

# Statistical Physics

**Matthew Wingate\***

19 January 2010

---

\*Comment & corrections to [M.Wingate@damtp.cam.ac.uk](mailto:M.Wingate@damtp.cam.ac.uk) . Notes and other information also available at <http://www.damtp.cam.ac.uk/user/wingate/StatPhys>

## 2 Canonical Ensemble

For now we focus on the canonical ensemble and derive a probability density with which we can carry out ensemble averages.

### 2.1 Probability density

Imagine a large number  $A$  of distinguishable replicas of a quantum system  $\mathcal{S}$

$$\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_A\} = \mathcal{E}^{(j)} \quad (2.1.1)$$

( $\mathcal{E}^{(j)}$  denotes the  $j$ -th sampling of ensemble  $\mathcal{E}$ .) These replicas could be envisioned as snapshots of a system as it evolves in time or, equivalently assuming the ergodic hypothesis, different subsystems of a larger system.

On each system we can, in principle, measure a complete set of commuting observables in order to determine its corresponding microstate. We could denote the microstate of  $\mathcal{S}_1$  by  $|\Psi_1\rangle$ , and so on. Let us denote the number of replicas which are in a general microstate  $|i\rangle$  to be  $a_i$ . In the physical interpretation of watching the time evolution of a system, each of these successive  $\mathcal{S}_b \in \mathcal{E}^{(j)}$  could be interpreted as snapshots at short intervals. Then  $a_i$  would answer the question, ‘‘How often does the state  $|i\rangle$  recur?’’ In the ensemble language, ‘‘How many of the  $|\Psi_b\rangle = |i\rangle$ ?’’

For clarity, let’s consider the toy example of a spin system which consists solely of  $N = 2$  spins. Each replica is in one of the 4 possible microstates  $|++\rangle, |+-\rangle, |-+\rangle, |--\rangle$ . We could imagine forming several samplings out of several sets of  $A = 5$  replicas, *e.g.*

$$\begin{aligned} \mathcal{E}^{(1)} &= \{|++\rangle, |+-\rangle, |+-\rangle, |--\rangle, |+-\rangle\} \\ \mathcal{E}^{(2)} &= \{|--\rangle, |--\rangle, |+-\rangle, |--\rangle, |--\rangle\}. \end{aligned} \quad (2.1.2)$$

Then

$$\begin{aligned} (a_{|++\rangle}, a_{|+-\rangle}, a_{|+-\rangle}, a_{|--\rangle})^{(1)} &= (2, 0, 2, 1) \\ (a_{|++\rangle}, a_{|+-\rangle}, a_{|+-\rangle}, a_{|--\rangle})^{(2)} &= (0, 0, 1, 4). \end{aligned} \quad (2.1.3)$$

In the canonical ensemble, we apply 2 constraints:

$$\sum_i a_i = A \quad (2.1.4)$$

$$\sum_i a_i E_i = A \langle E \rangle \quad (2.1.5)$$

The first is just a statement that numbers have to add up. If we count the number of replicas which are in a generic state  $|i\rangle$ , sum over states, then the total must equal  $A$  (each  $\mathcal{S}_b$  is in 1 and only 1 state, of course). The second can be taken as a definition of average energy  $\langle E \rangle$ , but we will implement it as a constraint later. Given a sampling of ensemble  $\mathcal{E}$  the chance of picking at random an  $\mathcal{S}_b$  which is in state  $|i\rangle$  is clearly

$$\rho_i = \frac{a_i}{A}. \quad (2.1.6)$$

Let us denote the collection of  $a_i$  for all possible  $|i\rangle$  by  $\{a_i\}$ . This way of keeping track of  $\{a_i\}$  discards any temporal interpretation, following the ergodic hypothesis. If we only know (2.1.3), we cannot reconstruct the ordering in (2.1.2). As long as the ergodic hypothesis is correct, it is only the  $\{a_i\}$  which are important.

