

Lesson 4 : The Boltzman distribution

Notes from Prof. Susskind video lectures publicly available on YouTube

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 chapter.

We look 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 sub-system 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 sub-systems. 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

$$P_i = \frac{n_i}{N} \quad (1)$$

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

$$S = - \sum_i P_i \log P_i \quad (2)$$

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

$$A = \frac{N!}{n_1! n_2! n_3! \dots} \quad (3)$$

Then we turned to the problem of finding the collection

(n_1, n_2, n_3, \dots) 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). 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 chapter with the following mathematical problem :

Maximize the function $S = -\sum_i P_i \log P_i$ under the two constraints

$$\begin{aligned} \sum_i P_i &= 1 \\ \sum_i P_i E_i &= E \end{aligned} \tag{4}$$

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 chapter. It is 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 constraints $G_j(x_i) = 0$, 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.

$$F(P_i) = \sum_i P_i \log P_i \quad (5)$$

And there are two constraints

$$\begin{aligned} G_1(P_i) &= \sum_i P_i - 1 = 0 \\ G_2(P_i) &= \sum_i P_i E_i - E = 0 \end{aligned} \quad (6)$$

Remember that F' is defined as

$$F' = F + \sum_j \lambda_j G_j \quad (7)$$

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.

Lagrange multipliers α and β

Let's continue working on the Lagrange multipliers. In our problem, equation (7) is rewritten

$$F' = F + \alpha \left[\sum_i P_i - 1 \right] + \beta \left[\sum_i P_i E_i - E \right] \quad (8)$$

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 $-E$ in the second constraint, in equation (8), won't play a role. We can omit them, and tackle the problem of minimizing

$$\sum_i P_i \log P_i + \alpha \sum_i P_i + \beta \sum_i P_i E_i \quad (9)$$

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

$$\frac{\partial}{\partial P_7} \left(P_7 \log P_7 + \alpha P_7 + \beta P_7 E_7 \right) \quad (10)$$

Noting that $\frac{\partial x \log x}{\partial x} = \log x + 1$, this is simply

$$\log P_7 + 1 + \alpha + \beta E_7 \quad (11)$$

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

$$\log P_i + 1 + \alpha + \beta E_i = 0 \quad (12)$$

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 know α and β , we will know P_i

$$\log P_i = -(1 + \alpha) - \beta E_i \quad (13)$$

or

$$P_i = e^{-(1+\alpha)} e^{-\beta E_i} \quad (14)$$

The first exponential is just a number. It has a standard notation in statistical mechanics. The exponential without the minus sign is denoted

$$e^{(1+\alpha)} = Z \quad (15)$$

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

$$P_i = \frac{1}{Z} e^{-\beta E_i} \quad (16)$$

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.

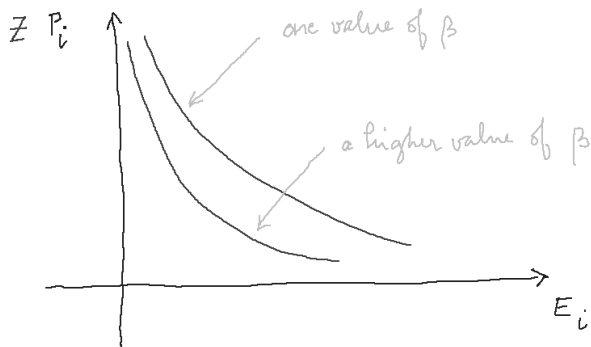


Figure 2 : Two distributions of probability, rescaled with Z , for two different values of β .

The higher β is, the more narrow and peaked the probabi-

lity 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. 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. From equation (16), we write it

$$\frac{1}{Z} \sum_i e^{-\beta E_i} = 1 \quad (17)$$

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

$$Z = \sum_i e^{-\beta E_i} \quad (18)$$

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

$$Z(\beta) = \sum_i e^{-\beta E_i} \quad (19)$$

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.

$Z(\beta)$ is called the *partition function*.

It is a function of β . And β is the inverse temperature, but we don't know that yet.

