't divide the system into two parts as in figure 14.

On the other hand, there are many contexts where the interaction energies between subsystems is negligible compared to the energy that the subsystems themselves have.

If we were, for instance, to divide the table in front of you up into blocks.

How much energy is in each block?

Well, the amount of energy that is in each block is more or less proportional to the volume of each block.

How much energy of interaction is there between the blocks?

The energies of interactions are surface effects.

They interact with each other because their surfaces touch.

And typically the surface area is small by comparison with volume.

We will come back to that.

But let's keep in mind that in many contexts, the energy of interaction between two systems is negligible compared to the energy of either of them.

When that happens we can say to a good approximation that the energy of a system made of two parts can just be represented as the sum of the energies of each part, plus a teeny little thing which has to do with their interactions and which we omit.

In those circumstances, the first law of thermodynamics is often expressed with equation (25).

In summary, equation (24) is always true, while equation (25) has that little caveat that we are talking about systems where energy is strictly additive.

Now let's return to entropy.

1.11 More on entropy

We are not finished with entropy.

We have introduced it.

We have also talked about energy.

Notice that we haven't talked about temperature yet.

Temperature comes in behind entropy.

Indeed temperature is a highly derived quantity.

By that we mean that, despite the fact that it is the characteristic of a system we can readily feel with our body, so it is very intuitive and appears fundamental, it is actually a concept derived mathematically from the motion and kinetic energy of molecules, less primitive and less fundamental than either energy or entropy.

It will be the subject of Part 2 of our discussions.

For the moment we shall go a bit deeper into entropy.

We defined entropy.

But we did it only for certain special probability distributions.

Let's represent it schematically, figure 15.

On the horizontal axis are laid out the different states labelled by their index i.

And vertically we plot their probability.

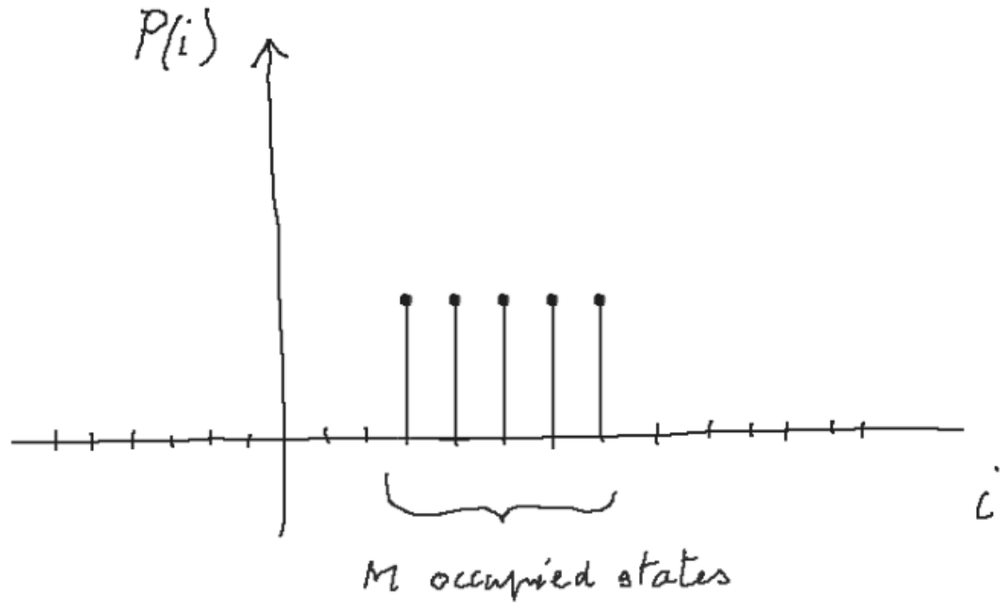


Figure 15: Equiprobable distribution over the set of occupied states.

The probability distributions we considered are those such that inside a subset of occupied states all the probabilities have the same positive value, and outside the probabilities are zero.

If there are M occupied states, the probability of each of them is $1/M$.

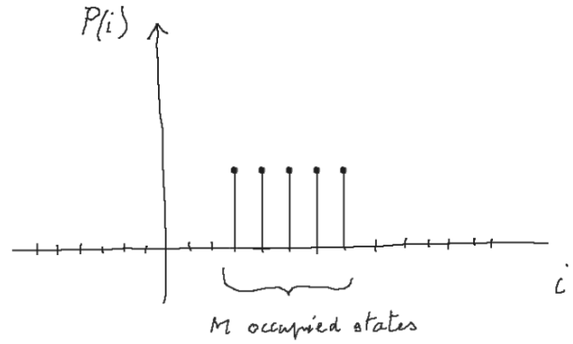
Then we defined the entropy as

$$S = \log M \tag{26}$$

That is, in the case where the distribution of probability is uniform over all the occupied states, the entropy is simply the logarithm of their number.

Generally speaking, however, we don't have probabilities evenly distributed like in figure 15.

We have probability distributions which are more complicated.



In fact they can be anything as long as they are positive and all add up to 1.

Let's draw another one, and for simplicity, even though we are in a discrete case, let's draw it as a continuous curve.

So the question is:

How do we define entropy in a more general context where the probability distribution looks something like in figure 16 or has an even more odd shape?

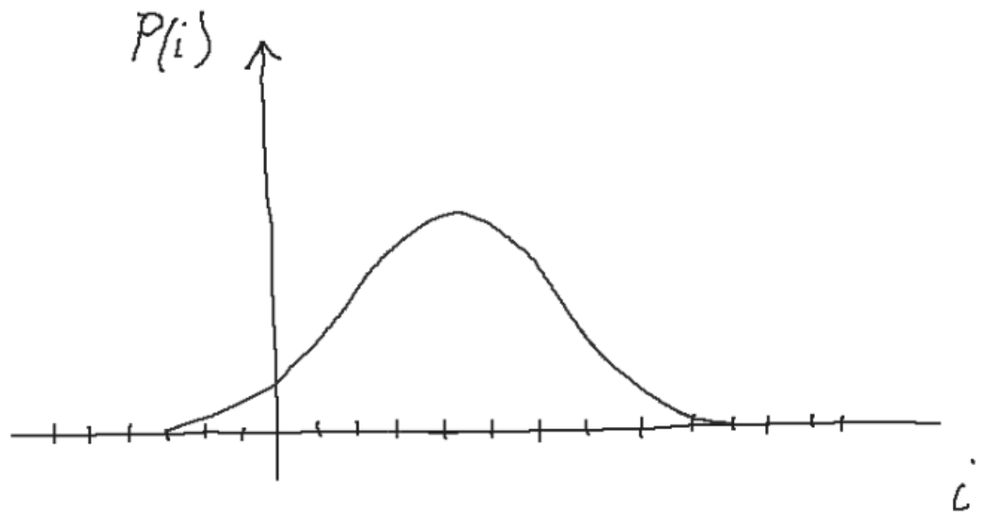


Figure 16: General distribution of probabilities over the states.

At this point in our discussion, we won't construct the formula logically.

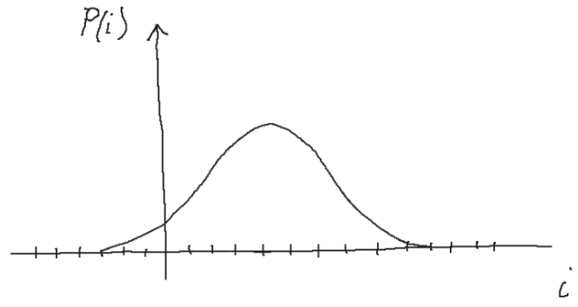
I will just give the definition.

We will check that, in the equiprobable case, it corresponds to equation (26).

Then our objective will be to get familiar with it and begin to see why it is a good definition.

It is representing something about the probability distribution.

It is in some sense the average number of states which are importantly contained inside the probability distribution in figure 16.



The narrower the probability distribution, the smaller the entropy will be.

The broader the probability distribution, the bigger the entropy will be.

In other words, a spread out distribution will correspond to a high entropy, while a narrow distribution over a small number of states will correspond to a small entropy.

The formula for the entropy S attached to a probability distribution of $P(i)$'s is

$$S = -\sum_i P(i) \log P(i) \tag{27}$$

Remembering formula (14) for the average of a function $F(i)$,

$$\langle F \rangle = \sum_i^n F(i) P(i)$$

we see that formula (27) can actually be read as the **average** of the function log $P(i)$.

It is worth emphasizing:

The entropy of a distribution P is the average of log *P.*

Let's see what this yields in the special case where the probability distribution is $1/M$, over M states.

For the unoccupied states the term $P(i) \log P(i)$ looks like it is going to cause a problem.

It is zero times logarithm of zero.

Of course logarithm of zero is not defined.

The limit when P tends to 0 of log P is minus infinity.

But the logarithm grows, in the negative numbers, to infinity much slower than P goes to zero.

So P log P goes to 0 (l'Hospitals rule).

$$\lim_{x \to a} \left[\frac{f(x)}{g(x)} \right] = \lim_{x \to a} \left[\frac{f'(x)}{g'(x)} \right]$$

So in general the contribution from states with very small probability will be very small, and by continuity the contribution of unoccupied states will be zero.

Now what about the states which have positive probability in the equiprobable case?

The probability of each one is $1/M$.

So formula (27) becomes

$$S = -\sum_{i=1}^{M} \frac{1}{M} \log\left(\frac{1}{M}\right)$$

Then, noting that $\log(1/M) = -\log M$, summing the M terms which are all the same, thus getting rid of the factor $1/M$, and cancelling the minus signs, we get

$$S = \log M$$

which is formula (26) again, as we want.

Notice that the minus sign in the general definition (27) for the entropy is there simply because $P(i)$ is always less than one, therefore all the $\log P(i)$ are negative.

If we want the entropy to be a positive number, we have to introduce a minus sign in its definition.

Our new definition now makes sense even when we have a more complicated probability distribution than just uniform over the occupied states.

We will discover, as we use it, that this average of $\log P$ is a good and effective definition.

Notice that entropy is associated with a probability distribution.

It is not something like energy which is a property of the system itself.

It is not a thing like momentum.

Entropy is associated with our imperfect knowledge of which state the system we are considering is in.

This idea requires some getting used to.

When we consider a system whose state is perfectly known, there is no concept of entropy.

Or rather the entropy is zero.

Your should check that it is consistent with formula (27).

When we consider a gas made of a very large number of molecules in a box, and it is in thermal equilibrium – assuming for the moment we intuitively know what that means –, at the same time we feel that we know well in which state it is, described by volume, pressure, temperature, etc. as classical thermodynamics tells us, but we must also admit that it is only a *statistical knowledge* over a large number of possible exact configurations of positions and velocities of the molecules.

And that is so even without taking into consideration the further limits imposed, at a much higher level of precision, by quantum mechanics on the simultaneous knowledge of positions and momenta.

So entropy is a special kind of measure.

It is a measure of our statistical knowledge only.

It is not a pure measure attached to the system itself.

It involves and reflects the incomplete knowledge we have of it.

We must develop a familiarity with this probability distribution representing our imperfect knowledge of the system.

Once in a while you should go back to the fundamental questions:

What is the random experiment that is implicitly referred to when talking about the $P(i)$'s?

What are the states i's?

In what sense do we have only an imperfect or incomplete knowledge of the system?

What are the $P(i)$'s? etc.

The beauty of statistical mechanics, or statistical thermodynamics, over classical thermodynamics, is that, once we have introduced the fundamental probabilistic description of the system, reflecting our imperfect knowledge of which state it is in, it offers new and stunning explanations of what is entropy and temperature, and their relations to energy and pressure.

We will discover, little by little, this beauty as we progress in our study.

This part of our discussion is devoted to the probabilistic framework and to the introduction of entropy.

Temperature will be treated in in the next part, pressure later in part 5.

And we will study many other subjects in statistical mechanics.

Hopefully you now understand why entropy is a somewhat more obscure quantity from the point of view of intuitive definition, than the other thermodynamics entities.

It is because its definition has to do with both the system and our state of knowledge of the system.

Let’s now do some examples.

1.12 Examples of entropy of simple systems

Let's calculate some entropy for a couple of simple systems.

Our first system will be with coins, not a single coin but a lot of coins.

We consider n coins, figure 17.



Figure 17: Collection of n coins, each showing heads or tails.

Each coin can be heads or tails.

Suppose that we have no idea what the state of the system is.

We know nothing.

The probability distribution, in other words, is the same for all states of this system consisting in n coins.

We are considering the case of absolute ignorance on our part.

What is the entropy associated to the system and us in such a situation?

Soon we will stop recalling that the entropy is related to us and our knowledge of the system.

It will be implicit, and we will drop the "and us".

All the probabilities are equal.

Under this circumstance, we just get to use logarithm of the number of possible states.

How many states are there altogether?

We have

$$N = 2^n \tag{28}$$

There are two states for the first coin, two states for the second coin, etc.

That makes 2^n possible states for the collection of n coins.

What is the entropy, given that we know nothing?

It is the logarithm of the number of possible states, therefore we have

$$S = n \log 2 \tag{29}$$

Here we see an example of the fact that entropy is kind of **additive** over the system.

It is proportional to the number of degrees of freedom in this case.

And we also discover a *unit of entropy*.

The unit of entropy is called a bit.

It is the abbreviation of binary digit.

That is what a bit is in information theory.

It is the basic unit of entropy for a system which has only two states, up or down, heads or tails, open or closed, or whatever.

The entropy is proportional to the number of bits, or in this case the number of coins times the logarithm of 2.

So log 2 plays a fundamental role in information theory as the unit of entropy.

It does not mean that in general that entropy is an integer multiple of log 2.

We will see it in a second.

Above our state of knowledge was zilch.

We knew nothing.

The value n log 2 has an important interpretation.

It is the *maximum entropy*.

With the same collection of n coins, let's try another state of knowledge.

Suppose, at the other extremity of the range of possible knowledge, that we know the state completely.

That is the case where the number of occupied states is $M = 1$.

The probability distribution now is shown in figure 18.

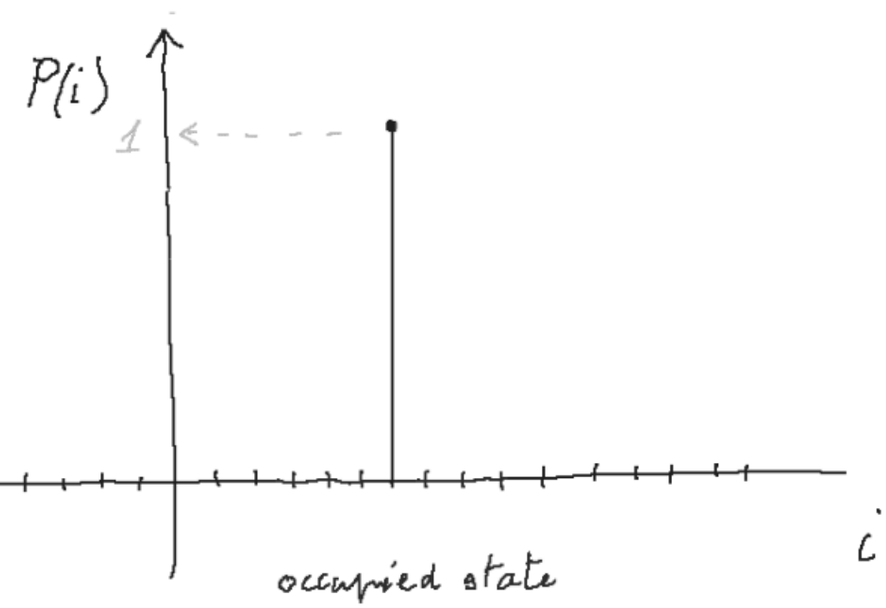


Figure 18: Case of perfect knowledge of the state of the system.

The entropy is now S = 1 log(1), and that is equal to 0.

So complete knowledge corresponds to zero entropy.

And, generally speaking, the more you know about the system, the smaller the entropy is.

With our collection of n coins, lets consider another interesting case.

We know that all of them are heads, except one that is tails.

But we don't know which one.

And we have no clue about where it could possibly be.

So our imperfect knowledge of the system corresponds to an even probability distribution over all the states with one tails and $(n-1)$ heads.

How many are there?

Well, clearly there are n such states.

Tails could be the first coin, or it could be the second coin, etc.

In other words, $M = n$.

Therefore, according to formula (27) or its simpler version (26), in this situation the entropy is

$$S = \log n \tag{30}$$

Notice that it doesn't have to be an integer multiple of log 2.

In general entropy is not an integer multiple of log 2.

Nevertheless log 2 is a good unit of entropy.

And, as I said, it is called the bit.

If the coins of our collection were objects with three possible states, formula (28) would become $N = 3^n$.

An example is a die where we consider the three states $\omega_1 = \{1 \text{ or } 2\}$, $\omega_2 = \{3 \text{ or } 4\}$ and $\omega_3 = \{5 \text{ or } 6\}$.

The maximum entropy would be n log 3.

And the example where we know that $(n-1)$ objects are in one state, and one of them is in another state, and we don't know which one, would have to be entirely reworked.

Now computer scientists of course, for a variety of reasons, like to think in terms of 2.

First of all the mathematics of it is nice.

2 is the smallest integer which is not 1.

Moreover, the physics that goes on inside your computer is connected with electric switches which are either on or off.

So it is easy to have them manipulate information expressed in bits(conventional computers).

1.13 Questions/answers

Question:

Is the logarithm in the definition of the entropy calculated in base 2 or in base e?

Answer:

Basically it doesn't matter because there is just a simple multiplicative factor between the two.

Remember that we have much freedom in the choice of our units.

The speed of light can be expressed in meters, or just taken to be 1.

The same is true with units of length, mass or time.

We have the MKS units system, the Planck units, etc.

And the same is true with the bit.

The usual choice depends on who you are.

If you are a computer scientist, or an information theorist, then you like logarithm to the base 2.

In this case log 2 is just 1.

And the entropy, for instance in the example leading to formula (29), it is just n.

That is, the maximum entropy of a collection of n coins is n.

If you are a physicist, then you usually work in base e.

But the relationship is just a multiplicative factor.

Logarithm to the base e and logarithm to the base 2 are just related by a numerical factor that is always the same.

Logarithm to the base e of 2 = 0.69314718056...

In our discussion, when we write log we mean logarithm to the base e.

But very little would change if we used some other base for the logarithms.

For us physicists the maximum entropy of the collection of n coins is approximately 0.7 x n.

But anyway we just write it n log 2.

1.14 Entropy in phase space

We are back in a continuous space of configurations of the possible states of the system, and the phase space, which combines positions and momenta, is continuous too.

Let's see what is the definition of entropy in this case.

We begin by supposing that the probability distribution is just uniform over some blob formed by the occupied states, and zero outside the blob.

In other words we are in a simple situation, figure 19.

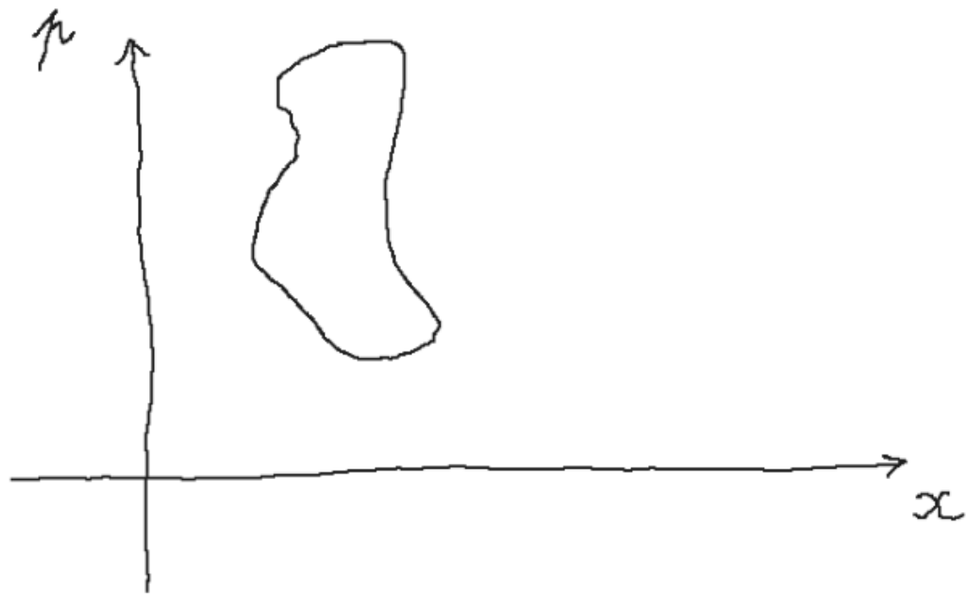


Figure 19: Region, or blob, in the phase space formed by the occupied states.

Then the definition of the entropy should be simple: we take the logarithm of the number of states in the blob – except that we are in the continuous case and the number of occupied states is infinite.

So **instead** we just say it is the logarithm of the volume of the probability distribution in phase space.

$$S = \log V_{PS} \tag{31}$$

V_{PS} is the volume in the phase space, which is high-dimensional.

Whatever the dimensionality of the phase space is, the volume is measured in units of position times units of momentum to the power of the number of coordinates in the system.

It is the volume of the region which is occupied and has a nonzero probability distribution.

This is the closest analog that we can think of to log M where M represented the number of equally probable states of a discrete system.

Here we didn't have to weigh the states with their probability or density of probability since we are in a uniform case.

More generally, if the distribution of probability over the blob in figure 19 is not uniform, we will introduce in formula (31) the *density of probability* $P(p, x)$ that the system be at point (p, x) in the phase space.

Remember that, because we are in a continuous case, we cannot deal with discrete probabilities and their ordinary sum.

The usual formula

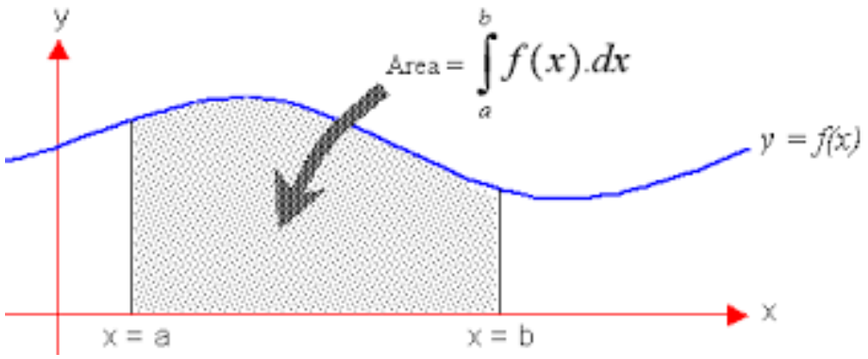
$$\sum_i P(i) = 1 \tag{32}$$

becomes now the integral of a probability density over the phase space

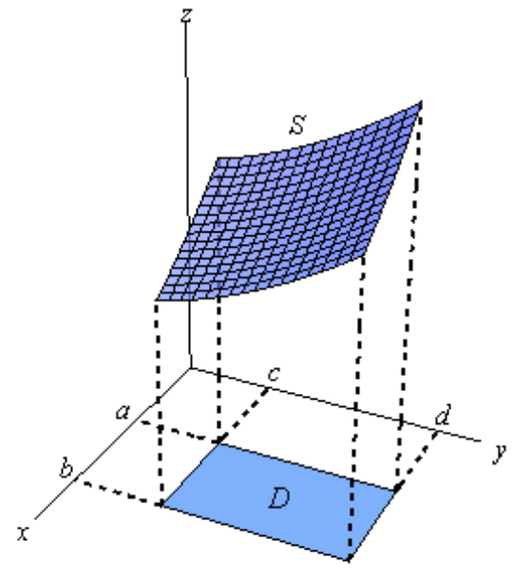
$$\int P(p,x)dpdx = 1 \tag{33}$$

The density $P(p,x)$ is non zero only over the blob of occupied states but that doesn't make any difference in formula (33).

Remember the 1-dimensional integral can be interpreted as the area under the curve….



and the 2-dimensional integral can be interpreted as the volume under the surface….



Now, in the continuous case with non uniform density of probability, the formula for the entropy is

$$S = \int P(p,x)\log P(p,x)dpdx \tag{34}$$

In first approximation – we don't mean first approximation numerically, but first **conceptual** approximation –, it is measuring the logarithm of the volume of the probability blob in phase space.

That was the meaning of formula (31) which we now extended to formula (34).

We have now defined, in the general continuous case, what is entropy, which as we have seen depends on the probability distribution, or density, over the blob of occupied states.

The next topics in our discussion are temperature and the Boltzmann distribution.

Temperature will be treated in part 2, maximizing entropy in part 3, and the Boltzmann distribution in part 4.

The Boltzmann distribution is the probability distribution for thermal equilibrium.

We haven't quite defined thermal equilibrium yet but we will.

Let's finish this part 1 in statistical mechanics with a some last questions/answers.

1.15 Questions/answers

Question:

In the example with n coins, do you assume that the tosses of each coin are independent?

Answer:

Well, I haven't made such an assumption.

I did not describe the experiment, or sequence of experiments, which produced the series of *n* elementary results, one for each coin.

Of course what you say could be the case, but the point is that the assumption in not necessary as such, and in fact would not always be true.

I just described what we know.

In the first example we know nothing, and the 2^n possible series of tosses are considered to be equally probable.

Then it is indeed equivalent to your assumption.

In the second case, we know that (n − 1) tosses are heads, and one is tails, and we don't know which one.

Now your assumption doesn't hold.

We just say what we know, without specifying how we got to our incomplete knowledge.

And our incomplete knowledge we translate into a probability distribution over occupied states.

To say, like in the first example, that all states are equally probable is closely related to saying that there are no correlations.

Let's go back briefly to elementary probability theory.

Consider two random variables X and Y, on the state space Ω.

For instance the experiment ε could be the tosses of two coins, coin 1 and coin 2.

X would be the result of the first toss, heads or tails.

And Y would be the result of the second toss.

The space of states here is

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\} \tag{35}$$

The random variable X is a function from Ω to the set (H,T).

It consists in picking the first value in the pairs forming Ω.

X has a probability distribution, which can be calculated from the fundamental probabilities for the four elements of Ω.

They don't have to be 1/4 each.

They could be, for instance, $P(\omega_1) = 1/2$, $P(\omega_2) = 1/4$, $P(\omega_3) = 1/4$, and $P(\omega_4) = 0$.

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$$

Then we can calculate the distribution of probability attached to X:

$$\begin{aligned} P\{X = H\} &= \frac{3}{4} \\ P\{X = T\} &= \frac{1}{4} \end{aligned} \tag{36}$$

This is denoted P_{X} , the probability distribution of the random variable X.

We can do the same for Y.

We can also define and compute a joint probability distribution for (X, Y). It is denoted P_{XY}.

For instance P_{XY} $(H, H) = 1/2$, P_{XY} $(H, T) = 1/4$, etc.

X and Y are said to be independent if and only if their joint distribution is the product of each of their distributions.

In other words, iff

$$P_{XY} = P_X P_Y \tag{37}$$

then, by definition, X and Y are independent.

Another way to define, and to feel, independence is as follows: X and Y are independent if knowing the value of X, after having performed ε , doesn't give any information on Y – technically, if the conditional distribution of Y , given a specific value of X, is the same as its marginal distribution (another name for P_{Y}).

Finally correlation is a probabilistic concept close to dependence.

However uncorrelation is less stringent than independence.

Strictly speaking correlation is defined only for numerical random variables.

We say that two random variables are uncorrelated iff the expectation of their product is the product of their expectations, that is iff (using the standard notations of probabilists)

$$E(XY) = E(X)E(Y) \tag{38}$$

Correlation is defined as follows

$$Corr(XY) = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y} \quad \textbf{(will discuss fully later)} \tag{39}$$

where σ_X is the standard deviation of X, and σ_Y is the standard deviation of Y.

Why introduce correlation and noncorrelation if dependence and independence is approximately the same concept?

Because it is simpler to compute or check.

It raises a little problem however: the two concepts are equivalent within the family of Gaussian random variables.

But they are not equivalent in general.

It is possible to construct counterexamples, where X and Y are not correlated but are not independent.

These counterexamples usually play no role in statistical mechanics, or even more generally in physics.

Let's go back to our series of n coins.

In the case where at first we know nothing, then knowing that coin 1 is heads, doesn't give any information on the others.

They are indeed uncorrelated and independent.

Now let's look at the case where what we know is: $(n-1)$ are heads, and one is tails, but we don't know which one.

If we look at the first coin and it is heads, does it give us any information on the second one?

Well, not much, but actually a little bit.

Instead of having probability $1/n$ of being tails, it is now $1/(n-1)$.

So the coin tosses cannot have been independent.

And in fact if the first coin is tails, then the second coin is surely heads, etc.

We gave a precise definition for correlation (not fully discussed yet).

For practical purposes, with the above little caveat, what is important to remember is this:

There is correlation between variables if measuring
a first variable gives us information on the others.

Or, said another way, there is correlation when the probability distribution for the other things is modified by measuring the first.

In the complete ignorance case, there is no correlation.

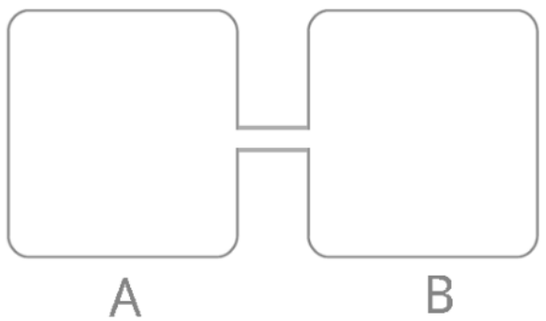
In any other kind of initial incomplete knowledge about the state of the system, in general there is some correlation.

Question:

In the system made of two parts A and B, in figure 14, once we measured one thing in A, which is a conserved quantity in the overall system, that gives us information on what it is in B, doesn't it?

Answer:

Oh yes. That is certainly true for energy.



Concerning entropy we have to be much more careful.

Incidentally, entropy is additive.

It is the sum of the entropies of all the individuals parts.

If your system is made of many identical parts, the overall entropy is proportional to the number of parts.

Entropy is additive whenever there is no correlation.

But this fact is more subtle than it looks, because entropy is not as simple a concept as energy.

The concept of entropy is related to our lack, or partial lack, of information.

So we will have to continue to develop our understanding of it.

In summary, uncorrelated systems have additive entropies.

But that is a theme that we will come back to.

Question:

Was this idea of entropy of a series of coins invented in 1949 with information theory?

Answer:

Shannon rediscovered, in the context of information theory that he was developing, the idea of entropy.

And he defined the entropy of a message made of a series of bits – that is the same as our collection of n coins.

However the formula

$$S = -\int P \log P dP \tag{40}$$

is due to Boltzmann.

Now of course Shannon discovered this entirely by himself.

He didn't know Boltzmann's work.

And he worked from an entirely different direction, from information theory rather than from thermodynamics.

But none of it would have surprised Boltzmann.

Nor do I think Boltzmann definition would have surprised Shannon.

They are really the same thing

If you put the minus sign in

$$S = -\int P \log P dP$$

it is called entropy.

If you don't put the minus sign, then you write

$$I = \int P \log P dP$$

and it is called information.

Question:

Is the incompleteness of our knowledge about the state of the system, in statistical mechanics, related to the Heisenberg uncertainty principle in quantum mechanics?

Answer:

No. These are two separate issues.

S doesn't have to do with the quantum mechanical uncertainty which we encounter when measuring an observable, if the state of the system is not one of the eigenstates of the observable.

S has nothing to do with the randomness of measures implicit in pure states.

Discussion: customary mistakes in elementary probabilities

When talking about probabilities, even educated people may say wrong things like:

> since having two accidents is very rare, and since I already had one,
> now the chances that I have another one are much smaller.

On a more sophisticated level, the paradox of Monty Hall, for instance, can befuddle even the best minds.

Here is how it goes:

There are two people, one is the guesser who must provide a best guess, the other is the operator of the game.

There are three closed doors, A, B and C, facing the guesser.

Behind one and only one is a prize.

Step one: the guesser must make a guess, for instance guess door A.

Step two: the operator, who knows where the prize is, doesn't open A, but selects among B and C a door where the prize is not, and opens it.

The guesser sees this new piece of information.

Step three: the guesser is invited to guess again where the prize is.

Question:

Should the guesser change his or her guess, and now choose the other non-opened door, or it doesn't matter?

Answer:

It does matter.

The guesser should change guess, and now choose the other non-opened door.

The probability of winning will go from 1/3 to 2/3.

Before reading the solution, try to solve it by yourself.

One way to see it is to note that, if the guesser follows this strategy, he or she will lose only when the prize was behind the initial guess.



2. Temperature

2.1 Units of temperature and entropy

Let's spend a few minutes talking about units, in particular the Boltzmann constant k_B.

What is the Boltzmann constant?

Like many constants in physics, k_B is a **conversion** factor.

The speed of light is a conversion factor from distances to times.

If something is going at the speed of light, and in a time *t* covers a distance *x*, then we have

$$x = ct \tag{1}$$

And of course if we judiciously choose our units, we can set c = 1.

Typically these constants are conversion factors from what we can call human units to units in some sense more fundamental.

Human units are convenient for scales and magnitudes of quantities which people can access reasonably easily.

We haven't talked about temperature yet, which is the subject of this part of our discussion.

But let's also say something about the units of temperature, and see where k_B – which will simply be denoted k when there is no possible confusion – enters the picture.

Units of temperature were invented by Fahrenheit and by Celsius, in both cases based on physical phenomena that people could easily monitor.

The units we use in physics laboratories nowadays are Kelvin units, which are Celsius or centrigrade degrees just shifted by an additive constant, so that the 0°C of freezing water is 273.15 K and 0 K is absolute zero.

There is no lower temperature.

Thus temperatures Kelvin are relative to absolute zero, while temperatures centigrade are relative to the freezing point of water at atmospheric pressure, as chosen by Celsius.

And the centigrade scale, the unit of 1 degree, was simply set by dividing the temperature scale between freezing and boiling of water at atmospheric pressure, into 100 units.

All these choices were highly arbitrary.

Fahrenheit used a mixture of water, ice and ammonium chloride for his 0°F, which turned out to be −17.78 °C.

Alcohol or anything else could have been used.

It would have produced different units.

The point is that these were units which people could manipulate.

They could measure temperatures.

It is not too hard to measure a change in temperature of one degree centigrade with a human scale thermometer.

By human scale I mean an object that you can buy in a store.

So the centigrade scale is a human construct that was invented for convenience.

In the 19th century, it was not fully understood – although Kelvin may have understood it – that the real physical units which temperature has to do with is energy.

Temperature is really an energy.

For example, the temperature of a gas determines the kinetic energy of its molecules.

And one measure of the temperature of the gas is just the kinetic energy of the molecules.

Well, you might say: wait a minute, that is not so good because if the gas is a composite of two gases, say oxygen and helium, the molecules of which have quite different masses, surely they have different kinetic energy.

But they don't!

We will eventually prove that in thermal equilibrium at a given temperature, all of the particles in a gas have basically the same kinetic energy.

In fact if you suspended a bowling ball into into a gas, or even a liquid, when the thermal equilibrium is reached, and everything is at the same temperature, the kinetic energy of the bowling ball would be exactly the same as the kinetic energy of one molecule of the gas.

That seems a little bit odd. But when you think about it, it is not that odd.

If the bowling ball has the same energy as a molecule, it would surely move a lot slower than a single molecule, for the reason that its mass is much larger.

If mv^2 for the bowling ball is the same as mv^2 for a molecule, then obviously a much bigger m will lead to a much smaller v.

The main point is that temperature is really a measure of an amount of energy.

So the natural units for temperature are really those of energy.

But at the time when temperature first started to be measured, nobody could measure a molecule.

After the atomic hypothesis appeared, under the impetus of people like Dalton, nobody had the vaguest idea how massive molecules were, or how much energy a typical molecule had at a given temperature.

And in fact, until the time of Maxwell, Stefan and Boltzmann, nobody had the idea that temperature could have to do with the energy of a basic constituent.

So the scale of temperatures was defined before there was a notion that temperature so to speak equals energy.

The natural unit for temperature could be Joules for example.

A Joule is a fine unit of temperature, although a rather large one.

Indeed the typical energy of a single molecule in thermal equilibrium at room temperature is a very tiny number of Joules.

And the temperature is essentially the kinetic energy of one molecule.

Pay attention to the fact that temperature is not an additive quantity.

If we take two volumes of gas, both at the same temperature, and put them together we still have the same temperature.

So we are indeed talking about the temperature of *one* molecule.

The conversion factor between human units and the basic units of energy is therefore a very small number.

To figure it out, let's start with the relationship between the energy E of a single molecule in a dilute gas and its temperature T.

We consider a dilute gas because we don't want to worry too much about the energies of interaction between the molecules.

The formula is

$$E = \frac{3}{2} k_B t_K \tag{2}$$

The 3 is easy to understand.

It has to do with the number of dimensions in space.

A molecule has a kinetic energy from moving along the x-axis, the y-axis and the z-axis.

So there is a factor of 3.

And the half is just a glitch of definition.

The temperature t_K in equation (2) is the human scale temperature, expressed in degrees Kelvin.

And k_B is the Boltzmann constant, that is the conversion factor.

In that case it has units of energy divided by degrees Kelvin.

And it is a rather small number as you might expect.

If the temperature is room temperature, t_K is about 300 degrees K.

The energy E of a single molecule, on the other hand, is clearly very small. So k_B (read k Boltzmann) is necessarily very small.

Its numerical value is

$$k_B \approx 1.4 \times 10^{-23} \ \frac{m^2\, kg}{s^2\, K} \tag{3}$$

$kg\ m^2/s^2$ is a unit of energy.

It is of course a Joule, and it is the standard unit of energy in the MKS system of units.

It is no accident that Avogadro's number is of order of magnitude 10^{23}.

Finally equations (2) and (3) give for the kinetic energy of a molecule in a gas at 300°K

$$E = 6.2 \times 10^{-21}\ J \tag{4}$$

As explained above, it is striking that this energy doesn't depend on the kind of molecule.

Whether big or small, the molecule will have the same kinetic energy.

For a given type of molecule we can calculate its speed.

Let's take, for instance, a molecule of nitrogen.

It is made of two nitrogen atoms.

One nitrogen atom is made of 7 protons, 7 neutrons and 7 electrons.

Its mass is $m \approx 2.3$ x 10^{-26} kg.

Thus writing $1/2\ mv^2 = E$, and solving for v, we get for the speed of a nitrogen molecule

$$v = 5.2 \times 10^2\, m/s \tag{5}$$

In other words, when we feel cozy at room temperature and ordinary pressure, it is because billions of billions of tiny molecules of nitrogen keep hitting us at a speed of about 520 meters per seconds

At freezing temperature their speed is still about 500 meters per second, since it varies like the square root of the temperature expressed in degrees Kelvin

The main point about equation (2) is that the fundamental quantity of interest involved, namely t_{K}, always comes in multiplying the Boltzmann constant.

This is just because the fundamental quantity is really energy.

So we define a new measure of temperature

$$T = k_B t_K \tag{6}$$

This measure is the Boltzmann constant multiplying the human temperature.

It has units of energy.

Once we use this measure T we get rid of k_B in all equations involving temperature.

We will never see it again.

But we must remember that what from now on we call temperature, namely T, differs from thermometer temperature.

It is thermometer temperature (expressed in degrees Kelvin) multiplied by an extremely small conversion factor.

That same factor k_B is also present in the units of entropy.

Remember what entropy is: it is a unit of information. It is measured in bits.

Entropy was discovered as an interesting quantity, which always increases, by a young steam engineer and scientist by the name of Carnot, who was working in the 1820's on improving the recently invented steam engines.

Carnot had a quantity, defined in terms of temperatures and energies and so forth, which wasn't easily measured, but played an important role in describing theoretically the functioning of steam engines.

He did not know anything about the underlying microscopic phenomena driving his quantity, let alone its relation to a probabilistic description of our ignorance of the exact state of the system.

At the time of Carnot, understanding the relationships between the concepts of temperature, heat, and work was just beginning.

The name entropy was given by Clausius in 1865.

And statistical thermodynamics was created by Boltzmann even later.

Entropy received the notation S.

For the moment let's use S_{carnot} for the measure discovered by Sadi Carnot.

That is the quantity which he did not call entropy but whose variation, at a given temperature, he understood, was equal to the variation of energy divided by the temperature.

S_{carnot} has units.

It is energy divided by temperature.

If the temperature is measured in Kelvin then it is Joule per Kelvin.

In other words, Joule per Kelvin was Carnot's units of entropy.

But in modern physics we think of entropy as being measured in bits.

A fundamental unit of entropy is logarithm of 2 for example.

And that is dimensionless.

So there is a conversion factor again.

Carnot's unit of entropy differs from what we have called entropy up to now.

Let's recall what entropy is from a statistical point of view.

We have some probability distribution over a discrete set of states, figure 1.

Then we define the entropy S – without the subscript Carnot now – as

$$S = -\sum_i P_i \log P_i \tag{7}$$

It is the same as equation (27) of Part 1.

It is a rough measure of the logarithm of the number of states which are important under the statistical distribution above.

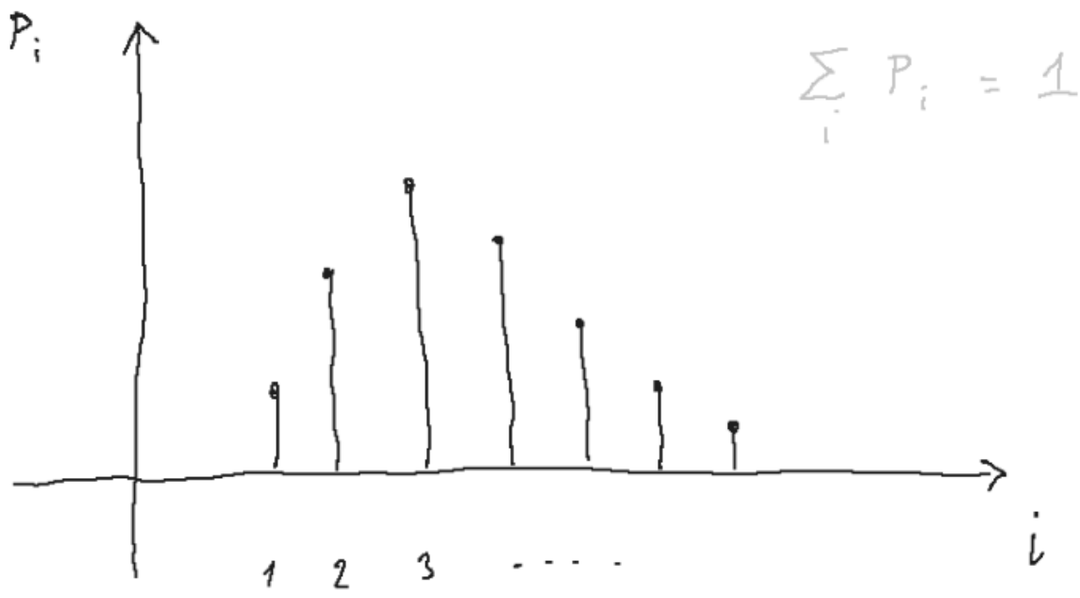


Figure 1: Probability distribution over a discrete set of states.

That is what we called entropy.

What is the connection between Carnot's entropy, and Boltzmann's entropy given by equation (7)?

They are simply related by the Boltzmann constant

$$S = \frac{1}{k_B} S_{\text{carnot}} \tag{8}$$

In other words when Carnot's entropy is an ordinary number like 6 in a steam engine, Boltzmann's entropy is enormous.

It is roughly proportional to the number of molecules in the system.

Once we know how to go back and forth between human units of temperature or human units of entropy and the more fundamental units of temperature and entropy – for temperature it is energy, and for entropy it is a dimensionless quantity – then we no longer need to use Boltzmann constant.

With fundamental units, the Boltzmann factor effectively goes out of all of our equations.

Conversely, if for some reason we need to go back to human units, it is easy to reintroduce.

Whenever you see capital T and you want to get it into human units think of it as k_B times temperature in Kelvin.

And whenever you see an entropy S which has a huge value typically, and you want to get it into steam engine units multiply it by k_B.

2.2 Questions/answers

$$E = \frac{3}{2} k_B t_K$$

Question: From equation (2), apart from the factor 3/2, now temperature is in Joules?

Answer: It depends of course now on what unit of energy you are using on the left-hand side.

If you are using the Joules for your units of energy, then the temperature T will be in Joules.

And in three dimensions a particle of energy 10 Joules, will have a temperature of about 7 Joules.

To be rigorous we should speak of "Joules- of-temperature" such that 0.67 Joule-of-temperature is the same as 1 ordinary Joule.

Then we would say that a particle of 10 Joules of energy has a temperature of about 7 Joules-of-temperature.

The point is that it is not really new units.

It is just like measuring something in units of 1.5 meters.

Although purists would maintain that it is a new unit.

What is the Joule?

It is the unit of energy in MKS units, that is meters, kilograms and seconds units.

1/2 Joule is the energy of a 1 kg object moving at a velocity of 1 m/s.

Question: Is *k* Boltzmann is always 1.4×10^{-23} ?

Answer: k Boltzmann is always a constant.

But its numerical value depends on the units you are using.

The same is true of the speed of light, or of any physical constant.

We like to use units of length and time such that c = 1, because is more convenient than its expression in meters per second.

And it is much more convenient that its expression, for example, in knots.

But nothing prevents us from using knots if we like.

The customary value of 1.4×10^{-23} for Boltzmann constant is only when it is measured in Joules per Kelvin.

If you use, say, BTU per whatever measure of temperature you like, you will get a different numerical value.

If instead of using Kelvin units we used Fahrenheit units (with the zero shifted to correspond to absolute zero, though), the numerical value of k_B would change.

Energy itself, like mass or length or many other characteristics of objects in nature, is a quantity with units.

Therefore you have to specify what units you are talking about.

When I say temperature has units of energy, it does.

But you have to pick your units of energy.

If you are working in the meter kilogram second set of units, then your natural unit of energy is called the Joule.

It is not a new unit.

It is the name of the unit of energy in MKS.

It is just $kg\ m^2/s^2$.

And that is the unit that you would use for temperature if you want to keep your formulas simple.

You could use any nice units you like, electron-volts or Planck masses or whatever.

If you change your units of energy, then you change your units of temperature, while still keeping the ratio $E = 3/2\ T$.

But, to keep simple formulas, your units of temperature and energy should be kept consistent.

Boltzmann's entropy on the other hand is dimensionless.

The quantity $-\sum_i P_i \log P_i$ doesn't have units.

Carnot's entropy has units, since it is related to Boltzmann's with the multiplicative factor k_B in equation (8).

We could also defend the point of view that degrees Kelvin are really units of energy.

Then k_B would be viewed as dimensionless, which doesn't mean that it does not depend on units.

Anyway we look at it, it is just a conversion factor.

Question: Is the value of $k_B \approx 1.4 \text{ x } 10^{-23}\ J/K$ something determined experimentally?

Answer: Yes, it is.

Boltzmann himself, same as Newton, never knew the value of his own constant.

I don't mean the usefulness, I mean the actual numerical value.

That is because nobody had measured the properties of atoms, the masses of atoms to any accuracy, until Einstein came along and figured out how to do it.

Boltzmann died just before the value of his constant was known.

I'm quite sure, though, that he had a rough idea.

And he certainly knew that it was connected to Avogadro's number.

But he probably never knew with any precision what the value of his number was.

The same is true of Newton.

Boltzmann did understand that, if only he could measure the properties of atoms, then energy could be measured using temperature and his constant.

That he understood.

Alright, thus far we have not said what temperature is.

It is time we turn to it.

2.3 What is temperature?

We have an intuitive sense of temperature.

That is because when we touch a hot thing it does something to our nerve cells and so forth and so on.

And we have a natural intuition about hot and cold.

But temperature is a rather derived mathematical quantity.

It begins with the twin notions of energy and entropy.

We will eventually define precisely what it means for a system to be in thermal equilibrium.

But, until we do so, it just means intuitively that a system, closed or exchanging with the outside, has been allowed to evolve for a long enough time that it just settles down to some equilibrium configuration.

However, as we shall see in a moment, even in equilibrium it is still evolving at a microscopic level between many states.

As I said, it may be closed, or it may be exchanging with the environment.

If the environment is not changing with time the system under consideration will equilibrate in a certain sense that will become clear.

During the process of equilibration, energy may flow from the system to the environment.

In that case we say the system is **hotter** than the environment.

The energy may also flow in the other way on the average, in which case we ay that the system is **colder** than the environment.

And in equilibrium on the average energy is flowing neither way.

Let's consider a system in equilibrium, figure 2.

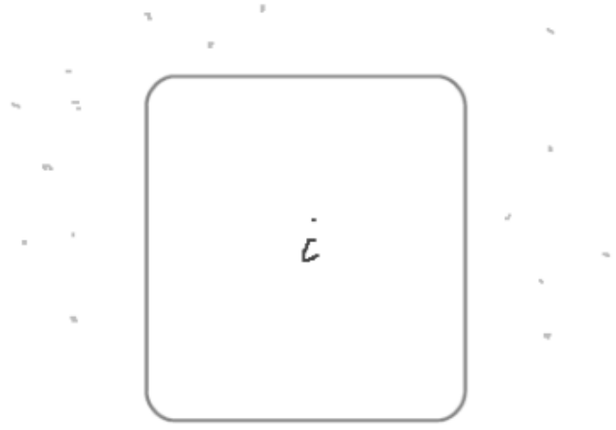


Figure 2: System in equilibrium with its environment. The system is in state i.

The system inside the box has a lot of degrees of freedom.

They define the collection of states.

The i-th state has an energy E_{i}.

Now, when a system comes to equilibrium it doesn't come to a particular state i.

It is interacting with its environment.

Energy is going back and forth.

Complicated things are happening.

Even in equilibrium there is some probabilistic process going on over time.

At time t, there is some probability that the system be in the i-th state, which has energy E_{i}.

So there is a probability $P(i)$ attached to the i state.

For lightness of notation, we will just denote it P_{i}.

And it doesn't depend on t, so we may forget about t for the moment.

The probabilities P_{i}'s satisfy some rules.

First of all, of course, they are positive and

$$\sum_i P_i = 1 \tag{9}$$

Secondly – it is not a rule but a definition –, the average energy of the system is

$$\sum_i P_i E_i = \langle E \rangle \tag{10}$$

This is the average energy of the system when it is in equilibrium.

When we say average what we really have in mind is that the energy is constantly changing.

It fluctuates.

Some extra energy goes out into the environment, and then it comes back in and so forth.

So it is constantly fluctuating.

When the system is in equilibrium, at any time t, when averaged over a short period of time around t, the average energy over the possible states is a constant which doesn't depend on t, and equation (10) is the equation for this average energy.

And finally we have a notion of entropy.

$$S = -\sum_i P_i \log P_i \tag{11}$$

Now what is the probability distribution P_{i}?

What does it depend on?

Among other things, but mostly, it depends on the average energy in the system.

The average energy may depend on the nature of the environment.

If the environment is very hot, the average energy in the box may be high.

If the environment is cold the average energy in the box may be low.

In other words, the environment determines the collection of states the system in equilibrium can be in, and their probability.

It also determines the E_i's and therefore $< E >$, which we will simply denote E.

The average energy E of the system can be thought of as a parameter.

For instance we can change the average energy of a kettle of water by putting it on a stove.

Or we can cool it down but putting it out in the snow.

So we can change the average energy.

In that sense the collection of states with non zero probability and the distribution of probability depends on E.

So we will denote the probabilities as

$$P(i, E) \tag{12}$$

For every given average energy, in thermal equilibrium, there is a distribution of probability $P(i, E)$.

Of course if we increase the average energy E, then we shift and reshape of the probability distribution toward larger energy values of E_{i}'s.

That means that there is not a single probability distribution P_{i}, that we call equilibrium, there is a family of them indexed by the value of the average energy E.

A given E corresponds to a given distribution $P(i, E)$, and conversely, figure 3.

On the horizontal axis are arranged the states i*'s* in ascending order of energy E_{i} .

Which distribution corresponds to more average energy?

Answer: the one that is peaked further out to the right.

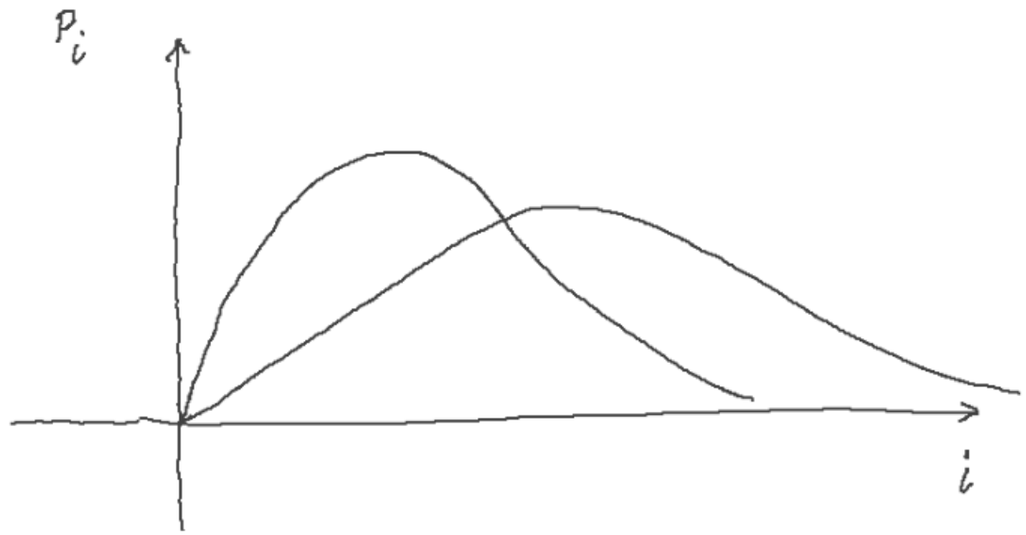


Figure 3: Two probability distributions corresponding to different average energies.

We haven't calculated yet what is the distribution $P(i, E)$ for a given E.

That will be one of our objectives.

So far, all we can say is that *as a matter of experience* for each value E of the average energy there is a probability distribution.

This forms altogether a family of probability distributions, two of them being represented on figure 3.

Every probability distribution $P(i, E)$ satisfies the three rules above: sum equal to 1, average energy equal to $< E >$, or E, and it has some entropy S.

We mentioned in part 1 of our discussion, that when talking about probabilities one must be totally clear about what is the experiment ε we have in mind, that can be – at least in theory – reproduced, and whose outcome will change from replication to replication, and is therefore driven by a probability measure.

So before proceeding, let's remind you what it is here.

We consider a system in thermal equilibrium.

The theoretical experiment ε is this: look at the system at time t, and suppose we have the power to observe everything about the system, in particular the position and velocity of every particle.

We call that state of the system ω_i, or simply i.

That state has an energy E_i, which is a random variable, since E_i depends on ω_i and will vary from replica to replica.

To reproduce ε means to look again at a new time later, and so forth, supposing the system always remains in the *same* thermal equilibrium.

If we reproduce ε a very large number of times, from the law of large numbers we can determine the distribution of probability over the possible states of the system.

These are the $P(i, E)$.

The average energy E is the following concept derived from experience:

in thermal equilibrium the system can exchange with the outside;
some energy goes back and forth;
but, averaged over even very short periods of time, the average energy doesn't change.

So it is a characteristic of the system in thermal equilibrium.

It is the E we are talking about above.

It is also a characteristic of the probability distribution attached to the thermal equilibrium we are observing.

A note on terminology: the outside which somehow stabilizes the system in thermal equilibrium is often called a *heat bath*.

The intuitive idea is that it will help the system reach the same temperature as the outside, while that of the outside doesn't change.

Hence the term heat bath.

Of course, there are many probability distributions that we could write down that have the same average energy E.

Where are we getting this one parameter family $P(i, E)$ from?

That will come later.

As a matter of experience if we know the average energy of a system, then there is some probability distribution associated with it - over the possible states it can take.

To give it a name, it is the Boltzmann distribution.

It will be the subject of parts 3 and 4 of our discussion.

For now we are just doing a mathematical exercise.

We have a one parameter family of probability distributions, parameterized by the average energy.

No two of them have the same average energy by assumption.

And the higher the energy is, of course intuitively the broader the probability distribution in figure 3 is likely to be.

And we will take that as a given.

To answer the question about whether two states could be correlated, notice that the concept of correlation doesn't apply to our setting.

Correlation is a technical term with a very precise meaning.

It is relevant when we consider *two random variables measured on the same experiment*.

Here so far we measure only one, namely the energy E_i.

General speaking, the idea that one state could be correlated to another state reflects a confusion.

In a random experiment ε, two measurements, one after the other, of any random variable X by definition are independent.

To establish it, we would have to define a new experiment $\mathcal{F}$ which consists in performing $\mathcal{E}$ twice in a row, and we would have to define a probability measure for $\mathcal{F}$.

But this probability measure would – by definition of how $\mathcal{F}$ is constructed – be the product of the probability measures for each of the replications of $\mathcal{E}$.

So there is no meaningful notion, in our present setting, of two states being correlated.

A more meaningful question would be what happens if there is more than one state with exactly the same energy – not average energy E, but E_i and E_j being the same.

What if, say, E_1, E_7, and E_{36} are exactly the same?

Well, first of all it never happens, but even if it did it would not raise any particular problem in the circumstances.

Degeneracy, which means equal energy for different states, is a very rare and unusual thing.

And, anyway, we don't have to worry about it.



For instance in figure 3, where on the horizontal axis the states are arranged with increasing energy E_i, we can assume that every state i has its own energy $E\mathrm{i}$, and every energy $E\mathrm{i}$ has its own state i.

If that makes things clearer, let's take it as a working assumption.

Let's emphasize: the i-th state is the precise microscopic description of the system.

It is the knowledge that an all-powerful, all-knowing observer would have.

Every molecule, position and velocity, or whatever it takes to describe the exact microscopic state of the system is specified in the i-th state.

The fundamental idea of statistical mechanics is that, at a time t, the system is in one and only one precise state, but we – the real life observer – have only an incomplete knowledge of it.

Our incomplete knowledge is modeled with a probabilistic framework.

We know that the system is in one of a collection of occupied states.

And we know the probability distribution P, or $P(i)$ over them.

NOTE: We use indifferently the notation P, or $P(i)$, or $P(i)$'s, for the whole distribution of probability.

The notation $P(i)$, or $P(i)$'s, for the whole collection may be a bit confusing, but it is standard practice.

It means implicitly that i runs over its range of values.

Sometimes of course, we do want to talk about one specific probability value of one i.

But the context should always make it clear.

The implicit random experiment ε underlying our incomplete knowledge is a replication of the whole experiment leading to a measurement at the same time t.

As we saw, what can change is the precise initial conditions, as well as the precise time t given by our timepiece.

We label the possible states i's in ascending order of the energy E_{i}.

And for simplicity, we take it as given that no two states have exactly the same energy – which is typically true except in very unusual situations.

So we have a clear notion of what we mean by i.

We have a notion of what we mean by a distribution of probability P.

And there is a one-to-one correspondence between an average energy E and a probability distribution P, in a family of probability distributions.

Thus E can be viewed as a parameter labeling the distributions in a family of probability distributions.

That is why we use the notation $P(i, E)$ for the distribution of probability corresponding to E.

The nature of $P(i, E)$, which is the distribution of probability corresponding to a thermal equilibrium with average energy E, will become clear when we work out the Boltzmann distribution in part 4 of our discussion.

Before that, our objective in the present part is to explain temperature.

Each distribution of probability in the family labeled by E has its own entropy.

So the entropy becomes a function of E

$$S(E) = -\sum_i P(i, E) \log P(i, E) \tag{13}$$

The entropy, which is a feature of a given probability distribution, now becomes a function of the average energy because for each average energy we have a different probability distribution.

So that tells us that there is a connection between entropy and energy if we know the one parameter family suggested in figure 3.

And, as we said, we will discuss later where this family of distributions comes from.

Having linked entropy and energy, now we can ask an interesting question.

2.4 Energy, entropy and temperature

Consider each one of the probability distributions, parametrized by a given average energy, figure 4.

The higher the average energy, the more they will spread out to the right.

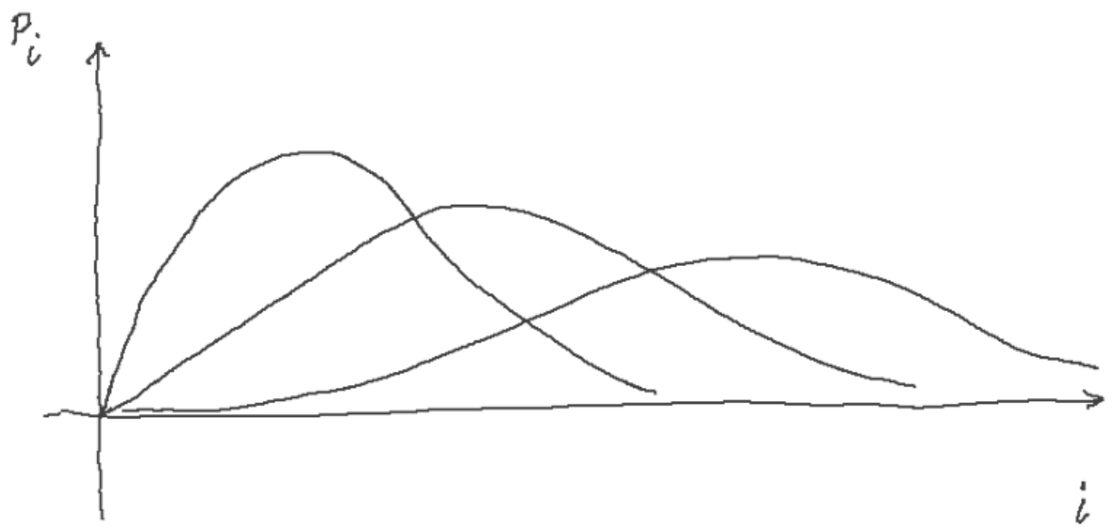


Figure 4: Family of probability distributions corresponding to different average energies.

We can ask: how much do we have to change the average energy in order to change the entropy by one bit or by a small amount ΔS?

To start with we can write the following identity from elementary calculus

$$\Delta E = \frac{\partial E}{\partial S}\Delta S \qquad (14)$$

where

$$\frac{\partial E}{\partial S} = \lim_{\Delta S \to 0} \frac{E(S + \Delta S, Z) - E(S, Z)}{\Delta S}$$

called the **partial derivative**

The letter Z is meant to represent the other variables, besides S, that E depends on.

If E only depends on one variable, i.e., S, then we have

$$\frac{\partial E}{\partial S} = \lim_{\Delta S \to 0} \frac{E(S + \Delta S) - E(S)}{\Delta S} = \frac{dE}{dS}$$

i.e., the partial derivative is the same as the ordinary derivative.

Equation (14) assumes nothing more than that energy and entropy are smooth functions that can be differentiated.

Equation (14) says that the change in the energy, in going from one probability distribution to the next, is the derivative of the energy with respect to the entropy times of change in the entropy.

We can also write it in the alternate and equivalent way

$$\Delta S = \frac{1}{\frac{\partial E}{\partial S}} \Delta E \tag{15}$$

Incidentally, we write partial derivatives because in some problems there might be more than one variable to keep track of.

But, as I showed earlier, there there is no difference with ordinary derivatives, and we could write dS/dE instead of ∂S/∂E.

So far equations (14) and (15) are trivial.

There is nothing insightful in them.

We just look at the variation of the energy when we vary the entropy, or equivalently the variation of entropy when we vary the energy.

By the way one over ∂E/∂S is nothing more than ∂S/∂E.

If we want to know for example how much do we have to shift the energy – that is go to a different point in the a probability distribution – in order to change the entropy by one bit, then we would just stick in for ΔS one bit, namely log 2, in equation (15).

That would tell us how much energy ΔE it takes to change the entropy by one bit.

Of course there is nothing special about 1 bit, all we want is to consider a small change in S and the related small change in E.

The main point is that the rate of variation

$$\frac{\partial E}{\partial S} = T \tag{16}$$

is called *temperature*.

The temperature is the rate of variation of the energy with respect to the entropy.

Or equivalently, 1/temperature is the rate of variation of the entropy with respect to the energy.

It is the more customary way to write it:

$$\boxed{\Delta S = \frac{\Delta E}{T}} \tag{17}$$

And this is the formula we see most frequently in classical thermodynamics books.

We even often see it written with human temperature and Carnot units of entropy, but as explained below it is equivalent, see equation (20).

We can also rewrite equation (17) with T upstairs, and drop the Δ's for d's instead.

Then we have

$$dE = TdS \tag{18}$$

As we said, temperature is a highly derived quantity.

Even though it is quite intuitive for us humans, in thermodynamics, be it classical or statistical, it is not a fundamental concept like entropy or energy, but a derived one, namely the rate of variation of the latter with respect to the former.

Equations (17) or (18) tell us that the higher the temperature, the more energy it takes to change the entropy.

Conversely at low temperatures, on an absolute scale where zero is the absolute zero, a small change in energy produces a big change in entropy.

In equations (17) or (18) the temperature T is the modern measure of temperature in physics.

If we switch to the human temperature t_{K} , measured in Kelvin, and to Carnot measure of entropy, equation (18) becomes

$$dE = t_K dS_{carnot} \tag{19}$$

To see that, simply observe that $T = k_{\mathrm{B}}\, t_{\mathrm{K}}$, equation (6), and $S = S_{\mathrm{carnot}}/k_{\mathrm{B}}$, equation (8).

If we introduce those in equation (18), the Boltzmann constant cancels and it yields equation (19), or equivalently

$$dS_{carnot} = \frac{dE}{t_K} \tag{20}$$

which is essentially what Carnot wrote in 1824.

He didn't have to know about Boltzmann constant because it cancels in going from (18) to (19), or from (17) to (20).

We mention this because if you go and look at any book on standard thermodynamics, you may wonder what happened to the constant k_B in our work?

It cancels out in the definition.

And equation (20) is the same as equation (17).

We have reached our objective of defining temperature.

It is an odd and abstract definition.

It is not in the least clear what equations (17) or (18) have to do with temperature as we experience it.

So let's see now what it has to do with our ordinary idea of temperature.

We are going to assume that entropy is a monotonic increasing function of energy.

That means that temperature, as defined in equation (17), is positive.

$dS/dE > 0$ by assumption for the moment.

This means that when we increase the parameter E in the family of distributions represented in figure 4, we spread the distribution over a larger number of states i, and more specifically its entropy increases.

We are going to prove most of these things, but let's take them as given for now.

So the question is: how is this definition of temperature in any way related to the familiar concept of temperature?

2.5 Link with the familiar concept of temperature

What is the familiar concept of temperature?

Of course we are familiar with how things can be hot, tepid, cold, etc.

We burn ourselves when we touch something too hot.

But can we do something that is more physics oriented?

The definition of ordinary temperature which comes closest to being general has to do with how two systems can exchange energy.

And it will actually be a measure of the difference in their temperatures.

Consider two systems connected as in figure 5.



Figure 5: Two systems which can exchange energy, but weakly.

And for simplicity let's suppose that the global system formed by the two of them is isolated.

We have system A and system B, connected by a thin pipe enabling them to exchange energy with each other, but only weakly.

Little bits of energy can go back and forth between them and cause the energies to fluctuate.

But on the average each one has an energy.

Imagine them having started out not in contact.

They each were in equilibrium, at different temperatures T_{A} and T_{B}.

These are the abstract temperatures defined in the preceding section.

It is the rate of change of energy with respect to entropy in each system

So each one of the boxes in figure 5 has been prepared with a given temperature by whatever it takes to make them have whatever temperature they have.

Now connect them.

This simply means allowing them to interact with each other and exchange energy back and forth

Which way does the energy flow?

In the ordinary concept of temperature, heat will flow from hot to cold.

In other words energy will go from the system of higher temperature to the system of lower temperature.

Now let's prove that this is also true with our abstract definition of temperature.

This will establish the link between ordinary temperature and the abstract definition we gave.

Suppose $T_B > T_A$. Then we will show that heat will flow from B to A.

We will also show that if $T_B = T_A$ then heat flows neither way.

Heat means energy now, and we use the two terms indifferently for the time being.

A very important assumption we will also make is the *second law of thermodynamics*.

Later on we will prove it, but for the moment let's take it as a postulate or a given.

The second law of thermodynamics says that the entropy of an isolated system, left to itself to evolve over time, can only increase or at best stay the same.

Remember that entropy is a measure of the incompleteness of our knowledge of the exact state of the system.

So the second law says that we will know less and less well in what state the system is.

Or, at best, the incompleteness of our knowledge will stay the same.

More specifically, in the case of the two boxes in figure 5 that have been prepared separately in thermal equilibrium, the second law of thermodynamics says that if we connect them, and let the global system to equilibrate again, heat will flow one way or another, until it eventually comes to some sort of compromise equilibrium.



But the entropy will have increased.

Let's now follow the line of reasoning showing that T_A and T_B are consistent with ordinary temperature.

The total entropy of the system is

$$S = S_A + S_B \tag{21}$$

Why is entropy additive?

It comes from the mathematics.

If you have two independent systems and each one has some sort of probabilistic description, the probability for the whole system to have some property multiplies the two.

If you take the logarithm of probabilities it adds.

That is essentially why entropy is additive

A second useful fact is that when the system equilibrates by exchanging energy the total energy is conserved.

This is the *first law of thermodynamics*.

So if the energy of A increases then the energy of B decreases by the same amount.

Or, written with an equation, we have

$$dE_A = -dE_B \tag{22}$$

Equivalently, the total change in energy of A and B is equal to zero

Next, the second law of thermodynamics says that

$$dS_A + dS_B > 0 \tag{23}$$

Entropy increases.

Let's stress that it is something we will need to justify.

Finally, for each box equation (18) holds, namely

$$\begin{aligned} dE_A &= T_A dS_A \\ dE_B &= T_B dS_B \end{aligned} \tag{24}$$

This will lead to the conclusion about temperature that we are looking for, but we are not yet ready to complete the argument - the discussion is taken up again and completed in part 3 of our discussion.

3. Maximizing entropy

3.1 Introduction

As usual we start with a brief review.

The entropy of a discrete probability distribution is defined as

$$S = -\sum_i P(i) \log P(i) \tag{1}$$

Here we are in the case where the system under study is in a state belonging to a discrete collection of possible states.

Our knowledge of which state the system is in is imperfect or incomplete.

This incompleteness is modeled with a probability framework [ε , Ω, P] which we described at length in part 1.

The probability that the system be in state i is $P(i)$.

Where the probability comes from, we have discussed a bit in part 1.

We will discuss it again later in greater detail in the case of a system made out of large number of molecules forming a gas in thermal equilibrium.

But for now we just assume that we know the distribution $P(i)$.

Attached to each state i is also an energy $E(i)$.

The assumptions, we should by now be familiar with, are

$$\sum_i P(i) = 1$$
$$\sum_i P(i)E(i) = \langle E \rangle \tag{2}$$

The second equation is pretty much a definition: the average energy of the system is $< E >$, which we will soon simply denote E.

The way to read the two equations is straightforward.

The sum of all the probabilities is equal to one.

And if we add all the energies, each weighted by its probability, we obtain what is called an average quantity in statistics or probability theory.

From equation (1), it looks like the entropy is negative.

But it is not.

Because each probability $P(i)$ is less than one, its logarithm is negative.

Therefore S is positive.

We saw earlier that the average energy E can be viewed as the parameter indexing a family of probability distributions $P(i, E)$.

For a given E the system has a given distribution of probability to be in the various states making up the set of possible states, which we called Ω.

Figure 1 shows three such distributions corresponding to three values of the parameter E.

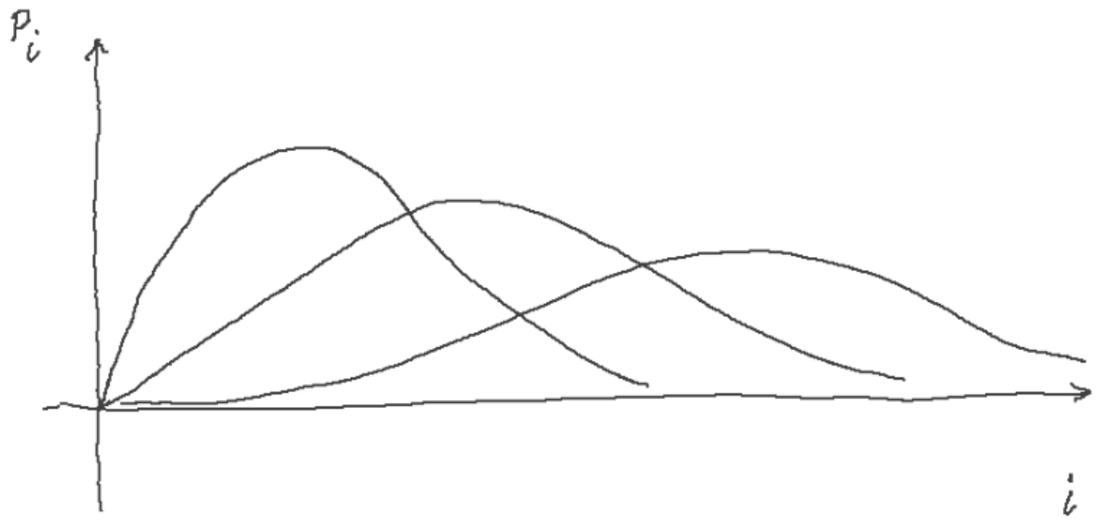


Figure 1: Family of distributions of probability. The states are arranged on the i-axis by increasing $E(i)$.

When the average energy gets larger, the probability distribution shifts to the right.

The distribution gets also broader.

And as a consequence its entropy gets larger too.

There is a particular probability distribution which is very special.

It is the one corresponding to the minimum value of the energy E.

It is of course the ground state of the system.

The ground state of a system is described by a probability which is zero everywhere except in the state of lowest energy.

In figure 1, it is the left-most state.

And the corresponding distribution is a peak of infinite height just at that value.

The entropy of the ground state is exactly zero.

Then as the one parameter family of probability distribution shifts to the right, first of all the energy becomes larger, and the entropy increases too.

Remember that the entropy is qualitatively simply a measure of the logarithm of the number of states which have a significant probability underneath the probability distribution.

When there are M occupied states with uniform probability, it is exactly true: $S = \log M$, as we saw in equation (23) of part 1.

When the blob of occupied states has a more general distribution of probabilities, the definition of S becomes equation (1) above.

$$S = -\sum_i P(i) \log P(i)$$

We checked in part 1 that it is consistent with the uniform case.

But it is more complicated.

It is now the average of the log $P(i)$'s.

And when we are in the continuous case, equation (1) transforms into its continuous analog $S = -\int \bar{P} \log P$, see equation (34) of part 1.

In summary, as the probability distribution gets wider and wider, in the family of distributions that can describe the system, see figure 1, the average energy E, which is also the indexing parameter in the family, goes up.

And at the same time the entropy goes up.

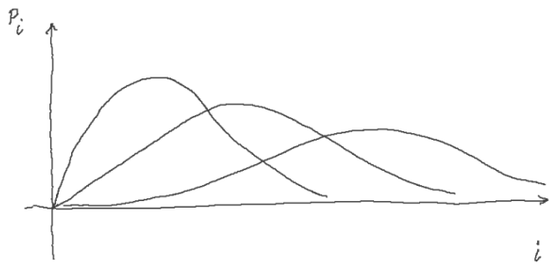
We will always assume that the entropy is a monotonically increasing function of the average energy.

Of course one can invent probability distributions for which that is not true.

It is possible to make up a family of probability distributions which get narrower and narrower as the energy increases.

But, as we will see, the probability distributions that govern thermal equilibrium, at different levels of energy, are not like that.

We will find out that exactly figure 1 is the picture that governs the probability distributions for the states in thermal equilibrium.



Thermal equilibrium is characterized by a temperature T of course, but we can also characterize it by an average energy E.

Suppose we have a box of gas, and we know the average energy in it, and it is in thermal equilibrium.

Thermal equilibrium means that whatever was going to happen has finished happening.

Now it is simply quiescent and the molecules are all thoroughly mixed up, thoroughly dispersed throughout the system.

Not much is happening other than microscopically of course.

It is perfectly tranquil, just every molecule moves at a few hundred meters per second

Then, in this state of thermal equilibrium, heat tends not to flow.

We will understand that soon enough.

Heat tends not to flow because all parts of the system are in equilibrium.

And, as we will see, the probability distributions $P(i, E)$ do in fact broaden as the energy E – or parameter E, if you like – goes up.

As a consequence the entropy is a monotonically increasing function of energy.

That is an important fact.

We are going to go now through some of the laws of thermodynamics.

At first again it is a review.

3.2 Laws of thermodynamics

The *first law* is just energy conservation.

The *zeroth law* was stated when people were axiomatizing thermodynamics in early 20th century .

In their approach it had to come before almost anything else.

But today we think of the zeroth law as a consequence of the first law and the second law in a sense.

We will talk about the zeroth law after we talk about the second law.

The *second law* of thermodynamics says that entropy increases.

Now what does that mean?

It means that if we have a system, described by a given probability distribution, and it is not in equilibrium, and it evolves toward equilibrium, generally speaking the probability distribution broadens.

We are going to prove that eventually, in a qualitative kind of way.

We will prove qualitatively the assertion that if we have a probability distribution and we allow it to evolve, being a little bit careless in our observation of it – that is called *coarse graining* –, we will find that the probability distribution always broadens.

It should not be surprising.

For a probability distribution to narrow, that would mean we found out more about the system than we originally knew.

Typically if we are considering a very complex system, with a large number of degrees of freedom, very difficult to keep track of, exactly the opposite tends to happen.

We lose track of things.

And because we lose track of things, the probability distribution tends to broaden.

Hence the increase of entropy with time.

So we have the first law: energy is conserved.

We have second law: entropy always increases with time.

If it doesn't increase, it stays the same.

But it never decreases.

Let's assume that the entropy always increases or stays the same.

Now let's talk about the *zeroth law*.

First of all the zeroth law asserts that there is such a thing as thermal equilibrium.

If we take a box, for instance a box of gas, and we wait long enough, it will come to equilibrium.

We take the example of gas because it is easy to think intuitively about how all the molecules positions and velocities should evolve.

But it doesn't have to be gas inside the box, it could be liquid, it could be solid, or any mix.

In fact it could be any system which is isolated and self-contained.

It will come to equilibrium.

What does that mean for it to come to equilibrium?

Imagine that the system is made out of separate subsystems.

Let's call them A and B.

And let's put in a third subsystem C.

Assume that they are all connected.

By that we mean that they can exchange energy with each other.

So we put some little pipes between them that allow energy to flow, figure 2.

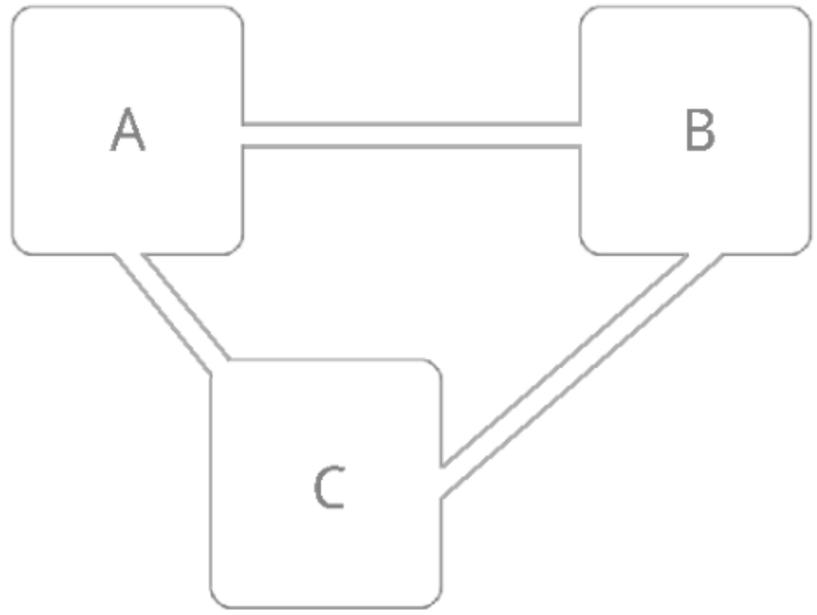


Figure 2: Subsystems exchanging energy through small interfaces.

The zeroth law also says that there is a concept called temperature attached to any equilibrium.

Temperature has the characteristic that, if the system is made of several parts separately in equilibrium, and connected together through small interfaces as in figure 2, the energy always flows from hotter parts to colder parts.

And if we wait long enough the flows will eventually come to an equilibrium, all the parts of the system having then one unique temperature.

Let's suppose there is only two systems A and B for a moment.

Energy will flow from the one with higher temperature, say B, to the one with lower temperature, say A.

Until what point?

Could T_A become higher than T_B?

No, that is not what happens.

The flow of energy from B to A will go on until they equilibrate.

And equilibrate – we will see – means the temperatures become equal.

When the temperatures on both sides have become equal the heat stops flowing.

That is the meaning of temperature, if we like: it tells us which way heat flows.

From this sort of axiomatic description of the laws of thermodynamics, we have the following rule:

two systems are in equilibrium with each other if and only if their temperatures are the same.

As a consequence, if A is in equilibrium with B, and B is in equilibrium with C, then $T_A = T_B = T_C$.

And therefore A is in equilibrium with C.

In other words, to be in equilibrium is an equivalence relation.

Note: The relation *is equal to* is the canonical example of an equivalence relation.

An example is "to share some (unspecified) quality".

We have simply replaced a qualitative notion of equilibrium by a quantitative one, through the use of temperature.

Then the logical properties of equality apply to equilibrium.

We can summarize the zeroth law with these three rules:

1. There is a notion of temperature. We have already introduced it, but we will use it to illuminate this law, see next section.

2. Energy flows from higher temperature to lower temperature.

3. In thermal equilibrium, the temperature of all parts of the system is the same. If the temperature was not the same in all parts of the system, energy would flow until it became the same.

That is the basic zeroth law: the existence of a *temperature* function that tells us which way energy flows; and the existence of an *equilibrium* when the temperature in all parts of the system is the same.

And isolated systems evolve toward equilibrium.

Let's see now if we can relate those ideas to the notion of temperature that we introduced in part 2, namely the rate of variation of energy with respect to entropy.

3.3 Energy, entropy and temperature

If we change the average energy of a system a little bit, the entropy will change.

The rate of change is called the temperature.

$$\frac{dE}{dS} = T \tag{3}$$

Let's see if we can figure out what this definition of temperature has to do with the idea of the flow of heat and the way it goes.

We begin with a system composed of two parts, A and B, and, as usual, a little pipe between them, figure 3.



Figure 3: Isolated system made of two subsystems, to study the direction of flow of heat.

Let's write down all the basic equations that we know.

For the moment we assume that the temperature of B is higher than the temperature of A.

Of course we could have $T_B = T_A$, but let's start with the case where they are not equal.

$$T_A < T_B \tag{4}$$

That implies that the subsystems A and B are not in equilibrium with each other.

Or said another way, the global system is not in equilibrium.

The first law tells us that when the system responds and does whatever it does over time, the total energy doesn't change.

Since the whole system is not in equilibrium, there will be some energy shift from one subsystem to the other.

Whichever way it happens, the following equation holds

$$dE_A + dE_B = 0 \tag{5}$$

For instance, we take our timepiece and we wait one second during which we watch the energy redistribute itself, figure 3.

We find out that the energy of A changes; and the energy of B changes.

But energy conservation tells us that the sum of the energies doesn't change.

Next, let's look at the change in the entropy of *A* plus the change in the entropy of *B*.

That is the change of the total entropy in fact.

It is greater than zero.

In some very special cases it can be equal to zero, but the general statement is that it is greater than zero

$$dS_A + dS_B > 0 \tag{6}$$

The probability distribution of the combined system shifts in such a way as to become broader.

That is what equation (6) says.

Now let's use one more statement: that the change in the energy is equal to the temperature times the change in the entropy.

And let's apply this to both A and B.

$$\begin{aligned} dE_A &= T_A dS_A \\ dE_B &= T_B dS_B \end{aligned} \tag{7}$$

Then equation (5), namely the first law of thermodynamics, rewrites as

$$T_A dS_A + T_B dS_B = 0 \tag{8}$$

Let's solve it for dS_{B}. We get

$$dS_B = -\frac{T_A}{T_B} dS_A \tag{9}$$

Now we take this formula and plug it in equation (6).

This is a general method when doing physics: you first think really hard about what you want to do, and then you blindly just go ahead with the equations and see where they lead.

So plugging equation (9) into equation (6) yields

$$dS_A - \frac{T_A}{T_B} dS_A > 0 \tag{10}$$

That is what the second law of thermodynamics says.

Let's multiply equation (10) by T_B.

We will assume temperatures are positive, so that doesn't change the direction of the inequality.

There are situations in which temperature can be negative, but not for us for the moment.

So we get

$$T_B dS_A - T_A dS_A > 0$$

or

$$(T_B - T_A) dS_A > 0 \tag{11}$$

We have already assumed that the temperature of *B* is greater than the temperature of *A*.

Then what does equation (11) say about dS_A?

It says that

$$dS_A > 0 \tag{12}$$

The entropy of the subsystem *A* increased.

The same reasoning leads to $\mathrm{d}S_\mathrm{B} < 0$.

But it is not because the total entropy is conserved.

The total entropy is not conserved, it increases.

The entropy of B decreases because of the presence of the cold heat sink A.

And A being colder than B, it requires more ΔS for the same ΔE than B lost.

Let's multiply equation (12) by T_A.

We get $T_\mathrm{A}\,\mathrm{d}S_\mathrm{A} > 0$.

But what is $T_\mathrm{A}\,\mathrm{d}S_\mathrm{A}$?

It is the change in energy of the subsystem A.

$$dE_A > 0 \tag{13}$$

So what have we found?

By a little bit of manipulation and the use of a couple of laws we found that the change in the energy of A is positive.

We used the first law of energy conservation, the second law of entropy increase, plus some simple algebraic operations.

It follows of course that the change in the energy of B is negative.

In other words heat has flowed – or energy has flowed – from B to A.

So we have proved that, with the definition of temperature as given here – namely the rate of change of E with respect to S –, it indicates a direction of heat flow.

Again when we say heat, so far we just mean energy.

Energy will flow from B to A until the temperatures become equal.

When the temperatures become equal of course if T_{A} equals T_{B} then there is no flow.

That is easy to see.

And that is when equilibrium is established.

So no flow of heat, all temperatures equal, and equilibrium are three equivalent situations.

So that is the zeroth law of thermodynamics.

3.4 Questions/answers

Question: We assumed that the subsystems A and B and C were each separately in equilibrium. What happens if A, for instance, is not in equilibrium?

Answer: Yes, that is true: we assumed that each subsystem was in equilibrium with itself so to speak.

But if one is not in equilibrium, you split it into subsystems which are in equilibrium and then apply the same analysis.

Question: You described the collection of possible states as discrete.

But if a state is the exact specification of the position and momentum of each molecule, doesn't it belong to a continuous space Ω?

Answer: Yes, I did discretize the problem.

But first of all in quantum mechanics we can take those variables to be discrete basically.

On the other hand, we could also use a continuous description of the space of states and the probabilities.

As you know, we would then use densities of probability.

But nothing would change essentially in all our reasonings and conclusions.

Most of the time, the two approaches are equivalent. And we choose the one that is most convenient mathematically.

The important point to remember is that a state always means the same thing.

It means "as much as you could possibly know about the system if you were an infinitely powerful observer".

Then how we decide to represent and manipulate it, is a purely mathematical question.

And we often have the choice of a continuous representation or a discrete representation.

Question: But if we have say a finite number of molecules in the box, their positions and velocities are not discrete variables.

Answer: In quantum mechanics they are.

In quantum mechanics there is a discrete state counting for a finite box.

What we will have to show is that if we do the correct quantum mechanical statistical mechanics, then in certain limits it becomes equal to classical mechanics.

In other words in certain limits you can replace sums by integrals. That is what it comes down to.

And that is not a one liner.

That is a subtle business.

But it is true.

Let's now turn to the probability distribution $P(\mathrm{i})$'s after equilibrium has been established.

3.5 Distribution of probability at equilibrium

We consider our box of gas, or whatever system we are interesed in, after equilibrium has been established.

There is some probability distribution $P(\mathrm{i})$ for the system to be in its different microstates i's.

What is that probability distribution?

What is the mathematics of it?

First of all, it is always a useful idea to imagine that our system is not closed, but is in contact with the outside world and can exchange energy with it.

It is not only useful, but it usually corresponds to the facts.

In particular, it is a useful idea, if we want to think about the probability distribution of the states of our system, to think of it as being embedded in a huge reservoir, of a much larger number of degrees of freedom, that provides a *heat bath*.

The heat bath allows energy to flow back and forth between itself and the system.

After a while, the system comes to thermal equilibrium with the big surrounding heat bath.



Figure 4: System plunged into a big heat bath.

The heat bath is so big that even if a little bit of energy flows into the system the temperature doesn't change much.

This is something we could also prove.

But let's just think about the physics: we have a huge system at a certain temperature – the heat bath –, and we have a little system plunged into it – our system.

A little bit of heat flows from the big system to the little system.

Typically the change in the temperature of the big system will be negligible.

So we can imagine that the big system is at some temperature and we simply wait until the small system comes to equilibrium with the big system and then is at that same temperature.

A particularly convenient choice of heat bath – we will see why –, which often actually corresponds to a fact, is just to imagine that the system in question is one of a very large number of identical systems, which are connected together by little pipes that allow heat to flow back and forth, figure 5.

Figure 5 is only suggestive.

It doesn't represent the exact connection between subsystems.

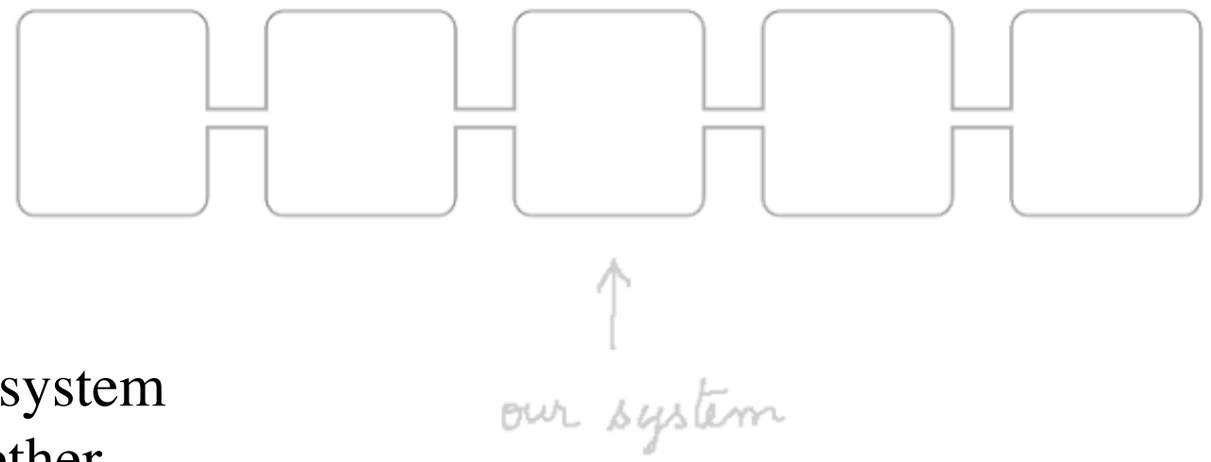


Figure 5: The heat bath is made of a large number of systems identical to the one under study. There are N identical systems.

The idea is that the system under consideration – our system – is one of many identical systems, all connected together.

And all but ours provide the heat bath.

Actually this sometimes corresponds more or less to facts.

In particular if we have a very big system and we divide it into numerous small subsystems, the small subsystems may be very much alike and the big system minus one of them then is just the heat bath to that one.

This is a useful trick to pretend that the heat bath is just a repetition of the same system over and over again.

How many such systems?

Eventually we will have a large number of them, large enough that we can think of the heat bath as very big.

But let's call it capital N for the moment.

Each one of the N systems is in a state.

The collection of possible states as usual is $\Omega = \{ \omega_1, \omega_2, \omega_3, \text{etc.} \}$.

To fix ideas, and prepare for some combinatorics, let's consider that there are $N = 9$ identical systems, and they each can be in one of three states ω_1, ω_2, ω_3, figure 6.

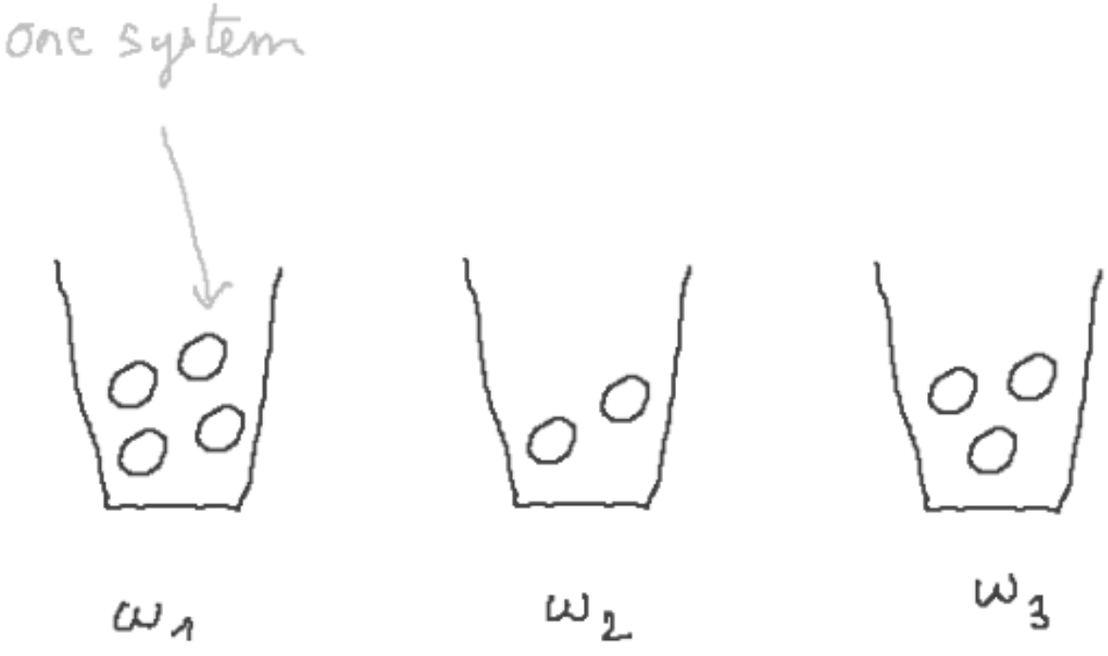


Figure 6: Nine systems distributed among three states.

The three states $\omega_1, \omega_2, \omega_3$ of Ω are not identical.

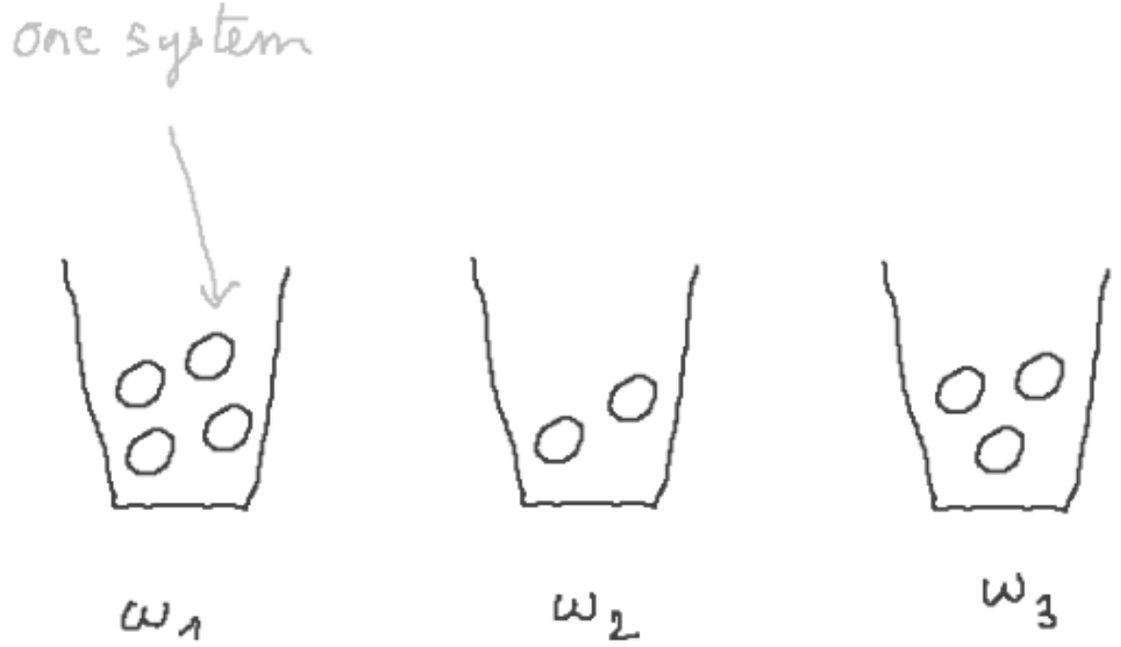
They have different energy for example.

Some of the N systems are in state ω_1.

In figure 6, there are four of them.

Generally speaking we will denote their number n_1.

Similarly there are n_2 systems in state ω_2.



In figure 6, $n_2 = 2$.

There are n_3 systems in ω_3.

In figure 6, $n_3 = 3$. And so forth.

The energy might increase from ω_1 to ω_3.

In other words in figure 6 the states might be viewed as forming an axis of increasing energy.

We represented three states, but there might be an infinite number of states (not to be confused with N).

For convenience I prefer now to represent the states of increasing energy on a vertical axis, leaving, for later on, the horizontal axis to time, figure 7.

In fact in general there will be an infinite number of possible states ω_i's.

But most of them will have so much energy that there is no chance at all that they be occupied.

Only a negligible fraction of the *N* systems will occupy states of very high energy – if anything simply because we may not have that much energy available.

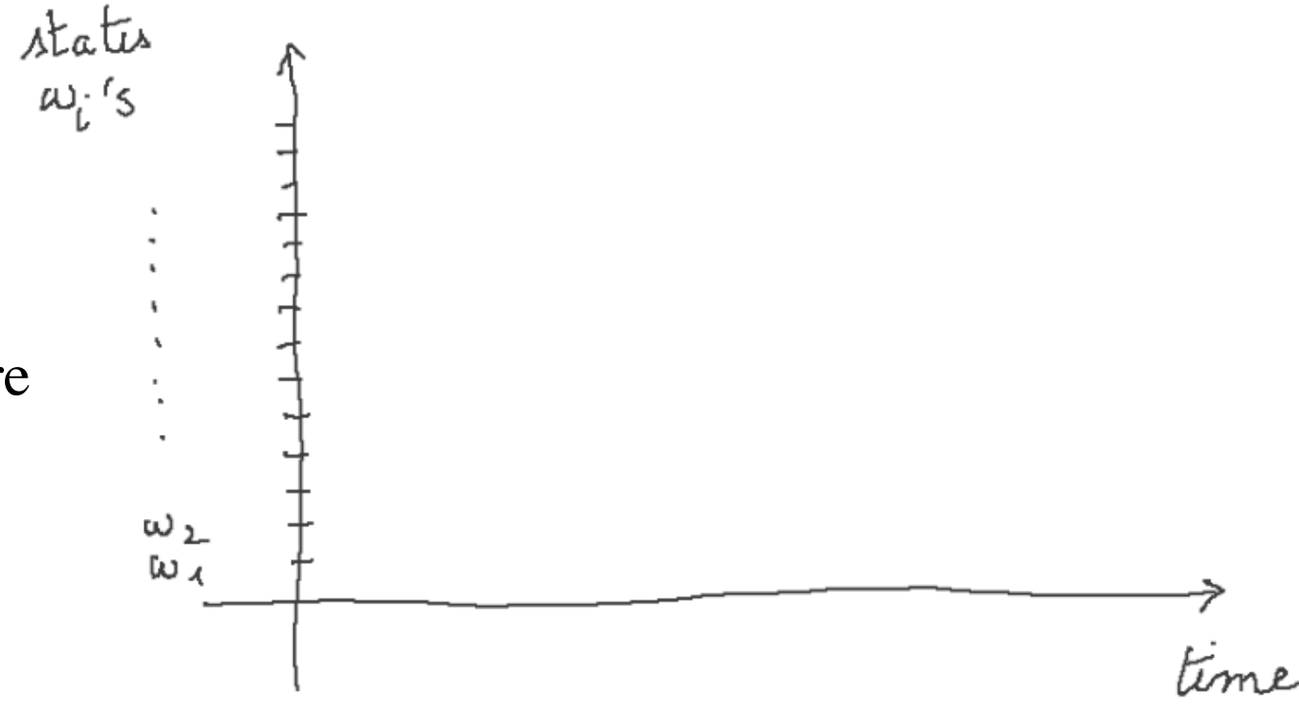


Figure 7: States of increasing energy.

The counts n_i's are called the *occupation numbers*.

A first question is: given the occupation numbers, how many ways are there of redistributing the *N* systems among the states so as to create that set of occupation numbers?

Obviously that is going to be related to the *P*(i)'s.

The more ways there are of distributing these *N* systems into a given set (n_1, n_2, n_3, ...) of occupation numbers, the more likely that set of occupation numbers is.

Or equivalently, the more likely are the probabilities $P(\mathrm{i})$'s defined as $P(\mathrm{i}) = n_\mathrm{i}/N$.

That is a symmetry argument.

If, in an imaginary random experiment distributing the systems into states, all possible redistributions are symmetric with respect to each other and therefore equally probable, then the most probable set of occupation numbers, is that which corresponds to the maximum number of ways of redistributing things.

Figuring out the most probable set of occupation numbers becomes a maximization problem where the unknowns are the n_i's.

Without further conditions it would be straightforward to see that the solution is a uniform collection. But there are constraints.

First of all we can see it as a constraint that the following must naturally hold

$$\sum_i n_i = N \tag{14}$$

A second constraint has to do with energy.

The total energy of the N systems is a fixed quantity.

We are going to denote that fixed quantity N times E, because it is natural to think of it as proportional to the number of boxes, or systems.

So let's write that constraint

$$n_1 E_1 + n_2 E_2 + n_3 E_3 + \cdots = NE$$

or

$$\sum_i n_i E_i = NE \tag{15}$$

where E is the total energy divided by the number of boxes.

So it is a fixed figure which is the average energy of a box.

It is the average, at a given time, over all the boxes.

And it is also the average, over time, of a given box.

Now suppose we have found the set $(n_1, n_2, n_3, \ldots)$ which corresponds to the maximum number of ways to distribute the N boxes into the states $\{\omega_1, \omega_2, \omega_3, \ldots\}$, under the constraints (14) and (15).

Picking one box at random uniformly over the N boxes, what is the probability that it be in the i-th state?

The answer is

$$P(i) = \frac{n_i}{N} \tag{16}$$

Remember that talking about probabilities only has a meaning if there is a clearly defined random experiment ε, a set Ω which is the space of states of the world after having performed ε, and a probability distribution, or measure, P over the elements or subsets of Ω.

As we saw in part 1, when we are interested in only one random variable X, namely, the state in which a box is, we have a choice of definition of Ω and P.

We can choose Ω to be the collection of boxes, and P to be, for instance, uniform over them.

Or we can choose Ω to be the set of possible values of X, here the states that the boxes can fall into, see figure 6, and modify P accordingly.

That is what we did with equation (16).

Let's now use equation (16) to rewrite the two constraints (14) and (15).

Simplifying our notation for P(i) into P_{i}, they become

$$\sum_i P_i = 1$$
$$\sum_i P_i E_i = E \tag{17}$$

So these are now the new expressions for our two constraints.

But if we prefer to think about them in terms of the occupation numbers, we can write them back as equations (14) and (15).

Before we worry about the total number of systems being kept constant, and the total energy being kept constant, let's just ask a more primitive question: for a given set of occupation numbers n_1 through n_p, or however infinitely many of them there are, how many ways are there of redistributing our N systems to correspond to that collection of occupation numbers?

That is a problem in combinatorics.

You can find the solution in any good book on the subject.

I will just give the answer, and then we will check it for a couple of cases.

The answer it not the same depending on whether the N boxes are distinguishable or not.

For instance, if the boxes are b_1, b_2, b_3, and the states are ω_1, ω_2, if the boxes are distinguishable, then (b_1, b_2) and (b_3) is not the same as (b_1, b_3) and (b_2).

When they are distinguishable, the number of arrangements, denoted A, of the N boxes into (n_1, n_2, n_3,...) states, as in figure 6, is

$$A = \frac{N!}{n_1!n_2!n_3!\cdots} \tag{18}$$

For any positive integer m, the symbol $m!$ – read m factorial – is by definition the product of all the integers from 1 to m.

Also by convention $0! = 1$.

Then, thanks to this convention, whether the number of possible states $\{\omega_1, \omega_2, \omega_3, \ldots\}$ is finite or countably infinite doesn't make any difference, because when $n_i = 0$, in formula (18) it corresponds to dividing by 1.

So far, equation (18) gives the number of arrangements for a fixed set of occupation numbers, with no regard to whether the constraints are satisfies.

We will come back to the question of the constraints being satisfied.