We will see that we need to count the number of ways  $\{a_i\}$  could be obtained. How many possible histories could give the same collection of  $\{a_i\}$ ? The answer is given by

$$W(\{a_i\}) = \frac{A!}{\prod_i a_i!} \quad (2.1.7)$$

The numerator is the number of ways of arranging  $A$  unique entities. This overcounts unique arrangements since some of the entries in  $\mathcal{E}$  are not unique. The degeneracies are given by the  $a_i$ , so we have to divide by the number of ways of arranging those identical systems in  $|i\rangle$ , for all  $i$ . In the toy example (2.1.3)

$$\begin{aligned} W^{(1)} &= \frac{5!}{2!0!2!1!} = 30 \\ W^{(2)} &= \frac{5!}{0!0!1!4!} = 5. \end{aligned} \quad (2.1.8)$$

( $0! = 1$ ; there is only 1 way to arrange no things.)

Now we come to an important point. As a consequence of each allowable microstate being equally likely, the probability of a given collection  $\{a_i\}$  occurring in a trial experiment, a sample ensemble, is proportional to  $W(\{a_i\})$

$$\boxed{\mathcal{P}(\{a_i\}) \propto W(\{a_i\})}. \quad (2.1.9)$$

This serves to count *with equal weight* each possible order of the  $\mathcal{S}_b$  in  $\mathcal{E}$  which could give a particular  $\{a_i\}$ . In the toy example there are 30 ways of getting 2 occurrences of  $|++\rangle$  and  $| - + \rangle$  plus 1 occurrence of  $| -- \rangle$ , but only 5 ways of observing the replicas in the  $| -- \rangle$  state 4 times and once in  $| - + \rangle$ , therefore  $\{a_i\}^{(1)}$  is twice as likely as  $\{a_i\}^{(2)}$ . Insofar as flipping spins, like flipping coins, and the motion of air molecules are the same, they can both be treated as random events from our macroscopic perspective, this toy example is a useful analogy.

Now we focus on a system in thermal equilibrium, that steady state we find after waiting sufficient time. The steady macroscopic state should spend most of its time visiting microstates consistent with constant or nearly constant macroscopic observables. Therefore we shall find those  $\{a_i\}$  which maximise  $\mathcal{P}(\{a_i\})$ , i.e. those which maximise  $\log W(\{a_i\})$

$$\begin{aligned} \log W &\simeq A \log A - A - \sum_i a_i (\log a_i - 1) \\ &\simeq A \log A - \sum_i a_i \log a_i \end{aligned} \quad (2.1.10)$$

In order to maximise  $W$  while satisfying the constraints (2.1.4) and (2.1.5) we introduce 2 Lagrange multipliers  $\alpha$  and  $\beta$ :

$$\boxed{\frac{\partial}{\partial a_j} \left( A \log A - \sum_i a_i \log a_i - \alpha \sum_i a_i - \beta \sum_i a_i E_i \right) = 0} \quad (2.1.11)$$

which is satisfied when

$$\log a_j + 1 + \alpha + \beta E_j = 0 \quad (2.1.12)$$

or

$$a_j = e^{-1-\alpha-\beta E_j}. \quad (2.1.13)$$

(A brief reminder regarding Lagrange multipliers is given at the end of this chapter.)

Since

$$A = \sum_i a_i = e^{-1-\alpha} Z \quad (2.1.14)$$

where

$$Z = \sum_i e^{-\beta E_i}. \quad (2.1.15)$$

Therefore the probability of a system in thermal equilibrium to be in state  $|i\rangle$  is

$$\rho_i = \frac{a_i}{A} = \frac{1}{Z} e^{-\beta E_i}. \quad (2.1.16)$$

As alluded to in Chapter 1, we find probability distribution which depends *only on the energy* of the microstate,  $\rho_i = \rho(E_i)$ . This distribution (2.1.16) is called the **Boltzmann distribution**.

Making contact with your quantum mechanics courses, we can write the average energy as the average over the energies of the microstates, weighted by how frequently they occur when in equilibrium:

$$\langle E \rangle = \sum_i \langle i | \hat{H} | i \rangle \rho_i = \frac{1}{A} \sum_i a_i E_i. \quad (2.1.17)$$

Note that the angled brackets (which we will drop in later chapters) refer to a *thermal average*, hence the average is weighted by the thermal probability distribution  $\rho$ .