Beginning to use the partition function $Z(\beta)$

Now we turn to the second constraint.

$$\sum_i P_i E_i - E = 0 \quad (20)$$

Using equation (16), we can rewrite it

$$\sum_i \frac{1}{Z} e^{-\beta E_i} E_i = E \quad (21)$$

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

$$Z(\beta) = \sum_i e^{-\beta E_i} \quad (22)$$

That is of course the partition function. And let's differentiate both sides with respect to β .

$$\frac{\partial Z}{\partial \beta} = \sum_i -E_i e^{-\beta E_i} \quad (23)$$

We get something which is a piece of equation (21). Let's divide by Z on both sides.

$$\frac{1}{Z} \frac{\partial Z}{\partial \beta} = - \sum_i \frac{1}{Z} E_i e^{-\beta E_i} \quad (24)$$

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).

So we arrive at

$$E = - \frac{1}{Z} \frac{\partial Z}{\partial \beta} \quad (25)$$

This is a remarkable formula for the energy in terms of the derivative of the partition function with respect to β .

What is this 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

$$E(\beta) = -\frac{\partial \log Z}{\partial \beta} \quad (26)$$

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. And it is a powerful formula.

Questions / answers session

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 energy¹. Equivalently *each box* has a fixed average energy E .

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 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.

1. 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.

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.

Q. : 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?

A. : 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.

Q. : 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?

A. : 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 energy².

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.

Q. : We are focussing on β . Hasn't α an important role too? Why did it disappear?

A. : 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

2. Remember that E is a given in the problem. But then it is also $\sum_i P_i E_i$.

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, to be 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.

Q. : 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.

A. : 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 : the partition function, which turns out to be a central tool ; the relationship between E and β ; the fact that β is the inverse of the temperature as we defined it in chapter 2 – 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.

Q. : In equation (26), where we write $E(\beta) = -\partial \log Z / \partial \beta$, should not we write $E(\alpha, \beta) = -\partial \log Z / \partial \beta$?

A. : The independent variables in our problem are initially the E_i 's and E .

We introduced α and β because we use the method of Lagrange multipliers to figure out the entropy. Furthermore, α and β turn out to have quite interesting significance. 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

$$E(\beta) = -\frac{\partial \log Z}{\partial \beta} \quad (29)$$

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

$$E(\beta) = -\frac{\partial \log Z(\beta)}{\partial \beta} \quad (31)$$

So we see that it would be meaningless to write $E(\alpha, \beta)$.

The new independent variable that we shall mostly use will be β . And that makes physical sense because it is like using the temperature³ as the independent variable.

So our next topic is the relationship between the energy⁴ at thermal equilibrium and temperature.

Q. : We don't pay attention to the E_i 's anymore? They still drive α and β , don't they?

A. : 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 mo-

3. Which is the inverse of β .

4. Here of course we mean the average energy. Even though at thermal equilibrium the system still goes randomly through the E_i , it is mostly E that we are concerned with. However, at high temperature, the collection of E_i 's is widely spread. That is also an important physical fact not to be overlooked. A system in thermal equilibrium with a heat bath goes randomly through a collection of energy levels.

ment the energy levels are fixed.

Q. : But since β depends ultimately on the E_i 's, how can we change β if we don't change the E_i 's?

A. : 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 and β , the two Lagrange multipliers⁵, 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 it is a lousy way. Or, let me say that differently : it is a great way! Just not helpful.

It is a little hard to measure the energy. It is much easier

5. It is a slight abuse of language to call Z one of the Lagrange multipliers. In truth it is α of course.

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.

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 :

$$\begin{aligned} S &= - \sum_i P_i \log P_i \\ &= - \sum_i \frac{1}{Z} e^{-\beta E_i} [-\beta E_i - \log Z] \\ &= \sum_i \frac{1}{Z} e^{-\beta E_i} [\beta E_i + \log Z] \end{aligned} \quad (32)$$

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 (32) becomes

$$S = \beta E + \frac{1}{Z} \log Z \sum_i e^{-\beta E_i} \quad (33)$$

Now, what is $\sum_i e^{-\beta E_i}$? It is Z . We can see that from the expression of P_i as $e^{-\beta E_i}/Z$ and the first constraint that all the probabilities add up to 1.