But let's study the right-hand side of equation (18) on its own.

$$A = \frac{N!}{n_1!n_2!n_3!\cdots}$$

It is an interesting expression.

We are going to want to approximate it when N and the n_i's become large.

Before that, let's check it for a couple of cases.

First of all, suppose $n_1 = N$ and all the other occupation numbers are 0.

Then formula (18) gives $A = N!/n_1! = 1$.

That seems reasonable.

Indeed, how many ways are there of putting all the N boxes into one given state? One.

So formula (18) is fine in this case.

Consider now the case where $N = 2$, that is there are two systems to distribute.

And suppose there are two states.

The case $n_1 = 2$ and $n_2 = 0$, we have already done.

So let's look at the case $n_1 = 1$ and $n_2 = 1$.

Calling the boxes, or systems, objects, and the states cups, as suggested in figure 6, how many ways are there to put two objects into two cups?

If the objects are distinguishable, there are two ways.

Formula (18) gives $A = 2!/(1!\ 1!) = 2$. So it is still ok.

Another case: $N = 3$, and $n_1 = n_2 = n_3 = 1$.

Binomial coefficient $\binom{n}{m} = \frac{n!}{m!(n-m)!}$

There are six ways.

And formula (18) gives that correct number.

One last case: $N = 4$, and $n_1 = 2$, and $n_2 = n_3 = 1$.

There are $\binom{4}{2} = 4 \times 3 / 2 = 6$ ways to fill the first cup, multiplied by two remaining ways to fill the other two.

This gives 12 ways.

Turning to formula (18), we have $A = 4!/(2!\ 1!\ 1!) = 12$. Again it is hunky dory.

So the general formula for A is formula (18).

It is not too hard to prove.

It is high school or freshman level.

It gives the number A of ways of distributing N objects into a collection of cups, with given occupation numbers $n_1, n_2, n3$, etc.

Although we haven't specified it with fancy notations, A is a function of the occupation numbers n_i's.

The interesting fact we are going to discover is that, in the circumstances that concern us, of a large number of systems distributed among a collection of states, with some constraints, the function A is highly peaked at a particular set of occupation numbers.

Namely, when N gets very big, the occupation numbers cluster very strongly about a particular set of occupation numbers – or better yet a particular set of fractions n_i/N.

The distribution will get tighter and tighter as N gets up, so that the probabilities n_i/N will be well-defined.

In order to do that we ought to have some approximation method.

3.6 Approximation using Stirling formula

We want to approximate

$$A = \frac{N!}{n_1!n_2!n_3!\cdots}$$

The first thing we need to approximate is factorials.

When we say approximate, we mean approximate under the circumstances that the n_i's are all very big.

We allow the number N of subsystems to get bigger and bigger in such a way that the occupation numbers also simultaneously get bigger and bigger.

So all the n_{i}'s are going to be assumed to be very big.

And let's approximate the number A.

We will use the well-known Stirling approximation for factorials.

$$\lim_{N\to+\infty} \frac{N!}{\sqrt{2\pi N} N^N e^{-N}} = 1 \tag{19}$$

In other words, as N gets large, $N!$ can be approximated by

$$N! \approx \sqrt{2\pi N} \frac{N^n}{e^N} \tag{20}$$

In order to prove this we are going to work with logarithms.

$N!$ is the product of the integers from 1 to N. So we have

$$\log N! = \sum_{k=1}^{N} \log k \tag{21}$$

Now we are going to use the curve of the continuous function log x to approximate the right-hand side, figure 8.

The proof works with the logarithm calculated with respect to any base, 2, 10, e, or whatever, it doesn't matter, and we don't need to specify it.

But we can think of Napierian logarithm if we like, that is the natural logarithm to the base e.

We see that $\sum_{k=1}^{N} \log k$ is slightly smaller than $\int_1^N \log x \; dx$

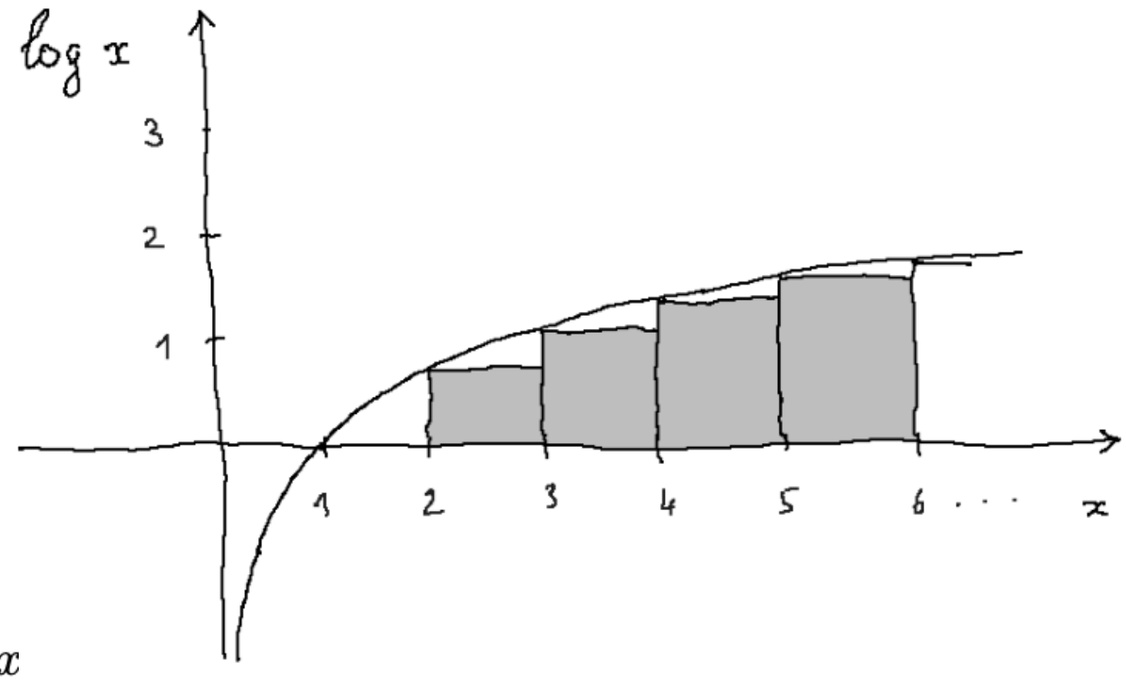


Figure 8: Curve $\log_e$ x and sum $\sum_{k=1}^{N}$ log k (in grey).

But also, had we drawn the grey rectangles shifted one unit to the left, we would see that it is slightly bigger than $\int_1^N \log x \; dx$

Thus we have

$$0 < \left[\sum_{k=1}^{N} \log k \; - \int_1^N \log x \; dx \right] < \int_N^{N+1} \log x \; dx \tag{22}$$

A integral of log x is x log x – x.

Therefore

$$\int_1^N \log x \; dx = x \log x - x \Big|_1^N$$
$$= N \log N - N - 1 \log 1 + 1$$
$$= N \log N - N + 1$$

And the upper bound on the right of expression (22) is smaller than log(N + 1).

So the whole expression can be rewritten

$$0 \; < \; \left[\sum_{k=1}^{N} \log k - \; N \log N + N - 1 \right] \; < \; \log(N+1) \tag{23}$$

If we now take the exponentials of the three terms, noting that $e^{N \log N}$ is the same as N^N, we get

$$1 < \; \frac{N! \; e^N}{N^N \; e} \; < \; N + 1 \tag{24}$$

In other words,

$$N! \; \approx \; \frac{N^N}{e^N} \; C(N) \tag{25}$$

where C(*N*) is a multiplicative factor which grows no faster than *N*.

With further calculations, which are not necessary for our purposes, C(N) can be refined into $\sqrt{2\pi N}$ giving the well-known approximation of equation (20).

And there exist even better approximations if need be.

In the subsequent calculations, we will simply use

$$N! \approx \frac{N^N}{e^N} \tag{26}$$

where the ≈ sign here doesn't mean that the ratio of the left-hand side to the right-hand side goes to 1, as it usually means, but that it is bounded by N itself – and in fact the multiplicative factor that we omit is $\sqrt{2\pi N}$

Now we can turn to approximating our monstrous combinatorial coefficient

$$A = \frac{N!}{n_1!n_2!n_3!\cdots}$$

Remember that it is the number of ways of rearranging N systems into into a collection of states $\{\omega_1, \omega_2, \omega_3,\ldots\}$, under the constraint that there be n_1 systems in the first state, n_2 systems into the second state, n_3 into the third, and so forth.

What we are going to do eventually is to maximize it.

We shall look for the set of n_i's which makes this number of ways of rearranging as big as possible.

And we will find a very sharp function. But we are not there yet.

First we need to deal with the coefficient A.

We want to maximize A with respect to the n_i's, but subject to the two constraints that we wrote before, namely

$$\sum_i n_i = N \tag{27}$$

$$\sum_i n_i E_i = NE$$

What is the set of occupation numbers which makes A as big as possible, under these two constraints?

That is the mathematical problem we want to solve.

Once we have found the solution, then we effectively know all of the probabilities $P_i = n_i/N$.

How do we maximize a quantity with respect to a variable?

We differentiate it.

Here our quantity is a big product.

That is unwieldy to work with.

But we can equivalently work on maximizing its logarithm, because logarithm is a monotonic function.

Since what we are interested in are the maximizing n_i's, it doesn't change the solution.

A second hurdle is that the n_i are not continuous variables but integer variables, so it looks like we cannot use calculus and are stuck.

But we shall replace the n_i's by the P_i's.

And since N is big, we can treat the P_i's as continuous variables.

But first of all, before going to logarithms, and to continuous variables, let's use Stirling approximation which we derived above in its simple form of equation (26).

A can be rewritten

$$A \approx \frac{N^N e^{-N}}{\prod_i n_i^{n_i} e^{-n_i}} \tag{28}$$

where $\prod$ is the standard symbol for a product of terms.

Notice that in the denominator we can factorize separately all the e^{-n_i}.

Taking into account that all the n_i's add up to N, this gives a factor e^{-N} downstairs.

It cancels with the one upstairs.

So equation (28) simplifies into

$$A \approx \frac{N^N}{n_1^{n_1}\ n_2^{n_2}\ n_3^{n_3}\ \ldots} \tag{29}$$

Now we take the logarithms.

$$\log A \ \approx\ N \log N - \sum_i\ n_i \log n_i \tag{30}$$

Now we can start to smell something interesting.

Remember that the n_i's, when divided by N, can be interpreted as probabilities.

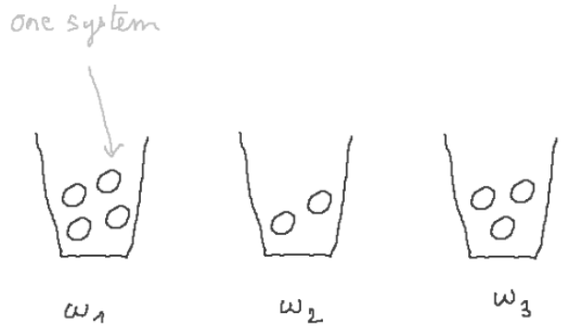
So equation (30) begins to look awfully interesting.

The second term on the right-hand side is not quite $-\sum_i P_i \log P_i$.

But it is somehow related to it, because the n_i's are related to the P_i's.

And what is $-\sum_i P_i \log P_i$?

It is the entropy of the probability distribution P_i.



So let's complete the intermediate steps, and see that what we are doing, when we are looking for the maximum number of ways of rearranging objects into cups as in figure 6, is maximizing some entropy.

We are finding the probabilities which maximize the entropy, subject to the constraints that first of all they add up to one, and secondly that the total energy is some fixed value.

Let's plug into equation (30) the n_i's expressed as probabilities multiplied by N.

We get

$$\log A \approx N \log N - \sum_i NP_i \log NP_i$$

$$\approx N \log N - \sum_i NP_i (\log N + \log P_i)$$

There is a nice cancellation again.

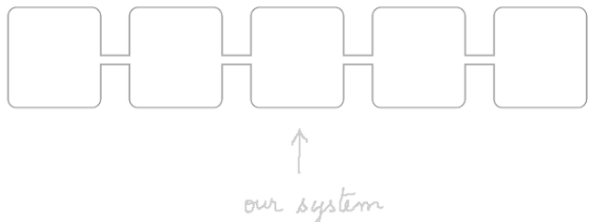
In the summation term, the sum of the $N P_i \log N$'s gives $N \log N$.

So the whole thing can be rewritten

$$\log A \approx -\sum_i N P_i \log P_i$$

or (31)

$$\log A \approx -N \sum_i P_i \log P_i$$



So we arrive at N times the entropy of any one of the systems in figure 5.

Indeed, remember that the P_i's now are the probabilities for a given system to be in the state ω_i, or said more simply state i.

Thus, apart from the factor N which appears in equation (31) because there are N subsystems altogether, what we want to maximize is the entropy.

Or said another way, if we want to find out the occupation numbers which maximize the number of ways that we can rearrange the N systems, keeping the occupation numbers fixed, it simply corresponds to maximizing the entropy.

3.7 Questions/answers

Question: If we look at the formula $N!/n_1!\ n_2!\ n_3!\ ...n_p!$, that would be maximized when all the n_i's are equal.

What is the meaning of that fact?

Answer: It means that, if there was no constraint related to energy, the distribution of probability over all the possible states would be uniform.

All the states would be equally probable.

But that mathematical result is of little interest, because it doesn't correspond to the physical situation we are examining.

We must add the constraint that $\sum_i n_i = N$.

Secondly and most importantly we must add the constraint that the total energy of the big system formed by the N identical little systems in figure 5 is a fixed quantity.

We denoted it NE, in order that E be the mean energy of any little system.

Then it is no longer true that the collection of occupation numbers, or equivalently, the probability distribution is uniform.

Question: In using Stirling formula we assumed that N gets large.

Do we have to justify that the n_i's get large too?

Answer: As N gets large, the n_i's will get large in the same proportion.

If you double the total number of systems, every average occupation number will also double.

I suppose we can justify it honestly as follows.

We chop the big system off into a number of small systems, and say: look I know that I don't have more energy than a certain amount.

So we can ignore states above that total energy anyway, see figure 7.

Then as N increases, the little n_i's can't but all increase, at least to maximize A under the two constraints.

Question: Is that what you meant when you mentioned coarse graining?

Answer: No. Coarse graining is an entirely different thing.

Coarse graining, so to speak, is what happens when you look at the world through somebody else's glasses.

Question: We arrived at equation (30) after having taken the logarithm of something we wanted to maximize.

But now in equation (31) we have a minus sign coming from something that was in the denominator.

Shouldn't we want to minimize it?

$$\log A \approx -N \sum_i P_i \log P_i$$

Answer: The minus sign does come from terms that were downstairs.

But don't forget that the P_i's are less than 1, so their logarithms are negative.

That is why the formula for the entropy has a minus sign in front of it.

We do want to maximize the right-hand side of equation (31).

We want to maximize the entropy, with respect to the P_i's, but subject to constraints.

It is certainly getting somewhat complicated.

But if we want to understand where the probability distribution, over the possible states which any of the little systems can occupy, comes from, we have to do it.

If you don't want to see only the slick general public scientific magazine version of it, but you want to see how it really works, you have to get your hands dirty.

We see that we cannot use a simple symmetry argument on the states $\{\omega_1, \omega_2, \omega_3,...\}$, which would lead us to some uniform distribution.

The states are different.

They have different levels of energy.

As time passes, in any one box in figure 5, the agitation of the molecules lead to some micro-exchanges of energy with the neighboring boxes.

The box goes through a collection of states, with energy E_i's.

But they have an average energy E, which is simply the total energy of the big system divided by N.

This transforms into a constraint when looking for the probability distribution with maximum entropy.

So our job then is to maximize $-\sum_i P_i \log P_i$ subject to constraints.

Maximize with respect to what?

Maximize with respect to the choice of probabilities P_i's.

If we go back, we are maximizing it with respect to the possible occupation numbers.

But that translates into maximizing the entropy.

Remember that entropy has a value which depends on the probability distribution.

Now we come to the conclusion that the most likely collection of occupation numbers corresponds to probabilities which maximize the entropy.

$$-\sum_i P_i \log P_i \tag{32}$$

But we want to maximize the expression subject to two constraints, namely

$$\sum_i P_i = 1$$

$$\sum_i P_i \; E_i = E \tag{33}$$

The average energy E of any subsystem is a fixed number, and so are the E_i's.

They are givens of the problem.

And of course the total probability must be equal to one.

And we look then for the P_i's which maximize the entropy.

That is the rule, and therefore the problem.

It is the second constraint in equations (33) which breaks the symmetry between the different states.

Some microstates have a big energy E_i and some have small energy E_i.

So we have reduced now our objective to a mathematics problem: find the collection of probabilities P_i's which maximize expression (32) subject to the constraints (33).

To solve it we need one more mathematical tool: *the method of Lagrange multipliers*.

Lagrange multipliers come up over and over in statistical mechanics, and more generally in any problem where we need to maximize a function subject to constraints on the variables.

The remainder of this part of our discussion is devoted to the general method of Lagrange multipliers.

Then in the next part we will use it in our problem of maximizing the entropy subject to the constraints we described.

Many thermodynamic quantities are Lagrange multipliers.

We will discover, in particular, that *temperature is nothing but a Lagrange multiplier*.

3.8 Maximization under constraints, method of Lagrange multipliers

Consider a function F of several independent variables, $x_1, x_2, \ldots, x_p$.

And F is nicely behaved, meaning that it is smooth – mathematicians would say continuously differentiable – with respect to the independent variables.

Suppose we want to maximize F, that is to find the set of x_i's at which it is maximum.

The x_i's will eventually be the P_i's of the preceding discussion.

But for the moment let's be very general.

If there are no further conditions in the problem, we know how to solve it.

We write the set of p equations

$$\frac{\partial F}{\partial x_1} = 0$$
$$\frac{\partial F}{\partial x_2} = 0 \qquad (34)$$
$$\dots$$
$$\frac{\partial F}{\partial x_p} = 0$$

We solve them simultaneously, and this gives us the solution for the p unknowns we were looking for.

But we want to add something to the problem. We want to add some constraints on what the x_i's can be.

If we think geometrically about the space in which the x_i's are, there will be some constraints on where they can be.

Let's draw a picture of the space of the x_i's and of F, figure 9.

For convenience, we are in two dimensions, $x_1 = x$ and $x_2 = y$.

F would be along a third dimension(a surface) which, instead of being represented in perspective, is described by its contour lines as shown.

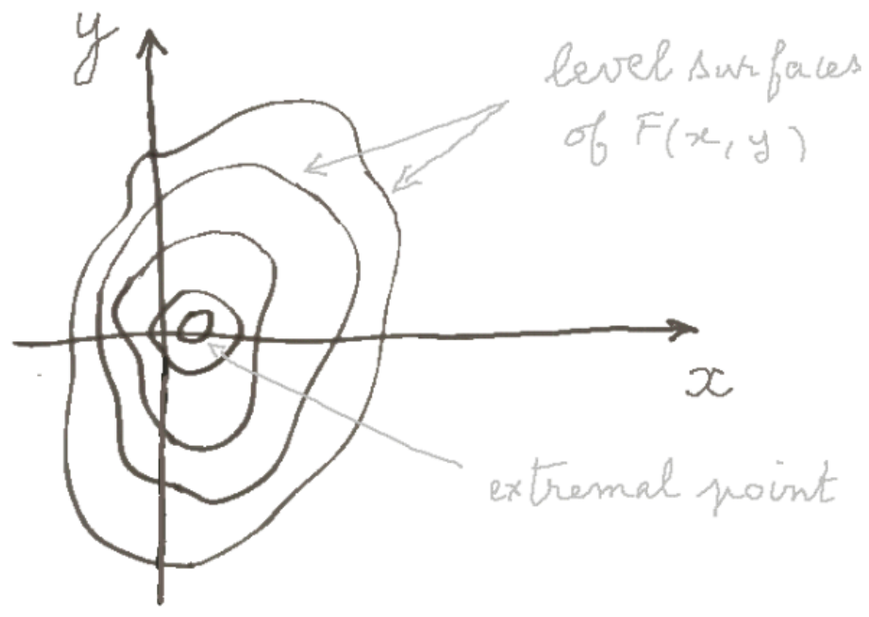
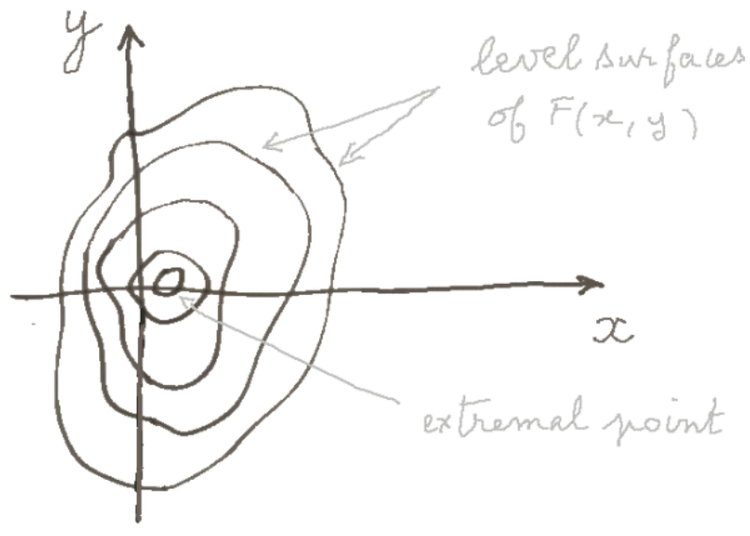


Figure 9: Contour lines of the function F (x, y).

Everybody is familiar with contour map representations.

They are used for instance in ordnance survey maps to represent the hills and valleys of a landscape.

A line in figure 9 corresponds to all the places where F has a given value.



When the contour lines become smaller and smaller the final point is an extremal point, either a peak or a hole.

When they are close to each other that corresponds to a steep slope in the terrain, etc.

Contour lines can be generalized to more than two dimensions.

They are then called level surfaces of the function F.

If there were three independent variables x_1, x_2 and x_3, the level surfaces would indeed be ordinary surfaces - beyond 3 dimensions I would not know the correct words to use.

In two dimensions, if somebody shows you the contour map of a function and you want to find the extremal points(no constraints), positive or negative, just look for the places that are surrounded by somehow concentric closed loops zeroing to points.

Now let's add one constraint. In the preceding discussion of maximizing entropy, we had two constraints.

But for the general explanations of the method of Lagrange multipliers, let's start with one.

Then we will add more.

We are now going to look for the place where $F(x_i)$ is maximum – x_i here stands for **all** the $x_1, x_2, \ldots, x_p$ – but given the constraint that some other function $G(x_i)$ must be equal to zero also.

What if the actual constraint is that G must be equal to 3?

Then subtract off the 3 from the original constraint function G, and call the new function G again.

Whatever the constraint is, you can always say that it is that some function must be equal to zero.

For instance, the constraint $G(x, y) = 0$ might correspond to some curve in the contour map of F, figure 10.

Along the curve $G = 0$, we are going to look at the various values F takes.

And we want to find the point where F is maximum **on** the curve $G = 0$.

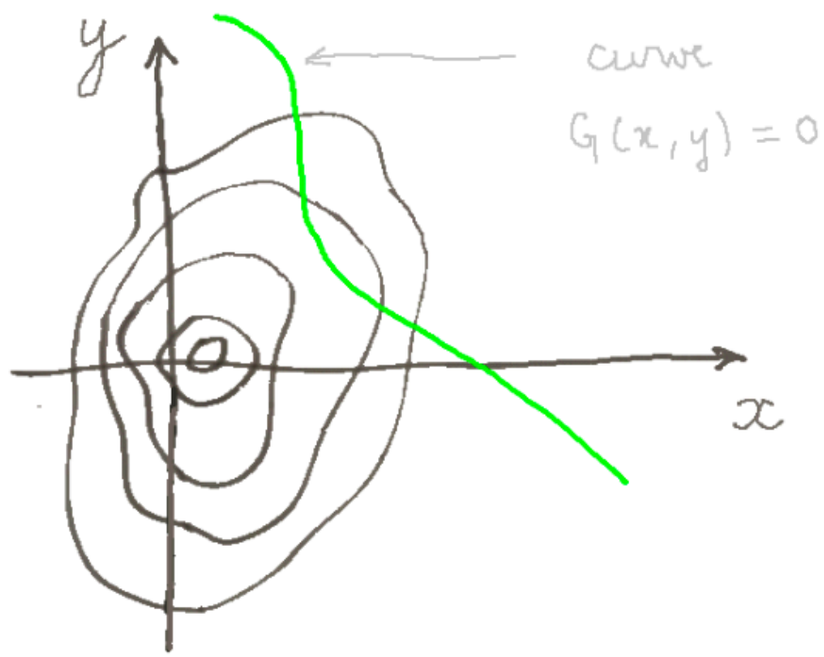
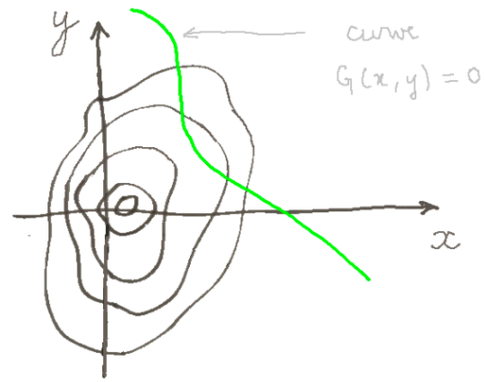


Figure 10: Contour map of F (x, y) and constraint G(x, y) = 0.

The first thing to note is that the point P on the curve $G = 0$, where F is maximum under the constraint, turns out to be a point where the corresponding contour line of F is tangent to the curve $G = 0$, which is itself a contour line of G.



Or equivalently, at P the gradient of F and the gradient of G are *colinear*.

This can be proved using only geometry.

The technique is to observe that **locally** every line is straight, every surface is a plane, and all contour lines are evenly spaced.

And that is true for F as well as for G, figure 11.

The requirement that at P the contour line of G must be parallel the contour line of F will be used in a moment.

So we want to find a method for solving the constrained maximization.

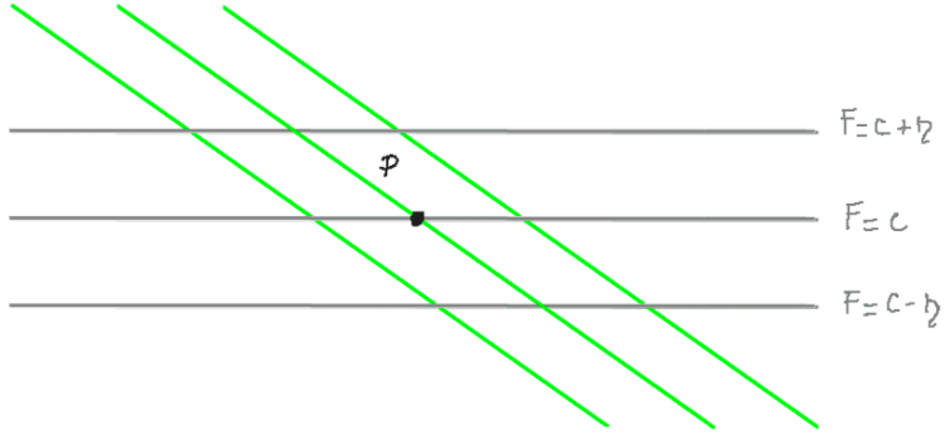


Figure 11: Local analysis, around P, of what would happen if the contour lines of G were not parallel to the contour lines of F , shows that in that case P would not be a maximum of F under the constraint G = 0.

At the point P, the solution to the problem, let's draw a line L perpendicular to the contour line of F, and therefore also to the curve $G = 0$, figure 12.

The point P is the solution to the maximization problem under the constraint of being on the curve $G = 0$.

But it is not necessarily the global maximum of F.

It could be, but it would be a fluke.

If we move along the line L away from P, we will cross other curves corresponding to G equal to different values, $G = -1$, $G = 0$ (at P), $G = +1$, etc.

Let's draw the value of F as function of G as we move along the line L, figure 13.

When $G = 0$ we are at the point P solution to the maximization of F under constraint, but the slope of F with respect to G at that point doesn't have to be zero.

There is no reason why should it be?

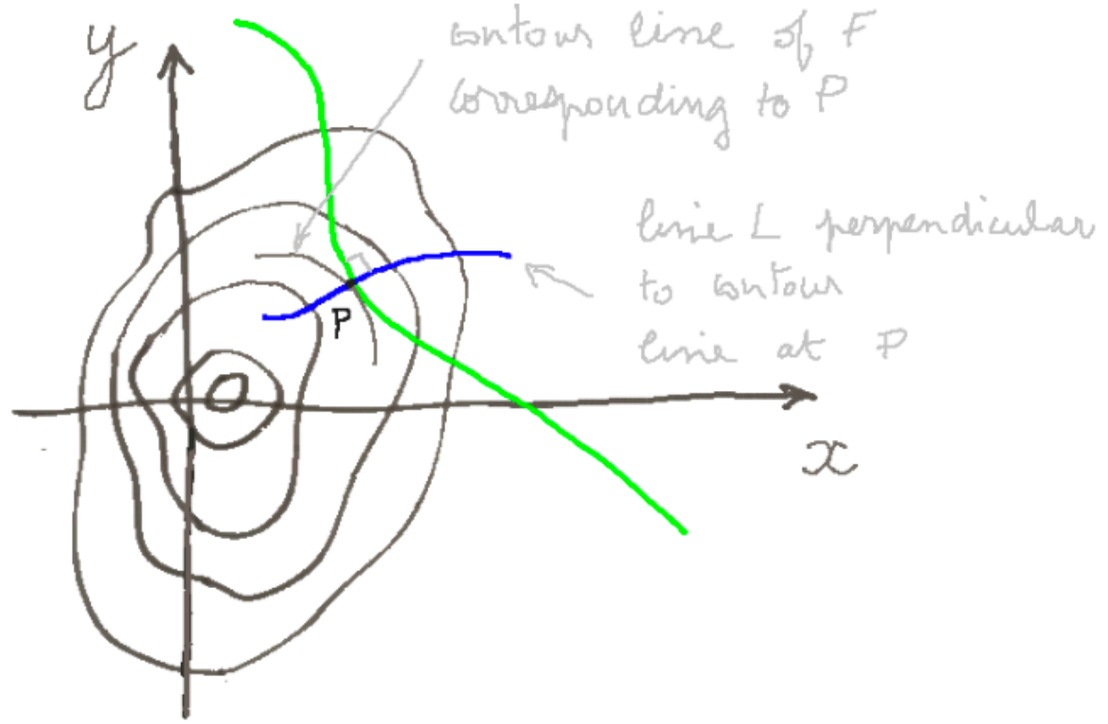


Figure 12: Point P solution to the maximization problem under constraint G = 0, and line L perpendicular to the contour line of function F at P.

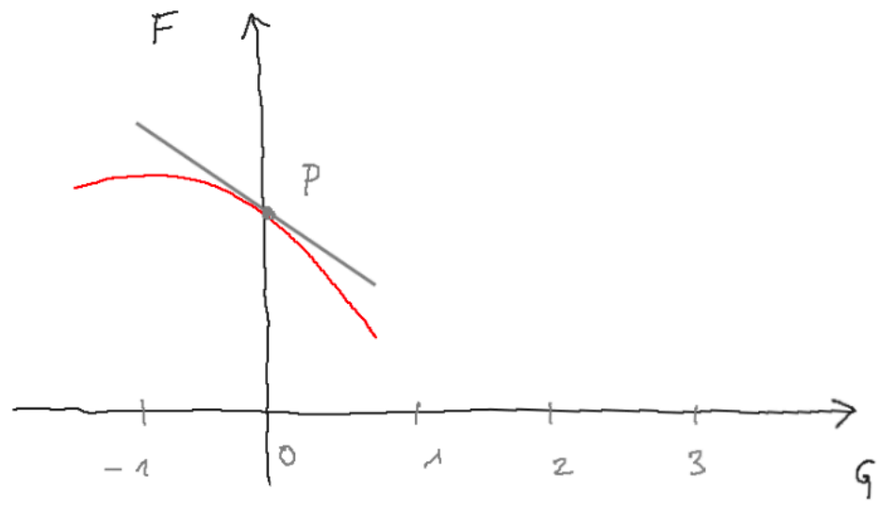


Figure 13: Value of F as a function of G when we move on line L perpendicular to the contour line of F and of G at P.

We were not looking for the place where F was globally stationary.

We were looking for the point where F was stationary subject to the constraint of moving along the line $G = 0$.

What can we do so that on the curve of F as a function of G, along line L, the point P will be a stationary point, in other words with flat slope?

One solution is to add to F something proportional to G. Thus let's define the function F' as

$$F' = F + \lambda G \tag{35}$$

F' is a function of all the x_i's, as are F and G.

And it is also function of some number λ.

The number λ can be chosen to make F' along the line L flat at P – with respect to G or to ordinary distance, it turns out that it doesn't matter.

And in the perpendicular direction, that is along the contour line of F, which is locally the contour line of G as well, F' is also flat at P, since $G = 0$ and F is maximum there on that curve.

So P must be a **global** stationary point of F'.

λ is a new unknown the value of which we will determine later.

But it makes the function F' have a global stationary point, without any constraint, at the same point P we were looking for.

So now it is going to be easy to find P, i.e., we have a definite process.

This procedure is due to Lagrange .

It is called the method of **Lagrange multipliers**.

And the coefficient λ is called the Lagrange multiplier.

It is a devilishly clever trick.

Before we see some examples, let's clarify one point.

We have added a new unknown λ to the problem.

Don't we have now more unknowns than equations?

No we don't, because we also have the new equation $G = 0$.

That is the whole point.

We have replaced a problem with p equations, p unknowns, and one constraint, by a problem with $p + 1$ equations, $p + 1$ unknowns and no constraint.

The proof that the method of Lagrange multipliers solves the problem of maximizing F under the constraint $G = 0$ is very intricate and we will not attempt it in this class.

Its use, however, is very straightforward, as we will see.

Let's use a couple examples of how to maximize or minimize a function under constraint to illustrate that the method work and that will be enough for us at this level.

Of course, remember that to maximize or minimize just depends on the sign.

It is not an important point.

It is the same technique for both.

Let us now look at an example.

Let's take $$F(x, y) = \frac{x^2 + y^2}{2} \tag{36}$$

I always like to put a factor of 2 downstairs, because we are going to differentiate F.

We want to minimize F.

If there is no further constraint, the answer is easy.

It is 0, and that minimum value is taken at $x = 0$ and $y = 0$.

But let's say we want to minimize F **given** $x + y = 1$.

What is G in this case?

$$G(x, y) = x + y - 1 \tag{37}$$

We are looking for the place, along the straight line $x+y = 1$, where $(x^2 + y^2)/2$ is minimum, figure 14.

Everybody can see by inspection where it is.

It is at the point, on that line, which is the closest to the origin.

It is of course x = 1/2 and y = 1/2.

But let's find it using the method of Lagrange multipliers.

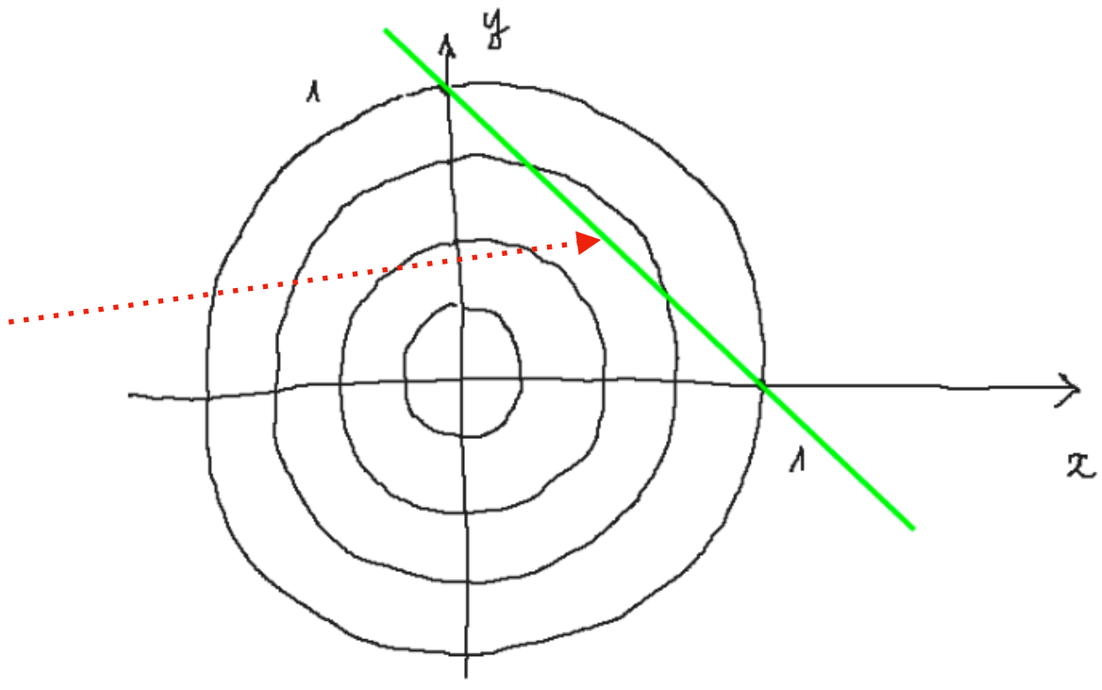


Figure 14: Contours of F , and constraint G = 0.

The rule is: add to F the quantity λ times G.

That defines F'.

Remember that F' here doesn't denote a derivative of F, but the new function we want to minimize.

$$\begin{aligned} F'(x, y, \lambda) &= F(x, y) + \lambda\, G(x, y) \\ &= \frac{x^2 + y^2}{2} + \lambda\,(x + y - 1) \end{aligned} \tag{38}$$

Next step, thinking temporarily of λ as a known coefficient, minimize F'.

That is the rule.

So that means instead of writing $\partial F/\partial x = 0$ and $\partial F/\partial y = 0$, we write

$$\frac{\partial F'}{\partial x} = 0 \tag{39}$$

$$\frac{\partial F'}{\partial y} = 0$$

This gives

$$x + \lambda = 0 \tag{40}$$

$$y + \lambda = 0$$

Hence, $x = -\lambda$ and $y = -\lambda$.

But, remember, we also have the extra equation $G = 0$.

By the way, it is nothing more, in the procedure of minimizing $F'(x,y,\lambda)$ without contraint, than the third equation $\partial F'/\partial\lambda = 0$.

It yields

$$-\lambda - \lambda - 1 = 0 \tag{41}$$

or $\lambda = -1/2$. Therefore, as expected,

$$x = \frac{1}{2} \tag{42}$$

$$y = \frac{1}{2}$$

That illustrates the **trick** of Lagrange multipliers.

What happens if we have more variables and more constraints?

Suppose we want to maximize, or minimize, a function F of three variables x_1, x_2 and x_3, and there are two constraints

$$G_1(x_1, x_2, x_3) = 0$$
$$G_2(x_1, x_2, x_3) = 0 \tag{43}$$

The way to proceed now is to define F' as F to which we add two terms, one for each constraint.

$$F' = F + \lambda_1 G_1 + \lambda_2 G_2 \tag{44}$$

So as usual F' depends on the three initial variables, x_1, x_2 and x_3, but also on the two new variables λ_1 and λ_2.

Then we minimize F' pretending that we knew what the lambdas were.

So for each variable x_i we write

$$\frac{\partial F'}{\partial x_i} \tag{45}$$

But we still have the two unknowns λ_1 and λ_2.

Indeed the solution to the set of three equations (45) will be a triplet of variables which depend on λ_1 and λ_2.

The two extra equations are again simply $G_1 = 0$ and $G_2 = 0$, which are nothing more than the partial derivatives of F' with respect to λ_1 and λ_2 set equal to zero.

Plugging the solution to (45) into the two extra equations gives a set of two equations for the two unknowns, λ_1 and λ_2.

This enable us to find λ_1 and λ_2, and in turn x_1, x_2 and x_3 without anything unknown anymore.

As I said, this seems like a rather complicated things to do.

But it is by far the easiest way to minimize something or maximize it when we have constraints.

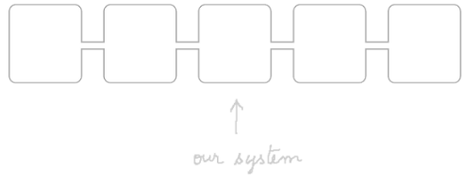
Why are we interested in the method of Lagrange multipliers in this class?

Because in many problems in statistical mechanics we want to maximize the entropy subject to some constraints.

The constraints might be that the total electric charge has some value, or the total energy has some value, or the total of something else has some value.

Those are the kinds of constraints that we impose.

In particular, in the problem of figuring out the probability distribution of the states of the N small systems making up the big system of figure 5, we want to maximize the entropy subject to the two constraints that the total energy is fixed, and of course that the total probabilities add to one.



Those are two constraints.

So let's now write our mathematics problem in this formalism.

The problem is

$$\text{maximize } F(P_1,\ P_2,\ ....) \tag{46}$$

under the constraints

$$\begin{aligned} G_1(P_1,\ P_2,\ ....) &= \sum_i P_i - 1 = 0 \\ G_2(P_1,\ P_2,\ ....) &= \sum_i P_i E_i - E = 0 \end{aligned} \tag{47}$$

That will give us the probability distribution as a function of i, that corresponds to the most likely possible values for the occupation numbers.

The occupation numbers will tend to cluster around these values.

A note on statistics:

What is the technical difference between a **probable** value and a **likely** value, or more generally between **probability** and **likelihood**?

We saw that we talk about probabilities when there is a well defined random experiment ε producing "states of the world" ω's belonging to a big set Ω, and there exist of measure of probability P on the subsets of Ω.

A random variable X is a function from Ω onto some set of things, numerical or not.

X enables us to define *subsets* of Ω, as for instance $\{X < 2\}$ when the set in which X takes its values is numerical.

And we can then talk about, calculate, or measure through a large number of replications of ε, the probability $P\{X < 2\}$.

Sometimes the problem we are confronted with is this: ε and Ω are well defined and known, but we don't know P, and we don't have the possibility of reproducing ε a large number of times.

However we know that P belongs to a known family of probability distributions indexed by a parameter θ.

And we have at our disposal some experimental data about a random variable X.

For instance, ε was reproduced n times, and we obtained the measures $x_1, x_2, \ldots x_n$ for X.

The problem is: What is θ?

Of course, we don't have enough information to know θ exactly.

But the experimental data we have do give us some information.

We can *estimate* θ.

The value of the parameter θ which gives the maximum probability to the data we observed is called the **maximum likelihood** of θ.

We don't talk of a "probable value" of θ because θ is not a random variable.

But we talk about a "likely value" of θ.

The maximum likelihood of θ is a well known and well respected estimator.

Let's denote it $\hat{\theta}$.

It is a random variable produced by the experiment $\mathcal{F}$ which consists in n replications of ε .

Indeed $\mathcal{F}$ produces $x_1, x_2, \ldots x_n$ and therefore $\hat{\theta}$.

If we reproduce $\mathcal{F}$, we will get a **new** set of data, and a **new** experimental value of $\hat{\theta}$.

The maximum likelihood estimator $\hat{\theta}$ is very natural estimator of θ.

If we know that the following measures $2, 5, -3, 2$ and 1 came from a Gaussian distribution of unknown mean θ and unknown width σ, we won't guess that θ was 50.

We will estimate it to be the mean of the five data values, namely 1.4, which happens to be the maximum likelihood of θ.

Maximum likelihood estimators have plenty of nice properties, and a few weird ones

That is why when we speak of the estimated value of some parameters, be they probabilities like P_1, P_2, ... P_n in our problem of entropy, we talk not about their probable values, but about their likely values.

Back to maximizing the entropy: again, it is not easy to solve.

But once we have set up the problem, the quantity to maximize, and the constraints, the rest is just pretty straightforward in principle.

It is then just probability theory and statistics.

It is the same set of rules which happens to govern many different problems.

Once the problem is set, we are not doing the theory of gases; we are not doing the theory of superconductors; we are doing basic probabilistic theory of maximizing some entropy, the number of ways of rearranging things, subject to some constraints.

It is not special to any particular kind of physical system.

It is only special to the kinds of physical systems which have so many degrees of freedom that you are forced to deal with them statistically and probabilistically.

But once you know that, the rules are always the same.

In the next part, we will solve the problem.

F will be the entropy. As usual it will be called *S*.

The problem will be to maximize

$$S(P_i) = -\sum_i P_i \log P_i \tag{48}$$

subject to the two constraints you should by now be familiar with.

The variables to shift around to maximize S are the P_i's.

In other words, we have a probability distribution and we are going to vary the probability distribution until we find that probability distribution P_i's which maximizes the entropy, given that the sum of all the probabilities is 1, and the average energy is E.

That is statistical mechanics in a nutshell: Solving the above problem for different systems.



4. The Boltzmann distribution

4.1 Introduction

We are now getting to the real heart of statistical mechanics : the Boltzmann distribution.

Then we will work out the example of the ideal gas.

But first of all, let's quickly review what we did in the last part

We looked at a big system made of a large number of replicated identical small systems all connected, figure 1.

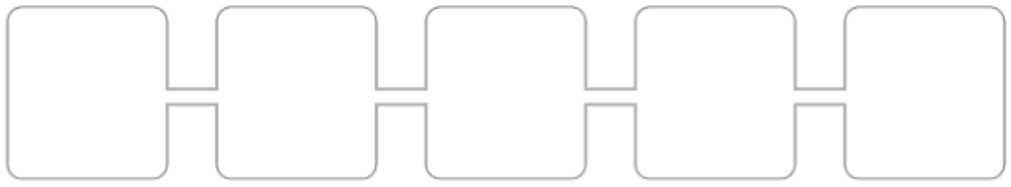


Figure 1 : Large number of replicated identical small systems.

Figure 1 is only suggestive.

We should think of each subsystem as an imaginary small volume in a big volume, each small system being connected to many others.

We are interested in one small system.

Any one will do.

The total energy of the big system is a fixed quantity which we denote NE, where N is the number of small systems, and therefore E is the average energy of a small system.

When the big system is in thermal equilibrium, there are still plenty of microexchanges of energy between the subsystems.

Thus, at any time, each small system is in a state i out of a collection of states Ω.

The number of possible states should not be mixed up with the number of small systems.

Ω can be finite, countably infinite, or, later on, a continuous phase space.

We denoted n_{i} the number of subsystems in the i-th state.

The n_i's are called the occupation numbers.

They satisfy . $\sum_i n_i = N$

The probability for a small system to be in the *i*-th state is

$$P_i = \frac{n_i}{N} \tag{1}$$

That gives us a probability distribution over the states of the system.

From now on, unless otherwise specified, when we talk about the system, we mean the subsystem we have chosen to follow.

From that probability distribution we can compute an entropy.

It is defined as

$$S = -\sum_i P_i \log P_i \tag{2}$$

Going back to the occupation numbers n_i's, we calculated the number of ways to arrange *N* objects into a given collection of occupation numbers.

For instance if we have four objects and three cups, and we want to arrange our four objects as two in the first cup, one in the second cup and one in the third cup, there are twelve ways to do it.

The general formula is

$$A = \frac{N!}{n_1!\, n_2!\, n_3!\, \ldots} \tag{3}$$

Then we turned to the problem of finding the collection ($n_1, n_2, n_3, \ldots$) which maximizes A under the constraint that $\sum_i n_i E_i = NE$, where E_i is the energy of the i-th state.

The E_i's are givens of the problem.

We used Stirling formula to approximate the factorials.

We discovered that to find what set of occupation numbers maximized A under the constraint – what we call the most likely occupation numbers – turned out to be same as to maximize the entropy S under the constraint $\sum_i P_i E_i = E$.

Let's take a closer look at the entropy S in equation (2).

$$S = -\sum_i P_i \log P_i$$

What is it a function of?

It is a function of the P_i's.

They are the independent variables, and S is the dependent variable.

Fortunately S is a rather simple function.

We ended last part of our discussion with the following mathematical problem :

Maximize the function $S = -\sum_i P_i \log P_i$ under the two constraints

$$\sum_i P_i = 1 \tag{4}$$

$$\sum_i P_i \, E_i = E$$

The E_i's are fixed.

To each possible state i there is attached an energy level E_i.

And E is fixed too.

It is determined once and for all by saying that there is a total amount of energy in the set of small systems, figure 1.

It is proportional to the number of subsystems.

E is that total energy divided by N.

Just out of habit, I prefer to talk about minimizing something rather than maximize it.

So we shall look at the problem of minimizing $-S = \sum_i \ P_i \log P_i$, subject to the same two constraints.

What is the technique for finding the minimum of a function given some constraints?

There are plenty of them depending on the problems.

But there is a celebrated general one which we presented at the end of last part.

It it the method of Lagrange multipliers.

It is a highly formalized mathematical construction, but it is extremely powerful.

In all sorts of contexts where you want to find the minimum of something, and you know some other things are constant, that is the way to do it.

In particular anybody who does statistical mechanics will use the method of Lagrange multipliers over and over.

We consider the function $F(x_i)$ that we want to minimize, and the contraints $G_j(x_i) = 0$, where j runs over the constraints.

The method consists in constructing a new function F', the minimization of which *without constraints* will lead to the solution to the initial problem with constraints.

That is the beauty of the method.

Here the independent variables x_i's are the P_i's.

$$F(P_i) = \sum_i P_i \log P_i \tag{5}$$

And there are two constraints

$$\begin{aligned} G_1(P_i) &= \sum_i P_i - 1 = 0 \\ G_1(P_i) &= \sum_i P_i E_i - E = 0 \end{aligned} \tag{6}$$

Remember that F' is defined as

$$F' = F + \sum_j \lambda_j G_j \tag{7}$$

It is a function of the initial variables P_i's plus the additional variables λ_j, one for each constraint.

These are called the Lagrange multipliers and the Greek letter λ is generally used for their notation.

In statistical mechanics, in our specific problem of maximizing the entropy S, or equivalently minimizing $-S$, there are two constraints.

The corresponding Lagrange multipliers have standard notations : α for the first constraint that the sum of the probabilities must add up to one, and β for the second constraint that the average energy of the system must be E.

The Lagrange multipliers α and β will play different roles.

And they will have important interpretations.

4.2 Lagrange multipliers α and β

Let's continue working on the Lagrange multipliers. In our problem, equation (7) is rewritten

$$F' = F + \alpha \left[\sum_i P_i - 1 \right] + \beta \left[\sum_i P_i E_i - E \right] \tag{8}$$

F' is a function of the P_i's and α and β.

We want to minimize F'. So we shall write that the partial derivatives of F' with respect to the P_i's are equal to zero.

That set of equations will lead to a solution for the P_i's.

They will still depend on α and β.

Then the constraints will provide two additional equations.

And that will enable us to determine α and β as well.

Let's now go into the calculations.

We start with the partial derivatives with respect to the P_i's.

So the -1 in the first constraint and the $-E$ *i*n the second constraint, in equation (8), won't play a role.

We can omit them, and tackle the problem of minimizing

$$\sum_i P_i \log P_i + \alpha \sum_i P_i + \beta \sum_i P_i E_i \tag{9}$$

We are very fortunate because this expression is a sum of terms, each of which contains only one P_i.

So the partial differentiations won't create a mess.

Suppose we want to differentiate expression (9) with respect to some particular P_i.

Just to focus attention on what we are doing let's pick P_7.

And we shall set that partial derivative equal to zero.

The partial derivative of the whole expression (9) with respect to P_7 simplifies into

$$\frac{\partial}{\partial P_7} \left(P_7 \log P_7 + \alpha P_7 + \beta P_7 E_7 \right) \tag{10}$$

Noting that $\frac{\partial\ x \log x}{\partial x} = \log x + 1$, this is simply

$$\log P_7 + 1 + \alpha + \beta E_7 \tag{11}$$

We have the same thing for P_8 and P_4 and P_9 ... and P_{152}.

And we want each of them to be equal to zero.

So going back to any i, we have the set of equations

$$\log P_i + 1 + \alpha + \beta E_i = 0 \tag{12}$$

where i runs over all the possible states for the small systems in figure 1.

Equation (12) is kind of amazing.

Statistical mechanics is full of things like this.

It is sort of magic.

Remarkable identities pop up all of a sudden in the course of calculations.

We were turning the crank, in autopilot, and suddenly we discover something important and unexpected.

What is unexpected?

It is that equation (12) will give us simply what P_i is.

Once we now α and β, we will know P_i

$$\log P_i = -(1+\alpha) - \beta E_i \tag{13}$$

or

$$P_i = e^{-(1+\alpha)}\ e^{-\beta E_i} \tag{14}$$

The first exponential is just a number.

It has a standard notation in statistical mechanics.

The exponential without the minus sign is denoted

$$e^{(1+\alpha)} = Z \tag{15}$$

The first Lagrange multiplier α, in the form Z given by this expression, has a name, which we will give in a moment, and will play a very important role.

We will come back to it.

Equation (14) becomes

$$P_i = \frac{1}{Z}\, e^{-\beta E_i} \tag{16}$$

The term $1/Z$ doesn't distinguish the different states.

It is just a number related to α.

Before we talk more about Z, let's observe that the relative probabilities of the different states are very simple.

P_i is just proportional to $e^{-\beta E_i}$.

β must have some physical significance.

It is the only variable in equation (15) that has any chance of telling us what the average energy is.

Let's draw the probability distribution of the P_i, rescaled with Z, for several values of β, figure 2.

The higher β is, the more narrow and peaked the probability distribution is.

Conversely the smaller β is, the more spread out the distribution is.

What are the different probability distributions going to be parameterized by?

Well they are parameterized by β.

But they also can be parameterized by the average energy.



Figure 2 : Two distributions of probability, rescaled with Z, for two different values of β.

It is quite clear that the more spread out the distribution is, the higher the average energy.

So β has something to do with energy.

It is the thing that we tune to change the average energy.

The smaller β, the higher the average energy.

We are going to be more specific about that.

So far it is just β, and its significance is a certain Lagrange multiplier.

What is its real significance ? We will see that *β is the inverse temperature*.

How shall prove that ?

We already have a definition of temperature.

We will show that β is indeed the inverse of the temperature we have already defined.

But for the moment let's continue to think in terms of β itself.

Equation (16) gives our probabilities P_i's in terms of two parameters.

One is Z and the other is β.

Z is a simple function of the first Lagrange multiplier α.

And β is the second Lagrange multiplier.

Now we have to come back to the problem of maximizing the entropy, or equivalently minimizing $S = -\sum_i P_i \log P_i$.

How do we fix the Lagrange multipliers ?

We fix them by making sure the constraints are satisfied.

The first constraint is that the sum of all the probabilities be equal to one.

equation (16), we write it

$$\frac{1}{Z}\sum_i e^{-\beta E_i} = 1 \tag{17}$$

Incidentally where do the E_i's come from?

Some laws of physics that have to do with whatever is inside the box tell us what the possible energy levels are.

So the E_i's, from our point of view, are simply numbers that are known.

Somebody has been smart enough to compute them and tell us what they are.

Equation (16) tells us what the probability to be in *i*-th level is as a function of β.

We saw that β is a thing that has to do somehow with the average energy, because as we vary β we shift the probability distribution in figure 2.

We saw that the bigger β is, the more peaked the function is and the more rapidly it falls off to zero.

If β is huge, then the probability distribution is confined to the lowest energy levels.

And the average energy is also very low.

Then, make the energy even a little bit bigger and we lose a whole bunch of probability over the lowest energy levels.

In other words our knowledge of the state of the system becomes fuzzier.

At the other end of the range of values of β, if β is near 0, then the probability distribution is practically flat.

And it is going to contain important contributions from large energy states.

So by tuning β, we tune the average energy.

And they vary in opposite directions.

We will come back to that and be more quantitative of course.

Equation (17) can also be written

$$Z = \sum_i e^{-\beta E_i} \tag{18}$$

The Lagrange multipliers, as the P_i's, are ultimately functions of the E_i's.

But – remembering that $Z = e^{1+\alpha}$ – equation (18) shows that α and β are related.

That is true in general : in a maximization or minimization problem the Lagrange multipliers are ultimately numbers, but they depend on each other in the sense that they can be linked by equations like equation (18).

Therefore we shall write equation (18) as

$$Z(\beta) = \sum_i e^{-\beta E_i} \tag{19}$$

From now on, we may forget where it comes from.

$Z(\beta)$ is by definition the sum over all of the states of the system of $e^{-\beta E_i}$.

It is a fundamental function.

Where does $Z(\beta)$ fit into the probability distribution?

It is a kind of normalization constant which guarantees that the total probabilities add up to one.

We emphasize the function $Z(\beta)$ in its own right because almost everything in statistical mechanics is buried in this function of β.

We are going to find that it is an extremely powerful thing.

Z(β) is called the partition function.

It is a function of β.

And β is the inverse temperature, but we don't know that yet.

4.3 Beginning to use the partition function Z(β)

Now we turn to the second constraint.

$$\sum_i P_i \, E_i - E \; = \; 0 \tag{20}$$

Using equation (16), we can rewrite it

$$\sum_i \frac{1}{Z} e^{-\beta E_i} E_i = E \tag{21}$$

Suppose we already know Z ; suppose we calculated it.

Then the rest of this equation determines β.

It becomes an equation for β in terms of the average energy.

If we know the average energy E and we are lucky, equation (21) might have a simple enough form that we can find out what β is in terms of the energy.

Or we can also read it the other way.

We can also say that, whatever β is, it determines the average energy.

Both are legitimate ways to look at it.

For the moment let's just stare at equation (21).

We are going to rewrite it.

There is a very famous trick, which goes as follows.

Let's begin with

$$Z(\beta) = \sum_i e^{-\beta E_i} \tag{22}$$

That is of course the partition function.

And let's differentiate both sides with respect to β.

$$\frac{\partial Z}{\partial \beta} = \sum_i -E_i \; e^{-\beta E_i} \tag{23}$$

We get something which is a piece of equation (21).

Let's divide by Z on both sides.

$$\frac{1}{Z} \frac{\partial Z}{\partial \beta} = -\sum_i \frac{1}{Z} E_i \; e^{-\beta E_i} \tag{24}$$

Now what is the sum on the right-hand side?

It is exactly the term that we identified with the average energy E, equation (21). $\sum_i \frac{1}{Z} e^{-\beta E_i} E_i = E$

So we arrive at

$$E = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} \tag{25}$$

This is a remarkable formula for the energy in terms of the derivative of the partition function with respect to β.

What this is suggesting to us is that the thing we want to compute is the partition function.

As I said, we are going to find that the partition function has everything in it.

In particular it has the average energy, as given by equation (25).

So if we know the partition function – which we will calculate in various cases –, we differentiate it with respect to β.

And that tells us the average energy.

E is also of course a function of β.

Everything on the right- hand side of equation (25) is a function of β.

The right-hand side of equation (25), being of the form f'/f, is the derivative of a logarithm.

So let's rewrite it

$$E(\beta) = -\frac{\partial \log Z}{\partial \beta} \tag{26}$$

So, it is really the logarithm of the partition function which is going to occupy us a lot.

Since β is the inverse of temperature – as we will show –, equation (26) will relate the average energy to the temperature via the partition function.

It is the first of many sort of magical formulas we shall encounter in statistical mechanics.

Statistical mechanics is at the same time extremely easy and extremely hard.

It is subtle.

It is full of surprises.

It is the application of very simple formulas which then yield extremely surprising and powerful results.

What makes it hard is that half the time you can't guess or you can't look at the formulas and say :

"Oh of course this is obvious. E is equal to the derivative of the logarithm of the partition function."

The whole proof here was complicated.

You couldn't have guessed that.

But it sure is a simple formula at the end.

And it is a powerful formula.

4.4 Questions/answers

Question: The E that we computed, is that the energy of the whole system, or the energy per box ?

Answer: It is the energy per box.

More precisely E is the average energy of any one box.

A large number N of similar and connected boxes form the whole system, figure 1.

To study one box in thermal equilibrium, we used a trick.

We thought of it as plunged into a heat bath formed by a large number of systems identical to ours.

The whole big system is also in thermal equilibrium.

Yet the small systems have microexchanges of energy.

Each subsystem, or box, is so to speak alive.

It goes through micro-evolutions over time.

But they all behave similarly.

Each box is in a random state belonging to a set Ω.

There is a probability distribution P_{i} over the states of the boxes.

You can think of the random experiment leading to the probabilities in two different ways.

Either, at a time fixed once and for all, look at the whole system and pick one of the subsystem at random, or keep to your own subsystem and look at it at a random time.

The two ways are equivalent.

The trick enables us to write a constraint on the probabilities.

Of course there is first of all the usual constraint of probabilities, namely $\sum_i P_i = 1$.

But also we consider that the whole system has a fixed energy .

Equivalently *each box* has a fixed average energy E.

That may seem like a paradox.

One box doesn't have a fixed energy level, but the whole system has a fixed quantity of energy.

The fact that the big system is made of a large number of boxes, and behaves like a heat bath has the consequence that we can study one box which has microexchanges.

In other words, thermal equilibrium means that the system concerned, here one box, goes through a collection of energy levels.

Each state in Ω has an energy E_i, which comes from physical considerations about the nature of what is going on in our initial box.

Don’t mix up the states in Ω and the boxes in the whole system.

The E_i’s are the givens in the problem.

Then we have the second constraint : $\sum_i P_i E_i = E$.

Finally we turn to entropy to figure out the P_i’s.

Don’t forget that the entropy of one box, $S = -\sum_i P_i \log P_i$, is a measure of our incomplete knowledge about the exact state, i.e. positions and momenta, of all the molecules of that box.

We apply a principle : in thermal equilibrium the entropy of the box is maximum.

You can also view it as a consequence of other axioms of thermodynamics :

There exists a concept of thermal equilibrium;

It has a temperature;

when the temperatures of two connected systems are not the same, energy goes from the hotter to the colder, till they reach equilibrium, etc.

So we have transformed a physical problem – what is the probability distribution of the states of our box? – into a mathematical one – maximize entropy under two constraints.

That leads us to the Lagrange multipliers, the partition function, the Boltzmann distribution, etc.

Question: From equation (22), can't we simply write that the logarithm of Z is the sum of the βE_i's with a minus sign?

$$Z(\beta) = \sum_i e^{-\beta E_i}$$

Answer: No, no. Z is a sum of exponentials.

The logarithm of the sum is not the sum of the terms in the exponents.

If $x = e^a + e^b$, it is not true that $\log x = a+b$ in general.

You cannot do that.

It would be true if we had a product. If $x = e^a e^b$, then $\log x = a+b$.

But in equation (22), we have a sum, not a product.

Question: All the E_i's and E are known.

To figure out the P_i's maximizing the entropy subject to two constraints, we introduce the Lagrange multipliers α and β, etc.

But once we have solved our mathematical problem nothing is left unknown.

The P_i's are now numbers, and so are α and β.

So how come in your formulation α and β are still variables ?

Answer: Sure, in the end, the P_i and α and β are just numbers.

That is correct.

And it looks like these auxiliary α and β can be forgotten again.

But it turns out that α and β have some interesting physical significance.

Remember there are other things interesting in the problem besides just the P_i's.

For example, there is the average energy.

Remember that E is a given in the problem.

But then it is also $\sum_i P_i E_i$.

Now what we are finding is that there is a relationship between average energy and β, equation (26).

You can either fix the average energy and then determine what β is.

That would be the standard way of looking at Lagrange multipliers.

But you can read equation (26) the other way.

You can say the average energy is determined in terms of β.

We are going to find, as I said, that β has a significance of its own.

It has a life of its own as the inverse temperature.

And that is the next thing we are going to do.

Question: We are focussing on β. Hasn't α an important role too?

Why did it disappear?

Answer: Well, α hasn't really disappeared.

We just replaced it by Z.

Remember

$$Z = e^{(1+\alpha)} \tag{27}$$

It is just a redefinition.

Of course α, β, E, the P_i's, etc. are all in the end functions of the E_i's and E.

Consequently there are relations between α, β, E, etc.

Z that we defined with equation (15) or (27) is also related to these other variables.

It turns out that the variable β is particularly fundamental, and we shall be interested in the functions $Z(\beta)$, $E(\beta)$, etc.

Let's look at the problem, for a minute, from a mathematical or logical point of view.

Suppose you have the three unknowns x, y and z satisfying the three equations

$$\begin{aligned} x - y + z &= 0 \\ y + z &= 2 \\ y - z &= 2 \end{aligned} \tag{28}$$

We can surely solve the problem for the three unknowns.

But that doesn't prevent the relationship $x = y$, that we can easily prove, from being interesting.

It means that irrespective of what the final value of y is in the problem, we have the possibly interesting dependence $x(y) = y$.

Similarly, in our physical problem, since the E_i's and E are all given parameters – given numbers if you like –, we can solve for the P_i's, α, β and even other intermediate variables like Z.

All these in the end are numbers.

But that doesn't prevent the expression of E as a function of β or of Z as a function of β to be interesting.

Question: It seems that the second constraint equation $\sum_i P_i E_i - E = 0$ presupposes a lot of stuff about the P_i's that may not be true.

Answer: This equation is the definition of the average energy.

In our problem, we assume that we know what the average energy is.

It is a given.

It may come from other physical considerations on the system, some specific knowledge, or whatever.

But anyway, in our problem it is a given.

And what we want to know first of all is the P_i''s.

Of course, the P_i's, even when theoretically solved for, remain somewhat abstract.

They form the Boltzmann distribution at thermal equilibrium.

But we don't intend to work on any specific value of, say, P_7.

On the other hand, the maximization of the entropy leads to very interesting things :

- partition function, which turns out to be a central tool ;
- relationship between E and β ;
- fact that β is the inverse of the temperature as defined in part 2 – now we shall prove it ;
- etc.

There may be other things going on in our physical problem, that transform into mathematical constraints.

For example, we may want to constrain the total momentum, or some other conserved quantity.

For the moment let's suppose that there aren't other conserved quantities.

Then there is nothing left to constrain.

Now you may come in and you may say that you are interested in changing the volume of your box that the gas is in, changing the shape of the box, changing all the other things that you could imagine varying.

For the moment we have supposed that all the parameters which define the system, such as the volume of the box, such as the shape of the box, all the other things that you can imagine varying are kept fixed.

Those are called *control parameters*.

Changing the volume of the box, changing the shape of the box, changing the magnetic field on the system, etc. that would be changing some control parameters.

And they are important.

But at the moment we are imagining they are absolutely fixed.

What do the control parameters determine incidentally?

The energy levels, the E_i's.

So when you change the control parameters, you in general change the energy levels.

And in so doing you change everything including the partition function.

Question: In equation (26), where we write $E(\beta) = -\partial \log Z/\partial\beta$, shouldn't we write $E(\alpha, \beta) = -\partial \log Z/\partial\beta$?

Answer: The independent variables in our problem are initially the E_i's and E.

We introduced α and β because we used the method of Lagrange multipliers to figure out the entropy.

Furthermore, α and β turn out to have quite interesting significance, i.e., α, when modified into Z, becomes the partition function.

And it turns out to be a central tool – the central tool, one might say – in statistical mechanics.

Then we found out that Z had an implicit dependence on β, and that the average energy E was equal to minus the partial derivative of log Z with respect to β.

That is why we write

$$E(\beta) = -\frac{\partial \log Z}{\partial \beta} \tag{29}$$

Including α in the variables driving E would only add confusion. Here is why :

The point is that α and β both depend on the initial independent variables E_i's and E.

But then, unless we are in a degenerate situation, we can interchange α, β and E.

In other words, leaving aside the E_i's which are once and for all given parameters, we can express any variable using any other.

We have

1. α depends on E
2. α depends on β, which we prefer to express as Z depends on β,
3. β depends on E
4. β depends on α
5. E depends on α
6. E depends on β

This last dependence, written even more explicitly than equation (29), specifically is

$$E(\beta) = -\frac{\partial \log Z(\beta)}{\partial \beta} \tag{30}$$

So we see that it would be meaningless to write E(α,β).

The new independent variable that we shall mostly use will be β.

And that makes physical sense because it is like using the temperature as the independent variable.

So our next topic is the relationship between the energy(average) at thermal equilibrium and temperature.

Question: We don't pay attention to the E_i's anymore?

They still drive α and β, don't they?

Answer: α and β depend on the E_i's and E.

And once you got the P_i's, you can view E as depending on the E_i's and the P_i's.

We are imagining the E_i's as a fixed set of numbers for the moment.

Now when we vary other things in the system, volume of the box, electric or magnetic field, or other things that we might vary, other control parameters, voltages, whatever we have, we may change the energy levels.

When we change the E_i's, we change Z.

But for the moment the energy levels are fixed.

Question: But since β depends ultimately on the E_i's, how can we change β if we don't change the E_i's?

Answer: The E_i's alone are not sufficient to determine β.

You need also E, the average energy.

With the same collection of possible energy levels, you can have different E, Z and β.

That is the whole point.

It depends on the probability distribution over the energy levels.

In the initial problem we introduce the constraint that the average of the energy levels must be E.

Then *Z(really α)* and β, the two Lagrange multipliers, are determined, because they depend on E.

β is a function of E.

Or E is a function of β.

Either one of them might be used as the parameter that you might vary to study the other.

Now in a box of gas, it is easier – much easier – to measure the temperature than it is to measure the energy in the box.

How do you measure the energy of a box of gas?

Well you can weigh it and use $E=mc^2$.

But that is a lousy way.

Or, let me say that differently(nicer way) : it is a great way !

Just not helpful.

It is a little hard to measure the energy.

It is much easier to measure the temperature.

You stick in a thermometer.

You read the temperature.

And if you know the connection between E and β, then you can determine the energy.

But there is one hang up, we haven't shown yet that β is connected with the temperature.

So let's do that now.

There is one preliminary step.

We must talk about entropy first.

Remember that entropy comes before temperature.

4.5 Entropy

Entropy logically comes before temperature.

So it behooves us to calculate the entropy.

Let's see if we can get an expression for the entropy in terms of the partition function.

Using equation (16) for P_i, here is more magic :

$$\begin{aligned} S &= -\sum_i P_i \log P_i \\ &= -\sum_i \frac{1}{Z} e^{-\beta E_i} \left[-\beta E_i - \log Z \right] \\ &= \sum_i \frac{1}{Z} e^{-\beta E_i} \left[\beta E_i + \log Z \right] \end{aligned} \tag{31}$$

Taking care of the two terms between brackets separately, this is the sum of two big sums.

The first one is nothing more than $\beta \sum_i P_i E_i$, that is β times the average energy.

So equation (31) becomes

$$S = \beta E + \frac{1}{Z} \log Z \sum_i e^{-\beta E_i} \tag{32}$$

Now, what is $\sum_i e^{-\beta E_i}$?

It is Z.

We can see that from the expression of P_i as $e^{-\beta Ei} / Z$ and the first constraint that all the probabilities add up to 1.

Thus, continuing on autopilot so to speak, equation (32) simplifies further into

$$S = \beta E + \log Z \tag{33}$$

We reached an interesting formula for the entropy.

We could write E in terms of a certain derivative of logarithm of Z.

But let's just leave it as the average energy.

So S is β times E plus the logarithm of Z.

The logarithm of the partition function is the quantity which comes in over and over in statistical mechanics.

It is related, by a factor of temperature, to something called the *Helmholtz free energy*.

But we don't need to know it right now.

Remember that the partition function Z depends on β.

If we can calculate the partition function, we can calculate the average energy E, and we can calculate the entropy S.

And in fact, if we can calculate the entropy and the energy, we can calculate the temperature.

4.6 Temperature

To find how to express temperature with our new partition function we have to remember the definition of temperature we gave in part 2.

The change in energy when we change the entropy by one unit is called the temperature.

$$dE = TdS \tag{34}$$

or

$$\frac{dE}{dS} = T \tag{35}$$

That is the way we defined temperature earlier.

When we used this definition of temperature, we found that heat always flows from hot to cold.

That is a pretty good indication that energy, in the form of heat, has something to do with temperature.

And it is the definition of the temperature of a system: the rate of change of energy with respect the entropy.

You have seen this stuff before, so you are familiar with the equivalent formula

$$dS = \frac{dE}{T} \tag{36}$$

That is, 1/T is the rate of change of entropy with respect to energy.

To put some flesh on this formula, let's print it in our minds as follows:

When the temperature is low, a small change in energy entails a big change in entropy.

At the other end of the temperature range, when the temperature is high, it requires a much bigger change in energy to create the same change in entropy.

In short, entropy is more sensitive to energy at low temperatures than at high temperatures.

Now that we expressed the entropy as a function of β, E and Z, we have the tools to calculate the derivative dS/dE.

We use equation (33) to differentiate S with respect to E.

Since β itself depends on E, from the product rule of differentiation, the first term βE will produce two terms.

And the second term will simply be dlog Z/dE.

So we have

$$\frac{dS}{dE} = \beta + E\frac{d\beta}{dE} + \frac{d\log Z}{dE} \tag{37}$$

The third term on the right-hand side can be rewritten

$$\frac{d\log Z}{dE} = \frac{\partial \log Z}{\partial \beta}\frac{d\beta}{dE}$$

So equation (37) becomes

$$\frac{dS}{dE} = \beta + \frac{d\beta}{dE}\left[\, E + \frac{\partial \log Z}{\partial \beta} \,\right] \tag{38}$$

Now all we have to do is to notice that the term in brackets is zero, because we established in equation (26) that E is $-\partial \log Z/\partial\beta$.

Therefore we have

$$\frac{dS}{dE} = \beta \tag{39}$$

And since the left-hand side is also the inverse temperature, we finally proved that

$$\boxed{T = \frac{1}{\beta}} \tag{40}$$

That is wonderful because we have now found out what the physical significance of β is.

β started out as simply a Lagrange multiplier.

We manipulated it, used a couple of identities, a little bit of calculus.

And we find out that β is the inverse of the temperature. Statistical Mechanics has a way of doing that!!

4.7 Questions/answers

Question: Doesn't S depend not only on E but also on β?

Answer: S depends only on one independent variable.

You can take E, or you can take β.

Or you can take even another variable.

The five variables S, β, E, Z and T are totally dependent.

There is only one independent variable among them.

It can be whichever you prefer.

When we say that equation (33) expresses S in terms of β, E and Z, we don't mean to say that β, E and Z are independent of each other.

We just mean to say that equation (33) is a useful expression.

Once again, let's emphasize that, for a system in thermal equilibrium, S, β, E, Z and T can all be expressed in terms of only one of them.

And in particular now we found that T and β are simply inverse of each other.

We consider the five variables because they each have an important interpretation in thermodynamics.

S is the entropy.

β is the second Lagrange multiplier, but it is also $1/T$.

E is the average energy of the system in thermal equilibrium.

Z is constructed from the first Lagrange multiplier α.

It is now the fundamental partition function.

And T is T.

Well, the temperature T is among other things the rate of change of energy with respect to entropy.

One paradox about temperature, is that it is at the same time the simplest physical quantity to perceive, to measure and to interpret, and it is also in some sense the least fundamental.

Question: In physics books we usually see the relation between T and β written as β equals 1 over kT.

What is the difference with equation (40) ?

Answer: If we expressed temperature in ordinary centigrade units, that is in degrees Kelvin, the formula

$$T = \frac{1}{\beta}$$

would include the Boltzmann constant and become

$$kT = \frac{1}{\beta} \tag{41}$$

But, as we said, unless for specific reasons we do otherwise, we no longer use k and the temperature, defined as the rate of change of energy with respect to entropy, is simply T.

In other words, in laboratory units you are right : β is $1/kT$, where k is Boltzmann constant.

But, in the theorist units where temperature has units of energy, T is just $1/\beta$.

Question: When we are dealing with a box of gas in thermal equilibrium, is the average energy E equal to n times the average energy of one molecule, where n is the number of molecules ?

Answer: Not in general.

If your system is defined to be a box of gas made of a collection of molecules, E is the average energy of all the molecules.

But it is not n times the average energy of one molecule.

Now you can go back and try to define your system as a single molecule.

And you might be able to get away with it under certain circumstances.

The reason you can't get away with it in all circumstances is because the molecules interact with each other.

And it was assumed that the boxes in figure 1 were extremely weakly interacting, only interacting enough to be able to exchange energy a little bit.

In the case of the ideal gas, the molecules don't interact very much.

So in that case you can reduce the system to be only one molecule.

We will study the ideal gas.

And we will actually see the relationship between T and the energy of one molecule.

Let's make a recap of the important equations we met and that we will use from now on without justifying them anymore

$$
\boxed{
\begin{aligned}
&(a) \qquad P_i = \frac{1}{Z} e^{-\beta E_i} \\
&(b) \qquad Z = \sum_i e^{-\beta E_i} \\
&(c) \qquad E = -\frac{\partial \log Z}{\partial \beta} \\
&(d) \qquad T = \frac{1}{\beta} \\
&(e) \qquad S = \beta E + \log Z
\end{aligned}
} \tag{42}
$$

Now we want to do an example drawn from real physics – approximate physics to be sure, meaning to say that we will make an approximation.

We now study the ideal gas.

4.8 The ideal gas

The ideal gas is a gas of molecules in a box where each molecule is a point molecule.

Furthermore the model is either that the molecules are so weakly interacting that we ignore the interaction between them, or that the gas is so dilute that the probability of two molecules being close enough together to interact is negligible.

Either way the approximation is that the molecules do not interact at all.

Another way of saying that is that the energy of the box, figure 3, is the sum of the kinetic energy of the molecules inside the box and nothing more.

The box has volume V.

The number of molecules in the box is N.

Therefore the density of molecules is N/V.

We will denote this density ϱ.

It should not be confused with the density of energy.

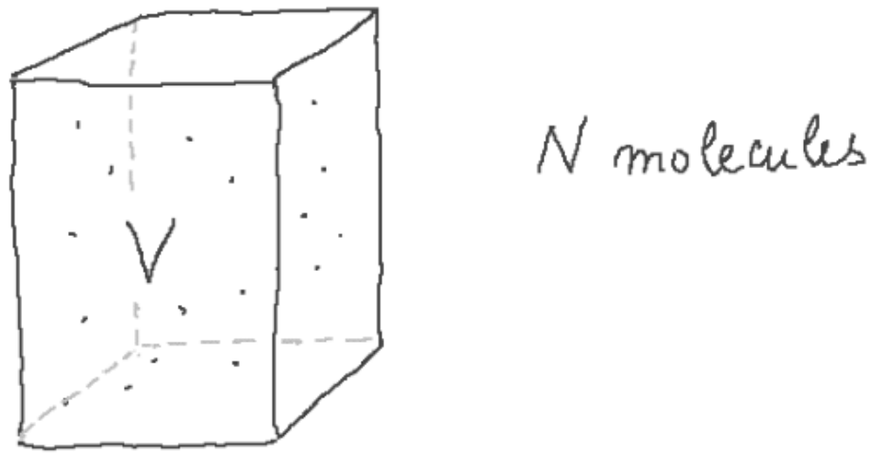


Figure 3 : Box of gas, of volume V. The molecules have kinetic energy, but don't interact with each other.

We want to calculate the partition function.

We want to calculate the energy.

We want to calculate the other thermodynamics variables.

But first of all, what are the states ω_i's or just i's?

The states are the collection of values of position and momentum for each molecule.

That is the way we label the states.

A state then is a collection of $3N$ coordinates for the positions plus a $3N$ coordinates for the momenta.

A molecule indeed has three position coordinates, and three momentum coordinates.

We can denote the $3N$ position coordinates x_1, x_2 ... all the way up to x_{3N}.

And for the momenta, p_1, p_2 ... up to p_{3N}.

A state of the system is then a point in the $6N$-dimensional space thus defined.

A set of 3 values of x and 3 values of p label one point.

What is a sum over states?

It becomes now an integral over the x's and the p's.

In other words we have to make use now of the correspondence between discrete sums and integrals(infinite sums).

We are not going to go through that in detail.

When there was a finite or countably infinite number of states, for any interesting function F of the states, its sum over all the states was denoted

$$\sum_i F(i) \tag{43}$$

Now, when the states form a $6N$-dimensional continuum, for any interesting function f of the states, the analogous integral is denoted

$$\int d^{3N}x \;\; d^{3N}p \;\; f(x,p) \tag{44}$$

The function f in this second formula is called a *density*.

Its mathematical nature is slightly different from that of the function F above.

For instance where the function F can have units of mass, the corresponding function f will have units of mass per volume, where the volume is in $6N$-dimensions.

So let's do this for the partition function.

In the discrete case we had the formula (42b) above, that is the sum over all the states of $e^{\beta E_i}$.

Now what is the energy of a state?

First of all, what it the energy of one molecule in one state?

It is just its kinetic energy.

The particle has coordinates, say, $x_1, x_2, x_3, p_1, p_2, p_3$.

Its kinetic energy is

$$\frac{1}{2m} \left[p_1^2 + p_2^2 + p_3^2 \right] \tag{45}$$

where m is the mass of one molecule.

Notice that the *x*'s don't appear, and that the three directions of momentum enter in exactly the same way.

What if we take all of the particles?

The total energy of *one state* becomes

$$\sum_{n=1}^{3N} \frac{p_n^2}{2m} \tag{46}$$

where *n* is the index labelling the positions or the momenta coordinates.

n runs from 1 to 3*N*.

Finally the partition function is

$$Z = \int d^{3N}x \;\; d^{3N}p \;\; e^{-\beta \sum_{n=1}^{3N} \frac{p_n^2}{2m}} \tag{47}$$

β is the second Lagrange multiplier we are familiar with.

It is one of the five important thermodynamics variables we talked about before, *E*, *S*, *Z*, β and *T*, which are all related.

They all depend on any one of them, which can be chosen as the independent variable.

And for the time being we use β as the independent variable.

We will meet, of course, many more interesting thermodynamics variables, for instance the pressure.

Let's assume that all the particles have the same mass m.

We can clean up equation (47) a little bit:

$$Z = \int d^{3N}x \;\; d^{3N}p \;\; e^{-\frac{\beta}{2m}\sum_{n=1}^{3N} p_n^2} \tag{48}$$

Despite its awe-inspiring look, it is a simple expression.

To start with the sum in the exponent of the exponential corresponds to a product of exponentials.

And they are all the same.

To see as clearly as possible what equation (48) is, we can spell out what is beneath the integral sign like this

$$d^{3N}x \quad dp_1 e^{-\frac{\beta}{2m}p_1^2} \quad dp_2 e^{-\frac{\beta}{2m}p_2^2} \quad \ldots \quad dp_{3N} e^{-\frac{\beta}{2m}p_{3N}^2} \tag{49}$$

In other words, next to the differential $d^{3N}x$, we have a product of $3N$ identical differentials terms

So it is simply the $3N$-th power of one of them.

Now since the integration in equation (48) is on differential elements which are neatly separate – meaning involving separate variables – we can treat the x integration separately, and also we can factorize the p integration into $3N$ identical ones.

So equation (48) simplifies into

$$Z = \left[\int d^{3N}x \right] \left[\int dp \;\; e^{-\frac{\beta}{2m}p^2} \right]^{3N} \tag{50}$$

Let's simplify this even further.

We focus first on the x integration.

It is the N-th power of $\int d^3x$.

And $\int d^3x$ is nothing but the volume V of the box, because we are integrating a small three-dimensional volume d^3x over the box.

So the first part of Z, that is the x integration, is simply V^N

The partition function now is

$$Z = V^N \left[\int dp \;\; e^{-\frac{\beta}{2m}p^2} \right]^{3N} \tag{51}$$

We shall divide the V^N in front by $N!$.

It is a little bit contentious whether this factor $N!$ ought to be there or not to be there.

Fortunately it doesn't matter whether it is there or not.

Here is why we worry about it.

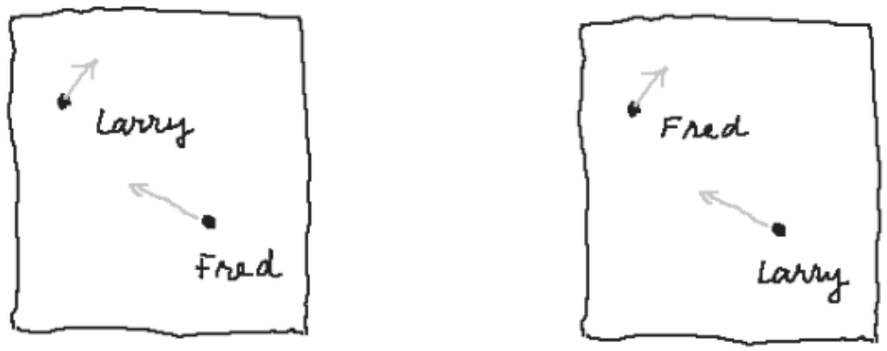


Figure 4 : Two states, where molecule Larry and molecule Fred are simply interchanged.

Consider a box with only two molecules in it.

And let's look at two possible states, figure 4.

Are they the same state or not?

It is a little bit ambiguous.

Do particles carry labels with names attached to them, so that putting Larry in (x_1, p_1) and Fred in (x_2, p_2) is different than putting Fred in (x_1, p_1) and Larry in (x_2, p_2) ?

Or are they sort of nameless things where you can just say the state corresponds to a particle over (x_1, p_1) and a particle over (x_2, p_2) ?

The answer in classical physics is that it doesn't matter.

In quantum mechanics it does matter.

And in quantum mechanics particles do not carry names.

An electron over (x_1, p_1) and an electron over (x_2, p_2), is the same as an electron over (x_2, p_2) and an electron over (x_1, p_1).

That means if there was two particles, we would say we overcount by a factor of 2 by considering all possible configurations.

Giving the particles names and then allowing them to have all possible configurations gives us a factor of two too many states.

How about 3, if there were three particles ?

The overcounting factor would be 3!.

That is 6.

For N particles, the overcounting factor would be $N!$.

So it is common practice to say that we overcounted the states in formula (51), and to divide by $N!$.

In the end it is going to make no difference – we see why, it is only a matter of how we label the elementary states ω_i, define the P_i's, etc. –, but it will do nice things to the partition function.

So now the partition function is

$$Z = \frac{V^N}{N!} \left[\int dp \;\; e^{-\frac{\beta}{2m}p^2} \right]^{3N} \tag{52}$$

It looks terribly complicated but it is not.

The term between brackets is just a definite integral.

It is a number.

What could it depend on?

It could depend on β.

It could depend on m.

It doesn't depend on p since p is the integration variable.

Let's calculate the integral $\int dp\ e^{-\frac{\beta}{2m}p^2}$ which is in the brackets.

It is called a Gaussian integral, because it is the same term, up to a factor, underneath the integral sign, which appears in a Gaussian density of probability with mean 0 and a standard deviation $\sqrt{m/\beta}$.

The expert will know that the integral is then $\sqrt{2\pi m/\beta}$, but let's do it anyway.

We begin by changing variable to make the exponent less unwieldy. So we define

$$q^2 = \frac{\beta}{2m}\, p^2 \tag{53}$$

The differentials satisfy

$$dp = \sqrt{\frac{2m}{\beta}}\, dq \tag{54}$$

Then we can rewrite the integral

$$\int dp\ e^{-\frac{\beta}{2m}p^2} = \sqrt{\frac{2m}{\beta}} \int dq\ e^{-q^2} \tag{55}$$

The term $\int dq\ e^{-q^2}$ is now really just a number.

It doesn't depend on any parameter.

Its value is π

The standard trick to see it is to look at its square.

Express it as a double integral over the entire plane $\int \int e^{-(x^2+y^2)} dx\ dy$

Then switch to polar coordinates. The double integral become $\int_0^{2\pi} \int_0^{\infty} e^{-r^2} r\ dr\ d\theta$

Now it is elementary to show that this is equal to π.

Finally we find that the partition function for an ideal gas made of *N* particles is

$$Z = \frac{V^N}{N!} \left[\frac{2m\pi}{\beta} \right]^{\frac{3N}{2}} \tag{56}$$

It is fine in this form.

But let see what the *N*! does for us, if we include it.

If we don't include it, we just go through the calculation the same way.

But still let's see where it goes.

Let's take $V^N/N!$ and make N big.

We are talking about something like Avogadro's number of molecules, which is 6.02 x 10^{23} molecules, in other words a huge number indeed.

Then $N!$ is even a huger number, and we can approximate it by Stirling's formula.

$$\frac{V^N}{N!} \approx \frac{V^N}{\sqrt{2\pi N}\ N^N e^{-N}} \tag{57}$$

Now N/V is ϱ.

So we can rewrite the partition function as

$$Z = \frac{1}{\sqrt{2\pi N}} \left[\frac{e}{\rho}\right]^N \left[\frac{2m\pi}{\beta}\right]^{\frac{3N}{2}} \tag{58}$$

Since we are going to take the logarithm, and then differentiate, we can forget about the multiplicative factor in front, which comes from the precise version of Stirling's approximation, and just write

$$Z = \left[\frac{e}{\rho}\right]^N \left[\frac{2m\pi}{\beta}\right]^{\frac{3N}{2}} \tag{59}$$

To arrive at this formula, what did we do ?

We knew what we wanted to calculate.

We went onto autopilot, started doing integrals.

And at the end of the day here is Z.

Notice first of all that it depends on the density ϱ.

The density, for a fixed number of particles in a fixed box, is just a number.

So it is worth noting that Z depends on the density.

But it is just a number.

2, m and π are numbers.

N is a number.

Only β in formula (59) at first is interesting.

It is also a number of course.

But it is an interesting one.

It is a variable we may want to change in the course of the problem.

For example we may want to differentiate with respect to it, which we will do momentarily.

Let's calculate the logarithm of Z.

We mentioned that the logarithm of Z is a thing which comes up over and over.

For instance it comes up in formula (42c) for the average energy E.

So let's calculate log Z.

So, incidentally, whenever you get a product of things and you want to differentiate it, there is an old saying – going back to Zarathoustra – that it is easier to differentiate a sum than a product.

So it is good to take their logarithm because they are easier to differentiate

Partition functions are products typically. It is easier to differentiate their logarithm.

So let's take the logarithm of Z defined by equation (59)

$$\log Z = N \left\{ \frac{3}{2} \log \frac{2m\pi}{\beta} - \log \rho + 1 \right\} \tag{60}$$

When we differentiate, the constants don't play any role.

The 1 will have no effect, neither will have the 2, the m, nor the π.

The only term which is important, when we differentiate log Z with respect to β, is the $-3N/2\log\beta$.

Everything else functions as a constant, including the term involving the density.

So we can write equation (60) as

$$\log Z = -\frac{3N}{2}\log\beta + \text{ constants} \tag{61}$$

Before we differentiate it, let's look at it for a moment.

In the factor 3/2, where does the 3 come from?

It comes from the three dimensions of space.

Suppose there was eleven dimensions of space.

We would get 11/2.

Where does the 1/2 come from?

It comes from the Gaussian integral, equation (55).

β appeared next to p^2.

To integrate the Gaussian we had to take, at some point, the square root of m/β.

That is where all the pieces making up log Z come from.

Now that we have log Z, what can we calculate ?

We can calculate the entropy, equation (42e).

But even more interesting for the moment and simpler is to calculate the total energy of the system, equation (42c).

All we have to do is differentiate log Z, given by equation (61), with respect to β, and a minus sign.

That gives us

$$-\frac{\partial \log Z}{\partial \beta} = \frac{3N}{2}\frac{1}{\beta} \tag{62}$$

The left-hand side is the average energy.

And on the right- hand side, 1/β is simply the temperature.

So we reach

$$E = \frac{3N}{2}\, T \tag{63}$$

This is remarkable.

For the ideal gas, *E* and *T* are the same thing with a simple proportionality factor.

We saw where the 3 and the 2 came from.

And regarding the *N*, it certainly makes sense to say that the total energy of the box of gas is proportional to the number of particles.

Therefore equation (63) also tells us that, for an ideal gas, the *energy per particle* is

$$\frac{E}{N} = \frac{3}{2}\, T \tag{64}$$

This is true for an ideal gas.

It is not true in general, although it is often a good approximation.

Where would Boltzmann constant be?

Whenever you see a formula like equation (64), and you want to convert it to laboratory units, the temperature becomes k_B times the temperature.

So in laboratory units, formula (64), becomes

$$\frac{E}{N} = \frac{3}{2}\, k_B T \tag{65}$$

k_B is of course a very small number.

And so if the temperature is 300 degrees K, we multiply 300 by this terribly small Boltzmann factor, and we get the average energy per particle in Joules.

Formula (64) – or its variant (65) – is where the idea comes from that the particles move around with a kinetic energy which is proportional to the temperature.

We can read equation (64) another way.

Since the kinetic energy is the sum of three term $p^2_{\mathrm{x}} + p^2_{\mathrm{y}} + p^2_{z}$, all divided by $2m$, we can say that for each direction of space the energy stored in the momentum in that direction is 1/2 kT. And so if there are three directions of space, this yields 3/2 kT.

We stressed that the relationship $E = N$ x 3/2 T holds for an ideal gas.

That is a gas where the particles only have kinetic energy.

They are so small or so dilute that they don't interact with each other.

Moreover they don't attract or repel each other.

A pair of particles doesn't have potential energy like an elongated or compressed spring.

We also ignored potential energy that could come from gravity.

How might equation (64) change if the gas has a potential energy term coming from gravity?

Suppose each particle vertical position was labelled by y_n, then its potential energy would be $-m\ g\ y_n$. The big formula (49) for Z would now include a term for this potential energy per particle and become

$$Z = \int d^{3N}x \;\; d^{3N}p \;\; e^{-\frac{\beta}{2m}\sum_n (p_{x,n}^2 + p_{y,n}^2 + p_{z,n}^2)} \;\; e^{-\beta \sum_n mgy_n} \tag{66}$$

The index n in the exponents now labels the N particles.

Let's see how the extra potential energy term, under the integral sign, changes the integration.

First of all, consider the integration over position.

The term $d^{3N}x$ can also be denoted $d^{N}x\, d^{N}y\, d^{N}z$. The extra energy term involving y doesn't change the integration over x and z.

So what had produced the volume of the box to the N-th power now produces the area A of the base of the box to the N-th power, times an integration over the vertical dimension y.

On the other hand the integration over momentum remains the same.

So the formula for the partition function, which was given by equation (52), now becomes

$$Z = \frac{A^N}{N!} \left[\int dp \;\; e^{-\frac{\beta}{2m}p^2} \right]^{3N} \left[\int e^{-\beta mgy} dy \right]^{N} \qquad (67)$$

The range of integration of the second integral is from 0 to the height of the box.

Since the extra term on the right-hand side of equation (67) is β dependent, it will give some contribution to the energy of the whole box, or to the energy per particle.

Exercise: In equation (67), calculate the two integrals.

Then calculate the average energy of the box.

Let's solve exercise in the case where the box is infinitely high.

This is literally the problem of the atmosphere between the Earth's surface and infinity.

The new integral term on the right, on the right-hand side of formula (67), is $1/(\beta mg)^N$.

And the middle integral is exactly what it was before, see equation (56).

So we have

$$Z = \frac{A^N}{N!} \left[\frac{2m\pi}{\beta} \right]^{\frac{3N}{2}} \left[\frac{1}{\beta mg} \right]^N \tag{68}$$

Notice that everything is to the *N*-th power.

That is nice because when we will take the logarithm, this will yield a proportionality factor *N*, corresponding naturally to the number of particles in the box.

So taking the logarithm of *Z*, and spelling out only the terms depending on β, we get

$$\log Z = -\frac{3N}{2} \log \beta - N \log \beta + \text{constants} \tag{69}$$

Then we take the derivative with respect to β, with a minus sign.

This yields

$$E = \frac{3}{2}NT + NT = \frac{5}{2}NT \tag{70}$$

This formula holds in the case where there is gravitational potential energy.

It may look disturbing because the temperature appears in the potential energy term which, it seems, should have nothing to do with it.

But let's not forget that the natural link, $E_{\text{particle}} = 3/2\ T$, between temperature and kinetic energy is only true in the ideal gas *without potential energy* of any sort.

When there is potential energy, temperature no longer has that simple interpretation.

4.9 Questions/answers

Question: All the equations we derived in this part rest on the Boltzmann distribution, is that right ?

Answer: Yes. And all this is in the case of thermal equilibrium.

We calculated the distribution of probability which maximizes the entropy subject to two constraints.

The first one is hardly a constraint ; it is simply $\sum_i P_i = 1$

But it has to be included in the mathematics.

The other is that the average energy is a given value E.

And the result we find is called the Boltzmann distribution.

The Boltzmann distribution is the probability distribution then over the possible states, that is essentially the energy levels, of the system.

It is given by equation (42a), which is reproduced below

$$P_i = \frac{1}{Z} e^{-\beta E_i} \tag{71}$$

We haven't expressed the Boltzmann distribution more explicitly, because there is no point.

Like all the other thermodynamics variables we have encountered – Z, β, S, T, E itself – it depends ultimately on the energy levels E_i's and the average energy E.

Moreover, as we saw, considering the energy levels as fixed parameters once and for all, all these variables, Z, β, S, T, and E, depend on any one of them, which can be selected as the independent variable.

When two variables appear in the same expression, as for instance Z and β in equation (71), we don't mean to say that they are independent from each other.

It is just a convenient way to express P_i.

The variable most often selected as the independent variable is β.

Statistical mechanics is pretty nifty.

After a while, you get familiar with the simple tricks.

You know which equation to use, one following the other.

You follow a set of procedures.

But you never get to the point where you are not surprised by the equations that pop up once in a while in the calculations.

It is always the same pattern.

You turn the crank, go through equations that seem totally uninteresting.

Then all of a sudden, you get to a thing and you say :

Whaaoow this is an interesting equation !

Statistical mechanics is a very curious subject.

Question: You introduced gravity potential energy.

Will we be able to apply this at some point to collapsing stars ?

Answer: If you mean a black hole, yes that is easy.

If you mean a collapsing star, or even a star not collapsing, that is not easy.

Stars are hard, not because they are conceptually hard, but only because there is a lot of details.

Question: So in a collapsing black hole, the entropy will increase ?

Answer: Yes, the entropy does go up to a point where it reaches the so-called Bekenstein entropy.

Then it cannot go higher than than – not unless you add more energy.

For a given amount of energy and a given volume, that is it.

Question: How can we use the partition function to calculate fluctuations ?

Answer: Let's look at the fluctuations in energy for example.

That is not a subject I intended to treat in this part of our discussion, but let me tell you what the trick is.

If you want to calculate the fluctuations in a quantity, what you want to first calculate is the average of the square of that quantity.

Fluctuations here means the width of the distribution of probability of the quantity we are considering.

That is what probabilists and statisticians call the *standard deviation*, or the square root of the variance of the probability distribution.

The random variable we consider here is E_{i}.

We already know its mean.

And now we are interested in its variance.

To calculate the square of the energy, you want to differentiate the partition function twice with respect to β.

You will find a formula involving second derivatives.

It is a very beautiful formula for the fluctuations of the energy.

It is also an important formula.

It relates the fluctuations in energy and the specific heat.

We will work it out.

Fluctuations were not a part of the original thermodynamics theory, as it was originally developed.

It was Einstein and Gibbs who really understood that the way to find out, in the framework of the statistical mechanics of a finite collection of atoms, as opposed to a complete continuum with no basis in statistics, was to study these fluctuations.

So we will do a little bit of that in the next part

5. Pressure of an ideal gas and fluctuations

5.1 Introduction

Over the last four parts of our discussion, you have seen how subtle statistical mechanics can be.

It is full of all these fancy calculus tricks, with sometimes very little intuition about what is really going on.

It is all in the calculus tricks then.

In this part of our discussion, we are going to study the problem of pressure of an ideal gas.

Intuitively we know where the pressure comes from.

It comes from the molecules bouncing against the walls of a system.

We could calculate it by just using the average energy of every molecule.

We already found that $E_{\text{particle}} = 3/2\ kT$.

So we know how fast they are moving on the average.

We can estimate how many of them are moving in each direction because they are pretty isotropic.

And that leads to a calculation of the force exerted on a wall using Newton's law in its form $F\Delta t = m\Delta v$.

We might wonder, incidentally, if the molecules near the wall are isotropic.

After all the presence of the wall influences what is going on near it.

And we might think : well, maybe there are more going horizontally to the wall than going vertically, that is parallel to the wall, and then going horizontally the other way.

It turns out not to be true.

Even very close to the wall, the distribution of velocities is still pretty isotropic.

But we might wonder whether that is a principle or just an accident.

We might also wonder whether the velocities near the wall happen to be different than the velocities in the interior far from the wall – in which case we would be making a mistake by just taking the good old 3/2 kT for every molecule no matter where they are.

It turns out we are not making a mistake.

But how do we know we are not making a mistake ?

We know from statistical mechanics.

We know from an honest set of principles, which don't use intuitive pictures of molecules hitting the wall and so forth, but which really use the basic foundations of statistical mechanics.

The foundations of statistical mechanics are mathematical.

They are not complicated as we have seen.

They are not hard.

But they are very mathematical.

So we emphasized over and over that in situations where we purposefully give up intuitions because we are not sure they are correct, such as for example whether the velocity distribution near the wall is the same as in the interior of the box, then we rely on the mathematics.

When we are writing the mathematics, we are sort of on autopilot using the rules.

We first figure out what we want.

Then we just start cranking, using here and there, whatever tricks we have, until we suddenly run across a formula that we recognize as meaningful, interesting and important.

That is just the way statistical mechanics is.

You get good at it after a while.

The great physicists that I have known all were very good at it.

I don't mean people who just studied statistical mechanics,

I mean all the very good physicists somehow love thermodynamics.

Why ?

It is fun and full of surprises.

It abounds with surprising relationships which suddenly fit together, and tell you some physics.

That is the attraction of statistical mechanics for many of us.

Einstein was the grandmaster of thermodynamics(Statistical mechanics is also called statistical thermodynamics.)

Feynman was a grandmaster of it.

I am a not really a master of it.

Well, maybe, maybe……..

And it is good stuff.

Question: Once we have done the mathematics of a problem, can't we check them with our intuition ?

Answer: Yes we can, but what we are really doing is checking our intuition not the mathematics.

For instance, it could conceivably have been wrong that the distribution of velocities near the wall was the same as in the interior.

We don't want to rely on that.

We want to use a set of fundamental principles, and then rules and mathematics to calculate the pressure.

You don't want to go through all the mathematics, then say :

Well, I think I proved that the rules of calculus work out ok, because they agree with my naïve intuitions about the way molecules behave near the wall.

It is the other way.

You are checking your intuitions from the mathematics.

So we could calculate the pressure on the walls as follows.

Let's not even talk about averages.

Let's say the room is filled with molecules, all moving with the same velocity v satisfying

$$\frac{1}{2}\, mv^2 = \frac{3}{2}\, T \tag{1}$$

using our favorite units such that temperature doesn't carry Boltzmann constant with it.

We know the mass of every molecule and we know the temperature.

Therefore we know the velocity of every molecule.

When we know the velocity of every molecule, we can compute the force on the wall by asking how many molecules per unit time hit the wall – if indeed they are distributed uniformly, which we don't know for sure.

If we know that the directions of their velocities are isotropic, we can calculate the number of collisions on the wall per unit time and the momentum delivered by each particle to the wall.

Particles bounce off the wall and the perpendicular component of momentum which was originally going to the wall is now going away from the wall.

We can calculate the change in momentum.

It corresponds to a force exerted by the wall on the particles, and therefore by the particles on the wall too.

Of course force is rate of change of momentum.

In fact, that defines the *average force over time*.

Indeed, everybody knows that hail does not feel the same as rain.

And so we can calculate the force on the wall.

Pressure is the force per unit area.

And so we can calculate that.

It is fun. See later…..

You can try to do it yourself.

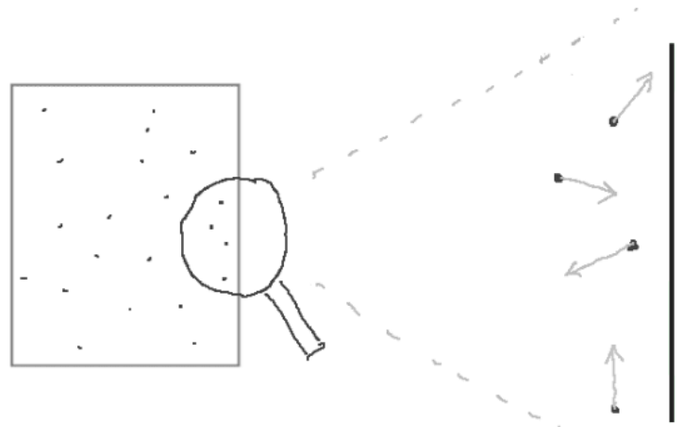
The method, when the molecules are assumed to be point masses isotropically distributed in position and velocity and completely free – no potential energy, no collisions between molecules, and so forth – yields the right answer.

But that is not really good enough.

We really want a set of rules which are good enough that they can take into account collisions between the molecules.

Collisions between the molecules can upset the whole thing a lot.

And they do.



Box filled with molecules exerting pressure on the walls. And view with a magnifier.

Derivation:

$$\Delta p_{molecule} = m\Delta v = 2mv$$

elastic collision with wall

$$F_{molecule} = \frac{\Delta p_{molecule}}{\Delta t} = \frac{2mv}{\Delta t}$$

volume of molecules that hits a wall in Δt = $V = (area\, of\, wall)v\Delta t = Av\Delta t$

Q = number of molecules in that volume = ρV where ρ = number density of molecules = N/V

$$F_{wall} = \frac{Q}{2} F_{molecule}$$

1/2 of molecules going in each direction

$$P = \frac{F_{wall}}{A} = \frac{N}{Q} \frac{2mv}{\Delta t} = 2\rho m v^2 = 2\rho E = \quad \rho T = \frac{2NT}{V}$$

putting k back in

$$PV = NkT$$

which is Boyle's law for an ideal gas

Now we want a set of rules which is robust, which doesn't depend on the detailed oversimplified model of a free gas (as in last discussion).

