Statistical Mechanics: A Bridge Too Far?

Throughout physical chemistry so far, you've learnt about two main areas of knowledge. Most physical chemistry education starts with a discussion of the things that concern **macroscopic** quantities like enthalpy, entropy, and heat capacity - you learn about these in thermodynamics, for example[1]. Normally then you proceed to learn about the **microscopic** - things that concern single molecules, like wavefunctions, or single particle energy levels. An obvious question for the inquisitive student is to ask '*how do we link these together?*' . After all, it would be a nice confirmation that all the quantum theory isn't just nonsense if it could predict measurable, macroscopic things that we can more easily see. In fact, if quantum theory is any good, then it must be able to do this! The way that you bridge together the microscopic and macroscopic worlds is using **statistical mechanics**.

> [1]One could (and I do) make a case that this is a bit backward, and learning things from a molecular level 'up' makes more sense than learning about phenomena but only explaining them in future years when you know the quantum/statistical mechanics.

To illustrate how we do this, let's imagine we want to find out how we can calculate the internal energy of a load of atoms in a container, as a nice example case.

What do we need to know?

So, we have a load of atoms. We need to know how many we have, and we can do this[2], and let's call this number N. We also need to know the possible energy states that each atom could occupy, and we can do this using quantum mechanics (the atoms are just particles in a box). Let's call the ith energy state ϵ_i, and the number of atoms in that state n_i.

> [2]Count them, or weigh them, or something.

Knowing this seems like a good starting point, but what we crucially do *not* know, is **how the atoms are partitioned amongst the possible energy states?**. After all, our internal energy would be very different if every atom was in the ground state, or every atom in an excited state. We need to know how the available population is split up amongst the different states. How it is **partitioned** amongst the states. At risk of labouring the point, we need to know how many are in state 1, how many are in state 2, and so on. If we knew this, we could just add up all the energies to find out our total internal energy!

The Most Probable Configuration

However, life gets a bit more complicated, because in general there are many different possible combinations of states that are in our box of atoms. The atoms keep moving around and this means that the number of atoms in each state at any one time varies. We call these combinations of atoms **configurations**. Helpfully, it turns out that provided we have enough atoms, then the behaviour of the **most probable configuration** describes the behaviour of the whole system. To understand this, imagine flipping a coin. You know that the chance of it

being a head or a tails is 50:50, but if you only flipped 5 coins then there's a reasonable chance they'd all be heads, or all be tails. From this, you might wrongly extrapolate that the most likely outcome of a coin flip is that it is heads, but if you flipped *enough* coins, then over time the average result would get nearer to the true result of 50:50. If we flipped a million coins, we would almost certainly measure exactly this result.

We can extend this argument to atoms in different energy states quite easily. If we have enough atoms (and we probably have on the order of 10^{23}), then the chance that enough of our atoms are in a state other than the most probable state to make a difference to the overall configuration is vanishingly small - just as we wouldn't expect 750,000 of our 1,000,000 flipped coins to all come out heads. The chance of our system being in anything other than the most probable configuration is essentially zero, and furthermore, even if it wasn't in the most probable configuration, it would be in a configuration that is very very close to the most probable configuration. So even if *wasn't* in the most probable configuration (and it almost definitely is), then it will behave essentially identically to it anyway. The upshot is that **we need to find out what the most probable configuration is.**

Finding The Most Probable Configuration

So how do we do this? Well, a helpful concept to have now is the **statistical weight** of one of our configurations. This just expresses how likely it is that we make a particular configuration if we randomise the states of our atoms. Equivalently, this is expressing how many ways there are to make a certain configuration - if there are more ways to make it, then the statistical weight is higher. The statistical weight, W, is given by:

$$W = \frac{N!}{\prod_i n_i!} \tag{1}$$

Where the symbols were defined above, and the capital Π denotes a product. The meaning of the statistical weight can be simply shown if we define a model system consisting of three particles, each with two possible energy states. All the particles could be in the ground state, which we call configuration (1,1,1), or the first could be in the excited state (2,1,1), and so on for configurations (1,2,1), (1,1,2), (1,2,2) (2,1,2), (2,2,1), and (2,2,2). Let's calculate the weight, W, for each of these configurations. In all cases $N = 3$, and so $N! = 6$, so we just need to calculate $n_1!$ and $n_2!$ - as we have two possible energy states. Calculate the weights now!