Thus, continuing on autopilot so to speak, equation (33) simplifies further into

$$S = \beta E + \log Z \quad (34)$$

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 over and over in statistical mechanics. It is related, by a factor of temperature, to something called the *Helmholtz free energy*⁶. 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.

6. named after Hermann von Helmholtz (1821 - 1894), German physician and physicist. Helmholtz free energy is usually denoted A , and we have $A = -kT \log Z$, if we keep Boltzmann constant in the expression of the temperature.

Temperature

To find how to express temperature with our new partition function we have to remember the definition of temperature we gave in chapter 2.

The change in energy when we change the entropy by one unit is called the temperature.

$$dE = TdS \quad (35)$$

or

$$\frac{dE}{dS} = T \quad (36)$$

That is the way we defined temperature in the second lesson.

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.

Many readers may be more familiar with the equivalent formula

$$dS = \frac{dE}{T} \quad (37)$$

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 (34) 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 $d \log Z/dE$. So we have

$$\frac{dS}{dE} = \beta + E \frac{d\beta}{dE} + \frac{d \log Z}{dE} \quad (38)$$

The third term on the right-hand side can be rewritten⁷

$$\frac{d \log Z}{dE} = \frac{\partial \log Z}{\partial \beta} \frac{d\beta}{dE}$$

So equation (38) becomes

$$\frac{dS}{dE} = \beta + \frac{d\beta}{dE} \left[E + \frac{\partial \log Z}{\partial \beta} \right] \quad (39)$$

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 $-\partial \log Z / \partial \beta$.

7. The notation with partial derivatives is not fundamental. It just reminds us that Z depends not only on β but on a collection of other variables, for instance the E_i 's, which play here the role of pure parameters. We could use partial derivatives for the other quantities in equation (38) as well. But we stick to customary notations.

Therefore we have

$$\frac{dS}{dE} = \beta \quad (40)$$

And since the left-hand side is also the inverse temperature, we finally proved that

$$T = \frac{1}{\beta} \quad (41)$$

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.

Questions / answers session (2)

Q. : Doesn't S depend not only on E but also on β ?

A. : 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 (34)

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 (34) 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.

Q. : In physics manuals we usually see the relation between T and β written as β equals 1 over kT . What is the difference with equation (41) ?

A. : 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

$$kT = \frac{1}{\beta} \quad (42)$$

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 $1/\beta$.

Q. : 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 ?

A. : 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

$$\begin{aligned} (a) \quad P_i &= \frac{1}{Z} e^{-\beta E_i} \\ (b) \quad Z &= \sum_i e^{-\beta E_i} \\ (c) \quad E &= -\frac{\partial \log Z}{\partial \beta} \\ (d) \quad T &= \frac{1}{\beta} \\ (e) \quad S &= \beta E + \log Z \end{aligned} \tag{43}$$

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 will study the ideal gas.

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.

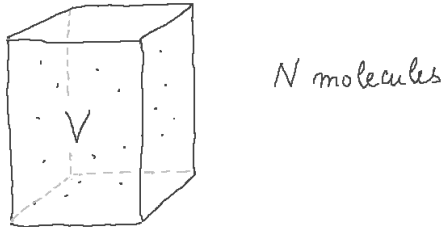


Figure 3 : Box of gas, of volume V . The molecules have kinetic energy, but don't interact with each other.

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.

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 $x_1, x_2 \dots$ all the way up to x_{3N} . And for the momenta, $p_1, p_2 \dots$ 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 integrals⁸. We are not going to go through that in detail. The reader who would like to brush his or her knowledge of integral calculus can go back to the volume 1, in the collection *The Theoretical Minimum*, where a whole interlude is devoted to it.

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

8. Remember that "integral" is a shorter name for "integral sum", and that the sign \int is the letter s written in an elongated form.

$$\sum_i F(i) \tag{44}$$

Now, when the states form a $6N$ -dimensional continuum, for any interesting function f of the states, the analogous integral is denoted

$$\int d^{3N}x \, d^{3N}p \, f(x,p) \tag{45}$$

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 (43b) 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 is 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

$$\frac{1}{2m} [p_1^2 + p_2^2 + p_3^2] \tag{46}$$

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

$$\sum_{n=1}^{3N} \frac{p_n^2}{2m} \quad (47)$$

where n is the index labelling the positions or the momenta coordinates. n runs from 1 to $3N$.