Free gas means a gas of particles where the particles are not interacting with each other, that is an ideal gas

So the ideal gas is a model that is easy.

It is easy to do things with.

But it is not very general.

The rules of statistical mechanics are very general.

And they would be sufficient – if you can do the calculations – to calculate the relationship between the pressure and other variables very generally.

What is it that we want?

We want something called the equation of state.

It will relate the pressure to other thermodynamical variables.

To establish that equation we need some preliminary concepts and results.

5.2 Control parameters

We would like to calculate or know what the pressure is as a function of temperature, volume, and possibly other relevant thermodynamics variables.

What is it that determines the pressure of the gas in the room where we are ?

Basically the temperature and the volume. That is about it.

If we consider that the number of molecules in the room is fixed, i.e., if we don't let it vary, then the pressure is a function of the temperature and the volume.

We would like to calculate the pressure in terms of those variables.

So we need a theory.

We need a set of principles.

We are going to describe those principles and work out the pressure of an ideal gas, with no intuition whatever, no picture in our minds of what we are doing.

I'm not saying that pictures in our mind aren't good things for understanding physics.

But there are times when you want to suppress the pictures and simply go with the principles and the mathematics.

And statistical mechanics is such a theory, because we want the rules to be general.

Let's recall a few statistical thermodynamics facts and formulas that we worked out before and that we will use.

The first one has to do with entropy.

Whatever the system and the model are, we have

$$S = -\sum_i P_i \ \log P_i \tag{2}$$

And if the distribution is the Boltzmann distribution, then

$$P_i = \frac{1}{Z} \ e^{-\beta E_i} \tag{3}$$

In that case equation (2) rewrites

$$S = -\sum_i \frac{1}{Z} \ e^{-\beta E_i} \ (-\beta E_i - \log Z) \tag{4}$$

The minus signs cancel, so this is also

$$S = \sum_i \frac{1}{Z} \ e^{-\beta E_i} \ (\beta E_i + \log Z) \tag{5}$$

Because of the two terms in the parentheses, this is a sum of two sums.

Pulling β, the first sum is the expression of the average energy.

So we get

$$S = \beta E + \sum_i \frac{1}{Z} \ e^{-\beta E_i} \ \log Z \tag{6}$$

Then in the remaining sum, the $1/Z$ and $\log Z$ can be pulled out of the sum too.

$$S = \beta E + \frac{\log Z}{Z} \sum_i e^{-\beta E_i} \tag{7}$$

Finally the last sum is the definition of Z itself.

It cancels with the Z in the denominator.

We arrive at

$$S = \beta E + \log Z \tag{8}$$

Now there is another standard way to write this.

We note that β is the inverse of temperature.

We multiply both sides by T.

Then we put on one side $E - TS$, and on the other side what it is equal to.

$$E - TS = -T \log Z \tag{9}$$

The left-hand side has a name.

It is called the *Helmholtz free energy*.

It occurs over and over in physics.

For us it is only important because it appears in equation (9) which is in another guise equation (8), and we are going to use it a lot.

The Helmholtz free energy is denoted A.

$$A = E - TS \tag{10}$$

Quantities are given names in physics when they occur over and over again.

We are going to find out that A plays a leading role in calculating the pressure, as well as all kinds of other things.

So A is another thermodynamic variable.

We already had the average energy E, the entropy S, the temperature T, etc.

We now have a new variable called the Helmholtz free energy.

Remember that all these variables are dependent on each other.

They are all dependent on any one of them, which can be chosen as the independent variable.

At times it is in turn E, or β, or T, etc.

We are going to introduce yet another variable into the problem.

Variables of this next kind are often called *control parameters*.

Control parameters are parameters that we, as experimenters, can easily change.

They are macroscopic parameters.

We are not trying to change details of one molecule at a time.

An example of control parameter is the volume of the container of our gas, figure 2.

We can move the piston back and forth, and in the process change the volume of gas.

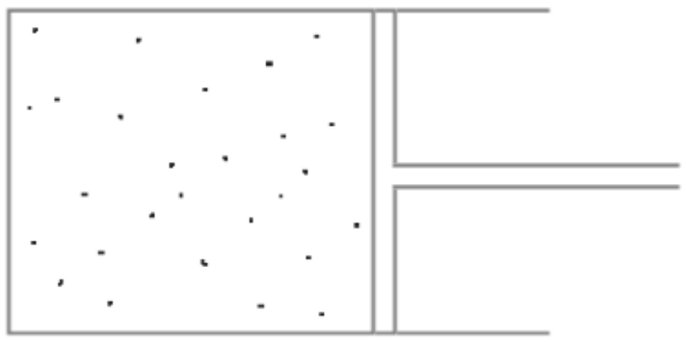


Figure 2 : Controlling the volume of the gas with a piston.

So volume is a control parameter.

Other control parameters could be the magnetic field on a system, the electric field on a system.

Is pressure one ?

No, pressure is not a control parameter.

We shall see that variables come in pairs.

One is usually called the control parameter and the other is called the *conjugate thermodynamical variable*.

Pressure and volume are closely connected.

They form a conjugate pair.

We can have a pressure gauge inside the volume of gas in figure 2, and by moving the piston change the pressure.

But of course what we are doing is changing the volume, and the response is a change in pressure.

On the other hand, changing the number of molecules inside the box would be a control parameter.

Notice that we talk about particles of gas inside the volume, because it is easy to imagine, particularly when we compress it.

But it is not a fundamental assumption.

It could be anything.

It could be a liquid.

It could be a solid.

We would be squeezing on the solid.

It doesn’t have to be homogeneous either.

We could have inside the box, in thermal equilibrium at the boiling temperature, a mix of steam and water simultaneously.

That would be neither a gas nor a liquid.

And if the box was in a gravitational field, the liquid might sit at the bottom and the gas at the top.

It would still be thermal equilibrium.

Same comment for a mix of two gases or anything.

That is why it is important to have general methods.

You cannot use the ideal gas law for a solid inside the box !

5.3 A mathematical result in differential calculus

We need a little mathematical theorem in differential calculus.

Although totally unintuitive, it is very simple.

So maybe it should only be called a lemma.

Suppose we have two functions of two variables.

One function will be called E.

Eventually it will be the energy.

And the other function will be called S.

That will be for entropy.

But the mathematical theorem we are going to prove is a very general result in calculus.

The independent variables are going to be called T and V.

There again, they will have interpretations that you can guess.

But for the moment let's just think of them as independent variables, that is, if we prefer, parameters that we can control from the outside.

If we knew the independent variables T and V, and knew enough about the system, we could calculate the dependent variables E and S.

The theorem says that the following relationship always holds

$$\left.\frac{\partial E}{\partial V}\right|_S = \left.\frac{\partial E}{\partial V}\right|_T - \left.\frac{\partial E}{\partial S}\right|_V \left.\frac{\partial S}{\partial V}\right|_T \tag{11}$$

The notations call for some comments.

The left-hand side means the partial derivative of E with respect to V keeping S fixed.

Usually we would talk about the partial derivative of E with respect to V keeping the other independent variable, namely T, fixed.

But here it is different, in the computation of the left-hand side the quantity being kept fixed is the other dependent variable S.

On the right-hand side, it begins with the more standard partial derivative of E with respect to V keeping T fixed.

And the other two partial derivatives are interpreted according to what the precise notations are, and the same explanations.

Here is the proof.

It is a lovely little proof.

I like it so much because I figured it out myself(although later I found out that it had been done).

You sort of fall in love with those little things that you figure out yourself when you are young.

On the plane of the two independent variables, T and V, let's draw one contour line of S, figure 3.

The contour line doesn't have to have necessarily this shape.

It can be any shape.

We just represented one shape to help us with the subsequent reasoning.

The important point is that S has a fixed value along the contour line shown in figure 3.

Now we want to calculate the rate of change of E, the other dependent variable, with respect to the volume V , but along that line of constant S.

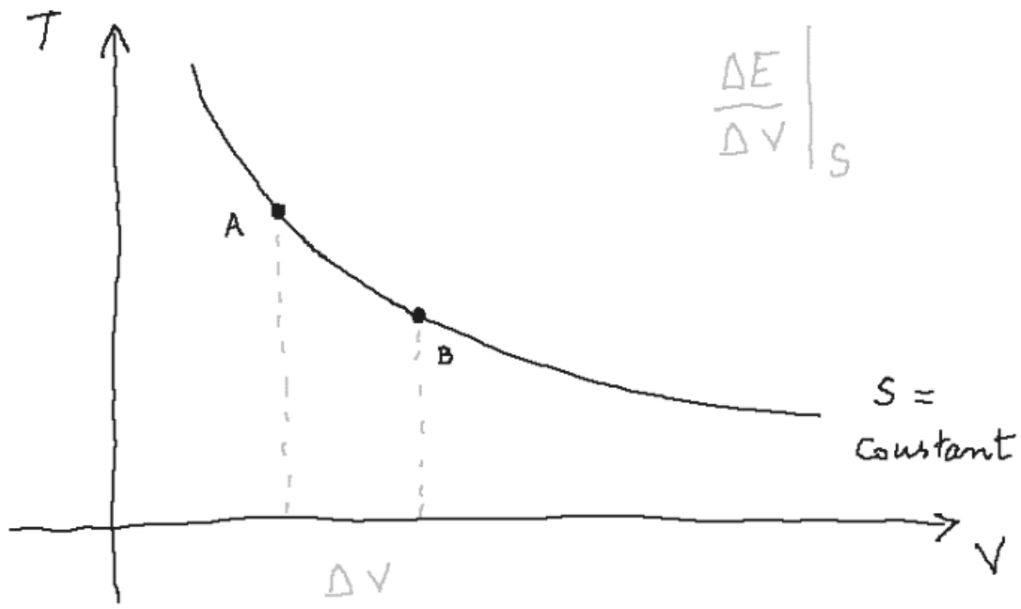


Figure 3 : One contour line of S(V,T). And two points at different values of V, on that contour line.

Let's begin with the following ordinary formula of calculus relating a small increase in the dependent variable E to small increases in the independent variables V and T and the corresponding partial derivatives

$$\Delta E = \left.\frac{\partial E}{\partial V}\right|_T \Delta V + \left.\frac{\partial E}{\partial T}\right|_V \Delta T \tag{12}$$

We use the notation ΔE instead of dE only to remind ourselves that we are looking at small variations and ratios of small variations.

We could use the notation dE, dV and so forth in equation (12).

And of course derivatives are limits of ratios.

Forgetting temporarily that A and B are points on a specific curve, equation (12) is valid for any points A and B nearby each other.

It says that the variation in E going from point A to point B, namely ΔE, is equal to the variation in E going from A to C, plus the variation in E going from C to B, figure 4.

That is the general meaning of equation (12).

$$\Delta E = \left.\frac{\partial E}{\partial V}\right|_T \Delta V + \left.\frac{\partial E}{\partial T}\right|_V \Delta T$$

Our objective is not to prove equation (12), which is an ordinary formula of calculus that we know is true, but to prove equation (11). $\left.\frac{\partial E}{\partial V}\right|_S = \left.\frac{\partial E}{\partial V}\right|_T - \left.\frac{\partial E}{\partial S}\right|_V \left.\frac{\partial S}{\partial V}\right|_T$

The first term on the right-hand side of equation (12) is the same as the first term on the right-hand side of equation (11).

So it is the second term of equation (12) that we want to manipulate, and see what it is when we move on a curve of constant S.

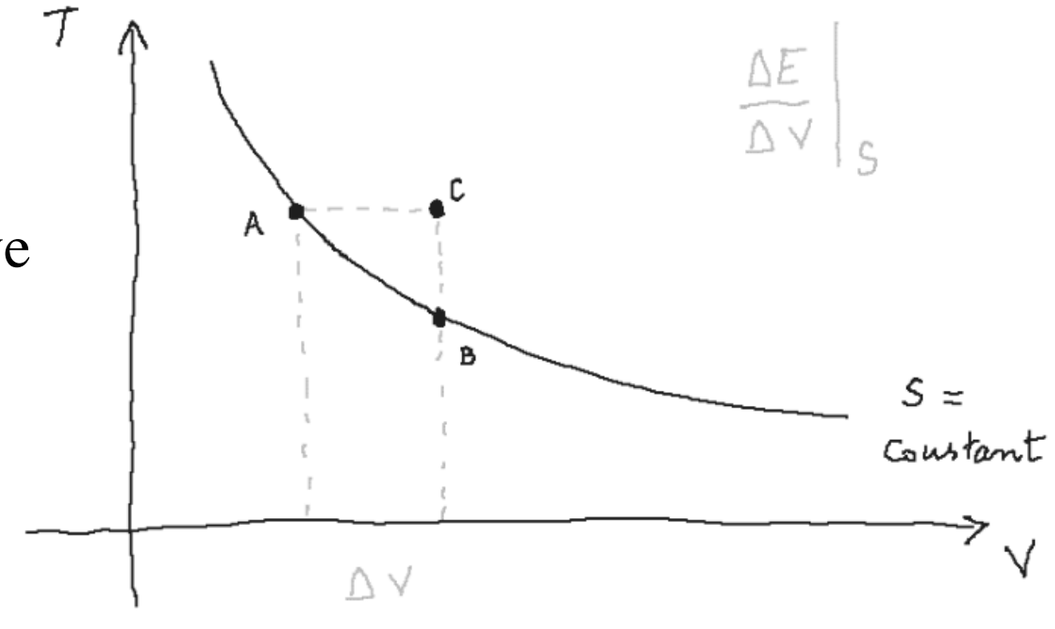


Figure 4 : E(B)–E(A) = E(C)–E(A) + E(B)–E(C)

For the second term on the right-hand side of equation (12), we can write

$$\left.\frac{\partial E}{\partial T}\right|_V \Delta T = \left.\frac{\partial S}{\partial T}\right|_V \left.\frac{\partial E}{\partial S}\right|_V \Delta T \tag{13}$$

This is also true if we divide by ΔV

$$\left.\frac{\partial E}{\partial T}\right|_V \frac{\Delta T}{\Delta V} = \left.\frac{\partial S}{\partial T}\right|_V \left.\frac{\partial E}{\partial S}\right|_V \frac{\Delta T}{\Delta V} \tag{14}$$

So far, this comes from equation (12), that is, it is true for any points A and B.

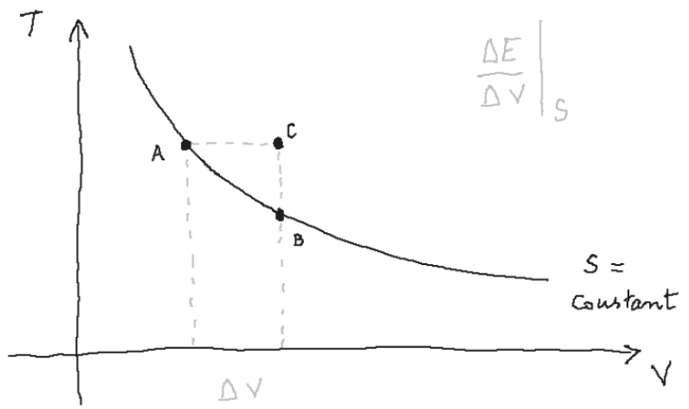
Nowhere have we used yet the fact that we are moving on a line of constant S.

Now we are going to use the fact that we are moving on a line of constant entropy.

Why we are interested in constant entropy?

That doesn't come from the calculus manipulations we are doing.

But it will come out soon enough.



We are moving from A to B in figure 4.

i.e., on a curve S = constant.

This curve is characterized by a slope between A and B.

That slope is simply $\Delta T / \Delta V$.

Let's use a little bit of a calculus to see if we can work out something about $\Delta T / \Delta V$.

So let's write down the condition that we are moving on a line of constant S.

Along that curve we have $\Delta S = 0$.

But we also have the general formula that always holds

$$\Delta S = \left.\frac{\partial S}{\partial V}\right|_T \Delta V + \left.\frac{\partial S}{\partial T}\right|_V \Delta T \tag{15}$$

So on the line of constant S we have

$$0 = \left.\frac{\partial S}{\partial V}\right|_T \Delta V + \left.\frac{\partial S}{\partial T}\right|_V \Delta T \tag{16}$$

Now we are going to be able to get a formula for $\Delta T/\Delta V$ in terms of the partial derivatives in equation (16).

Dividing by ΔV and reorganizing a bit – noting that the inverse of $\partial S/\partial T$ at constant V is simply $\partial T/\partial S$ at constant V – we get

$$\frac{\Delta T}{\Delta V} = -\left.\frac{\partial S}{\partial V}\right|_T \left.\frac{\partial T}{\partial S}\right|_V \tag{17}$$

Now we are going to plug it into formula (14) which comes from the second part of the right-hand side of equation (12).

$$\left.\frac{\partial E}{\partial T}\right|_V \frac{\Delta T}{\Delta V} = -\left.\frac{\partial S}{\partial T}\right|_V \left.\frac{\partial E}{\partial S}\right|_V \left.\frac{\partial S}{\partial V}\right|_T \left.\frac{\partial T}{\partial S}\right|_V \tag{18}$$

By a great stroke of luck, on the right-hand side, the first and the last partial derivatives are inverse of each other.

So their product is 1.

And formula (18) simplifies into

$$\left.\frac{\partial E}{\partial T}\right|_V \frac{\Delta T}{\Delta V} = -\left.\frac{\partial E}{\partial S}\right|_V \left.\frac{\partial S}{\partial V}\right|_T \tag{19}$$

Finally, going back to equation (12), dividing both sides by ΔV , we reach equation (11)

$$\left.\frac{\partial E}{\partial V}\right|_S = \left.\frac{\partial E}{\partial V}\right|_T - \left.\frac{\partial E}{\partial S}\right|_V \left.\frac{\partial S}{\partial V}\right|_T \tag{20}$$

Q.E.D

The last item was an abbreviation of *Quod Erat Demonstrandum*, the latin expression to mean *what was to be proved*.

The left-hand side stipulates that we are moving on a line of constant S. $$\left.\frac{\partial E}{\partial V}\right|_S = \left.\frac{\partial E}{\partial V}\right|_T - \left.\frac{\partial E}{\partial S}\right|_V \left.\frac{\partial S}{\partial V}\right|_T$$

And we used that fact on the right-hand side when we transformed equation (15) into equation (16) and then continued to turn the crank from there.

You may be concerned that the derivative of entropy with respect to temperature, which we eliminated in equation (18), should not be zero.

For most systems in most circumstances that is true.

There are some special cases where it is not true.

Then we have to be more careful.

But generally speaking entropy is a monotonic function of temperature, always increasing.

The reason is simple.

As you increase the temperature, the Boltzmann distribution gets broader.

As you decrease the temperature it gets narrower.

And entropy is just a measure of that width.

So we have established two facts : the link between Helmholtz free energy, temperature and the partition function, equation (9) ; and the above lemma.

The other facts that we need have to do with pressure.



5.4 Pressure

First of all we would like to emphasize that pressure is just one special case of a response to a control parameter.

We need to understand what pressure really is.

Going back to figure 2, we have a cylinder, gas inside it, and a piston.

The molecules of gas are hitting the walls – in particular the piston, exerting pressure on it.

Supposing we move the piston a little bit, the pressure in the interior does work on the piston.

If we move the piston out, the pressure pushes out in the same direction as the displacement.

Then there is positive work done on the piston.

Where does that work come from?

Work is a form of energy.

The other fundamental form of energy, in thermodynamics, is heat.

In our discussion, we shall understand the difference between work and heat.

And we shall see that heat, like entropy, is a very special characteristic of a system.

The work done on the piston is equal, with an opposite sign, to the change in energy inside the cylinder.

For convenience, let's redraw figure 2 with the piston moving vertically, figure 3

Imagine that the piston rises a little bit, by an amount *dx*.

That means, in this case, that there has been an increase in the gravitational potential energy of the piston.

That is called the work done on the piston by the pressure.

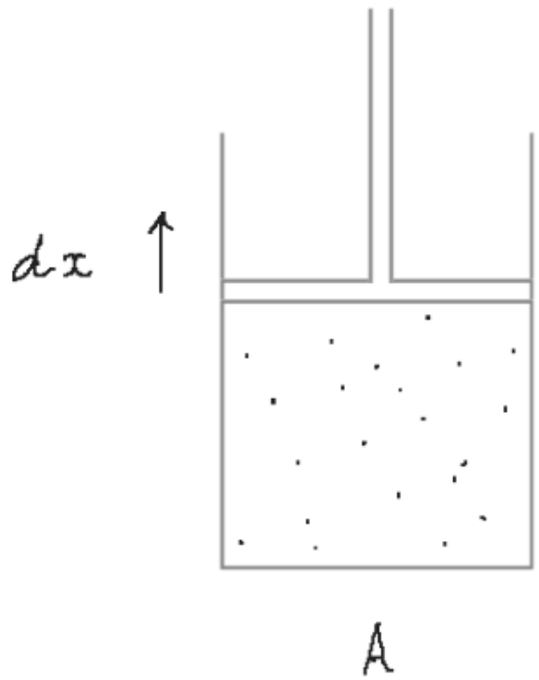


Figure 5 : Cylinder with a piston moving vertically.
The area of the base of the cylinder is A.

And by energy conservation, any work that is done on the piston must be taken away from the gas.

So energy conservation tells us that the change in energy of the gas is the opposite of the work done on the piston.

We also have to imagine something else : first of all, we have to do this slowly.

Some odd things can happen if we do it fast.

For example we could pull a piston out so fast that no molecule has a chance to hit the piston as it is moving.

Then, during that operation, there would be no pressure on the piston.

That is a rather extreme situation, of course, but it does illustrate that we want to define pressure by being able to average over many collisions.

So we want to slowly move the piston.

Secondly, we want to make sure that no energy comes into the system from outside.

In other words we want to insulate the walls of the system.

I remember, when I was student, I learned the word *adiabatic*.

I had no idea what it meant.

But my professor told me that it means that a system is insulated from heat coming into it or out of it.

Then later on, when I was in graduate school, another professor told me adiabatic means slowly.

Second try is correct idea — what adiabatic means is slowly and no heat comes into or out of the system.

If you move the piston adiabatically, then the change in the energy of the gas is, with a minus sign, the work done on the piston.

$$dE = -PA\ dx \tag{21}$$

P A is the force exerted on the piston, that is pressure times the area of the piston.

It is the elementary definition of pressure : force per unit area.

Then, multiplying that force by *dx*, we obtain the work done by that force.

With a minus sign, it is the change of energy *dE* in the cylinder chamber.

Now Adx is the change in the volume of the gas.

So we write Adx as dV , and equation (21) rewrites

$$dE = -PdV \tag{22}$$

We can use this equation to define pressure.

We then write

$$\frac{\partial E}{\partial V} = -P \tag{23}$$

Remember that Newtonian mechanics makes much use of the notions of force and work.

But then a more modern and more powerful presentation of mechanics, closer to Lagrangian mechanics, doesn't use force and work so much — it focuses on energy.

That is what we do here: pressure is no longer defined as force per unit area but as the rate of change of energy with respect to volume, with a minus sign.

Of course the two definitions are totally equivalent.

The partial derivative notation is here to remind us that other variables are kept fixed.

The only question is what do we keep fixed in considering this derivative.

The derivative is the rate of change of E with respect to V under the circumstance that we do the operation adiabatically.

There is another aspect of the notion of adiabatic change.

It has to do with the second law of thermodynamics, which we will come to.

The second law of thermodynamic says that the entropy of an isolated system in evolution always increases – except when it doesn't, in which case it stays the same.

But it never decreases.

Well, never say never, but never mind.

What are the processes where the entropy of an isolated system doesn't increase ?

Those are the adiabatic ones.

This requires some explanation.

We shall explain it from a quantum mechanical point of view.

The only aspect we are going to use of quantum mechanics is that energy levels are discrete, nothing else.

In other words, the E_i's of the possible states of the system are definite energy levels in a discrete collection.

So we have our box of gas shown in figure 5, and it has a collection of energy levels.

Let's plot them, figure 6.

The E_i's are not a single particle energy levels, they are the possible energies of everything inside the box.

There is some discrete collection of numbers, in quantum mechanics in any case.

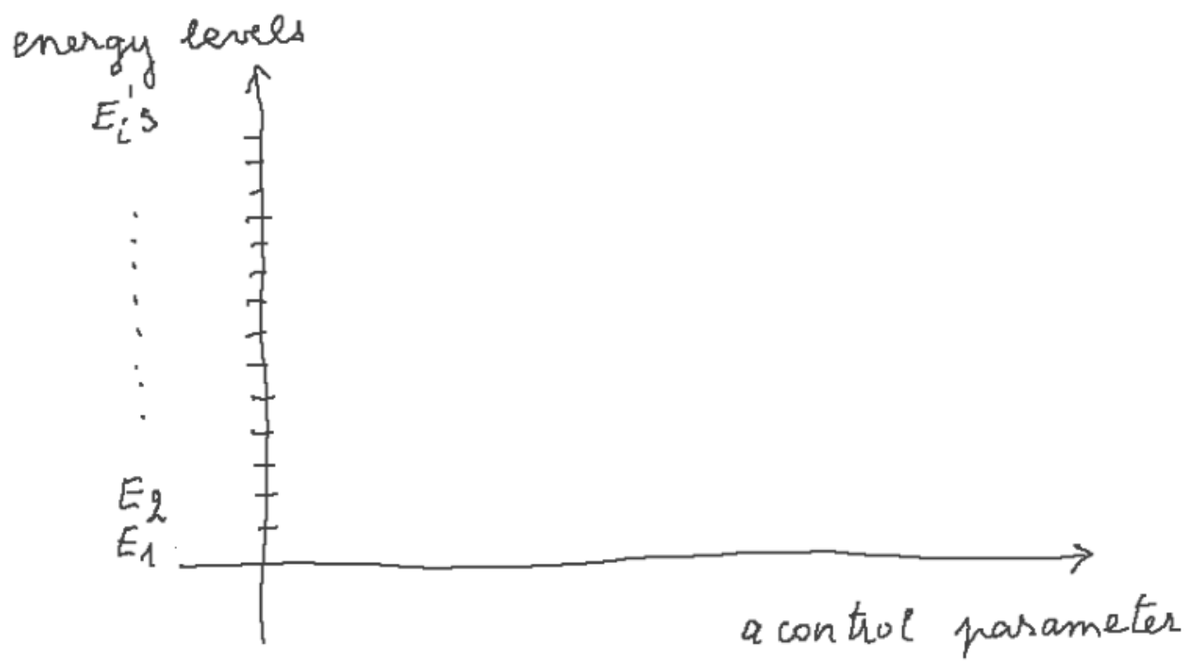


Figure 6 : Energy levels of a box of gas.

Now we are going to change the volume.

What happens to the energy levels ?

Do they stay the same ? No. Why should they ?

If we want to calculate the energy levels, we have to solve some problem, a Schrödinger equation or something like that.

The Schrödinger equation will depend on the size of the box.

And the energy levels will change, figure 7.

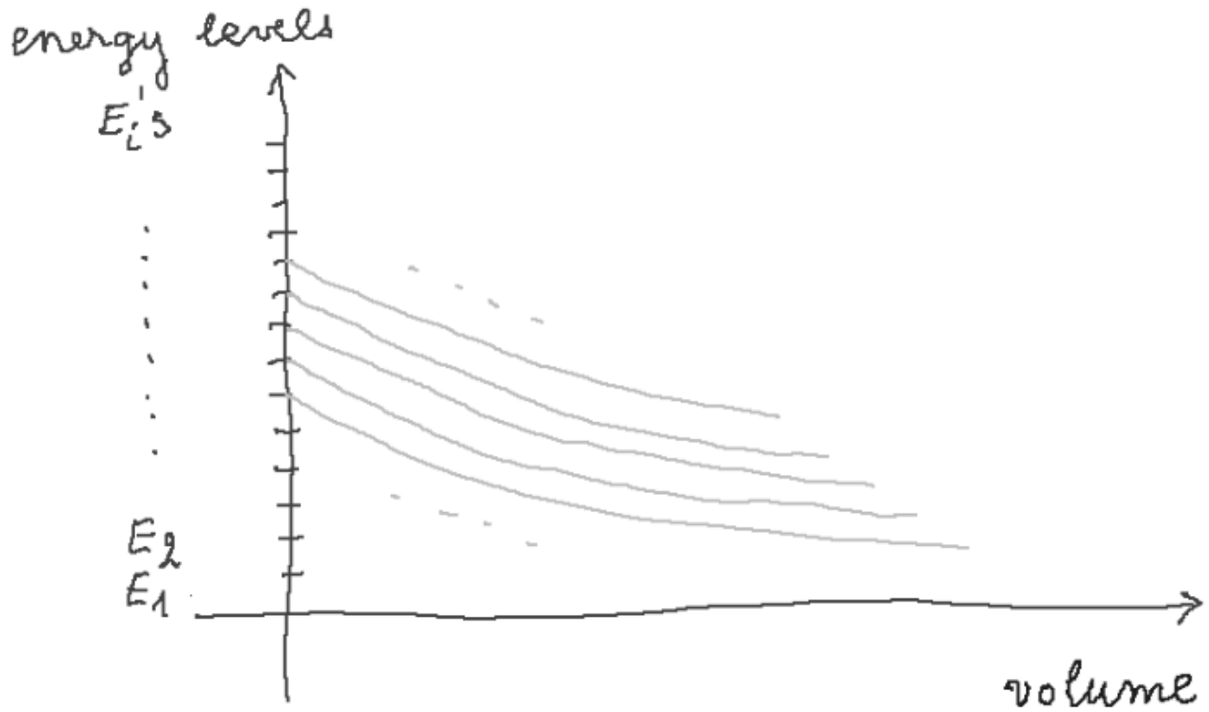


Figure 7 : Energy levels as a function of the volume of the box.

On the other hand, if you change the system rapidly this is not what happens.

Then you can have jumps from one energy level to another.

That is what is special about adiabatic processes.

The energy levels are remembered so to speak.

That has a beautiful mathematical consequence.

If we started with a collection of probabilities P_i's over the states of the system.

As we change V adiabatically, if we imagine the collection of energy states laid out on a vertical line which we shift to the right, that collection will slide along that vertical line.

It is not a simple translation, the states may get closer to each other.

But they will stay distinct and in the same order.

The crux of the argument then is that the probability distribution over the states i's – which represents our incomplete knowledge of the actual state of the system – won't change.

The energy levels may change, but the probabilities won't change.

If you had probability p to be at energy level E_{537}, that probability will stay the same.

If $P = 1$ then at any time you will be sure to be at that same energy level.

If you had probability 1/3 to be at or above energy level E_{129}, it will remain 1/3.

The probability distribution doesn't care what the values of the energy levels are.

Now the entropy is just built from the P_i's.

The entropy of a probability distribution is $S = -\sum_i P_i \log P_i$. .

Therefore if, as you change some control parameter, the probabilities stay the same, then you can be sure that the entropy stays the same too.

This is in fact the general definition of an adiabatic process :

An adiabatic process is a process in which the entropy doesn't change.

For that reason an adiabatic process is also sometimes called an *isentropic process*.

We see that defining an adiabatic process simply as a slow one would not be good enough.

If we were allowed to put heat into the system, that would corrupt it.

That would change the energy and other variables via an outside intervention, and that is not what we want to do.

In summary, adiabatic means slow with no added or subtracted heat.

And under those circumstances adiabatic also means constant entropy.

So now we know what to hold fixed in formula (23).

It is the change in the energy with respect to volume holding the entropy fixed.

$$\left.\frac{\partial E}{\partial V}\right|_S = -P \tag{24}$$

That is the reason we spent some time on the lemma, equation (20).

Formula (24) is perfectly right.

But it is hard to calculate things at fixed entropy.

What we are usually provided with, the easy quantities to calculate, are quantities at fixed temperature.

We stick in the temperature into the Boltzmann distribution and we calculate.

And what we get out is an answer which depends on the temperature.

Remember that at thermal equilibrium, the variables E, S, Z, β, T, and A, are all dependent on each other and any one of them can be chosen as the independent variable to express the others.

The volume V on the other hand is a control parameter.

What is easy to calculate usually is how things vary with temperature, simply because temperature directly appears in the Boltzmann distribution

$$P_i = \frac{1}{Z}\, e^{-\beta E_i} \tag{25}$$

So it is always easier to work with quantities as functions of temperature than as a functions of anything else.

Therefore what we would like to do is to convert equation (24) to something that involves derivatives with respect to temperature instead of entropy.

And then we can go back and see what we can learn from the Boltzmann distribution.

From equations (24) and (20), we can write

$$\left.\frac{\partial E}{\partial V}\right|_S = -P = \left.\frac{\partial E}{\partial V}\right|_T - \left.\frac{\partial S}{\partial V}\right|_T \left.\frac{\partial E}{\partial S}\right|_V \tag{26}$$

Now this looks terrible.

It looks too complicated.

Fortunately the last partial derivative on the right-hand side is something we have seen before.

Do you remember what the rate of variation of the energy with respect to the entropy is ?

It is the temperature.

When we hold the volume fixed, we hold the system fixed.

We are not changing the control parameter.

Our system is just the original system and its fixed energy levels.

And we are holding those fixed.

Under these circumstances $\partial E/\partial S$ is just the temperature, see equation (16) of part 2.

So equation (26) is not quite so bad.

It rewrites

$$P = -\left.\frac{\partial E}{\partial V}\right|_T + \left.\frac{\partial S}{\partial V}\right|_T T \tag{27}$$

Now the derivatives which remain are only derivatives with respect to the volume at fixed temperature.

That sounds promising.

Volume and temperature can be thought of as two independent variables.

And the derivatives that appear in equation (27) are sort of familiar rates of variation: how the energy changes with respect to volume at fixed temperature ; and how the entropy changes with respect to volume at fixed temperature.

And T is just temperature.

One more step : since the right-hand side of equation (27) involves only partial derivatives with respect to *V* at fixed temperature, as well as temperature itself, we readily see that it can be written

$$P \;=\; -\frac{\partial}{\partial V}\,\left(\,E - TS\,\right)\,\Big|_{T} \tag{28}$$

Indeed a priori the partial derivative with respect to *V* of the product *T S* is a sum of two terms: the partial derivative of *S* with respect to *V*, times *T*, plus the partial derivative of *T* with respect to *V*, times *S*, all at fixed temperature.

But what is the derivative of *T* with respect to whatever variable, at fixed temperature?

Zero.

So we can write

$$\frac{\partial}{\partial V}\,\left(\,TS\,\right)\,\Big|_{T} \;=\; T\,\frac{\partial S}{\partial V}\,\Big|_{T} \tag{29}$$

Whence, equation (28).

We have reached a fundamental theorem :

The pressure is equal to minus the derivative with respect to the volume of the Helmholtz free energy at fixed temperature.

$$P = -\frac{\partial A}{\partial V}\bigg|_T \tag{30}$$

This is much simpler than anything else we could imagine for the pressure.

In fact we even know a little more.

From equation (9), we know that the Helmholtz free energy is equal to $-T \log Z$.

So we can also write

$$P = \frac{\partial \left(T \log Z \right)}{\partial V}\bigg|_T \tag{31}$$

Since this is evaluated at a fixed temperature, we can pull T outside the partial derivative

$$P = T \frac{\partial \log Z}{\partial V}\bigg|_T \tag{32}$$

In other words, all we need to be able to calculate is how the partition function depends on the volume, at fixed temperature.

Then we can calculate the pressure.

Everything we have done up to now is completely general.

It doesn't matter if we are dealing with a liquid, a gas, a plasma or a solid.

It could be a solid inside the cylinder!

It doesn't matter what the chemical composition is either.

This is what statistical mechanics is, as opposed to the kinetic theory.

Kinetic theory is the intuitive picture of molecules bouncing around.

Whereas equation (32) is what statistical mechanics says: pressure is equal to temperature times the derivative of the logarithm of the partition function with respect to the volume, at fixed temperature.

Now we can carry out the same kind of derivation for basically any control parameter.

We used the volume as the control parameter simply because it is familiar.

For any control parameter there is always a derivative of energy with respect to the control parameter at *fixed entropy*.

That is called the *conjugate thermodynamic variable* of the control parameter.

Above, using the lemma, and doing some calculus manipulations, we then reached the pressure expressed as T times the derivative of log Z with respect to V at fixed temperature.

But let's emphasize that at the start what we considered is the derivative of energy with respect to volume at fixed entropy, with a minus sign, equation (24).

And that gave us the pressure.

Volume and pressure are conjugate variables thermodynamically.

There are many others, which we are not going to go into now.

But anytime we have a control parameter, then we can ask: How does the energy of the system depend on the control parameter at fixed entropy ?

And that defines another variable which is like pressure plays the role of pressure somehow.

One last point before we do some questions/answers.

Notice that in equation (32) P depends on log Z.

$$P = T \left. \frac{\partial \log Z}{\partial V} \right|_T$$

As we said earlier Z, and more specifically $\log Z$, appears all over the place in statistical mechanics.

Do you remember another important quantity which depends on $\log Z$?

The average energy.

Remember equation (42c) of part 4

$$E = -\frac{\partial \log Z}{\partial \beta} \tag{33}$$

The function $\log Z$ has a lot of stuff in it.

If you know $\log Z$ you know a lot of stuff.

5.5 Questions/answers

Question: It seems that the condition that the process we study be adiabatic is pervasive in all of thermodynamics.

What happens when the process is not adiabatic ?

Answer: Then it becomes a complicated non-equilibrium process.

I gave you an example.

If you pull the piston fast enough, no molecule will hit the wall.

There won't be any pressure.

What will happen is then, if you make a very sudden change ?

How will the system respond?

It will respond by a shockwave going through the fluid.

You can change the volume in either direction.

It is easier to think about it if you suddenly squeeze the fluid.

You suddenly push the piston and you make a shockwave going through the system.

You will also make a shockwave if you pull it away.

And very complicated things happen.

It is no longer controlled by simple statistical mechanics.

It is controlled by the non-equilibrium dynamics.

The only thing you can be sure of is that the entropy is going to increase.

Now you can actually work out the special case when you pull the piston so fast that no single molecule hits the wall.

Then you know that there is no work done on the piston.

And the change in the energy of the gas is absolutely zero

Of course, we have to imagine the whole system plunged into a void, so that it is isolated, and there is no external pressure from outside exerted on the piston.

If the change in energy is zero but the volume has changed, you can learn some things about that situation.

You can do some calculations.

You can figure out what the final temperature is for instance.

It is solvable.

But if you were to squeeze the gas suddenly then all kinds of very complicated things would happen.

Question: In the process just described of pulling the piston very suddenly vertically, as in figure 5, the process is not in equilibrium.

But it will return to equilibrium.

The total energy is conserved.

Some kinetic energy will have been transformed into potential energy as the molecules fill up the new available space.

Answer: Yes. But the question is: What is the new temperature of the system when it has returned to equilibrium ?

That can be a very complicated problem.

Generally it is not easy.

Sometimes it is easy.

Question: Concerning the lemma just proved, E and S are the dependent variables, and T and V are the independent variables.

What does that mean to look at the variation of the independent variable T with respect to the other independent variable V?

Answer: Well, when we wrote equation (14), reproduced below,

$$\left.\frac{\partial E}{\partial T}\right|_V \frac{\Delta T}{\Delta V} = \left.\frac{\partial S}{\partial T}\right|_V \left.\frac{\partial E}{\partial S}\right|_V \frac{\Delta T}{\Delta V} \tag{34}$$

that was a pure identity, like writing

$$\frac{a}{b} = \frac{a}{c}\,\frac{c}{b} \tag{35}$$

Equation (34) is true for any pair of displacements ΔT and ΔV of the independent variables.

But then we looked at the ratio $\Delta T/\Delta V$, given that we are moving along a line of constant entropy.

We calculated $\Delta T/\Delta V$ along such a line, imposing $dS = 0$.

We first wrote another identity that is always true

$$\Delta S = \left.\frac{\partial S}{\partial V}\right|_T \Delta V + \left.\frac{\partial S}{\partial T}\right|_V \Delta T \tag{36}$$

Then we imposed $\Delta S = 0$, or if you prefer $dS = 0$. That gave us an expression for $\Delta T / \Delta V$, because on an isentropy line, this ratio is no longer free.

It has a value depending on the point where we are in the plane V, T, figure 4.

It is the slope of the isentropy line at the point.

You can interpret it as a constraint in your calculations if you like.

Question: In an adiabatic process, when we say slowly, is there a physical parameter that has to be lesser or greater than some value?

In other words, could we quantify what we mean by slowly ?

Answer: Of course. But that will depend on the details.

Roughly speaking it means that many molecules are colliding with the piston, so that we can average over the details of the molecular collisions.

We could examine specific cases.

Some questions have very general answers.

And some questions depend on the details.

The question of exactly what time scales constitute slow could be one that depends on the details.

5.6 Ideal gas

In the introduction we said that for the ideal gas we can use kinetic theory, plus the assumption of isotropy everywhere including near the walls, to calculate pressure from the motion of the molecules.

On the other hand, the question of what happens if you move sufficiently slowly does not depend on details.

With our new more general theory, let's go back to the ideal gas.

Now we start from

$$P = T \left. \frac{\partial \log Z}{\partial V} \right|_T \tag{37}$$

When we work with the partition function, of course we fix the temperature.

We say what β is.

Typically when we calculate the partition function, it is Z as a function of other variables at fixed temperature.

We work it out at fixed temperature.

These other variables are not E, or S or A, which are also deterministically linked to the temperature.

The other variables that can still vary are control parameters.

So we turn to the partition function for the ideal gas.

This is going to be very easy from this point on.

Figure 8 shows a box filled with an ideal gas.

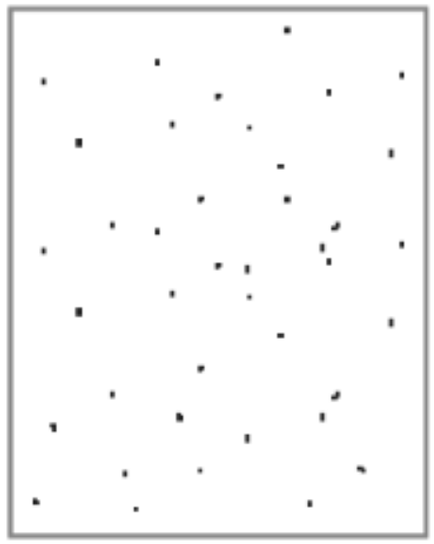


Figure 8 : Ideal gas.

We saw that the partition function is the following integral

$$Z = \int dx\ dp\ e^{-\beta \sum_n \frac{p_n^2}{2M}} \tag{38}$$

We are not going to write it in great detail.

Let's just recall that it is the integral over all the x's and all the p's of e to the minus β times the value of the energy at (x, p), which is

$$\sum_n p_n^2/2M$$

It depends on p, but not on x.

And the parameter M is the mass of one particle.

What is the actual dimensionality of dx ? Answer : $3N$.

And of dp? Also $3N$.

Where N is the number of particles in the ideal gas.

The integral defining Z, in equation (38), factorizes into a product of identical integrals for each molecules. Its logarithm is therefore proportional to N.

The x integration gives us the volume to the power N.

If there was only one particle, it would indeed just be the volume.

But the element, which we simply wrote dx, is N times a differential volume.

So we can write

$$Z = V^N \int dp\ e^{-\beta \sum_n \frac{p_n^2}{2M}} \tag{39}$$

The p integration depends on β, but doesn't depend on the volume.

So whatever it is, we can write

$$Z = V^N f(\beta) \tag{40}$$

Remember that we calculated it - see equation (55) in part 4.

Finally, we may or may not stick in the factor $N!$.

$$Z = \frac{V^N}{N!} f(\beta) \tag{41}$$

That doesn't make any difference. Why?

Because we are going to take the logarithm of Z and then differentiate with respect to the volume.

$$\log Z = N \log V + \log f(\beta) - \log N! \tag{42}$$

When we differentiate with respect to V , on the right-hand side only the first term matters.

The second term and the third term don't depend on V .

So we have

$$\frac{\partial \log Z}{\partial V} = \frac{N}{V} \tag{43}$$

Now we multiply that by the temperature, equation (37), and that gives us the pressure.

$$T\frac{\partial \log Z}{\partial V} = \frac{NT}{V} = P \tag{44}$$

The last equality is a version of the famous Boyle–Mariotte law already known in the 17th century.

$$PV = NT \tag{45}$$

Of course, if we work in laboratory units, that is, if we work in degrees Kelvin, T is replaced by kT, where k is the Boltzmann constant.

Notice that N/V is the density ϱ. So equation (45) can also be written

$$\boxed{P = \rho T} \tag{46}$$

This particularly simple form is the *equation of state of an ideal gas*.

This is exactly what we obtained when we used the intuitive picture of molecules bouncing off the wall – that is the kinetic theory of gas, with the assumption of isotropy everywhere, even near the walls.

But here we don't need that extra assumption.

It is implicitly justified

Indeed, in the kinetic theory, with a different assumption for the behavior of the molecules near the walls, we would not get equation (46).

And now we have proved that this different assumption would be mistaken.!

Moreover, there is power in this method.

Any time we can do the integrals necessary to calculate the partition function, we have a precise way then of calculating the derivative of its logarithm with respect to a control parameter.

So this was just the simplest of many examples where people can calculate the formula for a conjugate thermodynamical variable.

Here the pressure P , with a minus sign, is the conjugate thermodynamical variable of the volume V .

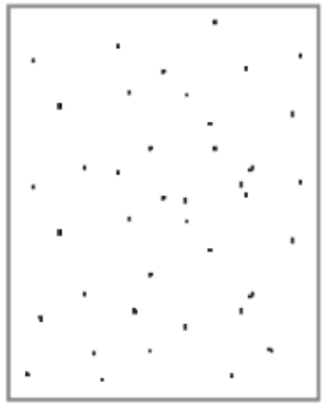
That is the meaning of equation (24). And the formula for P is as simple as it can be: $P = NT/V$.

We did this example in detail in order to see how the basic concepts fit together.

5.7 Fluctuations

What is a fluctuation?

If we look at a box of gas, like in figure 8, and it is in equilibrium with another system, for instance a heat bath, so that it is exchanging energy with the outside, the energy in the box won't be a definite value.



What we have called E so far is the *average energy*.

But the energy itself is a random variable.

It will fluctuate over the energy levels E_i's.

The fact that the energy fluctuates in certain ways is evidence that the system is really made up of molecules and so forth.

But we need a definition of the fluctuation of a quantity when it has a probability distribution, that is when the quantity is a random variable.

This part of our discussion is pure elementary probabilities.

It includes added parts to help you understand how probabilities work.

In the next part we shall get back to physics, apply the calculation of fluctuation to the energy, relate it to specific heat, etc.

Remember that to talk about probabilities, we need to have a well-specified random experiment .

Here the experiment is "wait until the box is in thermal equilibrium on its own and with the heat bath surrounding it", then, at a random time, and supposing that we have the adequate instruments, look at the *exact energy* of the system.

$\mathcal{E}$

It is one of the E_i's, and there is a probability distribution, which is the Boltzmann distribution.

For any probability distribution, here is the general definition of fluctuation.

Let's begin with a random quantity whose average is zero, figure 9.

We mix with the same picture the concepts of probability distribution and probability density, which will be clear as we proceed.

For a random variable to have mean zero doesn't mean that the picture of its distribution is symmetric around zero.

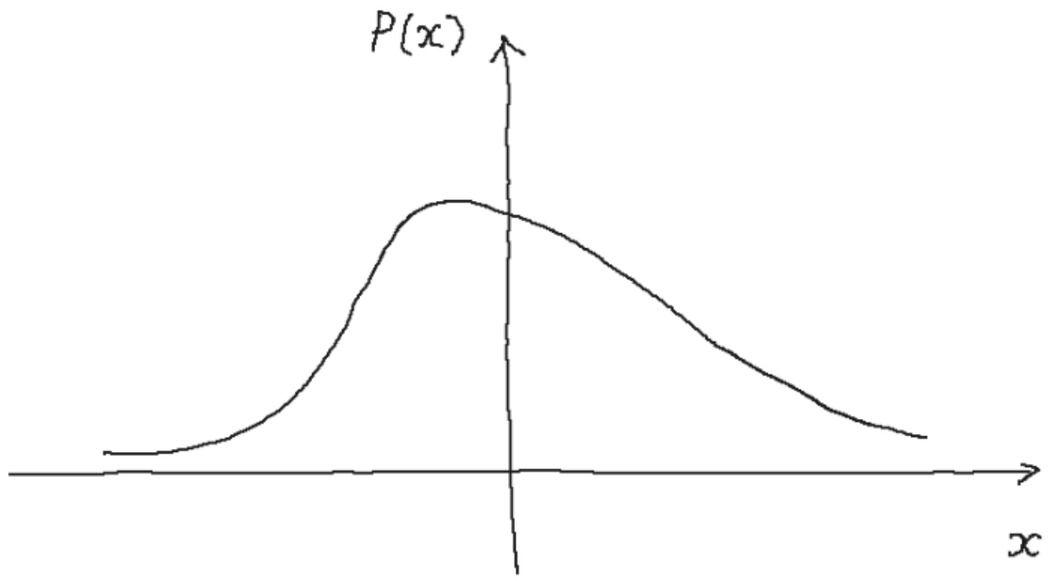


Figure 9 : Probability distribution of a random variable with mean zero.

It means that $\sum_i x_i P_i = 0$, if we are in the discrete case, or $\int xP(x)dx = 0$, if we are in the continuous case.

Fluctuation is a precise definition and measure of the intuitive concept of width of the probability distribution around its mean – how broad the distribution is.

For the sake of illustration, let's call the random quantity X, and let's denote its mean – another name for average – $< X >$.

This average is not a random variable, it is a number.

And we decide to talk first of all about a random variable whose average is zero.

$$< X > = 0 \tag{47}$$

Then its fluctuation is defined as

$$\Delta X = \sqrt{< X^2 >} \tag{48}$$

It is the square root of the average of the random variable X^2.

Probabilists call the average of X its first moment, the average of X^2 its second moment, the average of X^3 its third moment, etc.

It turns out that knowing the entire collection of moments of a random variable is *equivalent* to knowing its whole distribution of probability.

Knowing its first two moments gives a first approximate view of its distribution, namely where is its average, and what is its width, figure 9.

In fact for Gaussian distributions, the knowledge of $< X >$ and $< X^2 >$ determines completely the distribution.

We just defined the fluctuation of a random quantity X when its mean was zero.

Now let's gives the definition in the general case of a variable Y whose mean $< Y >$ can be any value, not necessarily zero.

We will shift the variable Y so that its new mean is zero, and then apply the first definition.

So we define the new variable

$$X = Y - < Y > \tag{49}$$

It is easy to see again that $<X>= 0$.

Then the fluctuation of Y is by definition the fluctuation of its "centered" version, that is the fluctuation of X.

$$\Delta Y = \sqrt{< X^2 >} \tag{50}$$

We can also write it

$$(\Delta Y)^2 = < X^2 > \tag{51}$$

Or even more elaborately, without reference to X,

$$(\Delta Y)^2 = < [\, Y - < Y > \,]^2 > \tag{52}$$

In probability theory the fluctuation ΔY of a random quantity Y is called its *standard deviation*.

It is often denoted σ_Y .

And the square of the fluctuation is called the *variance*.

We can now show:

$$\begin{aligned}(\Delta Y)^2 &= \langle [Y - \langle Y \rangle]^2 \rangle \\ &= \langle [Y^2 - 2Y\langle Y \rangle + (\langle Y \rangle)^2] \rangle \\ &= \langle Y^2 \rangle - 2(\langle Y \rangle)^2 + (\langle Y \rangle)^2 \\ &= \langle Y^2 \rangle - (\langle Y \rangle)^2 \end{aligned} \tag{53}$$

In summary, the square of the fluctuation of a random quantity is by definition the average of the square of its departure from its mean, equation (52).

And we have the simple result above saying that it can be calculated as the average of the square minus the square of the average, equation (53).

Now let's see if we can calculate the fluctuation of the energy.

5.8 Fluctuation of energy and specific heat

Having made a detour in elementary probability theory to define fluctuation, now we go back to physics. We want to calculate the fluctuation of the energy.

We will calculate it in terms of the partition function.

Generally speaking, in our problems in statistical mechanics, we are armed with one big tool: the partition function.

One always goes back to it.

It is the Swiss army knife of statistical mechanics.

In an honest real experiment, if we repeat it over and over again, and each time we are able to measure the exact energy inside the box, what fluctuation shall we observe ?

Why does it vary ?

It varies because there is energy coming in and out of the system.

It is not just a mathematical definition.

The energy in our box does fluctuate if it is in equilibrium with a heat bath.

We want to calculate the fluctuation of energy.

So we shall apply equation (53) to energy.

$$(\Delta E)^2 = < E^2 > - (< E >)^2 \tag{54}$$

Important caveat: in the earlier parts of our discussion, E was the average energy.

Now, in the above formula, E is the random variable, which can take values E_i's if we use a discrete model, or a continuum of values of we use a continuous model.

And its average is $< E >$.

What is $< E >$ in terms of the partition function ?

We go back to formula (42c) in part 4. It is

$$< E > = -\frac{\partial \log Z}{\partial \beta} \tag{55}$$

Thus we have the average energy already.

We don't have to do any more work on that term in equation (54).

But what about the average of the square of the energy?

To figure it out, let's remember the trick used to calculate the average energy, and see if we can use it again somehow.

To calculate $< E >$, we wrote

$$< E > = \sum_i \frac{1}{Z} e^{-\beta E_i} E_i \tag{56}$$

Then we noticed that the last factor E_i could be produced by taking the derivative of the exponential with respect to β.

So we wrote

$$
\begin{aligned}
< E > &= \sum_i \frac{1}{Z}\, e^{-\beta E_i}\; E_i \\
&= -\frac{1}{Z} \sum_i \frac{\partial}{\partial \beta}\, e^{-\beta E_i} \\
&= -\frac{1}{Z}\, \frac{\partial}{\partial \beta} \sum_i e^{-\beta E_i} \\
&= -\frac{1}{Z}\, \frac{\partial}{\partial \beta}\, Z
\end{aligned}
\tag{57}
$$

And the last equality became equation (55).

But suppose we want to calculate the average of the energy squared now.

What do we do ?

We want to calculate

$$
< E^2 > = \sum_i \frac{1}{Z}\, e^{-\beta E_i}\; E_i^2 \tag{58}
$$

We will differentiate twice.

$$
\begin{aligned}
< E^2 > &= \sum_i \frac{1}{Z}\, e^{-\beta E_i}\; E_i^2 \\
&= \frac{1}{Z} \sum_i \frac{\partial^2}{\partial \beta^2}\, e^{-\beta E_i} \\
&= \frac{1}{Z}\, \frac{\partial^2}{\partial \beta^2} \sum_i e^{-\beta E_i} \\
&= \frac{1}{Z}\, \frac{\partial^2}{\partial \beta^2}\, Z
\end{aligned}
\tag{59}
$$

Finally we can write the square of the fluctuation in the energy as follows

$$
(\Delta E)^2 = \frac{1}{Z}\, \frac{\partial^2 Z}{\partial \beta^2} - \frac{1}{Z^2} \left(\frac{\partial Z}{\partial \beta} \right)^2 \tag{60}
$$

Let's see if we can do anything with it.

But, first of all, observe that it tells us that we can calculate the fluctuation in the energy if we know Z.

Again that shows the power of the partition function.

All of statistical mechanics is about the power of the partition function, and the power of differentiation.

To try to improve expression (60), let's take a look at

$$\frac{\partial}{\partial\beta}\left(\frac{1}{Z}\frac{\partial Z}{\partial\beta}\right) \tag{61}$$

This is the derivative of the product $1/Z$ times $\partial Z/\partial\beta$.

So, applying the product rule, and noting that the derivative of $1/Z$ is $-1/Z^2$ times the derivative of Z, expression (61) can be rewritten

$$\frac{1}{Z}\frac{\partial Z^2}{\partial\beta^2} - \frac{1}{Z^2}\left(\frac{\partial Z}{\partial\beta}\right)^2 \tag{62}$$

This is exactly the right-hand side of equation (60).

So we can rewrite the square of the fluctuation of energy as

$$(\Delta E)^2 = \frac{\partial}{\partial\beta}\left(\frac{1}{Z}\frac{\partial Z}{\partial\beta}\right) \tag{63}$$

We could write the derivative of a log inside the parentheses.

But it is even simpler : inside the parentheses is minus the average energy.

So we reach

$$(\Delta E)^2 = -\frac{\partial}{\partial \beta} < E > \tag{64}$$

And what is the right-hand side of this equation ?

It is, with a minus sign, the rate of change of the average energy with respect to β.

Remember that β is the inverse of the temperature.

Massaging equation (64) a bit, we get

$$(\Delta E)^2 = -\left(\frac{\partial}{\partial T} < E > \right) \frac{\partial T}{\partial \beta} \tag{65}$$

The last derivative is

$$\begin{aligned} \frac{\partial T}{\partial \beta} &= \frac{\partial}{\partial \beta} \frac{1}{\beta} \\ &= -\frac{1}{\beta^2} \end{aligned} \tag{66}$$

Therefore equation (65) is simply

$$(\Delta E)^2 = T^2 \frac{\partial}{\partial T} < E > \tag{67}$$

What is the rate of change of the average energy of the system with respect to temperature?

It is called the *heat capacity* of the system.

How much energy you have to put into a bucket of water to increase its temperature by one degree?

That is the heat capacity of your bucket of water.

And if you look at one unit of water – meaning one kilogram – then it is called the *specific heat*.

Notice that, when writing formula (67), we keep the volume fixed.

We are not varying any control parameter.

We are not changing the volume.

On the right-hand side, next to T^2, we just look at the relationship between $< E >$ and T, and more precisely at the rate of change of the average energy with respect to the temperature.

The heat capacity at constant volume V is usually denoted C_{V}.

It depends on V.

It also depends on the number of particles in the box in figure 8.

But all these are fixed control parameters in our experiment.

We just look at the relationship with the temperature.

And so, for the square of the fluctuation in energy we finally reach the formula

$$(\Delta E)^2 \;=\; C_V \; T^2 \tag{68}$$

where C_{V} is the rate of change of the average energy with respect to temperature.

It is a famous formula, due to Gibbs and to Einstein.

So, if you know the heat capacity C_{V} of your system – which can be calculated as the specific heat times the mass –, you multiply it by the temperature squared and that will tell you how much fluctuation in the exact random energy there is.

Now we left out something : the Boltzmann constant.

If we work with laboratory units, in the definition of C_V as $\partial < E > /\partial T$, there will be a k_B downstairs, and from T^2 there will be ${k_B}^2$ upstairs, so formula (68) in laboratory units rewrites

$$(\Delta E)^2 \;=\; k_B \; C_V \; T^2 \tag{69}$$

That means it is small.

We expect fluctuations when there is a lot of particles.

But they will be fairly small.

Formula (68) or (69) is the fluctuation in the energy of any system.

It is not specific to a particular kind of system.

Inside the box, again, it can be gas, ideal or not, liquid, solid, etc.

Notice that C_V doesn't have to be constant when temperature changes.

So it would be more accurate to write equation (69) as

$$(\Delta E)^2 \;=\; k_B \; C_V(T) \; T^2 \tag{70}$$

The fluctuation depends on T not only because there is a T^2 in the formula, but also because C_V may depend on T.

A final comment : This is again part of the beautiful structure of statistical mechanics.

Formula (70) goes beyond classical thermodynamics.

Whenever there is a k_B appearing in the formula like here, it really does depend on the fact that systems are made up out of molecules and so forth, that is, not simply replacing T by kT.

It depends on Avogadro's number.

Boltzmann's constant k_B is closely related to Avogadro's number.

5.9 Questions/answers

Question: If we set pressure constant instead of volume, and therefore work is performed when the temperature changes, would formula (68) become different ?

Answer: Well, formula (68), or its variant, was the derivation of the fluctuation in the energy when all the control parameters are kept fixed, and the actual energy of the system ranges randomly over the E_i's, which themselves are also fixed energy levels.

We amply described the fact that, due to microexchanges with the heat bath, the actual energy inside the box fluctuates.

This fluctuation – another word for the standard deviation of the actual energy – happens to depend on the temperature T and on the heat capacity $C_{\mathrm{V}}(T)$.

One can define the coefficient C_{P} as

$$C_P = \left.\frac{\partial E}{\partial T}\right|_P \tag{71}$$

But if you use it instead of C_{V} in equation (68), you will get the wrong answer.

If you want to study the relationship between ΔE and C_{P}, you first of all have to describe in details what you mean.

What random experiment?

The fluctuation of which random variable precisely ? etc……

No doubt there are connections between the various coefficients of heat capacity that can be defined under various assumptions.

You will have to look them up, or figure them out by yourself.

Question: Isn't the specific heat a constant for a given material ?

Answer: Well, first of all, we don't use the specific heat in equation (68), but the heat capacity C_V of the system.

It is of course related to the specific heat.

What is the definition of specific heat?

It is the rate of change of energy per unit temperature per unit mass?

In other words, it is the heat capacity per unit mass of the material.

For instance, for water under ordinary conditions of temperature and pressure, it is 1 Calorie / K x kg (1 Calorie = 1000 calories = 4184 Joules).

Therefore the coefficient C_V in equation (68) is the specific heat times the mass of the sample.

Roughly speaking the heat capacity is proportional to the number of molecules in a sample.

Finally the specific heat of a given material is more or less constant over a sizable range of temperatures, but it may vary when the temperature gets very low or very high.

The important thing to note is that formula (68) and its variants don't depend on the details of what things are made out of.

Once you have measured the specific heat, you know the fluctuations without having to worry about what kind of material it is or anything else.

5.10 Ideal models in physics, linearity, and measurements

Last question was referring to a thermodynamical variable – the specific heat – that is presumably constant when temperature changes.

A consequence would be that heat capacity depends only on mass, linked to it by a constant proportionality factor.

Moreover for a gas at fixed density ϱ, by definition of density the mass m is proportional to the volume V .

So heat capacity in that case would be proportional to V.

If the gas is ideal, we saw that the following relation holds

$$P = \rho T \tag{72}$$

Then at fixed volume and mass, we can use the pressure to measure the temperature.

In other words, we could transform a pressure gauge into a thermometer.

And since at fixed volume and density the variation in heat would be proportional to the variation in temperature, it would also be an instrument measuring the quantity of heat transferred inside.

The general point we want to address here is that all these observations are based on linear relationships.

In physics we often assume linear relationships between variables, at least over a limited range of the independent variable.

The case of the ideal gas is an example.

It is ideal if its energy is only the kinetic energy of the molecules, if they don't collide with each other, don't interact, don't have potential energy like the Earth and the Sun, etc.

So we assume that it has low density, that the particles are small, etc.

This is often a fair assumption.

And for many calculations on gases the ideal gas approximation is a good approximation.

As a matter of fact, the air in the room is far from an ideal gas.

If we remember that the molecules of nitrogen move at a speed of approximately 500 meter per second – and that is before the rescaling – we are not surprised to know that there are plenty of collisions.

Yet for some calculations the ideal gas approximation is still satisfactory.

The notion of ideal device, or ideal relationship, or ideal conditions, is common in physics.

It refers to a model that we know is simpler than reality, but is useful when it is not too far off because then calculations become tractable.

Generally speaking, laboratory instruments are physical devices relying on a linear relationship between two variables.

Sometimes, like in the case of analog integrators, we rely on a more complicated relationship, but that is not the rule.

In truth however, as said, no physical device ever measures what it is intended to measure – the mathematical thing it is intended to measure.

At best what is usually true is that it measures what it is supposed to measure over a limited range of the independent variable which we want to measure by reading off the dependent one.

For instance we can use a spring to measure force, figure 10.

Over a limited range of forces, the elongation of the spring will be proportional to the force.

Therefore we will be able to measure the force by reading off the elongation.

But if we pull too hard, we will distend the spring, ruining the instrument

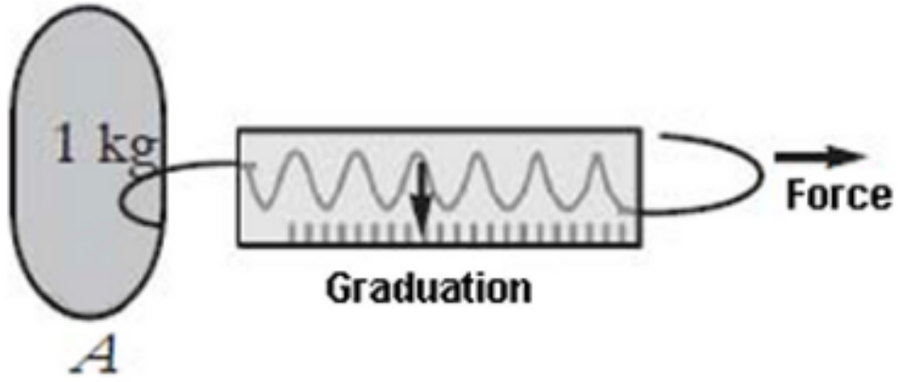


Figure 10 : Dynamometer. In this contraption, it is used to measure inertia, and check Newton's law F = ma.

So normally the situation is this : we have in the laboratory many quantities which we can measure easily.

It could be the length of a stretched spring, or it could be the height of the mercury in a thermometer, or the displacement of a needle on a dial in response to some input.

And we ask ourself : What could the quantity we want to measure depend on ?

Or, the other way around : What could the elongation of the spring depend on ?

It clearly depends on the force exerted on it, but it could depend on other things.

It could depend on the temperature.

It could depend on the air pressure.

But, really, that is not terribly important.

Those variations are pretty small so we ignore them.

And we say it depends on the force that is pulling the spring.

There is going to be some relationship between the force *F* and the elongation *X*.

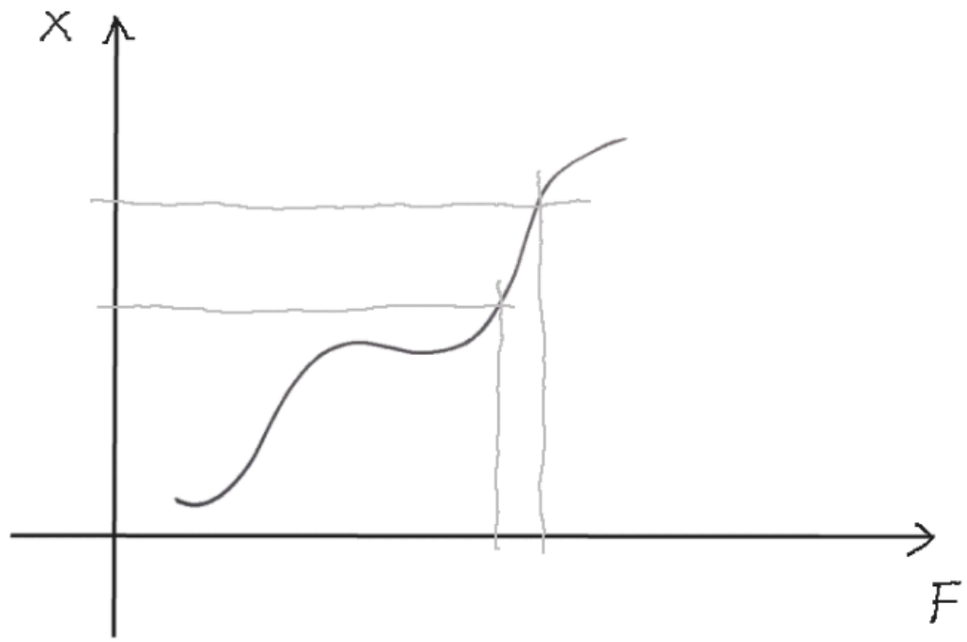


Figure 11 : Relationship between elongation X and force F.

It doesn't have to be a straight line.

It might be some crazy curve that looks like figure 11.

Over some limited range, if the function is smooth, it will be linear.

And, as you know, any function that is smooth enough, when looked at sufficiently locally is linear.

If we want to be more precise, it is like the arc of a circle.

And we can be even more precise.

That is the geometric interpretation of the Taylor-Maclaurin series expansion.

Anyway, for any device if over some range the relationship between the two variables is linear, then you can stay over that limited range, and use your device as a measuring instrument.

The relationship between the variable y you can read off, and whatever the independent variable x you want to measure happens to be, that relationship can be approximated by $y = ax + b$.

In the case of the spring balance it is just force and elongation.

For a thermometer it is the height of mercury in a column and the temperature, figure 12.

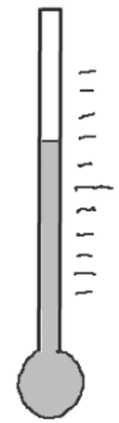


Figure 12 : Mercury thermometer.

Indeed what else can the height of mercury depend on ?

Can it depends on the air pressure in the room ?

Not much because the air pressure in the room is isolated from the mercury by the closed glass.

What else can it depend on ?

The local gravitational field ?

Yes a tiny bit.

Let’s formulate it in a different way : it can depend on the direction that we hold the thermometer.

Now for a common thermometer that won’t make any difference.

But we could imagine a thermometer that would behave differently if we held it horizontally than if we held it vertically.

That would not necessarily be a bad thermometer.

We would just add, for instance, the rule that the thermometer must be held vertically.

We would make it part of the definition of the thermometer.

What else?

The age of the thermometer?

Well, mercury and glass are pretty inert.

So we are in pretty good shape regarding that.

Of course over a very long time that is right.

The protons that make up the mercury could decay. But the order of magnitude of their half-life is 10^{35} years(????), so we are on the safe side here too.

We think about it for a while, and eventually we conclude : Gee, to a very very good approximation, the only thing that the height of the column is going to depend on is the temperature.

Or, equivalently, the energy that the mercury has absorbed from the thermal environment that it is in is proportional to the temperature.

That assumes of course that the thermometer is in thermal equilibrium with its environment.

Of course if the thermometer is not in equilibrium with the environment, it is not measuring the temperature of the environment.

It is only if enough time has been left for the environment to come to equilibrium with the mercury. Then the only variable around is the temperature.

At equilibrium the height of the column depends on the temperature in some way.

Suppose it is the curve in figure 13.

Suppose that A is the point of freezing water, B the point of boiling water, and between A and B the curve looks pretty straight.

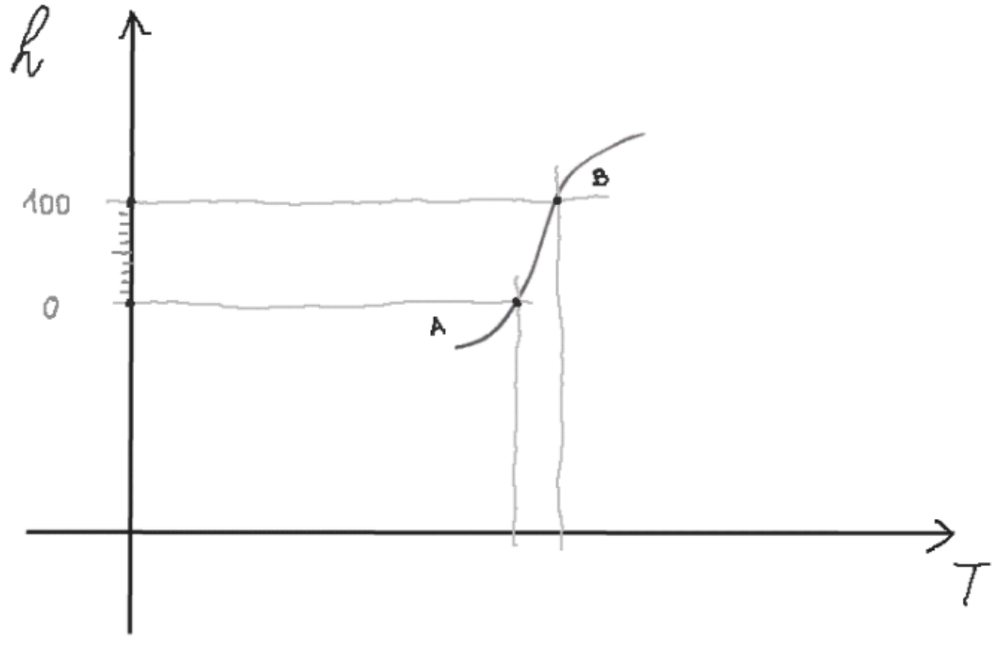


Figure 13 : Relationship between mercury height and temperature.

Remember that we already have a definition of temperature as the rate of variation of energy with respect to entropy, and that we know, at least in theory, the relationship between E and T .

So we can check that the curve is straight between A and B, and we can do this calibration.

Then we can arbitrarily call the ordinate of A zero, and the ordinate of B one-hundred.

We then divide the corresponding interval on the h-axis into a hundred little segments.

And we say that each segment is one degree – one rung, so to speak – on a centigrade scale.

Is the thermometer measuring temperature ?

Yes. It is measuring temperature to a high accuracy.

However it is not measuring temperature in the absolute sense.

We have called the point A temperature zero. It is not the absolute zero.

When you go out in the winter time and it says 0 degrees on your Celsius thermometer, it doesn't mean you are going out into absolute zero.

So there is a convention about where we put the zero.

But small incremental changes in the temperature will be proportional to small incremental changes in the height of the column.

5.11 Questions/answers

Question: When we use a thermometer to measure the temperature, we say that we need to wait until it is in thermal equilibrium with the environment.

When do we know that equilibrium has been reached ?

Answer: The answer is simple: when the height of the thermometer has stopped changing.

Depending on the heat conductivity of the glass that may take more or less time.

Typically, in ordinary conditions, a minute or so is sufficient

But that also depends on the density of the environment.

It is possible that the density be very low but the temperature be very high.

There is no reason why that can’t happen.

High temperature just means the kinetic energy of the molecules is high.

That doesn’t say anything about their number.

There may be very few molecules in the room.

If you bring the thermometer into the room, it is not going to equilibrate very rapidly because very few molecules are going to hit it.

Question: Can't there be a difference between the radiation temperature and the air temperature ?

Answer: Yes, there can be.

Usually there is.

The air and the radiations are usually not in equilibrium.

It is rather hard to bring radiation in the room in thermal equilibrium with the air.

The molecules of gases making up the air are neutral.

They don't interact much with photons.

They don't radiate much.

I can assure you that the optical radiation in the room is nowhere being in equilibrium with the air.

On the other hand there are very few photons in the room compared with the air molecules.

And so they really don't count very much.

Question: When does the ideal gas law break down?

Can it be calculated mathematically ?

Answer: Yes.

We shall now begin part 6 with such a calculation.

6. Weakly interacting gases, heat, and work

6.1 Introduction

Part 5 of our discussion ended on a question : when does the ideal gas law break down ?

Remember that for an ideal gas we assumed that the particles where point-like particles, not interacting with each other, not colliding into each other, with no potential energy, in other words no forces between them.

We are now going to study the case where there are forces between the molecules.

So let's consider a gas of particles that are weakly interacting.

Why weakly interacting ?

Because we have done the problem of the non interacting gas molecules.

And we want to take little steps.

Little steps are usually easy.

We start from a known solved problem, and change it a little bit.

More specifically, we take some of the parameters, change them a little bit and then try to expand the expressions in the small changes.

That is the usual trick.

We have some problem that we solved for some quantity ε equal to zero.

Then we reformulate the problem with ε not being zero and generally it is too hard to solve.

But if we are lucky we might be able to expand in a power series in ε.

Then we check if the succeeding orders are as big as the thing we started with, or if each one is an order of magnitude lower than the preceding one when ε is small.

That is the test to tell whether a series like that breaks down.

It breaks down when the second term is as big as the first term.

Of course, the gas we are going to study is weakly interacting because it is dilute, because the molecules are on the average rather far from each other.

We will assume the range of the forces is small by comparison with the distance between the particles.

We will assume that the potential energy between particles, or the forces, are small.

We will first do a calculation.

Then we will look at the results and ask : Where does this break down?

Where does the correction term become as large as the uncorrected term ?

We will check and find out.

It is going to depend on the density.

The bigger the density, the more likely molecules are going to crash into each other.

And if they crash into each other, they are going to experience the fact that they are interacting.

If we make the density too high, they are all jumbled in, squeezed on top of each other.

And the ideal gas is a very bad approximation.

On the other hand, we could keep the density low, but still, if the range of the forces is long, if particles were to interact with each other when they were six feet apart, that would also be a problem.

Considering for instance the molecules in this room, even if the interaction was pretty weak, if they interacted with molecules far away, that would still be a very serious modification of the ideal gas law.

So we start by assuming that we can make an expansion in the small parameters.

We will make the expansion – mathematically expand our formulas.

And then we will check how the first correction term compares with the term is it correcting.

Is it as big, or is it an order of magnitude smaller ?

Let's set up that problem.

As usual we will calculate the partition function.

Our system now is made of molecules whose total energy is not just the sum of the kinetic energies.

There are forces now.

Forces means of a potential energy.

We start with the formula for the energy of a set of molecules with fixed positions and momenta.

In other words it is one state of the system.

$$E(x,p) = \sum_n \frac{p_n^2}{2M} + \sum_{n>m} U(|x_n - x_m|) \tag{1}$$

How to read this equation :

- The notation $E(x,p)$ means the energy for a given set of x's and p's.
- The first sum on the right-hand side we already met in the ideal gas case. It is the kinetic energy of all the molecules.

$$E(x,p) = \sum_n \frac{p_n^2}{2M} + \sum_{n>m} U(|x_n - x_m|)$$

- M stands for the mass of one molecule.
- There are N molecules. In the first sum they are indexed by n. We haven't noted it explicitely, but n runs from 1 to N.
- The second sum is the potential energy of all the pairs of molecules.
- To track the pairs of molecules, we index the first molecule by n and the second by m.
- What is the meaning of $n > m$ under the sign sum ? We don't want to count n and n as a pair. And we don't want to count pairs twice. So the indexing of the pairs uses the two letters n and m and we sum over all the indices such that $n > m$.
- A pair of molecules has a potential energy U which depends on the distance between the two molecules.
- x_n is a vector. So is $x_n - x_m$. And the vertical bars mean that we take the length of $x_n - x_m$, in other words the distance between the n-th and the m-th molecules.
- Hence we use the notation $U(|x_n - x_m|)$ to mean the potential energy due to that pair.

Now what is going to be the small quantity?

It is going to be U.

The potential energy between molecules will be assumed to be a good deal smaller than their kinetic energies.

In other words, we start with the ideal gas.

And then we turn a knob which jacks up the potential energy of the configuration.

But we keep that knob set to a very small value, so that, at least at first, the potential energy stored by these pairs of molecules is much smaller than the kinetic energy of either of them.

That is our approximation, and where we are going to begin.

There is actually only one number regarding the potential energy which is going to be important.

It will appear in the calculation of the following integral concerning one pair of particles, when we will integrate over all their possible positions, say particle 1 and particle 2.

It doesn't matter which pair, we just call them 1 and 2 for convenience.

When we calculate the partition function, we will have to deal with

$$\int d^3x_1\ d^3x_2\ U(|x_1 - x_2|) \tag{2}$$

The notation d^3x_1 is there to remind us that x_1 is a three- dimensional vector, and the differential element is actually a small volume. Same comment for x_2.

So what we are doing in expression (2) is taking the two particles and integrating over all possible positions of the pair of particles.

In some manual of mathematics, this double integral would be written

$$\iint d^3x_1\ d^3x_2\ U(|x_1 - x_2|)$$

The integral in expression (2) is actually a double integral.

We can decompose its calculation into two steps.

First of all we keep x_1 fixed, and look at the simple integral when x_2 ranges over its domain :

$$x_1 = \text{constant} = c \qquad y = c - x_2 \qquad d^3x_2 = d^3y$$

$$\iint d^3x_1\ d^3x_2\ U(|x_1 - x_2|) \quad = \int d^3c \int d^3y U(|y|) \tag{3}$$

The term $\int d^3y U(|y|)$ is a number.

Now to complete the double integral, we integrate over the positions of x_1 or now the variable c.

That gives us simply, as a multiplicative factor, in front of expression (3) the volume V.

So we can write (changing the label y to x)

$$\int d^3x_1 \, d^3x_2 \, U(|x_1 - x_2|) \; = \; V \int d^3x \, U(|x|) \tag{4}$$

What are the units of the last integral on the right-hand side ?

Volume times energy.

So this last integral is telling us how important the energy is, because it is both telling us how strong the potential energy is and on how big a volume does U act – in other words how spread out it is.

The integral $\int d^3x \, U(|x|)$ is going to be our important numerical quantity which determines the strength of the potential.

As said, it is a combination of how strong the potential is and how widely spread it is distributed.

So let's give it a name.

We call it U_0 (read *u zero*, or *u naught*)

$$U_0 = \int d^3x \; U(|x|) \tag{5}$$

It is the only parameter that will come into our calculation.

We don't know offhand whether it is big or small.

It depends on the function U and the volume over which it is integrated.

It could be big, even if U is never very big, because it is integrated over a wide volume, like a large thin puddle can contain a lot of water.

Or the integration could be over a small volume, but where U is sufficiently big to make U_0 big.

One way or another U_0 is the quantity which will determine everything else.

Finally, as explained, when we do the double integration in expression (2), we end up with

$$\int d^3x_1 \; d^3x_2 \; U(|x_1 - x_2|) \; = \; VU_0 \tag{6}$$

This is the potential energy for one *pair* of particles, when integrated over all the possible positions of the pair.

To get a feel for what U_0 is, think for a moment of x_1 and x_2 as one-dimensional variables. Then fixing $x_1 = 0$ and letting x_2 vary, $U(|x_1 - x_2|)$ is the following function, figure 1.

Of course, in truth x_1 and x_2 are three-dimensional.

And the integration producing U_0 is the integral of U(x) over the whole volume of the gas.

The integral U_0 converges if U(x) falls off sufficiently rapidly when the distance between the molecules grows.

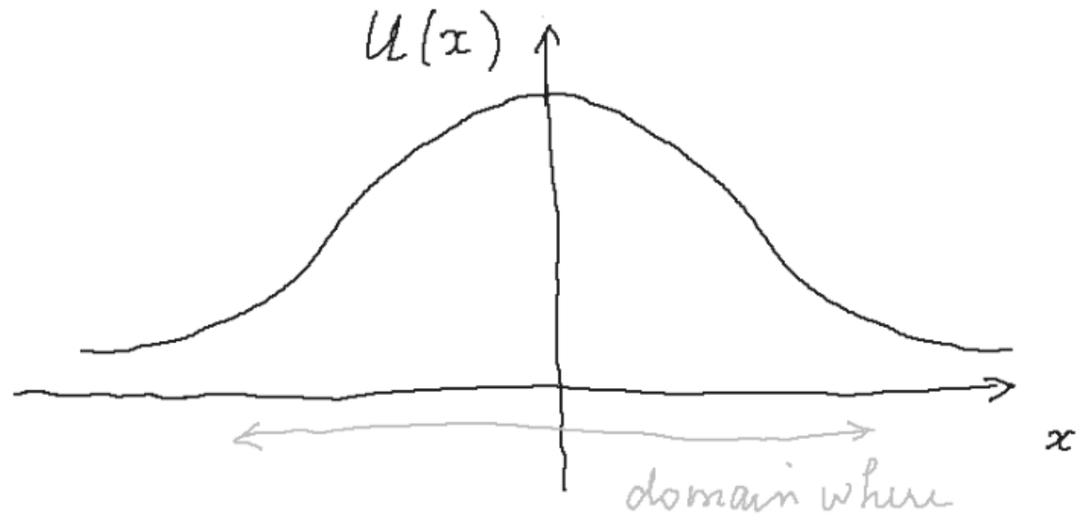


Figure 1 : Function U(x). U0 is the area under the curve.

In practice, for a weakly interacting gas, the function U(x) will become negligible when x is larger than a couple times the diameter of a molecule.

We can say that U_0 is "qualitatively equal" to $U\,\mathcal{V}$, meaning by that the height U of the function U(x) near zero, times the volume $\mathcal{V}$ over which the function U(x) is significant, figure 1. The volume $\mathcal{V}$ is a couple of times the volume of a molecule basically.

In other words, the potential energy between molecules – and therefore the force between them (remember that the force is minus the gradient of the potential energy – is only significant when the two molecules are within a couple of molecular diameters of each other.

There are two volumes we are talking about in our analyses.

The volume $\mathcal{V}$, in $U_0 = U\,\mathcal{V}$, is roughly the volume of a couple of molecules.

The other volume, V , on the right- hand side of equation (6), is the volume of the whole sample, that is the volume of the box of gas.

The small volume $\mathcal{V}$ gets absorbed into U_0.

We don't have to keep track of it anymore.

The quantity U_0 is the product of the small volume $\mathcal{V}$ where the potential energy between the molecules is significant, times the value U of that potential energy between the molecules.

Then we have to take into account the integral in equation (6). That gives VU_0, where V is the total volume of the system.

The next steps are to calculate the partition function, and then use it.

But let's first of all do some questions/answers to clear any cloudy point that may remain before we go on.

6.2 Questions/answers

Question: When we calculate the potential energy of a pair of molecules, shouldn't we take into account boundary effects ?

Answer: That is a good question.

The answer is : there are boundary effects, yet they can be neglected.

Let's see why.

We saw that U_0 is independent of the position x_1 where we fix the first molecule of the pair in order to compute the integral of $U(|x_1 - x_2|)$ over the position x_2 of the second molecule.

But of course that is not exactly true when x_1 is near a boundary of the box.

In all these thermodynamical problems there are always boundary effects.

And we are always using the fact that for a large vessel, the total volume is much bigger than the area of the boundary times the distance of the effect, in other words much bigger than the volume where there are boundary effects.

It is because we are accustomed to these boundary effects that we didn't consider a priori that the behavior of molecules in an ideal gas was isotropic everywhere including near the walls.

It turned out it is.

But we established $P = \varrho T$ for an ideal gas without assuming the absence of boundary effect.

And that was important because, in some sense, pressure on the wall is nothing but a boundary effect !

So in general when we take the molecule x_1, and move it near the walls of the box, we make a little mistake by assuming that the other molecule x_2 can be anywhere around it.

It can be anywhere around it except when it is banging to the wall.

The mistake of neglecting the boundary effects, however, is not important as long as the total volume V of the box is much bigger than the volume $\mathcal{V}$ where the potential energy is significant around a molecule.

In other words if the box is much bigger than a few molecules, evenly spread over the walls, then the error is negligible.

So, if the box is large enough the surface to volume ratio is very small.

Then, if the forces are not long range, we can safely neglect the boundary effects.

Of course, if the vessel if very small, of the order of magnitude of a few molecules, then we could not neglect boundary effects.

In fact they would **dominate** the experiment.

Question: The quantity $\int d^3x_1\, d^3x_2\, U(|x_1 - x_2|)$ is the integral of the potential energy of one pair of particles over all possible configurations of the gas in the box, is that correct ?

Answer: Yes, that is perfectly correct.

And as we saw it is equal to VU_0.

Its units are volume squared times energy, but it is only proportional to the volume V, the proportionality factor being the important quantity U_0.

We will use it in a moment.

Question: Shall we eventually take into consideration the collisions between particles ?

Answer: No, we won't.

Not that there are no collisions, but we won't have to introduce them specifically into our analysis.

They will be taken care of automatically.

That is the power and the beauty of applying general thermodynamical principles as opposed to sticking to a mechanistic analysis – a kinetic theory analysis – of the problem.

Again, we already saw that when we studied the ideal gas.

The general principles overcame the question of whether there was a boundary effect that would corrupt the simple application of $F\Delta t = m\Delta v$ onto the walls.

It turned out there was none. But we didn't care.

Similarly here there will be collision effects – and this time we won't discover in the end that there aren't any – but we don't have to be concerned with them.

All will be taken care of in the use of the partition function whose first brick we set with equation (1).

We are not ever going to have to think about collisions.

We are never going to have to think about following the particles.

We don't have to think about them.

This is the beauty of statistical mechanics : you lose all intuition about what is really going on, but you have a set of mechanical rules.

You go on to autopilot.

You follow the rules.

And you get the answers in a completely rigorous way.

You don’t have to worry about whether some assumptions that you may have made, such as whether the particles have the same average velocity, if they are near the walls or not near the walls, near each other or not near each other.

You don’t have to worry about all that.

Of course, once we reach some result, it is always good to step back and ask: What really are we doing ?

Are we getting the answer that we would have expected from naïve thinking ?

But if we want to really know whether we have the right answer, the best is to set up the rules, get on to autopilot, follow the equations and, at the end, look at the result and ask : Does this make sense ?

Question: We are basing our analysis on looking at pairs of particles.

Are we not losing anything by not looking at triples or other higher combinations of particles ?

Answer: First of all, when we wrote that the potential energy was

$$\sum_{n>m} U(|x_n - x_m|) \tag{7}$$

there could be terms in the energy which depend on three particles at a time.

The energy between two particles conceivably could depend on the presence of a nearby third particle.

So we have made an assumption when we wrote that the energy can be expressed as a sum over pairs, as in equation (7).

Now, the point is : Why was it a good approximation first of all to use the ideal gas?

And the answer is : It is a good approximation when the probability for two particles to be close enough together to feel each other is small.

When the gas is dilute, the particles spend most of the time being far enough away that they simply don't feel each other.

Some fraction of the time, two particles will get close enough together to interact.

That is when they are within the little distance where $U(x)$ is not negligible – when they are in the same little volume $\mathcal{V}$.

Then the forces between the two molecules become important.

Now what is the fraction of the time that three given molecules will be within the molecular distance of each other?

Much, much smaller.

You pick three molecules and ask what is the probability that all three of them be within a molecular diameter of each other.

It is vastly smaller, than for two molecules.

It depends on the overall volume V of the gas of course.

But, as said, we do assume that V is much larger than $\mathcal{V}$.

So if there were three-body forces, they would be less important not because they are weaker but because the probability of finding three particles in the neighborhood of each other for a dilute gas is even smaller than for two particles.

We are going to do an expansion.

And in the expansion, there are three-body effects even if the forces are only two-body effects.

The answer to the question is that we are going to follow the expansion and see what we find.

There are other terms.

But they won't be important as long as the gas of particles is reasonably dilute.

Let's go on.

Now we are going to go on autopilot.

What do we want to calculate ?

We want to calculate the partition function.

6.3 Partition function

The partition function Z is the integral over the positions and the momenta of $e^{-\beta E(p,x)}$.

It is a generalization to the continuous case of equation (42b) in part 4 which we established in the discrete, and even finite, case.

Now the space of possible states of the system is a $6N$-phase space of positions and momenta, where N is the number of particles.

Z will be therefore calculated as a multiple integral of high dimension.

dp will stand for $3N$ linear differential elements of momentum.

And so will dx for positions.

Introducing, to follow the custom, the factor $N!$ downstairs, although it is not going to make any difference, here is the formula for the partition function

$$Z = \frac{1}{N!} \int dp\, dx\, e^{-\beta \sum \frac{p_n^2}{2M}}\, e^{-\beta \sum U(|x_n - x_m|)} \tag{8}$$

The glorious and simplifying feature, that we have already seen, of this integral is that it factorizes.

We can rewrite it

$$Z = \frac{1}{N!} \int dp\, e^{-\beta \sum \frac{p_n^2}{2M}} \int dx\, e^{-\beta \sum U(|x_n - x_m|)} \tag{9}$$

The first integral, that is the momentum integral, we have already dealt with in the case of the ideal gas.

It is a product of identical Gaussian integrals.

We even calculated the Gaussian integral in part 4.

The second integral, that is the position integral, in the case of the ideal gas, was very simple.

There was no potential energy.

U was equal to zero.

And that gave us the volume V to the power N.

So let's multiply and divide expression (9) by V^N.

$$Z = \frac{V^N}{N!} \int dp\, e^{-\beta \sum \frac{p_n^2}{2M}} \int dx\, \frac{1}{V^N}\, e^{-\beta \sum U(|x_n - x_m|)} \tag{10}$$

The first part is just the partition function for the ideal gas.

See equation (39) of part 5, which we wrote before introducing $N!$ downstairs.

So we can call it $Z_0(\beta)$.

The index 0 now stands for the calculation when U was equal to zero, that is the ideal gas case.

So we write

$$Z = Z_0(\beta) \int dx \, \frac{1}{V^N} \, e^{-\beta \sum U(|x_n - x_m|)} \tag{11}$$

The remaining integral is the new thing. It is the factor we have to work on.

Notice that although the calculations may seem heavy, what we have done is really simple.

Starting from the ideal gas case, we added a potential energy term to the energy $E(x, p)$ of one configuration, equation (1).

Then we used this new $E(x, p)$ to write the partition function.

The sum in $E(x, p)$ becomes a product of exponentials under the integral sign in the expression of Z.

Because, as in the ideal gas case, the integral involves separately positions and momenta, it factorizes.

The first term, after multiplication by V^N, is the good old partition function of the ideal gas.

And the second term now is the new one due to the potential.

This remaining factor in equation (11) is an integral over all the x_n and all the x_m forming pairs, with $n > m$.

Let's write it temporarily

$$\int \frac{dx}{V^N}\, e^{-\beta W(x)} \tag{12}$$

What is the small quantity in our calculation, in terms of which we are going to expand?

It is the potential term W(x).

The potential energy is the quantity which is assumed to be small.

We are starting with an ideal gas.

Then imagine turning a knob which creates a potential energy.

We turn it slowly until we just reach a weak potential energy.

Then, in expression (12), having a weak potential energy, we can expand the exponential in a Taylor series in the strength of W.

It is very simple.

It is just equal to $1-\beta W$(x).

Those are the first two terms of the general expansion

$$e^s = 1 + s + \frac{s^2}{2!} + \frac{s^3}{3!} + \frac{s^4}{4!} + \ldots \tag{13}$$

This is the most famous power series.

It converges for all *s*'s positive or negative, real or complex.

The reason is that the factorials force a rapid decrease to zero after a while.

The power series (13) yields useful approximations of e^s when *s* is small.

First of all, when *s* is sufficiently small we can ignore everything but the 1.

The next approximation is 1 + *s*, that is in many cases adequate for small *s*.

In equation (12) we want to keep track of the correction term due to potential, so we keep the 1 and the first term

$$e^{-\beta W(x)} \approx 1 - \beta W(x) \tag{14}$$

Let's plug that in equation (12).

We get rid of the exponential and replace it by the approximation

$$\int \frac{dx}{V^N} \left[1 - \beta W(x) \right] \tag{15}$$

What does the integral over 1 give us ?

It gives V^N.

So we get

$$1 - \beta \int \frac{dx}{V^N} \, W(x) \tag{16}$$

Now let's go back to the messy sum that we had temporarily replaced by W(x).

There are N ways to choose 2 pairs x_n x_m.

The factor multiplying Z_0, in equation (11), due to the potential energy between molecules can therefore be rewritten

$$1 - \beta \binom{N}{2} \int \frac{1}{V^N} \, dx_1 \, dx_2 \, U(|x_1 - x_2|) \, dx_3 \, \ldots \, dx_N \tag{17}$$

Once we have put the factor $\binom{N}{2}$ in front, we don't have to sum anymore.

All the pairs playing the same role, we can integrate simply $U(|x_1 - x_2|)$.

The integrals over dx_3 dx_4 ... dx_N give V^{N-2} .

So, spelling also out the formula for $\binom{N}{2}$ we get

$$1 - \beta \, \frac{N(N-1)}{2} \, \frac{1}{V^2} \int dx_1 \, dx_2 \, U(|x_1 - x_2|) \tag{18}$$

Now remember, see equation (6), (in equation(6) the notations had stressed that each infinitesimal element *dx* was actually an infinitesimal volume which we had denoted d^3x and equation (6) and equation (19) are the same) so that we have

$$\int dx_1 \, dx_2 \, U(|x_1 - x_2|) \; = \; VU_0 \tag{19}$$

So, approximating $N(N-1)$ by N^2, the factor in the partition function due to potential can be written

$$1 - \beta \, \frac{N^2}{2} \, \frac{U_0}{V} \tag{20}$$

Finally we reached the partition function for the weakly interacting gas.

It is the partition function of the free gas times the factor we just calculated.

This gives

$$Z \;=\; Z_0 \left[\, 1 - \frac{\beta N^2}{2V}\, U_0 \,\right] \tag{21}$$

What do we want to do with it?

Well, again we go on to autopilot.

Everybody knows that what we do first with a partition function is to take its logarithm.

All the interesting formulas use the logarithm of the partition function.

$$\log Z \;=\; \log Z_0 \;+\; \log \left[\, 1 - \frac{\beta N^2}{2V}\, U_0 \,\right] \tag{22}$$

Since we are considering a weakly interacting gas, U_0 is small.

So we can expand in U_0.

U_0 is our expansion parameter.

Above we saw the Taylor series expansion of e^S, that was formula (13).

Now we want to use the Taylor series expansion of $\log(1 + s)$ - the natural logarithm.

It is

$$\log(1+s) = s - \frac{s^2}{2} + \frac{s^3}{3} - \frac{s^4}{4} + \ldots \tag{23}$$

Notice that it starts with s, the signs alternate, and there are no more factorials.

Unlike the power series of e^s, this one doesn't converge for all s.

Its radius of convergence is one.

That is to say, it is a meaningful series only for $|s| \leq 1$.

We are interested in only the first term in the expansion of $\log(1 + s)$.

So the log of the partition function rewrites

$$\log Z \;=\; \log Z_0 \;-\; \frac{\beta N^2}{2V}\, U_0 \tag{24}$$

Now, is this what we want?

No, we are never really interested in the log of the partition function itself.

We are interested in what we can get from it.

Let's pick something interesting that we can get from the log of the partition function.

For instance, energy.

Let's calculate the energy of the gas.

$$E = -\frac{\partial \log Z}{\partial \beta} \tag{25}$$

Using equation (24), this becomes

$$E = \frac{3}{2} NT + \frac{\partial}{\partial \beta} \left[\frac{\beta N^2}{2V} U_0 \right] \tag{26}$$

This first term is the good old formula of Maxwell and Boltzmann for the energy of an ideal gas of N particles in thermal equilibrium at temperature T.

Each particle has energy 3/2 T, and that is multiplied by N.

The second term is minus the derivative with respect to β of the potential term in equation (24).

The minus signs cancel, and, the only term depending on β being β itself, we get

$$E = \frac{3}{2}NT + \frac{N^2}{2V}U_0 \tag{27}$$

Now of course we always expect in a situation like this that the total energy should be proportional to the number of particles.

So the first term is alright.

But the second term is weird.

It is quadratic in the number of particles.

A priori to have an N^2 is crazy.

If I have 10^{23} particles, I get an energy of 10^{46}.

I don't care what units are used, 10^{46} is just too big.

But that is alright, because there is a V downstairs that is also big.

So let's keep one N over the volume V and pull the other one in front.

N/V is the density, which we denote ϱ.

Hence equation (27) becomes

$$E = \frac{3}{2}NT + N\frac{\rho}{2}U_0 \tag{28}$$

The density is neither big nor small.

It is what it is.

Well, it may be on the small side because we are talking about a dilute gas.

But equation (28) is the formula we obtain for the energy of a weakly interacting gas.

Let’s examine this formula.

It has two terms.

One coming from the ideal gas.

It doesn’t depend on the density at all.

Of course it is proportional to the number of particles.

And there is the other term that is also proportional to N.

It depends on the potential energy between particles. And it also depends explicitly on the density.

Why does it depend on the factor of the density?

In formula (28) the density appears, with the factor $U_0/2$, as an energy per particle.

But there shouldn't really be an energy per particle.

There should be an energy per particle only if there is another particle nearby.

So if we focused on one particle and ask : What is the probability that there is another particle nearby ?

It would be proportional to the density.

The higher the density, the more probable is that there is a neighboring particle.

Hence the correction term in formula (28), in addition to the ideal gas term.

It can also be viewed in formula (27) in the fact that there is an N^2.

N^2 is the number of pairs of particles roughly.

Remember, in fact, it is we who changed $N(N-1)/2$ into $N^2/2$ for simplicity.

We may also express formula (28) factorizing N in front of the energy per particle

$$E = N\left[\frac{3}{2}T + \frac{\rho}{2}U_0\right] \tag{29}$$

The energy per particle has a first term which is just proportional to the temperature.

And a second term is proportional to the potential energy between pairs of particles times another factor of the density.

Of course this can be confirmed for a weakly interacting gas.

However, despite the fact that we once in a while talked about measuring not only the average energy E but even the precise level of energy E_i of a gas, one doesn't easily measure the energy of a gas.

So this is not the way to confirm it.

What is an interesting quantity that is easy to measure?

I'm not alluding to the temperature.

Temperature is an input in formula (29).

Notice by the way that the correction term doesn't depend on the temperature.

The other interesting quantity that is easy to measure, beside temperature, is the pressure.

So what does our calculation say about the pressure ?

We have everything we need to calculate the pressure.

We built our tools in part 5.

Remember that we defined pressure as

$$P = -\left.\frac{\partial E}{\partial V}\right|_S \tag{30}$$

Then, since it is impractical to measure something at fixed entropy, we used a small result in calculus to reexpress the pressure as the derivative, with a minus sign, of the Helmholtz free energy A with respect to the volume at fixed temperature.

$$P = -\left.\frac{\partial A}{\partial V}\right|_T \tag{31}$$

Helmoltz free energy is $A = E - TS = -T \log Z$.

So formula (31) can be rewritten

$$P = T \left. \frac{\partial \log Z}{\partial V} \right|_T \tag{32}$$

We then applied this formula to the ideal gas, see formula (46) in part 5.

But formula (32) is true in general.

Now let's apply it to the weakly interacting gas.

In the weakly interacting gas case, we established, in formula (24), that

$$\log Z \;=\; \log Z_0 \;-\; \frac{\beta N^2}{2V}\, U_0 \tag{33}$$

To obtain P, we want to calculate at fixed temperature

$$T \frac{\partial \log Z_0}{\partial V} - T \frac{\partial}{\partial V} \left(\frac{\beta N^2}{2V}\, U_0 \right) \tag{34}$$

When we differentiate $\log Z_0$, what do we get, when multiplied by T in front ?

We get the pressure of the ideal gas that we calculated in formula (46) of last chapter.

That is ϱT.

Then let's differentiate the second term in expression (34).

But first of all, notice that we have the factor T in front, and the factor β inside.

Those two are inverse of each other and cancel.

So formula (34) yields the equation

$$P = \rho T - \frac{\partial}{\partial V}\left(\frac{N^2}{2V}\, U_0 \right) \tag{35}$$

This is

$$P = \rho T + \frac{1}{2}\frac{N^2}{V^2}\, U_0 \tag{36}$$

And finally

$$\boxed{P = \rho T + \frac{1}{2}\, \rho^2\, U_0} \tag{37}$$

That is the generalization to a weakly interacting gas of the formula we had for an ideal gas.

Notice, among other things, that we are expanding in powers of the density.

If the density is zero there is no pressure.

Indeed there is no particle, no temperature, no pressure, no nothing.

Then if the density is very low, the quadratic term is negligible and we are in the ideal gas case.

In that case the pressure is proportional to the density and to the temperature.

That makes sense.

If the temperature is zero, the molecules aren't moving at all.

They don't bounce off the walls at all.

And if the temperature is not zero, the linearity with the density also makes sense, as long as the density is small.

Indeed in that case the particles essentially don't interact with each other, so if we put twice as many in the same volume the pressure on the walls must double.

Then, as the density gets higher, a quadratic term appears.

It represents the effect of the interaction between molecules on the pressure.

The correction term is proportional to ϱ.

That makes sense : when ϱ is small, the correction term is a lot smaller than ϱT .

Interestingly enough, the correction term doesn't depend on the temperature.

It requires a bit of physical intuition to understand that.

But don't overlook that it is only the correction term to the pressure that is of course function of the temperature, i.e. the agitation of the molecules.

But it depends on the potential energy between molecules, or more precisely on the parameter U_0 defined as

$$U_0 = \int dx\ U(x) \tag{38}$$

where U(x) is the function of potential between two molecules with their distance as the independent variable.

And integrating $U(\mathrm{x})$ over the entire volume around one molecule gives U_0, which has units energy times volume.

The correction term in the pressure is proportional to U_0, which – remember – we assumed to be small, enabling us to approximate e^{s} as $1+s$, and $\log(1+s)$ as s, where s was essentially proportional to U_0.

In other words, the correction term depends on the fundamental parameter U_0 which governs the strength of the interaction.

And it has got ϱ^2 as a factor in front.

Laboratory experiments can confirm equation (37), and even produce a numerical measure of U_0.

We take a gas in a vessel, measure its temperature, its pressure and its density.

Then we change for instance the density.

We can change the density either by changing the volume or by introducing more matter into the vessel.

And we can keep the temperature fixed, if we like, by keeping the box in equilibrium with a heat bath.

Or we can also change the temperature.

If we do the experiment for several triplets (P, ϱ, T), which we can all control or measure, we shall not only be able to compute U_0 but we shall also be able to verify equation (37).

We will see that when the density remains reasonably low, the pressure varies according to equation (37) which we reproduce below

$$P = \rho T + \frac{1}{2}\,\rho^2\,U_0 \tag{39}$$

with a linear term in the temperature and the density, and a quadratic term in the density times $U_0/2$.

Of course the factor 1/2 won't show explicitly.