As part of our discussion above, in (2.1.15) we defined  $Z$ , which we call the **partition function**. We will see throughout this course that this is a very useful function. From it we can, in principle, calculate all the statistical properties of the equilibrated system. For example an easy way to calculate the average energy is to differentiate  $\log Z$  with respect to the Lagrange multiplier  $\beta$ :

$$\langle E \rangle = - \left. \frac{\partial \log Z}{\partial \beta} \right|_V. \quad (2.1.18)$$

It is important that the volume remain fixed since the  $E_i$  typically depend on the volume. We saw this last chapter in the example of particles in a periodic box.

Fluctuations about the mean can also be calculated from  $Z$ :

$$\left. \frac{\partial^2 \log Z}{\partial \beta^2} \right|_V = \frac{1}{Z} \left. \frac{\partial^2 Z}{\partial \beta^2} \right|_V - \frac{1}{Z^2} \left( \left. \frac{\partial Z}{\partial \beta} \right|_V \right)^2 \quad (2.1.19)$$

$$= \langle E^2 \rangle - \langle E \rangle^2 = \langle (E - \langle E \rangle)^2 \rangle = (\Delta E)^2. \quad (2.1.20)$$

For large systems, the total average energy scales with the particle number  $\langle E \rangle \sim N$ . Let us assume that  $E(\beta)$  is smooth, then the derivative of  $E$  also scales with  $N$ :

$$\frac{\partial \langle E \rangle}{\partial \beta} \sim N \Rightarrow \Delta E \sim \sqrt{N} \Rightarrow \frac{\Delta E}{\langle E \rangle} \sim \frac{1}{\sqrt{N}}. \quad (2.1.21)$$

Since we will be interested in the thermodynamic limit, where we make  $N$  and  $V$  arbitrarily large keeping  $N/V$  fixed, we can also see that  $\Delta E / \langle E \rangle \sim 1/\sqrt{V}$ . Keeping in mind that most of the time we have  $N$  of the order of millions, billions, even  $10^{23}$ , Avogadro's number, it is clear that fluctuations about the mean are very small. It is no wonder, then, that the same thermodynamics will emerge whether we use the canonical ensemble, where the total energy  $E$  fluctuates, or the microcanonical ensemble, where  $E$  is conserved.

## 2.2 Temperature

Consider 2 systems  $\mathcal{S}_a, \mathcal{S}_b$  with volumes  $V_a, V_b$  and particle numbers  $N_a, N_b$ . Put them in contact with each other, allowing the exchange of energy but not particle number.

First we note that this contact does not change the microstates with respect to the separated systems, for large systems. This is because any effects due to the contact scale like the surface area shared by the 2 systems, compared to the energies  $E_a, E_b$  which scale like the volumes  $V_a, V_b$ . This dimensional analysis is enough to let us proceed in straightforward manner. To repeat the statement that contact does not effect the microstates using quantum mechanics notation: the microstate of the joint system is just the product of the isolated systems' microstates

$$|ij\rangle_{ab} = |i\rangle_a \otimes |j\rangle_b \quad (2.2.1)$$

This joint state has energy  $E_{ij}^{ab} = E_i^a + E_j^b$ , which can be seen using a Hamiltonian  $\hat{H}_{ab} = \hat{H}_a \otimes \mathbb{1}_b + \mathbb{1}_a \otimes \hat{H}_b$ :

$${}_a b \langle ij | \hat{H}_{ab} | ij \rangle_{ab} = {}_a \langle i | \hat{H}_a | i \rangle_a + {}_b \langle j | \hat{H}_b | j \rangle_b = E_i^a + E_j^b. \quad (2.2.2)$$

Assume a long time has passed so that the joint system  $\mathcal{S}_{ab}$  is in equilibrium. Of course the subsystems  $\mathcal{S}_a$  and  $\mathcal{S}_b$  will be in equilibrium individually at the same time. Therefore we can equate

$$\rho^{ab}(E_i^a + E_j^b) = \rho^a(E_i^a) \rho^b(E_j^b) \quad (2.2.3)$$

for all  $E_i^a$  and  $E_j^b$ . We can repeat the analysis of the previous section, introducing separate Lagrange multipliers for each subsystem and the joint system, to rewrite (2.2.3) as

$$\frac{e^{-\beta_{ab}(E_i^a + E_j^b)}}{\sum_{i,j} e^{-\beta_{ab}(E_i^a + E_j^b)}} = \frac{e^{-\beta_a E_i^a} e^{-\beta_b E_j^b}}{\left( \sum_i e^{-\beta_a E_i^a} \right) \left( \sum_j e^{-\beta_b E_j^b} \right)} \quad (2.2.4)$$