You should find that $W(1,1,1) = W(2,2,2) = 1$, $W(\mathrm{oneinstate2}) = 3$, and $W(\mathrm{twoinstate2}) = 3$. This ought to make sense, as there are three ways that

each configuration where we have one particle in a different state from the other two can be reached, but only one way we can reach the configurations where all particles are in the same state. **This is what we mean by the statistical weight of a configuration**.

So, to find the most probable configuration, we want to find the configuration that has the **largest weight**. Or, we want to **maximise the weight**. As the weight is just a function, we can use calculus to find the maximum weight, but we need to impose some constraints on this maximisation, as otherwise we would end up finding out that the most likely configuration is when we have an infinite number of particles! The constraints we need to add are:

1. The total number of particles remains constant. $\sum_i n_i = N$. So we can't create or destroy particles to increase our weight.

2. The total energy of the system is constant. $\sum_i n_i \epsilon_i = E$, where E is the total energy of the system. This means we can't just create or destroy energy to increase our weight.

We will also find it convenient to maximise not the weight directly, but actually $\ln W$. This is fine, as a maximum in $\ln W$ is also a maximum in W[3]. Doing this means we can use Stirling's approximation[4] to make dealing with the factorials a bit easier. We can then use **Lagrange's method of undetermined multipliers** to maximise $\ln W$ subject to the above constraints. This is outlined below, note that we are maximising W *with respect to* n_i, i.e. we are trying to find the set of state populations n_i which give us the configuration with the largest weight.

First, an expression for the natural logarithm of the weight:

$$\ln W = ln N! - \sum_i \ln n_i \tag{2}$$

Where we have used some of the rules of logarithms - this part is left as an exercise for you! We now differentiate this with respect to n_i. Let's use partial derivatives because I'm not really sure whether to be partial or not, and a mathematician friend of mine once said 'if in doubt: partial'.

$$\frac{\partial \ln W}{\partial n_i} = \frac{\partial ln N!}{\partial n_i} - \sum_i \frac{\partial \ln n_i!}{\partial n_i} \tag{3}$$

The first term in the above equation is clearly zero, as the total number of particles N won't change if we change how the particles are split up amongst the different states - which is what changing n_i is doing. The second term isn't obviously zero, so let's apply Stirling's approximation to the second term and get rid of the nasty

---

[3]This is because $\ln x$ is a **monotonic** function, I think.

[4]$N! \approx N \ln N - N$

factorial.

$$\frac{\partial \ln W}{\partial n_i} = -\sum_i \frac{\partial (n_i \ln n_i - n_i)}{\partial n_i}$$

$$= -\sum_i \left(\frac{\partial n_i \ln n_i}{\partial n_i} + \frac{\partial n_i}{\partial n_i} \right)$$

$$= -\sum_i \left(n_i \frac{\partial \ln n_i}{\partial n_i} - \ln n_i \frac{\partial n_i}{\partial n_i} + 1 \right)$$

$$= \sum_i \left(- n_i \frac{1}{n_i} - \ln n_i + 1 \right)$$

$$= -\sum_i \ln n_i = 0$$

Where the last line is set to zero as we want a maximum (we can also therefore get rid of the annoying minus sign). However, now is time to apply **Lagrange multipliers** for our initial constraints. So let's define two Lagrange multipliers α and β. Look in some maths notes for more details of this (it's a bit fiddly), but when you do it you'll end up with the following:

$$\alpha + \beta\epsilon_i + \ln n_i = 0 \tag{4}$$

Which then rearranges to:

$$n_i = e^{-\alpha} e^{-\beta\epsilon_i} \tag{5}$$

For our population of state i. This is great, but we can also go a bit further use one of our initial conditions to get rid of the multiplier α (this is kinda the idea, that we often don't have to determine the *undetermined* multipliers). Using our constraint (1) from above, we can say that:

$$N = \sum_i n_i = e^{-\alpha} \sum_i e^{-\beta\epsilon_i} \tag{6}$$

And therefore that:

$$e^{-\alpha} = \frac{N}{\sum_i e^{-\beta\epsilon_i}} \tag{7}$$

And THEREFORE that:

$$n_i = \frac{N}{\sum_i e^{-\beta\epsilon_i}} e^{-\beta\epsilon_i} = \frac{N}{q} e^{-\beta\epsilon_i} \tag{8}$$

I'll let that sink in for a bit.