It is a matter of definition of U_0.

Question: How would the formula for the pressure change if we took into account the 3-body interactions between molecules ?

Answer: If we assume that the energy scale for the 3-body interactions between molecules is about the same as for the 2-body interaction, then there would be a further term in ϱ^3 in formula (39), with a factor depending on U_0 and also on T.

You will have to look up what the cubic coefficient is exactly.

In any case $1/2\ \varrho^2 U_0$ is the first correction term, for a dilute gas, to the formula $P = \varrho T$, when we want to take into account weak interactions between molecules.

It provides a good approximation for a dilute gas.

When

$$\rho^2 U_0 \ << \ \rho T \tag{40}$$

the sign << meaning "much smaller than", then we can use the ideal gas formula for the pressure.

Dividing by ϱ, condition (40) can be rewritten

$$\rho U_0 \ << \ T \tag{41}$$

So when the temperature is much bigger that ϱU_0, then we can use the ideal gas model to evaluate the pressure.

Temperature is a measure of the kinetic energy of each particle.

Remember, it is not related to the number of particles.

A unique particle has a temperature.

Furthermore in a vessel in thermal equilibrium, all the particles have the same kinetic energy, therefore the same temperature.

They don't have to be all the same particles.

You can have a mix of two gases, with different molecular masses.

The particles of one type will have one speed, and the particles of the other type will have another speed.

We pointed out that we can even plunge a bowling ball into the gas in the vessel.

When thermal equilibrium is reached, the ball will also have the same kinetic energy as the other molecules – which means of course that it will remain essentially still.

Remember that the units of U_0 are energy times volume.

And the units of ϱ are 1 over volume .

We are accustomed in physics to think of the density as a mass per volume, or an energy per volume.

That is for instance how one defines the various parameters ϱ in Cosmology.

But in statistical mechanics ϱ is most often defined as a number of particles per volume.

And that is indeed how we defined it in the previous parts and here.

So the units of ϱU_0 are energy.

Therefore in condition (41), we are comparing energy on the left with energy on the right

ϱU_0 is the potential energy per particle.

So if the potential energy per particle is much less than the kinetic energy per particle, then the ideal gas formula $P = \varrho T$ is a good approximation.

That is what we are doing when we use the ideal gas model.

We are really studying the gas in the range where the potential energy is much smaller than the kinetic energy.

The purpose of what we explained so far in this part was to show how the rules work, to show that we are not limited to the ideal gas, and finally to show that we reached a nice and simple formula, $P = \varrho T + 1/2\ \varrho^2 U_0$, that would have been difficult to guess from intuition.

Could we have guessed the correction term ?

We might have guessed that it was proportional to ϱ^2 because it involved pairs of particles.

We might have guessed that it involved the potential energy.

It would not have been so obvious, though, that it involved exactly the integral $\int U(x)\ dx$.

But it is plausible that U times the volume is there.

See figure 1 and the subsequent comments for what we mean by U times the volume.

Would we have guessed the factor 1/2?

May be if we are really smart, we might have guessed the 1/2.

Would we have been sure in the end that 1/2 $\varrho^2 U_0$ was right ?

Probably not.

The mechanical calculation, starting with calculating the partition function, is the way to do these things, and to be sure in the end that we are right

So we have done a couple of problems.

The next topic is heat and work.

But before that let us do some questions/answers.

6.4 Questions / answers

Question: In a typical situation of temperature and pressure, like the air in the room, how big are the numbers ϱ, U_0, and their product compared to temperature ?

Answer: In the room where we stand, at so-called room temperature and pressure, the temperature is much bigger than ϱU_0.

It is a little exercise to work out.

The potential energy U between a pair of molecules is of the order of let's say one hundredth of an electron-volt. That is 1.6 x 10^{-21} Joules.

The density ϱ is the number of particles per cubic meter in the room.

We can take Avogadro's number per 22.4 liters.

That gives $\varrho = 2.7 \text{ x } 10^{25}$.

The units are the inverse of cubic meters.

We need something else : the volume of a few molecules.

We can take for that a cube of side 10 Angströms.

One Angström is 10^{-10} meters.

Following the comments to figure 1, the formula $U_0 = U\mathcal{V}$ gives $U_0 = 1.6\text{x}10^{-27}$.

Thus $\varrho U_0 = 4.3 \text{ x } 10^{-23}$ Joules.

On the other hand 300 Kelvins is $4.1 \text{ x } 10^{-21}$ Joules.

You have to start getting into the range where the molecules are almost butting up against each other for the condition (41) to break down.

It is when they start having appreciable probability of being on top of each other that it becomes violated.

You can calculate how much you would have to shrink the room in order that the ideal gas approximation be no longer acceptable.

Question: Is it possible for the term U_0 to be positive or negative ?

Answer: Yes, it is.

Consider the potential energy function U(x) between a pair of molecules.

It can look like figure 2a or figure 2b.

Which represents attraction and which represents repulsion ?

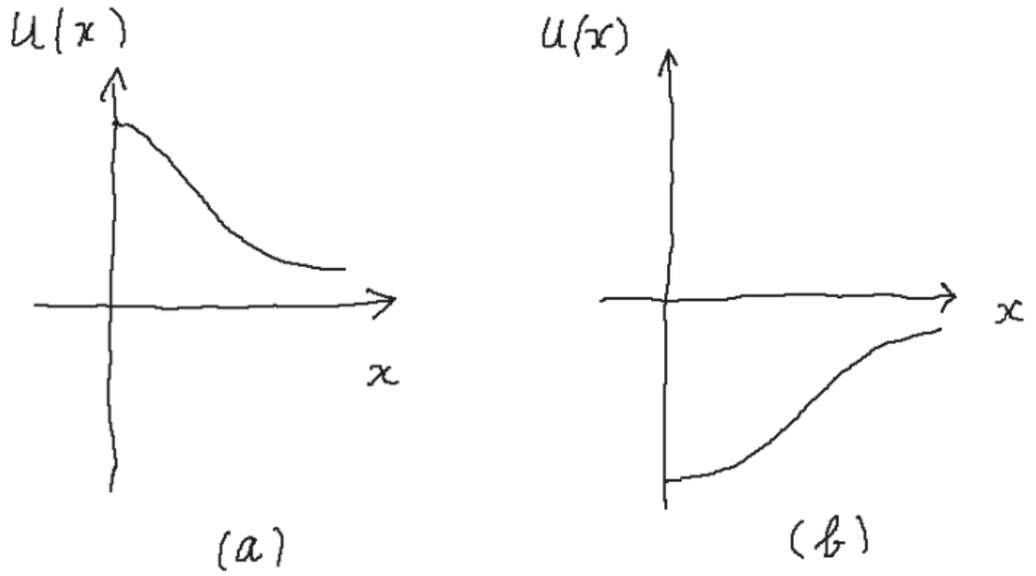


Figure 2 : Possible potential energy curves between two molecules at distance x of each other.

Curve (b) represents attraction, because the potential energy increases as we separate the molecules.

And curve (a) represents repulsion.

What should you think off hand should be the effect of repulsive forces on the pressure?

If there is repulsion between the molecules, and you try to squeeze them to the point where they scream to each other "I don’t want to get close to you, Buddy", the pressure is going to go up.

So in the repulsive situation, that is curve (2a), U_0 is positive.

It corresponds to an extra bit of pressure on top of the ideal gas pressure $P = \varrho T$.

If the molecules, on the other hand, attract each other that means U_0 is negative.

And the actual pressure is a bit smaller than in the ideal gas case.

Question: Can the forces between molecules be repulsive over a range of distances, and attractive over another range ?

And can there be gravitational forces ?

Answer: Forget gravitation.

It is totally negligible for a gas in a laboratory vessel.

Then yes the intermolecular forces can be repulsive at first and then attractive after a certain distance.

To know what the forces between two molecules are, you will have to know something about molecular and atomic physics.

The potential energy $U(x)$ between two molecules as a function of distance is given by the Lennard-Jones law.

A typical shape is shown in figure 3.

The forces are repulsive at short distances, and then there is an attractive tail.

The figure doesn't show it clearly, but usually the average force is repulsive.

Therefore in most molecular situations U_0 is positive.

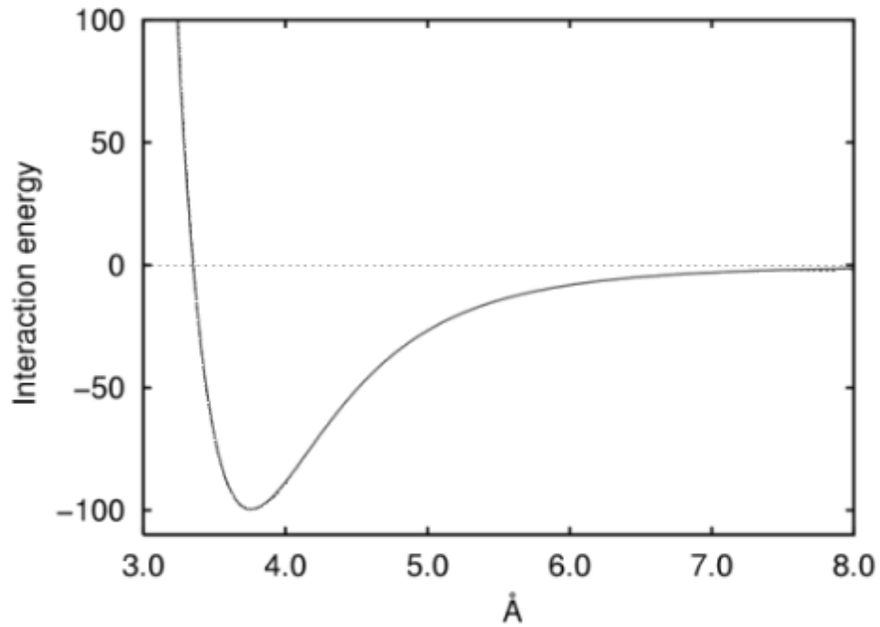


Figure 3 : Typical real life U(x). Lennard-Jones law.

And the actual pressure is a bit higher than in the ideal gas model.

But we are not doing molecular physics.

So I'm not going to tell you how to calculate the curve.

You need a handbook of physics or chemistry to learn what the actual potential energy is between a pair of molecules in various kinds of situations.

You will learn, for instance, what is the curve for a pair of hydrogen molecules, or a pair of helium molecules.

Helium is usually the first example presented, because helium is not reactive.

Question: What are the forces when we have a dipole ?

Answer: We should not mix up the forces of chemical bonds, called covalent forces, between atoms, which involve sharing electrons, and the forces between atoms or molecules resulting from other kinds of interactions, which are those that concern us here.

In other words, we are not concerned by dipoles in the chemical sense.

The forces we are considering are called the Van der Waals forces.

Beware that the term Van der Waals forces is used differently by different authors.

Sometimes they refer to all the non-covalent forces, attractive and repulsive, and sometimes only to the attractive ones.

They are then used to explain for instance how geckos can walk on ceilings.

And the curve U(x) is sometimes called the six-twelve potential because it involves an attractive force in x^{-6} and a repulsive force in x^{-12} which are competing to form the resulting curve in figure 3.

This is not the point though.

The point is, given the dynamics, given the formula for the energy, how do you proceed ?

Getting the formula for the potential energy is not a problem in statistical mechanics.

It is a problem in quantum mechanics, in molecular structure, in atomic physics, or whatever.

For us the potential energy is a given and then we start from to think about statistical mechanics.

That is not subtle.

It is very straightforward.

As I said, it is blind navigation.

Just write down the energy formula.

Start doing the integrals.

Here and there of course there are tricks.

But with experience, you know the tricks.

You know to multiply and divide by the volume to the *n*-th power, and things like that.

It doesn't take a hell of a lot of insight.

It is blind manipulation of the symbols.

Now at the end of the calculations, it takes some insight to check whether the results make sense.

We can ask the question : Am I getting a sensible result ?

Looking for instance at the equation

$$P = \rho T + \frac{1}{2}\,\rho^2\,U_0$$

what does this equation mean?

Has the correction term the right sign ?

Does it make sense to say that, if the forces are repulsive, it increases the pressure between things?

Is it reasonable that the correction be proportional to ϱ squared ?

Why the factor of 1/2?

Well, we know where that came from.

It came from the $n(n-1)/2$.

So you might in hindsight say: Yeah, there should be a 1/2 there.

We would overcount if we counted that 1 and 2 is different than 2 and 1.

And so that is the half.

Now we arrive at something that is much less straightforward.

I remember it really was confusing me when I learned thermodynamics.

Unfortunately the professor was more confused than I was and couldn't straighten me out.

It had to do with heat and work and something called exact differentials.

So let's take a little mathematical interlude to talk about exact and inexact differentials.



6.5 Exact and inexact differentials

Exact and inexact in this context have nothing to do with precise versus imprecise.

Exact is a technical mathematical term which probably predates many fancy things like that.

But it may have first turned up in thermodynamics.

Take a function of two variables $z = F\,(x, y)$, figure 4.

Think of $F(x,y)$ as the altitude z of the surface, at the point (x,y) on the horizontal plane.

In other words, the surface F represents a hilly terrain.

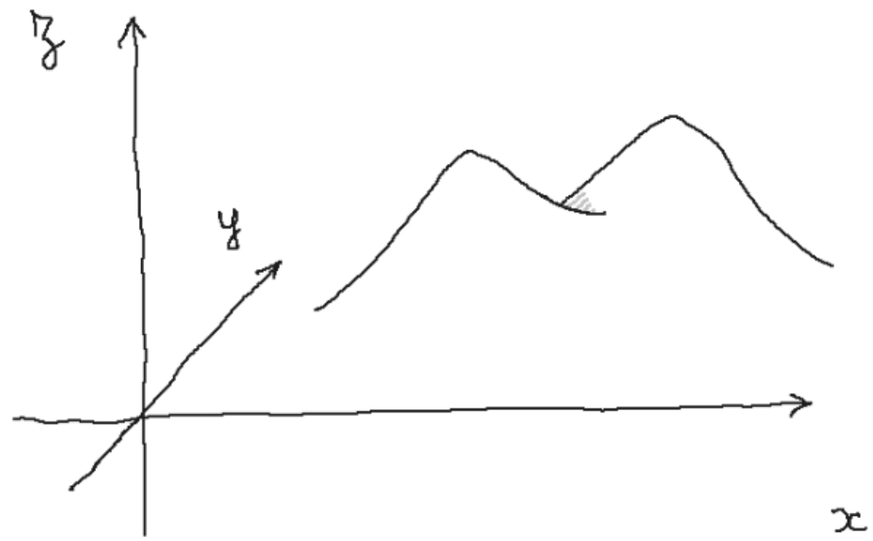


Figure 4 : Function of two variables. It is a surface.

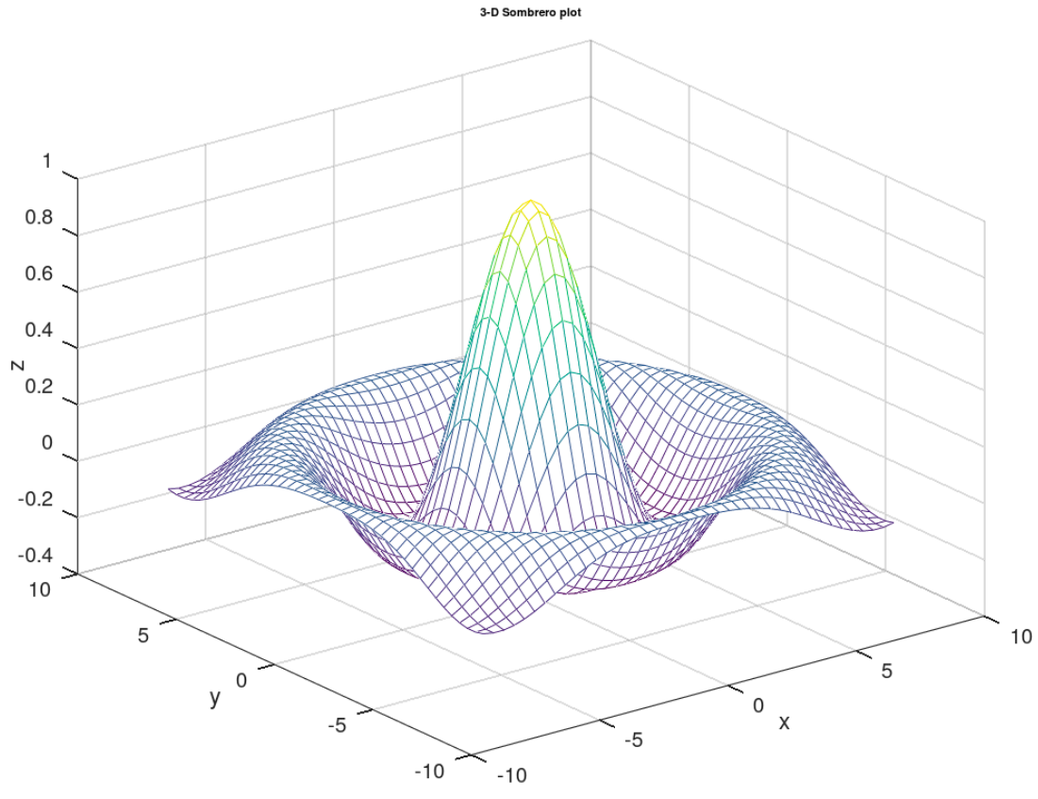
Now we take a little walk on the surface.

When projected onto the (x, y) plane, it corresponds to a dx and a dy.

How much does our altitude change during the walk ?

The answer is given by the familiar formula

$$dF = \frac{\partial F}{\partial x}\, dx + \frac{\partial F}{\partial y}\, dy \qquad (42)$$



Our walk doesn't have to be parallel to the *x*-axis, or parallel to the *y*-axis.

When projected onto *(x,y)*, we can be moving at an angle.

Formula (42) nevertheless is the answer.

Let's use the standard notation

$$dF = F_x\, dx + F_y\, dy \qquad (43)$$

Are there any restrictions on what F_x and F_y can be in order that the right-hand side really be the small change and some genuine function F ?

Yes there are.

There is basically one restriction.

The more mathematically inclined may have guessed that it has to do with the cross second derivative.

Consider the quantity

$$\frac{\partial^2 F}{\partial x\ \partial y} \tag{44}$$

This is the second derivative of F where we first differentiate with respect to y and then we differentiate the result with respect to x

It is a theorem of calculus that within a large class of functions it doesn't matter in which order we do the derivatives.

It is true for all smooth surfaces we are concerned within physics, eliminating only teratological examples of which mathematicians have the secret.

In order to carry out the next bit of mathematics, we will need more than the rudimentary definition of a partial derivative that we have been using… so

Digression on Partial Differential Calculus

Partial Derivatives

The calculus of multivariable functions is a straightforward generalization of single-variable calculus.

Instead of a function of a single variable t, consider a function of several variables.

To illustrate, let's call the variables x, y, z, although these don't have to stand for the coordinates of ordinary space.

Moreover, there can be more or fewer than three.

Let us also consider a function of these variables, V(x, y, z).

For every value of x, y, z, there is a unique value of V(x, y, z) that we assume varies smoothly as we vary the coordinates.

Multivariable differential calculus revolves around the concept of partial derivatives.

Suppose we are examining the neighborhood of a point x, y, z, and we want to know the rate at which V varies as we change x while keeping y and z fixed.

We can just imagine that y and z are fixed parameters, so the only variable is x.

The derivative of V is then defined by

$$\frac{dV}{dx} = \lim_{\Delta x \to 0} \frac{\Delta V}{\Delta x} \tag{1}$$

where ΔV is defined by

$$\Delta V = V(|x + \Delta x|, y, z) - V(x, y, z) \tag{2}$$

Note that in the definition of Δ V, only x has been shifted; y and z are kept fixed.

The derivative defined by Eq. (1) and Eq. (2) is called the *partial derivative of V* with respect to x and is denoted

$$\frac{\partial V}{\partial x}$$

or, when we want to emphasize that y and z are kept fixed,

$$\left(\frac{\partial V}{\partial x}\right)_{y,z}$$

By the same method we can construct the partial derivative with respect to either of the other variables:

$$\frac{\partial V}{\partial y} = \lim_{\Delta y \to 0} \frac{\Delta V}{\Delta y}$$

A shorthand notation for the partial derivatives of V with respect to y is $\frac{\partial V}{\partial y} = \partial_y V$

Multiple derivatives are also possible.

If we think of $\frac{\partial V}{\partial y}$ as itself being a function of x, y, z, then it can be differentiated.

Thus we can define the second-order partial derivative with respect to x:

$$\frac{\partial^2 V}{\partial x^2} = \partial_x \left(\frac{\partial V}{\partial x} \right) = \partial_{x,x} V$$

Mixed partial derivatives also make sense.

For example, one can differentiate $\partial_y V$ with respect to x:

$$\frac{\partial^2 V}{\partial x \partial y} = \partial_x \left(\frac{\partial V}{\partial y} \right) = \partial_{x,y} V$$

It's an interesting and important fact that the mixed derivatives do not depend on the order in which the derivatives are carried out.

In other words,

$$\frac{\partial^2 V}{\partial x \partial y} = \frac{\partial^2 V}{\partial y \partial x}$$

Exercise 1: Compute all first and second partial derivatives—including mixed derivatives—of the following functions. I hope you have done this!!

$$x^2 + y^2 = \sin(xy) \qquad \frac{x}{y}e^{x^2+y^2} \qquad e^x \cos y$$

Stationary Points and Minimizing Functions

Let's look at a function of y that we will call F (see Figure 1).

Notice that there are places on the curve where a shift in y in either direction produces only an upward shift in F.

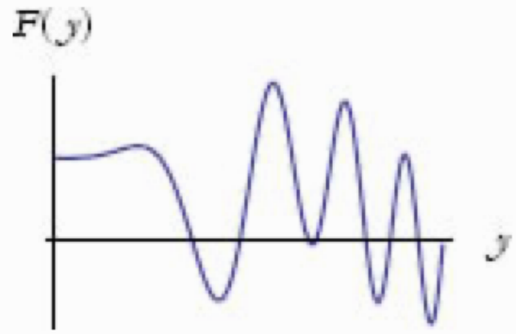


Figure 1: Plot of the function $F(y)$.

These points are called *local minima*.

In Figure 2 we have added dots to indicate the local minima.

For each local minimum, when you go in either direction along y, you begin to rise above the dot in F(y).

Each dot is at the bottom of a little depression.

The *global minimum* is the lowest possible place on the curve.

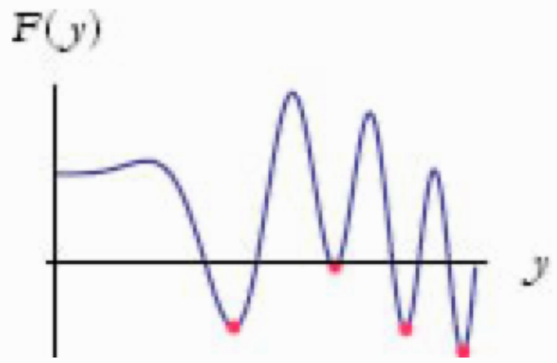


Figure 2: Local minima.

One condition for a local minimum is that the derivative of the function with respect to the independent variable at that point is zero.

This is a necessary condition, but not a sufficient condition.

This condition defines any *stationary point*,

$$\frac{d}{dy}F(y) = 0$$

The second condition tests to see what the character of the stationary point is by examining its second derivative.

If the second derivative is larger than 0, then all points nearby will be above the stationary point, and we have a *local minimum*:

$$\frac{d^2}{dy^2}F(y) > 0$$

If the second derivative is less than 0, then all points nearby will be below the stationary point, and we have a *local maximum*:

$$\frac{d^2}{dy^2}F(y) < 0$$

See Figure 3 for examples of local maxima.

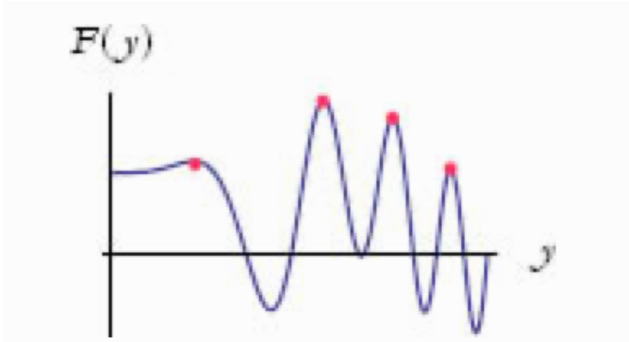


Figure 3: Local maxima.

If the second derivative is equal to 0, then the derivative changes from positive to negative at the stationary point, which we call a *point of inflection*:

$$\frac{d^2}{dy^2}F(y) = 0$$

See Figure 4 for an example of a point of inflection.

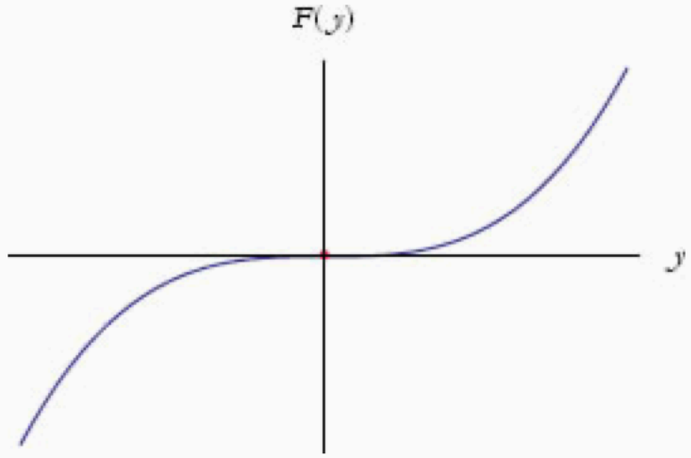


Figure 4: Point of inflection.

These are collectively the results of a *second-derivative test*.

Stationary Points in Higher Dimensions

Local maxima, local minima, and other stationary points can happen for functions of more than one variable.

Imagine a hilly terrain.

The altitude is a function that depends on the two coordinates—let's say latitude and longitude.

Call it A(x, y).

The tops of hills and the bottoms of valleys are local maxima and minima of A(x, y).

But they are not the only places where the terrain is locally horizontal.

Saddle points occur between two hills.

You can see some examples in Figure 5.

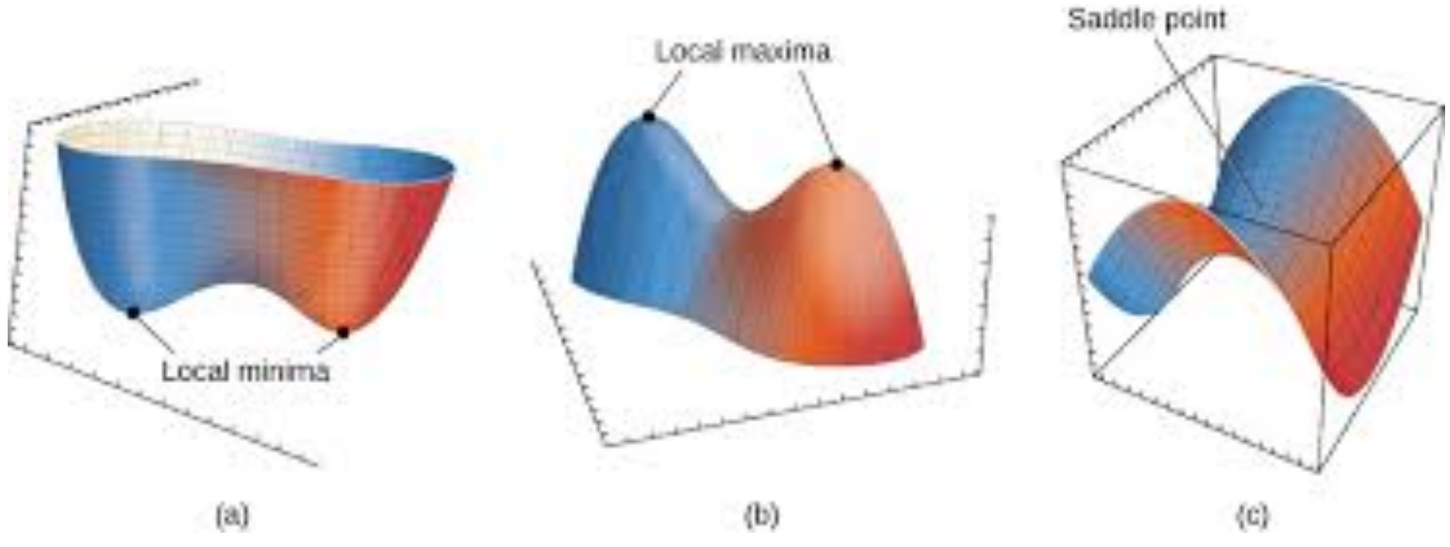


Figure 5: A function of several variables.

The very tops of hills are places where no matter which way you move, you soon go down.

Valley bottoms are the opposite; all directions lead up.

But both are places where the ground is level.

There are other places where the ground is level.

Between two hills you can find places called saddles.

Saddle points are level, but along one axis the altitude quickly increases in either direction.

Along another perpendicular direction the altitude decreases.

All of these are called stationary points.

Let's take a slice along the x axis through our space so that the slice passes through a local maximum of A, see Figure 6.

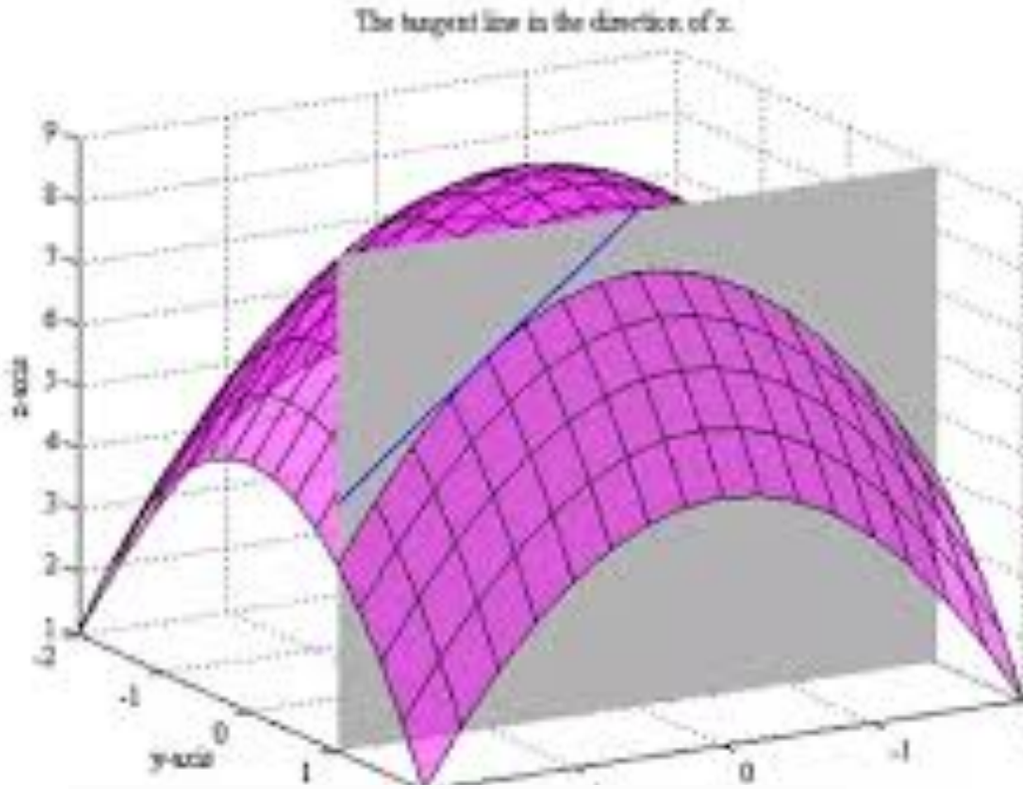


Figure 6: A slice along the *x* axis.

It's apparent that at the maximum, the derivative of A with respect to x vanishes, we write this:

$$\frac{\partial A}{\partial x} = 0$$

On the other hand, the slice could have been oriented along the y axis, and we would then conclude that

$$\frac{\partial A}{\partial y} = 0$$

To have a minimum, or for that matter to have any stationary point, both derivatives must vanish.

If there were more directions of space in which A could vary, then the condition for a stationary point is given by:

$$\frac{\partial A}{\partial x_i} = 0 \tag{3}$$

for all x_i.

There is a shorthand for summarizing these equations.

Recall that the change in a function when the point x is varied a little bit is given by

$$\delta A = \sum_i \frac{\partial A}{\partial x_i} \delta x_i$$

The set of Equations (3) are equivalent to the condition that

$$\delta A = 0 \tag{4}$$

for any small variation of x.

Suppose we found such a point.

How do we tell whether it is a maximum, a minimum, or a saddle.

The answer is a generalization of the criterion for a single variable.

We look at the second derivatives.

But there are several second derivatives.

For the case of two dimensions, we have

$$\frac{\partial^2 A}{\partial x^2} \quad , \quad \frac{\partial^2 A}{\partial y^2} \quad , \quad \frac{\partial^2 A}{\partial x \partial y} \quad \text{and} \quad \frac{\partial^2 A}{\partial y \partial x}$$

the last two being the same.

These partial derivatives are often arranged into a special matrix called the *Hessian matrix*.

$$H = \begin{pmatrix} \frac{\partial^2 A}{\partial x^2} & \frac{\partial^2 A}{\partial x \partial y} \\ \frac{\partial^2 A}{\partial y \partial x} & \frac{\partial^2 A}{\partial y^2} \end{pmatrix}$$

Important quantities, called the determinant and the trace, can be made out of such a matrix.

The determinant is given by

$$\det H = \frac{\partial^2 A}{\partial x^2}\frac{\partial^2 A}{\partial y^2} - \frac{\partial^2 A}{\partial y \partial x}\frac{\partial^2 A}{\partial x \partial y}$$

and the trace is given by

$$Tr\, H = \frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2}$$

Matrices, determinants, and traces may not mean much to you beyond these definitions, but they will if you go on to quantum mechanics.

For now, all you need is the definitions and the following rules.

If the determinant and the trace of the Hessian are positive then the point is a local minimum.

If the determinant is positive and the trace negative the point is a local maximum.

If the determinant is negative, then irrespective of the trace, the point is a saddle point.

However: One caveat, these rules specifically apply to functions of two variables.

Beyond that, the rules are more complicated.

None of this is obvious for now, but it still enables you to test various functions and find their different stationary points.

Let's take an example. Consider

$$F(x, y) = \sin x + \sin y$$

Differentiating, we get

$$\frac{\partial F}{\partial x} = \cos x$$

$$\frac{\partial F}{\partial y} = \cos y$$

Take the point $x = \frac{\pi}{2}$, $y = \frac{\pi}{2}$.

Since $\cos \frac{\pi}{2} = 0$, both derivatives are zero and the point is a stationary point.

Now, to find the type of stationary point, compute the second derivatives.

The second derivatives are

$$\frac{\partial^2 F}{\partial x^2} = -\sin x \qquad \frac{\partial^2 F}{\partial y^2} = -\sin y \qquad \frac{\partial^2 F}{\partial x \partial y} = \frac{\partial^2 F}{\partial y \partial x} = 0$$

Since $\sin\frac{\pi}{2} = 1$ we see that both the determinant is positive,

$$\det H = \frac{\partial^2 A}{\partial x^2}\frac{\partial^2 A}{\partial y^2} - \frac{\partial^2 A}{\partial y \partial x}\frac{\partial^2 A}{\partial x \partial y}$$

$$= (-\sin x)(-\sin y)$$

$$= \sin x \sin y$$

and the trace of the Hessian is negative.

$$Tr\, H = \frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2}$$

$$= -\sin x - \sin y$$

So the trace at the point $x = \pi/2$, is

$$Tr\, H = -1 - 1 = -2$$

thus we have a positive determinant and a negative trace, the point is a maximum.

Now to continue……

So we have

$$\frac{\partial^2 F}{\partial x\ \partial y} = \frac{\partial^2 F}{\partial y\ \partial x} \tag{45}$$

And that translates into the statement that

$$\frac{\partial F_y}{\partial x} = \frac{\partial F_x}{\partial y} \tag{46}$$

When we look at an expression like the right-hand side of equation (43), where someone has given us a function denoted F_x and another one denoted F_y, there is clearly no guarantee that equation (46) holds.

Or said another way, given any old functions $G(x,y)$ and $H(x,y)$, clearly in general *it won't be true* that

$$\frac{\partial H}{\partial x} = \frac{\partial G}{\partial y} \tag{47}$$

The condition that G and H have to satisfy, to be respectively the partial derivative with respect to x and the partial derivative with respect to y of a unique function F, is precisely equation (47).

We will see some examples in a moment.

Now take F_x and F_y which are indeed partial derivatives of F with respect to x and with respect to y.

They can be seen as the components of a vector.

This vector has a name.

It is the gradient of F.

If you like to think of F as a scalar field, then the gradient is a vector field, derived from the scalar field.

And if these F_x and F_y are the components of a gradient, what is the statement that $\partial F_y/\partial x = \partial F_x/\partial y$?

Let's write it this way

$$\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} = 0 \tag{48}$$

Notice that left-hand side is a *curl*.

It is the curl of (F_x,F_y).

i.e., for any vector field **A** we define a new vector field **B**, that is called the curl of **A**, as

$$\mathbf{B} = \nabla \times \mathbf{A}$$

Whether they come from a gradient, or are more general functions like G and H, doesn't matter.

It is true in general.

In this case (F_x, F_y) or (G, H) is only two dimensional.

If we had three functions F_x, F_y and F_z, or generally speaking (G, H, K), of x, y and z, then the left-hand side of equation (48) would be only the third component of the curl of (F_x, F_y, F_z), which would be a vector.

There would be other components of the curl.

Now to answer the question

"Are F_x and F_y really the gradient of a function F ?"

The test can be reformulated :

"Is the curl equal to zero or not ?" $$\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \tag{49}$$

There is another implication of the curl being equal to zero. It is intuitively obvious.

Suppose we take a walk on the hilly terrain of figure 4 or figure 5 and we come back first of all to the same point.

How much does F change ?

It doesn't change at all.

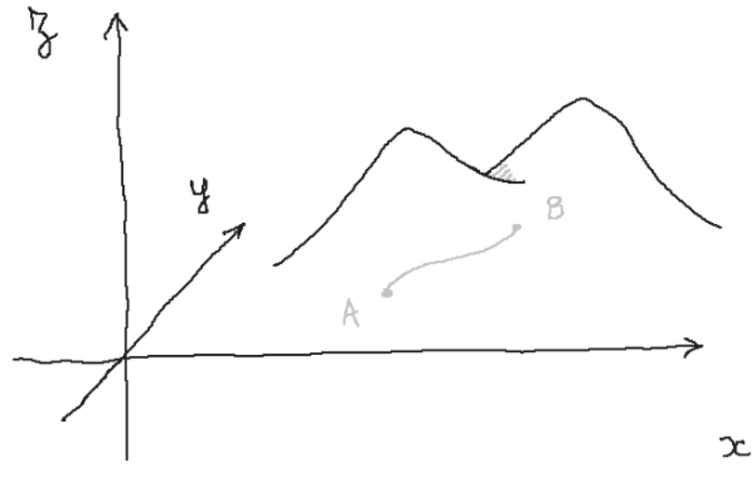


Figure 5 : Walk on the surface of F .

Now let's ask a slightly different question :

Supposing we take a walk from point A to point B, figure 5, how much does F change?

Well, one answer is just the difference of F at the end point and at the starting point of the walk, that is $F(B_x,B_y)-F(A_x,A_y)$ with obvious, although heavy handed, notations.

But another answer is to add up all of the little incremental changes in F as we move from one point to another.

This can be written

$$\Delta F = \int F_x dx + F_y dy \tag{50}$$

This integral, calculated along the curve between A and B in figure 5, is the change in F .

Now what would happen if we chose another path to go from A to B ?

Would we get the same answer for the change of F ?

Yes we would, because F is just the altitude.

If we go from A to B on one path or on another path, the change in the altitude is the same.

So that says that the line integral of equation (50) doesn't depend on what path we take.

Is it true in general for any F_x and F_y ?

In other words, if we take any function $G(x, y)$ multiplying dx and any function $H(x,y)$ multiplying dy, is the line integral

$$\int G dx + H dy \tag{51}$$

independent of the path used to calculate the integral ?

You have probably already figured out the answer.

But before we explain it, let's look at an example.

It is sort of a silly example but it is a concrete one.

We are going to do this operation by driving our car from one point to another along two different trajectories.

And what we are interested in is not the change in the altitude but the change in the reading of the gasoline gauge.

We want to know how much gasoline we use.

Does the amount of gasoline we use depend on the route that we take?

Of course it does.

For example, suppose A and B are at the same altitude.

If we take a route from A to B that keeps us at the same altitude at all times, we will use less gasoline than if we go over a hill to reach B from A.

Another way to say this, incidentally, is that if we are going around in a closed loop, starting from one place and coming back to the same place, the net altitude won't change.

But what about the gasoline in the tank?

Not only will it depend on the route we take, but it certainly won't come back to the same value afterwards.

Nevertheless, the change in the fuel – now F stands for fuel (take this with a grain of salt, because now F is longer a function of (x, y). It is the reading of the gasoline gauge)– can be written in the form of expression (51).

From any point (x,y), how much fuel we will use to go a little distance dx and then how much fuel we will use to go a little distance dy on the terrain?

It depends on the point (x, y) and can be written as a differential

$$G(x,y)\ dx + H(x,y)\ dy \tag{52}$$

But it is not the differential of some function.

If it were the differential of a function F, as in equation (43), then going around in a loop the total change would always have to be zero.

Another example, not using a car but walking, would be "how tired do we get, going from A to B ?"

Again, it would depend on the path used.

We can even make the approximation for simplicity, if we like, that we don't get tired at all if we don't change altitude along the path.

Then there would be some paths going from *A* to *A* consuming no energy.

Yet if we climbed over a hill before coming back to the same place, we would obviously get tired.

As all of us know, not only going uphill but going downhill is tiring.

Of course, if we could somehow get "untired" by going downhill, then maybe it would be possible to go around a loop, whatever path we take, and come back without net extra fatigue.

That is to some extent how electric bicycles work.

But then the measure of tiredness would be nothing more than the measure of potential energy according to altitude.

And the potential energy is simply proportional to altitude.

So we would be back in the case of looking at the function *F* in figure 5.

Thus in general, when $G(x, y)$ and $H(x, y)$ are not constrained in anyway, going around in a closed loop we do get a change.

The line integral $\int G dx + H dy$ is not zero.

Furthermore it depends on the path followed.

There is a **complete mathematical equivalent** to the question of whether line integrals around closed paths give zero.

It is the question of whether the curl of *(G, H)* is equal to 0.

The condition that, for all *(x, y)*, the following curl

$$\frac{\partial H}{\partial x} - \frac{\partial G}{\partial y} = 0 \tag{53}$$

is equivalent to the condition that, for any closed loop,

$$\int_{\text{loop}} G dx + H dy = 0 \tag{54}$$

They are the same mathematical statement.

And they are equivalent to the statement : G and H are respectively the partial derivative with respect to x and the partial derivative with respect to y of some function F of x and y.

In fact the curl on the left-hand side of equation (53) has to do with going around a *little loop*, figure 6, and asking how much things change.

The change is the curl itself.

And the line integral on the left-hand side of equation (54) has to do with going around a big loop.

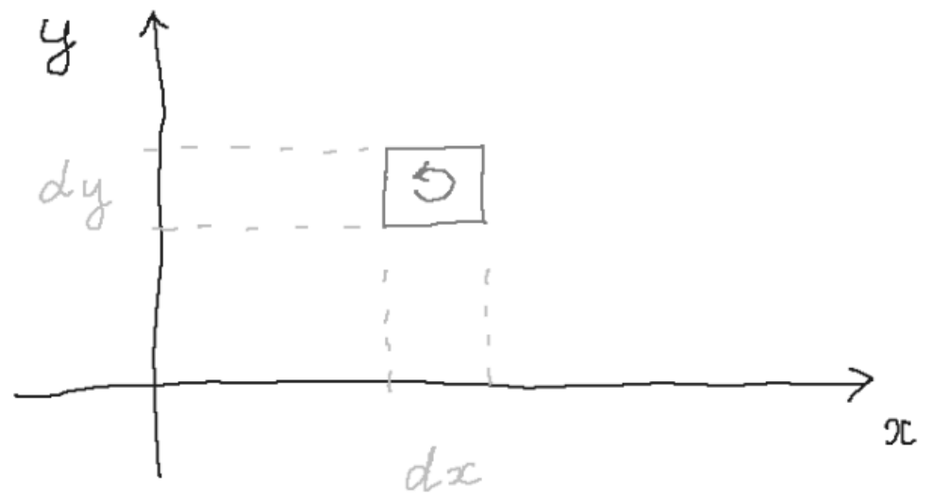


Figure 6 : Calculating the curl of (G,H) is like going around a little loop and calculating the total change in the differential given by expression (52).

The whole point of this section is that not every pair of functions F_x and F_y – which we prefer to denote then G and H – enable us to define the differential of a function F .

They can always be used to define the following differential element

$$G\ dx + H\ dy \tag{55}$$

but that differential is not necessarily the differential of a function.

When G and H are the partial derivatives F_x and F_y of a function F , and therefore expression (55) is the differential of a function F, it is called an *exact differential*.

It is also sometimes called a total differential.

And when the curl of *(G, H)* is not zero, that is when G and H are not partial derivatives of a function F, with respect to x and y, then the differential given by expression (53) is called an *inexact differential*.

Question: It this related to conservative forces ?

Answer: When F_x and F_y are the components of a force field, yes, the condition that their curl be zero is equivalent to the conservativeness of the force field.

But we don't want to get into forces here.

We want to concentrate on the idea that some differentials are exact and others are not.

Question: Is this related to curvature ?

Answer: Yes, it is.

But we don't want to get into curvature here either.

Again we are only interested in some differential calculus concepts that we shall use in statistical mechanics, when we study heat and work.

So some $G(x, y)\,dx + H(x, y)\,dy$ are exact differentials of a function F.

In that case they pass the curl test – the test being "Is the curl of (G, H) equal to zero ?", equation (53).

Others are just small changes involving two functions, $G(x, y)$ multiplying dx and $H(x, y)$ multiplying dy.

The gasoline in the car is a well-defined quantity.

You may be at some point (x, y) and read a certain value of the gasoline gauge.

Then if you move on the hilly terrain by a small dx and a small dy (when projected onto the horizontal plane), the fuel gauge will change by the amount $G(x,y)\,dx + H(x,y)\,dy$.

Yet, the reading of the gauge doesn't depend on (x, y).

You may come back, after a closed trip, to the same point (x,y).

The gauge will read a different value.

The gasoline in the car is a thing you carry with you.

When you move along a given path, it will change in a definite way, depending on the path.

The reading of the gasoline gauge is simply not a function of where you are.

It is a function of the whole path which you took to get you where you are.

Inexact differentials usually have to do with changes that depend not only on the end points of a trajectory, but depend on the whole trajectory, like getting tired or the amount of fuel in the tank.

We now know what exact and inexact mean.

Let's see some examples.

To start with let's try for instance

$$F_x = y$$
$$F_y = x \tag{56}$$

where we could have written $H(x, y) = y$, and $G(x, y) = x$, because we don't mean that necessarily F_x and F_y are partial derivatives.

The point is precisely to check whether they are the partial derivative with respect to x and the partial derivative with respect to y of a function F.

So let's do the curl test.

What is the partial derivative of F_x with respect to y ? It is 1.

And what is the partial derivative of F_y with respect to x? It is 1 too.

So we do have

$$\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} = 0 \tag{57}$$

The two functions defined by equations (56) pass the curl test.

It must mean that F_x and F_y are the partial derivatives of some function.

Can you guess what function, F_x and F_y are the gradient of ?

In this example F is the function xy, or the function xy plus any constant.

So (F_x, F_y) is really the gradient of some function.

And the line integral from A to B

$$\int_A^B ydx + xdy \tag{58}$$

doesn't depend of the path taken.

Second example : What about the functions F_x and F_y defined below ?

$$F_x = y$$
$$F_y = -x \tag{59}$$

It would be a perfectly good rule about the small differentials of the amount of gasoline in your car.

But it does not define a function.

How do we know that?

Because it doesn't pass the curl test.

If we took a little excursion on the horizontal plane coming back to the same point, and we calculated the line integral

$$\int_{loop} ydx - xdy \tag{60}$$

we would find that it is not zero in general.

Moreover we could not look at a corresponding round trip on a surface like in figure 5, because there is no such surface.

Now we are ready to talk about heat and work.

6.6 Heat and work

Heat and work have to do with energy.

And they have to do with the changes in energy.

So let's define them.

For convenience we are thinking about a gas, because it is easy to imagine the particles, their positions and movements, the possible states, the velocities, the energy, the volume, the pressure, the thermodynamical state of the system, and so forth.

But it doesn't have to be a gas.

It could be a liquid, a solid, or other things like a plasma, etc

So we consider the time-honored box of gas with a piston, figure 7.

We can push the piston or pull it.

It allows us to change the volume V of the gas, which is a control parameter.

What are the independent variables?

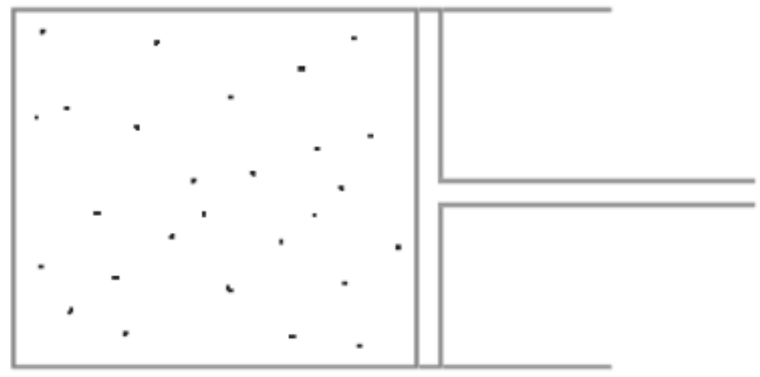


Figure 7 : Box of gas with a piston.

There are two independent variables.

We could take them to be the temperature and the volume.

Or we could take them to be the energy of the box and the volume.

In fact, besides volume, we could take any one among the variables Z, E, S, β, T or A.

Take this opportunity to make sure you know what each of them are, how they are calculated, what are their basic properties.

Any pair X and V is enough to determine the thermodynamic state of the gas.

This is not to be confused with the random states ω_i's.

What we call the thermodynamic state of the system is for instance its average energy E, and its volume.

It determines everything else if we look not at the molecular level but at a global level, i.e. at the human scale.

We are going to pick S and V as the independent variables.

Knowing how much entropy the box presents, and how much volume it has, completely determines the state of the system.

Let's think about the energy.

Consider the change in energy if we change the volume a little bit without changing the entropy.

Do you remember how we call a process in which the entropy doesn't change? An *adiabatic process*.

It is also called a *reversible process*.

We will study the link between entropy and reversibility in part 7 of our discussion.

In an adiabatic process the following relationship holds

$$dE = -PdV \tag{61}$$

Remember that it is even how we defined pressure in general, see equation (22) of part 5.

Formula (61) is the change in the energy of the gas if we change the volume slowly and without allowing any heat to flow in or out, that is with no change of entropy – reversibly if you like.

A more explicit way to write formula (61), stressing that P is the rate of variation of E with respect to V at fixed entropy, is

$$\left.\frac{\partial E}{\partial V}\right|_S = -P \tag{62}$$

Now what about the change in energy if we keep the volume fixed but we add a little bit of heat ?

It doesn't matter where that extra energy comes from.

It just doesn't come from compressing the gas with the piston, since we keep V fixed.

We have got it from somewhere and we dump it into the system in whatever way.

We could do it by putting a flame under the box and heating it a little bit.

Then what is the change in energy ?

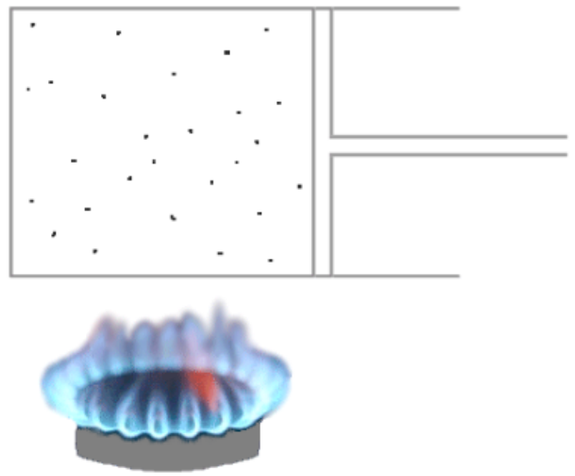


Figure 8 : Heating the gas at fixed volume.

In that case without changing the volume, it is equal to the temperature times the change in the entropy

$$dE = TdS \tag{63}$$

That is basically the definition of temperature.

Keeping all control parameters fixed, the change of energy is by definition the temperature times of change in entropy.

Entropy we defined independently, probabilistically.

And we showed that the temperature defined by equation (63) is also $1/\beta$, where β was one of the two Lagrange multipliers in the problem of maximizing the entropy, see part 3.

Now supposing we change both the volume and the entropy.

We could imagine the process as a sequence of little changes, each of them being a little change of the volume at fixed entropy followed by a little change of the entropy at fixed volume.

Along such a process, according to the first law of thermodynamics, we have for each little change of E

$$dE = -PdV + TdS \tag{64}$$

So we can do a general process involving both independent variables V and S.

Equation (64) gives the incremental change in E along the path we follow in the *(V, S)* space.

The equation giving dE just expresses energy conservation: if we inject a little bit of energy in the form of work and of eat, the sum is the change of energy in the box.

Incidentally $-PdV$ is the work part.

It is called dW.

And $T\,dS$ is the heat part.

It is called dQ.

$$dE = dW + dQ \tag{65}$$

Q stands for heat.

Where does the Q come from ?

Lots of possibilities.

At any rate, $T\,dS = dQ$.

And that is the definition of heat.

We can change the energy by changing the entropy, at fixed volume.

That is an irreversible change.

It could correspond to just dumping some heat in.

And $-P\,dV = dW$.

It is the change in energy from changing the volume at fixed entropy.

Who is doing work on whom incidentally ?

Is dW the work done on the gas, or is it the work done by the gas on the piston?

Answer : dW is the change in the energy of the box.

It is the work received by the gas.

Therefore it is the work done by the piston on the gas.

Concerning equation (65), the question is whether Q is really a function.

Or in other words, is dQ really an exact differential ?

Is there such a thing as the heat of a system ?

Let's stress first of all that if we know the volume and the entropy, we know everything about the system.

As said, volume and entropy is thermodynamically enough to know everything.

It tells us the temperature, it tells us the energy E, it tells us everything.

Now here is the problem.

If we think of going from one value of volume and entropy to another, taking a walk in a volume-entropy space, the question is whether the heat that we have to put in to go from one point to another depends on the path or doesn't depend on the path.

Another way of saying it is this : if we go around a closed loop – we put little bits of heat in, we put little bits of work in, we take little bits of work out, we take little bits of heat out –, at the end of the day, if we come back to the same point, is the total amount of heat that we put in 0 ?

Is it zero as it would be if the heat were a real function and we came back to the same point, or does it depend on the path that we took ?

The answer is : *it depends on the path*.

We are going to prove that.

We are going to prove that dQ is not an exact differential.

Neither is dW .

What is true is that, if we take a system through a cycle, or a series of steps, and come back to the same state, then the total energy that we put into it is 0.

If we bring it around and come back to the same point in the *(V,S)* space, the energy will come back to the same value. E is a genuine function of V and S.

But the heat that we put in, all the work that we take out, or whatever, will depend on the route that we took.

We are assuming that the path that we follow, whichever it is, is covered slowly, so that we stay in equilibrium at all times.

Let's prove the path-dependency of Q and W.

Remember that the independent variables we chose are the entropy S and the volume V .

First we rewrite equation (65) as

$$dQ = dE + PdV \tag{66}$$

Now we are going to change the independent variables from S and V to E and V.

However we move from one point to another, dQ will be equal to the small change in energy plus the pressure times the small change in volume.

Now let's assume that Q is a *real function of E and V* .

That would imply that dQ is an exact differential.

First of all, equation (66) would imply that

$$\frac{\partial Q}{\partial E} = 1 \tag{67}$$

Similarly

$$\frac{\partial Q}{\partial V} = P \tag{68}$$

That is what would hold if there really was a function Q the differential of which was equation (66).

Are equations (67) and (68) consistent ?

How can we test out whether there really is a function Q that satisfies this ?

Answer : we can test the curl.

If there is function Q, the curl of Q_E and Q_V must be zero.

Remember Q_E is the notation for $\partial Q/\partial E$ and Q_V is the notation for $\partial Q/\partial V$.

That is the application to Q, E and V of what we established in general in the previous section for F, x and y, see equation (48) from earlier in this part.

So we want to calculate

$$\frac{\partial}{\partial V}\left(\frac{\partial Q}{\partial E}\right) \tag{69}$$

and

$$\frac{\partial}{\partial E}\left(\frac{\partial Q}{\partial V}\right) \tag{70}$$

and compare them.

Expression (69) is equal to 0, while expression (70) is equal to $\partial P/\partial E$.

If there really was a function Q, in other words a function that didn't depend on how you move from one place to another, it would have to say that

$$\frac{\partial P}{\partial E} = 0 \tag{71}$$

That is what would be true if the curl of Q_{E} and Q_{V} was equal to zero.

This equation (71) says that the partial derivative of P with respect to E when we keep the volume V constant is zero.

Would you believe that if you have a box of gas and you change the energy in the box of gas, that the pressure doesn't change ?

No, of course it does.

Let's check it for an ideal gas.

We have done the calculation for an ideal gas.

We found for the pressure

$$P = \frac{NT}{V} \tag{72}$$

That is one equation.

And the other equation we want is the equation for the energy

$$E = \frac{3}{2}\, NT \tag{73}$$

So we can write

$$P = \frac{2}{3}\,\frac{E}{V} \tag{74}$$

In view of this equation for P in terms of the independent variables E and V , when we change the energy at fixed volume does the pressure change?

The answer is yes.

At fixed volume, the pressure is proportional to the energy.

So equation (71) is simply not true.

Therefore the curl test has failed, since following a reductio ad absurdum we saw that assuming that the curl was zero lead to a false statement.

Consequently,

There is no such thing as Q being a function of where we are, or what the thermodynamic state of the system is.

It depends on how we move from one place to another in the space *(E, V)*, or equivalently in the space *(S, V)*, or *(T, V)*, etc.

In other words,

> *The amount of heat that we have to put into the system, in order to change it from one state to another state, depends on the path that we are going to take.*

That is why heat is not a good description of the system.

It is as good as asking how tired you are to determine your altitude.

It depends on the path you took, not just on your change of altitude.

Heat is also like that.

And so is work.

Only the combination of the two of them is really a function of the state of the system.

A function of the state of the system means is a function of temperature and volume, or energy and volume, i.e., something that is determined by the state the system is in.

This is subtle stuff.

It is tricky.

It is natural to be a bit confused by it at first.

Only by going over it, studying it, asking yourself many small questions and answering them, will it becomes clear.

Why should it be the case that the heat that we have to put in, versus the work, to get from one place to another should not depend on how we go there ?

This fact, that the heat Q and the work W are not functions simply depending on the state of the system, is related to all sorts of other physical concepts and facts : the operation of steam engines, efficiency, hysteresis, memory in systems, etc.

At the end of this section on heat and work, the point to remember is that *heat is not a property of the system*, that would depend only on T and V for instance, figure 9.

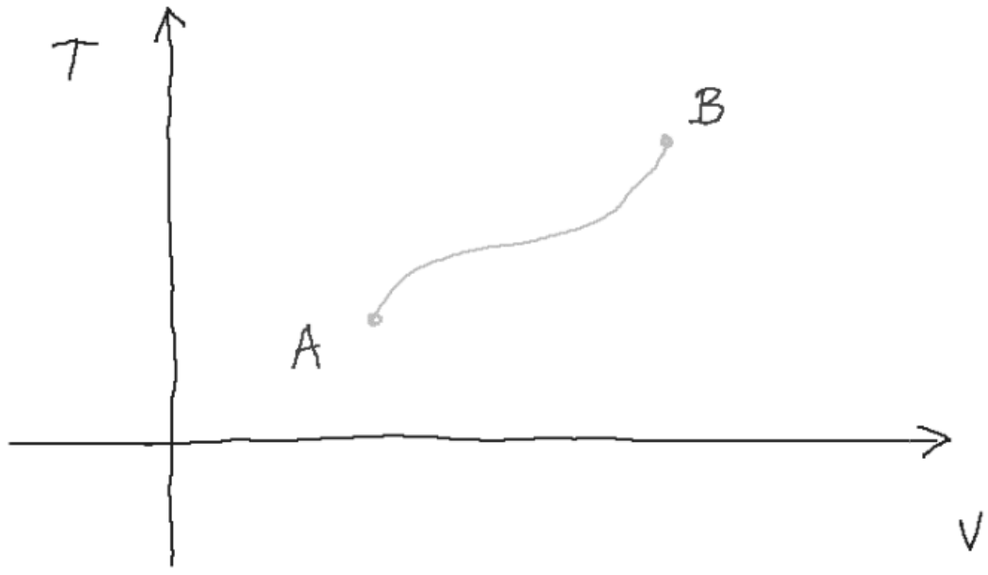


Figure 9 : Path in the (T, V) space.

The total energy E of the system depends only on where we are.

But ΔQ and ΔW, when we go from A to B, depend on the the path we follow.

It is similar to the following fact : Our state of tiredness after a trek is not a property of the landscape – in the sense that it is not a property of where we stand –, it is a property of us and the trek we went over.

Let's conclude the part of our discussion with this summary.

If we have a system in a state A at the beginning of an experiment, and in a state B at the end of the experiment, and in each of these two states we know everything about the system, we certainly can know the variation in its energy.

It is simply $E_B - E_A$.

It is naturally the integral over the path we followed of dE.

But, if we don't know the actual path we followed, if we just know A and B, then we cannot know how much heat ΔQ we put into the system and how much work ΔW we put in it.

We only know that

$$\Delta E = \Delta Q + \Delta W \tag{75}$$

That is what makes heat and work more subtle than simply energy.

7. Entropy versus reversibility

7.1 Introduction

In this part of our discussion, we will eventually get to the second law of thermodynamics, what the second law means, how it is consistent.

The subject is complex.

To complete the main text, there will be several questions/answers parts to help with understanding.

Before reaching the second law, however, we shall pursue the study of some physical examples which we have begun to examine with the tools of statistical mechanics.

First, we shall illustrate the physics of gases with some numerical data.

We shall study the speed of sound in a gas, and in particular in a fairly dilute gas.

As a second preliminary example, we shall study the behavior of a harmonic oscillator in a heat bath.

7.2 Speed of sound in a gas

What is the speed of sound?

There are two formulas for the speed of sound.

But if you think about it for a moment, and you start with a very dilute gas, what would you expect the speed of sound to be?

Let's say we make a little bit of overdense(excess) region.

How fast does that overdensity spread?

That is the speed of sound.

What would you guess?

Well, it sure isn't going to be faster than the time that it takes for a molecule to travel out to a certain distance.

In fact, for a very dilute gas, we might expect that that a little overdense region just spreads out with a velocity that is not too different than the average velocity of the molecules.

How does it spread out?

It spreads out by the molecules moving out.

How fast do the molecules move out?

They move out with their velocity.

What is their velocity?

Now, we have to work a little bit to find out what the velocity is.

So let's write down a formula.

In thermal equilibrium at a temperature T, in a dilute gas, every molecule has an energy

$$E_{part} = \frac{3}{2} k_B T \tag{1}$$

We are going to put the k_B in now, because we want to do some calculations in laboratory units.

So the temperature is measured in Kelvin.

Kelvin means relative to absolute zero, and 1 unit corresponds to 1 degree centigrade.

That is 9/5 of a Fahrenheit degree.

From formula (1), we can see that the temperature goes to zero when the energy E_{part} of one molecule goes to zero.

That energy must be the kinetic energy of the molecule.

It is the only energy that a molecule has for an ideal gas in a big box.

So we have got to have

$$\frac{3}{2}k_B T = \frac{1}{2}mv^2 \tag{2}$$

where v is the average velocity of one molecule.

So equivalently

$$v^2 = 3\,\frac{k_B T}{m} \tag{3}$$

You might expect that this is more or less the speed of sound in a gas.

And actually it is not too far off.

There is a more exact formula.

It is established by actually studying the mechanics of lumps of gas and the forces on them.

We are not going to do it in this discussion.

Newton did it.

It is not too hard.

You take a box of gas.

You take a cubic region in the gas.

And you ask:

What forces are around it?

The forces have to do with the pressure from the outside of the region .[1]

Now we do continuum mechanics.

The gas is taken to be a continuum.

There is a gradient of pressure.

There is some force due to the gradient of pressure.

What is the response to the variation in pressure?

It is F = ma, where a is the acceleration. So the box starts to move.

The acceleration depends on the mass in the box.

So the mass density comes into play.

But in the end you find a formula.

Newton's formula is

$$C^2 = \frac{1}{m}\frac{\partial P}{\partial \rho} \tag{4}$$

We write C instead of v because this is a fairly general formula.

It is more general than just for the ideal gas.

C^2 is the partial derivative of the pressure with respect to the mass density.

And since we used ϱ to denote the number of particles per unit volume, the mass density is $m\varrho$ where m is the mass of one particle.

Hence formula (4).

Let's see what formula (4) yields for the ideal gas.

We know that for the ideal gas we have

$$PV = Nk_BT \tag{5}$$

or

$$P = \rho\ k_BT \tag{6}$$

The derivative of P with respect to ϱ is

$$\frac{\partial P}{\partial \rho} = k_BT \tag{7}$$

And so, dividing by the mass of one particle, we get

$$C^2 = \frac{k_BT}{m} \tag{8}$$

It is not quite the same formula as formula (3).

The simplistic formula (3), when we take the square root to compute v, is too high by a factor of $\sqrt{3}$.

That is approximately 1.7.

So it is not so bad.

In order of magnitude, it is correct.

Let's work it out for the air.

We shall stick some numbers into formula (3).

The Boltzmann constant is

$$k_B = 1.3 \text{ x } 10^{-23} \tag{9}$$

Air is mainly made of nitrogen. So we shall take for m the mass of a nitrogen molecule.

Gas nitrogen has a diatomic molecule.

The atom of nitrogen has atomic number 7 and atomic weight 14. (There are 7 protons and 7 neutrons in the nucleus of the common isotope.)

We shall take

$$\begin{aligned} m &= 28 \text{ x mass of a proton} \\ &= 28 \text{ x } 1.7 \text{ x } 10^{-27}\ kg \\ &= 5 \text{ x } 10^{-26}\ kg \end{aligned} \tag{10}$$

And we need to know room temperature.

$$T = 300\ K \tag{11}$$

in the same units where k_B as the value given above.

Using formula(3), we get

$$\begin{aligned} v^2 &= 3 \text{ x } 1.3 \text{ x } 10^{-23} \text{ x } 300 \text{ x } 0.2 \text{ x } 10^{26} \\ &= 23.4 \text{ x } 10^4 \end{aligned} \tag{12}$$

Therefore, in round numbers,

$$v = 500 \ m \ s^{-1} \tag{13}$$

This is the correct value for the speed of the nitrogen molecules in the air at room temperature.

As an estimate of the value for the speed of sound, as expected, it is a bit on the high side.

The right answer, in the air, is about 340 meters per second.

7.3 Questions/answers

Question: In assimilating formula (3) to the square of the speed of sound, we also assume that when a sound wave moves through the gas, the temperature is constant, don't we?

Answer: Yes, we did assume that the temperature was constant.

I think it is a fairly good assumption.

For small amplitude sound waves, in a dilute gas, temperature doesn't vary very much as it vibrates.

I am not absolutely sure about it.

But I would guess that the temperature has very little variation, in a near ideal gas, for very low amplitude sound waves.

Then taking the temperature to be constant is probably OK.

Question: According to this reasoning, would not the variation of pressure also be small?

Answer: Yes the variation of pressure would also be small.

But the variation of ϱ is too.

And the rate of change in formula (4) – if that is what you are driving at – does not have necessarily to be small.

Question: If we increase the pressure, won't we expect the speed of sound to be faster?

Like in water or in a steel bar, it is much faster than in the air.

Answer: Well the $\partial P/\partial \varrho$, in formula (4), is calculated at the pressure you are working at.

$\partial P/\partial \varrho$ is not necessarily a constant.

The pressure depends on the density in a complicated way.

In particular, as the molecules get really squeezed together, the derivative of pressure with respect to ϱ gets very large.

So, the right answer is because $\partial P/\partial \varrho$ changes.

Question: Since the speed of sound is related to pressure, why doesn't pressure appear in formula (3)?

Answer: In general, the speed of sound does depend on the pressure.

If you use formula (4), $\partial P/\partial \varrho$ depends on P.

But if we are in the case of an ideal gas, the two following formulas hold

$$v^2 = 3\,\frac{k_B T}{m} \tag{14}$$

where v is the speed of a molecule, and

$$P = \rho T \tag{15}$$

The calculation yielding v = 500 m/s, was done for an ideal gas at T = 300K.

Being in the case of an ideal gas, ϱ and P are small.

Moreover, at fixed temperature they are proportional to each other.

As a consequence it eliminates them from the formula for the speed of sound, either in its simplistic formula (14) or in the more accurate formula (8).

So we can state:

In the ideal gas range, the speed of sound depends only on the temperature.

In other words, in the ideal gas range, at fixed temperature, the speed of sound doesn't depend on the pressure or the density.

But, as we increase the density and get away from the ideal gas range, formula (15) is going to break down.

For example P might start varying like ϱ^2, in which case the speed of sound will also depend on the pressure.

Then, for a given density, the speed of sound will depend on both temperature and pressure.

Question: Coming back to the idea that the speed of sound is equal to the speed of molecules, I imagine that, outside the range of the ideal gas, that idea will also break down?

Answer: It probably will. I never thought about it very much. What we studied is the depth that I have gone into it.

Let's move on to the harmonic oscillator.

We shall first study a classical harmonic oscillator, and then a quantum oscillator.

5.4 Harmonic oscillator

Now we do not have a free particle in a gas.

We are going to put into the system – the heat bath, or whatever the system happens to be – a single harmonic oscillator.

The harmonic oscillator will be part of the system.

And we are going to treat the rest of the system as a heat bath.

We shall look at the properties of that harmonic oscillator.

The harmonic oscillator under consideration could be a spring with a mass on it, figure 1, or anything else which naturally oscillates.

It could be an electromagnetic wave in a cavity.

What else oscillates?

A crystal for instance.

All sorts of oscillations take place in crystals.

There are lots and lots of oscillating systems in nature.

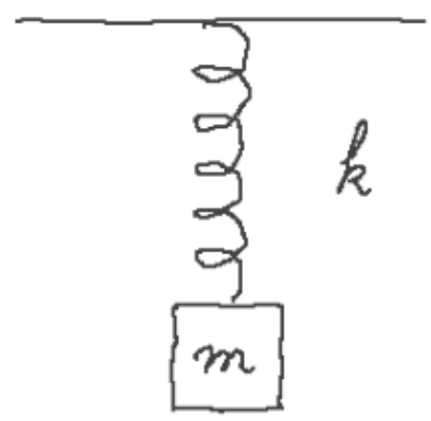


Figure 1: Harmonic oscillator.

Almost anything will oscillate if you just disturb it a little bit.

Therefore when we study the harmonic oscillator, we are studying an extremely wide class of systems.

We are studying anything which, when disturbed a little bit away from equilibrium, i.e. away from its ground state, will oscillate.

And that covers almost anything that you could think about.

So let's do the statistical mechanics of the harmonic oscillator.

An interesting question is:

What is the average energy of the oscillator?

How does it depend on its mass?

How does it depend on the spring constant?

And so forth.

Let's see if we can work out the answer to those questions.

So what do we do? We begin with a hamiltonian(the energy).

That is, we begin with an expression for the energy.

We are going to put it into the Boltzmann distribution.

And then we are going to calculate average quantities for the oscillator.

The energy of the system is its kinetic energy plus its potential energy

$$E(\dot{x}, x) = \frac{m}{2}\dot{x}^2 + \frac{k}{2}x^2 \tag{16}$$

where x is the coordinate of the oscillator away from its equilibrium position, and $\dot{x}$ the time derivative of x, that is the velocity of the little mass.

Caveat: k in equation (16) is not the Boltzmann constant, it is the Hooke's law spring constant.

We are finished with Boltzmann constant now.

It will fade in the background.

We no longer will keep track of the Boltzmann constant.

If instead of velocity we use momentum, equation (16) becomes

$$E(p,x) = \frac{p^2}{2m} + \frac{kx^2}{2} \tag{17}$$

This is our expression for energy when the system in figure 1 is in state *(p, x)*.

Now we are going to use it to calculate the partition function *Z* for the system.

The partition function *Z* is our best friend in statistical mechanics.

We use it, when we take its logarithm, to calculate all sorts of things.

In the discrete case, where the system is randomly in one of the states of $\Omega = \{\omega_1, \omega_2, \ldots \omega_i, \ldots \omega_n\}$, we saw in part 4 that

$$Z = \sum_i e^{-\beta E_i} \tag{18}$$

where E_i is the energy of the state ω_i, also called the *i*-th state.

And β is the inverse of the temperature.

Take this opportunity to refresh you memory of what is the probability that the system be in the *i*-th state.

You may also want to review how we defined or calculated α, β, the distribution formed by the P_i's, S, the average energy E, Z, T, etc.

When the set of states Ω is a continuous phase space, each state ω is labeled by two coordinates x and p.

Remember that x and p can each be of high dimensionality – although, in the present example, each is 1-dimensional.

And the energy of one state ω is denoted $E(p, x)$.

Then the partition function is

$$Z = \int e^{-\beta E(p,x)} \, dp \, dx \tag{19}$$

So let's write this partition function when the energy of the state *(p, x)* is given by equation (16):

$$Z = \int e^{-\beta \frac{p^2}{2m}} \, e^{-\beta \frac{kx^2}{2}} \, dp \, dx \tag{20}$$

As said, β is the inverse temperature.

Now, for our system, what determines the temperature?

The heat bath that it is immersed in.

So if the oscillator in figure 1 was in the room where we are, it would be 300 degrees K.

If it was in liquid helium, it would be 4 degrees.

If it was in the middle of the sun... well, it would evaporate.

But you get the point.

T or β are determined by the heat bath.

When writing the partition function Z in equation (19), we readily used the fact that E(p,x) is a sum of two terms, one depending only on p and the other depending only on x.

This enables us to write the exponential beneath the integral sign as a product of two exponentials.

Moreover, the integral in equation (19) factorizes into two integrals.

$$Z = \int e^{-\beta \frac{p^2}{2m}} \, dp \int e^{-\beta \frac{kx^2}{2}} \, dx \tag{21}$$

Now in previous situations the x-integration just gave us the volume of the container that the box was in, because there was no term in x in the energy.

But here the x-integral is not just a volume.

Anyway, all we have to do now is to do these two integrals.

Remember that multiplicative numerical constants in the partition function don't make any difference.

Why not?

Because when we take the logarithm of the partition function they correspond to additive constants.

And since we are always going to be differentiating the log of the partition function, we don't care about such numerical constants.

Let's look at the p-integral first

$$\int e^{-\beta \frac{p^2}{2m}} \, dp \tag{22}$$

We have already done it.

But let's review the procedure to calculate it.

We change variable

$$\beta \frac{p^2}{2m} = q^2 \tag{23}$$

Therefore

$$p = \sqrt{\frac{2m}{\beta}}\, q \tag{24}$$

The p integral becomes

$$\sqrt{\frac{2m}{\beta}} \int e^{-q^2}\, dq \tag{25}$$

The integral is a definite integral.

Its bounds of integration are $-\infty$ to $+\infty$.

It is a number.

If you happen to remember, the number is the square root of π^3.

But anyway it is just a multiplicative numerical constant that we won't care about in equation (21).

Having taken care of the p integration, equation (21) becomes

$$Z = \sqrt{\frac{2\pi m}{\beta}} \int e^{-\beta \frac{kx^2}{2}}\, dx \tag{26}$$

Now we play exactly the same game on the x integration.

Let's change variable

$$\beta \frac{kx^2}{2} = y^2 \tag{27}$$

That means that

$$x = \sqrt{\frac{2}{\beta k}}\, y \tag{28}$$

Again, the *y* integral yields

$$\sqrt{\frac{2\pi}{\beta k}} \tag{29}$$

Finally the partition function is

$$\begin{aligned} Z &= \sqrt{\frac{2\pi m}{\beta}}\, \sqrt{\frac{2\pi}{\beta k}} \\ &= 2\pi \sqrt{\frac{m}{k}}\, \frac{1}{\beta} \end{aligned} \tag{30}$$

Remember that, for a spring, the quantity $\sqrt{k/m}$ has an important physical meaning: it is its frequency of oscillation, usually denoted ω.

So we can write

$$Z = \frac{2\pi}{w}\, \frac{1}{\beta} \tag{31}$$

Of course the multiplicative constants in front of this expression won't matter, but at any rate that is the full expression of the partition function $Z(\beta)$.

Now let's calculate the energy(average energy) of the oscillator .

How do we do that?

We calculate the logarithm of Z and differentiate it with respect to β.

That is the formula that we learned to use and to trust, see formula (19) of part 4.

$$\log Z = c - \log \beta \tag{32}$$

where the constant c is the logarithm of the coefficient in front of $1/\beta$ in formula (31).

Now what we want to do is we want to calculate minus the derivative of $\log Z$ with respect to β.

We are cruising now.

We are not thinking, just chasing symbols.

We get

$$E = -\frac{\partial \log Z}{\partial \beta} = \frac{1}{\beta} \tag{33}$$

And of course this is the temperature.

So we end up with

$$E = T \tag{34}$$

This is not so different than $E_{\text{part}} = 3/2\ T$ of a particle in a gas.

The first question is: Why don't we have a factor 3 in formula (34)?

That is because we are considering a one- dimensional oscillator, oscillating along a single axis.

So there is no 3.

Secondly, what happened to the two downstairs?

Why isn't formula (34) $E = 1/2\ T$?

The reason is because there were two integrals, each of which, in the calculation of Z, gave a multiplicative factor proportional to square root of $1/\beta$.

In the end, in the calculation of E, this produced the addition of two $1/2\ \beta$.

If, for whatever reason – more dimensions, more degrees of freedom –, there had been more Gaussian integrals in the calculation of Z, like the two integrals we had in formula (21), there would have been more terms $1/2\beta$ added in the calculation of the energy E.

What are the two kinds of energies, incidentally, that we are talking about?

We are talking about the energy of an oscillator.

It is kinetic energy and potential energy.

For an oscillator, the average kinetic energy and the average potential energy are in fact equal.

And each one is equal to 1/2 temperature.

So the kinetic energy is $T/2$.

The potential energy is $T/2$.

Together they give us the temperature.

There are two interesting things to notice about formula (34).

The first is that E doesn't depend on the mass of the oscillator.

That was also true of the particle in the gas.

At a given temperature, the kinetic energy of a particle does not depend on the mass of the particle.

The velocity of the particle did depend on the mass.

The reason is that, at a given temperature, $1/2\ mv^2$ is the same for all particles.

So the higher m, the smaller v .

The same is true in the case of the oscillator.

The average kinetic energy is one half the temperature.

And it is independent of the mass of the oscillator.

The second observation is that E is also independent of the spring constant.

As we saw, the potential energy is also one half the temperature.

And k doesn't enter the formula.

That is a little surprising if you think about it.

Indeed, imagine making the spring constant so large that no known force on Earth, or elsewhere, can stretch the spring away from its equilibrium point.

One would think, under such circumstances, that it is not a spring, it is just a constrained object, which has a fixed length – the length possibly being zero – which cannot be changed.

In that case, any sensible person would say you can't excite it.

You can't give it any energy.

There is no way to start it vibrating.

There is no way to give it any structure at all.

You can't give it any kinetic energy because it is absolutely locked in place.

You can't give it any potential energy because you can't pull it away from its rest point, no matter how much force you exert on it.

Yet the formula doesn't seem to care.

There always seems to be energy $E = T$ for every oscillator no matter how unphysical or how hard it is to get that oscillator to move.

So there must be something wrong in the analysis.

The thing wrong is that we have ignored one very crucial feature of nature.

We have not ignored anything about classical mechanics.

These are correct classical mechanical conclusions.

On the other hand it is crazy.

This occupied people around the turn of the 20^{th} century.

They were very confused by it.

And the reason they were confused is because the ideal gas law and the energy per particle seemed to work pretty well when the particles were treated as a point particles.

Suppose, however, that the molecules could not be modeled as simple point particles.

Indeed molecules can be for instance diatomic.

Diatomic molecules can vibrate.

They are little oscillators.

In truth, they are rather stiff oscillators.

They have a good solid spring constant.

You have to give them quite a knock to start them vibrating.

Anyhow, the analysis seems to say that for a gas of diatomic molecules we ought to have not $E_{\text{particle}} = 3/2\ T$, it should rather be perhaps $5/2\ T$ to account for the extra energy due to the molecule vibrating like a harmonic oscillator.

But it was known that such extra energy wasn't there.

So you could think of two conclusions.

1. One conclusion is that the diatomic molecule is not really a diatomic molecule at all.
   It is a point particle.
   But that is ridiculous.

2. The other conclusion is that we are doing something wrong or missing some ingredients of physics in our reasoning, which tends to keep that oscillator from having that much energy.

The missing ingredient is of course quantum mechanics.

5.5 Quantum harmonic oscillator

So what we want to do is the corresponding quantum mechanical calculation for a quantum mechanical oscillator in equilibrium, and compare the results with equation (34), which we obtained doing classical mechanics calculations.

How much do we have to know about quantum mechanics?

Very little.

We need the expression for the energy of a harmonic oscillator.

You probably know that harmonic oscillator energies are quantized.

They come in discrete multiples of Planck's constant times the frequency.

So the possible energy levels of the quantum oscillator oscillating with frequency ω are

$$n\bar{h}\omega \tag{35}$$

where $\bar{h}$ is Planck's constant .

The set of energy levels of a quantum harmonic oscillator can be represented as in figure 2.

It is just a discrete set of states, all equally spaced.

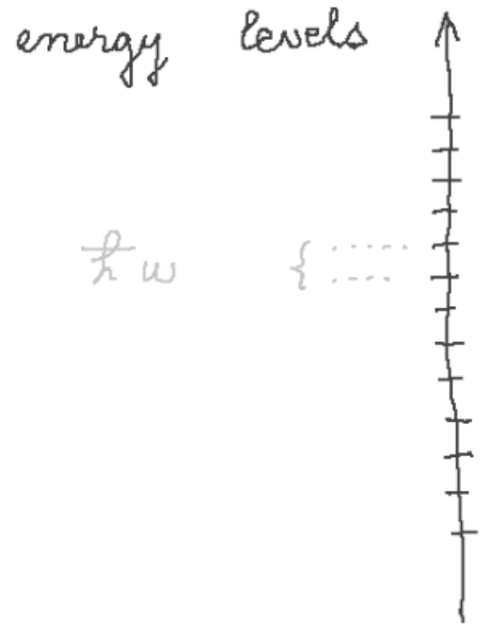


Figure 2: Energy levels of a quantum harmonic oscillator.

That is all that we have to know about the quantum harmonic oscillator: the energy of a single harmonic oscillator comes in discrete integer multiples of $\bar{h}\omega$.

Now we can calculate the partition function for the quantum mechanical oscillator.

As always, we sum $e^{-\beta E_i}$ over all energy levels E_i.

For the classical oscillator, the space of states was a continuous phase space, so we replaced the sum by an integral, equation (18). But for the quantum oscillator, it is truly a sum.

$$Z = \sum_n e^{-\beta n \bar{h}\omega} \tag{36}$$

At first sight this looks complicated.

But in fact it the simple series of powers

$$Z = \sum_n \left(e^{-\beta \bar{h}\omega}\right)^n \tag{37}$$

The first term, when $n = 0$, is 1.

And the series has the form

$$1 + x + x^2 + x^3 + ... \tag{38}$$

It is a **geometric** series.

When |x| < 1, it converges and is simply equal to

$$\frac{1}{1 - x} \tag{39}$$

Since the exponent $-\beta\bar{h}\omega$ is negative, the basic exponential term in the power series expressing Z is less than one.

The power series converges.

And we can rewrite equation (37) as

$$Z = \frac{1}{1 - e^{-\beta\bar{h}\omega}} \tag{40}$$

This is the partition function for the quantum harmonic oscillator.

There is a number of ways to write it. But this is fine.

What do we want to do with this partition function?

We want to calculate the average energy.

The general formula is

$$E = -\frac{\partial \log Z}{\partial \beta} \tag{41}$$

But here it is simpler to use the equivalent formula

$$E = -\frac{1}{Z}\,\frac{\partial Z}{\partial \beta} \tag{42}$$

What is the derivative of Z with respect to β?

There is an ugly thing in the denominator of the expression of Z, equation (40).

When you differentiate a function which is in the denominator, you get first of all, minus one over the denominator squared.

And that is multiplied by the derivative of the function in the denominator, so we get

$$\frac{\partial Z}{\partial \beta} = -\frac{1}{\left(1 - e^{-\beta \bar{h}\omega}\right)^2} \frac{\partial}{\partial \beta}\left(1 - e^{-\beta \bar{h}\omega}\right)$$

$$= -\frac{\bar{h}\omega \, e^{-\beta \bar{h}\omega}}{\left(1 - e^{-\beta \bar{h}\omega}\right)^2} \tag{43}$$

Now we have to divide by Z.

That means multiply by

$$\left(1 - e^{-\beta \bar{h}\omega}\right)$$

It reduces the power in the denominator of equation (43).

$$\frac{1}{Z} \frac{\partial Z}{\partial \beta} = -\frac{\bar{h}\omega \, e^{-\beta \bar{h}\omega}}{1 - e^{-\beta \bar{h}\omega}} \tag{44}$$

This is minus the energy.

We want plus the energy.

So we change sign, and finally we reach

$$E = \frac{\bar{h}\omega \, e^{-\beta \bar{h}\omega}}{1 - e^{-\beta \bar{h}\omega}} \tag{45}$$

After all these calculations, by sheer luck we land on a positive number.

Formula (45) is the answer for the average energy of the quantum harmonic oscillator.

It is to be compared with the average energy of the classical harmonic oscillator, which was simply $1/\beta$.

So what is the relationship between them?

Let's first look at high temperatures.

High temperature is a situation where the classical theory is good.

Each one of the little quanta of energy $\bar{h}\omega$ are thought of as very small.

A classical spring has a great many units of energy.

So in that sense the classical system in quantum units has a lot of energy.

That must mean that it has very high temperature in some quantum sense.

Classical systems are quantum systems that become classical when the temperature is very high.

The quantization of energy then becomes unimportant.

Let's look at equation (45) at high temperature, that means when β tends to 0.

Then the exponential term $e^{-\beta\bar{h}\omega}$ tends to 1.

We can replace $e^{-\beta\bar{h}\omega}$ by 1 in the numerator.

But of course we cannot do the same in the denominator.

So we will use the second term in the Taylor expansion of $e^{-\beta\bar{h}\omega}$

Remember that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \tag{46}$$

And the power series converges whatever *x* is, real or complex, with no bound on its magnitude.

Anyway, when *x* is small we can take

$$1 - e^x \approx -x$$

So the energy given by equation (45) can approximated as (47)

$$E = \frac{\bar{h}\omega}{\beta\bar{h}\omega} \tag{48}$$

The next terms after $\beta\bar{h}\omega$ in the denominator of the righthand side of equation (44) are proportional to β squared, β cube, β to the fourth, and so forth.

They are much smaller than βhω and they can be ignored when β gets very small.

So, at high temperature, equation (44) giving the energy becomes simply

$$E = \frac{1}{\beta} \tag{49}$$

At high temperature, it just reproduces classical physics: $E = T$, which is the same as equation (34).

Let's now ask about the low temperatures, that means when β becomes very large.

Then in equation (45), the exponential term $e^{-\beta\bar{h}\omega}$ becomes very small.

The denominator can be approximated by 1.

And we can write

$$E = \bar{h}\omega \; e^{-\beta\bar{h}\omega} \tag{50}$$

We see that, far from getting just the temperature as in equation (49), we get something that is exponentially small when the temperature goes to zero.

We discover that quantum mechanics tends to suppress the energy of an oscillator when the temperature is low enough.

At very low temperature, the oscillator doesn't behave like the classical oscillator at all.

It has much less energy than the corresponding classical oscillator would have had.

The question then is: Where is the crossover point?

When temperature increases, where does it go from being quantum to classical?

If we look at equation (50), we know that the crossover is where the exponential term $e^{-\beta\bar{h}\omega}$ goes from being small to being big.

The biggest $e^{-\beta\bar{h}\omega}$ ever gets is 1.

But we can consider that the crossover between the low temperature behavior and the high temperature behavior is when the exponent in $e^{-\beta\bar{h}\omega}$ is about 1.

So

- when $\beta\bar{h}\omega > 1$, quantum,
- when $\beta\bar{h}\omega < 1$, classical.

Re-expressed in term of temperature, this is: the behavior is quantum when $T < \bar{h}\omega$, , and it is classical when $T > \bar{h}\omega$.

The crossover point is when

$$T = \bar{h}\omega \tag{51}$$

If we step back and look at this result, it makes a lot of sense.

It says that the crossover is when the energy of the oscillator, measured classically, that is with *T*, is equal to one quantum worth of quantum mechanical energy.

If we try to make the temperature lower than that, then the oscillator has less than one quantum worth of energy.

The oscillator somehow doesn't want to have less than one quantum worth of energy.

So when the temperature, or equivalently the energy, crosses this threshold downward, the expression for the energy, given then by equation (50), becomes exponentially small.

To get a feel for the behavior of the energy given by equation (45), figure 3 represents the curve of *E*, for a given $\bar{h}\,\omega$, as a function of *T*

When T is large, E is simply approximately equal to T.

And when T is small, we see how E gets exponentially closer to zero than T.

Now we naturally know the answer to the puzzle about the diatomic molecule.

When the temperature is very low we are deep in the quantum regime.

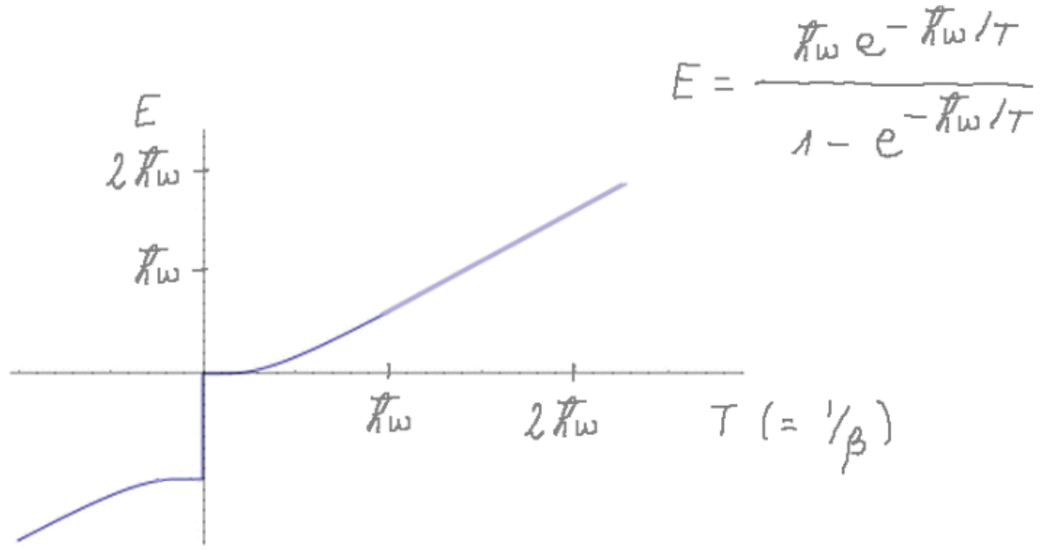


Figure 3: Curve of the energy given by equation (44) as a function of T. (Only the positive part is physically meaningful.)

Then the oscillator has exponentially small energy compared to what it would be classically.

On the other hand, when the temperature goes up, at some point the temperature will get larger than $\bar{h}\,\omega$.

At that point we begin to activate the oscillator.

We can't activate the quantum oscillator unless the temperature is such that E corresponds to more than one quantum's worth of energy.

That is what the threshold equation (51) says.

At low temperatures, up to some temperature, the diatomic molecule behaves like a monoatomic molecule.

The molecule doesn't vibrate like an oscillator.

When we raise the temperature, past a threshold where the temperature becomes bigger than the quantum of energy, the molecules start to behave like genuine diatomic molecules.

Let's turn to the characteristics of the molecular oscillator.

Its stiffness determines ω.

Remember that $\omega = \sqrt{k/m}$.

So the stiffer the oscillator, the higher the frequency of oscillation.

Therefore the stiffer the oscillator, the higher the temperature of the crossover point $\bar{h}\,\omega$ will be.

Molecules are pretty stiff on the scale of rubber bands.

So the threshold temperature is fairly high.

In fact, before we cross the threshold, something else happens.

Molecules start to rotate.

Other things start to happen.

Our calculations are just an illustration of happens when we go from quantum to classical behavior.

Then what happens if we raise the temperature even more?

We activate even worse things.

The atoms themselves start vibrating.

Eventually they will get ionized.

When an atom becomes ionized, it doesn't behave like an atom at all.

If it is an hydrogen atom, it start to behave like an electron and a proton

This is an example of quantum mechanics solving a deep problem that physicists have been very confused about around the turn of the 20^{th} century.

A similar problem came up in the following context.

Solids are simply systems of molecules where the molecules are arranged in a lattice.

They vibrate around their average lattice positions.

So, beside its possible internal structure deformation – which usually is much stiffer –, each molecule is also sort of an oscillator around its lattice position.

The energy of these oscillators as a function of the temperature determines the specific heat, or heat capacity, of the crystal.

However the theoretical estimations of their values by classical physics, for various crystals, were much larger than the actual measurements.

It was Einstein who figured out what was going on.

He explained that the quantum statistical mechanics of oscillators essentially suppressed them until the temperature got high enough.

Einstein published his paper on the specific heat of solids in 1907.

He had already thrown in some of the seminal ideas that would lead to quantum mechanics, for instance his explanation, in 1905, of the photoelectric effect with photons.

But his paper on specific heat came twenty years before people understood the foundations of quantum mechanics.

In 1907 Einstein just knew that the quantum energy levels had to be proportional to frequency and to n.

So it is another of Einstein's contributions – one of his minor ones.

Before we leave oscillators, let's step back to look at what we did and found.

We went from the classical harmonic oscillator to the quantum harmonic oscillator, because we followed the chronology of scientific progress.

Staying in classical physics, we saw that we could explain pretty well the behavior of an oscillator plunged in a heat bath.

But the theory ran into trouble for very stiff oscillators.

It predicted too high energy at low temperatures.

Then we went into quantum mechanics, staying at a very simple level, which corresponded to what was known in the first decade of last century.

No fancy complex vector space describing the states of a system, no Hilbert space, no distinction between states and observations, no fundamental randomness in the measurements made on a given superposition of states, when the state is expressed in the basis attached to the Hermitian operator of an observable, etc.

If you want more than take a full class on quantum mechanics.

We just assumed the quantification of the levels of energy.

And we saw it was sufficient to solve the problem.

The right way to go – or at least the logical way to go – is to study first the quantum mechanical description of nature, and then go to the classical description as a limit case when systems are made of a large number of particles, and the temperature is not too low.

The classical physicists were right in using a continuous phase space, calculating the partition function as the integral in equation (21), and showing that $E = T$.

They just had a problem at low temperature, that is low energy.

The oscillator had too much energy compared to experimental observation.

Elementary quantum mechanics solved the problem for temperatures below the threshold of $\bar{h}\,\omega$ – which happens to be high, and that is why classical physicists were aware of the problem and terribly puzzled by it.

But above the threshold, quantum mechanics confirms that $E = T$.

Then it does not depend on the stiffness of the oscillator or anything, only on the temperature.

In fact, elementary quantum mechanics did not change the partition function much.

In the classical case we obtained equation (31), reproduced below

$$Z = \frac{2\pi}{\omega}\frac{1}{\beta} \tag{52}$$

In the quantum case equation (40) gives

$$Z \approx \frac{1}{\hbar\omega}\frac{1}{\beta} \tag{53}$$

If we remember that $\hbar = h/2\pi$, we see that quantum mechanics simply replaces 1 by *h* in the denominator of the classical partition function.

And since multiplicative constants in partition functions don't play any role, at first sight they are quite similar.

The difference appears when we calculate the energy, with equation (31) in the classical continuous case, leading us to $E = T$ no matter what; and with the equations (40) to (44) in the quantum case, leading us to the more complicated equation (45).

Once it is above a quantum threshold, the energy is just equal to the temperature and it does not matter what the spring constant is.

Question: What happens to the oscillator in terms of kinetic and potential energy once we are above the threshold temperature?

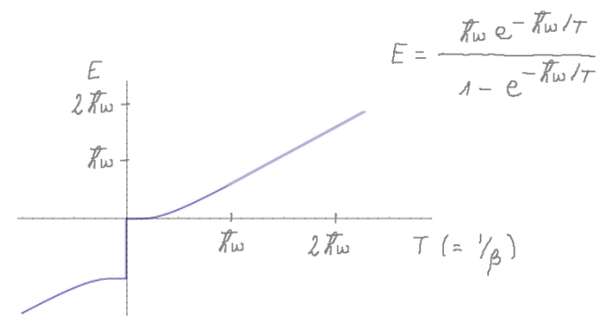
Answer: It speeds up and it makes larger and larger oscillations as we increase the temperature.

The point is that, as we heat it from low temperatures to high temperatures, the temperature at which the classical behavior sets in, at which the energy paradoxically becomes independent of how stiff the spring is, and becomes simply $E = T$, gets to be a higher and higher temperature as a function of the spring constant.

So if we take a very very stiff spring constant, let's say $k = 10^{10^{10}}$, then, yes, at sufficiently high temperature, it will have an energy equal to its temperature.

But at any kind of ordinary temperature, it would essentially have no internal energy.

It would be the motionless and energy-less solid we expect it to be – leaving aside the motion due to the heat bath.



Look again at figure 3, to see that it is exactly what the curve of E shows.

The threshold for the oscillator to start behaving classically would get higher and higher as the spring constant got stiffer.

This leads us to an interesting observation.

5.6 Degrees of freedom

The transition from quantum to classical behavior, when we cross a threshold in temperature, is interesting because it has the following consequence:

As we heat a system, we start to discover that it has more and more degrees of freedom.

To understand this better, let's make a crazy molecule.

It consists of a spring with two masses connected to it, at each of its extremities.

The spring is rather stiff.

Furthermore the mass on the right is itself actually made of two little masses linked by a secondary spring, much stiffer than the main one, as shown in figure 4.

How does it behave at low temperatures?

At low temperatures there is not enough temperature to overcome the quantum threshold for either of the two oscillators to start oscillating.

So the whole molecule behaves like a point mass molecule.

It behaves with the energy 3/2 T, like any other point mass particle in the heat bath, see equation (65) of part 4.

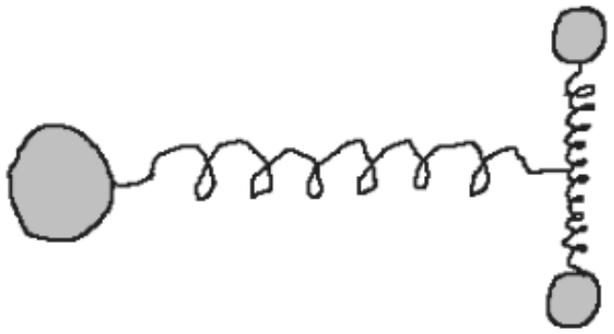


Figure 4: Molecule whose number of degrees of freedom increases with temperature. The secondary spring on the right is much stiffer than the main spring in the middle.

Then we heat it up to the point where we activate the main oscillator in the middle.

But still we don't give it enough temperature to activate the secondary super duper strong spring on the right.

So the molecule simply behaves like a diatomic molecule, with for instance 5/2 T .

We heat it more.

Eventually we reach the threshold where we also activate the secondary spring.

The molecule starts to behave like three masses connected by springs.

Instead of the 5/2 T of the diatomic molecule, it behaves now with for instance 7/2 T, or whatever the right number is.

The lesson to draw from this example is that, as we heat the system, more and more degrees of freedom become activated.

And we start to discover the complexity of things.

We don't discover the complexity of things at low temperature.

When the temperature is low, we discover the quantum behavior of things.

As we heat it, we start to discover the complexity of the system under study.

Degree of freedom by degree of freedom, unfreezing themselves from the quantum constraint, the system starts behaving more and more classically.

You may realize that this has a lot to do with the blackbody radiation.

We will come to that.

But we won't relate it right now to the blackbody radiation.

For the moment, let's relate it to something else: a violin string.

A violin string consists mathematically of a large number of oscillators.

The oscillators are the string oscillating in its various harmonics, figure 5.

In other words, from a fundamental mathematics point of view, a vibrating string is basically an infinite number of harmonic oscillators, of shorter and shorter wavelengths.

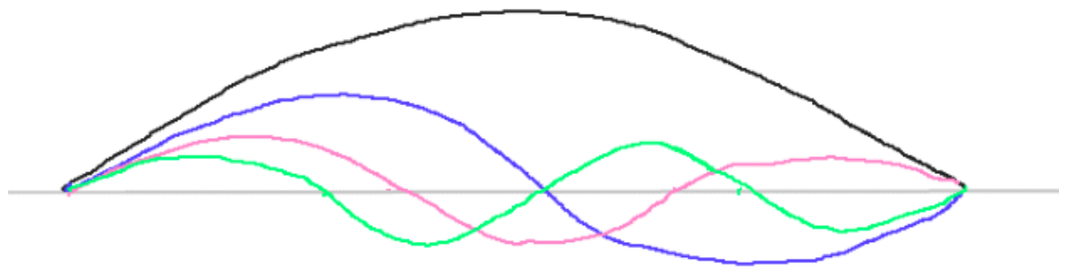


Figure 5: The first four harmonics of a violin string vibrating.

If each one of these oscillators had energy equal to the temperature, as given by equation (34), and there were an infinite number of them, then that string would have an infinite amount of energy when it came to thermal equilibrium.

So something is wrong.

The answer to the riddle is that most of these oscillations have very high frequency.

Indeed the shorter the wavelength the higher the frequency of the oscillator.

And therefore most of these oscillators, at any given temperature, are frozen out by quantum mechanics. i.e., for those frequencies, $\bar{\hbar} < \quad \omega$.

So they are not free to oscillate.

As the temperature goes up, more and more of these harmonic oscillations appear, representing more energy of the string.

That is another paradox similar the one discussed in the previous section.

5.7 Questions/answers

Question: If we want to know the values of these numbers k, T, $\hbar\,\omega$, etc., in the transition from quantum to classical behavior, can we observe and measure them experimentally, somewhere, somehow?

Answer: Yes, absolutely.

They can be observed and measured experimentally.

It is done just studying the specific heat of crystals.

The experimental phenomena are obvious.

And once physicists understood what was going on and had the right model, the calculations were easy.

It was obvious to experimental physicists around 1900 that there was something wrong, because their crystals did not behave according to classical mechanics.

So yes it is extremely easy to see the transition, and to make measurements about it.

Question: If we heat the crystal, we can make it change its behavior?

Answer: Yes.

If you heat the crystal, the specific heat starts to behave as if there was a certain number of T, or more precisely a certain number of T per oscillator.

This is really easy to see.

Question: This would translate into the partition function?

We could handle the multiple oscillators?

Answer: Yes.

The partition function would be the product of all the partition functions, one for each oscillator.

The logarithm would be the sum of the logarithms.

The logarithm being the sum implies that the energy of the system is the sum of all the oscillators energies.

When we calculate a partition function we take into account the exact energy level of each possible state of the system.

Then, when we talk about calculating the energy from the partition function, we mean the average energy.

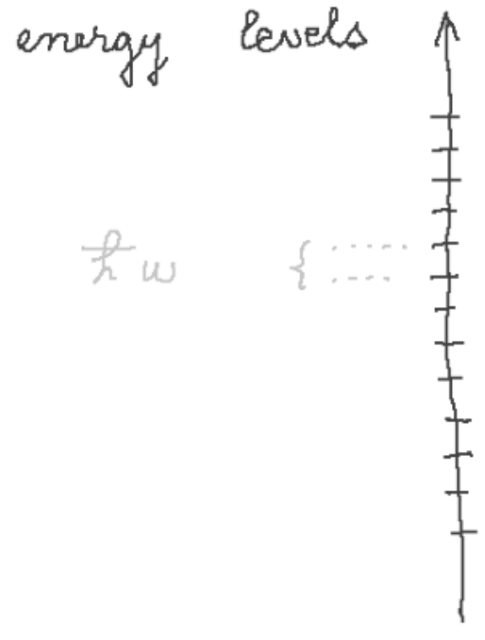
So the average energy of the system, would be the sum of the average energy of each oscillators.