Finally the partition function is

$$Z = \int d^{3N}x \, d^{3N}p \, e^{-\beta \sum_{n=1}^{3N} \frac{p_n^2}{2m}} \quad (48)$$

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 of course, for instance the pressure.

Let's assume that all the particles have the same mass m . We can clean a little bit equation (48) :

$$Z = \int d^{3N}x \, d^{3N}p \, e^{-\frac{\beta}{2m} \sum_{n=1}^{3N} p_n^2} \quad (49)$$

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 (49) is, we can spell out what is beneath the integral sign like this

$$d^{3N}x \quad dp_1 e^{-\frac{\beta}{2m}p_1^2} \quad dp_2 e^{-\frac{\beta}{2m}p_2^2} \quad \dots \quad dp_{3N} e^{-\frac{\beta}{2m}p_{3N}^2} \quad (50)$$

In other words, next to the differential $d^{3N}x$, we have a product of $3N$ identical differential terms $dp_n e^{-\frac{\beta}{2m}p_n^2}$. So it is simply the $3N$ -th power of one of them.

Now since the integration in equation (49) 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 (49) simplifies into

$$Z = \left[\int d^{3N}x \right] \left[\int dp \quad e^{-\frac{\beta}{2m}p^2} \right]^{3N} \quad (51)$$

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

$$Z = V^N \left[\int dp \quad e^{-\frac{\beta}{2m}p^2} \right]^{3N} \quad (52)$$

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. Consider a box with only two molecules in it. And let's look at two

possible states, figure 4.

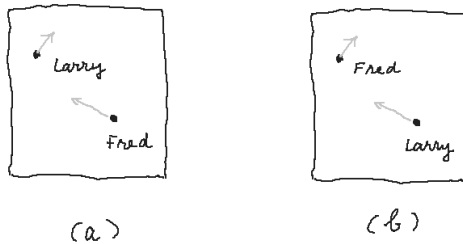


Figure 4 : Two states, where molecule Larry and molecule Fred are simply interchanged.

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 (52), 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

$$Z = \frac{V^N}{N!} \left[\int dp e^{-\frac{\beta}{2m}p^2} \right]^{3N} \quad (53)$$

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 familiar reader 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

$$q^2 = \frac{\beta}{2m} p^2 \quad (54)$$

The differentials satisfy

$$dp = \sqrt{\frac{2m}{\beta}} dq \quad (55)$$

Then we can rewrite the integral

$$\int dp e^{-\frac{\beta}{2m}p^2} = \sqrt{\frac{2m}{\beta}} \int dq e^{-q^2} \quad (56)$$

The term $\int dq e^{-q^2}$ is now really just a number. It doesn't depend on any parameter. Its value is $\sqrt{\pi}$ ⁹.

Finally we find that the partition function for an ideal gas made of N particles is

$$Z = \frac{V^N}{N!} \left[\frac{2m\pi}{\beta} \right]^{\frac{3N}{2}} \quad (57)$$

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

9. 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 becomes

$$\int_0^{2\pi} \int_0^\infty e^{-r^2} r dr d\theta$$

Now it is elementary to show that this is equal to π .

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 6.02×10^{23} molecules¹⁰, in other words a huge number indeed.

Then $N!$ is even a huger number, and we can approximate it by Stirling's formula.

$$\frac{V^N}{N!} \approx \frac{V^N}{\sqrt{2\pi N} N^N e^{-N}} \quad (58)$$

Now N/V is ρ . So we can rewrite the partition function as

$$Z = \frac{1}{\sqrt{2\pi N}} \left[\frac{e}{\rho} \right]^N \left[\frac{2m\pi}{\beta} \right]^{\frac{3N}{2}} \quad (59)$$

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

$$Z = \left[\frac{e}{\rho} \right]^N \left[\frac{2m\pi}{\beta} \right]^{\frac{3N}{2}} \quad (60)$$

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 .

10. It is the number of molecules in 22.4 liters of any gas under normal conditions of temperature and pressure. That is when we are in California or on the French Riviera.

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 (60) 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 (43c) for the average energy E . So let's calculate $\log Z$.