We should also now say that $\beta = 1/kT$ where k is the Boltzmann constant and T is the temperature. You can show this algebraically too - but it's easiest if you derive the whole equation slightly differently starting from a more 'microscopic' position. This derivation is done in McQuarrie's textbook, which is excellent and does a 'bottom up' approach to physical chemistry, starting you at the microscopic and leading up to the macroscopic.

The Partition Function, q

What we just did was pretty awesome, if we're honest. We just derived a formula for the populations n_i of our different energy states ϵ_i that will produce the configuration with the largest weight. We already know that this configuration is the one that dominates the behaviour of the system, from our arguments about tossing coins. Strictly this is due to a thing called *the principle of equal a priori probabilities*[5], which is bascially just saying that we assume that each configuration is equally likely to occur, and so the one that has the most possible **ways** of being made (the one with the highest weight), is the one that dominates the behaviour.

So, now we've basically solved our puzzle, right? Our challenge was to find expressions for n_i, i.e. how the available population N is partitioned among the different states i, and we just did this in Equation 8. To do this, we defined a new thing, q, which is called the **partition function**. It tells us how the available population is **partitioned** between the different states. I could rewrite Equation 8 as:

$$P_i = \frac{n_i}{N} = \frac{\mathrm{e}^{-\beta\epsilon_i}}{q} \tag{9}$$

Where P_i is the probability that we find ourselves in state i - the ratio of the population of state i to the total population N. So the probability of being found in state i is just the Boltzmann factor $\mathrm{e}^{-\beta\epsilon_i}$ divided by the partition function q. So if the partition function is big, then the probability of being found in one specific state is rather small, as we are **partitioning** our population over a lot of states. Let's look at q in more detail:

$$q = \mathrm{e}^{-\beta\epsilon_0} + \mathrm{e}^{-\beta\epsilon_1} + \mathrm{e}^{-\beta\epsilon_2} + ... \tag{10}$$

If you're a North Sea crossing intrepid explorer like me, you might have at one point learnt about this in a language like Danish or German. In those languages, q isn't called the **partition function**, instead it's called (in Danish) the **tilstandsummen**. This translates roughly to 'sum-over-states', which is a nice way to visualise what it physically represents. Looking at the equation above, it is clearly a sum of the Boltzmann factors for all our different states. We order the states such that $\epsilon_0 < \epsilon_1 < \epsilon_2$ and so on, so each term in the sum gets gradually smaller as the energy of the states increases (at a constant temperature).

Interpreting q

The partition function is **just a number** at the end of the day, but what does it actually mean? If I said that the partition of a system was 5, or 1, or 10 million, what does that mean?

To understand this, let's rewrite our expression for q in terms of energy differences by dividing by the first term in the equation above.

$$q = 1 + \mathrm{e}^{-\beta \Delta E_{10}} + \mathrm{e}^{-\beta \Delta E_{20}} + ... \tag{11}$$

Let's think about some physical cases now. What if the temperature is really really low, such that $\beta \to \infty$? In this case, every term except the first goes to zero, and $q = 1$. We know that at low temperature, everything will be in the ground state, and this is exactly what we've just observed here. We could arrive at the same conclusion if we said that the energy gaps to the ground state were huge too - do you agree?

What if we were at a really high temperature, so that $\beta \to 0$? In this case, every term in this sum goes to 1, and we end up that q is equal to the number of available states we have. We know that at really high temperatures, all states end up equally populated. Does this make sense here too?

Hopefully now you're seeing the interpretation of q. q tells you **at a given temperature, how many states are thermally accessible and therefore will have population**. If our q was 5, then we have 5 states that will have non-negligible population. Note that q doesn't need to be an integer, but it **cannot** be less than 1, as this implies we have less than one state populated! If you find that $q < 1$, this normally means that one of your assumptions in the calculation wasn't valid.