But those whose frequency is too high, above the threshold temperature, those are in the quantum regime.

And they are not appreciably excited.

Question: In describing the quantum levels of energy of the quantum oscillator, equation (35) and figure 2, in order to compute the partition function, why did we ignore the ground level energy?

Answer: That is right, we did ignore the ground level energy.



Instead of using the proper equation

$$E_n = \bar{h}\omega \ (n + 1/2) \tag{54}$$

we simplified it into equation (35) reproduced below

$$E_n = n \ \bar{h}\omega \tag{55}$$

The reason is that it doesn't matter.

Adding a constant to every energy levels is equivalent to multiplying the partition function by a constant.

And the result is no change in anything of interest.

Let's go to the next topic, which is the second law of thermodynamics.

We will first of all explain what is puzzling about the second law.

Then, after some further questions and answers, we will give one statement of the second law.

After that we will discuss further trajectories in the phase space, chaos and recurrence.

5.8 Second law of thermodynamics

Why is the second law a puzzle?

It is a puzzle because it says that the world is irreversible.

In isolated systems, something called entropy always increases, or stays the same.

In fact you would have to be infinitely careful, if the isolated system you examine is big and not in equilibrium, to keep its entropy from increasing.

Remember that the entropy of a system, unlike its energy, temperature, pressure, etc., is not a characteristic of the system in itself.

It is a characteristic of our incomplete knowledge of the exact state the system is in.

According to equation (30) of part 1, the entropy of a system, when we know exactly in which state it is, is zero.

$$S = \log V_{PS}$$

But in general, because it is difficult to follow exactly all the positions and motions of all the particles of a big system – for instance of a gas in a vase –, at time *t* the system has a certain entropy.

Moreover the entropy of an isolated system will increase, unless it stays the same.

For instance the entropy of a gas in a vase isolated from the exterior, and not yet internally in thermal equilibrium, will increase until it reaches a maximum value, corresponding to the Boltzmann distribution over all its possible states, see parts 3 and 4.

The puzzle stems from the fact that, on the other hand, Newton's equations of motion are completely reversible.

So anything that can happen in one direction of time can happen in the opposite direction of time – implying, in particular, that the entropy of an isolated system should be able to decrease spontaneously.

So there seems to be a contradiction.

At the least, we can say this:

> *There is a tension between reversibility of the fundamental equations, and irreversibility of the observational properties of complicated systems.*

That took some time to sort out.

And many people are still confused by that.

So let's talk about it.

Let's imagine some phase space.

A phase space is the space of states of a classical system.

It has got a lot of dimensions.

It has got a lot of particles, or a lot of whatever it is made of.

But, as usual, we are going to represent it simply as two-dimensional, figure 6.

One dimension is the coordinates.

The other dimension is the momentums.

There are a lot of them.



Figure 6: Phase space of a system.

So this is the sketchy representation of a high dimensional space.

We start out, at time t_0, with some probability distribution over a subset of all the possible states.

For example let's start out with some probability distribution over the subset represented in figure 7.

The picture doesn't show explicitly the probability distribution, or density, which would be in a third dimension toward us for instance, but we can imagine it to be a uniform distribution.

So there is zero probability for the system to be outside the blob, and uniform probability to be inside of the blob in figure 7.

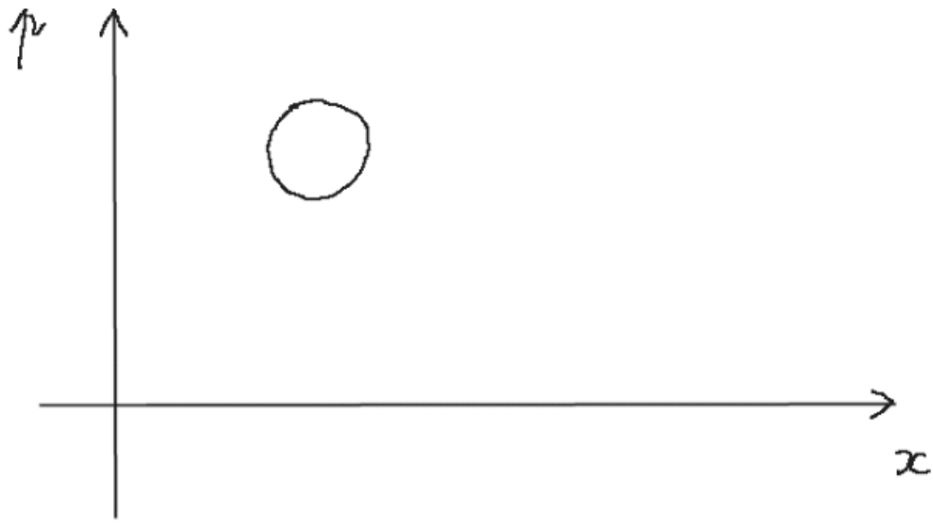


Figure 7: Initial subset of possible states.

This means that we know nothing about the exact state of the system, except that it starts inside the blob.

Since we don't know anything more about it, the probability density is constant within the blob.

Do you remember what is the entropy of that system?

Go back to part 1 if you need to refresh your memory.

When the distribution is uniform, the entropy is the logarithm of the volume of the blob, equation (31) of part 1.

$$S = \log V_{PS}$$

More generally, it is $-\int P \log P$, equation (40) of part 1.

It is not yet the maximum of anything, except, it is the maximum of our lack of knowledge within the blob.

It is not the result of some maximization under constraint.

We just happen to know that the system is within that blob.

We did our best possible measurements and we discovered that it is there.

In particular, the system is not in equilibrium.

We are not talking about the entropy in equilibrium, we are talking about the entropy of a certain probability distribution.

Notice that for a system in equilibrium the entropy doesn't change with time, although the molecules still zip around all over the place at a few hundred meters per second, and keep colliding with each other.

It is a pandemonium, and it is known that it has temperature.

So in thermodynamics, equilibrium doesn't have the same meaning as in Newtonian dynamics and statics.

The probability distribution is handed to us.

Somebody told us that the momenta and the positions of 10^{23} particles correspond to a point within the blob

But we are not told their exact values, we are only told they are in there, with a uniform distribution.

This situation is not that unrealistic.

We might start with all the air molecules in some air tank in the corner of the room.

Then we open up the tank, and let the air go.

Originally, at time t_0, we knew that the molecules were all confined in a small volume in the corner.

That corresponds to some section of the phase space.

Moreover we had a pretty good idea of what their energy was, because we knew what the temperature was in the tank.

So we had a pretty good idea of what the momenta of all the particles were.

That gives us a blob.

Don’t mix up the small volume in the corner of the room with the corresponding blob in the phase space.

What would be the blob corresponding to all the molecules in the air tank?

What would it be if we knew exactly the positions and momenta of all the molecules?

There might not be exactly a uniform density of probability over the blob, but it would certainly correspond to some blob-like configuration.

Make sure you clearly see that the more we know, the smaller the volume of the corresponding region in the phase space is.

Recall that, when the probability is uniform, the entropy is the logarithm of the volume.

So, as the volume gets smaller and smaller, the entropy goes down.

Now, what happens to the blob as time goes on?

Remember Liouville's theorem, which we studied earlier.

It says that the volume of the blob stays the same.

Therefore the entropy stays the same.

It is just another way of saying that if we were able to know the exact state the system is in at some time t_0 and we were able to follow what it becomes, it would stay unique in the future.

And that would be true also going backward in the past.

Each state comes from a unique state, and goes into a unique state. And if we could follow the system we would see it.

The discrete version of Liouville's theorem is easy to figure out intuitively.

If the system at time t_0 is, with respective probabilities 1/3, 1/3 and 1/3, in the three possible states ω_1, ω_2 and ω_3, then, at a later time t_1, it will still be in three possible states, not necessarily the same, but with the same equal probabilities 1/3. And it is even true with non-uniform probabilities p_1, p_2 and p_3.

That is one meaning of the following law, which we called the **law number –1**, because it is even more fundamental than the **laws number 0, 1, 2**, etc. :

In nature information never gets lost or erased.

We observers may lose information, but Nature doesn't.

Liouville's theorem – saying that the collection, or blob, of possible states may change from one time to the next, but that the shape of the distribution doesn't – seems to imply that entropy doesn't change.

And it is true.

This is the correct view of what is called *microentropy*.

But there is another concept of entropy which does increase.

This other concept of entropy has to do with something called *coarse graining*.

Typically our experiments are such that we cannot resolve (to resolve, here, means to distinguish, like in the resolution of a photograph) all points in the phase space.

Even if our experiments were ideal, i.e. perfectly precise, we still could not tell one state from another state if they were too close.

There is some maximal resolution of the state that we can achieve.

In quantum mechanics, it turns out that a limit to the resolution is naturally built in.

The smallest detail that we can know, in other words the smallest uncertainty, is given by Planck's constant, i.e., in quantum mechanics there exists something call Heisenberg's Uncertainty Principle, namely, that $\Delta X \Delta P \geq \hbar/2$.

But in other situations it may just be that our resolving powers are not good enough.

So, instead of talking about points of phase space, we talk about *resolved points*, or *coarse grained points*, in phase space.

Let's make an example. Supposing we started with the blob in phase space shown in figure 7, or on the left part of figure 8, and it has a certain volume.

Let's follow it.

We show what it evolves into, not solving any mathematics problem; we just make up an example of dynamics which preserves the volume.

So suppose the dynamics transform the initial blob into a long snake as shown on the right part of figure 8.

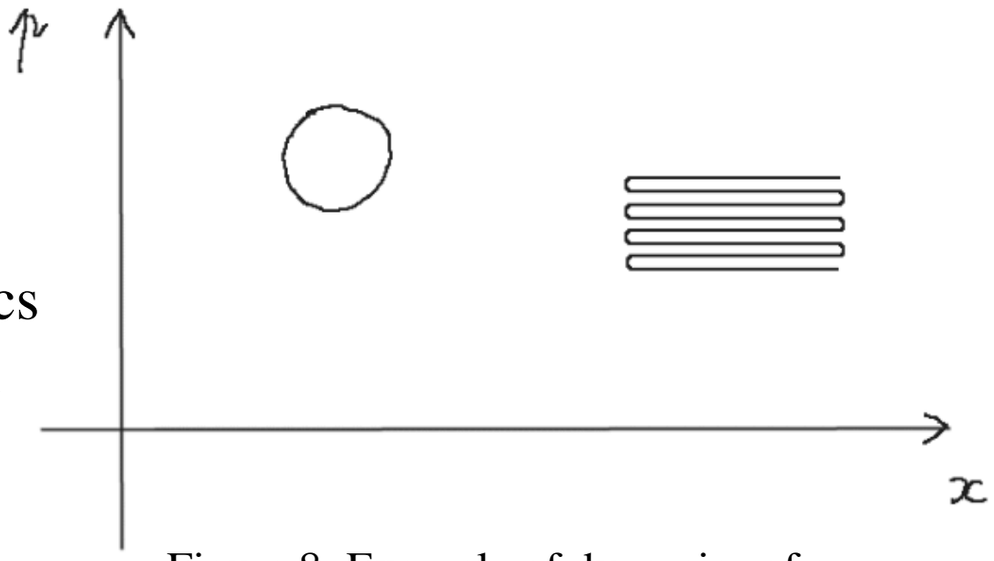


Figure 8: Example of dynamics of a blob: from the patch on the left it becomes a long folded snake on the right, *with the same volume*.

The snake is extremely thin, and the folds of the snake are extremely thin.

Yet the volume of the snake is the same as the volume of the initial blob.

But unfortunately, because of our limited resolution powers, we cannot tell a phase point from a neighboring phase point when they are too close.

What we do when this happens is we take the phase points of the snake and "blobify" them.

That is, we replace them by small blobs.

They all overlap.

And the result is that the initial blob now became not a snake but a bigger blob, figure 9.

We don’t violate Liouville.

The snake in figure 8 is still consistent with Liouville.

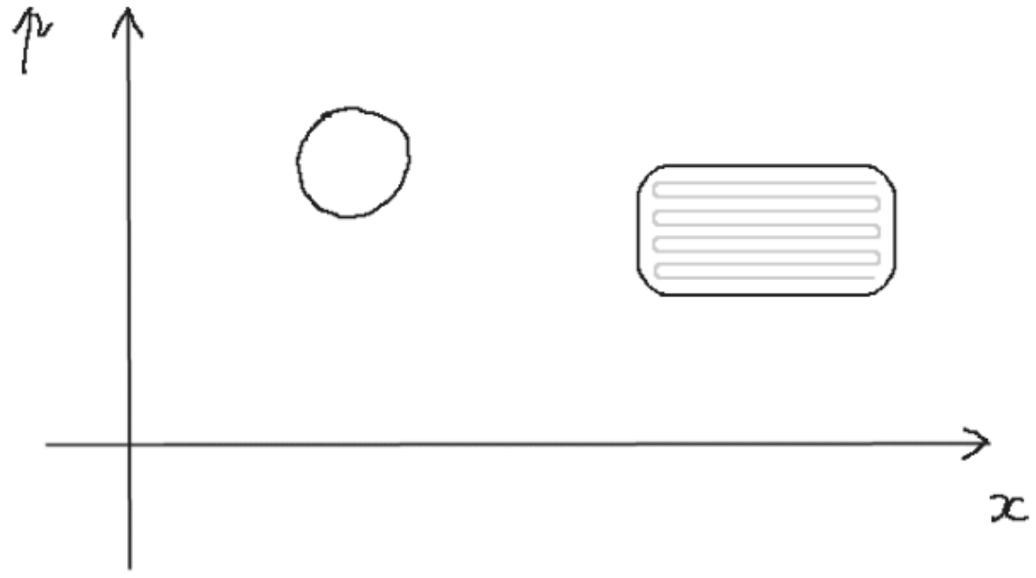


Figure 9: We view the snake as a blob.

We are just losing information in figure 9, because we can’t keep track of the fine structure of the snake.

We have a fuzzy resolution.

It is as if, so to speak, we were looking at these things with somebody else’s eyeglasses.

So instead of the snake we see a new larger blob.

If we were asked "where in phase space is the system now?", we would answer that, as far as we can say, its position is uniformly distributed over the larger blob.

The crux of the matter is that we couldn’t follow the system carefully enough.

So we lost information.

If we could follow the system with infinite precision, and we could figure out the initial probability density precisely too, then we would know that the system evolved into a point somewhere on the snake.

The volume would not change, representing the fact that we would not lose information.

But in real situations, we can't distinguish one state from a nearby state.

So we do this process of coarse graining meaning by that, replacing points in phase space by blobs in phase space.

Process may be too strong a word.

It does require any action on our part, except recognizing that we lost information, and now the blob representing our lack of knowledge is larger.

Once we do coarse graining, the volume increases.

The snake, which was actually a blob with the same volume as the initial one, becomes a blob with a bigger volume.

Here is an image which can help figure out what is going on in the way we perceive the dynamics of the system.

Imagine a piece of cotton wadding, figure 10.

What is the volume of this piece of cotton wadding?

If we mean the strict volume of the cotton fibers all added up, we can squeeze them down to something quite small.

If we let the cotton fibers grow, the strict volume of the cotton wad doesn't change very much incidentally.



Figure 10: Cotton wadding.

But the cotton wad sort of grows out.

And if we look at it again, it has a bigger apparent volume,

In the case of a blob of phase space points, when it evolves its strict volume doesn't change at all – that is Liouville's theorem –, unlike the wadding when we let the fibers grow.

But the image of the wadding is still appropriate to compare the strict volume of the snake to its apparent volume after coarse graining.

Returning to coarse graining, if we look at the cotton wadding through somebody else's glasses and we can't see the cotton fiber, we will say that the volume of that piece of fiber stuff has increased substantially.

In essence, that is what happens in the dynamics of the blob and the perception we have of it, figures 8 and 9.

5.9 Questions/answers

Question: Are we saying then that classically it only appears to increase, whereas quantum mechanically it does genuinely increase because of the uncertainty principle?

Answer: Let's forget quantum mechanics.

We are in classical mechanics.

The phenomenon is just due to the limits of our resolution power.

Remember that entropy is not really a property of a system.

It is a property of the system and what we know about the system.

So if we could follow in infinite detail every point in the phase space, and all we knew in the beginning was that the system was inside some blob, but we did not exactly where, and our lack of knowledge was represented by a uniform probability distribution, then we would say afterwards it is inside some other blob elsewhere with a comparable uniform probability distribution, and therefore the same volume.

We took the example where afterwards the new blob had the shape of a snake.

So again, if we could follow every point exactly, we would know afterwards it is in the snake somewhere.

But we would not know exactly where it is in the snake.

And entropy would be conserved.

But if we take into account that we can't tell one point from another, when they are too close to each other, because of limits of resolution, then we wind up saying:

"Look I know much less about where it is that I did to begin with."

Then entropy went up.

Question: So classically we can imagine that we might be able to keep track of the system in such a way that the entropy would remain the same.

Answer: Yes, that right.

But the problem is that, given any degree of resolution – whatever our resolution power is –, we will run into the same difficulty.

Let's suppose it was good enough to resolve the snake after 10 seconds.

So, for example, it is good enough to resolve the snake in figure 8.

But then, because of the way systems evolve in phase space, they start to grow even littler snakes, and littler snakes, and littler snakes...

Whatever our degree of resolution, this "snakiness" will appear.

Let's give its right mathematical name to the phenomenon: it is called *chaos*.

Let's now wrap up our description of the second law in terms of increasing chaos and coarse graining leading to a loss of information.

5.10 Second law of thermodynamics (continued)

The way systems evolve in phase space, from an initial blob, is they start to spread out sort of tentacles, without changing the volume.

They grow fibers if you prefer.

And the fibers grow thinner fibers.

And those thinner fibers grow even thinner fibers – the total volume staying always the same.

Eventually, no matter how good our resolution is, we will start seeing that the entropy increases.

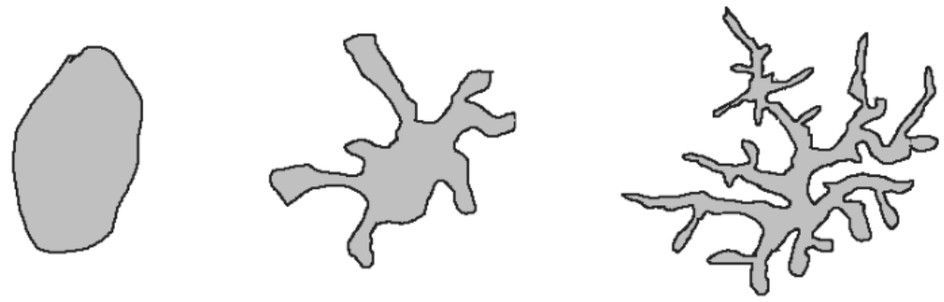


Figure 11: Chaotic growth, without changing the total volume.

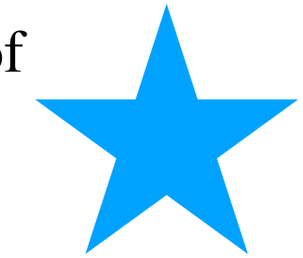
And that is what the second law of thermodynamics says.

One of the numerous equivalent statements of the second law of thermodynamics is:

> *The entropy of a closed system not in equilibrium increases until it reaches a maximum.*
>
> *At that point the system is in thermal equilibrium.*
>
> *And the uncertainty of our knowledge then corresponds to the Boltzmann distribution over all the possible states of the system.*

In the next part of our discussion, we will refine the above statement, adding that this evolution of the entropy is not absolutely sure but has an *extremely high probability*.

Let's continue investigating the interplay between the deterministic trajectory and our loss of information.



5.11 Trajectories in the phase space, determinism, chaos and recurrence

Let's imagine some phase space.

We started out knowing a great deal.

That is, we started out knowing that the system was somewhere inside some small blob.

Then the system evolves in the phase space.

Let's first of all talk about how big the entire available region of phase space can be.

Suppose the system is a gas made of a large number of particles in a box or a vase.

That means the x's are bounded.

For instance, in a one-dimensional analog, we would say that the available phase space is bounded within an interval.

Can the momenta be anything? No.

Usually the energy also constrains how much the maximum of the p's can be.

Even if all of the energy of the system went into one particle, that one particle would still itself have a finite energy.

Therefore the effective region of phase space which is available to us is some big box, figure 12.

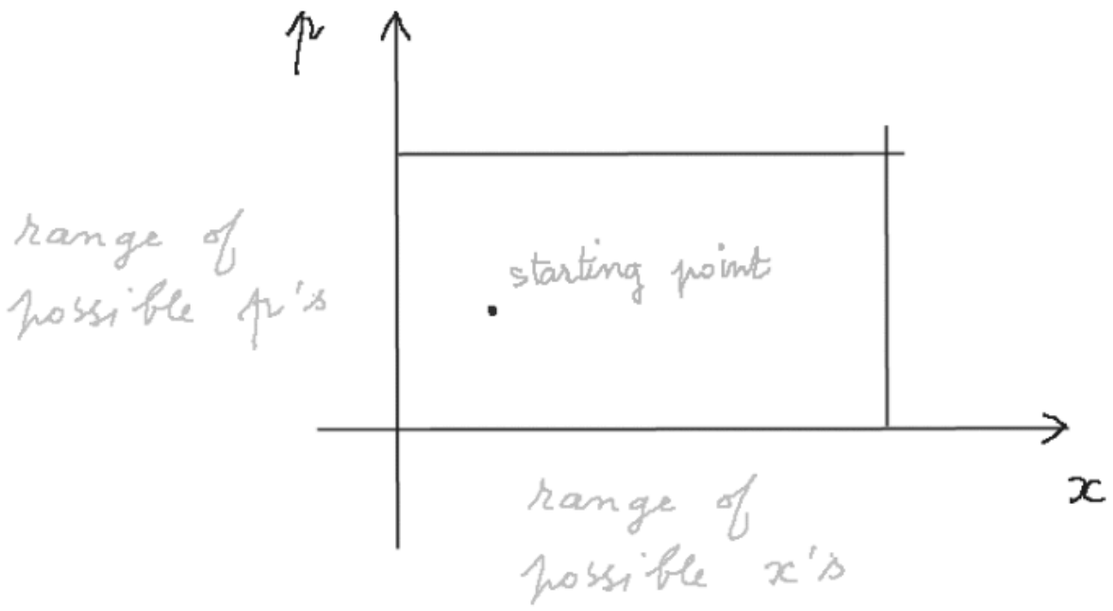


Figure 12: Maximum possible region in the phase space, for a gas in a box or a vase.

Now the phase point starts someplace.

The exact precise point is shown in figure 12.

And the system starts to evolve.

For most systems, what happens is that the subsequent path is extremely sensitive to the starting point.

So if we started initially just a little bit away from the first point, the paths would look the same for a little while, and then would be quite different, figure 13.

That is the essence of mathematical chaos: an equation of motion is classified as chaotic – or a system is classified as chaotic – if a small difference in the initial conditions leads quickly to quite different trajectories of evolution in phase space.

An example would be billiard balls.

Instead of the gas in a vase, think of it in two dimensions.

The vase is now an idealized billiard table with no friction.

The walls are the cushions.

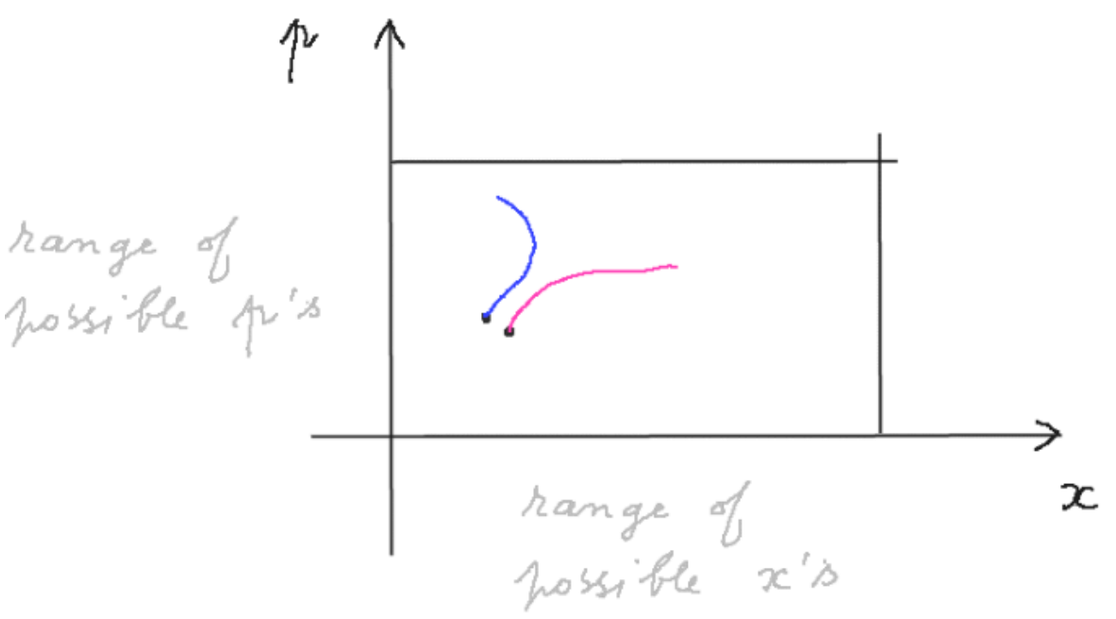


Figure 13: Sensitivity to initial conditions.

And the molecules are the billiard balls

Again, do not mix up the billiard table and figure 13.

When we represent a billiard table viewed from above, we represent only x_1 and x_2.

We don't represent the p's.

Whereas figure 13 represent all the x's as the horizontal axis, and all the momenta as the vertical axis.

So a trajectory on the billiard table doesn't have the same meaning as a trajectory in figure 13.

How many of you play billiard?

Do you play pocket pool? Chicago? Straight pool?

Let's leave three-cushion aside... too hard, too much chaos.

So we arrange the balls in a certain fashion and start with a cue ball.

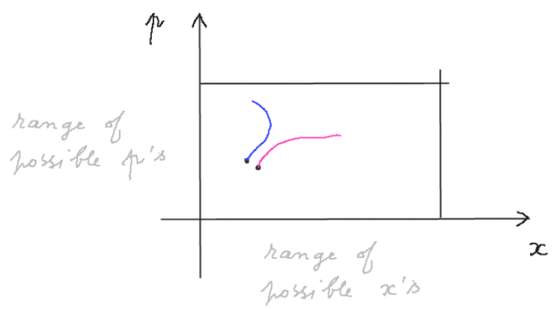
If, on two different occasions, we reproduce the same initial configuration extremely accurately and we shoot exactly the same, then we know that exactly the same things will happen.

That is indeed what Newton's laws of physics say.

However, everything else being kept perfectly fixed, what if we are a tiny bit off in our aim?

Well, if we are really just off by a truly tiny amount, the difference in the trajectory of the whole system will be small for a while (be it the spatial trajectories of the balls on the table bed, or the more abstract trajectory of the system in the phase space).

But sooner or later it is going to start to diverge exponentially, because errors compound.

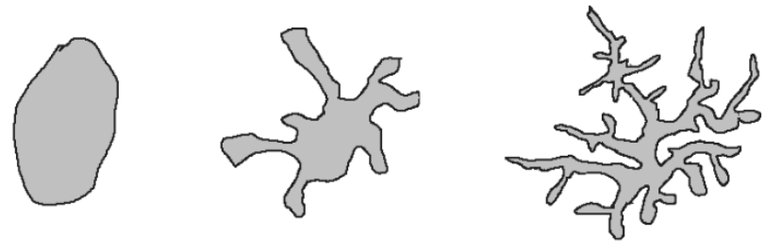


Sooner or later, the two trajectories staring from nearby points in figure 13 are going to deviate from each other.

As we said, that is the character of what is called *chaotic systems*.

And most systems in nature are chaotic.

So that means that, no matter how close we start together, we will eventually depart.

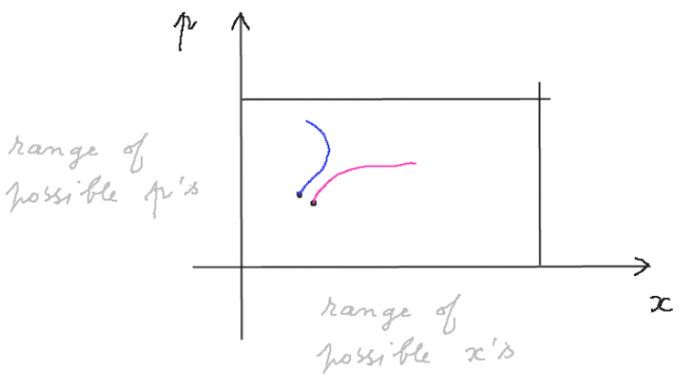


Moreover a small initial blob will grow into into a lot of phase space, even though its volume in the phase space strictly speaking stays the same, figure 11.

The phase space subset which the initial blob evolves into gets extremely complicated and ramified (split up into branches or constituent parts).

Eventually, if we wait long enough till the system gets to thermal equilibrium, the cottony stuff basically fills up the entire possible region of the phase space, figure 13.

And if we do any coarse graining at all, it covers it with uniform probability.



Now let's start from any point in the phase space.

To be practical we can start for instance from the usual initial configuration of the fifteen colored balls arranged at rest in a triangle and the cue ball with an initial impulsion.

But any other point where the sixteen balls have a position and a velocity would do.

Let's follow the trajectory of the system in phase space, figure 14.

Supposing the billiard is ideal, the table bed has no friction, for simplicity there are no pockets, and the balls run forever, then if we wait long enough we will come as close as we want to any possible configuration of positions and velocities of the sixteen balls in the phase space.

Every possible configuration in the phase space will get visited – not exactly but very close.

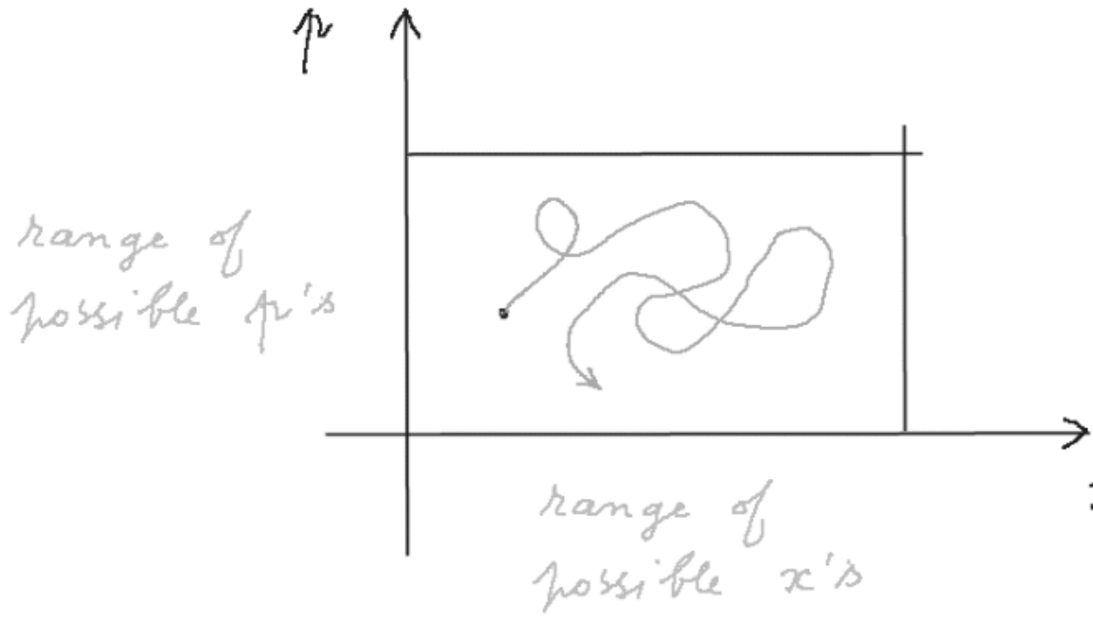


Figure 14: From any starting point, if we wait long enough the subsequent trajectory will ge as close as we want to any point.

That applies to any starting point, even those that are special.

For instance, going back to the example of the gas in the room, if we start with all the molecules in the corner of the room with a certain definite velocity going off in some direction, they will follow a trajectory which will fill up, in the sense explained, the whole possible phase space.

So, after a while, we don't know anything anymore.

The system has reached its maximum entropy.

Nevertheless, in a completely unpredictable and random way that we can't foresee – because we can't follow the system with enough detail –, the trajectory of the phase point will eventually come back as close as we like to the initial phase point where all the molecules were in the corner.

In the billiard analogy, if we start with the triangle of colored balls at rest and the cue ball with an impulsion, the balls will scamper in all directions, they will run all over the table bed, bumping into each other and against the cushions, but if we wait long enough at some time in the future, they will pass through being all arranged as initially and with only the cue ball having some velocity.

How long will it take?

Well it is not infinite.

It does depend on how precisely we are required to reproduce the original configuration.

If we require infinite precision, for most initial points it would never come back.

But if we allow a tolerance, even very small, then the time that it takes to pass back through that phase point to within tolerance epsilon is finite.

So, on the one hand, the entropy increases.

It increases in the *coarse graining sense* because we lose track of the exact position and motion of every particles.

On the other hand, that does not say that we can't return to the original configuration.

It is possible to estimate how long it will take for a system of many degrees of freedom to come back to within a certain tolerance of any phase point.

Take the air in the room, put it all up in the corner, with its kinetic energy, all the molecules for instance going in the same direction with the same speed, that is, a very special state.

Let them go.

They shoot off to the other wall, bounce off that wall, bounce off other walls, start hitting each other, etc.

They go zipping all over the place.

Such a system is very chaotic.

So, after a few minutes at the most, they fill the room in a normal pattern of thermal equilibrium.

To get a feel for what thermal equilibrium looks like:

In the air in the room, under normal conditions, there are thirty thousand billion billion molecules per liter.

At thermal equilibrium, each molecule runs freely, at five hundred meters per second, over a few nanometers before hitting another molecule, or the walls, or our skin.

Despite this pandemonium like picture, this gives us a nice feeling of cozy temperature.

Our body being at 37° Celsius, we contribute however to heating the room.

We are in a constant dynamic equilibrium of transferring heat generated by our basal metabolism, which we need to refuel once in a while with a fresh chocolate brownie.

We observers sit and wait, and wait, and wait.

How long will it take to see all the molecules back again in one corner?

We can be certain that, if the experiment lasts long enough, and the room is a closed system, or at least is kept at constant temperature and energy, the molecules at some point will come back to their initial state.

The question of how long it will take can be answered.

The answer is: a very long time!!

5.12 Questions/answers

Question: Don't we lose information because it is not a reversible process?

Answer: This question mixes up two things.

On the one hand we do lose information, because of the limitation of our resolution power.

If you can look at the trajectory of a minuscule ladybug on the table with perfect eyesight, you always know exactly where it is.

But then if you put on the spectacles of somebody else which blur badly your sight, you will no longer know exactly where is the bug.

You will only perceive a blurry red blotch.

And you will know the bug is somewhere in the blotch.

The same happens to us in the phase space of positions and momenta.

On the other hand the process is perfectly reversible.

So that is not why we lose information.

The process is completely reversible: anything that can happen, can also happen with time flowing in the opposite direction.

But we will talk more about it.

We will talk more about the second law.

In part 8, we will talk about reversibility, and how this paradox gets resolved.

It was one of the great classic paradoxes in the history of physics which puzzled a lot of people.

It drove Boltzmann to distraction.

I don't know if that is what drove him to suicide, but it certainly depressed him.

In fact Boltzmann eventually came to the right answer:

It is not that entropy always increases, but, given any configuration, it is most probable that the next configuration will have more entropy.

In other words, entropy probably increases.

The probability for the configuration aiming itself back to some unusual, unexpected configuration, is always small.

We will give more examples, and try to make this phenomenon a little more intuitive in the next part of our discussion.

We will go into the second law in more detail, and at least provide some kind of explanation.

Question: When we say that most systems in nature are chaotic, is that a principle or an empirical fact?

Answer: Well there is an infinite number of chaotic systems.

And there is an infinite number of non-chaotic systems.

Now we have to start talking about measures of the chaotic character of a system, measures on hamiltonians(energies) and things like that.

It is a good question.

I'm sure that the mathematical people have come up with a correct measure of what is a chaotic system versus a non-chaotic system.

Not all systems are chaotic, that is for sure.

In many cases you can perturb them a little bit, you can change them a little bit, and they will still remain non-chaotic.

It is not the sort of situation where, if you have a system and you change the initial conditions or the rule, meaning the hamiltonian, a very tiny bit, it will evolve quite differently.

What does chaotic mean?

It means that the phase trajectories tend to separate and become quite different, like in figure 13, after a certain amount of time.

I didn't want to define it precisely now, but it is this phenomenon of close by points in phase space following each other for a while then exponentially departing.

So that the system becomes effectively unpredictable.

It is a situation in which predictability effectively breaks down, because, in order to predict a system for a length of time your precision with the initial conditions and your knowledge of the rule of evolution of the system, has to get better and better as you increase the time over which you want predictability.

And necessarily at some point this exceeds the experimental precision we have access to.

In our mathematical and physical models, we often – most of the time – use precise measurements, precise positions, exact numbers expressing one thing or another.

This square has a diagonal of $\sqrt{2}$.

The energy required to lift this mass over one meter is 1 Joule, etc.

We are so accustomed to doing that, that we don’t see what could be wrong with it.

There is nothing wrong with it, but it is just a model.

Nature doesn’t provide exact measures. It provides perhaps exact counts, like "there are three peaches in this basket", but it doesn’t provide exact measures like "this railing is 1 meter high".

Furthermore our measures can vary a bit without affecting our reasonings or predictions, so it took a long time for scientists to see the problem.

Usually there is no problem – with one exception: chaotic systems.

Some scientists had begun to be aware of the chaotic character of certain systems at the end of the 19th century – for instance Henri Poincaré (1854 – 1912) when he studied the stability of the Solar system – and the problems it raised.

In 1963, Edward Lorenz (1917 - 2008), an American meteorologist at MIT, published a startling discovery.

He was studying numerically, with an early computer, the behavior of a set of differential equations linking a few variables describing the motion of the atmosphere.

Every morning he would enter into the computer the numbers spit out at the end the previous day, in order to let the computer continue to crunch all day, thus pursuing the calculations of a trajectory.

He noticed that if he changed a tiny bit (just rounding the numbers more or less) the input data, that would change dramatically the trajectory.

He had found a chaotic system in a remarkably simple set of equations.

The point not to oversee is that the system is and remains deterministic.

Let's take the weather for instance.

Just define it as a collection of molecules doing whatever they are doing instantaneously.

And let's ask how long can we predict the weather?

It depends on how precisely we know the initial conditions (ideally, the exact position and velocity of every molecule!).

The longer the period of time we want, the more precise the initial data must be.

Even longer, even more precise.

Now that is quite different than the harmonic oscillator.

The harmonic oscillator, in phase space, just moves around in a circle with constant frequency, see figure 15, because $E^2 = p^2/2m + kx^2/2 = \text{constant}$.

If we start from initial conditions A in phase space, the trajectory will be the inner circle.

If we start from B, the trajectory will be the outer circle.

The two trajectories are close to each other.

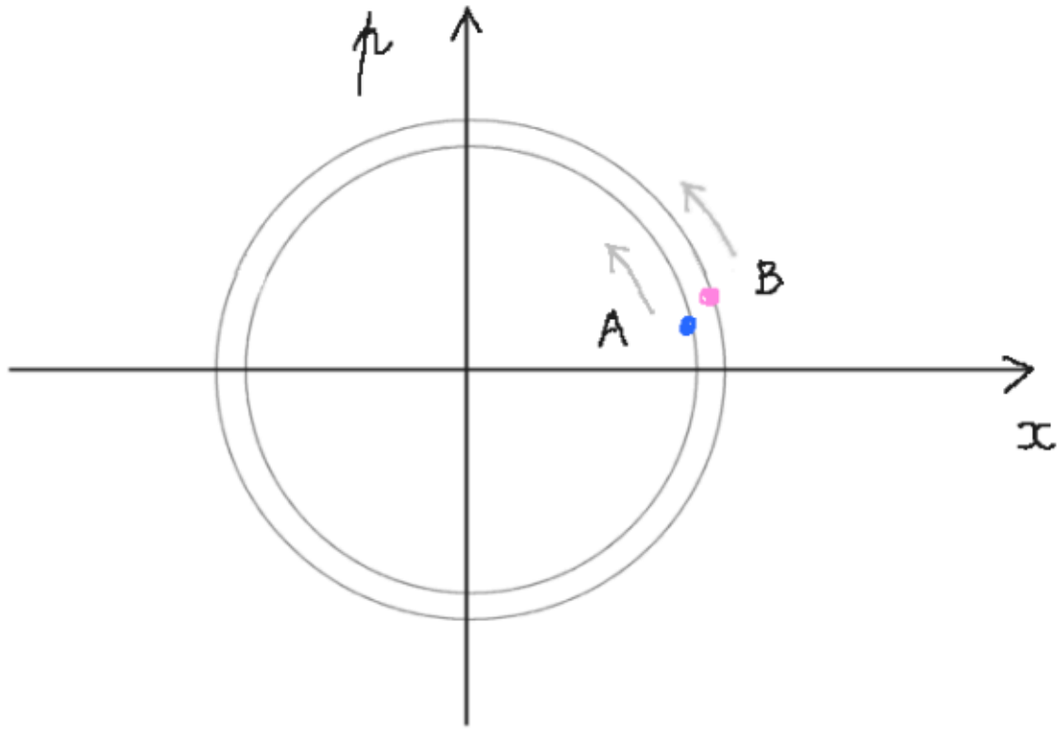


Figure 15: Harmonic oscillator. Representation in phase space. This system is a very non-chaotic. If we start from close by initial conditions, the trajectories are close to each other and remain so forever.

Going to a slightly more complicated system than the simple harmonic oscillator, a particle in an orbit around the Sun, in a true inverse square law, where we do exact Newtonian physics, is not chaotic.

Two close by orbits will remain close by basically forever.

The three-body problem is chaotic in general.

Neighboring orbits will eventually depart from each other.

In that case, predictability is a function of how accurately we know the initial conditions.

To predict satisfactorily for a length of time t, we must know the initial conditions to within ε(t).

And we have

$$\lim_{t \to +\infty} \epsilon(t) = 0 \tag{56}$$

So, in real life situations, necessarily at some point in terms of t, ε(t) will be smaller than the precision our instruments can achieve.

That is what it means to be chaotic.

Question: Don't you also need to know the laws of motion with high precision?

Answer: Yes, you also need to know the laws of motion with high precision, both the laws and the initial conditions.

Question: What is the difference between the three-body problem and the two-body problem?

Answer: One is chaotic and one is not. It is a mathematical property.

In the study of mechanical systems (classical mechanics), one discusses the simple pendulum and the double pendulum.

The simple pendulum is not chaotic. It is very predictable.

But the double pendulum is chaotic.

Suppose we start from two slightly different initial conditions, figure 16.

We can release the double pendulum in configuration (a), or in configuration (b) – not even bothering to give it any initial momentum –, then it will almost immediately behave quite differently.

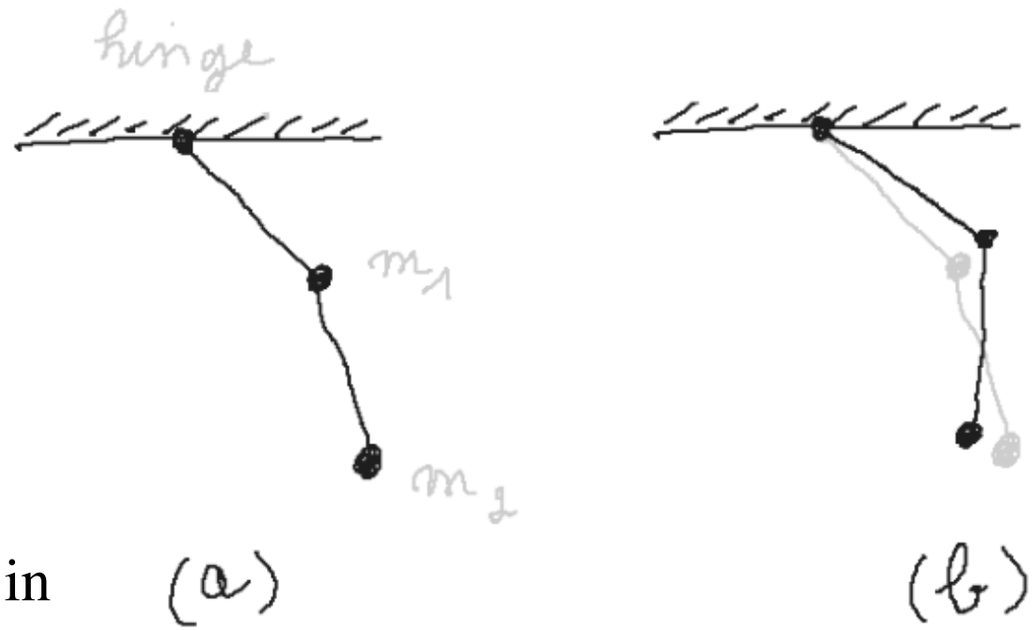
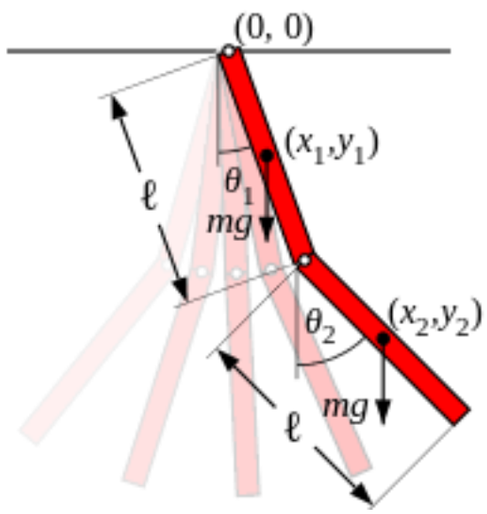


Figure 16: Double pendulum in two slightly different initial conditions.

Have you ever seen a double pendulum in action?

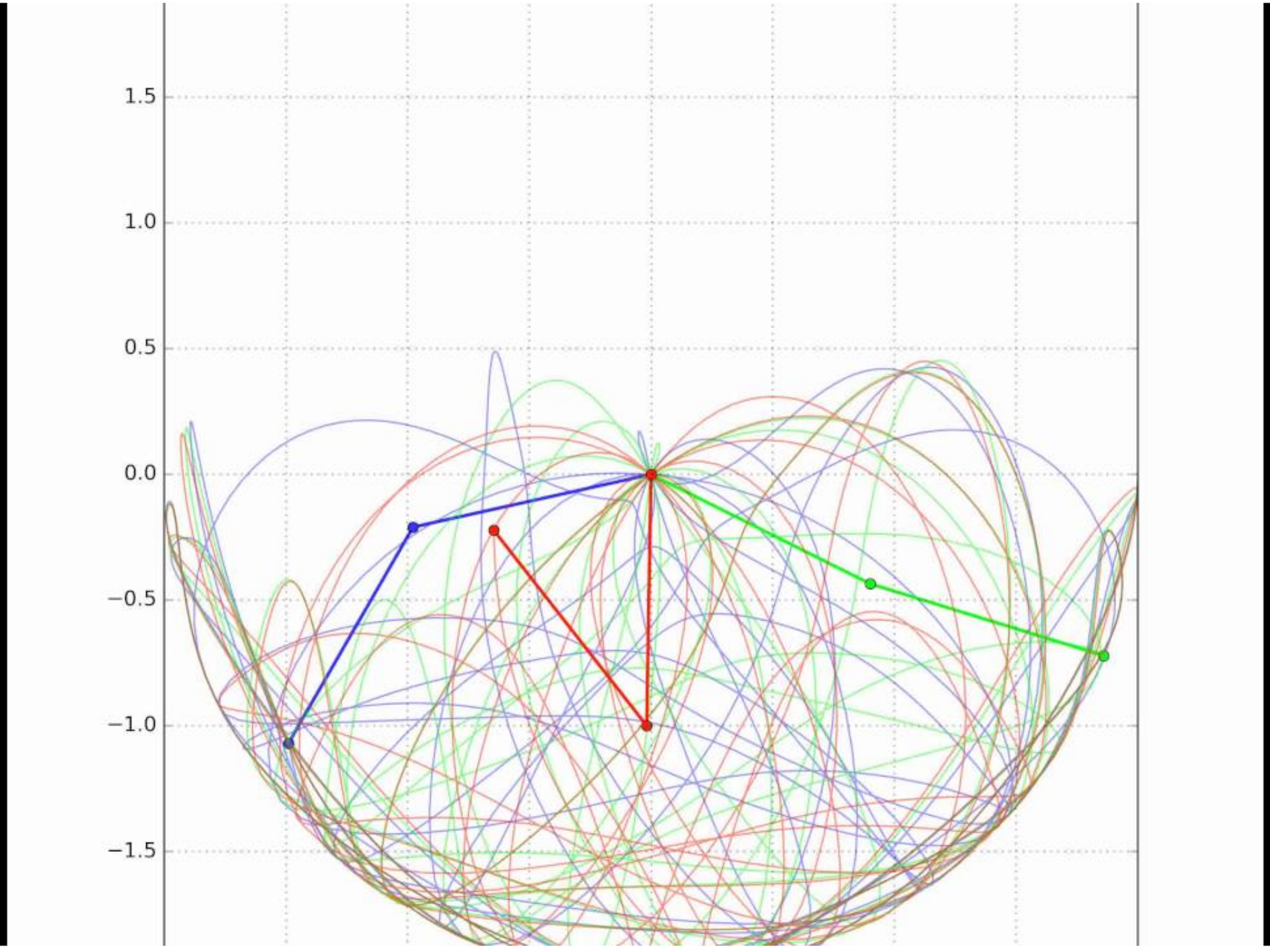
It is quite fascinating. You can follow it. After a little bit of time, it will do something really weird.



For instance, at some time t, all of the kinetic energy may unexpectedly go into one of the masses.

Then the two will move again, not necessarily in the directions you expected.

Very chaotic things happen.


1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

Question: How do we see it mathematically?

Answer: You have to study classical mechanics and write down the equations of motion.

Once we solve them, with numerical methods, we can represent in the phase space the trajectory depending on the initial conditions.

We see that two trajectories starting from close by phase points depart radically

But mathematically speaking, this is a very hard question.

Why are some systems chaotic and other systems aren't chaotic?

And can we readily distinguish them from some general characteristics of their equations?

I would say the state-of-the-art is still primitive.

It is very hard to look at a system and know whether it is chaotic or not, except for the fact that if it is complicated it is usually chaotic.

There are complicated systems, however, which are not chaotic.

And, in 1963, the meteorologist Edward Lorenz found one which was not particularly complicated and was chaotic.

So this is not an easy question.

It is something that mathematicians and mathematical physicists study.

They don't have answers that are very general.

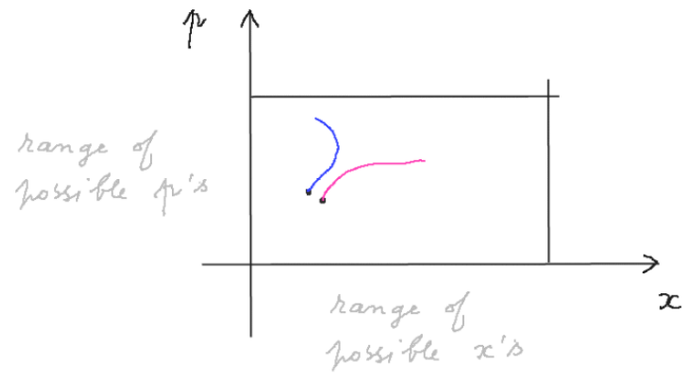
Question: For the double pendulum to be chaotic, we need no friction, don't we?

Answer: Right, the double pendulum with no friction.

Question: So if we knew the initial conditions with infinite precision, there would be no chaos?

Answer: If we knew the initial conditions to infinite precision – whatever that means – we could predict the motion to infinite time.

But I think that is probably a meaningless statement.



The right statement is:

There is a connection between the time period over which you want to predict the behavior of your system and how precise you have to know the initial conditions.

That has to do with how fast the trajectories in figure 13 depart from each other.

But if you give it the slightest perturbation, it will leave that state and not come back.

If the trajectories tend to stay close to each other for a long period of time, the system is rather predictable for a long period of time.

But eventually they will separate.

So most systems are unpredictable in that sense.

Another way to say it is:

Given any length of time that you want to predict, there exists a tolerance or a precision ε, such that, if you knew the phase point within that precision ε, it would allow you to predict for that length of time.

You know, it is the kind of reasoning:

Whatever t, there exists an ε such that and so forth and so on – the kind of mind twister some maths teachers in high school pretend to teach you maths with.

There is a theory by Lyapunov, developed well before chaos had become an active domain of research in mathematics, which studies the stability of motions, and which has also relevant things to say about chaotic equations.

The Lyapunov coefficient is the coefficient of the exponential growth of the separation between trajectories.

Question: Is saying that a system is chaotic the same as saying that its equations don't have a closed solution?

Answer: Chaotic systems certainly don't have closed solutions to the equations.

But the converse is not true.

It is possible to have a system which has no closed solution, and yet which is not chaotic.

Question: What is the relationship between being unstable and being chaotic?

Answer: Chaotic systems may or may not have unstable points.

And an unstable system is not necessarily chaotic.

As we saw, the double pendulum is chaotic.

If we think of one which can rotate 360° around its fixed hinge, it also has an unstable point.

If you release it at rest with the two sticks vertical above the hinge, it is an unstable position.

In theory it will stay there.

Furthermore the stick attached to the fixed hinge can start to rotate on the right or on the left.

You can also think of a pendulum with complete spherical possible rotation if you want to enrich the example.

But being chaotic is a different than being unstable.

It means that from any small blob in the phase space trajectories depart quickly.

There are plenty of systems which have unstable states, but which are not chaotic.

Orbits in the two-body problem are unstable, but they are not chaotic.

The simplest kind of instability is illustrated in figure 17.

There is just one point of instability.

It is at the very top of the hill.

And, if you move the marble ever so slightly from that point, it can roll down in any direction.

So from a small blob at that point in the phase space (position at the top, and zero velocity), the possible trajectories do depart from each other.

But from a blob around any other point in the phase space of that system, possible trajectories don't depart from each other, and are predictable over a long time unrelated to the precision of the initial conditions.

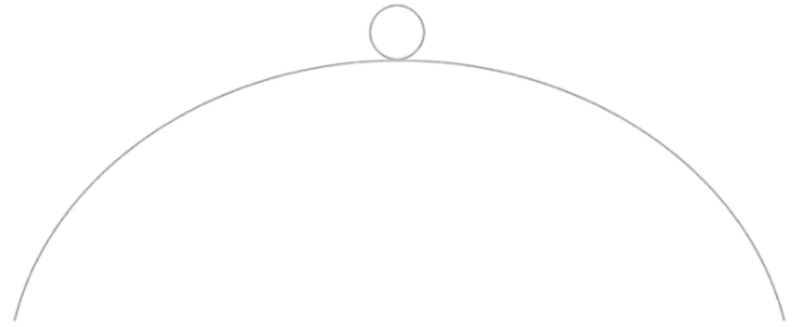


Figure 17: Unstable point in a non-chaotic system. A small perturbation on the marble will let it roll down the hill. And the possible trajectories from there are quite different.

The double pendulum, with a 360° fixed hinge, is of a different nature.

It has one unstable rest point, the two sticks vertical above the fixed hinge, and one stable rest point, the two sticks hanging vertical below the fixed hinge.

And it is chaotic everywhere, meaning that from any small blob around any point, the trajectories will quickly be different.

You can call any point – except the stable rest point – unstable if you like, although some people prefer to talk of an unstable point if, when perturbed the system won't come back to it, but which it won't leave spontaneously.

On the other hand a point is called stable if, when moving a bit away from it, the system comes back to it, whatever trajectory it follows to return.

So being unstable is different than being chaotic, which has to do with the behavior of trajectories leaving small blobs anywhere in the phase space.

In a chaotic system, trajectories starting from any pair of nearby points in phase space depart radically from each other after some time.

8. Entropy, reversibility, and magnetism

8.1 Poincaré recurrences

Let's go back to the second law and talk about Poincaré recurrences – which might also have been named Boltzmann recurrences because Boltzmann (1844-1906) studied them too, somewhat before Poincaré (1854-1912).

Here is a question we might ask.

Let’s suppose we start out with the air all on one side of the room.

This is not an impossible thing to do.

We put up a dividing wall in the middle.

We evacuate the air out of one side of the room, stick it in the other side, and start with initial conditions where all the air is on, say, the right side of the room.

And now we let it go.

What happens ?

It fills the room pretty much uniformly.

It comes to thermal equilibrium.

Entropy increases.

But what if we sit there and wait, and wait, and wait ?

Sooner or later the unlikely will happen.

By accident or by just waiting long enough in time, all of the air will reappear on the right side of the room, or the top half of the room, or wherever.

But let's say the right half of the room.

That is called a *Poincaré recurrence*.

And that is really no different than saying that, if we flip a coin enough times, we will eventually get a million heads in a row.

It is very unlikely if we flip a coin, say, a couple million times.

But, if we do it long enough, such a fluctuation will happen.

So the question is :

How long do we have to wait for the air to all appear on one side of the room again?

Is it a year? Is it ten years? Is it a hundred years?

Is it the age of the universe ?

So let's see if we can get a handle on that.

We start out with the idea of phase space, figure 1.

The phase space is the space of the spatial coordinates of the molecules and the momenta of the molecules.

Of course it is a very high dimensional space.

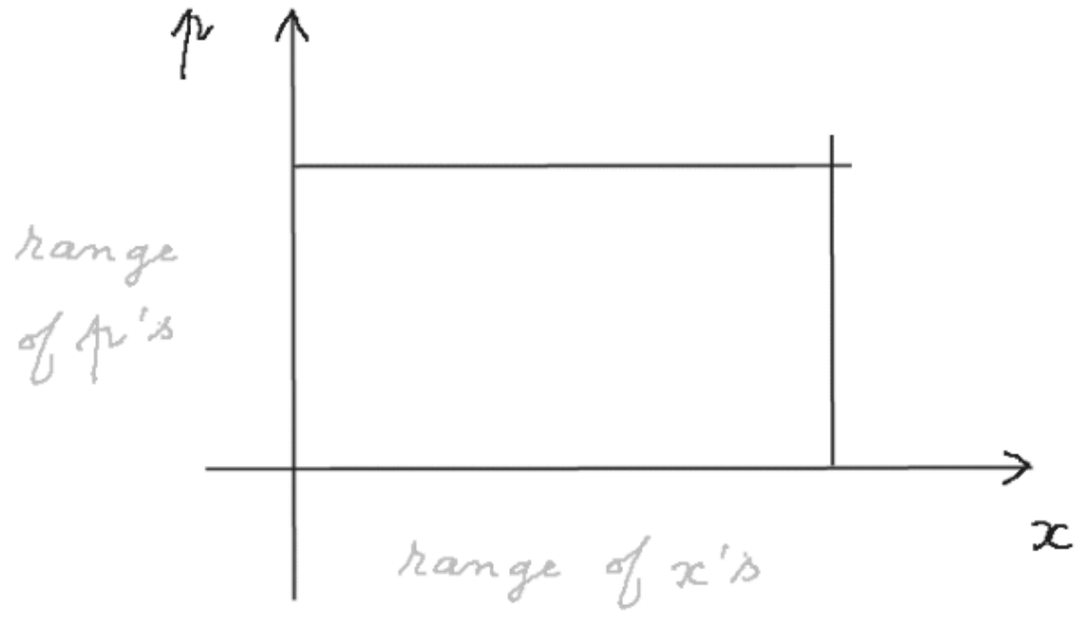


Figure 1 : Phase space of the air in the room.

It has $6N$ dimensions, because each particle spatial coordinate has 3 components, and its momentum also has 3 components.

And there are N particles.

So the x-axis in figure 1 is a schematic representation of $3N$ dimensions, and so is the p-axis

As far as the momentum space goes, it is bounded because if any particle has an enormous momentum, it will also have an enormous energy, and we put only a certain amount of energy in the box and no more.

So the momentum, on the p-axis, is definitely within some bounded range.

It doesn't matter how many particles we have.

It doesn't matter how big the system is.

It does matter what the temperature is.

The higher the temperature, the larger the momentum range.

But if the temperature stays reasonable, then the momentum direction in figure 1 is bounded.

And as far as the x-axis is concerned, it runs from the left part of the room to the right part of the room.

So to say that we started out in one half of the room meant that the phase point was in the region shown in figure 2.

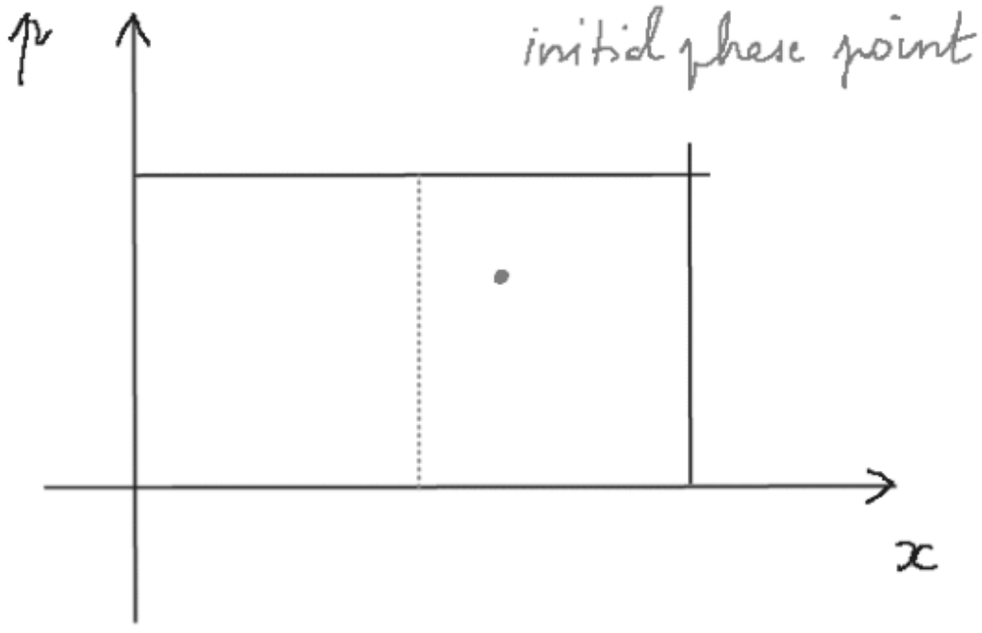


Figure 2 : Initial phase point.

In fact, let's not be detailed about where it is.

It is somewhere in the region corresponding to all the particles on the right side of the room.

The probability distribution is spread out over that region.

Remember also that the entire possible region in phase space, for the point representing the air in the room, is not a rectangle as in figure 2 but a high dimensional parallelepiped.

So imagine splitting the parallelepiped in two parts.

The first part corresponds to all the particles in one half of the room, and the second part is the complement.

And the initial point is in the first part.

That is what figure 2 wants to suggest.

But pay attention to the fact that in reality the two volumes are not equal at all !

Now we wait for a while. What happens ?

The phase point starts to move.

It moves chaotically.

From last part of our discussion, we know precisely what chaotically means.

But for our purposes here it simply means pretty unpredictably.

It is unpredictable not because the laws of physics are unpredictable in principle, but because, in trajectories like for example that of a billiard ball, errors or slight differences in initial detection tend to magnify themselves after a while.

And so, even if we started out along two very similar trajectories, they would quickly depart.

Thus we can imagine that the phase point moves around in the entire range in very complicated ways, and covers pretty much uniformly the phase space, figure 3.

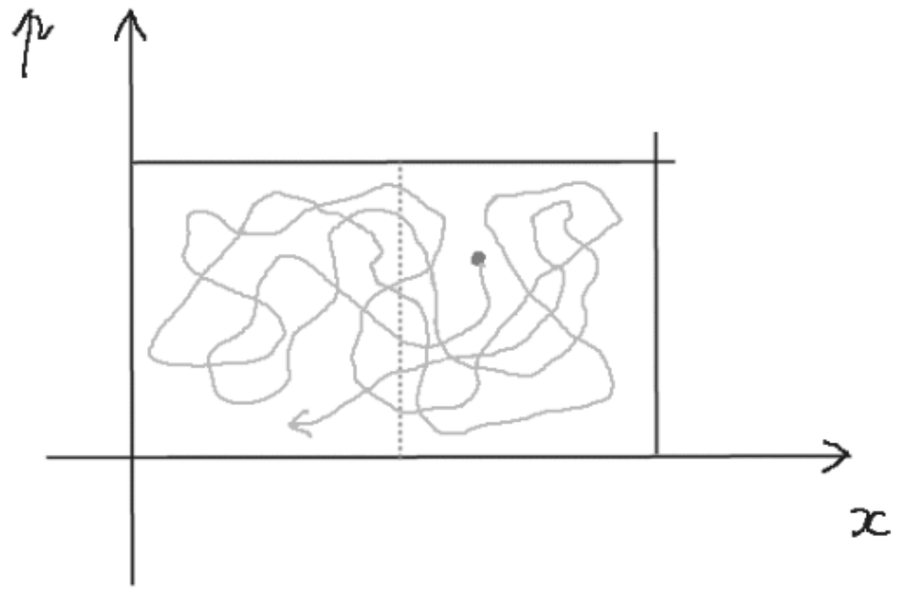


Figure 3 : The trajectory of the phase point covers pretty much uniformly the entire phase space.

We say it covers pretty much uniformly the phase space in the sense that if we coarse grain it, if we blur our sight a little bit, wearing somebody else's glasses, it will look like it has pretty much filled up uniformly the phase space, figure 4.

Then what percentage of the time would you expect the phase point to reside in the region corresponding to all the particles in one half of the room ?

It looks like half the time.

But that is crazy.

We don't expect half the time the molecules to be in the left half of the room or the right half of the room.

The mistake we are making, as said, is we are drawing a picture in just two dimensions, one of position and one of momentum.

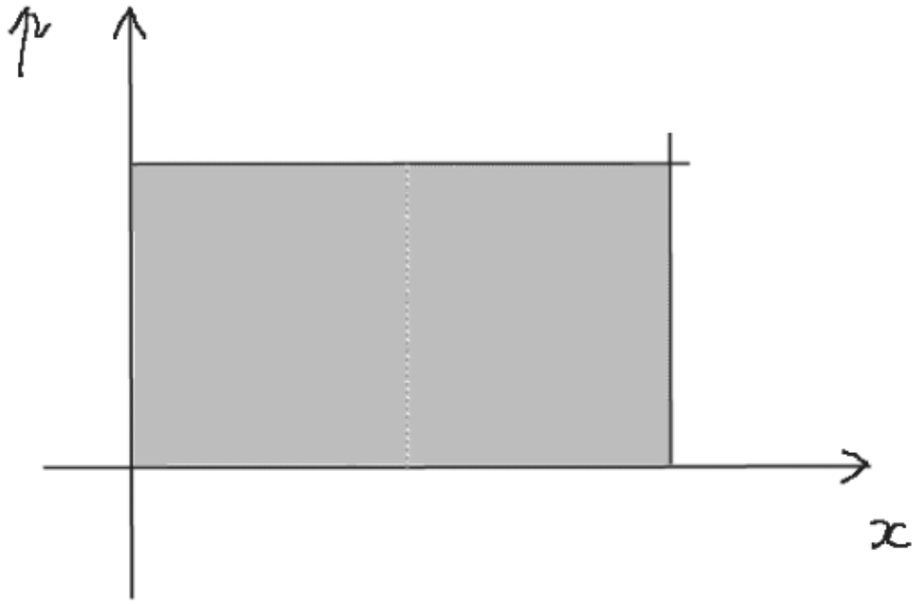
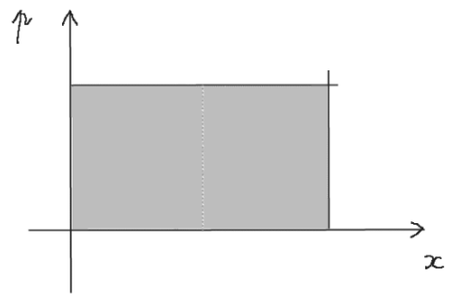


Figure 4 : Coarse graining the trajectory after a while.

If there were indeed only one spatial dimension, then it would be correct to split the area in two equal halves as shown in figure 4.



But what if we have N particles ?

We are talking about the region where, for each particle, say its first spatial coordinate is on the right half of its range.

Let's reason on two particles, i.e. $N = 2$.

And let's forget about their momenta.

Suppose the first particle has spatial coordinates (x_1, x_2, x_3) and the second has (y_1, y_2, y_3).

To say that the first particle is in one half of the room is to say that x_1 is in one half of its range.

Same for the second particle and y_1.

So the volume of phase space where both particles are in the same half of the room is 1/4 of the total volume of phase space.

Consequently, the percentage of the time when the air – made of two particles – is all in one half of the room is 1/4.

What if there are three particles ?

Then the volume of the region in phase space where all three particles are in one half of the room is $1/2^3$.

And if there are N particles, it is

$$\frac{1}{2^N} \tag{1}$$

There is another way to think about it.

Let's take the phase space and identify a subregion of it.

We first consider the room.

The volume of the room is V , and the volume of a small region of interest in the room is v.

Then the spatial volume of the entire phase space is V^N , and the volume of the subregion in phase space corresponding to all the particles in the small region of interest is v^N , figure 5.



Figure 5 : Entire phase space and small region of interest.

So the percentage of the time when all the particles are in v is

$$\left(\frac{v}{V}\right)^N \qquad (2)$$

If $v = V/2$ we get back formula (1).

Observe is that V^N is the exponential of the entropy of the whole system

To be rigorous, we should multiply by a constant factor coming from an integral over the momenta axes, but it is irrelevant in the subsequent reasonings.

Indeed, remember that the entropy of a region of phase space is the logarithm of the volume of that region.

Or, said another way, if all we know is that the system is in some region of phase space with uniform probability, then the logarithm of the volume of that region is the entropy of the system.

We saw that in part 1 and used it in our discussion in part 7.

What is the entropy of the molecules in thermal equilibrium in the whole room ?

It is log V times the number N of molecules.

The order of magnitude of N is between 10^{27} and 10^{30} depending on the size of the room.

But anyway it is a huge number.

And when we take any number less than 1 to that power, we get essentially 0.

Therefore, at any time the probability for the molecules to be all in one part of the room is essentially 0, even if the part of the room is 99.999999% of it

Note that: 99.999999% to the power 10^8 is approximately 0.4.

So that number to the power 10^{20} is absolutely negligible.

What may make the above reasoning dizzying is that we alternate between considerations of volumes in the room and in the phase space.

As soon as we consider a volume v in the room that is smaller than the entire volume V of the room, the corresponding volume v^N in the phase space is absolutely tiny compared to V^N , which is the exponential of the entropy of the system in thermal equilibrium.

So the probability, in a random draw of points in the phase space, that we find ourself in the tiny volume v^N in figure 5, in case we look at half of the room, is

$$\left(\frac{1}{2}\right)^{10^{30}} \tag{3}$$

It is a pretty small number.

If globally, over an immense length of time, the fraction of time that we spend in that odd little region v^N is 1/2 to the 10th to the 30th, starting from that point, or from any point in the phase space, how long do we have to wait till we find ourself back in that region ?

It is the question of return time in probability theory, and more precisely in the theory of random walks or the theory of Markov processes.

It is intuitively natural that the mean return time is of the same order of magnitude as the inverse of the probability at any one time to be in that region

So it is 2 to the 10th to the 30th.

In what units, years, seconds ?

We could make it more precise, taking into account the speed of molecules and things like that, but it doesn’t matter to conclude that we will never see it.

Why did we do this exercise?

We did it to develop our intuition about how gases of molecules behave, and more precisely to understand in what sense systems are reversible.

We see that, if we wait long enough, they will reverse themselves.

If we consider a sealed room, and we let the gas evolve, let's say starting from the odd state where all the molecules are in one corner of the room, or in one half of the room, it will quickly fill the room and come to thermal equilibrium or what looks like thermal equilibrium.

It will spend a long long time there.

But every so often, every once in two to the 10th to the 30th years or whatever, we will find the molecules in half of the room.

Then it will again fill the whole room, equilibrate, look conventional.

And we wait another eon and all of a sudden we find the molecules in the other half of the room or in another corner of the room.

The huge number of molecules in a gas give a peculiar meaning to the probabilistic reasonings about its behavior.

Thermodynamics was developed in the 19^{th} century at first without any considerations of probabilities.

Entropy was identified as a useful quantity in the equation $dS = dE/T$, see equations (17) or (18) of part 2.

But when Carnot wrote that equation, in his investigation of the workings of steam engines, he did not relate it to any probability that would represent our imperfect knowledge of the state of the system.

Highly improbable events like all the molecules going back to one half of the room are to be compared with picking the black ball in the urn in figure 1 of part 1of our discussion.

And the states of thermal equilibrium are to be compared with the series like H, H, T, H, T, T, H, H where the counts of heads and tails are about equal.

In most of the 2^N series of N tosses, they are about equal.

That is the meaning of the law of large numbers : when the number of trials is large enough, the experimental frequency of heads is close to 1/2.

One could say that thermodynamics is about the law of very large numbers.

So probabilities are always equal to experimental frequencies.

In some sense, values observed are always mean values or typical values, states are always typical states, etc.

Thermodynamics was developed at a time when the atomic hypothesis, that is the idea that matter was composed of very small particles, was only a vague proposal made by chemists like Dalton to try and explain the neat ratios of reactants that one always observed in chemical reactions.

Therefore thermodynamics did not relate to molecules, let alone to random behavior.

In fact it is the development of statistical thermodynamics by Maxwell and Boltzmann, but chiefly Boltzmann, and its great successes in explaining so clearly the classical results of thermodynamics, which reinforced the atomic hypothesis.

But you should be aware that it was still disputed by many eminent physicists and chemists as late as 1900.

So we understand how systems like gases of molecules in a box evolve toward maximum entropy.

It is an instance of the law of large numbers.

Yet they are still reversible.

And once in an eon they go back to an atypical state corresponding to the black ball in the urn.

In other words, if you integrate the system or study it over sufficiently long times, you will discover that the entropy sometimes goes down.

Or, if you prefer, the oddness goes up and down, up and down, in a completely time symmetric way.

What is not time symmetric is what happens if you knowingly start in a very odd configuration.

If you start in a tiny volume of phase space, most likely the next thing is to find yourself out of that volume.

If you start in an odd situation with all the molecules in the corner of the room, you expect the next thing to find is the molecule spread out.

Perhaps the following simple example will make it even clearer.

You start from H, H, T, H, T, T, H, H and you randomly change or leave unchanged each of the trials, the next series will most likely also be typical, in the sense that it will also have roughly half heads and half tails.

That is also true if we start from an unlikely series, like all heads.

We naturally go from the typical to the typical.

We also naturally go from the atypical to the typical. But we rarely go from the typical to the atypical.

That is the sense of irreversibility in thermodynamics.

The systems however remain perfectly deterministic and reversible.

These considerations do not exhaust the fascinating topics of information theory and complexity theory.

What is a typical state?

Is it having half heads and half tails?

What about H, H, H, H, T, T, T, T ?

Is it having that approximate equi-repartition also in any subsequence ?

What about H, T, H, T, H, T, H, T ?

Are the decimals of π random ?

What is a meaningful message, as opposed to a garbled one ?

Those questions in which Gödel made great progress didn't drive him to suicide, but did drive him to what is generally described as insanity.

In the later part of his life, when he lived at Princeton near Einstein, Gödel was convinced that he had a mathematical proof of God.

Would that be the (probable) explanation for their similar names ?

So, starting from an odd state and going to a typical one, then to another typical one, and to another typical one, and so forth, sounds like it violates the reversibility of the physical laws.

But in fact, if you waited long enough, you would find the system reversing itself and doing everything imaginable for a closed system.

That is more than an interesting point.

It is a deep conceptual point that enabled Boltzmann to resolve the puzzle of time being apparently only one-way toward more disorder and more entropy, and the laws of motion being two-way.

Of course we will then have to understand why the universe started in a little corner of phase space.

That is a separate issue.

Boltzmann knew that, and he pointed out the problem.

8.2 Questions/answers

Question: Does this have application in cosmology ?

Answer: Absolutely. And there is still this open question:

Why did the universe start in a small corner of phase space?

Cosmologists and physicists work on it.

Question: So the universe is globally going toward more entropy, and will eventually reach a state of thermal equilibrium.

But, after it has spread out, if we waited long enough, we could see it come back once in a while close to its initial state ?

Notice that "the state of thermal equilibrium" for a thermodynamical system is not one state, it is the collection of states $\Omega = \{\omega_1, \omega_2, \ldots\}$, calculated from the maximization of entropy under constraints that we did in part 3, and over which it then roams in a random process.

And, in classical physics, the randomness doesn't come from nature but from our incomplete knowledge of the exact state of the system.

Answer: Yes, under the assumptions of the above reasoning, the universe is globally evolving toward thermal equilibrium.

Then it will "once in a while" go to an odd state, like for instance its initial one.

In other words, if the universe is a closed system, it will just recur and recur and recur.

To avoid this somewhat strange conclusion, and its strange implications (see below), I think the answer is the universe must not be a closed system.

A closed system doesn't make for a good statistical explanation of the world.

Question: We said that the laws of motion are symmetric toward the future and toward the past.

Then we said that time appears to be only one-way, only toward more entropy, and that was the puzzle that troubled physicists.

But we added that in fact time is symmetric too.

What do we mean by time is symmetric ?

Answer: Consider a closed system.

And you have some measure of the localization of the particles.

Suppose the system starts in a corner of phase space.

That is the particles start concentrated around some location, and therefore their localization measure begins with a high value.

If you sit and follow your measure of localization as the sys-evolves, you will see that the particles become quickly less localized.

And then they will be pretty delocalized for a long time.

Suddenly at some unexpected time they will localized again.

So, over a very long period, you will see the localization measure go down and up, down and up...

However the timescale to discover this reversibility – and therefore this symmetry of time –, that what goes up must come down so to speak, or the other way around if you prefer, is a time length of 2 to the 10th to the 30th.

Question: Would the process of going back to some order be symmetric with that of going into disarray ?

Answer: Yes, they would look the same, except with time reversed.

Every now and then you would find the molecules in exactly the right configuration to swoosh into the corner, pretty much the same way they swooshed out of the corner.

In fact you will see everything happen, every possible states and every possible evolutions.

And they will happen in both directions of time.

Question: So this has nothing to do with the fact that we started in a small region of phase space ?

Answer: Yes, that is correct.

The evolution to an odd state doesn't require that we started from an odd state.

But precisely, if we are trying to understand why the world looks the way it does, we must explain why it seems to have started not from any state, but from a very orderly one.

The problem can be formulated as follows – and Boltzmann already knew it, but it has known a renewal in recent years with the discovery of exoplanets.

Suppose the universe really were a closed box and you asked :

What is the most likely configuration to find a planet with people on it ?

What is the most likely possibility ?

You would discover that the most likely possibility is to have uniform gas everywhere, except the smallest possible amount of gas, necessary to make up a planet, having condensed into the planet.

The chances that you see one planet are very small, say ε.

And the chances that you see two planets are vastly more negligible.

In a simple model it is ε^2.

But we won't even have to resort to that latter probability to explain why what we see doesn't really make sense in a closed universe.

If you are asking the question : What should astronomers expect to see out there – conditional on the fact that there is already a planet in the universe with astronomers on it doing astronomical observations ?

What is the most likely thing they will see ?

The most likely thing is they would look out and see nothing, because, even conditionally on the existence of their planet, the chances that they see a second planet are very small.

They are again ε.

Maybe they will see some gas out there.

But they will not see it condensed into another planet.

So by far the most likely thing we should see is our own planet and no others.

What is the probability to see the universe filled with stars ?

Absolutely negligible, unless you knew that in a fairly recent past you started from some very exceptional and unusual starting point.

And then the flow out from that starting point is likely to have certain kinds of structure that a random fluctuation would not have.

It is called the problem of Boltzmann brains.

It bears that name because people are a little bit excessive on it.

They say the most probable astronomer would be a single astronomer's head, looking like Boltzmann's, disconnected from anything else.

The crux of the matter is trying to use statistics and random fluctuations to understand why our world is the way it is.

We consider an unlikely event X – us in the world, with probability ε – and then another unlikely event Y – anybody else out there?

We ask what is the conditional probability of Y given X, or things like that.

We may even use ε^2 if we don't want to take ourselves for granted.

Anyway we try to make it work, it is not a good theory to think that we are just the result of a random fluctuation.

Observe, by the way, that if we were the result of a random fluctuation where things just assemble themselves sort of accidentally, or what is apparently accidentally, into a planet with people on it, we would have no good explanation for the coherence of history.

Why would there be a consistency to the sequence of historical facts if the universe had materialized by a random fluctuation?

Why would all history books describe history the same way ?

Let's not worry about it.

It is not a good theory of nature.

Question: Can't we just say that we live in world that takes a very long time to reach equilibrium ?

Answer: Yes, you could say that.

But you would nevertheless run into the problem of having to explain how we got where we are from a random fluctuation, to explain what we see, to explain the coherence of history, etc.

Serious theoretical cosmologists do worry about it.

But I am not going to try to sell you anything.

It is something that is of concern, that if we want to use statistics and ask what is the most probable thing we should see, given that we are here, we have to take into account all the ways we could have gotten here.

Most of the ways we could have gotten here would be by random fluctuation.

History would not be coherent, etc.

Ok, that is it for the problem of Boltzmann brains.

Question: Isn't the problem solved by the fact that life creates order and decreases entropy ?

Answer: No, no, not at all.

Life doesn't decrease entropy.

You may see order brought about by life, and see something decreasing its entropy, but it is always at the cost of something else increasing its entropy.

The second law does not say that some subsystem of the world can't decrease its entropy.

But it will always be at the cost of some other subsystem increasing its entropy even more.

Question: A living process, however, seems to last for a very long time.

It is not something that comes together then goes away, like, in phase space, the random passage of the system through an odd state.

The fact that life is auto-maintained is a little strange.

Answer: Life on Earth is caused by the flow of energy from the Sun.

The Earth is not in equilibrium.

It is in a stationary configuration.

Stationary means that it stays the same, but there is a flow.

If you have a system which has a flow moving through it, it can create interesting structures.

For example a flow of water through a pipe can create vortices.

Those vortices have a structure, little eddy currents and so forth.

You could imagine that the little eddy currents have enough structure to display interesting properties.

But if you stop the flow by turning off a tap or sealing off the ends of the pipe, then what happens ?

The eddy currents disappear.

The system just returns to a quiescent, dull, boring equilibrium.

Let's also stress that, even in that steady flow situation, of course the total entropy of the whole system is increasing.

Even though you are making a pump and a sink, water coming in one end, going out the other end, interesting eddy currents in the middle, etc., the water comes out warmer than it came in.

So altogether the second law is not being violated.

The same thing is true on Earth.

What is the flow, in the case of life, that allows this kind of apparent violation of the second principle ?

It is the flow of energy from the Sun.

If we sealed up the Earth, didn't let sunlight in, didn't let sunlight out, everything would eventually come to thermal equilibrium.

It would be dull.

There would be no life.

And there would just be a featureless thermal equilibrium.

Life is a kind of eddy currents or little vortices that appear in a moving fluid, the fluid being energy from the Sun.

That is it for questions/answers now. We discussed some vast questions!

Now we can get back to some more mundane things : magnets.

8.3 Magnets

When we talk about magnets in statistical mechanics, we are usually not talking about pieces of iron.

We are talking about mathematical models for a class of systems that have certain features which resemble magnetism.

So, first of all, what is a magnet?

Whatever an ordinary magnet is, it is made up of lots of little magnets.

The little magnets could even be as small as single atoms.

Or they could be little crystal grains.

But, whatever, a magnet is made up of a large number of little magnets.

Typically, at ordinary room temperature – such temperature being considered rather high in this context –, and certainly at very high temperature of a thousand degrees or more, those little magnets are randomly oriented in such a way that the whole sample doesn't have a net orientation.

The orientation is random.

And not only is the orientation of the whole thing random, but the relative orientation of the parts of it are random.

And so there is no net magnetization.

We don't see a macroscopic magnetic field from it.

Now we cool it down.

If the energy which is stored in pairs of these little magnets is such that the magnets like to line up in the same direction, then, as we start to cool it down, we find out that groups or lumps of little magnets tend to be an alignment.

Other lumps of magnets will also tend to be in alignment but in other directions.

So we will find sort of domains which are magnetized, meaning they tend to point in the same direction.

But these domains are still fairly small.

These are experimental facts.

And they are not too hard to understand.

In the energy considerations, when we say that things like to be in the same direction, "like" means that the energy is lower if the little magnets are parallel and pointing in the same directions.

So as we cool the sample down more and more, if the energy is lower when the magnets are parallel, then as we suck energy out of the system more and more of them will want to come into alignment.

And these domains will start to grow.

Eventually, at some low temperature, but not zero temperature, we may or may not hit a point at which all of a sudden these domains become infinitely big, so that the magnets tend to be somewhat lined up everywhere in the same direction.

That is called a ferromagnetic transition.

It is a phase transition.

And it is basically the simplest kind of phase transition.

Certainly at zero temperature we expect the elementary magnets to be all lined up.

Why is that ?

Because, at zero temperature, the only state of importance in the Boltzmann distribution is the lowest energy state.

When the temperature is zero the system is with probability 100% in the lowest energy state.

And in the lowest energy state all the microscopic magnets line up.

This tendency to alignment of the microscopic magnets when we lower their energy is counterintuitive.

Actually it depends on the details of the material.

In a piece of iron they like to align.

This runs contrary to the macroscopic observation that the North Pole wants to grab the South Pole.

It is just a little more complicated.

And that is partly why there aren't that many magnetic materials.

It is right that in a microscopic pair there is also a tendency for them to want to anti-align.

But there are competing phenomena going on.

Question : Don’t we need an external magnetic field for them to line up ?

Answer : No. When their temperature is lowered they will line up spontaneously.

But, without external field, which direction they line up in maybe random.

The lining-up in itself is called *spontaneous symmetry breaking*.

We are going to talk about simple magnetic systems, and their tendency toward order as we cool them – "order" meaning parallelism in this case.

When the microscopic magnets line up, they can line up in this direction, or that direction, or whatever direction.

So which way do they wind up lining up ?

That itself might be determined by the tiniest little stray magnetic field.

Just one elementary microscopic magnet of the sample, one molecule, even one atom, being in an external magnetic field, which will tend to orient it, may govern the way the whole system lines up.

There is a symmetry, which is which way each little thing can point.

If they wind up all pointing in one direction, that symmetry is broken.

That is called breaking the symmetry.

There is no more symmetry – or at least it looks like there is no more symmetry –, but it is spontaneous.

It is not the result of a magnetic field pushing everything in one direction.

It just had to pick a direction, apparently randomly.

It is like when we flip a coin.

We call it random, but the side the coin that ends up showing may be the result of the tiny action of some molecules during its motion.

It is not like the air forcing all the molecules of the coin to arrange themselves to show heads, or tails.

The direction picked may be because of a tiny little stray magnetic field acting on even a single microscopic element.

These are the things we are going to talk about.

We are going to examine what happens at the point where the symmetry is broken, the point where the magnets tend to line themselves up in some direction.

That is a phase transition.

In the case of magnets, it is called a magnetic phase transition.

First of all, don't think about literal magnets.

We use models, which draw their inspiration from magnetism, to apply them to situations in statistical mechanics.

But these model systems would often be quite unrealistic as actual theories of ferromagnetic chunks of iron.

What makes these models inspired by magnets interesting is of course that they resemble a lot of other things in nature.

Moreover they are mathematically simple enough to be studied, and interesting enough to exhibit features like phase transitions.

Let's start with the very simplest magnet – and, as said, don't think of it as a real magnet.

This kind of magnet either points up or it points down, figure 6.

It doesn't get to point in all possible 3D directions.

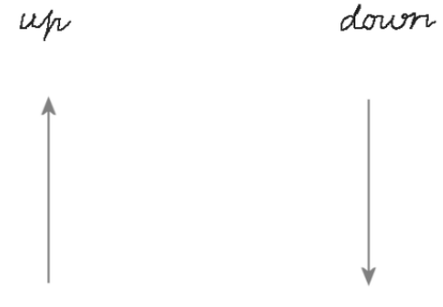


Figure 6 : Simplest magnet. It has only two possible positions : up or down.

You can think of it as heads and tails if you like.

For obvious reasons, I may sometimes myself lapse into calling it a spin.



Figure 7 : Magnet made of a collection of N elementary magnets.

Now let's consider a collection of these elementary magnets.

It doesn't matter how they are laid out on the page, but for simplicity let's lay them out in a line, figure 7.

Some of them are up, some of them are down.

You can think of them as atoms in a crystal.

The atoms have electrical currents, or maybe they have spins.

Electrical currents make little electromagnets.

So each atom is a magnet with a North Pole and a South Pole.

The same applies to spins.

But for the simplest model that has ever been studied, and which we are going to begin with, the atoms point up or point down.

They can't point any other way.

The purpose of this is to be simple.

We want to make some statistical mechanics of this, and ask such questions as :

What is the relative percentage of ups and downs?

What is the energy of the magnet?

And so forth.

So before we begin, if we are going to be talking about statistical mechanics and the Boltzmann distribution, we must have an energy function.

If we want to compute things like $Z = \sum e^{-\beta E_i}$ or $P_i = e^{-\beta E_i}/Z$ we need to know what the energy is.

The magnet of figure 7 is made of N elementary magnets, N being a large number.

And we are going to start with the following very simple model for the energy :

- Each elementary magnet i has a magnetic moment μ, oriented either up or down.
- The orientation of i is represented by a function $\sigma(i)$ which takes the values either $+1$, for up, or -1, for down.
- There is no interaction between the elementary magnets at all.
- Moreover, independently of the little magnets, there is a global magnetic field H in which the system is plunged.
- μ and H are positive numbers. But think of the magnetic field H as pointing downward.
- So the potential energy of the little magnet i is $\sigma(i)\mu H$. It is higher if i is pointing upward than if it is pointing downward.
- n little magnets are up, and m little magnets are down. Thus $n + m = N$.

The energy of the whole system, that is of the magnet formed by the N elementary magnets, is

$$E = (n - m)\mu H \tag{4}$$

which could also be written

$$E = \mu H \sum_{i=1}^{N} \sigma(i) \tag{5}$$

But we won't use the notation with the σ's for the moment.

Moreover since the quantities μ and H are given constants entering the model only through their product, sometimes μH is simply called h.

It is clearer, however, to spell out how the model is built.

How many configurations are there with n ups and m downs, without imposing which ones are which ?

There are [N choose n] denoted by $\binom{N}{n}$.

And the formula is

$$\binom{N}{n} = \frac{N!}{n!\ m!} \tag{6}$$

remembering that $m = N - n$.

That is the number of configurations which have the given energy E of equations (4) or (5).

$\binom{N}{n}$ is called a binomial coefficient.

To see how the binomial coefficients work let us do this exercise.

Prove that

$$\sum_{n=0}^{N} \binom{N}{n} = 2^N \tag{7}$$

Now if $N = 0$, then we have

$$\sum_{n=0}^{0} \binom{0}{n} = \frac{0!}{0!0!} = 1$$

Now if $N > 0$, then we have

$$\binom{N}{n} = \binom{N-1}{n-1} + \binom{N-1}{n}$$

Finally

$$\sum_{n=0}^{N} \binom{N}{n} = \sum_{n=0}^{N} \left[\binom{N-1}{n-1} + \binom{N-1}{n} \right] = \sum_{n=0}^{N} \binom{N-1}{n-1} + \sum_{n=0}^{N} \binom{N-1}{n} = 2^{N-1} + 2^{N-1} = 2^N$$

showing the statement is true for arbitrary N. By induction then it is true always!

Now what is the Boltzmann weight corresponding to one configuration?

By definition it is $e^{-\beta E}$, where E is the energy of that configuration.

The partition function is the sum of all the Boltzmann weights.

So we have

$$Z = \sum_{n\ m} e^{-\beta\mu H(n-m)} \tag{8}$$

where the sum is meant over *all* the configurations, with the appropriate multiple counting.

Now, for each value of $(n-m)$, there is going to be a certain number of configurations.

And that number of configurations is the combinatorial coefficient of equation (6).

In order to simplify expressions, and to trigger memories from high school, let's define the two variables

$$\begin{aligned} X &= e^{-\beta\mu H} \\ Y &= e^{+\beta\mu H} \end{aligned} \tag{9}$$

Then equation (8) for the partition function can be rewritten

$$Z = \sum_{n=0}^{N} \binom{N}{n} X^n\, Y^{N-n} \tag{10}$$

And lo and behold, that is the binomial expansion of $(X + Y)^N$.

So we have figured out that Z is:

$$
\begin{aligned}
Z &= (X+Y)^N \\
&= \left(e^{-\beta\mu H} + e^{+\beta\mu H} \right)^N \\
&= 2^N \left(\frac{e^{-\beta\mu H} + e^{+\beta\mu H}}{2} \right)^N \\
&= 2^N \left(\cosh(\beta\mu H) \right)^N
\end{aligned}
\tag{11}
$$

The factor 2^N is a multiplicative constant.

We know that, in a partition function, it has no importance, because we shall take the log and then differentiate.

And when we differentiate a constant we get zero.

But let's leave it there for the sake of completeness.

Now, suppose we are interested in the question :

What is the relative percentage of up spins and down spins?

This quantity has a name.

It is called the *magnetization*.

The magnetization is 0 if there are as many up spins as down spins.

It is positive if there are more up spins than down spins.

And it is negative in the opposite situation.

The number $(n - m)$ is the difference between up spins and down spins.

So it is linked to the magnetization.

The usual definition of the magnetization is

$$M = \frac{(n-m)}{N} \tag{12}$$

So it becomes a number between -1 and $+1$.

We can view it as the magnetization per little magnet if we like.

It is the bias for each elementary magnet toward up or toward down.

The magnetization is clearly related to the energy.

We will just write a few equations and then we will be able to use the partition function.

To start with

$$E = NM\mu H \tag{13}$$

We will use definition (12) and equation (13).

Let's keep them in mind.

We will come back to them.

We have introduced many definitions, but *magnetization* is an important one.

It is roughly speaking the probability of being up minus the probability being down for a given spin.

How can we calculate the magnetization ?

One easy way is to calculate the energy of the system.

If we know the energy of the system, and we know the number of particles (we use indifferently the terms particles, little magnets, or spins, to describe the elementary magnets making up the system in figure 7), and we know μ and we know H, from equation (13) we can calculate the magnetization.

So the first thing we shall calculate from the partition function is the average energy.

From that we can read off the magnetization.

Of course in the same way there is an energy per configuration and there is an average energy, there is a magnetization per configuration and there is an average magnetization.

Equations (12) and (13) hold for each configuration.

And they also hold for the average, or expected, values.

What we are interested in is the average magnetization of the system under various circumstances which we shall look at.

So it should be denoted, more properly,

$$M = < \frac{(n-m)}{N} > \tag{14}$$

It is the expected value of the magnetization of the configurations when they have the Boltzmann distribution of probability $P_i = e^{-\check{\beta} E_i}/Z$.

Remember that generally speaking in statistical mechanics we consider a system which can be in a collection Ω of states, see part 1, with a probability distribution, or measure, P over Ω.

Each state, or configuration, has various characteristics, like its own energy, or its own magnetization.

But unless otherwise specified, when we talk about the energy, or the magnetization, or any other characteristics of the system, we mean the average value over the probability distribution P.

And, unless otherwise specified too, the probability distribution P is the Boltzmann distribution at equilibrium – which is the one which maximizes the entropy, see parts 3 and 4

Unlike energy, or magnetization, which can be defined for each state, the entropy is only defined for the collection Ω of states and the probability measure P on it.

P is a measure of our incomplete knowledge of the exact state $\omega \in \Omega$ the system is in.

The probability measure which maximizes the entropy under the constraint that the average energy is a given, is called the Boltzmann distribution.

The Boltzmann distribution should not be confused with another fundamental concept of statistical thermodynamics introduced by Boltzmann too, and which we haven't talked about : the Boltzmann equation, which governs the time evolution of the probability distribution for a system not in equilibrium.

Now what do we do to calculate the average energy ?

We take the logarithm of the partition function, differentiate it with respect to β, and change the sign of the result.

Before we do this calculation, let's build our intuition about this system.

Remember that it is a very simple model of N little magnets plunged into a magnetic field H as described in the comments to figure 7.

The energy of each particle is positive if it is up, negative if it is down.

Systems like to have lower energy, meaning to say that the Boltzmann distribution favors lower energy.

Which way will it be at zero temperature?

All little magnets down.

So, at zero temperature, what do we expect the magnetization to be ?

If everybody is down, from equation (14) that means the magnetization will be -1.

What do we expect at infinite temperature ? Fifty fifty.

At infinite temperature everything is maximally random.

All states are equally probable.

Therefore, at infinite temperature, we expect the magnetization to be 0.

So M goes from -1 at zero temperature, to 0 at infinite temperature.

This is what we expect.

At no point will M be positive.

The average magnetization will not be positive because of the bias down.

The infinite temperature will defeat the bias.

At infinite temperatures it is just so random that a little bit of magnetic energy is unimportant.

As we saw, instead of calculating the average magnetization, we are going to calculate the average energy, using the equation

$$E = -\frac{\partial \log Z}{\partial \beta} \tag{15}$$

There is a little bit of algebra to do here.

From equation (11), and dropping the constant coming from 2^N, we get

$$\log Z = N \log \left(\cosh(\beta \mu H) \right) \tag{16}$$

Notice first of all that is proportional to N.

That is a good thing because typically energies when we differentiate will be proportional to N.

Let's differentiate equation (16) with respect to β.

We apply the chain rule for differentiating a function :

$$(f \circ g)' = (f' \circ g) \times g' \tag{17}$$

also written

$$\frac{df(g)}{d\beta} = \frac{df(g)}{dg} \frac{dg}{d\beta} \tag{18}$$

to explain heuristically the rule.

It requires, however, a bit of elementary differential calculus to justify cleanly.

In equation (16), f is the function logarithm, whose derivative is $1/f$.

And g the hyperbolic cosine, whose derivative is hyperbolic sinh.

So we get

$$\frac{\partial \log Z}{\partial \beta} = N \frac{\sinh(\beta\mu H)}{\cosh(\beta\mu H)} \mu H \tag{19}$$

Is that the average energy ? Not quite.

We must put a minus sign

$$E = -N \tanh(\beta\mu H) \mu H \tag{20}$$

Now, from equation (13), to get the average magnetization, we divide by N and by μH.

$$M = -\tanh(\beta\mu H) \tag{21}$$

Now all we have to do to understand the system is understand what the *tanh* function looks like.

Incidentally β is one over the temperature.

So we just want to plot this function as a function of the temperature.

μ times H is only a number.

It is not so interesting.

We could absorb it into β and plot *tanh(βμH)* as a function of $\beta\mu H$.

They come in together.

But first of all, simply, what does the *tanh* function look like?

You can do it by yourself.

It can be figured out easily from what *cosh* and *sinh* are.

Remember that

$$\cosh x = \frac{e^x + e^{-x}}{2}$$
$$\sinh x = \frac{e^x - e^{-x}}{2} \tag{22}$$

The functions *sinh* and *cosh*, for very large values of the argument, become equal to each other, because the term e^{-x} becomes negligible.

They are then basically both exponential of x, divided by two.

On the other hand, when the argument becomes very small, the *sinh* tends to 0, and *cosh* tends to 1. Therefore *tanh* tends to 0.

To see what is the slope of *tanh* at 0, we can expand *sinh*.

It is easy to see that $sinh \approx x$.

So the first derivative of *sinh* as well as *tanh*, at zero, is one.

Here is the graph of *tanh*, figure 8.

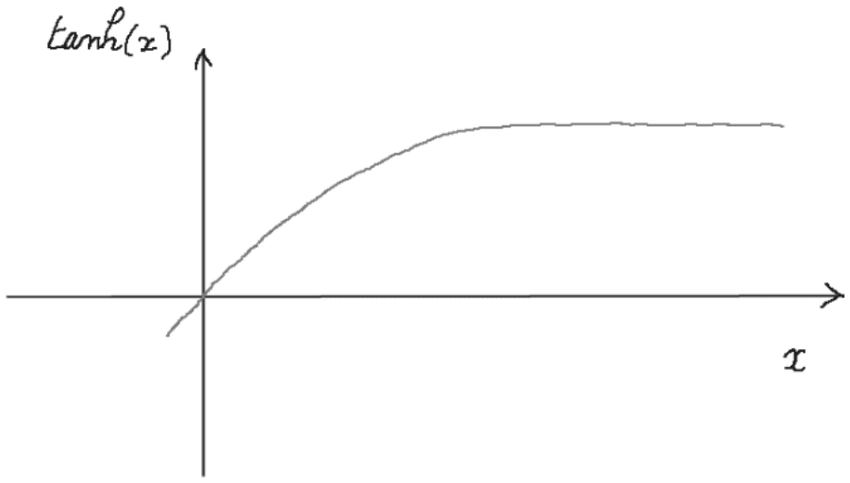


Figure 8 : Graph of hyperbolic tangent.

The function is symmetric about the origin, but we are not concerned with negative values of the argument in our study of M.

Indeed we want to study the value of $-tanh(\beta\mu H)$ as a function of temperature.

Since $\beta = 1/T$, we see that when the temperature becomes very high, $\beta\mu H$ becomes very small, and the average magnetization is zero.

Whereas when the temperature be-comes very low, β and $\beta\mu H$ become very large, and then $-tanh(\beta\mu H)$ tends to -1.

Going back to the physical meaning of these results, when the temperature is very low, all the spins are down, the energy is minimum, and $M = -1$.

When the temperature is very high, $\beta = 0$, the entropy is maximum, and the magnetization 0 as expected.

The graph in figure 8 just fills in the exact details for the problem of how the magnetization goes from -1 at low temperatures to 0 at high temperatures.

Notice that the evolution of the magnetization from 0 to -1 when we go from high to low temperatures is continuous.

There is no jump in the magnetization at some temperature.

This is an important point we will now take up, and pursue in the next part of our discussion.

The simple system we just studied has no *phase transition*.

The reason is that it is too simple.

The first interesting system that has a phase transition is the two-dimensional Ising model.

But first of all we are going to do the one-dimensional Ising model, because the system we just studied doesn't belong to the class of Ising models.

It was even simpler than them.

In the 1920's Ernst Ising was as student of Wilhelm Lenz at the University of Hamburg.

Lenz gave Ising, as a subject for his Ph.D. thesis, the study of a model of magnet made of a linear chain of small magnetic moments coupled by interactions between neighbors.

What Ising had to determine was whether in that model there was a phase transition.

Ising solved the problem and gave the correct answer that in one dimension there is no phase transition.

Then he wrote that he thought in two dimensions there would not be a phase transition either.

That was wrong.

Actually it was a much harder problem to tackle.

Two physicists, Kramers and Wannier, solved it in the 1940's.

Ising models turn out to be very important in statistical mechanics and other branches of physics.

And the Ising name is most famous.

It might be thought that his fame is somewhat larger than his merits.

So it often goes in science.

The interesting thing about Ising models is that they are symmetric between up and down.

Therefore, if there is any magnetization, it is because somehow the system has **spontaneously** broken the symmetry.

In the one-dimensional Ising model that does not happen.

In the two-dimensional Ising model it does happen.

Let's define all of the Ising models.

The energy is not stored particle by particle.

And there is no external field.

So, if the little magnets didn't interact with each other, there would be no energy.

And, if there's no energy, all configurations are equally likely.

In the simplest Ising model the magnetic field that each spin sees is due to its two neighbors.

If both neighbors are up, it feels the magnetic field up.

If both neighbors are down, it feels the magnetic field down.

And, if one is up and one down, it feels no magnetic field.

In other words, the energy is associated with **pairs** of neighboring spins.

Now we have to make a choice of model.

We can choose the energy of a pair to be minimum when the two spins are pointing in the same direction – we call them *aligned*.

Or we can choose the energy of a pair of spins to be minimum when they are *antialigned* or *unaligned*.

We will come back to the variables σ's defined above to express that.

Let's decide that the energy of a pair is minimum if they are aligned, and maximum if they are unaligned.

Then we express the energy of a pair as follows

$$E = -j\ \sigma(1)\ \sigma(2) \tag{23}$$

j is just a number with units of energy.

The particles are on a lattice which is just a line as in figure 7.



And particle 1 and particle 2 are two neighboring particles.

Later on, we will allow the lattice to be two-dimensional or more.

1 and 2 are called neighboring sites on the lattice.

Equation (23) is the energy of the 1-2-pair.

If the two spins in the two sites are unaligned, the product of their σ's is -1.

So their energy is j.

Whereas, if they are aligned, the energy of the pair is $-j$.

Therefore the energy is lower when they are aligned.

Now supposing we have a line of them, and each one is interacting with its neighbor, then we can write for the energy of the whole system

$$E = -j \sum \sigma(n)\, \sigma(n+1) \tag{24}$$

This is the sum of the energy in neighboring pairs.

And each of these neighboring pairs is counted once.

Let's pore over equation (24) for a moment.

What do we expect to happen at infinite temperature ?

The general rule is that infinite temperature is just random chaos.