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 differentiate¹¹.

Partition functions are products typically. It is easier to differentiate their logarithm. So let's take the logarithm of Z defined by equation (60)

$$\log Z = N \left\{ \frac{3}{2} \log \frac{2m\pi}{\beta} - \log \rho + 1 \right\} \quad (61)$$

When we differentiate, the constants don't play any role.

11. And if we differentiate because our objective is to find a stationary point, considering that logarithm is monotonic, it is ok.

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 (61) as

$$\log Z = -\frac{3N}{2} \log \beta + \text{constants} \quad (62)$$

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 (56). 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 (43e). But even more interesting for the moment and simpler is to calculate the total energy of the system, equation (43c).

All we have to do is differentiate $\log Z$, given by equation (62), with respect to β , and a minus sign. That gives us

$$-\frac{\partial \log Z}{\partial \beta} = \frac{3N}{2} \frac{1}{\beta} \quad (63)$$

The left-hand side is the average energy. And on the right-hand side, $1/\beta$ is simply the temperature. So we reach

$$E = \frac{3N}{2} T \quad (63)$$

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

$$\frac{E}{N} = \frac{3}{2} T \quad (64)$$

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

$$\frac{E}{N} = \frac{3}{2} k_B T \quad (65)$$

k_B is of course a very small number¹². 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

12. $k_B = 1.4... \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1}$. When there is no ambiguity, we denote the Boltzmann constant simply k . And formula (65) then reads $E_{particle} = 3/2 kT$.

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 $p_x^2 + p_y^2 + p_z^2$, 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 $E = N \times 3/2 T$ 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

$$Z = \int d^{3N}x \ d^{3N}p \ e^{-\frac{\beta}{2m} \sum_n (p_{x,n}^2 + p_{y,n}^2 + p_{z,n}^2)} \ e^{-\beta \sum_n m g y_n} \quad (66)$$

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 (53), now becomes

$$Z = \frac{A^N}{N!} \left[\int dp e^{-\frac{\beta}{2m}p^2} \right]^{3N} \left[\int e^{-\beta mgy} dy \right]^N \quad (67)$$

The range of integration of the second integral is from 0 to the height of the box.

Exercise 1 : In equation (67), calculate the two integrals. Then calculate the average energy 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.

Let's solve exercise 1 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 (57). So we have

$$Z = \frac{A^N}{N!} \left[\frac{2m\pi}{\beta} \right]^{\frac{3N}{2}} \left[\frac{1}{\beta mg} \right]^N \quad (68)$$

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

$$\log Z = -\frac{3N}{2} \log \beta - N \log \beta + \text{constants} \quad (69)$$

Then we take the derivative with respect to β , with a minus sign. This yields

$$E = \frac{3}{2}NT + NT = \frac{5}{2}NT \quad (70)$$

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, $E_{particle} = 3/2 T$, 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.

Those are a couple of examples of very simple calculation of partition functions.

Exercise 2 : Compute the entropy of the box in the case of no potential energy from gravitation.

Compute it too in the case of potential energy from gravitation.

Questions / answers session (3)

Q. : All the equations we derived in this lesson rest on the Boltzmann distribution, is that right ?

A. : 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 (43a), which is reproduced below

$$P_i = \frac{1}{Z} e^{-\beta E_i} \quad (71)$$

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 : Whaaooow this is an interesting equation !

Statistical mechanics is a very curious subject.

Q. : You introduced gravity potential energy. Will we be able to apply this at some point to collapsing stars?

A. : 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 because there is a lot of details.

Q. : So in a collapsing black hole, the entropy will increase?

A. : Yes, the entropy does go up to a point where it reaches the Bekenstein¹³ entropy. Then it cannot go higher than that – not unless you add more energy. For a given amount of energy and a given volume, that is it.

Q. : How can we use the partition function to calculate fluctuations?

A. : Let's look at the fluctuations in energy for example. That is not a subject I intended to treat in this lesson, but let me tell you what the trick is.

If you want to calculate the fluctuations in a quantity, what

13. Jacob David Bekenstein (1947 - 2015), Israeli-American theoretical physicist.

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 twice the partition function 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 chapter.