Back to the Point

Ok, so we understand how to find the populations of our states, and we understand what q is and how it helps us to do this. But our original question was to find out how we can calculate the internal energy of a load of atoms in a box, so let's carry on with this. We now know everything we need to know: how many atoms we have; how many states are available to them, and how the population is distributed among those states. So how do we calculate thermodynamic properties?

[6]You can also derive these directly, but we're just going to quote the results here for brevity.

We can do this with the help of things called **bridging functions**[6]. These are things that link together thermodynamic functions with the partition functions, for example:

$$U = NkT^2 \left(\frac{\partial \ln q}{\partial T} \right) \tag{12}$$

For the internal energy. This is fairly straightforward to derive if we note that $U = \sum_i n_i \epsilon_i$. We can also say that:

$$A = -Nk \ln q \tag{13}$$

For the Helmholtz free energy. We can find expressions for entropy, Gibbs' free energy, pressure, etc.. too. The main idea is that we have some functions that give

us a **bridge** between the partition functions and the thermodynamic properties. To finish off, let's see how we can use the formula for U to calculate the internal energy of our gas.

We need to find the partition function for our atoms, but we know this! All we need is an expression for the possible states ϵ_i. A load of atoms in a box only have translational energy, so we can consider them just as particles in a box, where:

$$\epsilon_i = \frac{i^2 h^2}{8mL^2} \tag{14}$$

Where i is our quantum number now, but other symbols have their usual meanings. L is the length of one side of the box (as this is an expression for energy in a 1D box). So our partition function is given by:

$$q = \sum_i \exp\left(\frac{i^2 h^2}{kT 8mL^2})\right) \tag{15}$$

Where we expanded the β into the fraction. The trick now is to say that because we're probably at room temperature (and we are, I just forgot to write that at the start, and it's up to us to define the problem anyway), then the levels are closely spaced enough that we can change this from a sum over i to an **integral** over i. Do this integral (remembering some standard integrals!), and you'll find:

$$q = \sqrt{\frac{L^2 2\pi mkT}{h^2}} \tag{16}$$

But this is actually only the equation for one dimension of the motion, as we used the expression for energy of a particle in a 1D box. So we can just cube this to find the partition function for 3D motion as the degrees of freedom are separate.

Having cubed the expression, take it and bang it into the expression we had for U above. This isn't as bad as it looks, because as we're taking the derivative of $\ln q$ we can just ignore everything that isn't temperature dependent in our expression for q (try it and see). What you should find is that:

$$U = NkT^2 \times \frac{3}{2T} = \frac{3}{2}RT \tag{17}$$

Isn't that satisfying?

Epilogue: Complications

A couple of things I haven't mentioned here are probably worth flagging up.

Firstly, I've kept talking about the distribution of **energy states** not **energy levels**. This is deliberate. The difference between a state and a level is that a level can contain many states, depending on the degeneracy of the level. If we

wanted to change the formulae used to refer to energy levels, then we also have to include a factor to account for the degeneracy of each level (normally called g). This doesn't add any complexity, but it's worth being aware of. For example, for a sum over **states:**

$$q = \sum_{\text{states},i} e^{-g_i \beta \epsilon_i} \tag{18}$$

Where g_i is the degeneracy of state i. Whereas for a sum over **levels:**

$$q = \sum_{\text{levels},i} e^{-\beta \epsilon_i} \tag{19}$$

Secondly, is the difference between q and Q, which you may have already seen. q is the **molecular partition function**, which describes the behaviour of each particle in the system separately. Q is the **canonical partition function**, which describes the behaviour of the entire **system** together as an ensemble. If the particles are indistinguishable, which is when you could take a particle out and put it somewhere else without *really* noticing[7], then:

[7]There is a better definition than this probably.

$$Q = \frac{q^N}{N!} \tag{20}$$

But if the particles are distinguishable, then:

$$Q = q^N \tag{21}$$

In the gas or solution phases, then particles are indistinguishable. The classic case of distinguishable particles is if they are in a solid lattice (like a crystal, for example). The difference arises because if the particles are indistinguishable then we have to avoid 'over counting' states that are the same, but just arises from permuting two particles.