Chaos here is used in its general English sense, not in the technical sense used for the behavior of certain equations when we change initial conditions a bit.

Everything is equally likely.

The spins are randomly up or down, and this leads to zero magnetization

The products $\sigma(n)\ \sigma(n+1)$ on the average will be equal to zero.

This looks disturbing because $\sigma(i)$ is either -1 or $+1$.

So how come the product of $\sigma(n)$ and $\sigma(n+1)$ can be zero?

But we are talking about its average.

And don't forget that the average of a random variable is not necessarily one of its possible values.

The average of a well balanced die is 3.5 for instance.

So high temperature corresponds to zero energy – which is high energy in our model – and to zero magnetization.

What about zero temperature?

The spins will be aligned.

But which way are they going to be aligned ?

Are they going to be all up, or all down ?

Unlike in the first model studied above, where figure 7 was plunged in an external field H, we can't tell offhand.

There are two ground states now.

Ground state means state of minimum energy.

And the two ground states will come out with equal probability.

Now let's add one more thing.

Let's suppose that there is an external magnetic field.

But it is only acting on one particle.

One out of 10^{23} particles is subject to a little stray magnetic field.

And let's say that magnetic field is along one axis, i.e. up or down.

Then what is the ground state ?

Now the ground state has a definite orientation.

Even if that magnetic field is very small, the ground state still has a definite orientation.

And at strictly zero temperature, the Boltzmann distribution always favors infinitely strongly the lowest energy state.

That means that even with the tiniest little stray magnetic field, the Boltzmann distribution will favor all of the spins pointing along one direction.

If you were to apply that tiny magnetic field, let the system come to equilibrium at zero temperature, and then remove the magnetic field, then the system would remember it.

Everybody is holding everybody else in place, and the possibility of them all simultaneously jumping to the opposite state is remote if there is enough of them.

That is what is called *spontaneous symmetry breaking*.

In this case it is very simple.

And this example, characterized by equation (24), has a symmetry.

We can actually say quantitatively, or mathematically, what that symmetry is.

A symmetry is usually represented by a mathematical operation on the degrees of freedom.

What operation would you do on σ to change from up to down ?

Simply, multiply by -1.

Let's go back to the earlier case where the energy was proportional to just σ itself, not σ times a neighboring σ.

Does this have a symmetry ? No.

The energy itself changes sign when you change σ.

And that is not a symmetry.

Symmetries are actions that you can do that don't change the energy.

Now what about the simplest Ising model that we studied next ?

Furthermore, suppose there is only one pair.

In equation (23), you change the sign of one spin, just one of them, for instance $\sigma(1)$ but not $\sigma(2)$.

Is that a symmetry? No.

The energy changes, figure 9a.

But what if you change them both ?

Then the energy of the pair doesn't change sign, figure 9b.

For instance if you take them from two up to two down that is a symmetry, figure 9b lower case.

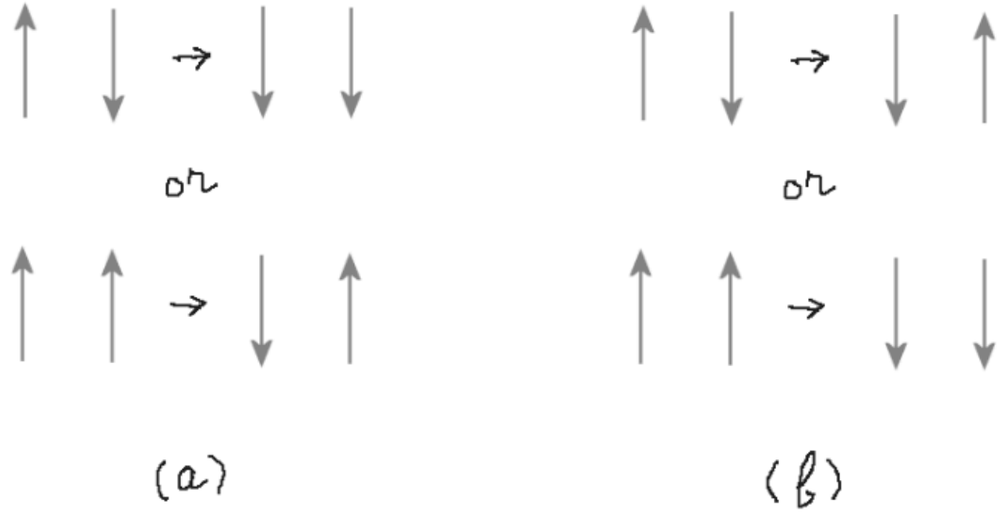


Figure 9 : Ising model, one-dimensional, and with only one pair. On the left part, we change only one spin. On the right part we flip both spins.

But now consider the whole vast array of spins, or magnetic moments, figure 10.



Figure 10 : Ising model, one-dimensional, many elementary magnetic moments.

What if we change all of the sigmas simultaneously?

We can write formally the equation of the transformation

$$\text{for all } i,\ \sigma(i) \rightarrow -\sigma(i) \tag{25}$$

In other words, we replace every σ by minus its value.

Then the energy doesn't change.

And that is what it means to have a symmetry.

> *A symmetry is an operation we can do on the coordinates of a system such that, no matter what state we started from, the end state has the same energy.*

That ensures that in some sense there is no bias to up or down.

If the system is going to flop itself all simultaneously into up at zero temperature, that means it could have also flopped itself into down.

There is no way to predict in advance, unless you know that tiny little stray magnetic field.

There is no way to predict in advance which way it is going to go.

But it is going to go one way or the other, because the Boltzmann distribution says we have got to be in the ground state for 0 temperature.

As we said, the little tiny stray magnetic field will determine which one it is (like the result of flipping a coin will be determined by tiny, hard to measure, atmospheric conditions, but it will end up heads or tails with equal probability), but it will be one of them.

The next thing we want to do is solve the one-dimensional Ising model.

From equation (24), we want to calculate the partition function for the system.

We will do that in part 9 of our discussion.

We will work it out.

And we will see that it does not have a phase transition at a finite temperature.

Nothing funny happens at finite temperature in the one-dimensional case as Ising showed in his Ph.D. thesis.

Then it took a few more years for Kramers and Wannier to prove that if the system is a two-dimensional lattice, then, contrary to what Ising believed and wrote, there is a phase transition.

That is a beautiful story, of which we will try to get a whiff.

9. The Ising model

9.1 Introduction

We began to study the Ising model in the last part of our discussion.

It is actually a class of models sharing certain features.

They began as the following problem given by the physicist Wilhelm Lenz to his doctoral student Ernst Ising in the early 1920's at the University of Hamburg :

In a model of magnet made of a linear sequence of little magnetic moments where only neighbors are energetically coupled is there a phase transition in the magnetism of the system ?

Ising solved the one-dimensional problem correctly, and showed that it didn't have a phase transition.

On the basis of that he believed and wrote in his thesis that the Ising model in any dimension does not have a phase transition.

But that was incorrect.

We are going to solve the Ising model in one dimension.

We are not going to solve the higher dimension Ising models exactly.

They are hard maths and not within the scope of this class

But we will use an approximation method that is physically very intuitive.

And we will see that in higher dimensions, when their number is sufficiently high, there is a phase transition.

As stated earlier, there is a phase transition, when we lower the temperature, as soon as the number of dimensions is two, which is still very difficult to show exactly mathematically.

We will show this using a **computational** model approach.

Generally speaking, a phase transition means a sudden change in the properties of the system as you vary a parameter, for example the temperature.

But let's go slow, taking steps.

So let us first go back to the simple example – even simpler than an Ising model – that we studied in the last part of our discussion.

We studied a problem in which there was a collection of little magnets independent of each other, i.e. with no coupling, plunged in a magnetic field H.

Each magnet could be up or down.

And there was an elementary energy function E for each little magnet.

The energy function differed between up and down.

Its expression was

$$E = \mu H \sigma \tag{1}$$

where μ was the magnetic moment of the little magnet, H was the magnitude of the external magnetic field, and σ was $+1$ or -1 depending on whether the little magnet pointed up or down.

We shall change a little bit the notation for this energy function.

First of all, since μ and H appear only through their product, we denote it j.

It is a standard notation for what we will be doing later.

And by convention we will put a minus sign in front.

Thus the elementary energy function becomes

$$E = -j\sigma \tag{2}$$

The minus sign doesn't make any difference to the physics, because it really is just a redefinition of what we mean as up and down.

If the original energy favored down, then, if we change the sign, it favors up.

But, other than that interchange of up and down, there really is no difference.

Physical systems always tend to favor lower energy.

In equation (2), when σ is negative, E is positive.

And when σ is positive, E is negative, that is lower.

So the system favors σ positive, or the little magnet pointing up.

We studied a whole group of such little magnets, figure 1.

And we did some combinatorics.

How many states are there, given that there are n little magnets up, and m down ?

It is the customary number $\binom{N}{n}$, etc.

We did it that way to see carefully all the details.



Figure 1 : System made of a collection of elementary magnetic moments. The elements are uncoupled among each other. The whole system is plunged in a magnetic field.

But we could have done something much simpler.

We have already learned that we can always think of a system as being a small system plus a heat bath.

So focus on any one of the elementary magnetic moments, for instance the one highlighted in figure 2.

We assume that it is in thermal equilibrium with its environment – its environment being the rest of the magnets.

Those magnets of the environment are the heat bath.

And they bring the highlighted little magnet to equilibrium.



Figure 2 : System viewed as a small system plus a heat bath. The highlighted magnetic moment is the small system. The rest is the heat bath.

Now that the system of interest has become only one spin which we focus on, we can write the partition function for it.

It is very simple.

From equation (2) we get

$$Z = \sum_i e^{+\beta j \sigma_i} \tag{3}$$

The plus sign in the exponent is because there is a minus sign in equation (2), and, as we are familiar with, the general form of the partition function is $Z = \sum e^{-\beta E_i}$.

In equation (3), the sum is over all the possible configurations for the little magnet.

There are only two : σ equals +1 or −1.

So equation (3) simplifies into

$$Z = e^{\beta j} + e^{-\beta j} \tag{4}$$

That is the partition function for just one spin.

When we calculated it, thinking of the whole system as one system, do you remember the answer we got ?

We got equation (11) of part 8, which was

$$Z = \left(e^{\beta j} + e^{-\beta j} \right)^N \tag{5}$$

It is also the right-hand side of equation (4) to the *N*-th power.

This makes sense because, whenever we have independent systems, the partition function of the big system they form is simply the product of the partition functions of each system.

And indeed, in the simplest model we studied in the last part of our discussion, each spin is independent of the other spins.

There is no coupling between spins.

So we got a factor $e^{\beta j} + e^{-\beta j}$ for each spin.

And the partition function for the whole collection of spins is that factor to the *N*-th power.

In other words, we can really short-circuit all that stuff about combinatorics and focus on one spin at a time.

So let's concentrate on the partition function of one spin, that is equation (3).

It can be rewritten

$$Z = 2\cosh(\beta j) \tag{6}$$

When we did it in the last part of our discussion, we got *2cosh(βj)* raised to the power the number of spins.

When we take the logarithm of that – which is the interesting thing – it simply becomes the sum.

For example the energy of the collection of spins, which is just minus the derivative of the logarithm of their partition function, is nothing but the sum of the energies of the individual spins.

Here we are concentrating on one spin at a time, one small magnet at a time, and we calculate exactly the same energy for each of them.

So let's calculate the energy :

$$E = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} \tag{7}$$

which is the same as minus the derivative of the logarithm.

The derivative of Z is twice the derivative of the *cosh*.

That is *2 sinh*.

But then we have to multiply by the derivative of the argument βj with respect to β, which gives another j.

So we get

$$\begin{aligned} E &= -\frac{1}{Z}\ 2\ \sinh(\beta j)\ j \\ &= -j\ \tanh(\beta j) \end{aligned} \tag{8}$$

That is the average energy of one of these little spins in figure 2.

We can also ask :

What is the probability that the spin is up versus down ?

Or what is the average ?

σ is the variable that can take the value +1 or −1.

But what is the average σ?

Well, from equation (2), the energy in each configuration of the spin is $-j\sigma$.

Equation (8) is the average energy.

Then it is pretty clear that the average σ is the right-hand side of equation (8) without the $-j$.

When we write E, unless specified otherwise, we talk about the average energy.

For the average σ, for maximum clarity, we can go back to our heavier notation $< \sigma >$.

So let's summarize this

$$\begin{aligned} E &= -j \ \tanh(\beta j) \\ < \sigma > &= \tanh(\beta j) \end{aligned} \tag{9}$$

We are going to need this.

But let's just draw a picture of it.

Figure 3 is a picture of the function hyperbolic tangent.

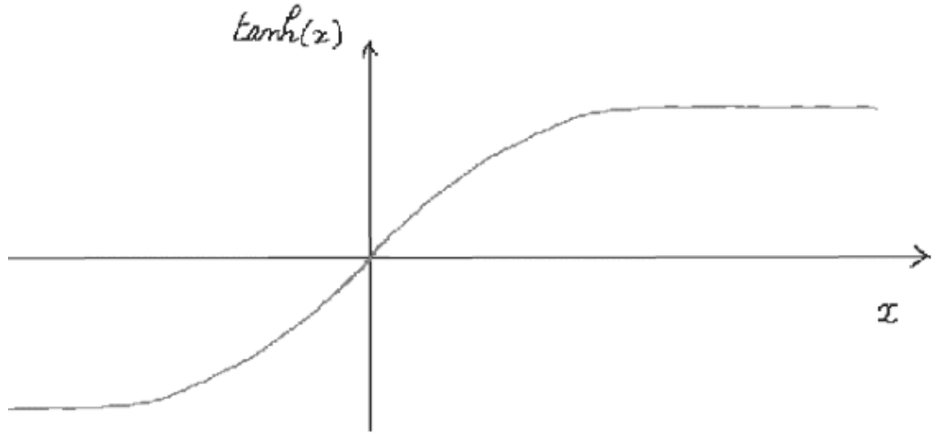


Figure 3 : Function hyperbolic tangent.

This is hyperbolic tangent of x.

In the expression of the average of σ, x is βj.

j is just a number.

We could have taken it to be equal to 1.

So $< \sigma >$ is hyperbolic tangent of something proportional to the inverse temperature.

The slope of *tanh(x)* at $x = 0$ is 1.

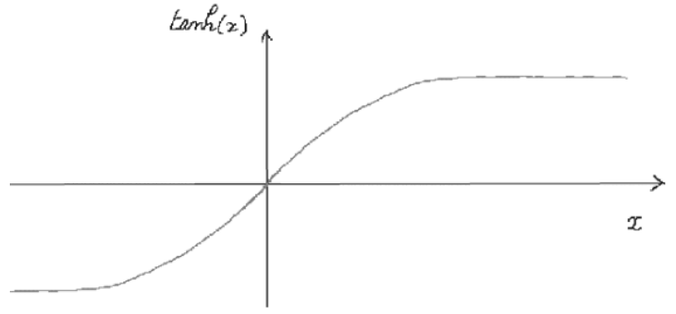
Then it has the two asymptotic values -1 and $+1$, when x respectively goes to $-\infty$ and $+\infty$.

But, if we took j to be positive, which is entirely conventional and up to us, we are only concerned with the positive branch of its graph.

Now let's look at what that says physically.

When β is very large, that means very low temperature, and since j is positive, then βj is very large and positive.

We are way out on the right in figure 3.



What it is telling us is that at very low temperature the spin will point up with probability very close to 1.

Indeed, if the two possible values of σ are -1 and $+1$ and its average is close to $+1$, this implies that the probability that it take the value $+1$ must be almost 100%.

And we built the model, with equation (2), in such a way that this corresponds to the low energy state of the spin.

Let's turn to very high temperatures.

Then βj is very small.

We are near the origin.

So $< \sigma >= tanh(\beta j)$ is near zero too, on the positive side.

The magnetic spin is being constantly kicked around with equal probabilility between its values $+1$ and -1, and its average value is near 0.

At very high temperature it overcomes any bias it could have.

That was a very simple system.

5.2 One-dimensional Ising model

Now we want to come to the next least complicated big magnet.

It is the one-dimensional Ising model, which we began to talk about at the end of the last part.

Again, at each point along a one-dimensional array, there is a little magnet with a σ, figure 4.



Figure 4 : One-dimensional Ising model.

It is either up or down, corresponding to $\sigma = +1$ or -1.

It looks like the preceding model, and indeed figure 4 is the same figure as figure 1, but the new thing is that the energy now depends on the relationship between neighboring spins.

And there is no more external magnetic field.

We could do it with an external magnetic field too.

And we will discuss that at some point.

But for the moment, no external magnetic field.

Equation (8) was the average energy of one spin.

But now the energy of the system comes from the coupling of neighboring spins.

So we can no longer simply look at the energy of each individual spins and add them up.

Hence we look directly at the energy of the whole shebang.

It is given by

$$E = \sum_i -j \ \sigma_i \ \sigma_{i+1} \tag{10}$$

As before, j is only a positive number.

And the minus sign means that the model favors neighboring spins pointing in the same direction.

In that case, we say that they are parallel, or aligned.

Note that at macroscopic scale this is not what ordinary ferromagnets do.

But in ordinary ferromagnets, at microscopic scale, the spins do line up when the temperature and the energy go down.

It is a matter of very intricate detail whether the little elementary magnets in the material prefer to be aligned or antialigned.

Most of the times they prefer to be antialigned, but in iron they prefer to be aligned.

We are not going to go into these details, which are not relevant to this class.

Remember that the "big magnets" made of "elementary magnets" we study here are only mathematical models, remotely related to real magnets.

We use them as simple examples displaying or not displaying phase transitions.

Thus the energy of one pair is lower when the two spins are parallel, either both up or both down.

And it is higher when they are unaligned, i.e. when one is pointing up and the other down.

If all the spins are parallel, all the products $\sigma_i \, \sigma_{i+1}$ are equal to +1.

Then because of the $-j$ in front, the total energy of the system is minimum.

So what is the ground state of the system like ?

It is when all the spins point in the same direction.

But which direction? Answer :

It doesn’t matter.

There are two ground states, with the same lowest energy.

Now what would happen if we changed the sign in front of j, that is if we consider another model where the total energy was $\sum j \sigma_i \sigma_{i+1}$?

Is it really a different system? No.

It is really the same system.

All you have to do to see it is the same system is to redefine every other spin by changing its sign.

On every other spin what you called up you now call down.

You redefine the variable.

And then what you find out is the ground state wants to have them all antiparallel.

And again there are two ground states.

So it is mathematically identical.

In other words, the minus sign in front of j is really only a convention, justified because it is simpler to think of grounds states with all the spins aligned, up or down, rather than to think of the two states where they perfectly alternate.

What we want to calculate is the usual thing : our friend the partition function.

What is it now ? It is

$$Z = \sum e^{\beta j \sum \sigma_i \sigma_{i+1}} \tag{11}$$

The small sign Σ in the exponent is the sum over i for one configuration.

And the big sign Σ, is the sum over all possible configurations.

Equation (11) looks terrible and very hard.

But it is the partition function we must work with.

Let us first present a question that we can ask.

We would like to know the answer to the following.



Figure 5 : We know that the spin at position i is up.

Suppose we know that the spin in one of the positions along the chain is up, figure 5.

Then, knowing that, we can ask : What is the probability that n links down the chain it is also up ?

It is a **conditional** probability, given that the spin in figure 5 is up, that at some other specified place it is also up ?

It can be expressed in terms of correlation.

The correlation between the random variable X_i and the random variable X_j is by definition

$$\mathrm{corr}(X_i,\ X_j) = \frac{< X_i X_j > - < X_i >< X_j >}{\mathrm{sd}(X_i)\ \mathrm{sd}(X_j)}$$

where < ... > stands for mean, and *sd(...)* for standard deviation.

We could also ask it in a different way, but it is the same question really : What is the average of the product of spins at two different locations ?

Note that, if for all i σ_i is equally probably +1 or −1, then all the means are zero, and all the standard deviations are one.

In that case the above formula reduces to

$$\mathrm{corr}(\sigma_i,\ \sigma_j)\ = < \sigma_i \sigma_j >$$

paying attention that here σ_i and σ_j are not the usual notation for standard deviations, but are random variables.

Now you might think – and in this case you'd be right – that a spin being up at some position *i* would have very little effect on a spin at some position *j* far down the chain.

You might expect then that the average of the product would be zero.

Because whatever σ_i is, σ_j would have equal probability to be up or down, independently of the value of *i*.

So their correlation would be zero.

But it might be wrong too.

There might be an effect of just having one spin up at i that would propagate all the way through the system and tell you that there is a net bias throughout the whole sample.

You could diagnose that by looking at the average value of the spin at i times the value of the spin at *j*, where *j* is *n* positions down from position *i*, that is $j = i + n$. In other words, you would be looking at

$$< \sigma_i \ \sigma_{i+n} > \tag{12}$$

If there existed this kind of memory where, if the spin at location *i* was up, it biased the system all the way to infinity, then expression (12) would not go to 0 with large distances.

If, on the other hand, that bias does go to 0, then σ_i being up would not bias σ_{i+n}.

And on the average the product with a distant spin would be 0.

So this is an interesting diagnostic test of the effectiveness of one spin being up on its neighbors, and how far that propagates down the chain.

Let's see if we can guess what the answer is in our one- dimensional model.

Instead of a chain of little magnets, think of it as a game of telephone, the game where you whisper to your neighbor.

So we have a long chain of people and a message is going to start at one end somewhere.

It could be anywhere, for instance at location i.

And the message is going to be a very simple message. The message is either "0", or "1".

We are not going to make fancy messages like the kids who play this game – the kind of message like : "my dog had a heart attack and therefore I'm off dog food for the rest of my life".

Just "0" or "1".

And people hear pretty well, as well as talk pretty well, among this group.

But the fidelity is not absolutely perfect.

It is pretty good, yet there is a small probability that the information is changed.

The question is : How far down the chain does the signal propagate before it gets lost?

In other words, how far do we have to go before it just becomes equally likely that the person agrees with the starting point or disagrees.

The answer is no matter how good the fidelity, as long as it is not absolutely perfect, you lose memory sufficiently far down.

If there is a probability for an error of 1% in each time, then the probability for no error is 99%.

But, when we go two steps down the chain, it is 99% of 99%.

And that is 98.01%.

Each time it goes down by a factor of 0.99.

And if you go far enough down the chain that product will get arbitrarily small.

The probability a hundred units down the chain that the signal is remembered with good fidelity is only 36.6%.

And down 10 000 units it is 2.3 x 10^{-42} %, that is essentially zero.

That is basically no memory.

It is the same thing in our model.

There is no signal to be transmitted, but there is an energy bias.

If the spin σ_i is up, the system preferring lower energy, the spin σ_{i+1} will have a bias to be up too.

At zero temperature, it is like the perfectly accurate signal, infinite fidelity.

That is like the situation where there is no loss of information whatsoever.

If one spin is up, then everybody just lines up.

It is perfect positive correlation.

But if there is the slightest systematic bias, or infidelity, the correlation far down will go to zero.

That is what will happen in the Ising model.

So let's see if we can calculate E in equation (10).

There is a marvelous trick.

Instead of focussing on the spins, it is to focus on the links between the spins.

We are going to imagine for the moment that the chain is finite.

Then we will let it get very big.

The first spin at the left extremity can be up or down.

Let's start by assuming it is up.

And later on we will add back the configuration where it is down.

Thus we are going to write the partition function as the sum of two terms.

In the first term, the first spin is up.

In the second term, the first spin is down.

But let's concentrate on the term where the first spin σ_1 is up to begin with, figure 6.

Now when considering the second spin, we can either work with the value of the spin itself or we can work with the product $\sigma_1\sigma_2$.

Since σ_1 is known to be up, $\sigma_1\sigma_2$ will tell us everything we want to know about σ_2.

For any i, let's define

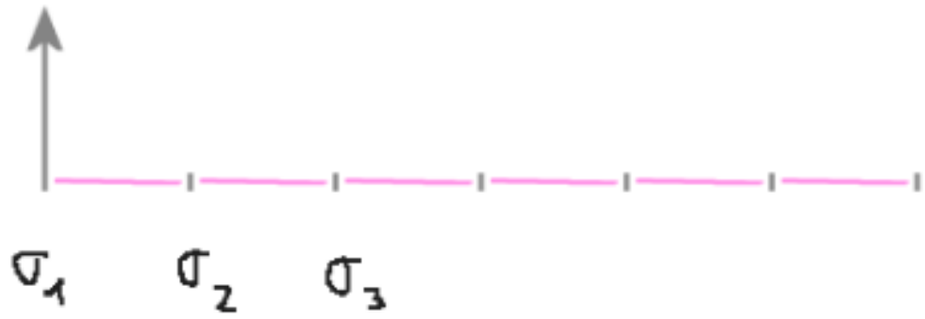


Figure 6 : Collection of spins, and their links. We know that the first spin is up.

$$\mu_i = \sigma_i \ \sigma_{i+1} \tag{13}$$

Thus μ_1 is a variable that has to do with the relationship between the first spin and the second spin.

But if we know it, then you know what the second spin is.

What about the next one ? $\mu_2 = \sigma_2\,\sigma_3$.

Instead of thinking of the spins, think of the links.

The links have two possibilities : either parallel or antiparallel.

Knowing the links would tell us nothing about any individual spin, unless we knew the value of the first spin.

But if we know σ_1 is up and we know μ_1, then we know σ_2.

Indeed, if $\sigma_1 = +1$, then, from $\mu_1 = \sigma_1\sigma_2$, we deduce that

$$\sigma_2 = \mu_1 \tag{14}$$

Now supposing we know μ_1 and also μ_2, then we know what σ_3 is.

What is it ?

To calculate it, we can start from

$$\begin{aligned} \mu_2 &= \sigma_2\sigma_3 \\ &= \mu_1\sigma_3 \end{aligned} \tag{15}$$

Multiply both sides by μ_1

$$\mu_1\mu_2 = \mu_1^2\sigma_3 \tag{16}$$

But whatever μ_1 is, its square is 1. So we get

$$\sigma_3 = \mu_1\mu_2 \tag{17}$$

The general formula stated below.

Given that $\sigma_1 = +1$, one can show that for all n

$$\sigma_n = \mu_1 \mu_2 \ \cdots \ \mu_{n-1} \tag{18}$$

In other words, knowing the first spin, if we know the μ's for all of the links in between, we know all the spins.

And there is no redundancy, there is no double counting.

As long as we know that the first spin is up, it is equally good to know the μ's, which live so to speak on the bonds between the spins, as it is to know the spins themselves.

That is very useful. Why ?

Because the energy is just made up out of these bond variables.

From equation (10), we can write

$$E \ = \ -j \sum_i \sigma_i \, \sigma_{i+1} \ = \ -j \sum_i \mu_i \tag{19}$$

We are not even multiplying anything anymore.

With $-j$ in front, the energy E of a particular configuration is just the sum of the values of the individual bonds.

That is a powerful new expression, because the individual bonds are all independent, and that will come in handy when we calculate the partition function of the system.

What are the possible values of the bonds ?

It is +1 or −1.

It is +1 if the two spins they link are aligned, and −1 if the spins are antialigned.

Going back to the partition function Z, from equation (11) we can write

$$Z = 2 \sum e^{\beta j \sum_i \mu_i} \tag{20}$$

The factor 2 in front is because our above reasoning was only in the case σ_1 was up.

But, as said, we must add the case when σ_1 is down.

The values of the two terms are the same, whence the factor 2.

But, as we know, a multiplicative factor is no big deal in a partition function.

As before, in equation (20), the sum in the exponent is over the bonds, and the big sum is over all the possible configurations.

Notice there is one fewer bond than spins, although, if the number of elementary magnets is very large, it doesn't matter much.

Equation (20) resembles equation (3) we obtained in the very simple model studied at the outset, before the Ising model.

The differences are that the σ_i which appeared in equation (3) is replaced by $\sum_i \mu_i$ in equation (20).

This $\sum_i \mu_i$ is for one configuration.

Then there is the big sum of $\exp(\beta j \sum_i \mu_i)$ over all the possible configurations, the calculation of which appears daunting.

And there is the factor 2.

True, equation (20) looks much messier than equation (3) or even equation (3) to the power N, that is equation (5), but the important point we are going to use is that the μ's are independent of each other.

Remember that we established equation (5) following two methods : $E = \mu H \sum_{i=1}^{N} \sigma(i)$

First, in the previous part of our discussion, we looked at all the configurations of the N spins ;

Second, we looked only at one spin, and wrote that the partition function for the whole collection was that of one spin raised to the N-th power.

We could do that because the spins in that simple model were all independent.

Now in the Ising model, where the coupled spins were astutely replaced by the independent bonds, we use the second method.

That allows us to avoid working with the messy equation (20), and reach directly the simple answer.

The partition function for the collection of bonds, is that of one bond raised to the $(N-1)$-st power, because there are $N-1$ bonds.

And we have a factor 2 in front.

So at one fell swoop we get

$$Z \;=\; 2\;\left[\;2\cosh(\beta j)\;\right]^{N-1} \tag{21}$$

We have reduced our problem to exactly the same problem as before.

The only difference is the physical meaning of the degrees of freedom in the collection of spins or bonds that is a little bit different.



Now what is the average ?

Let's not ask about the average of one spin, but about the average of one bond, that is the average of the product of two neighboring spins.

We can either write it as $< \mu_i >$, for any given μ_i, or as $< \sigma_i \, \sigma_{i+1} >$, for any given pair of adjacent spins.

In any case, it is

$$< \mu_i > = < \sigma_i \, \sigma_{i+1} > = \tanh(\beta j) \tag{22}$$

It is not zero.

And if we take *j* to be positive, it is positive.

The fact that the average is non 0 and positive tells us that there is a net – in the sense of average – tendency for the *(i + 1)*-st spin to line up in the same direction as the *i*-th spin.

That is the same thing as saying that there is a bias for μ_i to be positive.

If the first spin is found up, the next one has a better than even chance of also being found up.

And what is the better than even chance ?

Using the fact that

$$< \sigma_1 \, \sigma_2 > = \tanh(\beta j)$$

we can prove that

$$\text{Prob} \{ \, \sigma_2 = +1 \mid \sigma_1 = +1 \, \} = \frac{1}{2} + \frac{\tanh(\beta j)}{2}$$

The left hand side is the usual notation for the conditional probability that $\sigma_2 = +1$, given that $\sigma_1 = +1$.

This tells us that there is a correlation between neighbors.

Remember, from elementary probability theory, that two random variables are independent if and only if the unconditional distribution of one of them – also called it *marginal distribution* – is the same as its *conditional distribution* given that we know the value of the other random variable.

You might have come across the nagging little fact that uncorrelation and independence are not exactly equivalent concepts.

They are equivalent within the family of Gaussian random variable's.

But they are not quite equivalent in general.

It is possible to build two random variables which are dependent but whose correlation coefficient is 0.

Counterexample : take $X = \text{Uniform}(-1, +1)$, and $Y = X$.

Then you can check that $\text{Corr}(X, Y) = 0$, but they are obviously not independent.

However such counterexamples don't concern us here.

But let's now go far down the line.

And let's ask exactly the question which we asked when we started :

What is the correlation between the *i*-th spin and one *n* units down the chain ?

How are we going to get at that ?

The answer is very simple.

We want to calculate

$$< \sigma_i \ \sigma_{i+n} > \tag{23}$$

We can write this as follows

$$< \sigma_i \ \sigma_{i+1} \ \sigma_{i+1} \ \sigma_{i+2} \ \sigma_{i+2} \ \cdots \ \sigma_{i+n-1} \ \sigma_{i+n-1} \ \sigma_{i+n} > \tag{24}$$

Why is that?

Because inside we multiplied by σ_{i+1}^2, σ_{i+2}^2, etc.

But these squared random variables are always equal to 1, so we didn't change anything.

The nice thing is that expression (24) can be rewritten $< \mu_i \ \mu_{i+1} \ \mu_{i+2} \ \cdots \ \mu_{i+n-1} >$ (25)

Also, notice that we could also have used expression (23) and equation (18) to establish expression (25).

Now again we shall use the fact all of the μ's are independent of each other.

The energy is the sum of the energies.

The problem completely factorizes.

$$< \mu_i > < \mu_{i+1} > < \mu_{i+2} > \ \ldots \ < \mu_{i+n-1} > \tag{26}$$

And each of these averages is *tanh(βj)*.

So we have finally found that

$$< \sigma_i \ \sigma_{i+n} > = \ \tanh(\beta j)^n \tag{27}$$

n being the number of steps between *i* and *(i + n)*.

This formula is expressing exactly the idea of the game of telephone.

The coefficient *tanh(βj)*, with is akin to fidelity, is always positive and less than or equal to 1.

When *tanh(βj)* = 1, β is infinite, that is zero temperature, perfect fidelity.

Then, no matter how far down the line you go, $< \sigma_i \sigma_{i+n} >$ in equation (27) will be one.

This is the statement that at zero temperature if you know that μ_i is lined up, every one of them is lined up.

There are only two possibilities, everybody parallel upward or everybody parallel downward.

And in either case the product (26) is one.

But if there is any loss of fidelity at all, which means in this case the temperature is not absolutely 0, then $tanh(\beta j) < 1$.

And each time we go another step we multiply by a factor of $tanh(\beta j)$.

It like what we saw when fidelity in the telephone game was not perfect but only say 99% at each step.

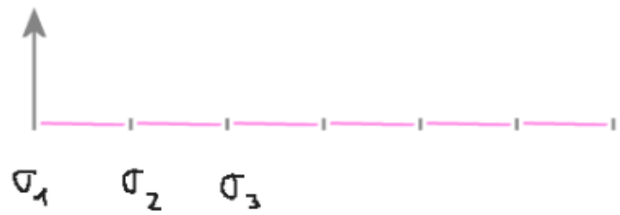
After two steps it was only about 98%, etc.

So, no matter how close $< \mu >$ is to 1, as long as it is less than 1, when we go sufficiently far down the chain the correlation $< \sigma_i \sigma_{i+n} >$, as a function of n, will become arbitrarily small.

There would be no or negligible memory. Indeed $tanh(\beta j)^n$ is an exponential function of n.

And when the temperature is above zero, it is a number less than 1 raised to the n-th power.

So the correlation falls exponentially fast with n.



Incidentally, the big trick of focussing on the bonds instead of the spins in figure 6, thereby replacing messy coupled random variables – the spins – by nice independent ones – the bonds –, and transforming the problem of finding the partition function, the energy, etc. into a tractable one, is a pattern in physics.

It is called a *duality*.

In modern physics this could be thought of as the first duality, that is equivalence of different systems.

We found an equivalence between a theory of spins which are connected to nearest neighbors, with another theory of just spins which are uncoupled, because we saw that the bonds in figure 6 behave like the spins of the very first problem where they were independent of each other, just plunged in a magnetic field.

This involved a clever change of variables which basically interchanged sites to bonds.

Here is, for instance, a picture of sites and bonds in a 3D lattice, figure 7.

The sites and bonds are also sometimes called vertices and edges.

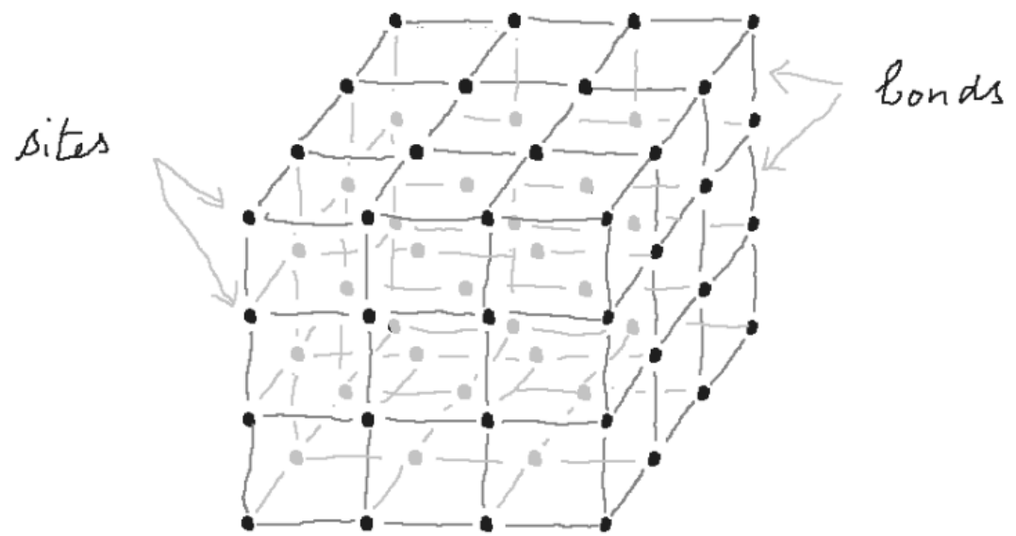


Figure 7 : 3D cubic lattice, with sites and bonds.

So we met the first example of a duality between different statistical mechanical systems.

It is not always an interchange between edges and vertices, but in this case it is.

We found out some things about the *one-dimensional Ising model*.

We found out that it is never magnetized, except at zero temperature.

Magnetized can translate into the statement that there is this long-range memory :

if one spin is up all the others will be biased up till very far away.

That is called **magnetization**.

It is equivalent, in fact, to the statement that if we put a tiny magnetic field on one site of the system to bias it, it will cause everything to line up.

We will see that with a little calculation in a moment.

In summary the one-dimensional Ising model is a boring system.

It has no phase transition.

It does exactly what we might have expected it to do.

Correlations fade as we go down the line.

And the reason is simple.

It is the same reason as in the game of telephone.

There is a probability that we make an error when we go from location i to location i + 1.

Once we make an error, we start over again with a new message.

And we wait until we make another error.

Then we start over again, and so forth and so on.

And that is the way it goes.

Nevertheless there will be some clumping, a tendency to display clumps of spins in the same directions.

It is just like in the game of telephone where there will be long stretches of agreement.

And then statistically, on the average, every so often, there is a switch.

But then a long chain of agreement again.

Then a switch, etc.

Of course high temperatures are like the fidelity being very bad.

The clumps or stretches are short then.

Infinite temperature is like the situation where you just can't hear your neighbor.

You get no information.

He or she whispers in a din, and on top of that there is a dense fog, so that you can't even read his or her lips.

You wind up, whatever your neighbor says, making a random guess.

That is the statement that there is not even correlation between nearest neighbors.

$T = +\infty$, $\beta = 0$, $tanh(\beta j) = 0$, the probability calculated earlier is 1/2, and $\mathrm{Corr}(\sigma_i, \sigma_{i+1}) = 0$.

How do we make the system more interesting?

What is wrong in a linear chain is the dimension is too low.

Once we make a mistake our neighbor doesn't have any support to know what the right answer is.

What about higher dimensions ?

5.3 Ising model in two or more dimensions

Now, instead of playing telephone, let's play a different game.

Players are sitting in a room, for example a lecture hall, with rows after rows of seats covering the entire floor.

So there are also columns of seats.

Each player has four neighbors.

They can be seen as the vertices of a two- dimensional square lattice.

Kevin over here starts a message.

He sends it to the four fellows respectively to his right, his left, in front of him, and behind him.

Then they each send the message to all their neighbors.

In particular they will send their message to some nev people who will receive them from several different directions, for example Seth in figure 8.

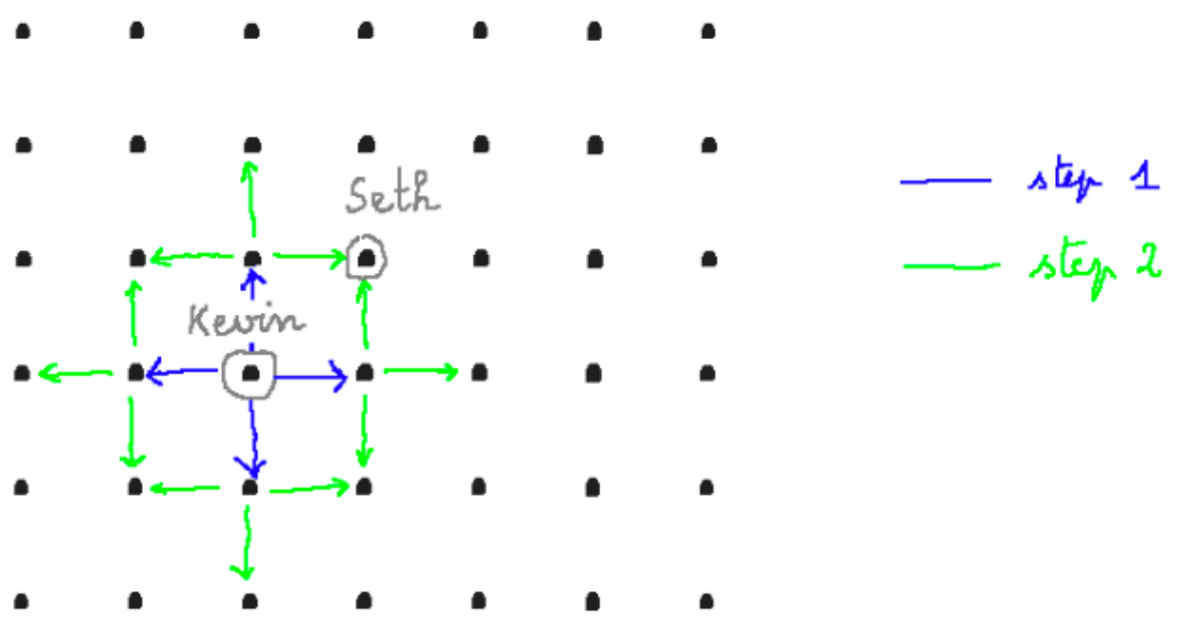


Figure 8 : Message transmission in a 2D square lattice. Players receive messages from several different directions.

As the message spreads throughout the lattice, new people will be getting messages from different sides.

They will have to make their judgments about what is the right message, and therefore in turn what to transmit.

If after having sent his message a player becomes somehow new again and can again receive messages and transmit them, then as the game unfolds everybody will get basically four messages from all the people around them.

We are a player, and suppose three of those messages arriving at us say one thing, but one of them says another thing, what do we do ?

We will do what computer scientists call error correction.

It just means we will take the majority vote.

This system works much better.

And in fact, if the fidelity is reasonably good, the initial message will spread throughout and off to infinity.

The same thing happens in the two-dimensional Ising model.

The bias of anyone spin will bias the rest of the sample.

Said another way, putting a little magnetic field on one spin in two-dimensions will bias the whole sample.

This is not completely obvious.

Surprisingly, there is no easy solution.

One way to prove it is going through the analysis of the two-dimensional Ising model following what Kramers and Wannier did.

That Ising got it wrong in two-dimensions is forgivable.

But he thought and stated that in every dimension, if the fidelity was not 100%, then the message would always eventually get lost, as in one dimension.

That is also wrong.

And in high dimensions, it is not too hard to disprove.

If we think about it for a minute, we realize that it is very dimension dependent.

Why is that ?

In one dimension each person has only two neighbors

That is not very many.

The player at site *i* doesn't have much of a support system.

He gets the message really only from one direction.

How many neighbors does a two-dimensional system have ?

Four if it is a square lattice like in figure 8.

So that means we will be getting messages from four people.

We have a pretty good chance, using the majority rule, of registering the right answer even if the fidelity is not so good.

If we use some weighing procedure such as taking the majority, we do pretty well.

We do much much better than if we were just getting the message from one person.

How about in three dimensions?

We have six neighbors, see figure 7.

How about 10 000 dimensions.

Whatever it is, 20 000 neighbors if the lattice is square.

I used to have this friend.

Anything you asked him, he would say : A hundred men can't be wrong.

I don't know what he had in mind, but in high dimensions it is true a hundred neighbors won't be wrong very often.

Fluctuations among large numbers of variables tend to be very small in comparison with the net average.

By that, we mean precisely the following.

Consider n Bernoulli trials, that is, n random variables X_i, independent of each other and each one taking value +1 with probability p, and −1 with probability $(1 - p)$.

Then the average of each of them is $2p-1$, and so is the average of $\sum_i X_i/n$.

But *the fluctuation of the average of the* X_i's *is much smaller than that of each* X_i.

By fluctuation we mean the standard deviation.

The variance of X_i is *4p(1 − p)*.

So the standard deviation of X_i is $2\sqrt{p(1-p)}$.

Then, because the X_i are independent with the same distribution, the variance of the sum of the X_i's is the sum of the variances, that is

$$\text{Var}\left(\sum_i X_i\right) = 4np(1-p) \tag{28}$$

And if we divide the sum by n, the variance will be divided by n^2.

$$\text{Var}\left(\frac{1}{n}\sum_i X_i\right) = \frac{4p(1-p)}{n} \tag{29}$$

Finally, taking the square root, we reach

$$\text{Std dev}\left(\frac{1}{n}\sum_i X_i\right) = \frac{1}{\sqrt{n}}\ \text{Std dev}(X_i) \tag{30}$$

If $n = 100$, we see that the fluctuation of the average of the X_i is ten times smaller than the fluctuation of each X_i.

This is in fact also a direct consequence of the law of large numbers, which is proved along similar lines, using Chebyshev inequality.

$$\mathrm{P}(|X - \mathrm{E}(X)| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2},$$

If we have a positive random variable Y with mean μ, we prove first that, for any $t > 0$, $Pr\{Y \geq t\} \leq \mu/t$.

Then, if X is a random variable, which can be positive or negative, with mean μ and variance σ^2, Chebyshev inequality says that, for any $t > 0$,

$$\Pr\{\ |X - \mu| \geq t\ \} \leq \frac{\sigma^2}{t^2}$$

It is the familiar and intuitive result that the fluctuation of the average of n measures is smaller than that of each measure, and goes to zero like $1/\sqrt{n}$.

If p is higher than 1/2, then $EX_i = 2p - 1$ is positive, and standard deviation of $X_i = 2\sqrt{p(1-p)}$.

The expectation of the average of the X_i is also $2p-1$.

But, as we saw, the standard deviation of $\sum_i X_i/n = 2\sqrt{p(1-p)/n}$

So it is highly likely that the random average will be positive and therefore the message we will register in a transmission, which we select to be either $+1$ or -1, will be $+1$.

So, the higher the dimensionality, the better our shot at being able to propagate the information of one spin throughout the entire lattice, because then n will be high in the above calculations.

That leads us to ask the question :

What would the Ising model be like in very high dimensions?

We consider some very large number of dimensions, ten thousand or whatever it is.

Then what can we say about the propagation of the information ?

We will observe a phenomenon.

You might even ask :

Above what value is the number n of dimensions considered to be large ?

It turns out 2 is high enough.

For the study of the propagation of the information, using the majority rule, which is reinforced by the number of dimensions, it turns out that 2 behaves like 10 000.

As soon as the number of dimensions is more than one, there is a *phase transition*.

Only the one-dimensional Ising model does not have a transition.

When $n \geq 2$, if the probability p, i.e. the fidelity, is high enough, the information is transmitted throughout the lattice.

That is not obvious, and could not have been predicted easily.

It also turns out that 3 is already very large in this game.

That can be calculated, or checked numerically.

But anyway $n = 2$ is enough.

As p goes down from 1 to 1/2, at some point there is a phase transition from order to disorder.

We won't prove it for $n = 3$, let alone for $n = 2$(will show it in a computer model), because it is hard.

But for sufficiently high dimensions there is an awfully good argument to say that there is a transition.

At high temperatures everything is random.

As we lower the temperature, everybody wants to line up.

And that tendency to lineup propagates throughout the system to infinity.

There are many possible tricks to do this thing.

We are going to use something called mean field approximation.

Instead of n, the number of dimensions will be denoted d.

Each site is surrounded by how many neighbors ?

In 3D the answer is 6, figure 9.

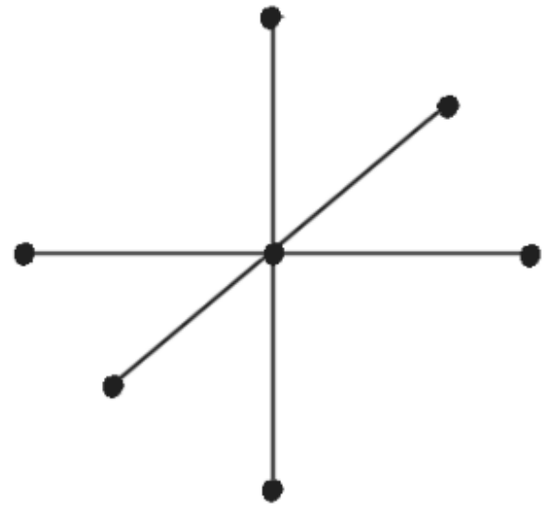


Figure 9 : Neighbors of a site when d = 3.

Each site is surrounded by 2d neighbors. Imagine that d is very large.

And because d is very large, imagine that all of the neighbors define a *field* whose fluctuation is much smaller than its average value.

As we spelled out in the case of the Bernoulli trials, if we have a large number of variables, and they are biased, so that their biasses(the bias of a random variable here means the departure of its expectation from 0, we use the word bias to emphasize that, unlike when it has mean 0, the variable carries information) add up to something of order the number of variables, then the fluctuation of the sum will typically be much smaller than its bias, because the fluctuation of the sum will typically be much smaller than its bias, because the fluctuation will grow like $\sqrt{n}$ compared to the bias which will grow like n.

So that is what we are going to do.

Let's focus on the spin in the center of figure 9. For all the other spins surrounding it, we are going to make an approximation about them.

Let's write the energy of the spin in the center. It is

$$E = -j\sigma \sum_{neighbors} \sigma_{neighbor} \tag{31}$$

where σ is the value of the spin at the center.

This formula is the generalization of $-j\sigma_i\sigma_{i+1}$ which appeared in equation (10).

But be careful, the sum in equation (31) still concerns only one spin, whereas the sum in equation (10) came from the fact that we were looking at the whole sample of spins.

Now let's suppose that there is a bit of bias and that the average of the spins is not 0.

Of course we are going to check that.

We are going to write a formula, that would allow the spin to be 0 too.

But let's suppose throughout the lattice that the spin has an average denoted $\overline{\sigma}$.

$\overline{\sigma}$ is a notation, read *sigma bar*, meaning the average spin in the whole sample.

It is lighter than $< \sigma >$, but is the same thing.

We don't know what the value of $\overline{\sigma}$ is.

It might be 0, it might be positive, it might be negative.

We don't know.

But, in equation (31), we are going to replace the sum over neighbors by simply $2d$ times this average.

That is a pretty good approximation if the number of neighbors is large.

And the larger the number of neighbors, the better the approximation.

When we have large numbers, as a typical rule we are allowed to average and the fluctuations are small.

So equation (31 becomes

$$E = -2dj\sigma\overline{\sigma} \tag{32}$$

That is the energy of one particular spin sitting in the bath, or in the field, of all the others.

This is called **mean field approximation**.

Mean is not in the sense of nasty, but in the sense of average.

And the field now is the field experienced by one spin in the field of all the others.

We can do the partition function of that one spin at the center of figure 9.

In fact we don't even need to do the calculation.

We know how to do the partition function.

It is exactly the same calculation we did starting from equation (2), except wherever we wrote j, we instead write $2dj\overline{\sigma}$.

So we can immediately write down the average value of the spin at the center of figure 9.

Let's denote it $\overline{\overline{\sigma}}$, read *sigma double bar*.

This average spin is going to be

$$\overline{\overline{\sigma}} = \tanh(2d\beta j\overline{\sigma}) \tag{33}$$

The spin at the center of figure 9 is moving in the background of all the others.

And the others all constitute a constant field that we just call $\overline{\sigma}$.

This mean field approximation is also sometimes called *self-consistent field approximation*.

Why self-consistent – hopefully self-consistent?

Because the spin at the center of figure 9 is no different than all the others.

It is just one of the many spins in the system.

And if each of the neighbors in figure 9 has an average equal to $\overline{\sigma}$, the physical intuition would require that the one at the center also should have the same average.

That has got to do with translation invariance.

Everyone is really the same as everyone else in a big sample.

So the self-consistent field theory implies that $\overline{\overline{\sigma}}$ is the same as $\overline{\sigma}$.

That gives us an equation for $\overline{\sigma}$:

$$\overline{\sigma} = \tanh(2d\beta j\overline{\sigma}) \tag{34}$$

$2d\beta j$ is simply a constant factor, inside *tanh*, in front of $\overline{\sigma}$.

And remember that β is the inverse of the temperature.

Equation (34) is an implicit equation for $\overline{\sigma}$.

We want to solve it.

How do we get an intuition about what it says ?

Question: Regarding the factor *2d*, is it only the number of nearest neighbors, or could it include next nearest neighbors ?

Answer: We don't look at next nearest neighbors.

2d is the number of nearest neighbors in *d* dimensions.

In one dimension it is two.

It two dimensions it is four, see figure 8.

In three dimensions it is six, see figures 7 or 9.

In four dimensions it would be eight, etc.

Next nearest neighbors would go in the same direction as going in higher dimension.

That would just change the factor inside *tanh* in equation (34).

We pick a particular model and study it as a function of dimension.

We could study it as a function of the nature of second nearest neighbor couplings, third nearest neighbor couplings.

We can do all these things.

The physics of high dimensions is not something we realize in the laboratory.

But we do this procedure all time : We want to prove something about a system.

It is a little too hard. We can't get our hands on it.

But we can go, mixing mathematical and physical reasoning, to some limit where it becomes much easier.

And we prove it in the limit.

Notice though that the experiments which we can do in the laboratory – as opposed to the gedanken experiments Einstein was fond of – are always studied, analyzed, interpreted with models of a mathematical nature.

Sometimes these models involve high or even infinite dimensions.

Nature provides only raw facts, not elaborate models or interpretations.

And even those raw facts are not devoid of puzzles.

Is it really possible to perceive without any implicit models?

Can there be raw facts or phenomena we cannot perceive?

It is tempting to answer no.

But then we are part of the experiments in some uncomfortable way.

These considerations may eventually lead to the next scientific paradigm which would look differently at Man looking at Nature, and would solve the riddles of the hypothetical multiverses and such things.

See the discussion at end of class on the anthropic principle for more.

We haven't proved it for the physical case of interest but at least we can prove that somewhere between, in our case, one dimension and a hundred thousand dimensions, a certain new kind of behavior happens that wasn't present in one dimension – namely the phase transition that we are going to show.

Once we see how that works, we can then ask where does the change happen?

And in our case it happens between dimension one and dimension two.

So when *d* is large, we get a nice simple example where we can view what happens physically without having to solve a difficult mathematical problem.

So let's try to solve equation (34). $$\overline{\sigma} = \tanh(2d\beta j\overline{\sigma})$$

The first thing to do is to change variables.

To change variables often seems to obscure a problem, because we go from a variable we have an intuition for to one which seems artificial and we can't interpret as easily.

But there is a good reason.

Equation (34) involves hyperbolic tangent of a complicated thing.

We don't want a complicated thing inside *tanh*.

That is a nuisance.

So we are going to make 2dβj $\overline{\sigma}$ our new variable :

$$y = 2d\beta j\overline{\sigma} \tag{35}$$

Equation (34) becomes

$$\frac{y}{2d\beta j} = \tanh y \tag{36}$$

Now to solve this equation, we are just going to graph both sides and see where they intersect.

That is the crudest way to solve an equation : to draw a graph of both sides and see where they intersect.

On the left-hand side, instead of β in the denominator, let's put the temperature *T* in the numerator because it is a more intuitive quantity to follow.

Equation (36) becomes

$$\frac{Ty}{2dj} = \tanh y \tag{37}$$

So we shall draw on the same graph *tanh y* and the simple linear function *Ty/2dj* for various values of the coefficient *T/2dj*.

We begin with the first case when the temperature is high.

Then the straight line *Ty/2dj* intersects *tanh y* only at the origin, figure 10.

Of course, since d is a large number, T must be large too.

How high the temperature must be depends on d.

But eventually for a high enough temperature the slope of the straight line is more than 1, i.e. more than the slope of the hyperbolic tangent of y at 0.

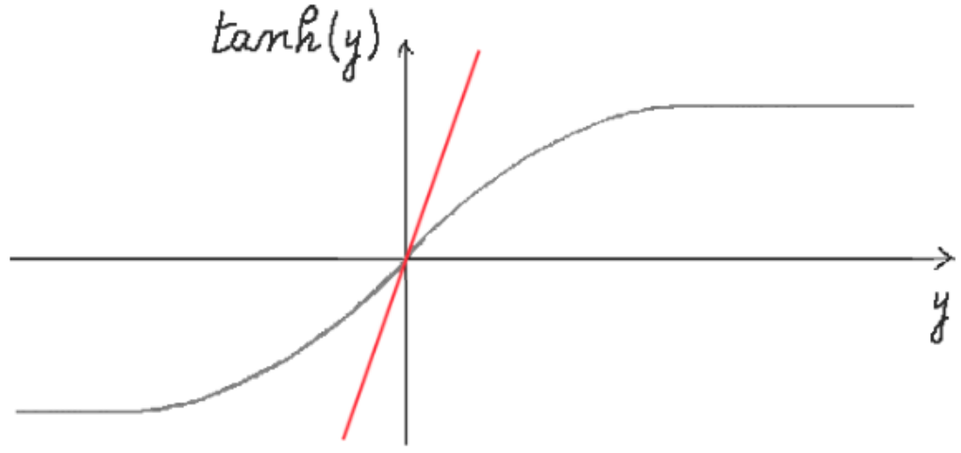


Figure 10 : tanh y and T y/2dj when T is high.

So we see that for high temperature, the only possible answer to the self-consistent field theory is that $y = 0$.

But what is y ?

y is proportional to the average of σ.

Thus we learned the first result that at high enough temperature, the average of σ must be zero, as expected.

That is not too surprising.

Now let's start lowering the temperature.

As T goes down, the slope of the straight line decreases.

We eventually reach a temperature where the slope is equal to 1.

We see that when

$$T = 2dj \tag{38}$$

the straight line becomes tangent to the hyperbolic tangent.

Something happens at that temperature.

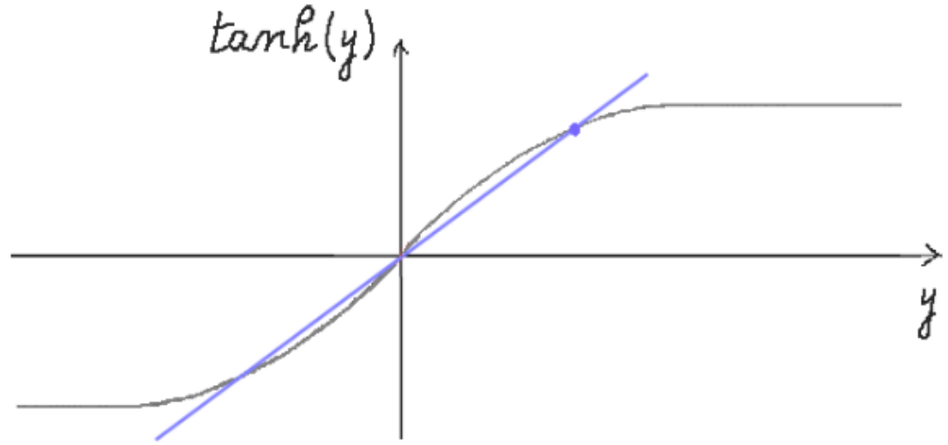


Figure 11 : tanhy and Ty/2dj when T < 2dj.

When we go beyond that, there is some new solution to equation (37), or equivalently to equation (34), figure 11.

Just to be clear let's assume j is positive, so we are only interested in the right part of the graph.

So now, beside the solution $y = 0$, there is another solution with y strictly positive, therefore where $\overline{\sigma}$ is not zero.

In other words the system can have an overall average of spin pointing all in the same direction.

That is the phenomenon of magnetization.

What happens when T crosses $2dj$ is called a *phase transition*.

In terms of T , below that point there is an average magnetization, an average field that just permeates the whole system forever.

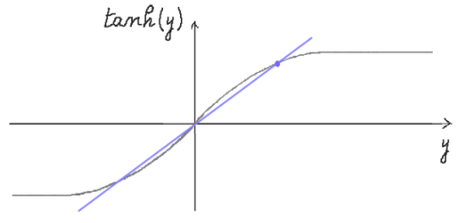
Now you might say :

Hold on, how do I know that the right physical solution wasn't the one at the origin ?

Indeed for any value of the constants, the straight line does also intersect the *tanh* curve at the origin.

So how do we know which one is right ?

We are going to do a little calculation.



We shall add to this problem a teeny little magnetic field that biases the system.

And we shall discover that the teeniest little magnetic field will tell us that we should be on the right branch in figure 11, and not at the origin.

But we can see it another way.

Let's go to zero temperature.

Zero temperature is way out to the right in figure 11.

And we know what to expect at zero temperature.

We expect alignment.

Everybody must be locked into a parallel configuration.

Then, as we raise the temperature, we slide along the right branch of *tanh* leftward toward the origin.

Eventually when the increasing T reaches $2dj$, we hit the origin, that is the transition point.

Beyond that temperature, the straight line is steeper than 1.

Thence, there is only one solution, $y = 0$, or equivalently $\overline{\sigma} = 0$.

That is the nature of a phase transition, which we'll talk about it a little more.

We have been working with *d* high so that we could apply the mean field approximation.

We haven't shown that there is a phase transition even for low *d* as long as it is greater than or equal to 2.

We have already seen that there is no phase transition in one dimension, because as soon as $< \mu >$ is less that one – see equation (22), the correlation between distant spins falls off to zero, so they cannot obey like an infinite army of robot soldiers.

But let's see, with a different explanation, why a phase transition can't happen in one dimension.

At absolute zero temperature everybody wants to align, figure 12.



Figure 12 : One dimensional collection of small magnets at zero temperature.

Now let's imagine writing the partition function as a sum over all the configurations.

How can we enumerate the configurations starting with the configuration in figure 12 ?

We can enumerate it by the number of spins that are flipped relative to this initial configuration.

So writing the partition function, we would start with *e* to the minus beta times the energy of the initial configuration.

We are now familiar with this.

$$Z = \sum_i e^{-\beta E_i} \tag{39}$$

Starting from zero, figure 12 is the dominant configuration.

Then we raise the temperature a little tiny bit.

Some new configuration start to become important too.

For example we flip one spin in figure 12.

To understand what will happen and not happen, let's also do the same thing in two dimensions.

We start with a two dimensional square lattice at zero temperature.

All the spins are parallel, for instance all up, marked with a plus sign, figure 13.

This is one of the two lowest energy states.

Remember that in the Ising models the energy is stored in the bonds linking the sites.

And two sites pointing in the same direction correspond to a bond with low energy, see equation (10) and the comments thereafter.

So here we start with everybody pointing in the same direction.

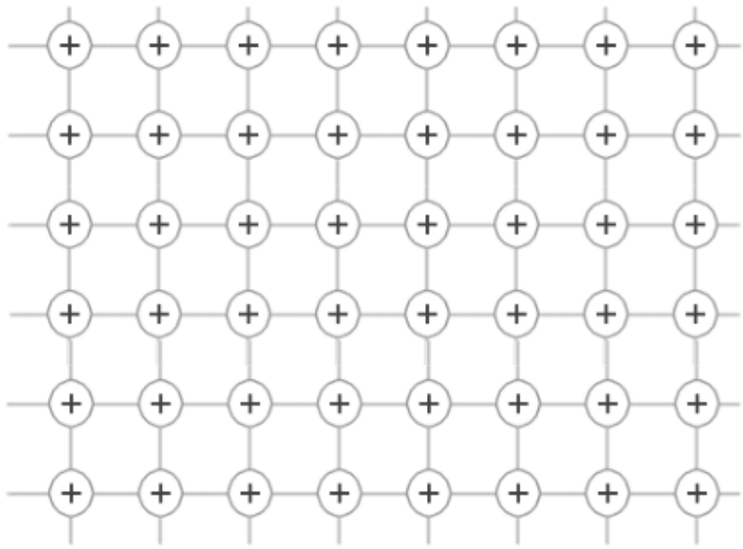


Figure 13 : Two dimensional square lattice of small magnets at zero temperature.

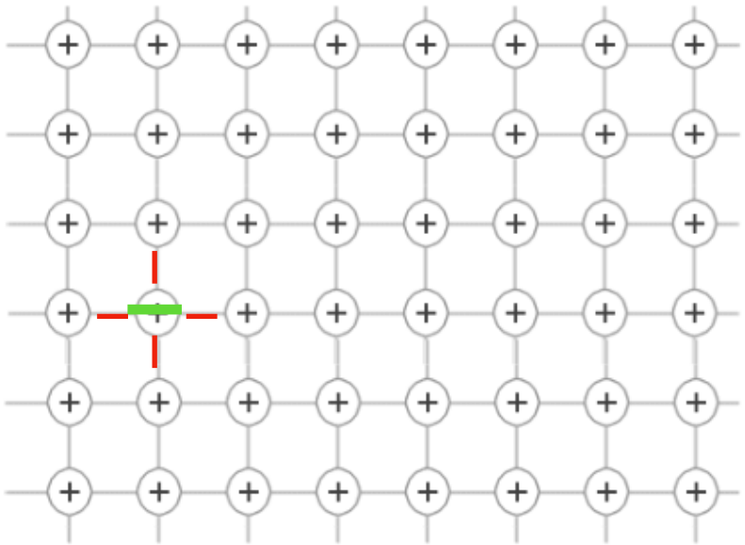
Now let's ask what is the next state, the next higher energy ?

What happens if we flip one spin ?

How much energy do we put ?

How many bonds have we broken ?

Clearly if, in figure 13, we flip only one spin, we break four bonds.



Imagine of course the spin among all the spins, not along a border of the figure.

And breaking a bond means to make it have two anti-aligned spins.

In this case the energy of the bond becomes high.

And how many places are there where we can do that ?

Basically on each site.

The number of places where you can do that is the total number of sites.

So that means there is a number of configurations that we can add in with four extra units of energy.

Now supposing we want to flip two spins, the next configuration.

If the spins are next to each other, as in figure 14, then we break six bonds.

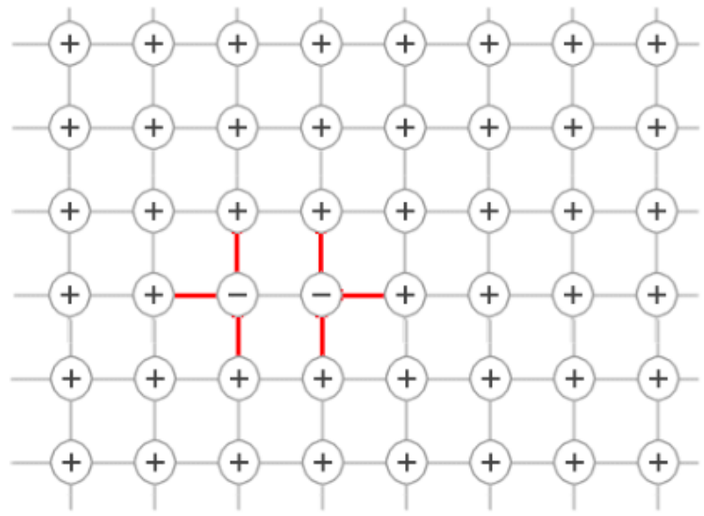


Figure 14 : Two dimensional square lattice, two adjacent spins flipped down, and the six broken bonds.

Notice, in that case, that the bond between the two flipped spins is parallel again, i.e. not broken.

Each broken bond costs a certain amount of energy.

So we have increased the energy by six broken bonds.

What would have happened if we would have put the second minus sign away from the first one ?

Then we get eight broken bonds.

So each time we increase the number of flipped spins, it costs us something.

For each new configuration, a further Boltzmann factor, e to the minus beta times the energy, is added in the partition function.

Each time we flip another spin, the energy goes up.

Now let's compare that with the one-dimensional case.



From the configuration shown in figure 12, how much energy does it cost to flip one spin ?

How many bonds get broken ? Two.

How about to flip two spins ?

If they are adjacent, it is still just two broken bonds.

How about three spins ?

If they are adjacent, it is still two.

You can flip any number of them and it still only costs two units of energy.

That means there is a lot of configurations all with the same energy.

In the two-dimensional case, the number of configurations with four extra units of energy would be just proportional to the number of sites.

In the one-dimensional case, the number of configurations with two extra units of energy is proportional to the square of the number of sites, because you can pick any two bonds and flip all the spins in between them.

So it is a lot of configurations where a lot of spins are flipped.

That is why the one-dimensional case is unstable – in the sense that it doesn't require more energy – with respect to flipping lots and lots of spins.

And as soon as the temperature is turned on, it costs very little energy to flip whole big loads of them.

Whereas in the two-dimensional case to flip one more spin always costs some energy.

That is the basic mathematical effect that is going on.

Question: Can't we see the effect of the difference in the number of dimensions from equation (37) which in one case is $Ty/2j = tanhy$ because $d = 1$, and in the other case is $Ty/4j = tanhy$ because $d = 2$?

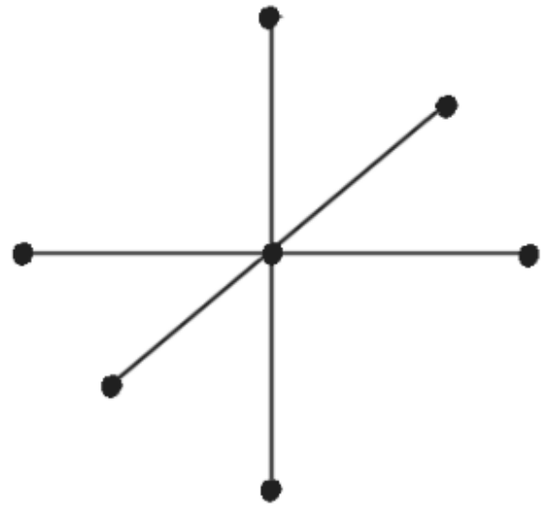
Answer: No we can't.

Equation (37) only makes sense when d is large.

We cannot use it for $d = 1$ nor $d = 2$.

The whole physics of the mean field approximation, the whole argument only makes sense in high dimension.

5.4 Effect of a small external magnetic field



Now let's turn on a tiny magnetic field and see what happens.

Let's still concentrate on one spin, the one at the center of figure 9 for instance.

We saw that its energy function was

$$E = -2dj\overline{\sigma}\sigma \tag{40}$$

But now, in addition, the whole system has an external magnetic field.

That means that each spin has an extra energy not related to its neighbors.

Let's denote the magnetic field B(earlier it was denoted H, but it is the same).

The new energy function for one spin is

$$E = -2dj\overline{\sigma}\sigma - B\sigma \tag{41}$$

where $B\sigma$ is favoring up, that is, B is positive.

However we are going to imagine eventually that B is very small.

$B\sigma$ is the extra energy of the spin under consideration, not related to its neighbors as I said.

And each of the neighbors in figure 9 also has such an extra term.

That is the whole difference : instead of the hyperbolic tangent term, in equation (37, being tanh(2dj $\overline{\sigma}$), it is going to be tanh(2dj $\overline{\sigma}$ + Bβ).

Equation (36) now becomes

$$\frac{Ty}{2dj} = \tanh(y + B\beta) \qquad (42)$$

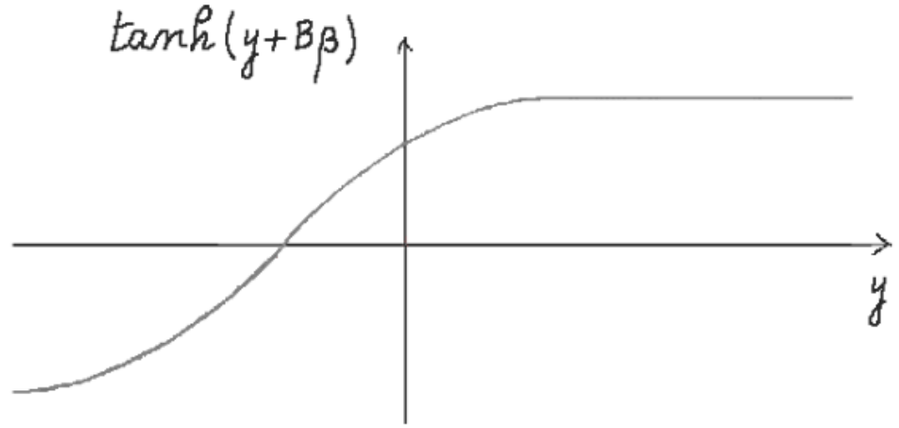


Figure 15 : Graph of tanh(y + Bβ).

The right-hand side has a new graph, figure 15.

We have assumed $B\beta > 0$, so that shifts the curve to the left, compared to that of figures 10 or 11.

Let's draw again the straight lines which we drew in figure 10 and figure 11, figure 16.

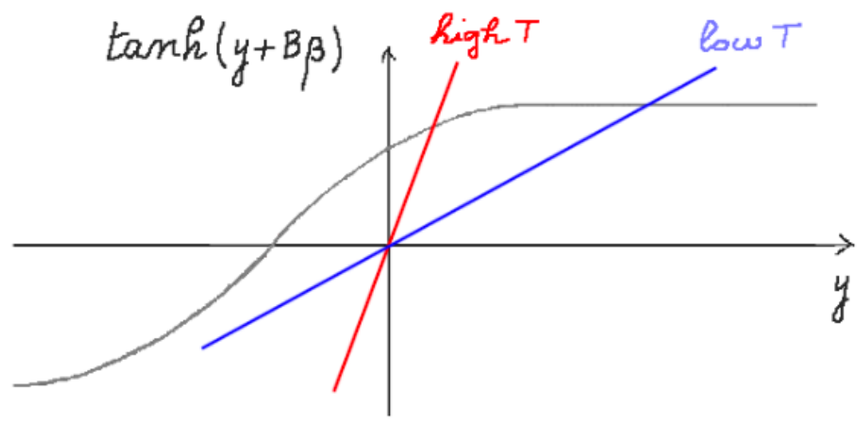


Figure 16 : Graphic solution to $\frac{Ty}{2dj}$ = tanh(y + Bβ).

Equation (42) is a little more complicated than equation (37) because β is actually $1/T$.

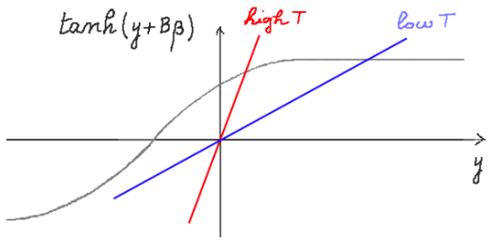
We can rewrite it

$$\frac{Ty}{2dj} = \tanh\left(y + \frac{B}{T}\right) \tag{43}$$

But anyway, at any given positive temperature, and therefore slope of the straight line, β, and shift B/T, there is no solution at the origin anymore.

Only the intersection point on the right branch, for positive j, is possible.

To be accurate, in figure 16, the *tanh* curve should be slightly different, with a different shift, for two different temperatures.



But this doesn't change the conclusion.

Remember that, when we rotate it clockwise, *the straight line goes from vertical to horizontal as the temperature goes from infinity to zero*.

The clockwise knob, so to speak, turns the temperature down.

In the high-dimensional case without a magnetic field, we saw that there is a phase transition : starting from infinite temperature, i.e., the straight line vertical, and slowly rotating it, i.e., lowering the temperature, we remained for a while in a zone of no magnetization, until T reached $2dj$.

At that point there was a phase transition.

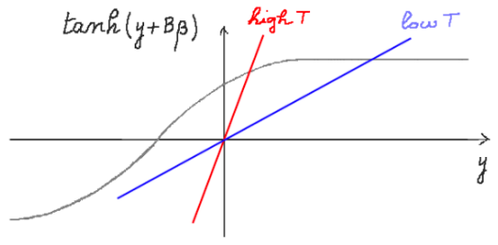
Magnetization appeared, and of course remained till T was zero.

Now, when the whole system is plunged into a magnetic field, however tiny, we no longer have this phenomenon of phase transition appearing at some temperature.

At any positive temperature, there is magnetization, that is there is a bias in $\overline{\sigma}$.

When T is infinite, the slope of the straight line is vertical, and at the same time the *tanh* curve in figure 16 is not shifted at all, then $y = 0$ is still the solution.

There is no magnetization, as expected.



But as soon as T is less than infinite, the only solution is magnetization.

Now we can understand better, in the case of no outside magnetic field what happens when the temperature is below the phase transition point, figure 11.

The three solutions $y = 0$, $y > 0$, and the symmetric $y < 0$, are possible.

But *the solution y = 0, i.e., no magnetization, is unstable*.

Add the tiniest little magnetic field, no matter how small it is, and it will bias one way or the other the whole collection of spins.

It is called *spontaneous symmetry breaking*.

The argument is that the solution where each spin is equally likely to be up or down is unstable.

The solution is that there is fifty fifty probability of everybody being up or everybody being down.

If you turn on the tiniest magnetic field all the spins will turn one way, and the difference of energy to go to the other way can be enormous. Why?

Because if you have a tiny stray magnetic field up and you have a zillion spins, the total energy of up will be much lower than down, just because you have this great number of spins .

The symmetric argument works if you have the tiniest stray magnetic field pointing down.

That is what happens with a real magnet too.

And that may help understand what is going on.

You could imagine that you have a real ferromagnet which has equal probability of pointing everywhere. And then you bring it into the field of the Earth.

Pretty quickly that magnet knows that the Earth's magnetic field is there.

And it orients itself in that direction.

So the configuration where it is equally likely to be in every direction is unstable.

A small stray field will orient it.

And the stronger or the bigger the magnet, the more it is unstable.

In fact the Earth magnetic field itself flips once in a while.

There must be an instability there too.

But that is for geologists to explain why.

Numerical Solution of 2-Dimensional Ising Model

Exact solutions for spin systems can be difficult to obtain.

So one often resorts to computer simulations, of which Monte Carlo is one of the most popular.

Monte Carlo simulations are used widely in physics, e.g., condensed matter physics, astrophysics, high energy physics, etc.

Typically in Monte Carlo simulations, one evolves the system in time.

The idea is to visit a large number of configurations in order to do statistical sampling.

For a spin system we would want to obtain average values of thermodynamic quantities such as magnetization, energy, etc.

More generally, Monte Carlo is an approach to computer simulations in which an event A occurs with a certain probability P_A where $0 \leq P_A \leq 1$.

In practice, during each time step, a random number x is generated with uniform probability between 0 and 1.

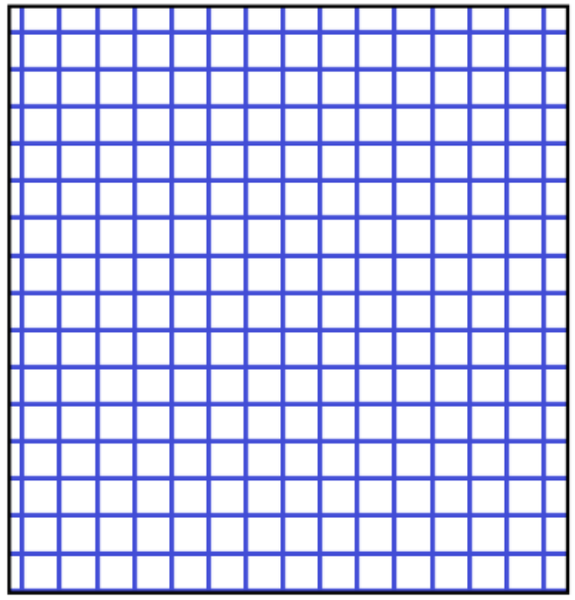
If $x \leq P_A$, event A occurs; if $x > P_A$, event A does not occur.

Monte Carlo is also able to handle cases where multiple outcomes are possible.

For example, suppose there are three possibilities so that either event A can occur with probability P_A or event B can occur with probability P_B or neither occurs.

Then if $x \leq P_A$, A occurs; if $P_A < x \leq (P_A + P_B)$, B occurs; and if $(P_A + P_B) < x \leq 1$, neither occurs.

Consider a lattice with N^2 (N x N) sites and the connectivity of a square lattice.



Each lattice site has a single spin variable: $s_i = \pm 1$.

The energy is:

$$E = -\sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij} s_i s_j$$

where J is the nearest neighbor (i,j) coupling:

The important computational technique used is the Monte Carlo Method is based on the Metropolis Algorithm.

The algorithm has the following steps.

0. Start with some spin configuration $\{s_i\}$.

1. Randomly choose a spin s_i

2. Attempt to flip it, i.e. $s_i := -s_i$ (trial).

3. Compute the energy change ΔE due to this flip.

4. If $\Delta E < 0$, accept the trial.

5. If $\Delta E > 0$, accept the trial with probability $p^{\text{acc}} = e^{-\beta \Delta E}$.

6. If trial is rejected, put the spin back, i.e. $s_i := -s_i$.

7. Go to 1, unless maximum number of iterations is reached.

Every spin in the system is given a chance to flip.

One pass through the lattice is called a “Monte Carlo step/site” (MCS).

This is the unit of time in the simulation.

Some sample Octave code is shown on the next slide.

```
% ising.m
% input parameters
size=256;iters=5;tstart=10.0;tfinal=1.0;tstep=-0.1;h=0.0;
% initialization of spin array
netspin=0;numspins=size*size;
num=((tfinal-tstart)/tstep) + 1;
temp1 = rand(size,size);
temp1= (temp1 >= 0.5) - (temp1 < 0.5);
spins =  temp1;
% calculation of lattice netspin and energy
netspin = sum(sum(spins));
e=sum(sum(-spins.*(spins(:,[[2:size],1])+spins([[2:size],1],:))));
e=e-h*netspin;
% zero plot arrays
tp=[];
sh=[];
sc=[];
mag=[];
eng=[];
figure('Name','Spin Array','Position', [50,400, 300,300])
% start simulated annealing run from tstart
for cnt=1:num
  temp=tstart+(cnt-1)*tstep
  eav=0;
  mav=0;
  % iters thermalization runs at each temperature
  for i = 1:iters
    for y = 1:size
      for x = 1:size
        shere=spins(x,y);
        % periodic boundary conditions on lattice
        sleft=spins(((x-1)*(x ~= 1)+size*(x ==1)),y);
        sright=spins(((x+1)*(x ~=size)+(x ==size)),y);
        sup=spins(x,((y ~= 1)*(y-1)+(y == 1)*size));
        sdown=spins(x,(y ~= size)*(y+1)+(y ==size));
        % Metropolis algorithm
        deltae=2*shere*(sleft+sright+sup+sdown+h);
        om=exp(-deltae/temp);
        if (rand(1) < om)
          spins(x,y)=-shere;
          netspin=netspin-2*shere;
          e=e+deltae;
        end
      end
    end
  end
  % accumulate observable values
  elist(i)=e/numspins;
  mlist(i)=netspin/numspins;
  eav=eav+elist(i);
  mav=mav+mlist(i);
end
  % calculate observables for given temperature
  eav=eav/iters;
  edev=sum((elist-eav).*(elist-eav));
  mav=mav/iters;
  mdev=sum((mlist-mav).*(mlist-mav));
  cv=edev/(iters*temp*temp);
  sus=mdev/(iters*temp);
  %  accumulate plot variables
  tp=[tp,temp];
  sh=[sh,cv];
  sc=[sc,sus];
  mag=[mag,mav];
  eng=[eng,eav];
  % plot image of spin array (2 color)
  image(100*(1+spins))
  colormap(waves)
  drawnow
end
% finished
% plot specific heat versus temperature
figure('Name','Specific Heat','Position', [50,50, 300,300])
plot(tp,sh)
% plot susceptibility versus temperature
figure('Name', 'Susceptibility','Position', [400,50, 300,300])
plot(tp,sc)
figure('Name', 'energy')
plot(tp,eng)
figure('Name', 'magnetization')
plot(tp,mag)
```

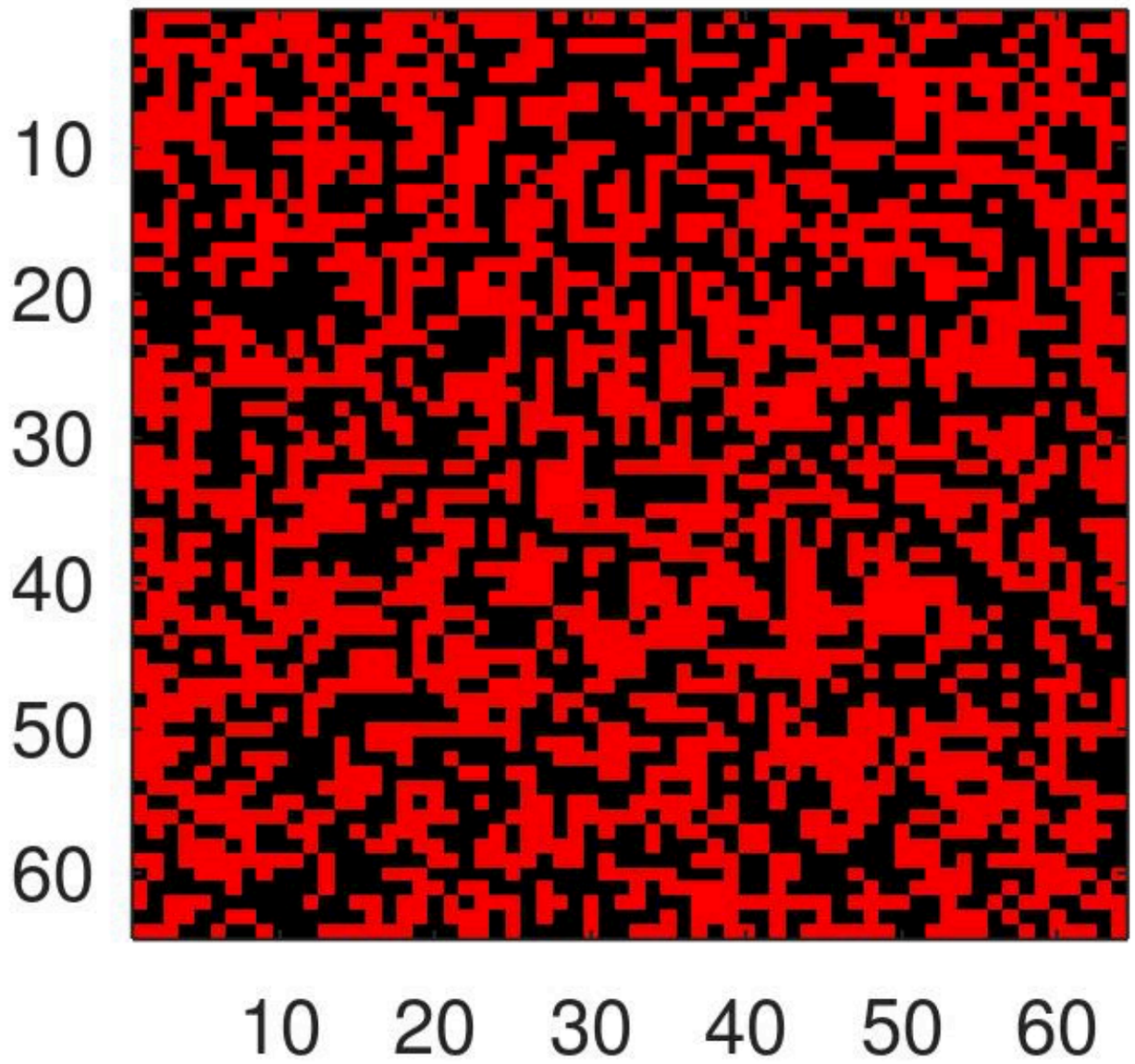
The Monte Carlo technique is very important to learn if you are into doing simulations.

Another sample run on a 64 x 64 lattice.

Sample run on a 64 x 64 lattice.

Temperature starts at 10 and decrease to 1.

The array of spins begin as uniform and at one point a phase transition occurs.

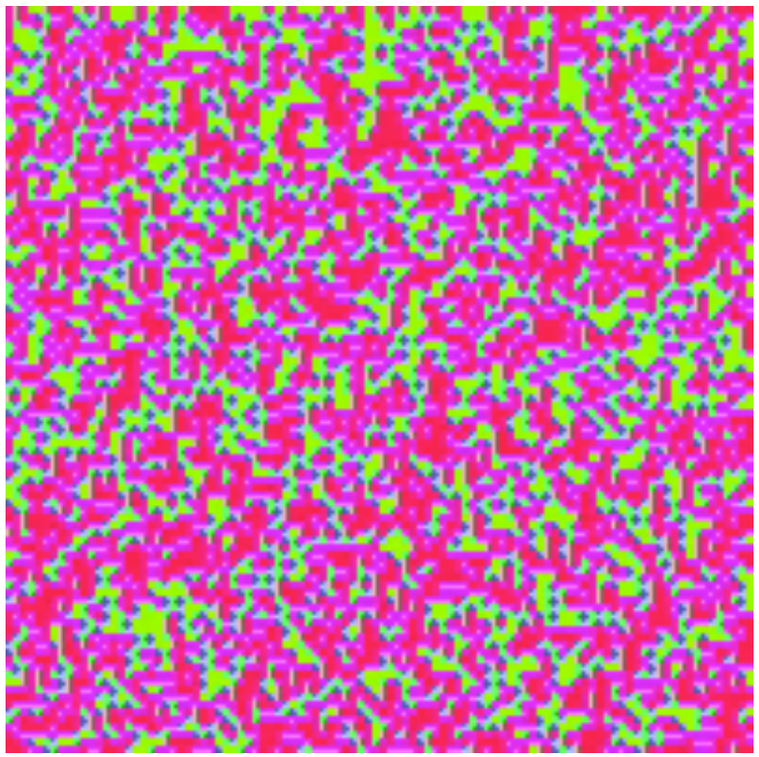


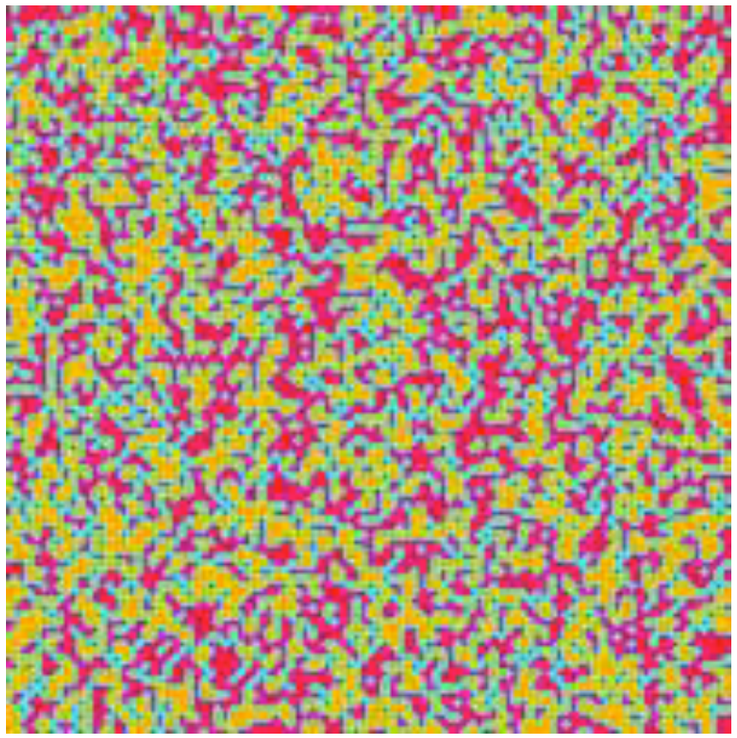
Sample runs on a 256 x 256 lattice.

Temperature starts at 10 and decrease to 1.

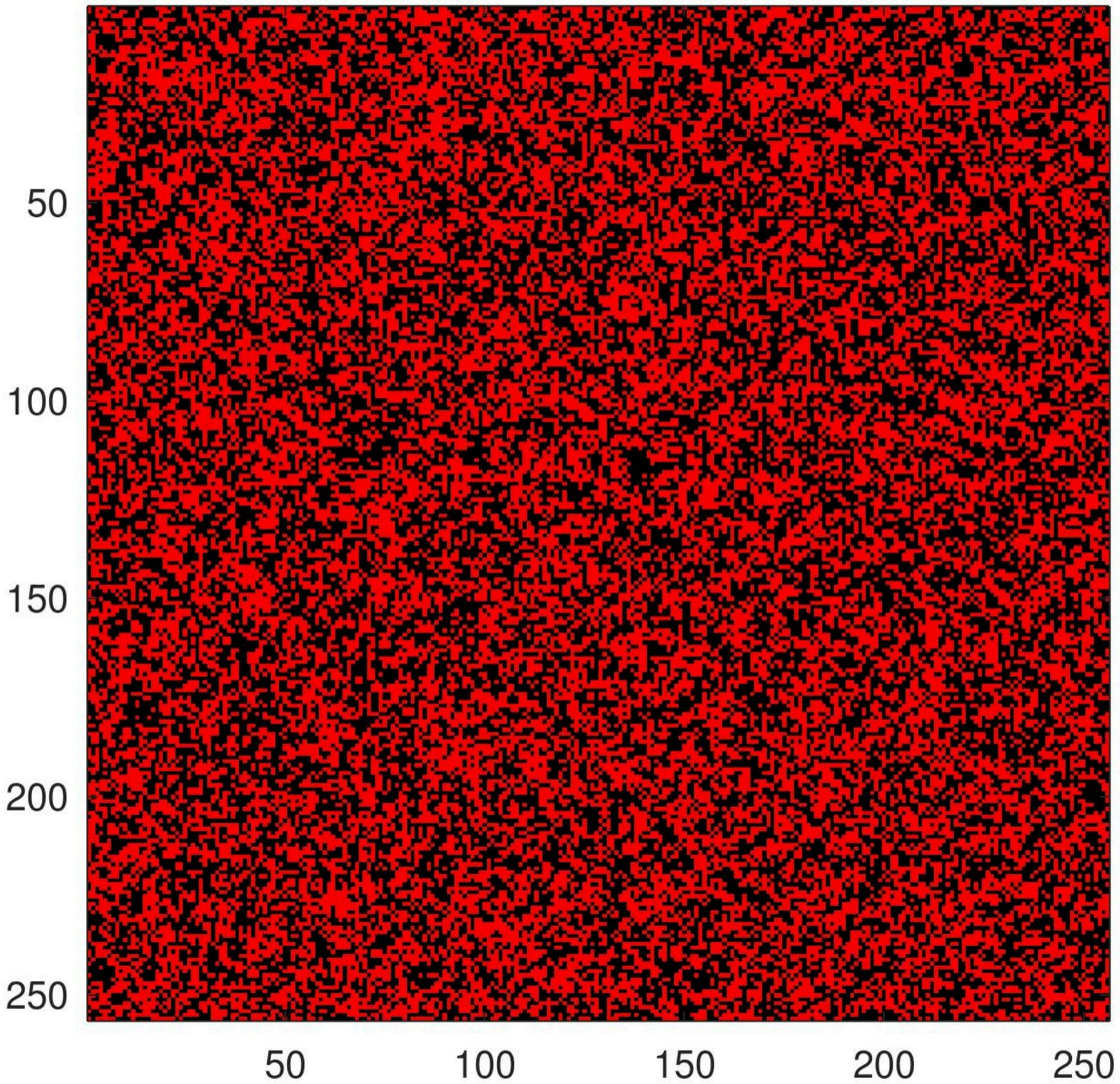
The array of spins begin as uniform and at one point a phase transition occurs.

Points = 8x8 areas



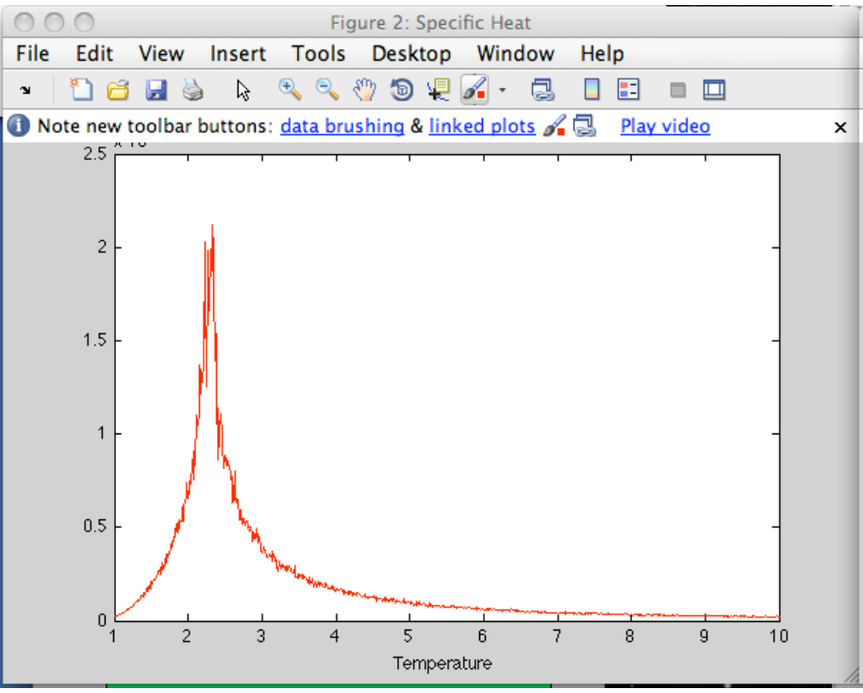


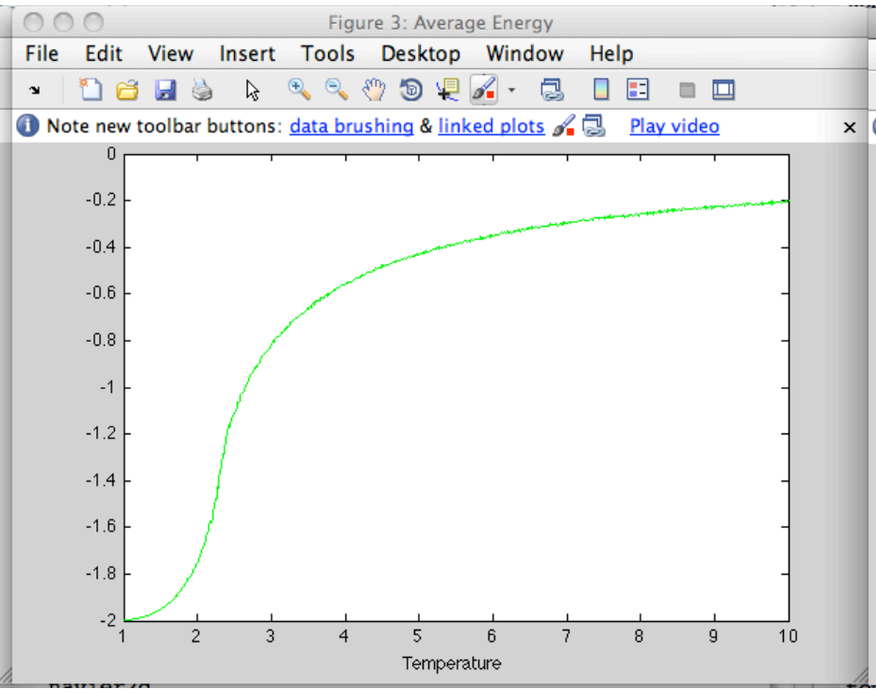
Sample run on a 256 x 256 lattice show all points - maximum detail.



Plots are generated for specific heat, average energy and magnetization as a function of temperature.

The phase transition is clearly shown





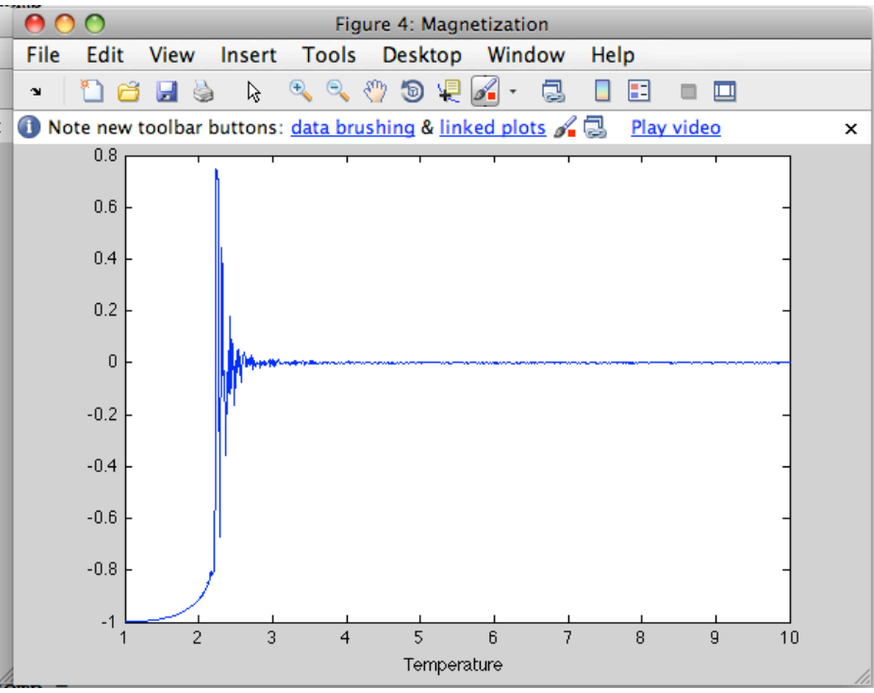
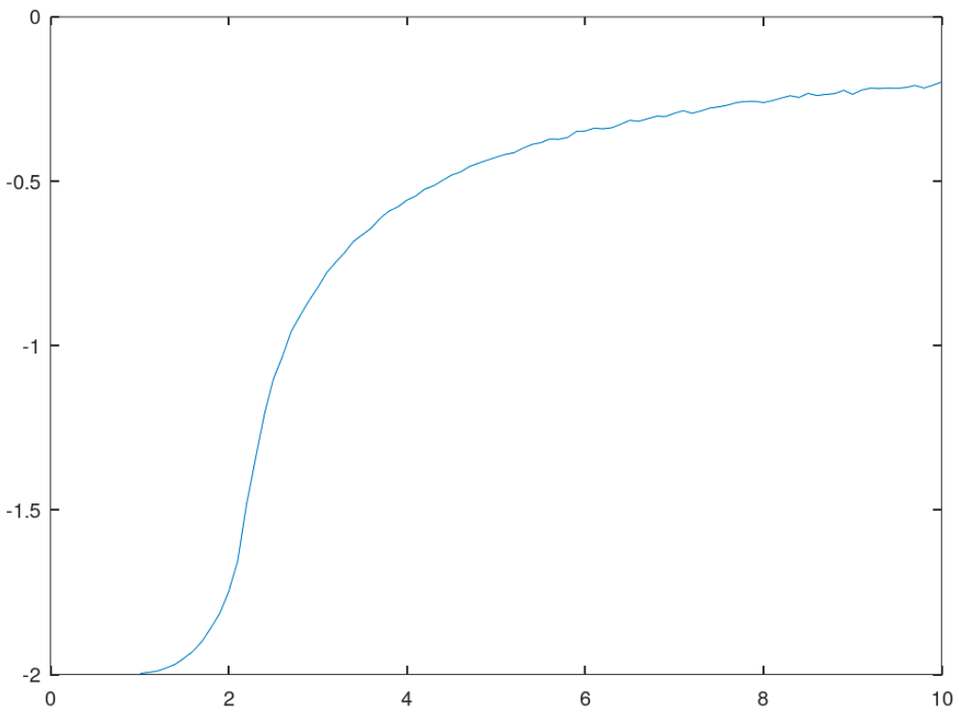
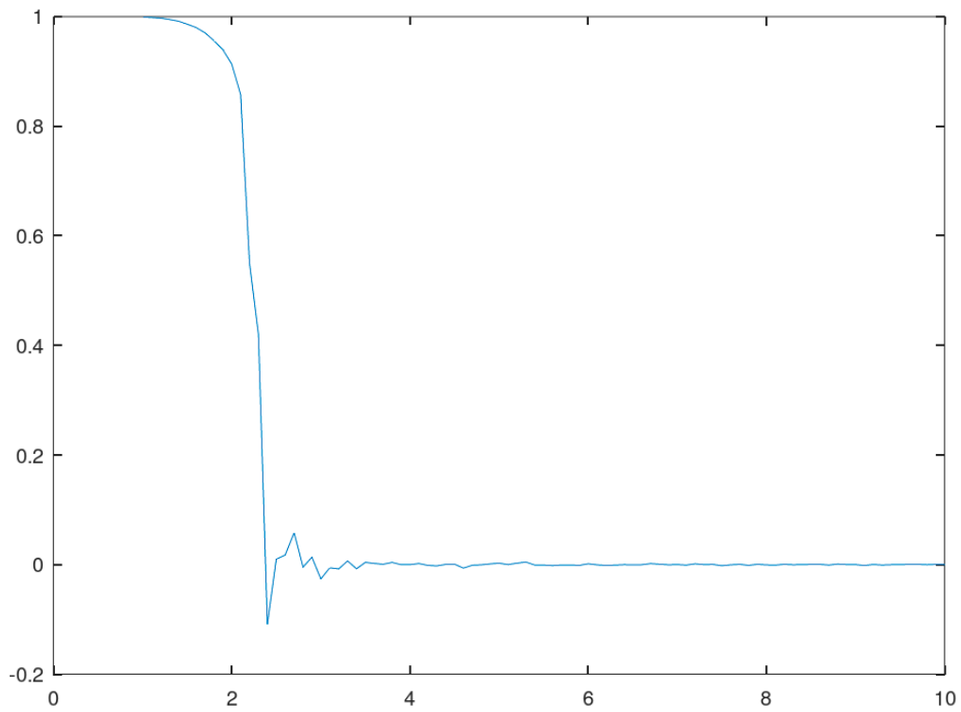


Figure 4: energy



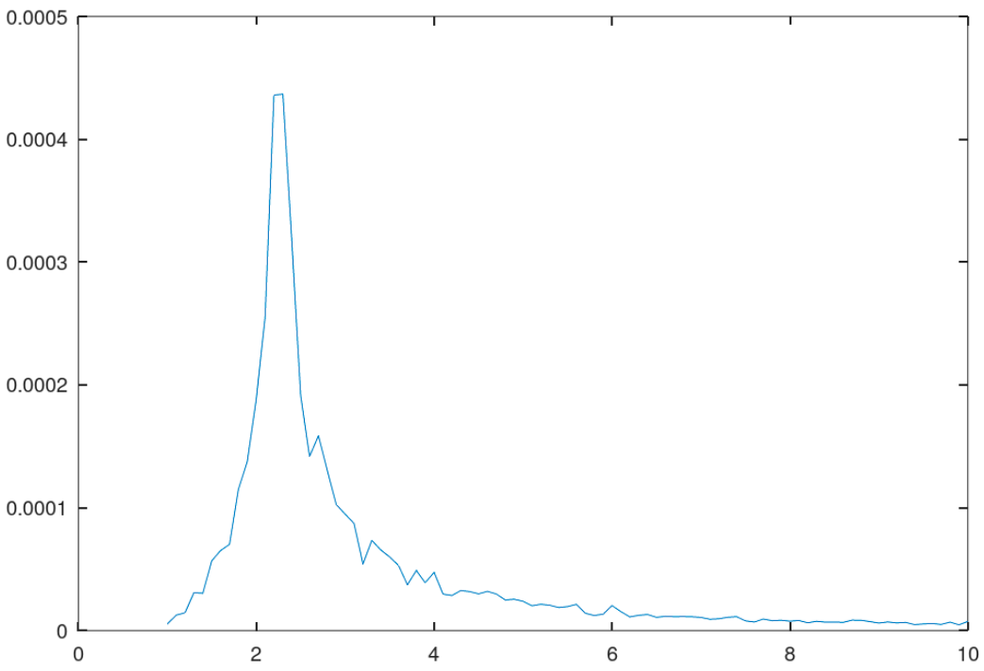
(4.0829, -1.5331)

Figure 5: magnetization



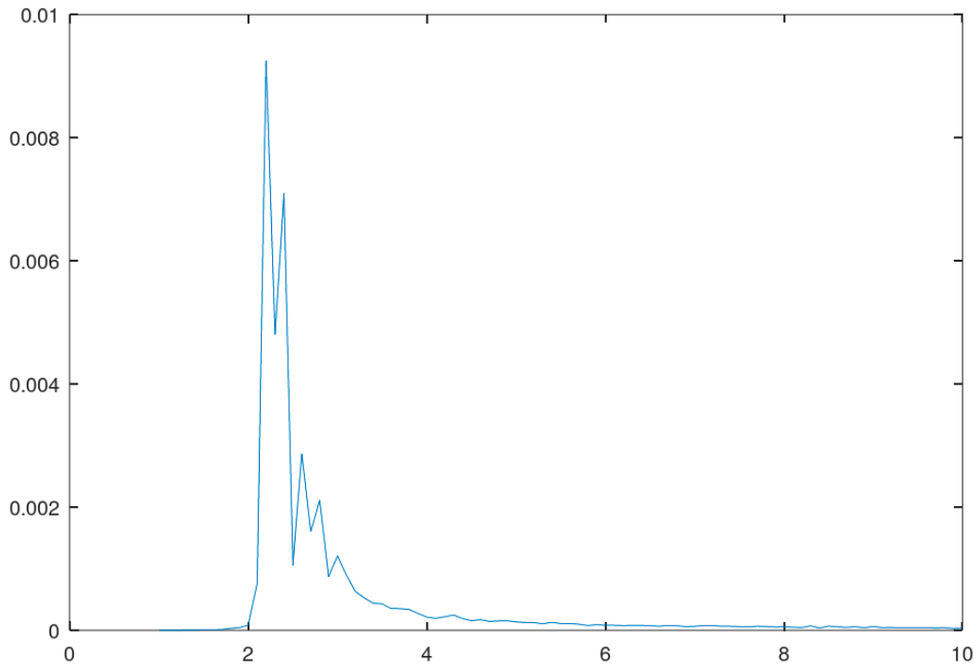
(7.7465, 0.52971)

Figure 2: Specific Heat



(9.9355, 0.00046062)

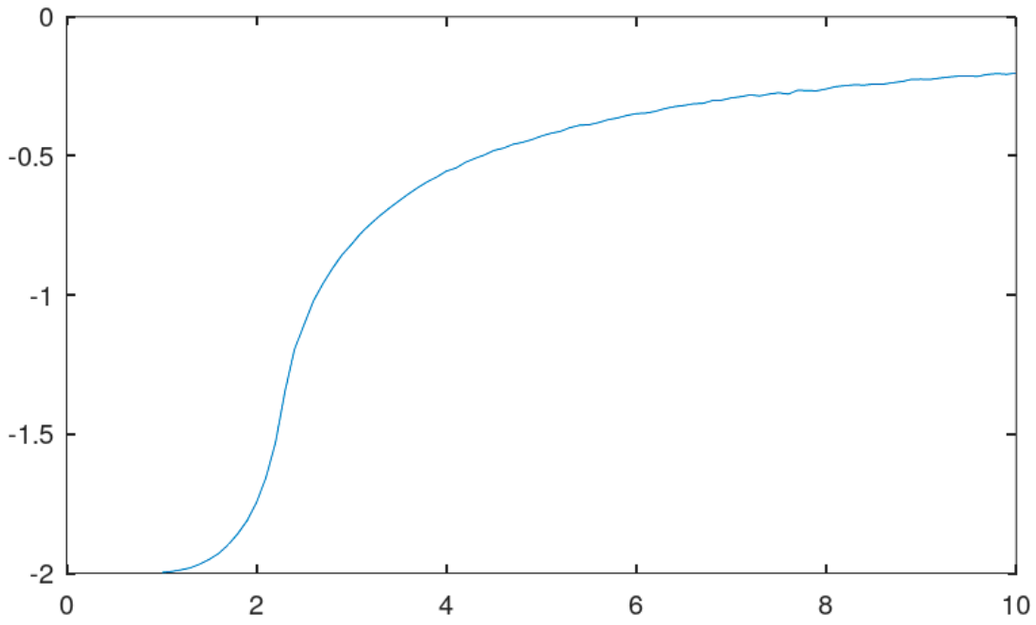
Figure 3: Susceptibility



(9.2903, 8.8541e-05)

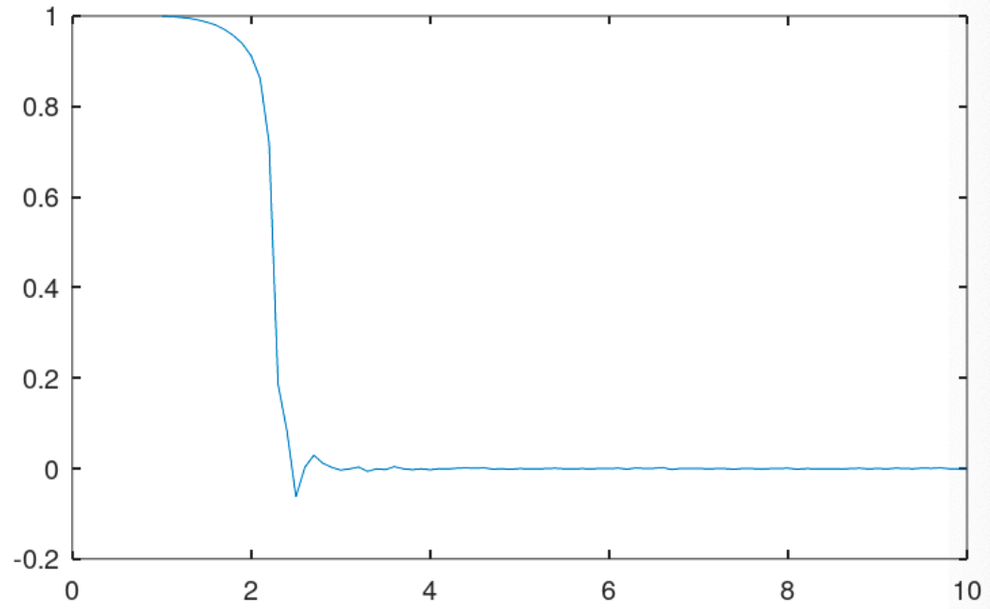
64-200

Figure 4: energy



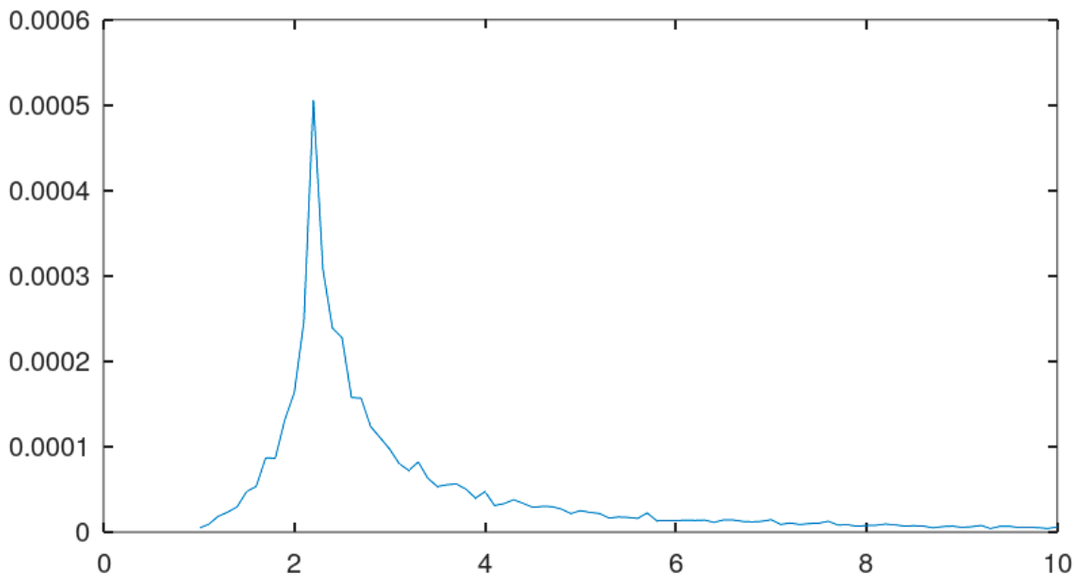
(9.513, -1.2085)

Figure 5: magnetization



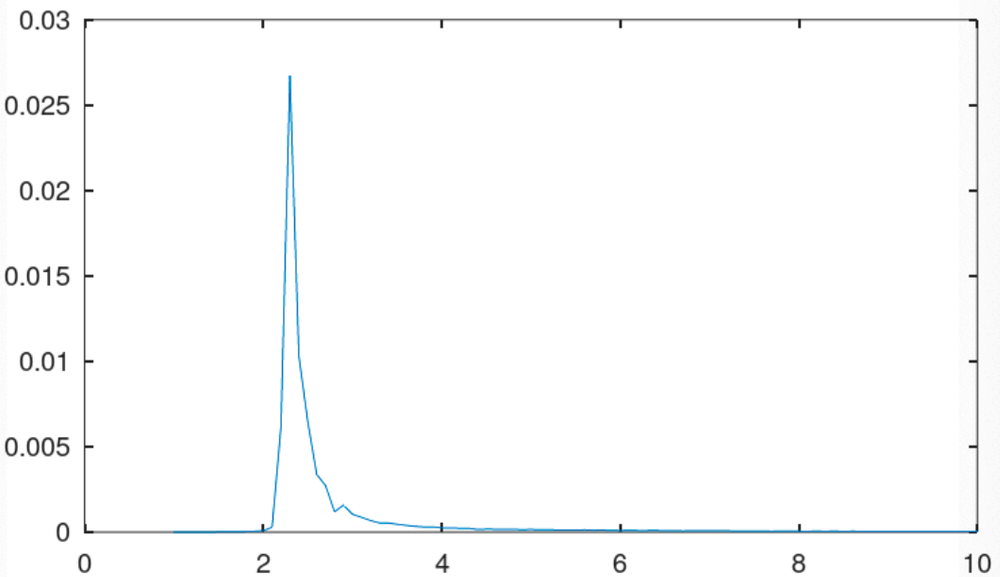
(9.6205, 0.57773)

Figure 2: Specific Heat



(6.65, 0.0005981)

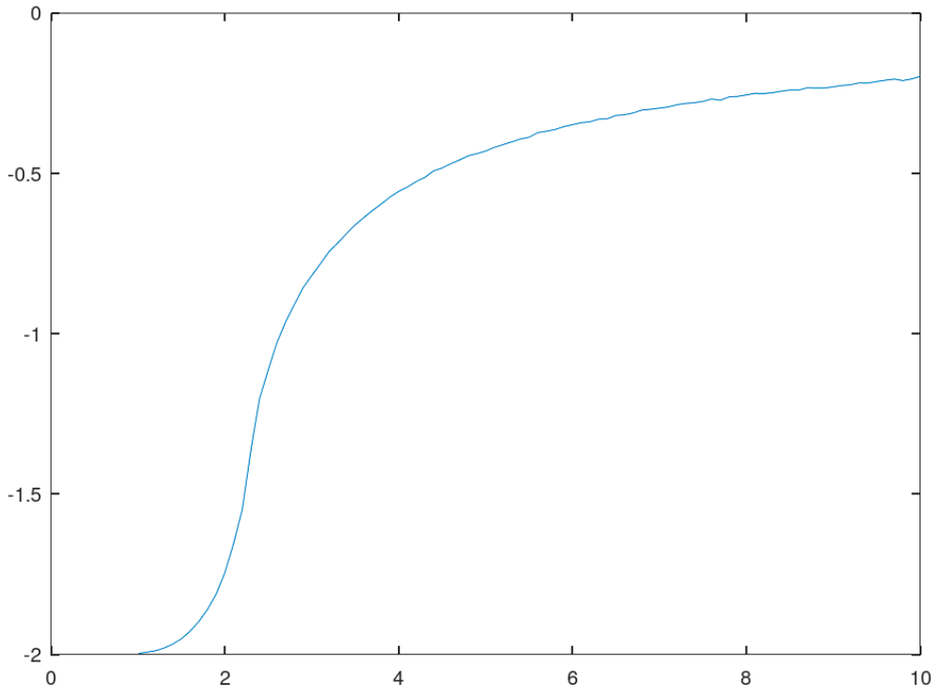
Figure 3: Susceptibility



(0.12903, 5.9159e-05)

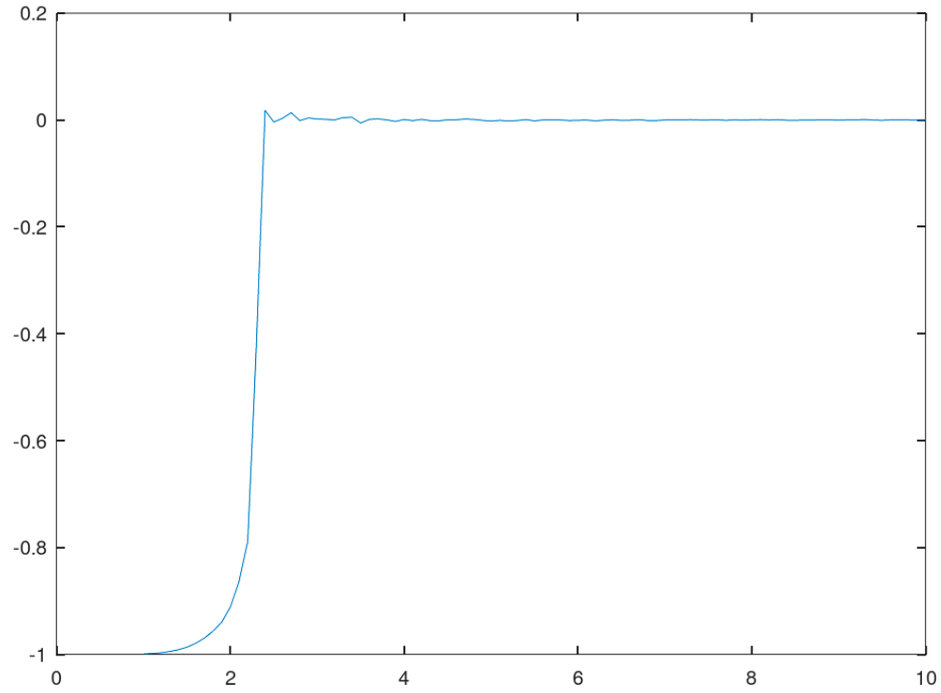
64-300

Figure 4: energy



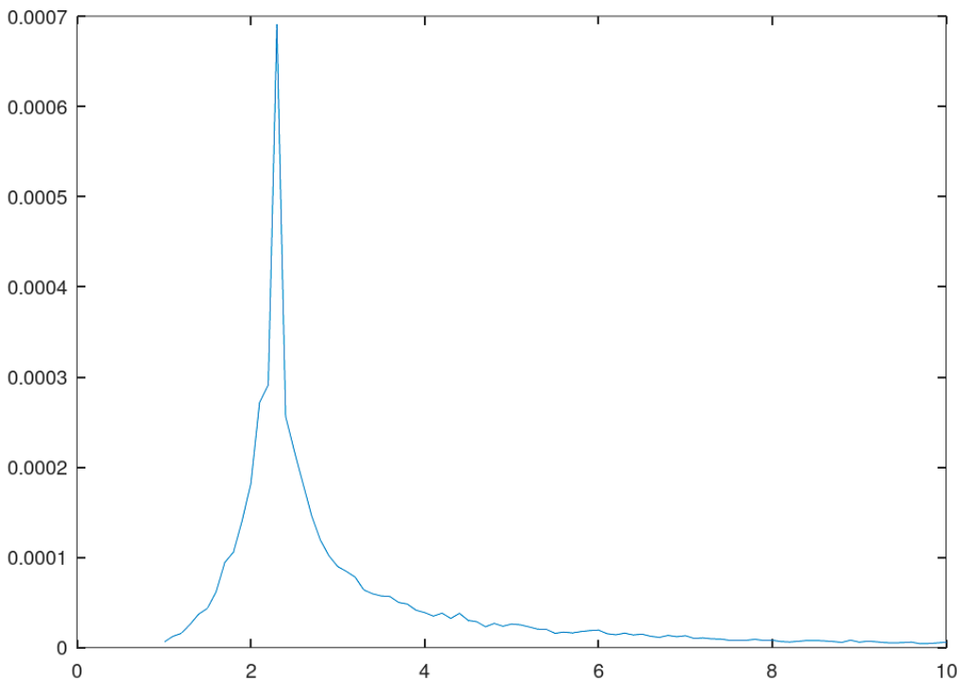
(6.1567, -0.046988)

Figure 5: magnetization



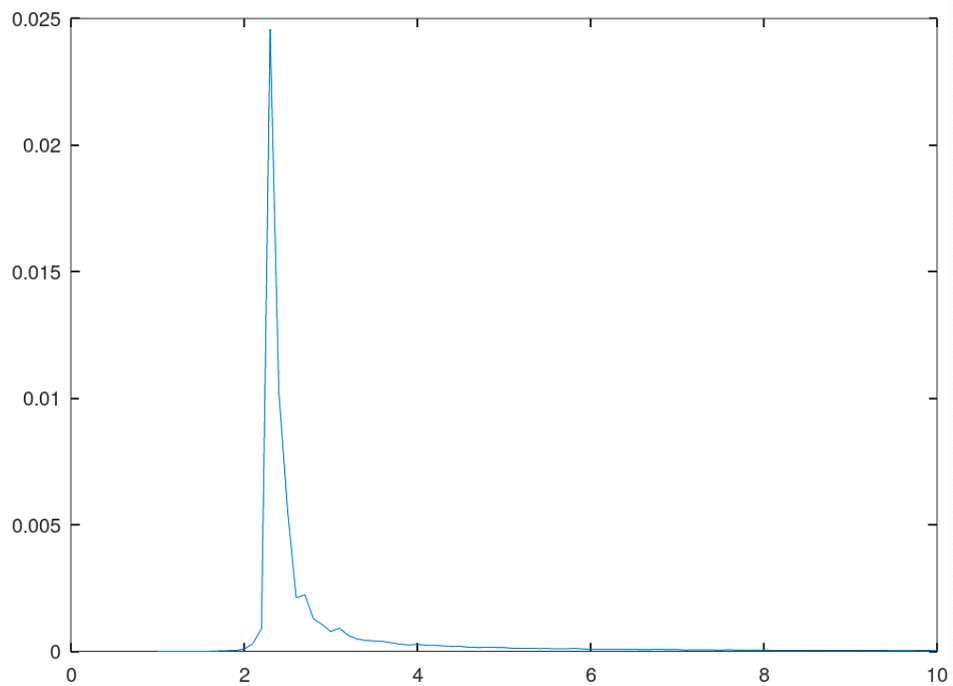
(3.5991, 0.19054)

Figure 2: Specific Heat



(9.4977, 0.00064901)

Figure 3: Susceptibility



(0.17251, 0.00065777)

64-500

Part 10 - Quantum statistics of ideal gases - an idea of the $SM_{\text{next level}}$ using standard methods.

In this section we return to a familiar system: the ideal gas.

But while our previous considerations were limited to the classical limit, in this section we will study the ideal gas at a much more fundamental level that includes in full the effects of quantum mechanics.

New phenomena will thus be brought to light.

Before getting on with our discussion of the quantum statistics of ideal gases, however, it will be necessary to sharpen our statistical-mechanics tools so as to make them easier to use.

We will thus be led toward a new formulation of statistical mechanics, one which is entirely equivalent to the one encountered before.

A major advantage of this new formulation will be the removal of any danger of over-counting the number of states by not taking into account the **indistinguishability** of the particles - important in QM.

You may recall that this problem was overcome earlier by the ad hoc procedure of dividing the statistical weight by $N!$, the total number of permutations among the N particles.

In our new formulation, the indistinguishability of identical particles in quantum mechanics will be taken into account from the very start.

This reformulation will take a while to get accomplished, and during this time, our considerations will be largely formal.

We will derive results that will appear to be very abstract, and the advantages of doing all this will not be clear immediately.

Please bear with it!

The payoff will come.

10.1 Quantum statistics

We want to formulate, in a very precise manner, the statistical mechanics of a system of N identical, noninteracting particles confined to a volume V and in thermal contact with a heat bath at a temperature T.

The relevant statistical ensemble for this problem is the **canonical** ensemble, whose most important quantity is the partition function $Z(\beta,V,N)$, where $\beta = 1/kT$.

To compute this we must sum the Boltzmann factor $e^{-\beta E}$ over all states **accessible** to the entire system.

The major difficulty in doing this is to specify what these states are, in a way that takes into account the fact that the particles are indistinguishable.

This is where our reformulation of statistical mechanics comes in.

10.1.1 Microstates

Instead of thinking about the states of the entire system, focus for the time being on the states accessible to a *single particle*.

These will be called *single-particle states*, and they are extremely easy to characterize: We already know that these states are labeled by three integer-valued quantum numbers, n_x, n_y, and n_z.

This, however, is only a partial description, because it ignores the internal states that arise because of particle spin.

For example, an electron has a spin $s = {}^1/_2$, which can have either one of the two projections $m_s = \pm 1/2$.

The full characterization of the single-particle states therefore requires the specification of four quantum numbers, n_x, n_y, n_z, and m_s.

We may introduce the abstract index r to label the single-particle states.

The index r stands for the set of all four quantum numbers that are necessary to completely specify the states:

$$r = \{n_x, n_y, n_z, m_s\}$$

There is obviously an infinite number of single-particle states, because n_x, n_y, and n_z can all take an infinite number of values.

The quantum number m_s, on the other hand, ranges from $-s$ to $+s$ with integral steps.

The zero-energy state $(n_x, n_y, n_z, m_s) = (0,0,0,-s)$ will be symbolically represented by $r = 0$.

Similarly, $r = 1$ stands for $(0, 0, 0, -s + 1)$, and so on.

The energy eigenvalue corresponding to the single-particle state r is denoted ε_r.

We know that this is given by

$$\varepsilon_r = \frac{(2\pi\hbar)^2}{8mV^{2/3}} \left({n_x}^2 + {n_y}^2 + {n_z}^2\right).$$

which shows that ε_r is independent of m_s.

How do we go from the single-particle states to the states of the entire system?

The key is to introduce the quantity n_r, called the *occupation number* of the single-particle state r.

By definition, this is the total number of particles which happen to occupy the particular single-particle state r.

Notice that we do not care which particles among the N particles happen to occupy the specific state r.

All we are asking is how many particles do, and the answer is provided by the occupation number n_r.

This is where the indistinguishability of the particles is taken into account.

Once we have the complete list of all the occupation numbers, we know precisely how many particles occupy each and everyone of the single-particle states r.

Since this is all that can be known about the system without violating the indistinguishability postulate of quantum mechanics, the state of the entire system is now completely determined.

If we use the abstract index R to label the states of the entire system, then R stands for the complete set of all the occupation numbers, one for each of the single-particle states r.

In mathematical notation,

$$R = \{n_0, n_1, n_2, \ldots\} = \{n_r\}$$

We now have a way of characterizing the states of the entire system in terms of the single-particle states.

The total energy of the system is given by $\varepsilon_0 (= 0)$ times n_0, the number of particles in the state $r = 0$, plus ε_1 times n_1, the number of particles in the state $r = 1$, plus all other contributions.

If we denote the total energy of the state R by E_{R}, then

$$E_R = \sum_r n_r \, \varepsilon_r$$

where the sum is over all single-particle states.

Similarly, the total number of particles in the state R is given by the sum of all the occupation numbers:

$$N_R = \sum_r n_r$$

In our case, the number of particles is fixed, so $N_R = N$ for all states R.

This shows that the occupation numbers are not all independent, because their sum must always reproduce the total number of particles.

10.1.2 Bosons and fermions

It is one of the most fundamental facts of nature that elementary particles, atoms, and molecules are divided into two groups called bosons and fermions.

Bosons are particles possessing integer spin ($s = 0,1,2,\cdots$).

For the purpose of this discussion, the most important property of bosons is that an *arbitrary* number of them can exist in the same single-particle state r.

In other words, their occupation numbers can take any value between 0 and N:

$$n_r = 0, 1, 2, \cdots, N \qquad \text{(bosons)}.$$

Fermions, on the other hand, are particles possessing half-integer spin ($s = {}^{1}/_{2}, {}^{3}/_{2}, \cdots$).

The most important property of fermions is that two or more fermions *cannot* occupy the same single-particle state r.

Their occupation numbers, therefore, can only take the values 0 and 1:

$$n_r = 0, 1 \qquad \text{(fermions)}.$$

For both bosons and fermions, the occupation numbers are further restricted by the condition $\sum_r n_r = N$.

10.1.3 Partition function

Let us now attempt to calculate the partition function of an ideal gas consisting of bosons.

Using our previous results, the partition function is given by

$$Z = \sum_R e^{-\beta E_R} = \sum_{\{n_r\}} \exp\left(-\beta \sum_r n_r \varepsilon_r\right),$$

where the sum is over all the individual occupation numbers n_r. Because of the restriction on $\sum_r n_r$, this becomes

$$Z = \sum_{n_0=0}^{N} e^{-\beta n_0 \varepsilon_0} \sum_{n_1=0}^{N-n_0} e^{-\beta n_1 \varepsilon_1} \sum_{n_2=0}^{N-n_0-n_1} e^{-\beta n_2 \varepsilon_2} \times \cdots$$

These sums are very difficult to evaluate, because of the explicit dependence of the upper limits on the occupation numbers.

There seems to be little hope that this calculation could be pushed forward.

Evaluation of the partition function is made difficult by the constraint $\sum_r n_r = N$, which enforces the condition that we are dealing with a fixed number of particles.

If, however, this condition were to be removed, then the constraint would disappear, and the sums could presumably be evaluated much more easily.

Could we possibly do this and get away with it, i.e., still get the correct result?

10.1.4 Grand partition function

The answer is yes.

Let us pretend that the number of particles is not fixed, and that the system's interaction with the reservoir allows for an exchange of particles.

In effect, we are switching from a canonical (Boltzmann) description to a grand canonical (Gibbs) description.

10.1.4.1 Gibbs distribution

The Gibbs distribution applies to a system interacting with a reservoir, when the interaction allows for an exchange of both heat and particles.

Here it is convenient to introduce, apart from $\beta \equiv 1/kT$, the parameter

$$\boxed{\alpha = -\frac{\mu}{kT}} \tag{10.1}$$

where μ is the chemical potential of the reservoir. We discussed these ideas earlier in thermodynamics.

The Gibbs distribution of probabilities is then given by

$$\boxed{p_r = \frac{1}{\mathcal{Z}} e^{-(\beta E_r + \alpha N_r)}, \qquad \mathcal{Z} = \sum_r e^{-(\beta E_r + \alpha N_r)}} \tag{10.2}$$

The quantity $\mathcal{Z}$ is called the *grand partition function*, and it plays essentially the same role that the partition function plays in the Boltzmann distribution.

In particular, we will see that all thermodynamic quantities can be computed directly from the grand partition function.

10.1.4.2 Thermodynamic quantities

We define the thermodynamic energy E exactly as before, as the system's mean energy.

The only difference is that now, the average is computed with the Gibbs distribution.

So

$$E \equiv \overline{E_r} = \sum_r p_r E_r = \frac{1}{\mathcal{Z}} \sum_r E_r e^{-(\beta E_r + \alpha N_r)}$$

It is easy to show, as we have done before for the Boltzmann distribution, that E can be conveniently expressed as

$$\boxed{E = -\frac{\partial \ln \mathcal{Z}}{\partial \beta}} \tag{10.3}$$

The first law of thermodynamics contains the term $\mu\, dN$, where N is the system's macroscopic number of particles.

Since our system does not have a fixed number of particles, we must also define N as an averaged quantity.

More precisely, we define N as

$$N = \overline{N_r} = \sum_r p_r N_r = \frac{1}{\mathcal{Z}} \sum_r N_r e^{-(\beta E_r + \alpha N_r)}.$$

The same steps that lead to Eq. (3.20) can be applied here, and we find

$$\boxed{N \equiv -\frac{\partial \ln \mathcal{Z}}{\partial \alpha}} \tag{10.4}$$

We now calculate the entropy from the definition.

Expressing $\ln p_r$ as $-(\ln Z + \beta E_r + \alpha N_r)$, we obtain

$$S = k \sum_r p_r (\ln \mathcal{Z} + \beta E_r + \alpha N_r)$$

or

$$\boxed{S = k(\ln \mathcal{Z} + \beta E + \alpha N)} \tag{10.5}$$

This implies that the Landau potential $\Omega = E - T S - \mu N$ is very closely related to the grand partition function.

Indeed, a few steps of algebra yield

$$\boxed{\Omega(T, V, \mu) = -kT \ln \mathcal{Z}(T, V, \mu)} \tag{10.6}$$

That Ω, and therefore also $\mathcal{Z}$, depend on the quantities T, V, and μ follows from the relation

$$d\Omega = -S\,dT - P\,dV - N\,d\mu$$

was derived when we covered Thermodynamics.

This relation also gives definitions for S, P , and N in terms of partial derivatives of Ω:

$$\boxed{S = -\left(\frac{\partial\Omega}{\partial T}\right)_{V,\mu} \qquad P = -\left(\frac{\partial\Omega}{\partial V}\right)_{T,\mu} \qquad N = \left(\frac{\partial\Omega}{\partial\mu}\right)_{T,V}} \tag{10.7}$$

Notice that the equations $S = -\partial\Omega/\partial T$ and $N = -\partial\Omega/\partial\mu$ are entirely equivalent to Eqs. (10.5) and (10.4), respectively.

So, if the grand partition function is known as a function of T , V , and μ, then the Landau potential is known, and all other thermodynamic quantities can be obtained by partial differentiation.

In other words:

> For a system in thermo-chemical contact with a reservoir at temperature T and chemical potential μ, every thermodynamic quantity can be calculated directly from the grand partition function $\mathcal{Z}\ (T, V, \mu)$.

This is highly analogous to the situations encountered before.

10.1.4.3 Fluctuations

A system described by the Gibbs distribution interacts with a reservoir, and the interaction allows for an exchange of both energy and particles.

We ask: how strongly can the mean values E and N deviate, at some instant of time, from the actual values for the energy and the number of particles?

In other words, how large are the fluctuations?

In the case of the Boltzmann distribution we saw that the fluctuations in energy are very small.

Here we will show that the fluctuations are small also in the case of the Gibbs distribution.

We define ΔE and ΔN by

$$\boxed{(\Delta E)^2 = \overline{(E_r - E)^2}, \qquad (\Delta N)^2 = \overline{(N_r - N)^2}} \tag{10.8}$$

In the case of the Boltzmann distribution,

$$\boxed{(\Delta E)^2 = -\left(\frac{\partial E}{\partial \beta}\right)_{\alpha, V}} \tag{10.9}$$

This result is valid also for the Gibbs distribution.

Instead of proving this explicitly, we will derive the analogous result for ΔN; the steps are virtually identical.

We first express *N* as

$$N = \frac{\sum_r N_r e^{-(\beta E_r + \alpha N_r)}}{\sum_r e^{-(\beta E_r + \alpha N_r)}}$$

and compute its partial derivative with respect to α:

$$-\frac{\partial N}{\partial \alpha} = \frac{\sum_r {N_r}^2 e^{-(\beta E_r + \alpha N_r)}}{\sum_r e^{-(\beta E_r + \alpha N_r)}} - \left(\frac{\sum_r N_r e^{-(\beta E_r + \alpha N_r)}}{\sum_r e^{-(\beta E_r + \alpha N_r)}} \right)^2$$

We may now identify the first term of the right-hand side with $\overline{{N_r}^2}$, and the second term with $\overline{N_r}^2$.

Since

$$\overline{(N_r - N)^2} = \overline{{N_r}^2} - \overline{N_r}^2$$

we have established that

$$\boxed{(\Delta N)^2 = -\left(\frac{\partial N}{\partial \alpha} \right)_{\beta, V}} \qquad (10.10)$$

To demonstrate that the fluctuations are typically very small, we evaluate Eq. (10.10) for the specific case of an ideal gas.

We must first express N in terms of α and β; otherwise, we might get confused when taking the partial derivative.

We recall that for an ideal gas,

$$\alpha = -\frac{\mu}{kT} = \ln\left[\frac{V}{(2\pi\hbar)^3 N}\left(\frac{2\pi m}{\beta}\right)^{3/2}\right]$$

which implies

$$N = \frac{V}{(2\pi\hbar)^3}\left(\frac{2\pi m}{\beta}\right)^{3/2} e^{-\alpha}$$

This immediately gives us $\partial N/\partial\alpha = -N$, and

$$\frac{\Delta N}{N} = \frac{1}{\sqrt{N}}$$

We see that the fluctuation in the number of particles is indeed very small.

We shall now work out an expression for the energy fluctuation.

To evaluate Eq. (10.9) we express $E = {}^{3}/_{2}\, NkT$ as

$$E = \frac{3}{2}\,\frac{V}{(2\pi\hbar)^3}\,\frac{(2\pi m)^{3/2}}{\beta^{5/2}}\,e^{-\alpha}$$

which implies that $\partial E/\partial\beta = -\,{}^{5}/_{2}\, E/\beta$.

Finally, we arrive at

$$\frac{\Delta E}{E} = \sqrt{\frac{5}{3N}}$$

This is slightly different from the result with the Boltzmann distribution.

The difference is due to the fact that the fluctuations in the number of particles induce additional fluctuations in the energy, so that $(\Delta E)_{\text{Gibbs}} > (\Delta E)_{\text{Boltzmann}}$.

We conclude that for all practical purposes, the physical predictions obtained using the Gibbs distribution are virtually identical to those obtained using the Boltzmann distribution.

Even though the number of particles is subject to fluctuations, these are so small that they can be safely ignored.

$$p_r = \frac{1}{\Omega}$$

where Ω is the statistical weight.

The canonical ensemble designates a large number of identical systems in thermal interaction (V and N fixed) with a reservoir; in the canonical ensemble, probabilities are given by the Boltzmann distribution,

$$p_r = \frac{1}{Z} \exp\bigl(-E_r/kT\bigr).$$

Finally, the grand canonical ensemble designates a large number of identical systems in thermo-chemical interaction (only V fixed) with a reservoir; in the grand canonical ensemble, probabilities are given by the Gibbs distribution,

$$p_r = \frac{1}{\mathcal{Z}} \exp\Bigl[-\bigl(E_r - \mu N_r\bigr)/kT\Bigr]$$

10.1.4.4 Brief summary of statistical mechanics

We use the following symbols: $\alpha \equiv -\mu/kT, \beta \equiv 1/kT$.

The index r labels the microstates accessible to the system; these have energy eigenvalues E_r and particle numbers N_r.

The sum $\sum_r$ is over all these microstates.

10.1.4.5 Boltzmann distribution

Valid for systems in thermal interaction (V and N fixed) with a reservoir.

- Boltzmann distribution:

$$p_r = \frac{1}{Z} e^{-\beta E_r}$$

- Partition function:

$$Z(\beta, V, N) = \sum_r e^{-\beta E_r} = \int \omega(\mathcal{E}) e^{-\beta \mathcal{E}} \, d\mathcal{E}$$

- Helmholtz free energy:

$$F = -kT \ln Z$$

- Mean energy:

$$E = -\frac{\partial \ln Z}{\partial \beta}$$

- Entropy:

$$S = k(\ln Z + \beta E)$$

- Pressure:

$$P = kT \frac{\partial \ln Z}{\partial V}$$

- Chemical potential:

$$\mu = -kT \frac{\partial \ln Z}{\partial N}$$

- Energy fluctuation:

$$(\Delta E)^2 = -\frac{\partial E}{\partial \beta}$$

10.1.4.6 Gibbs distribution

Valid for systems in thermo-chemical interaction (only V fixed) with a reservoir.

- Gibbs distribution:

$$p_r = \frac{1}{\mathcal{Z}} e^{-(\beta E_r + \alpha N_r)}$$

- Grand partition function:

$$\mathcal{Z}(\alpha, \beta, V) = \sum_r e^{-(\beta E_r + \alpha N_r)}$$

- Landau potential:

$$\Omega = -kT \ln \mathcal{Z}$$

- Mean number of particles:

$$N = -\frac{\partial \ln \mathcal{Z}}{\partial \alpha}$$

- Mean energy:

$$E = -\frac{\partial \ln \mathcal{Z}}{\partial \beta}$$

- Entropy:

$$S = k(\ln \mathcal{Z} + \beta E + \alpha N)$$

- Pressure:

$$P = kT \frac{\partial \ln \mathcal{Z}}{\partial V}$$

- Fluctuations:

$$(\Delta N)^2 = -\frac{\partial N}{\partial \alpha}, \qquad (\Delta E)^2 = -\frac{\partial E}{\partial \beta}$$

Thus, this switch to Gibbs should not produce any changes in our physical predictions:

As we have seen, the fluctuation $\Delta N/N$ in the number of particles in a system described by the Gibbs distribution is extremely small, of the order of $N^{-1/2}$.

So, although we shall no longer *formally* impose that the number of particles be fixed, this number will still be fixed for all practical purposes.

The advantages associated with the removal of the constraint $\sum_r n_r = N$ will thus be obtained at virtually no cost at all.

The physical picture is now the following.

We have a system of identical, non-interacting particles confined to a volume V and in contact with a reservoir at a temperature T and with a chemical potential μ.

As a result of the interaction between the system and the reservoir, energy and particles are exchanged, but only in microscopic amounts, as the fluctuations are extremely small.

The most important quantity of this description is $\mathcal{Z}(\alpha, \beta, V)$, the grand partition function, where $\alpha = -\mu/kT, \beta = 1/kT$.

This is given by

$$\mathcal{Z} = \sum_R e^{-(\alpha N_R + \beta E_R)} = \sum_{\{n_r\}} \exp\left[-\left(\alpha \sum_r n_r + \beta \sum_r n_r \varepsilon_r\right)\right]$$

and becomes, after a few lines of algebra,

$$\boxed{\mathcal{Z} = \sum_{n_0} e^{-(\alpha+\beta\varepsilon_0)n_0} \sum_{n_1} e^{-(\alpha+\beta\varepsilon_1)n_1} \sum_{n_2} e^{-(\alpha+\beta\varepsilon_2)n_2} \times \cdots} \tag{10.11}$$

We shall leave the grand partition function in this form for the time being.

The main advantage of this expression over what we had previously obtained for Z is that the occupation numbers are now all independent.

For this reason, the sums will be easy to evaluate.

This alternative description of our problem has introduced a new quantity, α, which is related to the chemical potential μ by $\alpha = -\mu/kT$.

At the same time, the quantity N seems to have disappeared from the description.

However, because of the fact that the fluctuations are small, we can identify N with N_{R}, the system's mean number of particles.

This is related to α through the relation, $N_{\mathrm{R}} = -\partial\ ln\mathcal{Z}\,/\partial\alpha$.

The parameter α is therefore determined implicitly in terms of N by the relation

$$\boxed{N = -\frac{\partial \ln \mathcal{Z}}{\partial \alpha}} \tag{10.12}$$

10.1.5 Mean occupation numbers

We have seen earlier that once the grand partition function has been obtained, the thermodynamic quantities of the system follow immediately by straightforward manipulations.

In the following it will be convenient to deal not with $\mathcal{Z}$ directly, but with the quantities $\bar{n}_r$, the *mean occupation numbers* of the single-particle states r.

The reason is that these quantities are much simpler than the grand partition function itself.

From the mean occupation numbers we will be able to directly compute the thermodynamic variables, without the need to compute $\mathcal{Z}$ explicitly.

The mean occupation number of a particular single-particle state r^* is obtained by averaging n_{r*} over the Gibbs distribution:

$$\bar{n}_{r^*} = \frac{1}{\mathcal{Z}} \sum_R n_{r^*} e^{-(\alpha N_R + \beta E_R)}$$

where the sum is over all states accessible to the entire system.

Writing this out more explicitly, we have

$$\bar{n}_{r^*} = \frac{1}{\mathcal{Z}} \sum_{\{n_r\}} n_{r^*} \exp\Big\{-\big[\alpha(n_0 + \cdots + n_{r^*} + \cdots) + \beta(n_0\varepsilon_0 + \cdots + n_{r^*}\varepsilon_{r^*} + \cdots)\big]\Big\}$$

But we see that the sum on the right-hand side is equal to $-\beta^{-1}\partial\,\mathcal{Z}\,/\partial\varepsilon_{r*}$.

We therefore have

$$\boxed{\bar{n}_r = -\frac{1}{\beta}\frac{\partial \ln \mathcal{Z}}{\partial \varepsilon_r}} \tag{10.13}$$

for any single-particle state r.

This equation shows that the mean occupation numbers can all be obtained from the grand partition function.

We will show the converse below.

10.1.6 Thermodynamic quantities

We now show how such thermodynamic quantities as the mean number of particles N, the mean energy E, and the pressure P can all be obtained directly in terms of the mean occupation numbers, without the need to first compute $\mathcal{Z}$.

Summing over all the mean occupation numbers should obviously return N, the mean number of particles.

This is easily verified if we rewrite Eq. (10.2) in the form

$$N = -\frac{1}{\mathcal{Z}}\frac{\partial}{\partial\alpha}\sum_{\{n_r\}}\exp\Big\{-\big[\alpha(n_0+n_1+\cdots)+\beta(n_0\varepsilon_0+n_1\varepsilon_1+\cdots)\big]\Big\}$$

Taking the derivative over α brings down the factor $-(n_0 + n_1 + \cdots)$, and we obtain $N = \bar{n}_0 + \bar{n}_1 + \cdots$.

In other words,

$$\boxed{N = \sum_r \bar{n}_r} \tag{10.14}$$

which confirms our initial expectation.

A similar calculation reveals that the mean energy E is obtained by summing $\bar{n}_r\,\varepsilon_r$ over all possible single-particle states:

$$\boxed{E = \sum_r \bar{n}_r\,\varepsilon_r} \tag{10.15}$$

This result also appears quite reasonable.

The pressure follows easily.

We recall that $P = kT\partial ln\ \mathcal{Z}\ /\partial V$.

Since the grand partition function depends on V only through the energy eigenvalues, $\varepsilon_r \propto V^{-2/3}$, we have that

$$P = kT \sum_r \frac{\partial \ln \mathcal{Z}}{\partial \varepsilon_r} \frac{\partial \varepsilon_r}{\partial V}$$

where we have applied the chain rule of differential calculus.

Now, the first factor within the sum is equal to $-\beta \bar{n}_r$.

The second factor is equal to $-2\varepsilon_r/3V$, and the previous equation becomes $P = (2kT\,\beta/3V)\ \sum_r \bar{n}_r \varepsilon_r$.

Using Eq. (10.15), we arrive at our final result,

$$\boxed{P = \frac{2E}{3V}} \tag{10.16}$$

Equations (10.11)–(10.16) are quite general: they are valid for any system of noninteracting particles, whether they are bosons or fermions.

These equations, however, are still rather formal, and we cannot yet use them in practical calculations.

To get there is the purpose of the following two sections.

10.2 Bose-Einstein statistics

10.2.1 Grand partition function

We now specialize to an ideal gas consisting of boson particles, and we want to calculate the grand partition function for this gas.

We therefore return to Eq. (10.1), and observe that $\mathcal{Z}$ is given by an infinite product of factors of the form

$$\sum_{n_r} e^{-(\alpha+\beta\varepsilon_r)n_r} = \sum_{n_r=0}^{\infty} \left[e^{-(\alpha+\beta\varepsilon_r)}\right]^{n_r}$$

where we have indicated that for bosons, there is no restriction on the value of the occupation numbers.

This sum is easily evaluated, if we recall that for $x < 1$ $\sum_{n=0}^{\infty} x^n = 1/(1-x)$.

For $x > 1$, on the other hand, the sum does not converge.

The grand partition function is therefore defined if and only if $\exp[-(\alpha+\beta\varepsilon_r)] < 1$, which implies $-\alpha < \beta\varepsilon_r$.

This condition must be valid for all single-particle states r, and since $\varepsilon_0 = 0$ is the smallest energy eigenvalue, we conclude that $\mathcal{Z}$ is defined if and only if $-\alpha \equiv \mu/kT < 0$.

This means that *the chemical potential of an ideal boson gas is necessarily negative*.

Evaluating the sums, we find

$$\mathcal{Z} = \frac{1}{1-e^{-(\alpha+\beta\varepsilon_0)}} \frac{1}{1-e^{-(\alpha+\beta\varepsilon_1)}} \frac{1}{1-e^{-(\alpha+\beta\varepsilon_2)}} \times \cdots$$

or

$$\ln \mathcal{Z} = -\ln\left[1-e^{-(\alpha+\beta\varepsilon_0)}\right] - \ln\left[1-e^{-(\alpha+\beta\varepsilon_1)}\right] - \ln\left[1-e^{-(\alpha+\beta\varepsilon_2)}\right] + \cdots$$

In other words,

$$\ln \mathcal{Z} = -\sum_r \ln\left[1-e^{-(\alpha+\beta\varepsilon_r)}\right]$$

We will express $\ln \mathcal{Z}$ in a tidier form in a moment.

10.2.2 Mean occupation numbers

As Eq. (10.13) shows, the mean occupation number of the single-particle state r^* is obtained by differentiating $\ln \mathcal{Z}$ with respect to the corresponding eigenvalue ε_{r*} .

From our previous expression for $\ln \mathcal{Z}$, we see that the only term involving this eigenvalue is

$$-\ln\left[1 - e^{-(\alpha+\beta\varepsilon_{r^*})}\right]$$

and differentiating yields

$$-\frac{\beta e^{-(\alpha+\beta\varepsilon_{r^*})}}{1 - e^{-(\alpha+\beta\varepsilon_{r^*})}}$$

According to Eq. (10.13), we must now multiply by $-1/\beta$.

Rearranging slightly, we arrive at the following result: For an ideal gas of bosons, the mean occupation number of the single-particle state r is given by

$$\boxed{\bar{n}_r = \frac{1}{e^{\alpha+\beta\varepsilon_r} - 1}} \tag{10.17}$$

This distribution of mean occupation numbers is called the **Bose-Einstein distribution**, and bosons are therefore said to obey the **Bose-Einstein statistics**.

Contrary to our previous results which were quite formal, Eq. (10.17) provides a very concrete relation between $\bar{n}_r$ and ε_r, for given values of α and β.

Equation (10.17) can be inverted to give

$$e^{-(\alpha+\beta\varepsilon_r)} = \frac{\bar{n}_r}{1+\bar{n}_r}$$

This means that we can express the logarithm of the grand partition function as

$$\boxed{\ln \mathcal{Z} = \sum_r \ln(1+\bar{n}_r)} \tag{10.18}$$

This proves a statement made earlier: the grand partition function can be expressed in terms of the mean occupation numbers, as can the mean number of particles, the mean energy, and the pressure.

10.3 Fermi-Dirac statistics

10.3.1 Grand partition function

We now turn to the case of an ideal gas consisting of fermion particles.

Our first task is to evaluate the grand partition function, which is given by an infinite product of factors of the form

$$\sum_{n_r} e^{-(\alpha+\beta\varepsilon_r)n_r} = \sum_{n_r=0}^{1} \left[e^{-(\alpha+\beta\varepsilon_r)}\right]^{n_r}$$

where we have indicated that for fermions, the occupation numbers can take the values 0 and 1 only.

This sum is trivial, and we find that the grand partition function is given by

$$\mathcal{Z} = \left[1 + e^{-(\alpha+\beta\varepsilon_0)}\right]\left[1 + e^{-(\alpha+\beta\varepsilon_1)}\right]\left[1 + e^{-(\alpha+\beta\varepsilon_2)}\right] \times \cdots$$

or

$$\ln \mathcal{Z} = \ln\left[1 + e^{-(\alpha+\beta\varepsilon_0)}\right] + \ln\left[1 + e^{-(\alpha+\beta\varepsilon_1)}\right] + \ln\left[1 + e^{-(\alpha+\beta\varepsilon_2)}\right] + \cdots$$

In other words

$$\ln \mathcal{Z} = \sum_r \ln\left[1 + e^{-(\alpha+\beta\varepsilon_r)}\right]$$

We will rewrite this in terms of the mean occupation numbers in a moment.

10.3.2 Mean occupation numbers

Calculation of the mean occupation numbers proceeds exactly as in the previous section, starting from Eq. (10.13).

The end result is the following: For an ideal gas of fermions, the mean occupation number of the single-particle state *r* is given by

$$\boxed{\bar{n}_r = \frac{1}{e^{\alpha+\beta\varepsilon_r}+1}} \tag{10.19}$$

This distribution of mean occupation numbers is called the ***Fermi-Dirac distribution***, and fermions are therefore said to obey the **Fermi-Dirac statistics**.

Again, Eq. (10.19) provides an explicit relation between $\bar{n}_r$ and ε_r, for given values of α and β.

Equation (10.19) can be inverted to give

$$e^{-(\alpha+\beta\varepsilon_r)} = \frac{\bar{n}_r}{1-\bar{n}_r}$$

This means that we can express the logarithm of the grand partition function as

$$\boxed{\ln \mathcal{Z} = -\sum_r \ln(1-\bar{n}_r)} \tag{10.20}$$

It is worth spending of moment to appreciate the similarities and differences between Eqs. (10.17) and (10.19), and between Eqs. (10.18) and (10.20).

10.4 Summary

We have now completed our reformulation of the statistical mechanics of ideal gases, and it is time to summarize the major ideas and results.

In this formulation, the focus shifts from the states of the entire system, which are difficult to characterize, to the single-particle states r, which can be fully characterized with just four quantum numbers.

The link between these two very different sets of states is provided by the set of all occupation numbers, $\{n_{\mathrm{r}}\}$.

Indeed, because identical particles are indistinguishable in quantum mechanics, all that can be said about the state of the entire system is how many particles occupy each and everyone of the single-particle states; this information is provided by the occupation numbers.

We recall that n_{r} is equal to the number of particles that occupy the single-particle state r.

Our ideal gas is not isolated.

Instead, it interacts with a reservoir, in such a way that both energy and particles can be exchanged.

Our system therefore adopts the Gibbs distribution of probabilities.

Because of this interaction, the number of particles occupying a specific state r does not stay constant, but keeps on fluctuating.

We are therefore lead to consider the average value of the occupation numbers, which is given by

$$\boxed{\bar{n}_r = \frac{1}{e^{\alpha+\beta\varepsilon_r} \mp 1}} \tag{10.21}$$

where the upper sign applies to bosons, while the lower sign applies to fermions.

Here, α and β are parameters related to the temperature T and the chemical potential μ of the reservoir: $\alpha = -\mu/kT$ and $\beta = 1/kT$.

As we have shown, α is always positive for bosons, but there is no such restriction for fermions.

From the mean occupation numbers we may calculate all thermodynamic quantities of interest.

Using the same sign convention as above, the logarithm of the grand partition function is given by

$$\boxed{\ln \mathcal{Z} = \pm \sum_r \ln(1 \pm \bar{n}_r)} \tag{10.22}$$

The mean total number of particles is

$$\boxed{N = \sum_r \bar{n}_r} \tag{10.23}$$

and the mean total energy is given by

$$\boxed{E = \sum_r \bar{n}_r \, \varepsilon_r} \tag{10.24}$$

In these expressions, the sums are over all the single-particle states r, and ε_r denotes the corresponding energy eigenvalues.

It is precisely this property — that the sums are over the single-particle states and not the states of the entire system — that makes this reformulation so useful.

The preceding four equations are the fundamental equations of quantum statistics.

From these we can derive the following expression for the entropy:

$$\boxed{S = k \sum_r \left[\pm (1 \pm \bar{n}_r) \ln(1 \pm \bar{n}_r) - \bar{n}_r \ln \bar{n}_r \right]} \tag{10.25}$$

We also recall that the pressure can be obtained directly from the general relation

$$\boxed{PV = \frac{2}{3} E} \tag{10.26}$$

which holds both for bosons and fermions.

The present formulation of quantum statistics therefore has the virtue of being very economical.

Its other important virtue, that it is easy to use in practical computations, will reveal itself in the following sections.

10.5 Classical limit

As our first application of the methods of quantum statistics, it is perhaps a good idea to make sure that we can at least reproduce the familiar results pertaining to the classical limit.

After taking this tentative first step, we will be in a good position to generalize our discussion to more interesting situations.

10.5.1 Classical approximation

We must first specify what we mean by the classical limit.

The classical picture of an ideal gas is that of little billiard balls moving around in a box.

In this classical picture, there is no difficulty of principle in specifying the position of each particle with arbitrary accuracy.

The quantum picture, of course, is drastically different.

Here, the position of each particle cannot be specified with arbitrary accuracy, because of the Heisenberg uncertainty principle.

The best we can do is to picture our particles as fuzzy wave packets moving around in the box.

The uncertainty Δx in the position of a given wave packet is related to Δp, the uncertainty in its momentum, by the Heisenberg relation

$$\Delta x \, \Delta p \sim \hbar$$

The fuzzy nature of the gas particles never disappears completely, even in the classical limit.

What happens, however, is that as we approach the classical description, the wave packets become increasingly localized.

In other words, the wave-packet extension Δx becomes very small compared with the mean distance between gas particles.

Because the mean volume per particle is V/N, we have that the condition

$$\Delta x \ll (V/N)^{1/3}$$

describes the classical limit.

We now want to convert this condition into something more useful.

For this purpose, we use the Heisenberg relation to convert Δx into $\hbar/\Delta p$.

Now, Δp is at most of the order of p itself, which is approximately equal to $\sqrt{2m(E/N)}$, where E/N is the mean energy per particle, which is of the order of kT.

Gathering the results, and discarding meaningless numerical coefficients, we obtain

$$\Delta x \leq \frac{\hbar}{\sqrt{mkT}}.$$

Substituting this into the preceding equation, we find that the classical limit corresponds to the mathematical statement

$$\lambda^3 \ll V/N,$$

where $\lambda \equiv \Delta x$ is the length scale associated with the fuzziness of our particles:

$$\boxed{\lambda = \frac{2\pi\hbar}{\sqrt{2\pi mkT}}} \tag{10.27}$$

For convenience, numerical factors have been inserted in this expression for λ.

We now express the condition $\lambda^3 \ll V/N$ in a different and more useful form.

In these terms, our earlier work says that the chemical potential of a classical gas is given by

$$\mu = -kT \ln\left[\frac{V/N}{\lambda^3}\right].$$

This tells us immediately that the condition for the classical limit can be written as

$$\boxed{e^{\alpha} \gg 1 \qquad \leftrightarrow \qquad \text{classical limit}} \tag{10.28}$$

This is the form that we will use in our calculations.

In fact, we know already that this condition is true: a helium gas at atmospheric pressure and at room temperature has $\alpha = -\mu/kT \simeq 13$, which yields $e^{\alpha} \simeq 10^5$.

This is indeed much larger than 1!

Notice that we have cheated a little.

We have used a previously derived expression for the chemical potential to give a precise formulation of the classical approximation, which we will now use to calculate the thermodynamic quantities of a classical gas, including the chemical potential!

This is not as bad as it seems.

We may now forget that the chemical potential is given by the expression above.

All we shall need is Eq. (10.28), which merely tells us that in our calculations, we can treat e^{α} as a very large quantity.

This is far less information than the detailed knowledge of how μ varies with temperature T and number density N/V .

10.5.2 Evaluation of α

What better proof of this statement than a rederivation our previous expression for the chemical potential as earlier?

In the classical limit, where e^{α}, and therefore also $e^{\alpha+\beta\varepsilon r}$, are much larger than one, Eq. (10.21) reduces to

$$\boxed{\bar{n}_r = e^{-(\alpha+\beta\varepsilon_r)}} \tag{10.29}$$

which implies that the mean occupation numbers are all very small.

This means that in the classical limit, the particles are distributed over an extremely large number of single-particle states.

Equation (10.29) also shows that in the classical limit, there is no distinction between bosons and fermions.

We now use Eq. (10.29) to calculate N, the mean number of particles within the gas:

$$N = \sum_r \bar{n}_r = e^{-\alpha} \sum_r e^{-\beta\varepsilon_r}$$

By inverting this relation, we will be able to express the chemical potential $\mu = -kT\alpha$ in terms N, which is exactly what we want.

But we must first evaluate the sum $\sum_r e^{-\beta\varepsilon_r}$.

This is nothing but the partition function of a single particle confined to a box of volume V.

To evaluate this, we first observe that a number $\Omega(\varepsilon_r) \neq 1$ of single-particle states r will share the same energy eigenvalue ε_r.

The quantity $\Omega(\varepsilon_r)$ is of course the statistical weight for a single particle in a box.

The sum over single-particle states can therefore be written as

$$\sum_r = \sum_{\varepsilon_r} \Omega(\varepsilon_r)$$

where we are now summing over the distinct energy eigenvalues.

Because these eigenvalues are quantized, this is a discrete sum.

But because the vast majority of the energy eigenvalues are so extremely close to each other, we may in fact treat ε_r as a continuous variable.

We shall therefore replace the discrete sum by an integral,

$$\boxed{\sum_r \to \int_0^\infty \omega(\varepsilon)\, d\varepsilon} \qquad (10.30)$$

by re-introducing the density of states $\omega(\varepsilon)$.

This is defined so that $\omega(\varepsilon)\, d\varepsilon$ is the number of single-particle states with energy eigenvalues in the interval between ε and $\varepsilon + d\varepsilon$.

The density of states has already been calculated earlier:

$$\boxed{\omega(\varepsilon) = g\,\frac{2\pi V}{(2\pi\hbar)^3}\,(2m)^{3/2}\,\varepsilon^{1/2} = \frac{2gV}{\sqrt{\pi}\lambda^3}\,\beta^{3/2}\,\varepsilon^{1/2}} \tag{10.31}$$

where Eq. (10.27) was used to obtain the second expression.

The factor g is new.

We need to insert it to account for the number of internal states associated with the particle's spin, something which we had neglected to do before.

The number g is equal to the number of values that m_s can take between the bounds $-s$ and s.

Clearly, this is

$$\boxed{g = 2s + 1} \tag{10.32}$$

where s is the particle's spin

We are now ready to calculate the chemical potential.

Gathering our previous results, we have

$$\begin{aligned} N &= e^{-\alpha} \int_0^\infty e^{-\beta\varepsilon} \omega(\varepsilon)\, d\varepsilon \\ &= \frac{2gV}{\sqrt{\pi}\lambda^3}\, e^{-\alpha}\, \beta^{3/2} \int_0^\infty e^{-\beta\varepsilon} \varepsilon^{1/2}\, d\varepsilon \\ &= \frac{2gV}{\sqrt{\pi}\lambda^3}\, e^{-\alpha} \int_0^\infty e^{-x} x^{1/2}\, dx, \end{aligned}$$

where, in the last line, we have put $x \equiv \beta\varepsilon$.

We have encountered integrals of this sort before, and in fact, this integral is .

We have therefore obtained the simple result $N = gV\, e^{-\alpha}/\lambda^3$, or $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$.

$$\boxed{e^{-\alpha} = \frac{N\lambda^3}{gV}} \tag{10.33}$$

This means that $\mu = -kT\, ln(gV/N\lambda^3)$, and apart from the new factor g this is exactly what we had written.

10.5.3 Other thermodynamic quantities

The same method can be used to calculate other quantities of interest.

Let us start with the mean energy, which according to Eqs. (10.24), (10.29), and (10.30), is given by

$$E = \sum_r \bar{n}_r \, \varepsilon_r = e^{-\alpha} \sum_r \varepsilon_r \, e^{-\beta \varepsilon_r} = e^{-\alpha} \int_0^\infty \varepsilon e^{-\beta \varepsilon} \omega(\varepsilon) \, d\varepsilon$$

Substituting Eqs. (10.21) and (10.23), this becomes

$$E = \frac{2N}{\sqrt{\pi} \beta} \int_0^\infty e^{-x} x^{3/2} \, dx,$$

where $x \equiv \beta \varepsilon$.

The integral evaluates to $\Gamma(\frac{5}{2}) = \frac{3}{4}\sqrt{\pi}$, and we arrive at the familiar result

$$\boxed{E = \frac{3}{2} N k T} \tag{10.34}$$

From this and Eq. (10.26) we immediately obtain the classical limit for the equation of state of an ideal gas:

$$\boxed{PV = NkT} \tag{10.35}$$

Let us see if we can also compute the entropy.

From Eq. (10.35) and the fact that the occupation numbers are all much smaller than unity — we use the approximation $ln(1 + \varepsilon) \simeq \varepsilon$ when ε is small — we obtain

$$S = k\sum_r (\bar{n}_r - \bar{n}_r \ln \bar{n}_r).$$

But Eq. (10.29) also gives $\ln \bar{n}_r = -\alpha - \beta\varepsilon$, so the entropy becomes $S = k(N + \alpha N + \beta E) = Nk(\alpha + {}^{5}/_{2})$. SubstitutingEq.(10.33),we arrive at

$$\boxed{S = Nk \ln\left(\frac{gV}{N\lambda^3}\right) + \frac{5Nk}{2}} \tag{10.36}$$

This result differs from earlier only by the inclusion of the factor g.

10.5.4 Conclusion

We see that we are able to reproduce all of the familiar classical results.

It should be appreciated that calculations were handled quite easily and efficiently in our current formulation of quantum statistics.

The major advantage of this new formulation is that it will allow us to go beyond the classical limit.

We will be able to predict how quantum effects affect such relations as $E = {}^{3}/_{2}NkT$ and $PV = NkT$, and this will bring to light interesting new phenomena.

It will be useful, for what comes next, to introduce a terminology to characterize the importance of quantum effects in an ideal gas.

We shall use term *nondegenerate gases* for those which behave essentially classically.

The criterion for this was derived in subsection a: it is $\lambda \ll (V/N)^{1/3}$, or $e^{\alpha} \gg 1$.

We shall use term *slightly degenerate gases* for those which quantum effects are still quite small, but not negligible.

The criterion for this is $\lambda < (V/N)^{1/3}$, or $e^{\alpha} > 1$.

Finally, we shall use term *highly degenerate gases* for those which quantum effects are extremely important.

The criterion for this is $\lambda \geq (V/N)^{1/3}$, or $e^{\alpha} \leq 1$.

10.6 Slightly degenerate ideal gases

We now generalize the discussion of the preceding section to ideal gases for which quantum effects are becoming important.

As was mentioned before, we refer to this situation as the slightly degenerate limit, in which e^{α} is still larger than one, but only moderately so.

Thus, we want to compute the leading order corrections to the classical results derived in the preceding section.

10.6.1 Evaluation of α

We first recall that for bosons (upper sign) and fermions (lower sign), the mean occupation numbers are given by

$$\bar{n}_r = \frac{1}{e^{\alpha+\beta\varepsilon_r} \mp 1}$$

This is an exact result.

But if e^{α} is fairly large, then the mean occupation numbers are given approximately by

$$\boxed{\bar{n}_r = e^{-(\alpha+\beta\varepsilon_r)} \pm e^{-2(\alpha+\beta\varepsilon_r)} + \cdots} \tag{10.37}$$

where the dots represent terms of order $e^{-3\alpha}$ and smaller.

These will be neglected throughout the calculation

We will first use Eq. (10.37) to express $e^{-\alpha}$ in terms of N, the mean number of particles.

We start with the relation $N = \sum_r \bar{n}_r$, in which we substitute Eq. (10.37).

After converting the discrete sums into integrals, we obtain

$$N = e^{-\alpha} \int_0^\infty e^{-\beta\varepsilon} \omega(\varepsilon)\, d\varepsilon \pm e^{-2\alpha} \int_0^\infty e^{-2\beta\varepsilon} \omega(\varepsilon)\, d\varepsilon + \cdots$$

We already know from earlier that the first integral gives gV/λ^3.

To evaluate the second integral, we substitute Eq. (10.31) and find

$$\int_0^\infty e^{-2\beta\varepsilon}\omega(\varepsilon)\,d\varepsilon = \frac{2gV}{\sqrt{\pi}\lambda^3}\,\beta^{3/2}(2\beta)^{-3/2}\int_0^\infty e^{-x}x^{1/2}\,dx$$

where $x \equiv 2\beta\varepsilon$. The integral gives $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$, and gathering the results, we arrive at

$$N = \frac{gV}{\lambda^3}\,e^{-\alpha}\left(1 \pm 2^{-3/2}e^{-\alpha} + \cdots\right)$$

It is not difficult to invert this relation between N and $e^{-\alpha}$, especially if we keep in mind that all terms of order $e^{-3\alpha}$ and smaller must be neglected.

We obtain

$$\boxed{e^{-\alpha} = \frac{N\lambda^3}{gV}\left(1 \mp 2^{-3/2}\frac{N\lambda^3}{gV} + \cdots\right)} \qquad (10.38)$$

This gives the chemical potential of a slightly degenerate ideal gas.

Equation (10.38) shows that for bosons, quantum effects produce a slight decrease in $e^{-\alpha}$ with respect to the classical result.

For fermions, quantum effects produce a slight increase.

The physical interpretation of this result will come more easily when we consider the other thermodynamic quantities.

10.6.2 Energy and pressure

The mean energy of the gas is calculated as ealier.

After substituting Eq. (10.37) into $E = \sum_r \bar{n}_r \, \varepsilon_r$ and converting into integrals, we obtain

$$E = e^{-\alpha} \int_0^\infty \varepsilon e^{-\beta\varepsilon} \omega(\varepsilon) \, d\varepsilon \pm e^{-2\alpha} \int_0^\infty \varepsilon e^{-2\beta\varepsilon} \omega(\varepsilon) \, d\varepsilon + \cdots$$

After using Eq. (10.31), the first integral gives $3gV/2\lambda^3\beta$ as before, while the second integral is

$$\int_0^\infty \varepsilon e^{-2\beta\varepsilon} \omega(\varepsilon) \, d\varepsilon = \frac{2gV}{\sqrt{\pi}\lambda^3} \beta^{3/2} (2\beta)^{-5/2} \int_0^\infty e^{-x} x^{3/2} \, dx,$$

where $x \equiv 2\beta$.

The integral gives $\Gamma(\frac{5}{2}) = \frac{3}{4}\sqrt{\pi},$, and gathering the results, we obtain

$$E = \frac{3}{2} kT \frac{gV}{\lambda^3} e^{-\alpha} \left(1 \pm 2^{-5/2} e^{-\alpha} + \cdots\right)$$

Although this is a perfectly valid expression for the energy, we would prefer to express E in terms of N.

Using Eq. (10.38) gives $$E = \frac{3}{2} kT \, N \left(1 \mp 2^{-3/2} \frac{N\lambda^3}{gV} + \cdots\right) \left[1 \pm 2^{-5/2} \frac{N\lambda^3}{gV} (1 \mp \cdots)\right]$$

Recalling that we must neglect all correction terms beyond $N\lambda^3/gV$, this simplifies to

$$\boxed{E = \frac{3}{2} NkT\left(1 \mp 2^{-5/2} \frac{N\lambda^3}{gV} + \cdots\right)} \tag{10.39}$$

This result shows that the classical relation E = $^{3}/_{2}$NkT is modified when quantum effects are taken into account.

Equation (10.39) further shows that the energy is slightly decreased for bosons, while it is slightly increased for fermions.

Equations (10.26) and (10.39) give the equation of state for a slightly degenerate ideal gas:

$$\boxed{PV = NkT\left(1 \mp 2^{-5/2} \frac{N\lambda^3}{gV} + \cdots\right)} \tag{10.40}$$

This result is quite interesting, because it shows that the classical relation $PV = NkT$ is modified by the quantum nature of the gas.

We see that the pressure is lower for bosons, and higher for fermions.

It is worth spending some time trying to understand this result.

Why is the pressure higher for fermions?

The answer comes from the fact that fermions do not like to occupy the same state, so that they will resist compression more vigorously than classical particles.

The extra pressure provided by quantum effects can therefore be understood as being there to help enforce Pauli's exclusion principle.

This effect is most dramatic in the highly degenerate limit, to which we turn in the next three sections

Why is the pressure lower for bosons?

This is not as easy to understand.

Bosons, of course, do not obey an exclusion principle, and this explains why the pressure of a boson gas must be lower than that of a fermion gas.

This, however, does not explain why the pressure must be lower than the classical result.

The answer lies with the fact that unlike fermions and classical particles, bosons prefer to be close together, so that they will resist compression less vigorously than fermions and classical particles.

This effect is most dramatic in the highly degenerate limit, which we will discuss later.

10.6.3 Higher-order corrections

Higher-order corrections to the equation of state can be calculated in the same way, by simply keeping more terms in our expansions in powers of $e^{-\alpha}$, or in powers of $N\lambda^3/gV$.

We will not go through this exercise, but it should at least be clear that the end result will be an expansion of the form

$$\begin{aligned} PV &= NkT\left[1 + a_1\left(\frac{N\lambda^3}{gV}\right) + a_2\left(\frac{N\lambda^3}{gV}\right)^2 + \cdots\right] \\ &= NkT\sum_{n=0}^{\infty} a_n\left(\frac{N\lambda^3}{gV}\right)^n . \end{aligned}$$

Such an expression for the equation of state of a gas is called a *virial expansion*, and the coefficients an are called *virial coefficients*.

The first few are given by $a_0 = 1$, $a_1 = \mp 2^{-5/2} \simeq \mp 0.17678$, $a_2 \simeq -0.00330$, and $a_3 \simeq \mp 0.00011$.

In this section we have calculated only the first-order correction.

10.7 Highly degenerate Fermi gas. I: Zero temperature

We have seen in the preceding section that quantum effects modify the behavior of the thermodynamic quantities of an ideal gas, with respect to their classical behavior.

For example, the pressure of a slightly degenerate Fermi gas was found to be slightly larger than NkT/V , while it was found to be smaller than this for a Bose gas.

In this and the following sections we will explore the depths of quantum statistics by going to the highly degenerate limit, where quantum effects truly dominate.

By doing so we will encounter drastic deviations from the classical predictions, and interesting new phenomena will be discovered.

We recall from earlier that the highly degenerate limit corresponds to a situation in which the quantum wave packets representing the gas particles are no longer well separated from each other.

The opposite situation, in which the wave packets are well separated, is the classical limit.

We also recall that the typical extension of the wave packets is given by λ, as defined by Eq. (10.27).

What is noteworthy here is that λ is inversely proportional to the square root of the temperature: $\lambda \propto T^{-1/2}$.

The wave packets therefore increase in size as the temperature decreases, and there comes a point where λ becomes comparable with $(V/N)^{1/3}$, the mean distance between wave packets.

At this point the gas can no longer be pictured in terms of individual particles: the wave packets all overlap with each other.

In this situation, clearly, quantum effects play a very important role.

We shall begin our discussion of the highly degenerate limit by examining a fermion gas at zero temperature, where things are quite extreme.

10.7.1 Fermi gas at $T = 0$

It should be quite clear that at $T = 0$, a system of fermions will be found in its state of lowest energy, or ground state.

This, however, does not mean that each particle will be found in a single-particle state of zero energy.

Indeed, Pauli's exclusion principle prevents this from happening:

Once the first g particles have been put in the single-particle states of zero energy, additional particles must be put in the excited states, until all N particles are accounted for.

Obviously then, the state of the entire system will not be a state of zero energy, although it is still the state of lowest energy.

Following these considerations, it is easy to predict what values the occupation numbers will take at zero temperature.

Because the first particle is placed in the state $r = 0$, we automatically have $n_0 = 1$.

The second particle, on the other hand, is placed in the state $r = 1$, so we have $n_1 = 1$.

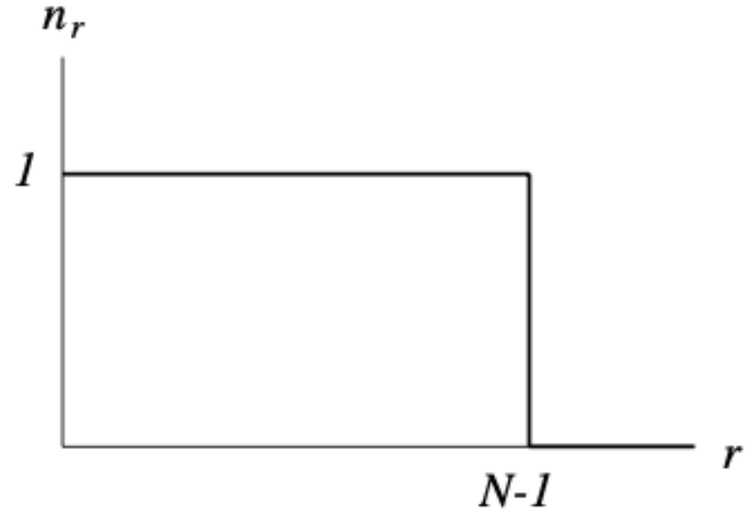
Proceeding similarly for all N particles, we find that every occupation number up to n_{N-1} is equal to 1.

Because we now have accounted for all the particles, all other occupation numbers must be zero.

So we have

$$n_r = \begin{cases} 1 & r \leq N-1 \\ 0 & r \geq N \end{cases}$$

The occupation numbers behave as a step function.



The energy of the last occupied single-particle state ($r = N - 1$) is called the *Fermi energy*, and is denoted ε_F.

This step-function behavior should also be apparent in the mean occupation numbers, which for a Fermi gas are given by

$$\bar{n}_r = \frac{1}{e^{(\varepsilon_r - \mu)/kT} + 1}.$$

In the limit $T \to 0$, we have to pay close attention to the sign of $\varepsilon_r - \mu$.

Suppose first that $\varepsilon_r < \mu$.

In this case $e^{(\varepsilon r - \mu)/kT}$ is an extremely small number, and we find that $\bar{n}_r = 1$.

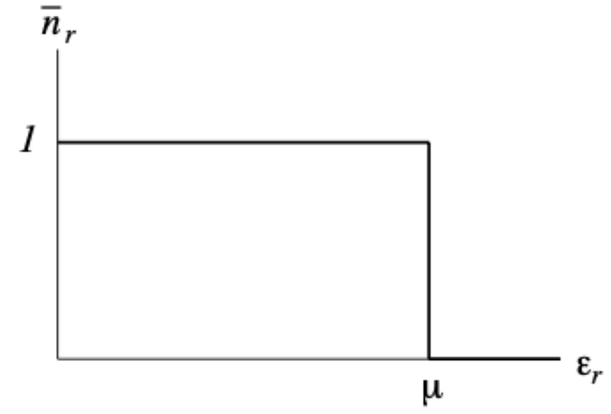
Suppose now that $\varepsilon_r > \mu$.

In this case $e^{(\varepsilon r - \mu)/kT}$ is an extremely large number, and we find that $\bar{n}_r = 0$. .

Thus, the mean occupation numbers jump r abruptly from 1 to 0 as ε_r reaches the value μ.

This is clearly the same behavior as what we have found for the occupation numbers themselves, and we may identify the Fermi energy with the chemical potential at zero temperature:

$$\boxed{\varepsilon_F = \lim_{T \to 0} \mu(T)} \qquad (10.41)$$



10.7.2 Evaluation of the Fermi energy

It is not difficult to express the Fermi energy in terms of other properties of the gas, most importantly, its density.

We start from the basic expression for the mean number of particles,

$$N = \sum_r \frac{1}{e^{(\varepsilon_r - \mu)/kT} + 1},$$

which we convert into the integral

$$N = \int_0^\infty \frac{\omega(\varepsilon)\, d\varepsilon}{e^{(\varepsilon-\mu)/kT} + 1}$$

Such an integral would be difficult to evaluate in general.

But we know that the function $1/(e^{(\varepsilon-\mu)/kT} + 1)$ behaves as a step function in the limit $T \to 0$, so the integral becomes simply

$$N = \int_0^{\varepsilon_F} \omega(\varepsilon)\, d\varepsilon$$

This is quite easy to evaluate if we substitute Eq. (10.31).

A few lines of algebra yield

$$\boxed{\varepsilon_F = \frac{(2\pi\hbar)^2}{2m}\left(\frac{3N}{4\pi g V}\right)^{2/3}} \tag{10.42}$$

This result shows that the Fermi energy increases when N/V, the density of the gas, increases.

We will find it useful to rewrite Eq. (10.31) in the form

$$\omega(\varepsilon) = \frac{3N}{2{\varepsilon_F}^{3/2}}\,\varepsilon^{1/2}.$$

10.7.3 Energy and pressure

The mean energy of the gas is easily calculated in a similar way.

We find that the mean energy of a Fermi gas at zero temperature is equal to

$$\boxed{E = \frac{3}{5} N\varepsilon_F} \tag{10.43}$$

From this it follows that the gas also has a pressure given by

$$\boxed{PV = \frac{2}{5} N\varepsilon_F} \tag{10.44}$$

A Fermi gas at zero temperature has pressure!

Clearly, this pressure is not created by the thermal agitation of the gas particles.

This behavior could not happen at $T = 0$.

Instead, the pressure is kindly provided by Professor Pauli: the fermions fight hard not to occupy the same single-particle state, and it is this fight that is responsible for the pressure.

In other words, Pauli's exclusion principle provides all the pressure in a Fermi gas at zero temperature.

We shall see later that this degeneracy pressure is put to good use in nature.

11.7.4 Entropy

Because our Fermi gas is at zero temperature, its entropy is automatically zero:

$$\boxed{S = 0} \tag{10.45}$$

This explains why the chemical potential at zero temperature is the Fermi energy.

Recall the fundamental definition of the chemical potential as an injection energy:

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{S,V}$$

In words, the chemical potential is equal to the change in energy when new particles are added to the system, keeping the entropy constant.

Now, at zero temperature, a particle can be added to the system if and only if its energy is larger than, or equal to, the Fermi energy.

If it is larger than ε_F , then the system's temperature will increase, and so also will its entropy.

This does not correspond to the operational definition of the chemical potential.

If, on the other hand, the particle's energy is equal to the Fermi energy, then the system will stay at $T = 0$, and its entropy will not increase from $S = 0$.

This does correspond to the operational definition of the chemical potential.

We may then conclude that the Fermi energy must be equal to the chemical potential at $T = 0$.

10.8 Highly degenerate Fermi gas. II: Low temperature

We now move away from $T = 0$, and ask how our previous results are affected by finite-temperature effects.

For simplicity, we shall demand that kT be much smaller than the Fermi energy ε.

By so doing, we will still be dealing with a highly degenerate Fermi gas.

10.8.1 Fermi-Dirac integrals

In the following calculations we will encounter integrals of the form

$$I[f] = \int_0^\infty \frac{f(\varepsilon)\, d\varepsilon}{e^{(\varepsilon-\mu)/kT} + 1}$$

where $f(\varepsilon)$ is some function.

We have already recognized the fact that such integrals are difficult to evaluate in general, but that they simplify drastically when $T = 0$.

When T is different from zero but small, *I[f]* can be evaluated approximately.

The result, which is derived in many advanced books, is

$$\boxed{I[f] = \int_0^{\mu} f(\varepsilon)\, d\varepsilon + \frac{\pi^2}{6} f'(\mu)\, (kT)^2 + \cdots} \tag{10.46}$$

where $f' \equiv df/d\varepsilon$.

The derivation of this result is not particularly difficult, but it is quite long and tedious.

This is why we have chosen to omit it.



10.8.2 Chemical potential

We first compute the finite-temperature corrections to the chemical potential μ.

We start once again with the following expression for the mean number of particles:

$$N = \int_0^{\infty} \frac{\omega(\varepsilon)\, d\varepsilon}{e^{(\varepsilon-\mu)/kT} + 1} = \frac{3N}{2{\varepsilon_F}^{3/2}} \int_0^{\infty} \frac{\varepsilon^{1/2}\, d\varepsilon}{e^{(\varepsilon-\mu)/kT} + 1}$$

The integral is $I[\varepsilon^{1/2}]$, and substitution of Eq. (10.36) yields

$$N = \frac{3N}{2{\varepsilon_F}^{3/2}} \left[\frac{2}{3} \mu^{3/2} + \frac{\pi^2}{12} \mu^{-1/2} (kT)^2 + \cdots \right]$$

A little cleaning up then gives

$${\varepsilon_F}^{3/2} = \mu^{3/2} \left[1 + \frac{\pi^2}{8} \left(\frac{kT}{\mu} \right)^2 + \cdots \right]$$

or

$$\varepsilon_F = \mu \left[1 + \frac{\pi^2}{12} \left(\frac{kT}{\mu} \right)^2 + \cdots \right]$$

Finally, by inverting this series we obtain the desired expression for the chemical potential:

$$\boxed{\mu(T) = \varepsilon_F \left[1 - \frac{\pi^2}{12} \left(\frac{kT}{\varepsilon_F} \right)^2 + \cdots \right]} \tag{10.47}$$

We see that finite-temperature effects decrease the chemical potential from its value at $T = 0$.

10.8.3 Energy and pressure

The mean energy is calculated in a similar way.

We start from

$$E = \int_0^\infty \frac{\varepsilon\,\omega(\varepsilon)\,d\varepsilon}{e^{(\varepsilon-\mu)/kT}+1} = \frac{3N}{2{\varepsilon_F}^{3/2}} \int_0^\infty \frac{\varepsilon^{3/2}\,d\varepsilon}{e^{(\varepsilon-\mu)/kT}+1}$$

and we identify the second integral as $I[\varepsilon^{3/2}]$.

Proceeding as before, we obtain

$$E = \frac{3N}{5{\varepsilon_F}^{3/2}}\,\mu^{5/2}\left[1+\frac{5\pi^2}{8}\left(\frac{kT}{\mu}\right)^2+\cdots\right]$$

If we now substitute Eq. (10.47) raised to the 5/2 power, this becomes

$$E = \frac{3N}{5{\varepsilon_F}^{3/2}}\,{\varepsilon_F}^{5/2}\left[1-\frac{5\pi^2}{24}\left(\frac{kT}{\varepsilon_F}\right)^2+\cdots\right]\left[1+\frac{5\pi^2}{8}\left(\frac{kT}{\varepsilon_F}\right)^2+\cdots\right]$$

After multiplication, we obtain

$$\boxed{E = \frac{3}{5}N\varepsilon_F\left[1+\frac{5\pi^2}{12}\left(\frac{kT}{\varepsilon_F}\right)^2+\cdots\right]} \tag{10.48}$$

We see that finite-temperature effects increase the mean energy of the gas.

Not a surprising result!

From Eq. (10.48) we immediately obtain the finite-temperature correction to the pressure,

$$\boxed{PV = \frac{2}{5} N\varepsilon_F \left[1 + \frac{5\pi^2}{12}\left(\frac{kT}{\varepsilon_F}\right)^2 + \cdots\right]} \tag{10.49}$$

which is also increased with respect to the zero-temperature result.

10.8.4 Heat capacity

We conclude this section by computing C_V, the heat capacity at constant volume.

This is defined by $(đQ/dT)_V$, which is clearly equal to $\partial E/\partial T$.

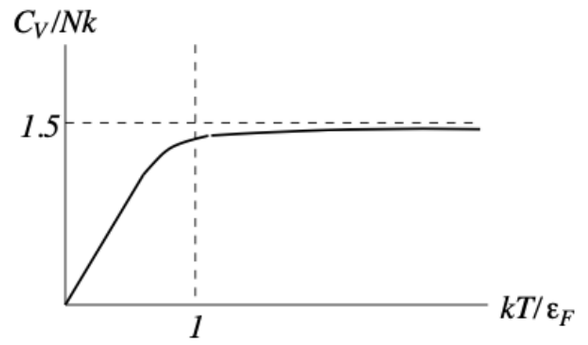
Differentiating Eq. (10.48) with respect to T yields

$$\boxed{C_V = \frac{\pi^2}{2} Nk\left(\frac{kT}{\varepsilon_F}\right)} \tag{10.40}$$

The heat capacity of a Fermi gas is therefore proportional to T in the highly degenerate limit $kT \ll \varepsilon_F$.

At high temperatures, on the other hand, the gas behaves classically, and the heat capacity saturates at the constant value

$$C_V = \frac{3}{2} Nk.$$



10.9 Highly degenerate Fermi gas. III: Applications

Highly degenerate Fermi gases occur quite commonly in nature.

We describe a few examples now.

10.9.1 Conduction electrons in metals

Consider a typical metal, copper.

When copper atoms combine to form a solid, each atom loses one or two electrons.

The lost electrons move (more or less) freely within the metal, and therefore form a gas.

It is this electron gas which is responsible for the thermal and electrical properties of the metal (heat conduction, current conduction, etc.).

We may model the electron gas as an ideal Fermi gas, thereby neglecting all interactions with the copper ions, and all mutual interactions.

We first calculate the Fermi temperature of this gas,

$$T_F \equiv \frac{\varepsilon_F}{k} = \frac{(2\pi\hbar)^2}{2km}\left(\frac{3N}{4\pi g V}\right)^{2/3}.$$

Since we are dealing with an electron gas, $g = 2$ and $m = 9.1\times10^{-28}$ g.

The density of copper is $N/V = 8.4 \times 10^{22}$ electrons/cm^3.

This gives

$$T_F \simeq 80\ 000\ \mathrm{K}$$

which is much higher than room temperature!

In fact, the Fermi temperature is also much higher than copper's melting point, which is approximately 1000 K.

The conclusion is therefore that conduction electrons in metals form a highly degenerate Fermi gas.

The heat capacity of metals has long been measured to grow linearly with temperature.

This experimental fact was at first very puzzling, because before the advent of quantum mechanics, C_V was expected to behave classically, $C_V = 3/2Nk$.

This would indeed be the correct behavior if all the conduction electrons were contributing to the thermal properties of the metal.

But this is not the case: Only those electrons which lie above the Fermi level actually contribute, and this constitutes only a very small fraction of all the electrons.

This fraction is of the order of kT/ε_F , and therefore,

$$C_V \sim Nk\,\frac{kT}{\varepsilon_F}$$

Apart from the numerical factor of $\pi^2/2$, this is exactly the result derived in the preceding section.

10.9.2 White dwarfs

During their normal lives, stars are supported against gravity by radiation pressure, which is provided by the nuclear reactions occurring in the core.

When these reactions cease, the pressure drops, and the core collapses.

The outer layers are then blown away and form what is known as a planetary nebula.

If the core mass is less than a few solar masses, then the endpoint of the collapse is a white dwarf.

The density inside a white dwarf is so high that the electrons form a highly degenerate Fermi gas.

Because the electrons move with relativistic speeds, the laws of relativistic statistical mechanics must be used to work out the details.

This lies somewhat beyond the scope of this class.

It is now the degeneracy pressure, provided by Pauli's exclusion principle, which supports the dead star against gravity.

The first white dwarf ever discovered was Sirius B, an invisible companion to the normal star Sirius A.

Sirius B has a mass approximately equal to the solar mass, but its radius is comparable to that of the Earth: it is only 4800 km.

10.9.3 Neutron stars

White dwarfs can exist only if their mass is less than the Chandrasekhar limit of 1.4 solar masses.

For masses larger that this, the density becomes so high that inverse beta decay occurs,

$$p + e \rightarrow n + \nu$$

the white dwarf becomes neutron rich.

Because the lost electrons are no longer there to provide pressure, the dwarf collapses further to form a *neutron star*.

Neutron stars are supported against gravity by the neutrons, which also form a highly degenerate Fermi gas.

A typical neutron star has a mass comparable to that of the sun, but its radius is of the order of 15 km.

Neutron stars can exist only if their mass is less than approximately 3 solar masses.

For larger masses, the neutron degeneracy pressure is no longer sufficient to support the star against its own weight.

The neutron star then collapses completely, and becomes a *black hole*.

10.10 Highly degenerate Bose gas. I: Zero temperature

Because bosons are not bound by Pauli's exclusion principle, a highly degenerate boson gas behaves very differently from a highly degenerate fermion gas.

This is the topic now, which will conclude our discussion of the quantum statistics of ideal gases.

A Bose gas behaves quite simply at zero temperature: because there is no obstacle in doing so, all N particles just occupy the single-particle zero-energy states.

Because of spin degeneracy, there is a number g of zero-energy states.

Recalling from Eq. (10.11) that the mean occupation numbers of a Bose gas are given by

$$\bar{n}_r = \frac{1}{e^{\alpha+\beta\varepsilon_r} - 1}$$

where $\alpha = -\mu/kT$ and $\beta = 1/kT$, and using the fact that only the states for which $\varepsilon_r = 0$ are occupied at $T = 0$, we have that

$$N = \frac{1}{e^{\alpha} - 1}$$

at zero temperature.

This can be inverted to give

$$\boxed{e^{\alpha} = 1 + \frac{1}{N} \simeq 1} \tag{10.51}$$

Thus, the highly degenerate limit of a Bose gas is characterized by the fact that $\alpha \equiv -\mu/kT$ is a very small number, approximately equal to $1/N$.

It is obvious that if all the particles occupy the zero-energy states, then $E = 0$ at zero temperature.

This immediately implies that $P = 0$, and $T = 0$ also implies $S = 0$.

Not a lot happening in a Bose gas at $T = 0$!

10.11 Highly degenerate Bose gas. II: Low temperatures

The relation $e^{\alpha} \simeq 1$ stays valid at low temperatures, and we will use it to derive the following results.

We are mainly interested in two things: How many particles stay in the zero-energy states when $T \neq 0$, and how does the system's mean energy increase with increasing temperature.

To simplify the terminology, in the following we will refer to the zero-energy states as the single-particle ground state.

10.11.1 Bose-Einstein integrals

Before we begin, we must quote the following mathematical result:

$$\boxed{I_\nu \equiv \int_0^\infty \frac{x^\nu \, dx}{e^x - 1} = \Gamma(\nu+1)\,\zeta(\nu+1)} \tag{10.52}$$

Here ν stands for any number, and Γ(z) is our old friend the Gamma function.

The function ζ(z) is new. It is called the *Riemann zeta function*, and it is defined by

$$\boxed{\zeta(z) = \sum_{k=1}^{\infty} \frac{1}{k^z}} \tag{10.53}$$

By directly evaluating the sum, which converges quite rapidly, it is easy to show that

$$\zeta(\tfrac{3}{2}) \simeq 2.612, \qquad \zeta(\tfrac{5}{2}) \simeq 1.341$$

The integrals $I_{1/2} = \frac{1}{2}\sqrt{\pi}\,\zeta(\frac{3}{2})$ and $I_{3/2} = \frac{3}{4}\sqrt{\pi}\,\zeta(\frac{5}{2})$ will be required in the following calculations.

10.11.2 Number of particles in ground state

We want to calculate N_0, the mean number of particles that stay in the single-particle ground state when $T \neq 0$.

To do so, we go back to the fundamental relation

$$N = \sum_r \bar{n}_r = N_0 + \sum_{\text{excited states}} \bar{n}_r$$

As usual, we replace the sum by an integral, which gives

$$N = N_0 + \int_0^\infty \frac{\omega(\varepsilon)\, d\varepsilon}{e^{\beta\varepsilon} - 1}$$

where we have used the approximation $e^{\alpha} = 1$.

It may appear objectionable to the set the lower limit of integration to zero, as we are supposed to sum over the excited states only.

However, because $\omega(\varepsilon) \propto \varepsilon^{1/2}$, the lower limit actually contributes nothing to the integral, and we make no significant error in setting it to zero.

Substituting Eq. (10.31) for the density of states gives

$$N_0 = N - \frac{2\pi g V}{(2\pi\hbar)^3}\,(2m)^{3/2}\,\beta^{-3/2}\int_0^\infty \frac{x^{1/2}\,dx}{e^x - 1}$$

where $x \equiv \beta\varepsilon$.

The integral is just $I_{1/2}$, and we obtain

$$N_0 = N - \frac{2\pi g V}{(2\pi\hbar)^3}\,(2mkT)^{3/2} \times \frac{1}{2}\sqrt{\pi}\,\zeta(\tfrac{3}{2})$$

Cleaning this up, we arrive at the desired result

$$\boxed{N_0 = N\left[1 - \left(\frac{T}{T_c}\right)^{3/2}\right], \qquad T \le T_c} \tag{10.54}$$

where the constant T_c, called the *condensation temperature*, is given by

$$\boxed{T_c = \frac{(2\pi\hbar)^2}{2\pi mk}\left[\frac{N}{g\zeta(\frac{3}{2})V}\right]^{2/3}} \tag{10.55}$$

Equation (10.54) tells us that as the temperature increases, the mean number of particles in the ground state decreases, until the condensation temperature T_c is reached, at which point all the particles occupy the excited states.

This result certainly makes sense physically, as we expect that increasing the temperature will give some energy to the particles.

What is remarkable, however, is that until $T = T_c$, there is still a macroscopic number of particles that stay in the ground state.

In effect, the thermal energy is not shared by all the particles, but by only a fraction of them.

As was indicated, Eq. (10.54) is valid only for $T \leq T_c$.

Beyond this point, the approximation $e^{\alpha} \simeq 1$ start s breaking down, and this prevents N_0 from going negative, a nonsensical result.

10.11.3 Energy and pressure

We now turn to the evaluation of the mean energy.

We start with the integral

$$E = \int_0^\infty \frac{\varepsilon\, \omega(\varepsilon)\, d\varepsilon}{e^{\beta\varepsilon} - 1}$$

in which we substitute Eq. (10.21).

After using Eq. (10.55) to simplify the resulting expression, we obtain

$$E = \frac{2N}{\sqrt{\pi}\zeta(\frac{3}{2})} \frac{1}{(kT_c)^{3/2}} \beta^{-5/2} \int_0^\infty \frac{x^{3/2}\, dx}{e^x - 1}$$

where $x \equiv \beta\varepsilon$.

The integral is just $I_{3/2}$, and we arrive at

$$\boxed{E = \frac{3\zeta(\frac{5}{2})}{2\zeta(\frac{3}{2})} NkT \left(\frac{T}{T_c}\right)^{3/2}, \qquad T \le T_c} \tag{10.56}$$

This is clearly very different from the classical result, $E = {}^3/_2 NkT$, and also the Fermi-Dirac result (10.48).

From Eqs. (10.26) and (10.56) we immediately obtain an expression for the pressure,

$$\boxed{PV = \frac{\zeta(\frac{5}{2})}{\zeta(\frac{3}{2})} NkT \left(\frac{T}{T_c}\right)^{3/2}, \qquad T \le T_c} \tag{10.57}$$

Again, this is markedly different from the classical equation of state, $PV = NkT$, and also the Fermi-Dirac relation (10.49).

Also interesting is the fact that because $T_c \propto V^{-2/3}$, the right-hand side of Eq. (10.57) is actually proportional to V.

This means that the pressure of a highly degenerate Bose gas is *independent* of volume!

10.11.4 Heat capacity

Finally, we evaluate C_V, the heat capacity at constant volume, by differentiating Eq. (10.56) with respect to T.

This gives

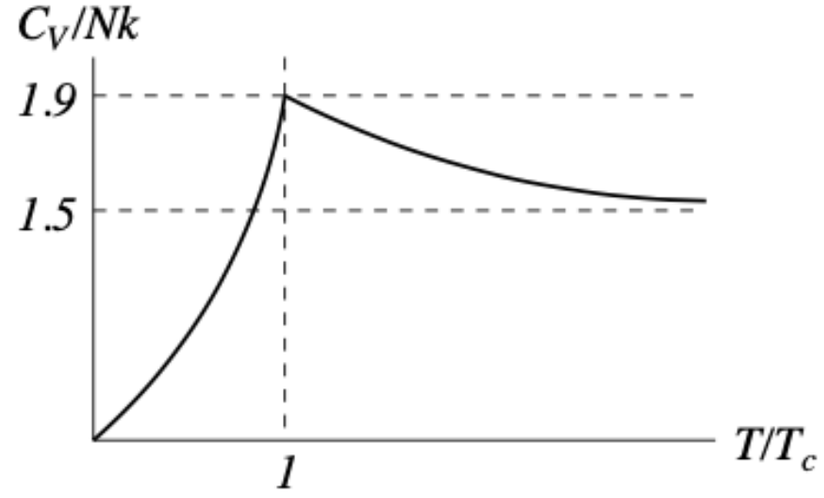
$$\boxed{C_V = \frac{15\zeta(\frac{5}{2})}{4\zeta(\frac{3}{2})} Nk \left(\frac{T}{T_c}\right)^{3/2}, \qquad T \leq T_c} \tag{10.58}$$

The numerical factor is approximately equal to 1.9.

The heat capacity is therefore proportional to $T^{3/2}$ at low temperatures.

At high temperatures, it settles to the classical value $3/2Nk$.

It can be shown that although the heat capacity is continuous at $T = T_c$, its slope $\partial C_V/\partial T$ is not.



This discontinuity of the first derivative signals the occurrence of a *phase transition* at $T = T_c$: An ideal Bose gas cooled below the condensation temperature goes into a state in which a *macroscopic* number of particles have precisely zero energy.

This transition is known as *Bose-Einstein condensation*, and the term "condensate" refers to that part of the system which occupies the single-particle ground state.

10.12.1 Superfluidity

Unless the pressure is extremely high, Helium 4 does not freeze when cooled toward T = 0.

Instead, it undergoes a phase transition — the lambda point — at $T = T_\lambda = 2.17$ K.

Below T_λ, the liquid becomes a *superfluid*, a liquid with no viscosity at all.

This phase transition bears a close resemblance to Bose-Einstein condensation.

First, Helium 4 atoms are bosons, having zero spin.

Second, the lambda-point temperature is very close to the condensation temperature,

$$T_c = \frac{(2\pi\hbar)^2}{2\pi mk}\left[\frac{N}{g\zeta\left(\frac{3}{2}\right)V}\right]^{2/3}$$

with $g = 1$, $m = 6.65\times10^{-24}$ g, and $V/N = 27.6$ cm^3/mole, this yields $T_c \simeq 3.1$ K.

However, there are significant differences between Bose-Einstein condensation and the lambda point.

The most noticeable is in the behavior of the heat capacity near the transition temperature: Whereas the heat capacity climbs to the finite value 1.9Nk in the case of Bose-Einstein condensation, it climbs to infinity in the case of the lambda point.

This λ-shaped behavior of the heat capacity is at the origin of the term “lambda point”.)

These differences can be attributed to the fact that unlike the ideal Bose gas, Helium 4 atoms have interactions that cannot be neglected.

Otherwise they would not form a liquid!

Nevertheless, the highly degenerate Bose gas provides a useful basis for our understanding of superfluidity.

10.12.2 Superconductivity

Some metals undergo a phase transition to a superconducting state as they are cooled below some critical temperature.

This temperature is distinct for each metal, and ranges from approximately 1 K to 100 K.

When the metal is in the superconducting state, it offers no resistance whatsoever to electric currents, and it expels any applied magnetic field (Meissner effect).

Again, in spite of differences due to interactions, the superconducting phase transition has strong similarities with Bose-Einstein condensation.

But where are the bosons?

We already have mentioned that the thermal and electrical properties of metals are due to the conduction electrons, which are fermions.

In the superconducting state, interactions within the metal allow electrons to form bound pairs, called Cooper pairs.

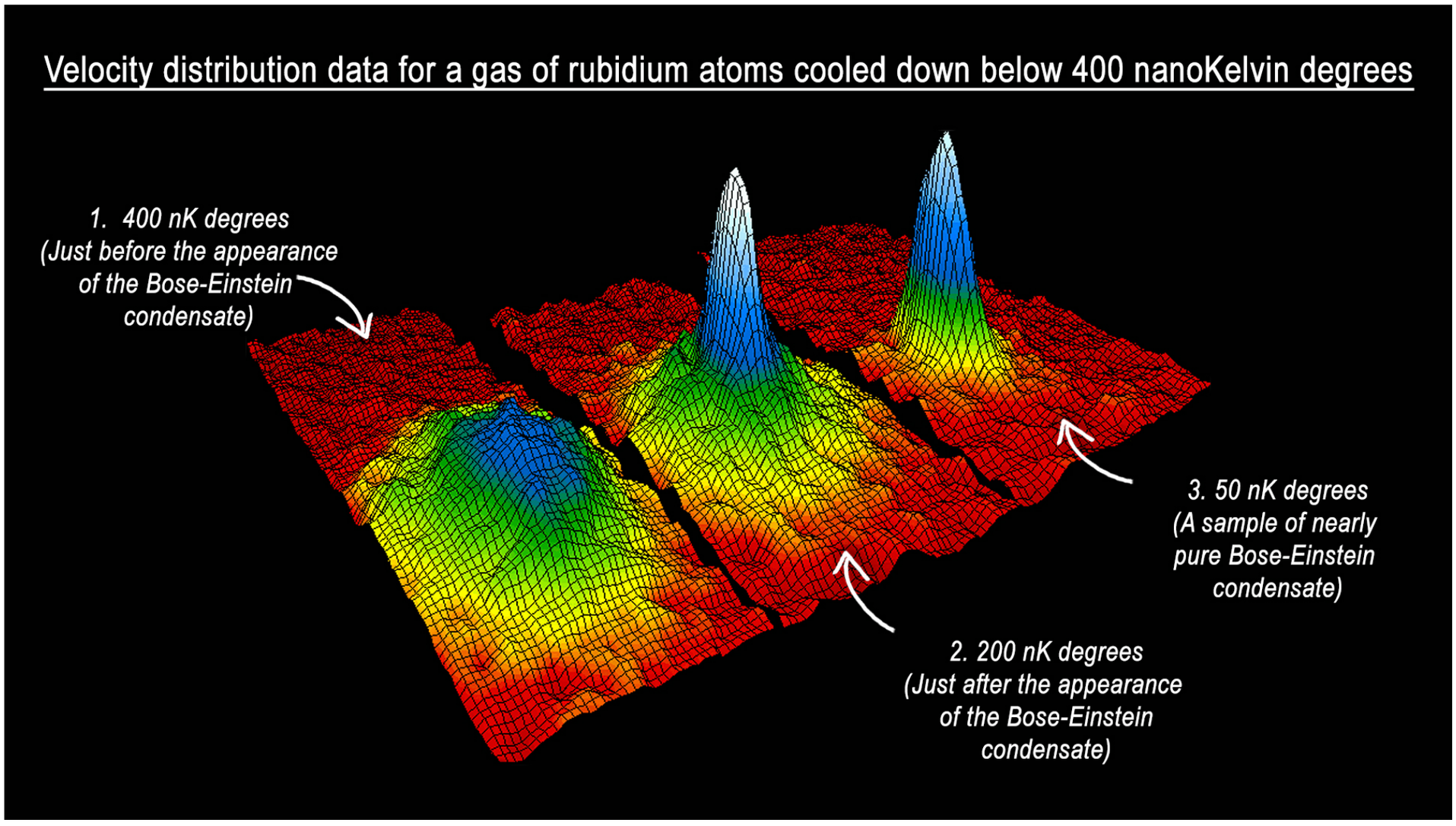
These electron pairs obey Bose-Einstein statistics, and they are the bosons associated with the condensation phenomenon.

10.12.3 Holy Grail

To produce a near-ideal Bose gas and cool it down below T_c has long been considered a Holy Grail of experimental physics.

This Grail was finally achieved in 1995: Bose- Einstein condensation was observed directly by an experimental group in Boulder, Colorado - Eric Cornell and Carl Wieman(2 friends) of the University of Colorado Boulder using rubidium atoms.

The bosons are Rubidium 87 atoms. For the very low densities involved, $T_c \simeq 170$ nK.



11. Liquid-gas phase transition - back to the classical way of thinking about problems…..

11.1 Introduction

We all know about phase transition from ice to water, then from water to steam.

It has been understood qualitatively for more than a hundred years.

Van der Waals explained it from Boltzmann statistical mechanics.

We are now going to see how phase transition from liquid to gas works.

We will not follow the way Van der Walls approached it.

His approach was very elegant, very nice, and equivalent to what we are going to do.

Our approach, however, will be based on a system that we have already studied: magnets, more specifically, the Ising magnet.

We will see in what way the Ising magnet is a model for the liquid gas phase transition.

Remember that the Ising magnet was more a simple mathematical model for any kind of phase transition, when elementary building blocks have two possible states, and adjacent pairs are coupled in some way, than a real model of magnetism in nature.

It is too crude for that.

But it is quite adapted to illustrating how sudden changes in the properties of certain systems can appear when some parameters are changed continuously.

Before we do that, let us review quickly what we learned in the last part of our discussion.

We studied the mean field approximation, not for the liquid-gas but for the Ising model.

It enabled us to see how and why a phase transition can occur in certain systems.

We also know that it does not work in one dimension.

In fact it then gives the wrong answer.

Indeed, before studying several dimensions, we showed that in one dimension there was no phase transition.

The mean field approximation always tells us that there is a phase transition.

We saw why it is appropriate to use it when the number d of dimensions is large.

We also stated, and gave some heuristic justifications, that there is a phase transition as soon as $d \geq 2$.

By the time we get to three dimensions, every particle in a lattice has enough nearest neighbors – six of them in the basic cubic lattice case – that the mean field approximation becomes very good.

It is not accurate in the sense of 1% but accurate in the sense of probably 10%.

So the mean field approximation is a good way qualitatively to understand how phase transition takes place.

Let's consider a two-dimensional square lattice.

At each site, also called vertex, there is a little magnet that can be up or down, figure 1.

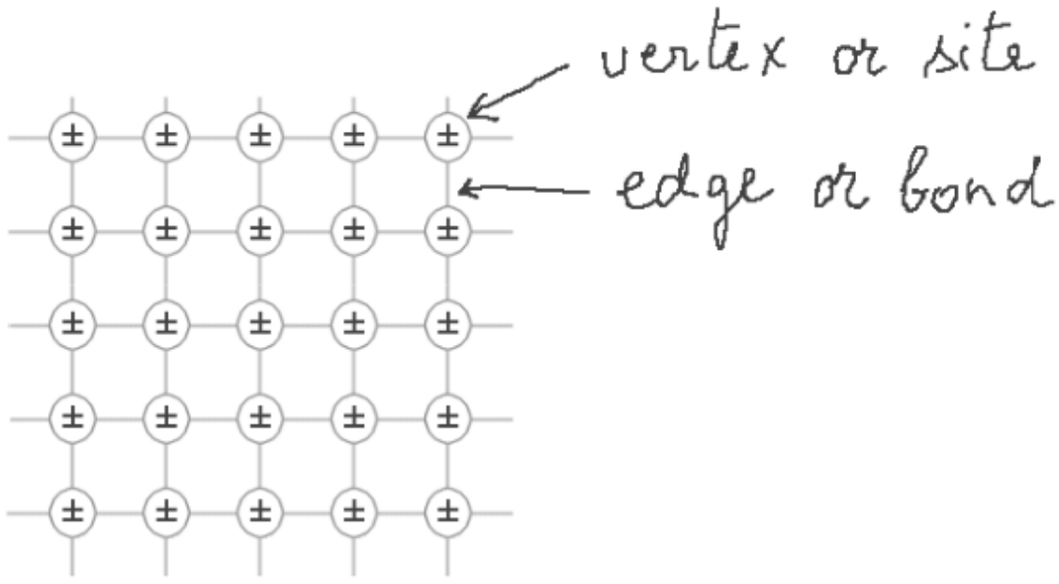


Figure 1: 2D square lattice of little magnets, sometimes called spins. They are located at the sites of the lattice. Each spin is either up ($\sigma = +1$) or down ($\sigma = -1$). And the edges between the sites are also called the bonds, or the links.

Each little magnet either points up, in which case its corresponding σ is equal to $+1$, or points down, in which case its corresponding σ is equal to -1.

The magnets are positioned at the sites, and they are linked by bonds, also called edges, or links.

For each magnet, σ is a random variable.

For the i-th magnet the random variable is more precisely σ_i.

They all have the same distribution of probability, and therefore in particular all the same mean $\overline{\sigma}$.

When $\overline{\sigma} \neq 0$, by definition there is magnetization.

We talked about two different kinds of models for the magnets.

Before the Ising model, we first considered a very simple model where the little magnets have no relation among each other.

We say that they are uncoupled.

They can be pictured as forming a big linear sample, or a multidimensional one.

It doesn't make any difference in this model.

Moreover the whole system was plunged in a magnetic field B.

Then there is no magnetization only when the temperature is infinite.

We saw that there is of course magnetization at T = 0, because then all the elementary magnets must point in the same direction, which, taking B into account, makes the energy of each of them minimal.

As soon as the temperature is less than infinite there is some magnetization.

And there is no phase transition.

Then we turned to the Ising model, which is more difficult.

We studied it first in one dimension, then is several dimensions.

In all Ising models, anyway, the energy in the system is no longer stored independently in each spin, but is stored in pairs of adjacent spins.

Equivalently we can say that it is stored in each bond, or link.

The energy in a bond is lower if the two spins are parallel, and higher if they are antiparallel

If two neighboring spins are *parallel*, we call that an *unbroken bond*.

Then the energy of that bond is low.

If they are *antiparallel*, we call that a *broken bond*.

Then the energy of the bond is high.

We wrote an expression to represent that.

The energy of the whole system is

$$E = -j \sum_{links} \sigma_n \sigma_m \tag{1}$$

We don't try here to make the notation fancy.

Remember that in physics, as well as in mathematics, notations are only shorthands for descriptions and explanations in plain English.

What makes them sometimes a bit cryptic is that, whenever we can, we devise them in such a manner that they lend themselves to algebraic manipulations and rules which, more or less automatically, express operations, like if $x+y = 8$ then $y = 8-x$.

When I was 10 years old, one day I witnessed someone come to my math class and ask: if $x = 5$ and $x+y = 8$ what was y? I was in a rapid progress, tracked elementary school program.

Since all the pupils who were not already bored said y must be 3, the inspectors concluded that we were ripe for higher algebra.

This unfortunately stressed the mechanical rule to the expense of common sense, and contributed to the esoteric aspect of mathematics, which many of my friends etched forever in their minds as their principal feature.

All too often official education consists in teaching a formalism which makes obscure what is simple.

That is also true in physics.

And that is why, in my classes, I explain higher physics as plainly as possible, with models described as straightforwardly as I can, and the least amount of arcane equations.

The values σ_n and σ_m simply mean the two ends of one link.

And the sum of is over all the links, in figure 1, or whatever lattice we are talking about, including the one-dimensional one.

The constant j is just associated to how tightly the spins are bound, how hard it is to break a bond.

How much energy does it cost to break a bond?

If we start with two aligned spins, then $\sigma_n\sigma_m$ is +1.

If they are anti-aligned, the product is −1.

So the difference, to go from +1 to −1, in the sum in expression (1) is two units, with a negative sign.

But there is a −j in front.

That means every broken bond costs an energy 2j.

Then we introduced a possible term in the energy that would correspond to a uniform magnetic field.

It is a term which acts not on neighbors, not on links, but on the individual sites.

It is an extra energy if a site is up or down.

It doesn't depend on what its neighbors are doing.

So let's add it in equation (1).

The total energy becomes the sum of two terms

$$E = -j \sum_{links} \sigma_n \sigma_m - \sum_{sites} h\sigma_n \tag{2}$$

In the second term, if h is positive, then $-h\sigma_n$ is lower when σ_n is positive.

In the calculation of E, if the second sum is all there was – which incidentally is the first model we studied –, then it would favor each σ_n being positive, or being up.

That would lower the energy of each spin σ_n.

And of course, with h negative, or changing the second sign in equation (2), it would be the other way around.

Nevertheless thermal fluctuations will cause σ_n to randomly point down sometimes,

And at high temperature, it will be down as much as up.

So, in the complete expression (2) for the energy E, the first term wants the spins to point in the same direction, and the second term wants them to point up.

Notice that, since it is only the difference between various levels of energy which matter, we could add a constant k on the right-hand side of equation (2) without changing anything.

$$E = -j \sum_{links} \sigma_n \sigma_m - \sum_{sites} h \sigma_n$$

You may also observe that the higher j is, the harder it is to break a bond.

And the higher h is the more energy is stored in one spin when it is down.

But let's not pay attention to that for the moment.

We are in a simple abstract model to study phase transition, and j and h are just given numbers.

How much energy do we have in the first term if all the spins are aligned?

That is the lowest possible energy coming from the bonds.

It is just the number of links times −j.

Incidentally, how many links are there in a 3D cubic lattice?

Every site has 6 neighbors .

But, if we want to avoid double counting, it is the number of sites multiplied not by 6 but by 3.

We neglect the border effects if the size of the whole lattice is big compared to the size of the bonds.

It is easy to see in 2D, figure 2.

Starting from the lower left site, say, $\sigma 1$, draw two links.

That leads us to the two sites $\sigma 2$ and $\sigma 3$.

From each of these sites, there are only two new links, etc.

So altogether, without double counting, each site corresponds to only two possible new bonds.

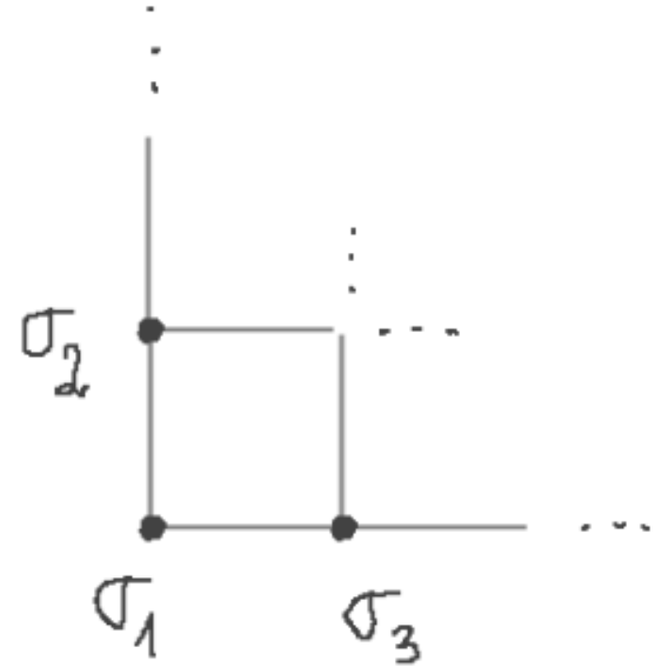


Figure 2: In a 2D square lattice, count of the number of bonds compared to the number of sites.

The same is true in 3D, with three new bonds for each site.

And generally speaking, the number of bonds is the number of sites times de number of dimensions.

In any case, looking at the two terms on the right-hand side of equation (2), we are in the ground state, i.e., the state with lowest energy, when everybody is aligned, and in the direction of h.

That energy is a certain quantity which we could subtract from the right-hand side of equation (2), to make the ground state energy equal to zero if we liked.

However, as said, the actual energy of the ground state is unimportant.

Only differences in energy levels are important in statistical mechanics.

Here what is important is the difference in energy between a spin up or down, and an unbroken or a broken bond, and how they sum up.

Let's remember what we found out in the mean field theory.

We calculated the partition function for one spin.

It has the usual form

$$Z = \sum e^{-\beta E} \tag{3}$$

where the sum is over the possible energy levels attached to that spin.

Each energy level is a part of the right-hand side of equation (2), which was for the whole system.

For one spin, an energy level is

$$E = -j\sigma \sum_{links} \sigma_m - h\sigma \tag{4}$$

Then, in order to make equation (4) tractable, we used the mean field approximation.

Let's go over it quickly, illustrated with a picture in 2D, although in truth the number of dimensions should be large.

So we consider one spin, or site, at the center of figure 3.

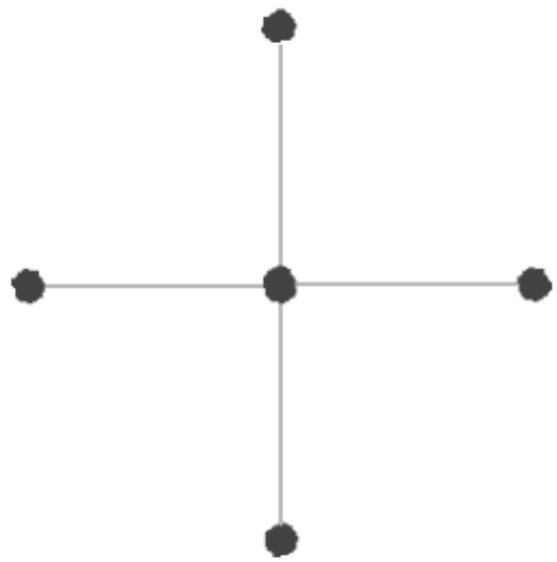


Figure 3: One site and its neighbors, leading to the mean field approximation.

The center site is surrounded by a bunch of neighbors, four in 2D, six in 3D, and 2 times d in d dimensions.

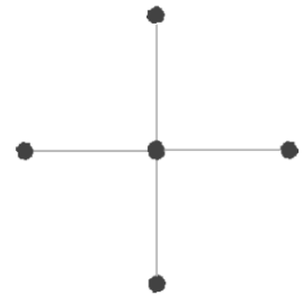
The spins of the neighbors have an average value.

That average might be zero, or it might be a net non zero value.

The more neighboring sites we have, that is the higher the number of dimensions, the better approximation it is to just say that the average of the neighbors is $\overline{\sigma}$, which is the average of all the spins in the system.

Moreover we can say, without making a statistical mistake, that the spin under consideration sees a collection of neighbors, each of which has spin $\overline{\sigma}$.

So the energy of all the bonds in figure 3, is



$$-2dj\overline{\sigma}\sigma \tag{5}$$

where d is the number of dimensions, j is the usual factor measuring how difficult it is the break a bond, $\overline{\sigma}$ is the average spin in the system, and σ is the spin under consideration at the center of figure 3.

Secondly, what about the sum over sites on the right-hand side of equation (2)?

Well, that term doesn't care about neighbors.

For one site, it is just the extra –hσ appearing in equation (4).

So, for one site, the complete energy level E, as a function of σ, is

$$E = -2dj\overline{\sigma}\sigma - h\sigma \tag{6}$$

And, with the average spin $\overline{\sigma}$ being just a number, we pretended that this is the statistical mechanical problem of *one spin*, the energy of which is given by expression (6).

Of course, equation (6) is simply

$$E = [\ -2dj\overline{\sigma} - h\ ]\,\sigma \tag{7}$$

That is the calculation we did when we first looked at these magnets.

And we found out that the *average of the spin under consideration* in the center of figure 3, which we called at first $\overline{\overline{\sigma}}$, satisfies

$$\overline{\overline{\sigma}} = \tanh\left(\ [\ 2dj\overline{\sigma} + h\ ]\ \ \beta\ \right) \tag{8}$$

Now comes the second crucial step in the mean field approximation: we say the $\overline{\overline{\sigma}}$ must be equal to $\overline{\sigma}$.

This of course would be true if we had a big enough sample and we were deep in the interior of it (and, as said, d was high).

This *self-consistency* trick leads to

$$\overline{\sigma} = \tanh\left(\ [\ 2dj\overline{\sigma} + h\ ]\ \ \beta\ \right) \tag{9}$$

That is our equation which determines $\overline{\sigma}$.

The factors 2, d, j and h are just numbers, so it is an equation in $\overline{\sigma}$.

Let's just remind ourselves what the solution looks like.

We did another small manipulation.

We defined the new variable

$$y = 2dj\overline{\sigma}\beta \tag{10}$$

And we wound up with the equation

$$\frac{y}{2dj\beta} = \tanh(y + \beta h) \tag{11}$$

Now since β = 1/T, this is also

$$\frac{Ty}{2dj} = \tanh\left(y + \frac{h}{T}\right) \tag{12}$$

First exercise: what happens if h = 0?

Then it is simply *Ty/2dj = tanhy*.

And now, just by inspection, we get a pretty good idea of what the solution looks like.

On the right side, *tanh y* looks like this, figure 4.

And on the left side we have a linear function of y which passes through the origin.

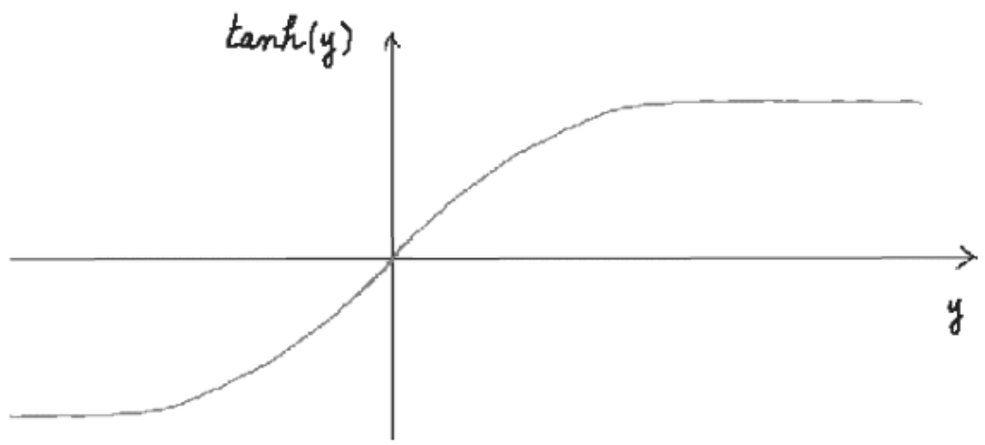


Figure 4: Graph of tanh y.

It has the slope *T/2dj*, where T is the temperature.

So, in particular, when the temperature is high, the slope is steep.

And the only place where these curves intersect is right at the origin, figure 5.

Therefore, for large temperature, there is no magnetization.

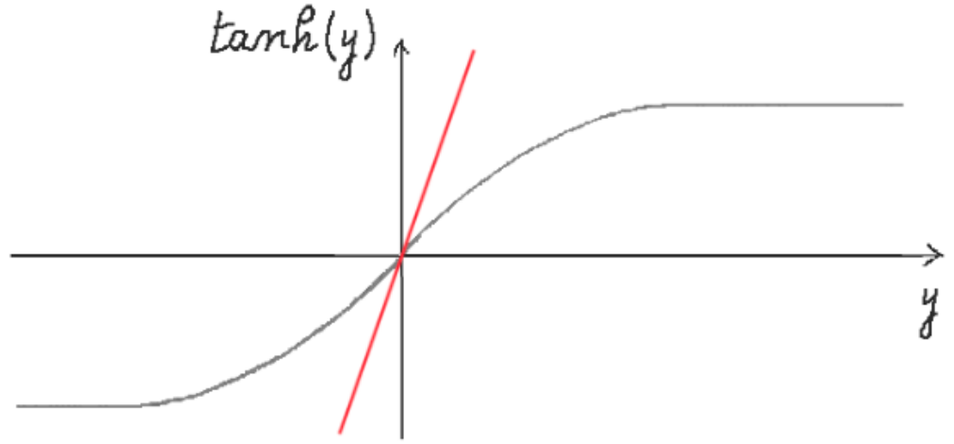


Figure 5: Intersection between T y/ 2dj and tanh y when the slope of the straight line is steep.

Indeed the solution to equation (12), when h = 0, is y = 0.

But y is proportional to $\overline{\sigma}$.

Therefore $\overline{\sigma} = 0$ when the temperature is high, as expected.

Now we start lowering the temperature.

The straight line rotates clockwise while still going through the origin.

Its slope decreases.

When T reaches 2dj it becomes tangent to the hyperbolic tangent curve at the origin.

We saw all that in part 9 of the lectures.

We are only going through a brief review, before turning to liquid-gas transition, because sometimes a second presentation may clarify points which had remained unclear or which we had missed.

So that means, at temperature T = 2dj, when the slope of the straight line is 1, which is also the slope of *tanh* at the origin, that it is the magic point where the two curves lie on top of each other.

Go past that.

Decrease the temperature even more.

That is, decrease the slope of the straight line even more.

And now we have three solutions to equation (12).

There is still mathematically the solution at the origin.
But now there are also two more solutions, with $y > 0$ and $y < 0$, figure 6.

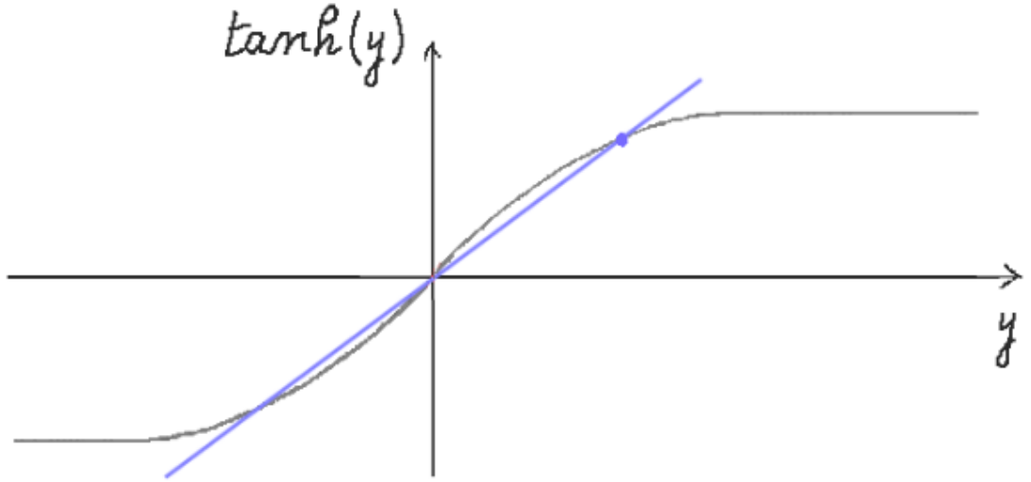


Figure 6: Intersections between T y/2dj and tanh y when the slope of the straight line is less that 1.

Those solutions where y ≠ 0 are magnetized. Indeed, $\overline{\sigma}$ being equal to $y/2dj\beta$, see equation (10), it is non-zero.

That is a phase transition.

Past that point, when now $T < 2dj$, the system doesn't quite know what to do.

It doesn't stay at origin because it is unstable.

But should it go to the solution point with y > 0, or should it go to solution point with y < 0?

In other words, there is an ambiguity.

It seems like the spin at the center of figure 3 wants to have a non zero average value.

But is that average value up, or is it down?

On the right side of figure 6 the average value of the spin is up, $\overline{\sigma} > 0$.

On the left side it is down.

And the problem is that there is a complete symmetry in the situation.

If you take all the spins and, whatever each one is, turn it over, you end up with the same situation and therefore the same problem back again.

So there is no way to decide whether the spin will prefer to point up or to point down, that is whether it will choose the solution with $y > 0$ in figure 6, or the solution with $y < 0$.

To decide we add a tiny little bit of magnetic field h.

The effect of that is to shift the *tanh* curve a tiny bit either to the left or to the right.

But it is enough so that the point $y = 0$ is no longer a possible solution, even unstable.

And only one of the two where $y \neq 0$ will remain, depending on the orientation of the tiny extra magnetic field, because one solution will have much less energy than the other.

That is what happens.

That is magnetization.

If the extra tiny external magnetization, however small, pushes the spins to be up on average, because that is where they have less energy, see equation (4), then up will win.

If the extra magnetization pushes them to be down, then down will win.

We shall draw a graph of what the magnetization of the system looks like as a function of two variables: the temperature and the external magnetic field.

Those are the two parameters that can vary in the problem.

We can turn the temperature up or turn it down.

We do that by heating up the heat bath or cooling it down, with a little fire under it or with cold air.

And we can change the magnetic field by running a current through a coil, and make the magnetic field larger or smaller, in the direction we like.

What about j?

The parameter j is also something which we could change.

In order to change it what we usually have to do is change the material.

In fact we can also change j by changing the pressure or other things.

But we are going to consider j to be a fixed number.

It is fixed for the material has a fixed chemical composition, at a fixed pressure, fixed distance between the atoms, etc.

The parameter j is a well-defined number which we can look up in a chemistry handbook.

But β and h are the parameters that are easily under our control.

So we might want to ask: as an experimental fact, what does the magnetization look like as a function of temperature and magnetic field?

The temperature $T = 2dj$, where a transition appears, is called the *critical temperature*.

It is not exact.

It is a mean field approximation.

It becomes exact when d becomes large.

For d = 2 it is off by a significant amount.

For d = 3 it is already very good.

But in any case it captures qualitatively what really is happening.

Remember that for d = 1 we saw that there is no phase transition, no critical temperature.

So let's denote the critical temperature T_c, read *T critical*.

And we are in a lattice of dimension $d \geq 2$.

Below T_C, the system is spontaneously magnetized.

Everybody wants to line up.

The lining up tendency wins, and everybody lines up that way.

Not absolutely.

They are not a 100% lined up.

It fluctuates but there is a net, in the sense of average, tendency for everybody to line up.

Above T_C, thermal fluctuations win.

And the system doesn't magnetize.

When we talk strictly about real magnets, T_c is also called the Curie temperature.

Here, however, we are talking about the general phenomenon of phase transition.

So it is called the critical temperature.

The Curie temperature is a special case specific to real magnets.

The "lambda point", in liquid helium, is another special case.

They all fall into a class of transitions of things which are very similar in many respects.

We will see why, or at least how the transition in magnets is related to another interesting transition.

Now let's draw a picture to study the magnetization as a function of T and h, figure 7.

We positioned T_c on the T axis, because it is the critical temperature, defined when there is no external magnetic field, i.e. h = 0, and at which things happen.

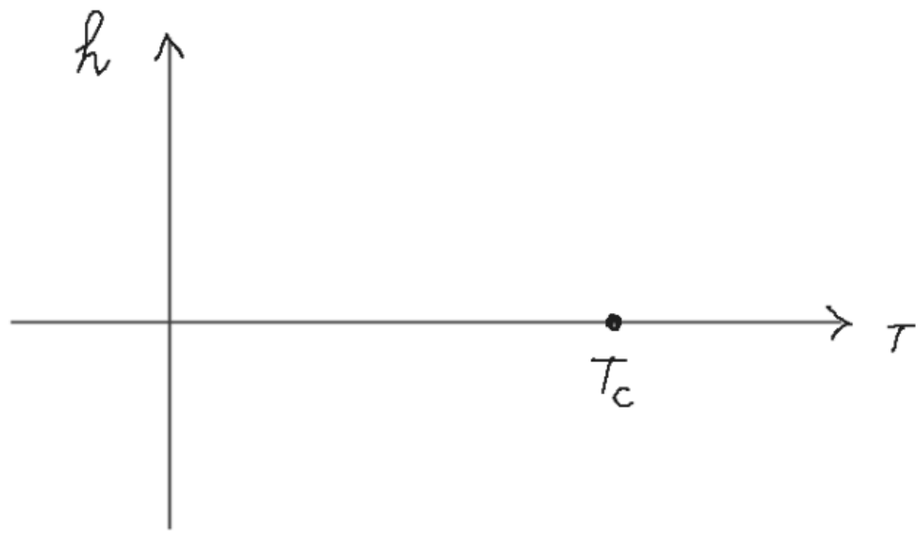


Figure 7: Study of magnetization in the (T , h) plane.

First of all, if we are at a higher temperature than T_c, and there is no magnetic field, h = 0, then the magnetization is just $\overline{\sigma} = 0.$

There is no bias due to an external magnetic field.

At high temperatures the statistics of ups and downs will equally win.

And so, on the T-axis, on the right of T_c, $\overline{\sigma} = 0.$

Now, sliding along the T-axis starting from a point farther than T_c, we begin lowering the temperature, that is we move leftward.

When we get to the critical point, the system just doesn't know what to do.

Should it go up or should it go down?

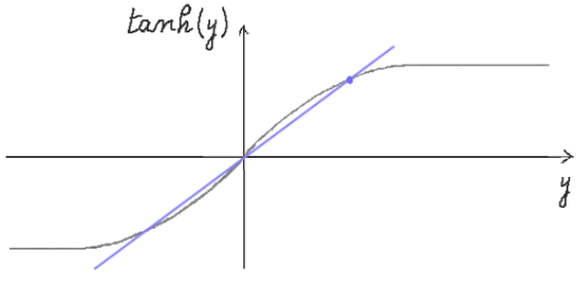
The tiniest little bit of stray magnetic field h will tell us which way to go.

If h is positive the magnetization will be positive.

Right above the T -axis it won't be zero, and right below it won't be zero either.

Spontaneous magnetization has happened, figure 8.

It corresponds to the two possible roots, with the straight line in figure 6 either intersected on the right side or intersected on the left side.



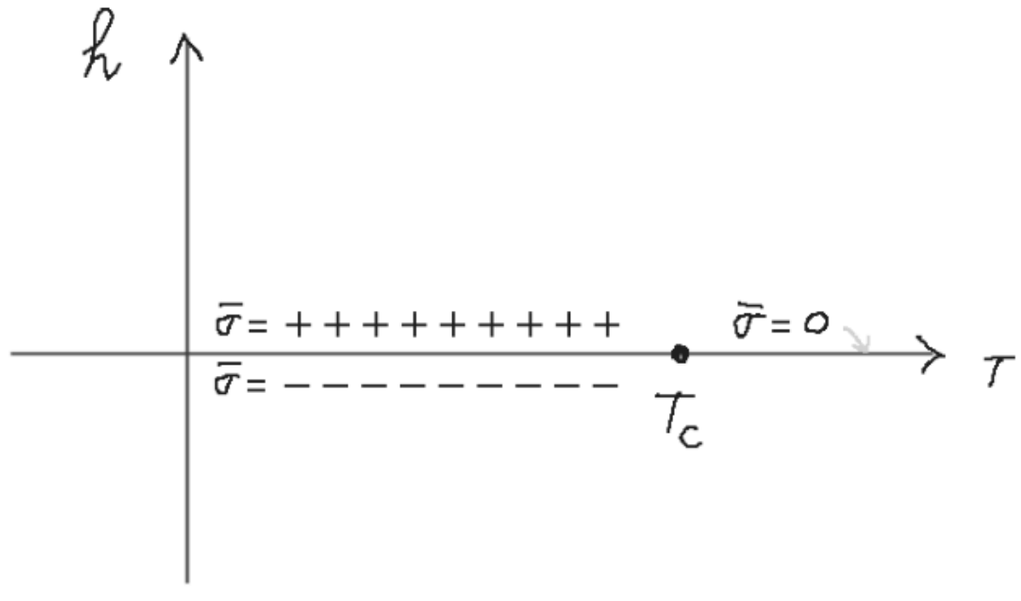


Figure 8: Magnetization on or close to the T -axis.

Physically, it is just that the extra little bit of magnetic field has broken the symmetry between ups and downs and has biased the sample.

That tiny little bit of biasing h is enough to make $\overline{\sigma} > 0$ if h points up, and $\overline{\sigma} < 0$ if h points down.

See equation (2) for the effect of the sign of h.

Now we can ask : what is the jump in the magnetization going across the T-axis?

Remember that $\overline{\sigma}$ is not necessarily $+1$ or -1.

$\overline{\sigma}$ is the average magnetization.

It can be 0 or it can be 2/3.

It can be $-1/4$ or any number between -1 and $+1$.

At zero temperature, i.e., at the origin, the absolute value of $\overline{\sigma}$ doesn't depend on the magnetic field.

When T = 0, all the spins are lined up.

$\overline{\sigma}$ is either $+1$ or -1.

Where $h < 0$, all the spins point downward, therefore $\overline{\sigma} = -1$.

And where $h > 0$, they all point upward, and $\overline{\sigma} = +1$, see figure 9.

That comes from the fact that at zero temperature the spins all choose the lowest energy, and there is no fluctuation.

What about life at the critical point?

Imagine going up and down from $h > 0$ to $h < 0$ at $T = T_C$.

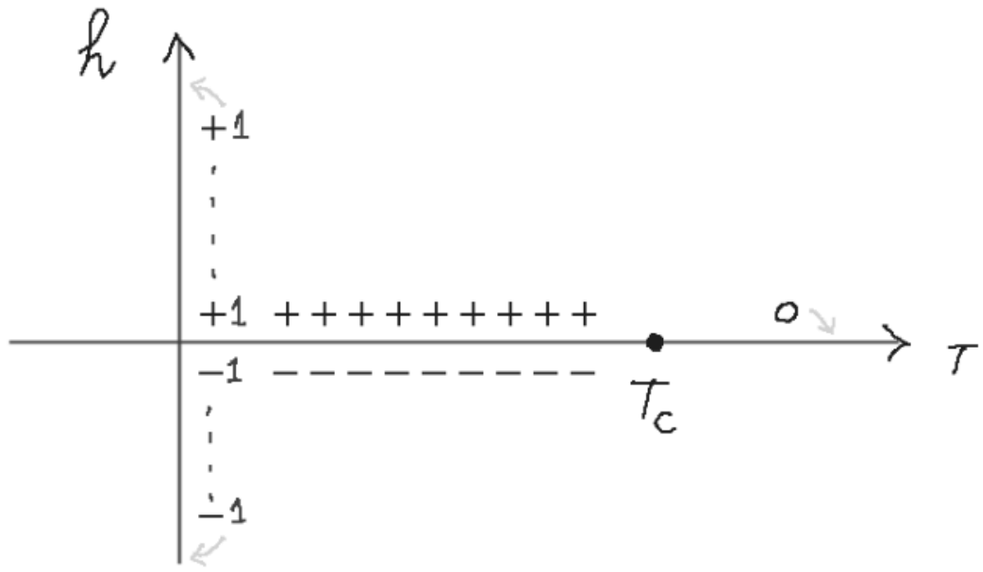


Figure 9: Magnetization along the h-axis.

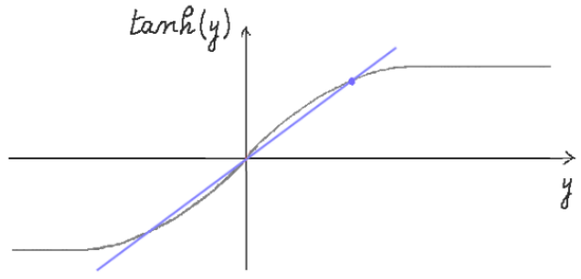
At exactly the critical point, the straight line in figure 6 has exactly the slope 45°.

So at the critical point, the magnetization is clearly 0.

Now, from T_C, we decrease the temperature a little bit, while staying on $h = 0$.

That is, we don't turn the external field.

We lower the slope of the straight line in figure 6.



The two branches of *tanh y*, the one on the right and the one on the left, now play a role.

As we saw, the magnetization $\bar{\sigma}$ doesn't stay 0, because it is unstable.

The reason is because *there is always a stray magnetic field h, however tiny, the consequence of which is that* $\bar{\sigma}$ *becomes either positive or negative*.

Of course, as we move vertically upward, staying just to the left near the critical temperature, the magnetization increases.

But it does so continuously in the sense that, in the vicinity of the point ($h = 0, T = T_c$), to the left of it, when $h > 0$, $\bar{\sigma}$ is positive but very small, and, below the T-axis, $\bar{\sigma}$ is negative but very small.

In summary, as we come in to the critical point ($h = 0, T = T_c$), the magnetization tends to 0 from every direction, including arriving from the left.

However, once we are away to the left of the critical point, there is a jump in the magnetization.

There is a shear in the magnetization surface $\bar{\sigma}(h, T)$ along the half line $h = 0, T < T_c$.

But, as just explained, even that shear goes to zero as we approach T_c.

Let's pursue our investigation of the surface $\bar{\sigma}(h, T)$ in some detail at different points.

For example, let's look first at a temperature above the critical one.

If we had our external magnet turned on, and we slowly varied the magnetic field h from a positive value to zero, while staying at a fixed temperature T above the critical temperature T_c, then the magnetization $\bar{\sigma}$ would get smaller and smaller, going to 0 continuously, figure 10.

And if we passed into $h < 0$, $\bar{\sigma}$ would increase again in the negative values (that is $|\bar{\sigma}|$ would increase, while $\bar{\sigma} < 0$).

And all this would happen completely smoothly, just as we might expect.

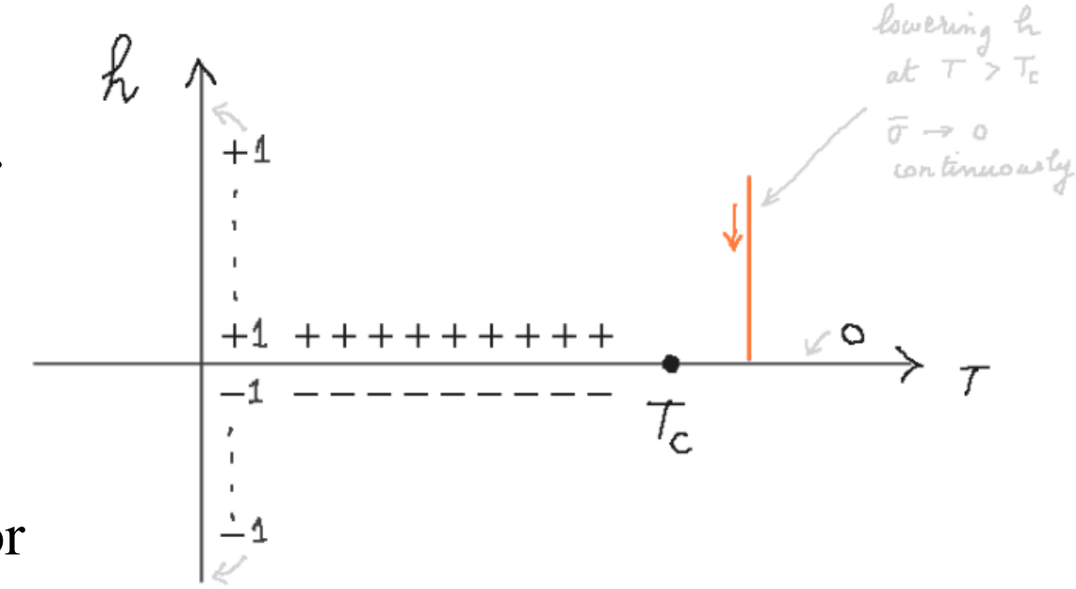


Figure 10: Lowering h while staying at $T > T_c$.

Now let's look at a temperature below the critical one, and for the sake of variety let's position ourselves in the $h < 0$ half plane.

Actually it is a quadrant, because T is always greater than or equal to zero.

As the experimental physicists manipulate the magnetic field at fixed temperature, the magnetic field varies.

At first the magnetization is nice and smooth until it reaches the T-axis, figure 11.

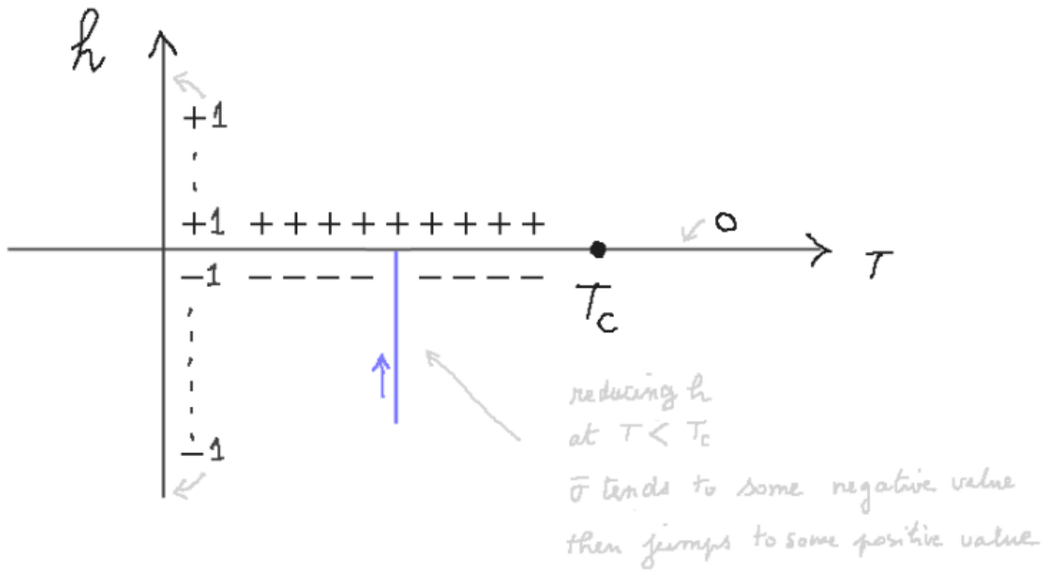


Figure 11: Varying h while staying at $T < T_c$.

The limit at h = 0 *coming from negative values*, is some magnetization $\overline{\sigma}$ negative.

If we are not afraid of heavy notation, we can denote it $\overline{\sigma}_{-}(T)$.

Then change a tiny bit the magnetic field h making it positive: the magnetization changes discontinuously to some positive value, $\overline{\sigma}_{+}(T)$.

The jump is biggest at the origin, and goes to zero as the temperature approaches T_C.

Phase transitions are classified by their order, which depends on how smoothly the free energy varies with respect to the parameter.

Technically what happens in figure 10, when we cross the T-axis somewhere between 0 and T_C, is a second-order phase transition.

But we don't need to distinguish them because we are not going to get a chance to talk about first-order phase transitions.

In summary it is the jump which takes place on the segment where h = 0 and $0 \leq T \leq T_C$, i.e., where there is a shear in the surface $(h, \overline{\sigma})$.

This jump in magnetization from $\overline{\sigma}_{-}(T)$ to $\overline{\sigma}_{+}(T)$, when we cross the T-axis at some temperature T below the critical one, is a symptom that in a sense there are two phases.

Suppose we didn't know anything about what happens on the right part of figure 11.

Suppose the critical temperature was too high for our experimental physicist to ever study and become aware that things behaved differently beyond it.

Then he would say that there are two phases.

There is a phase of negative magnetization; there is a phase of positive magnetization; and, as we vary the parameter, it jumps.

It is not that all the spins flip from down to up.

It is the average that goes from some value down to some value up.

And, as we stressed, the jump is biggest near $T = 0$ and goes to zero as T approaches T_c.

Now there is another way we can move around the space (h, T).

We might wonder: is there a way to move from the positive zone to the negative zone without a sharp change?

And of course there is.

We can go *around the critical point*.

Then we experience no jump.

But we do get from positive magnetization to negative magnetization.

So, on the one hand, experimentalists starting from a point (h, T) in the lower part, and left of T_c, and who have never heard about the critical point, will say:

"*There are two phases. They are distinct. And not only are they distinct,*

but there is a sudden jump as we vary the magnetic field h."

On the other hand, someone with more experience with the system would say:

"*oh, no, no, no. There is really not a sharp difference between the two zones.*

Let me show you how to go from one to the other without jumping."

And both would be right.

Both statements are correct statements.

There is a jump if we stay at fixed temperature below T_c.

But if we allow ourself to explore the whole region of (h, T), there are ways to go from the upper zone to the lower zone which don't involve the jump, which are smooth.

Now we have everything we need to know about magnets.

But we are actually interested in the liquid-gas phase transition.

11.2 Liquid-gas phase transition - I want to do this analysis using only the ideas we have already introduced! It will be a tortuous path but we will learn more about the transition!!

What does the above study of the surface $\overline{\sigma}$ (h, T) have to do with liquid-gas transition?

There is a way to think about exactly the same system as being the transition between liquid and gas.

It is not good for liquid to solid.

That case is very different.

So let's talk about the former and what it possibly can have to do with a bunch of particles.

Consider a box of particles.

It can be gas or liquid, so let's say a box of fluid.

And let's plot its potential energy, figure 12.

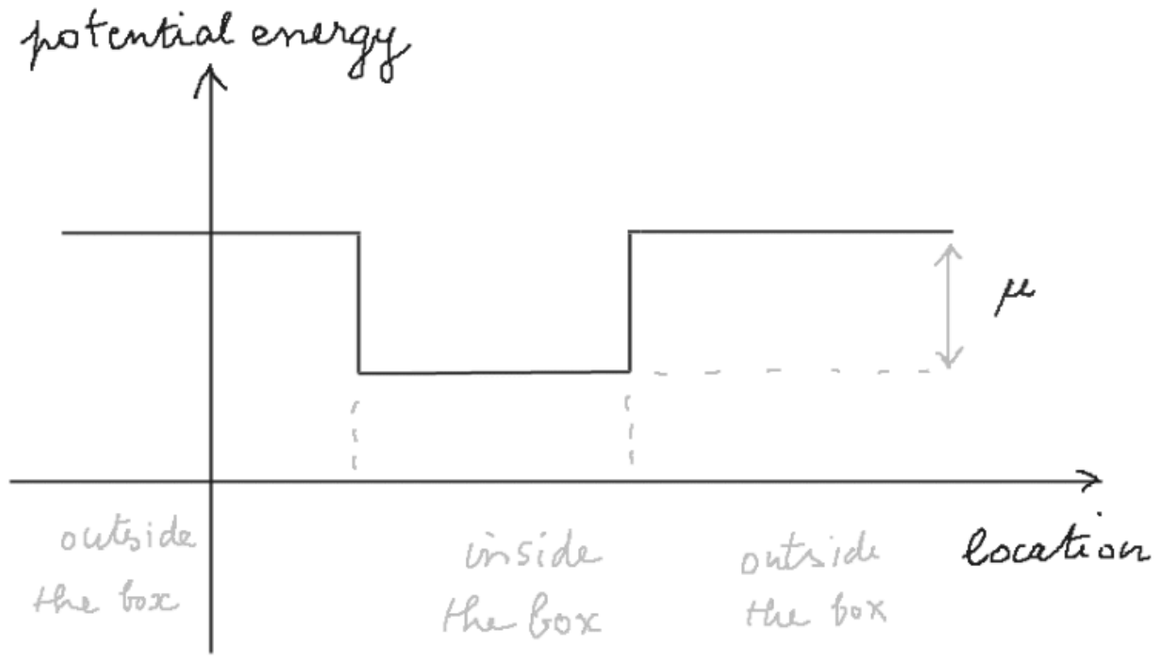


Figure 12: Potential energy of a box of fluid.

The vertical edges correspond to the walls of the box.

They don't represent the walls though.

This is only a graph of the potential energy.

In between, we are inside the box.

Elsewhere, we are outside.

If it turns out the box really corresponds to some macroscopic material, or whatever it happens to correspond to, then the energy of a particle is lower when it is inside the box than it is when it is outside the box.

Now we could manufacture this system by having the inside of the box and the outside of the box being of slightly different elevation.

We could have a flat box with a lower bottom over some area.

But for whatever reason we are going to imagine that a particle at rest inside the box has less energy than relative to what it has outside the box.

That difference in energy has a name.

It is called *chemical potential*.

And it is usually written μ.

It is actually the amount of energy that it takes to remove the particle from the box.

In this particular case it is not important to know where the difference comes from or why there are different energies.

We just take it that the energy is lower inside the box.

Furthermore, there are particles inside the box and outside the box

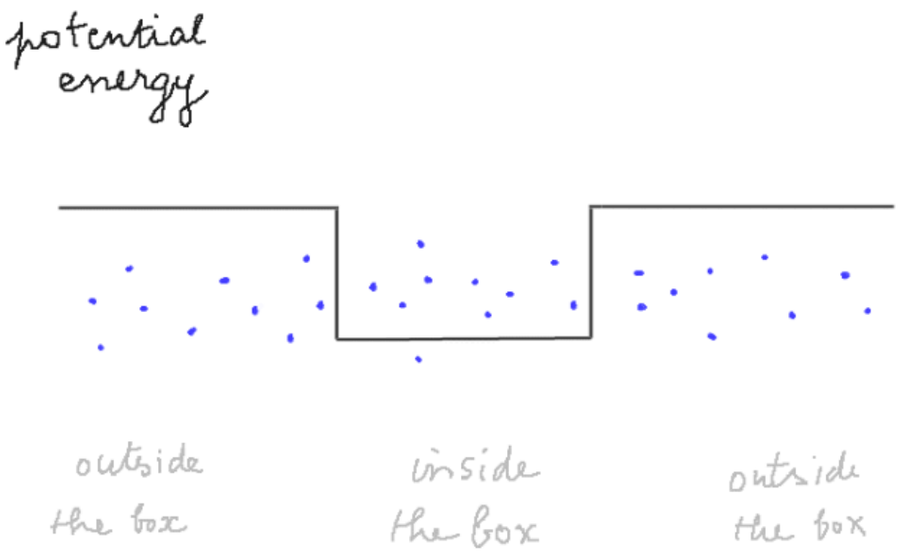


Figure 13: Particles inside and outside the box.

They are all over the place, figure 13.

One more thing: the walls of the box are permeable.

Particles can come in and go out.

They can cross the boundaries of the box.

This could be, for instance, a bunch of molecules in the following system.

We dig a hole in the ground.

And we deposit some particles in as well as outside the well.

If the particles are in the hole, they have less potential energy than if they are out of the hole.

We can ask what is the density inside the box and outside the box.

We are thinking of thermal equilibrium.

Where is the density of particles likely to be larger, in the hole or out of the hole?

In the hole.

Indeed, the energy is less in the hole.

And so it is natural that the density there be higher, because the Boltzmann distribution always favors lower energy.

Remember, the Boltzmann distribution is $P_i = e^{-\beta E_i}/Z$, where P_i is the probability of the state i, which has energy E_i.

That is the Boltzmann distribution – at thermal equilibrium by definition – in the discrete case, which is often easier to conceive.

Remember, there is a big space of states Ω.

The individual possible states, which are the elements of Ω, are labelled ω_i.

The i-th state has energy E_i.

And the probability that the system be in the i-th state is P_i.

We constructed the whole model in parts 1 to 4. You are invited to refresh your memory by going back over them.

In view of all we have done since, they may make more sense, and be easier to follow.

The function – that is the probability – decreases as the energy increases.

So it always prefers lower energy.

So the answer will be that the system will adjust itself so that the density in the box is in a certain kind of equilibrium.

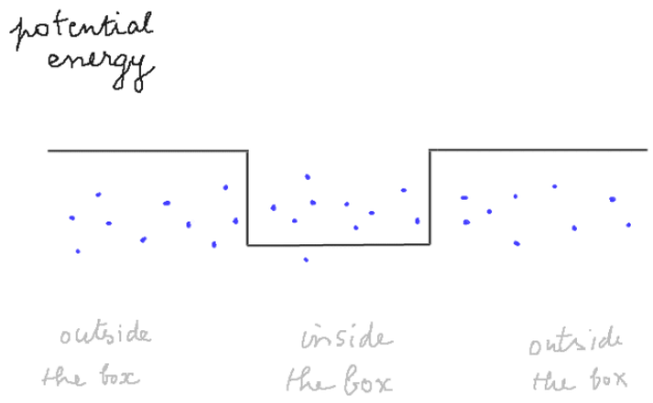
And the equilibrium will depend on the energy cost of taking a particle outside the box.

Now we can set this up in a very definite way and do ordinary statistical mechanics.

There is a potential energy which depends on position.

But it depends on position in a very special way.

It is flat inside of the box of interest, that is between the walls in figure 13.



And it jumps when we go outside the box.

We focus on inside the box and ask the question: what is the density of the fluid as a function of the depth of the chemical potential μ, see figure 13?

Note that μ is called the chemical potential because in general it is different for different kinds of molecules.

So we could imagine a situation where one kind of molecule prefers to be in the box, and the other kind of molecule doesn't prefer to be in the box.

Turning on the chemical potential then can change chemical composition in various ways.

However we are not interested in the chemistry here.

We are interested just in this idea that we can tune the density by changing the chemical potential.

All that μ is here is a term in the energy which depends on the number of particles in the box.

There is a factor μ for each particle.

If there are N particles in the box, then Nμ.

Now in this problem the number of particles is a variable, because particles can come in the box and go out of the box.

That means a configuration depends not only on the position and momentum of every particle inside the box, but also on the number of particles.

That number is free to change because particles move in and out of the system.

And part of determining thermal equilibrium is determining how many particles are on average inside the box.

We are not going to go through this in detail, but we will outline the calculations.

The usual formula for the partition function, e to the minus β times the energy of each state, properly summed, will in this case look like this

$$Z = \sum_{PXN} e^{-\beta E + N\mu\beta} \tag{13}$$

The term E in the exponent represents the energy of one state.

As usual it is multiplied by $-\beta$.

But now to it is added the term $N\mu\beta$, which is the potential energy of that state.

There is a plus sign because the potential energy of one particle inside the box is $-\mu$ compared to one outside.

And there are N of them.

Finally the sum is over all configurations (X, P) and all N.

We are not going to work it out, it would make this part of the discussion too lengthy, but we shall see what the meaning of the chemical potential is.

The factor μ is the thing that we adjust if we want to adjust the density of the fluid.

Actually we could do a lot of things to adjust the density.

We could seal the box and not let the particles go in and out.

And we would put in just the number of particles we wanted.

But we may be interested in the problem where particles can go freely in and out of the box.

In that case, we change the density by changing the chemical potential.

By varying it, we can vary the average number of particles in the box.

In this language what is a liquid-gas transition?

We take our box, we keep it at a fixed temperature, and we start varying the chemical potential.

The more we increase the chemical potential, the more it wants to pull particles back into the box.

Of course it depends on the size of the box, but there is a density that is a natural function of the chemical potential.

And how does the transition materialize?

As we vary the chemical potential, at some point the density of the fluid suddenly changes.

The fluid changes from being a gas to being a liquid.

The box, suggested in figure 13, of course might be a region of three-dimensional space.

We can also imagine a real three-dimensional box.

When a particle goes in, we actually get energy out of it, because of the chemical potential difference. And when a particle goes out, it costs energy.

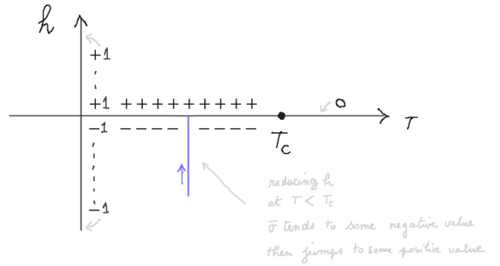
The extra energy released, which is the same for every particle going into the box, will tend to increase the density inside the box.

Even though the atmosphere around the box might be at a temperature and density corresponding to vapor, if we keep lowering the energy inside the box, sucking particles in, then all of a sudden we hit a transition point where everything becomes liquid.

The difference between liquid and vapor is some discontinuity in the density, liquid being denser than vapor.

So if we could do this experiment with the box, vary the potential energy of particles inside the box, while outside the conditions were such that we had steam vapor, particles would congregate inside the box as we lower their energy.

And all of a sudden at some point we would jump to the liquid phase.



That liquid-gas transition is very much like the earlier transition studied in the previous section, see for instance figure 11.

Let's spell out exactly what the relationship is.

First of all, what are the conditions required to have a liquid-gas phase transition?

Two conditions need to be satisfied:

1. There must be a hard core repulsion between molecules. In fact, molecules always have hard core repulsion. You cannot stick two molecules into the same site. For a number of reasons, they will repel. So it is natural to view the molecules of our fluid like little billiard balls. They don't want to interpenetrate. Roughly speaking you cannot put two in the same site. And that is a form of really hard core repulstion.

2. The next thing important is that there be a little bit of attraction between molecules when there are not quite touching.

So we need a potential energy that is big, repulsive, positive when the molecules try to get into each other.

And then, when we separate them a little bit, all of a sudden they want to attract a little bit.

That is an extremely common feature of molecular interaction:

hard core repulsion and very short range attraction.

That is what is needed to have the standard liquid-gas transition.

We shall see how the magnet produces exactly that situation.

Let's start with the magnet with all of its little elementary magnets pointing down.

When they are all down, that is the ground state.

And let's call that empty space.

This is a change from equation (2), where the spin lower energy was up.

It is entirely conventional.

We just have to think now of h being a negative number, or in the equation change the sign in front of it.

In fact what we are going to say is something like this: on each site of the lattice we can have a particle or not.

This is our aim.

Out of our liquid gas problem we are going to make a lattice version of it.

The particles, or molecules, live on a lattice.

At a site, as said, there can be a particle or no particle.

When a site has no particle it is assimilated into *down*.

And when all the sites are empty, that is the *ground state*, figure 14.

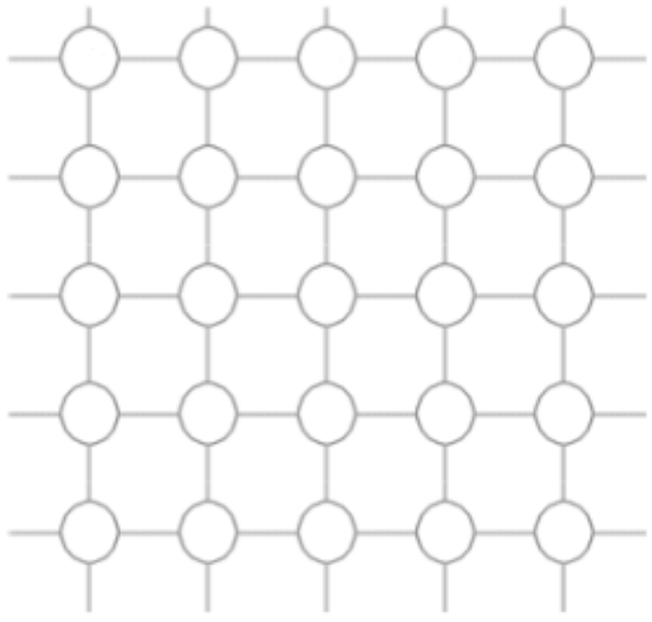


Figure 14: Fluid modeled as a lattice. Represented is the ground state with all the sites empty.

So in terms of σ at a site we can have

- $\sigma = -1$: no particle at the site
- $\sigma = +1$: one particle at the site

What about two particles at a site?

Well, the magnet doesn't allow that.

The value of σ can't be 2.

It can be only −1 or 1. It cannot be 4.

It cannot be anything else.

Thus, immediately the model forbids the possibility of two particles on a site.

And that is exactly what we want.

We are trying to model a gas where we cannot put two particles on a site because molecules have a hard core.

Automatically built into the system of particles, if they are described by the same kind of variables $\sigma = \pm 1$ as the magnet, they have an infinite hard core repulsive barrier when they try to get onto the same site.

So we satisfy above condition number one.

What about condition number two?

Do they attract?

Here we are going to use for the energy exactly the Ising model of energy plus a little magnetic field.

And we shall see how this translates in the language of our particles of fluid which are hard core and have short range attraction.

In fact we can keep on using the magnets idea or we can use the particles idea.

The point is it is the same mathematical system.

So let's not worry about it, and start with the term already given by equation (1)

$$\sum_{links} -j\sigma_n\sigma_m \tag{14}$$

It is the energy contained in each pair of neighboring sites, summed over all the links.

How much energy is there if there are no particles whatsoever?

That is the situation where all the σ's are down.

Then there is an energy −j for each link.

But this total ground state's energy, which is −j times the number of links, what do we do with it?

We don't do anything with it.

That is just the energy of the system when there is no particles at all.

Does it matter to anything?

No, it doesn't. It is always there.

It is the ground-state energy. And we can throw it away.

We could get rid of it by subtracting –2j times the number of links.

We would get

$$\sum_{links} -j\sigma_n\sigma_m + 2jN_{links} \tag{15}$$

But we might as well just ignore it.

It is just a number.

And the energy differences are the important thing.

So we will continue to use expression (14) as our energy function when there are no particles.

In other words, we set the zero of energy when there are no particles.

That gives us a certain energy.

To nail the point – and use the prerogatives of physicists – let's just call it zero.

In other words, we set the zero of energy when there are no particles

Now, what happens if we put in one particle?

That means flip one of the spins over, figure 15.

All the other spins are still down.

But one is up.

Again, to flip one spin up or to introduce one particle are mathematically equivalent operations.

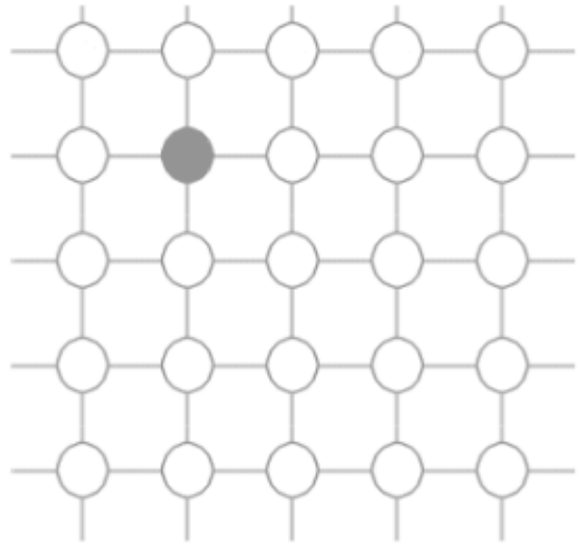


Figure 15: One site occupied by a particle. It is equivalent to one spin flipped to positive.

So you should not be disturbed if we go back and forth between the two ways to describe what is going on in the model.

Let's calculate the energy that we get.

How many bonds have we broken?

It depends on the dimensionality.

Let's stick to the model in 2D because it is easy to visualize.

If we flip one site over, then we break four bonds, figure 16.

The amount of energy necessary to break one bond is 2j.

That yields 8j for all four bonds in the figure.

Later on when we add a little magnetic field, we will add 2h.

But for the time being, 8j is the energy it costs to create one particle at the point shown in figure 16.

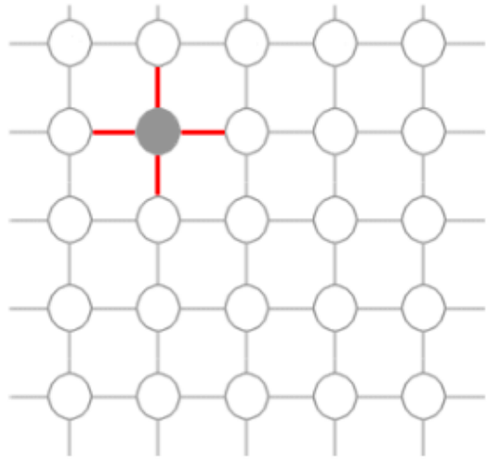


Figure 16: Bonds broken from flipping one site over.

What about two particles?

Suppose they are not adjacent.

We don't need to draw another figure.

It is easy to see that, in that case, it costs 16j to put in the two particles – even if they are separated by only one diagonal interval.

And if they are neighbors, if they are adjacent?

When we introduce 2 adjacent particles, we break only 6 bonds.

The one between them is no longer broken, because both sites have flipped.

So in terms of energy we pay only 12j.

If we plotted the energy as a function of the distance between particles, we would find that just the mere act of putting two particles in the lattice, farther apart than one link, costs us 16j.

But that is alright; that energy is always there, no matter where we move the particles around.

However, if we move the two particles within a bond length, the energy decreases.

That is like a potential energy which decreases when they get very close together.

So there is a potential energy, which depends on the positions of the particles.

And when the particles get within one bond length of each other, the energy decreases by 4j.

That is like having a short-range attractive energy.

When the two particles are close to each other, the energy is negative relative to what it would be if they were far apart.

So particles attract. They like to be close together.

How close together?

One bond length away.

But that is not an overwhelming attraction.

It doesn't say that the energy is infinitely negative if you put them next to each other.

That is the case of attraction laws in $1/r^2$.

Then the delta potential energy is proportional to $-1/r$.

So it is infinitely negative for instance when point masses are at distance zero of each other.

There is just a modest amount of attraction, a modest saving of energy of putting them close together.

And it is very much like the molecular attraction of a pair of molecules when they are close together.

So, we have a system of elements now (σ's, spins, small magnets, particles, whatever) which is mathematically isomorphic to a system of hard core particles on a lattice, which have a short-range attractive force between them, and where we could have any number of particles.

The particles number is something that can change.

So it is like a system of molecules where they can come in and go out of the system.

It doesn't have a definite number of molecules. Their number itself is a variable.

There is short-range growing negative potential, that is short-range attraction.

Yet, there is very short-range infinite repulsion.

We cannot put two elements on the same site.

But when they get a little bit apart, they are slightly attractive.

That is exactly what we need for liquid-gas phase transition.

Remember the two conditions spelled out above.

Now let's add something else.

By the way, does this system have a chemical potential?

In a chemical potential the energy is stored in just having one particle.

That is what it is.

The chemical potential is the energy in having a particle present.

Just by virtue of having the particle, if there is an energy that wouldn't be there if the particle weren't there, that is called the chemical potential.

So the answer is yes.

Just having a particle, with no other particle around, gives us an energy 8j.

Therefore there is a chemical potential.

But we want to be able to vary that chemical potential, without varying j incidentally.

We would like to hold j fixed, and vary the chemical potential separately from it.

That is easy to do.

Put a magnetic field in.

From equation (14), this takes us back to equation (2).

But, since we want the energy to be lower when σ_n is negative now, i.e.. when there is no particle, we modify it a little bit.

The sign in front of the second sum becomes plus.

The energy of the whole system is

$$E = -j \sum_{links} \sigma_n \sigma_m + \sum_{sites} h\sigma_n \tag{16}$$

How much does this give us when we add a particle to the brew?

At the n-th site, σ_n starts down, that is -1, when there is no particle.

When we introduce a particle, σ_n switches to +1.

So it jumps two units.

It gives us two units worth of energy for every spin which is flipped up, in other words for every particle that we put in.

That is 2h.

So for each isolated particle, we can add +2h.

The total energy needed to introduce an isolated particle is

$$8j + 2h \tag{17}$$

Now we have everything we need to study the liquid-gas transition in a fluid of molecules.

From our Ising magnet model we have built a system that is equivalent to a collection of particles on a lattice.

It has the following three specifications:

1. It has a variable chemical potential. Notice how we vary it. We do it by varying the magnetic field. The magnetic field gave us an energy per particle.

2. It has the hard core billiard ball potential which is repulsive.

3. And it has a short-range attraction, the coefficient of which is $4j$ (going from $16j$ when the two particles are far apart, to $12j$ when they are separated by only one link).

We want to keep the molecular properties fixed, therefore we are not going to play with j.

That has to do with the molecular properties.

By design, it tells us what the potential energy between them is.

All we can vary is the chemical potential.

And that is the same thing in our model as varying h.

So the problem of the behavior of the molecules of a fluid, with the three specs listed above, has become exactly the same as the previous problem of a magnet plunged into an external field, which was summarized in figure 11.

They obey precisely the same model.

There is a temperature that we can vary.

Particles are at some temperature.

And there is a chemical potential that we can vary.

Now h is not exactly the chemical potential.

There is an offset by an amount 8j.

What about the particle density?

We haven't discussed it yet. So let's do it.

How many particles are there, on the average, at a particular point?

The answer is somewhere, first of all, between 0 and 1.

Of course the number of particle is either 0 or 1, but we are talking about the average.

The average number of particles at a site is

$$\frac{1+\overline{\sigma}}{2} \tag{18}$$

Let's see why that is true.

The number of particles itself is the random variable $(1 + \sigma)/2$.

Indeed, if there is a particle there, that means $\sigma = +1$.

One plus one is two, divided by two is one.

So the random variable has the right value.

And if there is no particle? $\sigma = -1$.

In that case $(1+\sigma)/2$ is 0. It is ok too.

And the average of $(1 + \sigma)/2$ is $(1 + \overline{\sigma})/2$.

Since the density ϱ of particles (here we talk about the average density, not about an random variable) is proportional to the average number of particles, with units properly chosen, we can write

$$\rho = \frac{1 + \overline{\sigma}}{2} \tag{19}$$

That is the density.

Notice that it is not $\overline{\sigma}$ itself, but is given by formula (19).

There is no subtle fact here.

It is just that we have naturally chosen our spins to take their values in $\{-1, +1\}$, rather than in $\{0, 1\}$. So, to produce 0 or 1 from the two possible values of σ, we have to use the function $(1 + \overline{\sigma})/2$.

So the density of the fluid is just 1/2 plus half the average magnetization.

Now we are all set to see what would happen if we were to vary the chemical potential, or equivalently h, while keeping the temperature fixed.

Instead of following the evolution of the magnetization $\overline{\sigma}$, we will follow the evolution of the fluid density $\varrho = (1 + \overline{\sigma})/2$, figure 17.

If we are above the critical temperature, the density ϱ varies continuously, with no jumps.

In a box of gas, at fixed temperature above T_c, as we decrease the chemical potential, i.e., as we lower the energy, as we pull it down, nothing very exciting happens when we cross the horizontal axis in figure 17.

The horizontal axis corresponds to $h = 0$, or to a chemical potential of $8j$, because it has the general form $8j + 2h$.

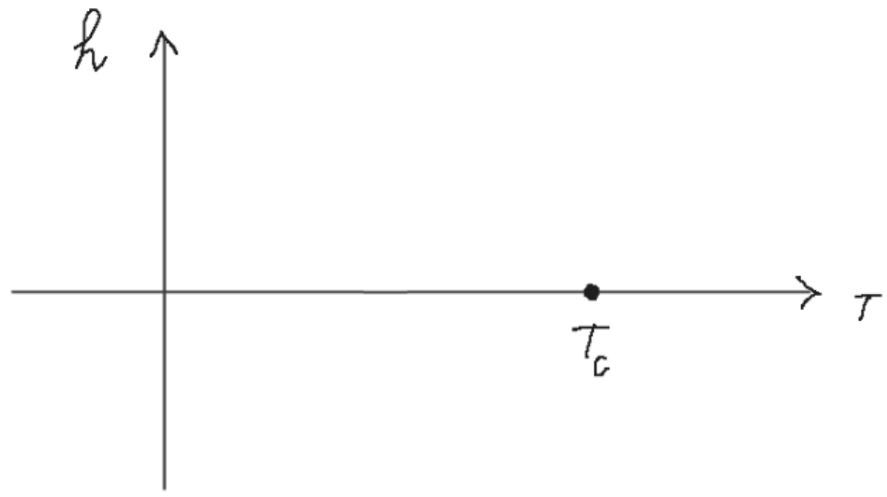


Figure 17: Space of variation of the fluid density $\varrho(h, T)$. This is exactly the same picture as in figure 7, except that the quantity $\overline{\sigma}$ we were looking at is now replaced by $\varrho = (1 + \overline{\sigma})/2$.

Below the critical point, by contrast, as we vary the chemical potential h, all of a sudden, when we cross $h = 0$, the density jumps.

Reducing h reduces the density, see equation (16).

And crossing the horizontal axis is the phenomenon of boiling.

For a long time, experimentalists had difficulties experimenting with the chemical potential at temperatures above T_c.

So they only knew of the phase transition.

What is the critical temperature for water?

It is not 100°C.

It is about 374°C.

But the chemical potential is linked to the density, therefore it is pressure dependent.

And it is hard to experiment above T_c.

Below the critical temperature, when we cross $h = 0$ from positive to negative is when boiling happens.

And going the other way, it is liquefaction.

Above the critical temperature, when we cross h = 0 no boiling nor liquefaction happen.

There is a smooth continuous transition of the density.

11.3 Questions/answers

Question: It is easy to understand how we change the magnetic potential.

But how do we change the *chemical potential*?

Answer: Really by varying the *density*.

We do whatever it takes to vary the density of the fluid.

For example we could create a difference of potential in energy between outside the box and inside the box, and we would literally suck particles in.

We would have to devise something in practice.

And it would not be very accurate.

Actually that is not the way we do it in practice.

What we do instead is vary the pressure, figure 18.

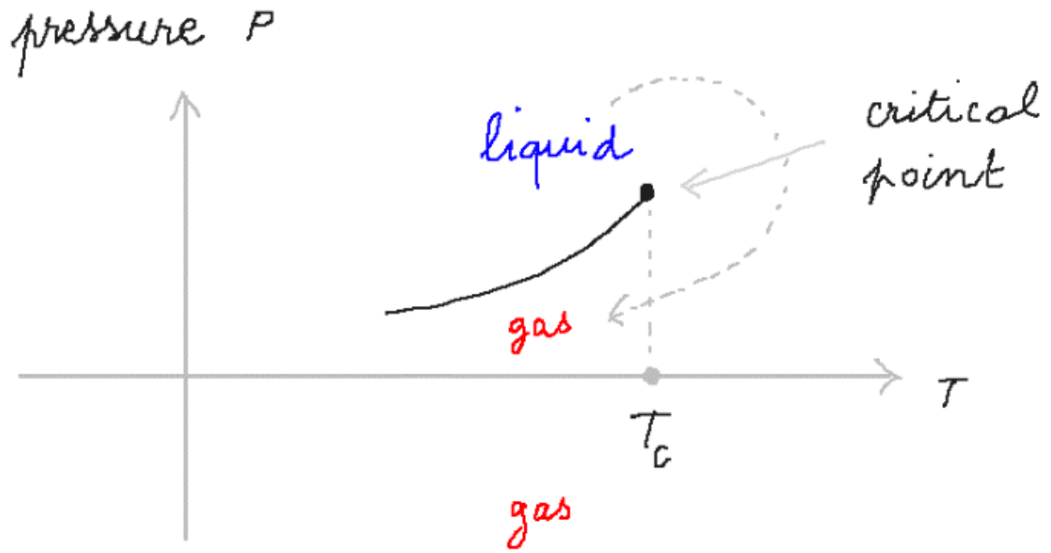


Figure 18: Simplified temperature/pressure diagram. The solid phase and the "triple point" are not shown. But we see the frontier between the liquid and gaseous phases. The dotted path shows how we can go from one to the other continuously, with no jump, by going around the "critical point".

Varying the pressure is the simplest solution to vary the chemical potential.

When we do it while the temperature remains below the critical one, we witness a sudden jump in the density of the fluid.

On the left of T_c, there is a frontier between liquid and gas.

Below it is gas, above it is liquid.

But notice, as before, that we can also go from gas to liquid, and back, without any jump, by going around the critical point.

Question: When you say you are changing the chemical potential by changing the density, are you basically changing the effective distance for Van der Waals potential?

Answer: Well, we don't change the properties of the molecules. Those stay the same.

But when we increase the density, we do shorten the average distance between molecules.

At some point the system realizes, so to speak, that a negative potential energy is there.

Then it tries to pull all the molecules together.

That is when it suddenly makes the transition.

But this description is valid only if we stay below the critical temperature.

If we go around it on its right we can go from the liquid phase to the gaseous phase without jump.

Question: In a liquid-gas transition, what is the relationship between the critical temperature and the boiling point?

Answer: The boiling point depends on the pressure.

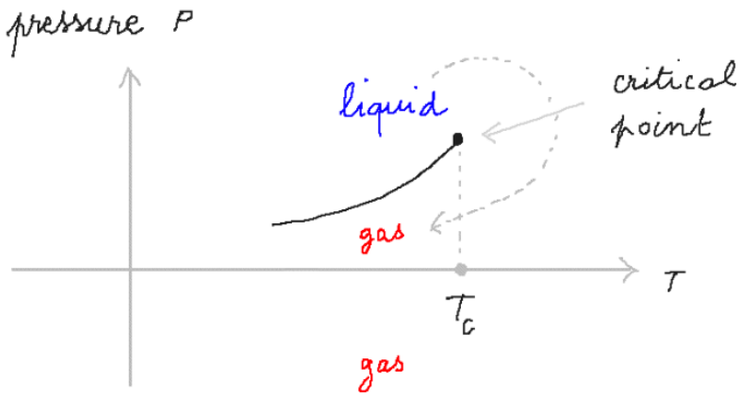
This in turn means it depends on the density.

In figure 17, we cannot see it, because the coordinates are T and h, and the frontier is flat and horizontal.

But the pressure is a function of the temperature and the chemical potential.

When we represent the average "magnetization" $\overline{\sigma}$ in the T and P coordinates, the frontier is no longer a flat interval along the horizontal axis.

It is the curve shown in figure 18.



Then, for each pressure below the ordinate of the critical point, we can read off at what temperature the transition from liquid to gas happens.

The lower the pressure, the lower the boiling point.

On the top of Mount Everest water boils at 68°C.

One of the very fascinating things is the properties of the system near the critical point.

There is a whole theory on the behavior of the system and there are lots of experiments to study what happens as we approach the critical point.

The phenomena happening there are all characterized by what is called *critical exponents*.

Every quantity that you can think of, that is interesting, in its dependance on temperature will generally go as

$$(T - T_c)^{\gamma} \tag{20}$$

where γ is some power.

And those powers are not 1's and 2's.

They are various odd numbers, rational, irrational or only experimental.

Of course they vary from one kind of phase transition to another, but they are exactly the same for the magnetic transitions and the liquid-gas transitions.

They fall into exactly the same class.

The behaviors at the critical points are rather insensitive to the details.

They depend on features that don't care whether we put the particles on a lattice.

They don't care whether it is nearest neighbor or second nearest neighbor.

They all behave the same way.

And the magnet and the fluid transitions are in the same class.

Question: Do the critical points form a continuum, or is it a discrete finite or countable collection of points?

Answer: The set of critical points is discrete, but it is not finite.

Question: How do you define gas versus liquid?

If when we go around the critical point, there is no transition.

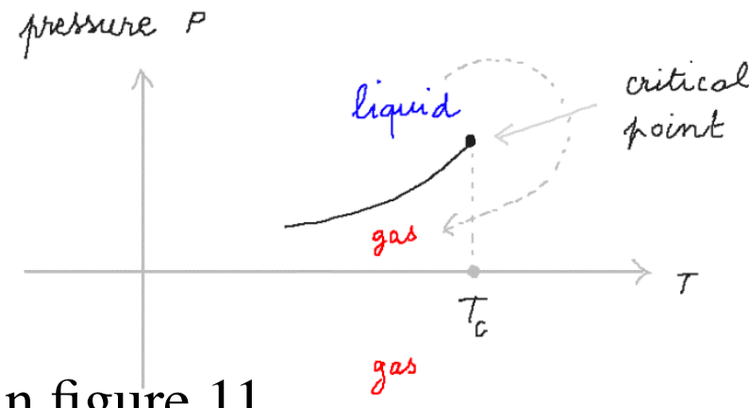
So it seems that they should be the same.

Answer: Indeed there is no sharply defined difference.

The density is obviously different in the lower part of figure 18 than in the upper part.

If we try to go across the transition frontier left of T_c, the density jumps.

If we go around the critical point, the density varies continuously.



The density of the fluid in figures 17 or 18 behaves like the magnetization in figure 11.

The precise connection is $\varrho = (1+ \overline{\sigma})/2$.

On the left of T_c there is a frontier where it jumps.

But we can go from a point below the frontier to a point above it by going around the critical point, i.e., passing to the right of T_c, then there is no jump.

Question: So liquid and gas coexist if we go around?

Answer: Well, they are not distinguishable.

There is a continuous transition from one to the other.

Question: What do we need to add into the theory to get liquid and gas together in a phase?

Answer: We need for instance gravitational potential energy to separate them.

Question: Could it also be surface tension?

Answer: No. In most of these kinds of analyses as above we think of an infinite volume of fluid.

So surface tension is not relevant here.

In an infinite volume of fluid, however, a gravitational field would create a liquid-gas physical separation.

They would be separated by a plane, as we are familiar with.

There would be an altitude-related chemical potential.

So varying the altitude varies the chemical potential.

And there is a transition at some height

Question: In this model, do we get a heat of vaporization as we go across that transition?

Answer: In general yes. No doubt

The problem is that, in the magnetization model, the energy near the transition frontier is exactly the same just above and just below it.

So it is a good question.

In the fluid case some refinements, at least in the analysis, and perhaps in the model, would have to be introduced.

Question: Does this have implications for superconductors and magnets, in particular for highly sensitive detectors, for example detectors capable of revealing the presence of dark matter?

Answer: A material going from ordinary conductor to superconductor does undergo a phase transition. It has similarities with what we studied.

Now concerning detectors, if you can adjust your superconductor to be very near the phase transition, then by a tiny change in some external parameter you can have it jump from superconductor to non-superconductor.

And that gives you the possibility of constructing highly sensitive detectors.

For instance some energy is deposited in the detector.

And even a small amount of energy can create a little local bubble of changed phase.

So it can be useful as a very sensitive kind of detector.

That is true in fact for any phase transition.

Think of cloud chambers for instance.

There are no detectors for dark matter since it does not exist!!!

Closing Discussion: the anthropic principle

In view of the wide interest in it, I would like to say a few words about the anthropic principle.

We will treat it as **questions/answers.**

Question: What is exactly the anthropic principle?

Answer: Can I give a crisp answer to that question?

No, because the anthropic principle means many things to many people.

But I can explain what is, in my view, a rational use of it

I won't claim that my explanation is right, but I'll say that in some circumstances it can give a rational explanation of some of the puzzling things we observe in the universe.

First we have to understand what *fine-tuning* means.

There is a small number of instances of fine-tuning in physics.

The classic case of an extreme fine-tuning is of course the *cosmological constant*.

The cosmological constant is very small.

In natural units, that is in Planck's units, it is 123 decimal places small.

What is the difference between being small and being fine-tuned?

They sound like the same thing but they are not the same thing.

I'll give you an example that I cooked up to explain this once to a bunch of condensed matter physicists who didn't know what fine-tuning was.

As soon as I talked about it, they said: "oh we know that."

But in truth they didn't make the difference with just being small.

It is a silly example.

And it is the only example I ever made up like this.

It doesn't have fish in it, but it does have submarines and also balloons.

Let's suppose that you are a certain kind of creature that can only exist at a certain altitude where the air was just at the right temperature – well, not only the right temperature, but the right temperature, the right density and so forth.

So you find yourself at that temperature and density and so forth.

You are in a blimp floating in the atmosphere, figure 1.

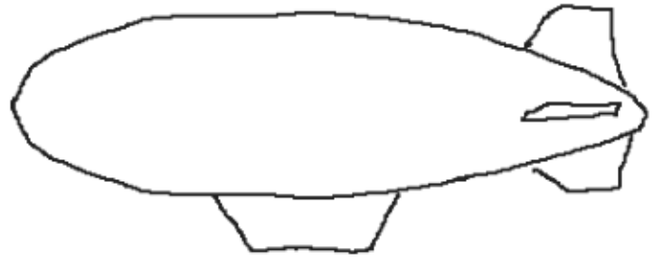


Figure 1: Blimp floating in the atmosphere, with creatures in the nacelle.

And you notice that you are neither rising nor falling.

Isn't that lucky?

You are at some stationary height.

And at that stationary height you have been there for several billion years, long enough to have evolved and evolved in just the right way to live at that temperature and that density.

Isn't it very lucky that this blimp happens to be exceptionally light so that it floats?

Yes, you may say, you are lucky.

The blimp happens to be light enough that it floats with you in it.

And there is a small number involved.

It is the density of the material inside the blimp.

In this case it is not horrendously small, but it is small.

It might be the density of hydrogen. Or it might be the density of helium.

Notice, though, that the mass of helium and the mass of hydrogen differ by a factor of 4.

So the density can vary over a wide range.

Still, you float somewhere.

Your helium will support you. Or your hydrogen will support you.

So you are ok.

Indeed, since the density of the air varies, you will find an equilibrium at some altitude.

And you will just hover at that altitude of equilibrium forever.

That is, you will stay there long enough to evolve your species and adapt it to the conditions.

That is an example of a small number.

The small number in this case being that by luck your system happens to have a *low density* – just low enough that you float in your blimp.

There is a small number, but there is no fine-tuning.

It didn't matter if it was helium or hydrogen.

And it would not matter if it was a mixture of helium and hydrogen.

You could change the parameters of the gas quite a lot, by several hundred percent, as long as it stays lighter than air.

So, if there is luck, it is that the density of the blimp is a small number.

But there was no exceptional fine-tuning of anything.

You didn't have to fine-tune in any detail the percentage of hydrogen and helium.

So that is the case when there is a small number.

Now let's talk about another situation.

The object in fig. 2 is not a blimp, it is a submarine, a yellow submarine.

We all live in a yellow submarine.

The submarine has no motor, it just floats somewhere at an intermediate depth.

If the submarine approached the sea bottom or the sea level, we would die.

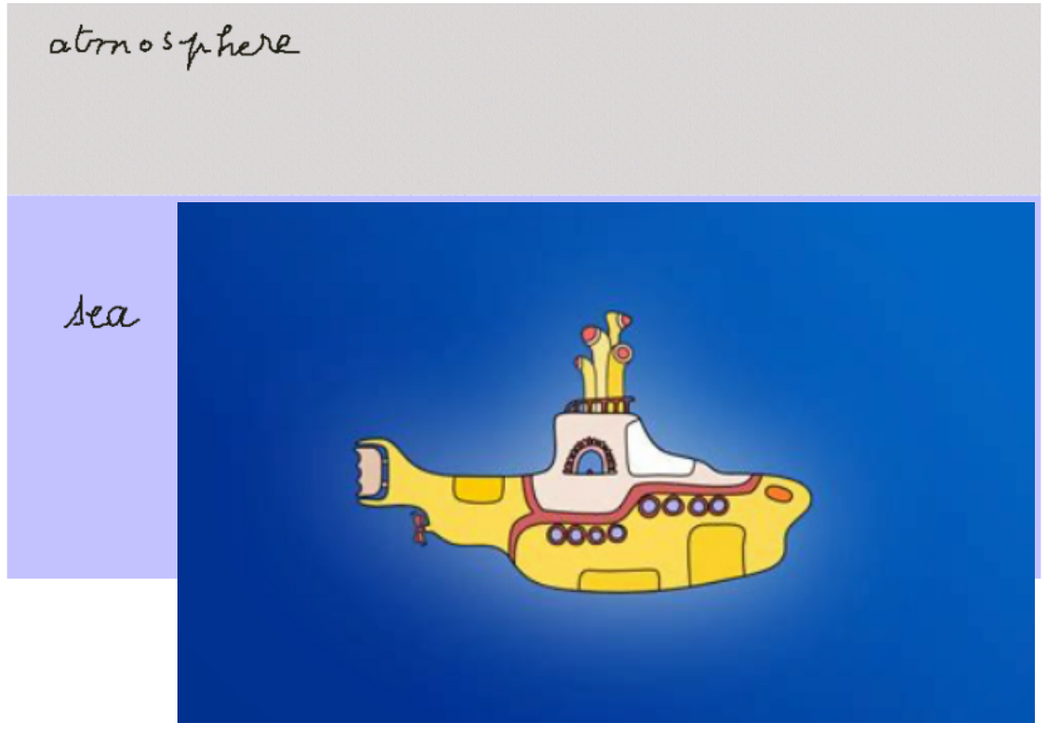


Figure 2: Submarine floating in the sea, with creatures inside.

Down at the bottom the submarine would be crushed by the pressure.

And near the sea level, we could not survive either, for the water is too acid there or there are harmful radiations or whatever.

For us to live, the submarine has to float in between.

Maybe it doesn't matter exactly how high we are.

But we have got to be neither near the top nor the bottom.

Our submarine has existed in that configuration long enough for us to evolve.

other words, for billions of years it has been hovering in the water without falling nor rising.

We know that because we are here.

Or these people who live in the submarine are here.

And they know that, if there was any significant tendency to fall, they would be dead.

If there was any tendency to rise, they would be dead.

Let's also assume for simplicity that the water has the same density from the bottom of the sea to the top of the sea, which is approximately true in reality.

What do these people in the yellow submarine conclude?

They conclude that whatever the submarine is made out of, the density of that submarine is finely tuned.

It is made out of some iron, together with some brass, maybe a little bit of concrete and a few other things.

Plus it is hollow in the interior.

And all that – with us inside – is such that its average density is very very close to that of water.

If it was any denser than water it would sink.

If it was any lighter than water it would rise.

And, what is more, it has been sitting there at that compromise position for billions of years.

From that you can conclude that, whatever the chemical composition is, it is very finely tuned.

If you were to change the ratio of iron to concrete by one part in a thousand or even one part in a million, you would be dead.

Remember this submarine has been sitting there for a long time.

Let's say 4 billion years, or 13 billion years.

For 13 billion years, it has been right somewhere in the middle of the vertical zone.

That is extraordinary!

It is extraordinary because, unlike in the case of the blimp in the atmosphere, there is no feed-back mechanism whereby if it started to sink something would tend to make it rise, and conversely.

If it was just a little bit too heavy it would slowly sink to the bottom.

If it was a little too light it would slowly rise.

And so somebody very fortunately has tuned the chemical composition, and the amount of hollowing out, and the weight of everything inside it, to one part in a fairly large number, to make sure that the submarine hovers.

Now the submariners are curious about it.

They want to know why is this true.

They have no explanation for it at all.

Their first reaction is: this must be something like the whatchamacallit, the dirigable, the Zeppelin, the blimp.

It the blimp goes too high, the air density goes down and the blimp sinks back.

When it goes too low, the air density increases and the blimp rises up again.

But the mariners do a few calculations, and there is no question of feed-back mechanism.

The density of water is too constant over the range from the bottom to the top.

There is no possibility to use that to explain the stability of the depth of the submarine.

They try all kinds of mathematical tricks, symmetry principle, whatever they have, to explain their still being here alive.

Nothing works.

And the reason nothing works is because this is one of these very fine-tuned things.

If you change the chemical composition a tiny bit this way or a tiny bit that way, the whole thing breaks down.

What kind of conclusion could they come to?

They do a little more calculations using whatever laws they know.

And they discover something very interesting about the nature of the fluid that they are in.

It has instability: it makes bubbles.

And the form of these bubbles is submarines.

At various places and times in the fluid, "pop", bubbles "nucleate", i.e. appear. And when they appear, they morph into submarines.

But the trouble is the submarines which they can nucleate into are generally not of the right kind.

In fact, as it turns out, the submarines can nucleate with different chemical compositions.

Some of them have a little more brass.

Some have a little more iron.

They can be more or less hollowed out.

And so there is just a whole huge variety of different species of submarines which nucleate.

The mariners figure that out.

They make these calculations.

And they discover that it is a property of the fluid they live in.

So they find this answer: there must be zillions of different kinds of submarines, a huge number of them.

And, moreover, they keep nucleating and nucleating and nucleating in this fluid.

So the fluid just keeps producing more and more and more of them, of every conceivable type, 10% iron and 90% brass, 20% iron and 80% brass, and all other possible compositions.

Most of them either sink to the bottom and lie crushed on the sea floor, or they rise to the top and rot at the surface.

You may even add, if you like, the feature that the creatures inside, despite their remarkable adaptive capabilities, don't have time to adapt to the extreme environments of the bottom or the top.

But a tiny fraction of the bubbles will nucleate with the right properties, just because everything that could happen will happen.

And so they are not particularly lucky .

They just are in a submarine which nucleated with exactly the right density to hover.

To talk about luck or not luck in this setting is a matter of point of view.

If we ask a zillion people to flip a coin a thousand times, most of them will produce approximately half heads and half tails, but some will turn 1000 heads.

We won't be particularly lucky to witness that.

But the ones who produced such a series by definition will.

If all of that really happens, and these mariners really have a genuine scientific reason to believe in submarine nucleation in their sea, and they could see that the bubbles could nucleate in a vast variety of different types of submarines, they would have what – I would say – is a reasonable explanation of what is going on:

there is a lot of stuff;

the sea is very big;

a long time has passed;

every conceivable kind of things have nucleated many many times; they are just in a kind where survival is possible.

That is the logic that, I think, the anthropic principle, when it is rational, is making use of.

That fine-tuning is a consequence of many many possibilities and many actual events taking place creating the submarine that we live in.

Some fraction of them will be liveable.

And we happen to be in one of them.

Then it becomes a non-mystery why we live in a fine-tuned universe.

That is a logic that I find personally acceptable.

There are alternative possible answers.

Another answer is that some intelligence designed the submarine with exactly the right properties.

There are some people who like to believe that.

Most of my scientific friends don't take that as an acceptable solution.

Are there other explanations?

At this time, within our current framework of thinking, there aren't any other possible answers to why certain things are fine-tuned.

What is it in our universe that is playing the role of the density of the submarine in the funny little illustration?

It is the cosmological constant.

If the cosmological constant is slightly negative, the universe will recollapse.

However, if the negative cosmological constant is really very small, that is like the density of the submarine being only a tiny bit higher than that of the water.

Then the universe will take a long time to recollapse in a very similar way that it would take a long time for the submarine to fall to the bottom of the sea if the difference between water density and submarine density was very small.

So if the cosmological constant, assuming it is negative, is small enough, the universe will last long enough without recollapsing for us to be here and survive and so forth.

Then we have the case that the cosmological constant could be positive.

In that case the universe doesn't crunch, but expands.

And it expands in an exponential way.

But if it expands too fast, then it prevents structures from forming.

If the universe was accelerating very rapidly, then the material in it would be carried along with the acceleration and would simply be incapable of contracting to galaxies and stars and so forth.

So there is some window analogous to the window of submarine density inside which the submarine could last for billions of years.

The cosmological constant is like the parameter density of the water minus density of the submarine.

And if it is small enough in absolute value, the universe can recollapse, or expand, sufficiently slowly that in either case we can be here.

The window of possible values of the cosmological constant allowing us to exist is extremely small.

It is extremely small because the time scale is so long.

It is the same reason the window is very small for the submarine.

To become what we are, we have to be able to survive for eons.

That implies that the possible imbalance in the cosmological constant, i.e. its difference from zero measured in natural units, has to be unbelievably tiny.

Now there is another sense in which the cosmological constant is like the density of the submarine.

The density of the submarine is made up out of a composite of densities of several materials in various proportions.

It is not just that some odd number happens to be small. It is a compound of numbers that are very finely adjusted.

The same is true of the cosmological constant in quantum field theory.

It gets contributions from many many different sources.

Basically it gets a contribution from every quantum field that exists.

Bosons give positive.

Fermions give negative.

Shift a little bit the mass of the particles that are associated with these quantum fields, and the cosmological constant that results from them shifts a little bit.

So the whole thing, all of the constants, all of the particles, all of the contributions, making up the cosmological constant have to balance very finely, exactly the same way the densities making up the submarine have to balance.

If there is a little too much iron, we sink and are crushed.

If there is a little too much brass, we go belly up.

Same thing with the cosmological constant: put in a extra fermion species in the theory and the cosmological constant gets too negative and the universe implodes;

put an extra boson, the cosmological constant gets too big and we are out of luck in a tiny fraction of a second.

Question: What is the meaning of "making a change that compensates"?

Is it just a different submarine, but still floating?

Answer: Yes, you could change some of your parameters in physics, such as various masses, coupling constants, and so forth; you could change one thing one way a little bit, which might increase the cosmological constant, and change something else another way a little bit, which might decrease the cosmological constant.

And you might be able to keep its value the same.

It is not that there is just only one chemical composition of the submarine that can survive.

There is a thin surface in some space of different configurations.

It is the same thing in our universe.

You can change things.

But you have to change them in a very very fine-tuned way.

You change one thing by this amount, you ought to change this other thing by that amount.

And it has to be adjusted to 123 decimal places!

Question: So will this sea water, that produces every kind of submarine, make various submarines that have the same suitable density, but look different than the one represented in figure 2?

So these correspond to alternate universes?

Answer: Well, they are *all* alternate universes.

But some of them are capable of sustaining themselves for long enough time, and have people.

Others are not.

The fact is, we know very little about this.

Our knowledge of the space of possibilities is extremely limited.

String theory seems to say that there are endless huge numbers of possibilities, analogous to every possible chemical compositions.

So that is a plus according to believers in string theory.

I don't know whether you regard that as a plus for string theory, or a plus for the anthropic principle, or a plus for cosmology.

But string theory does seem to fit together with the idea of the anthropic principle - on purpose!

And the fact is that there is no other known explanation.

For many years now people have been looking for explanations, without turning up anything convincing.

There is one more thing to stress about the submarine.

The submariners, as I said, might look for a mathematical explanation, some symmetry in the equations which would make the submarine density exactly equal to the surrounding water's.

They have this theory called "rope theory" or something.

They can't quite prove it, but they sense, if they only had the right idea, that rope theory would say that the submarines always come out with exactly the right density.

Some symmetry keeps water and submarine at the same density.

We feel that it is within reach. We don't quite understand it.

But we know that if we understood that theory better, it would tell us why the cosmological constant, or the density of submarines, has exactly the right value.

But then, to their horror and dismay, the submariners discover that their own submarine density doesn't exactly have the right value.

In fact their submarine is slowly sinking.

It really is slowly sinking.

How long does it take for it to sink?

More than 10 billion years.

Their finest measurements tell them: nope, things are not perfectly adjusted.

So we don't have to explain why they are perfectly adjusted, because they are not.

They are only adjusted to 123 decimal places.

And at the 124-th decimal place, they are sinking.

Question: Isn't this precision violating the uncertainty principle?

Answer: No. Any measurement taken alone can be as precise as we want.

Question: You say this is a use of the anthropic principle, to explain our universe, which you are comfortable with.

What are other uses that you are not comfortable with?

Answer: I have no idea whether there is an intelligence who created the universe.

I do not know, and I don't pretend to know.

Question: So you say that some people give that idea spin?

Answer: I suppose. Not many scientists do.

There are other uses of it which I find much less compelling.

There are all sorts of modest coincidences.

If any of them were not satisfied, they might be enough to make our universe inhabitable.

If we were to make a list of the elementary particles, we would find that almost all of the known ones are necessary for our own existence.

On the other hand, our understanding of particle physics would allow just about anyone of them to not be there, or other ones to be there.

For example, let's start thinking about the particles we are familiar with.

Photons are light.

Where would we be without photons?

There would be no light, no heat, no Sun.

Worse than that, there would be no photons jumping back and forth between the atomic nucleus and the electrons to hold them together.

No chemistry.

We would not be here.

What about no neutrinos?

If there were no neutrinos, there would be no nuclear reactions of the kind which created the elements out of the Big Bang.

So the existence of carbon and all these other elements that make ourselves would not be here without neutrinos.

What if there were no electrons?

With no electrons, it would be a clear disaster.

There is no question of that.

What about no quarks?

No quarks, then no protons or neutrons.

What if you changed the parameters a little bit?

For example it was a puzzle for many years, and it still is a puzzle: why is the proton lighter than the neutron?

In the early days of nuclear physics this was a great puzzle because the proton is charged.

We would have thought its electrostatic energy would have made it a little bit heavier than the neutron.

Neutrons and protons are twins.

They are almost exactly the same, except that the proton has a little bit of positive electric charge.

Ordinarily we would say that this increases its energy.

It takes energy to pull together a charge.

So the protons should have been a little bit heavier than the neutrons.

Yet, if we take the mass of the neutron as the unit, the respective masses are

$$\begin{aligned} \text{neutron} &= 1 \\ \text{proton} &= 0.99862349... \\ \text{electron} &= 0.00054386734... \end{aligned} \tag{21}$$

In fact it would have been a disaster if the proton had been heavier than the neutron, because proton would decay to neutron, instead of neutron decaying to proton.

If the proton decayed into neutron, there would be no hydrogen.

There would be nothing.

We would not be here.

So the little tiny difference between proton and neutron – which is indeed very small, and which happens to go in the wrong way to expectation –, is absolutely essential to our own existence.

All sorts of other things appear to be lucky.

Notice that this meditation about the luck we have to be in a habitable universe, can also be taken up by any individual – you or me for instance – wondering that he is lucky to be alive considering all the chances he had not to, for instance if the spermatozoid that lead to his birth didn't hit the ovule.

What would happen if the electromagnetic coupling constant, the fine-structure constant which is 1/137.035999139..., what would happen if it was 1/10 instead?

Then the atom would be very very tightly bound and ordinary chemistry would go awry.

And what would happen if the electron were heavier than the proton?

They would be very deep inside the center of the atom instead of having their present cloud structure.

So you can look at physics as we know it, which means largely particle physics, and say: there is so many different things which could have changed by 10 or 20% or even 1% here and there, and any one of them would have destroyed our existence.

I can easily imagine 1% accidents. I can even imagine ten 1% accidents, so it gets a little harder.

Are those accidents sufficient to really push us in the anthropic direction?

I am not going to answer that; I don't have an answer to that.

It is largely a matter of taste.

But the accidents of 1 part to 10^{123} or one part to 10^{30}, in another case having to do with the Higgs boson, those are not the kind of accidents that you can imagine are accidental.

So I think, if it wasn't for the cosmological constant and the other comparable fine-tunings, we would not be considering the anthropic explanation.

Most serious physicists would be rejecting the anthropic explanation, maybe not completely but with less hesitation.

For the time being, it is a working explanation for want of better one.

That doesn't mean every physicist approves of it.

There is just no other around.

Question: Isn't this like the inflationary paradigm?

It is the best explanation available, although many people say basically most inflationary models don't work?

Answer: Who says that?

– Paul J. Steinhardt, Abraham Loeb, and Anna Ijjas

– Ok. These are contrarians.

The vast vast vast overwhelming majority of cosmologists and physicists agree with the cosmic inflation theory.

– They say that the data from Planck satellite, that were presented in 2013, seem to support better another model.

– Sure, people make up a lot of wild models.

Then we sift through them and keep only the sensible ones, provided they are not in contradiction with observations of course.

Plank satellite data don't knock out the inflation theory.

As early as the 14th century, when modern scientific thinking, resting on experimental observation, had finally taken off, William of Ockham recommended, among the various consistent explanations of a phenomenon, to retain the simplest.

This principle is called "Ockham's razor".

There are people with axes to grind.

Physics is not devoid of politics.

Even in our community, there are people who get very highly politicized about their own ideas.

Going around saying things like Planck data do not confirm the inflation theory, I think, is irresponsible.

Question: Couldn't the cosmological constant change over time?

Answer: There is no good theory in which it varies gradually.

It can vary in a jump, if one of these submarines nucleates with a different value of its density.

But there is no good theory in which it varies smoothly over such a long time scale, slow enough to be consistent with what we observe.

There is a theory with a very slowly varying cosmological constant, but it doesn't seem to fit with anything we know.

In any case, if we had a slowly varying version of it, it would require fine-tuning of several parameters.

Not only the vacuum potential energy but several of its derivatives would have to be fine-tuned.

So it is not a very satisfactory explanation.

The truth is that nobody at present knows where the theory of the birth of the universe is going to go.

There are contrarians.

There are people who have a political agenda.

I don't.

I'm not the inventor of the cosmic inflation theory.

I don't have a big stake in it.

I'm not going to earn the Nobel prize if it is confirmed.

The cosmic inflation theory is highly confirmed.

Somebody has earned the Nobel prize for it.

The logic of the cosmic inflation theory, for me, is extremely compelling.

And I honestly cannot understand these people who say that Planck satellite's data is ruling it out.

Question: Does inflation theory imply nucleation of universes?

Answer: It doesn't fit with it very well.

Go back to the submarines.

All the anthropic principle itself says is that we live in a submarine which is of the right type for us to live in.

Which is cause, and which is effect?

Do we explain the fact that we live in the right kind of submarine by the fact that we are here?

$$\textit{We are here}. \Rightarrow \textit{We live in the right kind of submarine}.$$

Or do we explain the fact that we are here by the fact that we live in the right kind of submarine?

$$\textit{We live in the right kind of submarine}. \Rightarrow \textit{We are here}.$$

I prefer the latter.

That is I prefer to explain why we are here as a consequence of there being enough right kind of submarines.

The former seems scientifically backward.

It would put us again at the center of Nature to some extent.

In other words, we misuse the anthropic principle, in my view, when we start to try explaining why the laws of Nature have to be the way they are because we are here.

It seems to me much less egregious to say that the laws of Nature and the possible environments are highly variable – which theories tend to suggest anyway –, and that we live in a temperate zone.

We don't live in a inhospitable universe for the same reason we don't live in Antartica.

It is just too damn cold there.

The real question is: how are we going to find out?

The agenda of those who hate the idea and those who love the idea should be identical.

Take it seriously enough to either kill it or make it into science.

And the fact is that it looks very very hard to do either.

It looks extremely hard to find signature of these kinds of things in data, that will either kill or confirm the hypothesis of many universes.

That is the real problem.

The problem is not that it is bad philosophy, or that it is religious, or that it corrupts the youth.

One of my friends thinks that the anthropic principle is a terrible thing to propose because young people will have their minds deformed by it and they will stop thinking about other explanations.

Other people think it is bad because it enables the religious right to say: ha, you see, the world was created with just the right properties.

They are political reasons.

But none of this is science.

The scientific question is: is it true, or is it not true?

And how are we going to find confirming or destroying evidence?

Question: We want to observe other submarines nucleating?

Answer: Right.

So, if you are very lucky – or unlucky, as the case may be – you might bump into another submarine.

You would be lucky if it doesn't kill you and you thus find out there are other submarines out there, confirming that you win the Nobel prize.

Or it could be bad, the other submarine could make a hole in your submarine and then you would sink to the bottom.

In principle there could be collisions between regions of space with different properties.

Computation seems to say that it is extremely rare.

Basically computation seems to say that the ocean is exceedingly big.

But the nucleation of submarine is very rare.

It doesn’t happen very often in time or space.

So most submarines will last a long time without banging into another.

We find no evidence in the cosmic microwave background for collisions of this kind.

The only question is then: what other possible kind of signatures could there be for the physics that goes into this kind of theory?

And it is hard.

Nobody has a really good idea.

That is the worst aspect of the anthropic principle: it looks overwhelmingly hard to find convincing data for it or against it

Thanks everyone for listening to me present my way of thinking about theoretical physics in many contexts!!