Define the denominator of the left-hand side to be  $Z_{ab}$  and the product in the denominator of the right-hand side to be  $Z_a Z_b$ . This equality holds for all  $E_i^a, E_j^b$  only if

$$\beta_{ab} = \beta_a = \beta_b \quad (2.2.5)$$

which further implies  $Z_{ab} = Z_a Z_b$ . The fact we only have one common Lagrange multiplier is a consequence that there really is only one constraint for the joint system

$$(A + B)E = \sum_i a_i E_i^a + \sum_j b_j E_j^b = \sum_{ij} c_{ij} (E_i^a + E_j^b). \quad (2.2.6)$$

This  $\beta$  is a useful parameter, as we have already seen at the end of the previous chapter. More intuitive is its inverse

$$\beta = \frac{1}{kT}. \quad (2.2.7)$$

where  $k$  (often written  $k_B$  when notation might be redundant) is Boltzmann's constant, equal to  $1.38 \times 10^{-23}$  Joules per Kelvin or  $8.6 \times 10^{-5}$  electron volts per Kelvin. In many branches of physics, we just set  $k = 1$  and quote temperature in energy units.

Equilibrium is characterised by the distribution

$$\exp\left(-\frac{\text{energy}}{k \cdot \text{temperature}}\right). \quad (2.2.8)$$

As you might expect intuitively, if we make the temperature higher, we make higher energies more likely. It is the ratio of the two quantities that controls probabilities.

### 2.3 Summary

Having demonstrated some concepts through the canonical ensemble, we summarize the concepts of thermal equilibrium.

- An isolated system reaches a macroscopic steady state after sufficient time which called equilibrium and is characterised by a temperature  $T$ .
- Two systems in contact with each other eventually come into equilibrium with each other at a common temperature.
- If  $\mathcal{S}_a$  is in equilibrium with  $\mathcal{S}_b$ , and  $\mathcal{S}_b$  is in equilibrium with  $\mathcal{S}_c$ , then  $\mathcal{S}_a$  is in equilibrium with  $\mathcal{S}_c$ , all at the same  $T$ . This is called the **Zeroth law of thermodynamics**.

### 2.4 Lagrange multipliers

Since some people in the lectures may be unfamiliar with Lagrange multipliers, here is a short summary.<sup>1</sup>

Let's say we want to maximise the function  $f(x, y)$  subject to the constraint that another function  $g(x, y) = c$  (constant  $c$ ). We aim to solve

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = 0. \quad (2.4.1)$$

---

<sup>1</sup>E.g. see K F Riley, M P Hobson, and S J Bence, *Mathematical Methods for Physics and Engineering*, (Cambridge Univ. Press, 2006).

However  $x$  and  $y$  are not independent. Since  $g$  is constant, we have

$$dg = \frac{\partial g}{\partial x} dx + \frac{\partial g}{\partial y} dy = 0. \quad (2.4.2)$$

Multiply  $dg$  by a number  $\lambda$ , to be determined, and add, so that

$$d(f + \lambda g) = \left( \frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} \right) dx + \left( \frac{\partial f}{\partial y} + \lambda \frac{\partial g}{\partial y} \right) dy. \quad (2.4.3)$$

In order for  $d(f + \lambda g) = 0$ , each of the 2 terms on the right-hand side above must separately be zero

$$\begin{aligned} \frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} &= 0 \\ \frac{\partial f}{\partial y} + \lambda \frac{\partial g}{\partial y} &= 0. \end{aligned} \quad (2.4.4)$$

These equations are the analogue of (2.1.11).

### Further reading

1. L D Landau and E M Lifshitz, *Statistical Physics*, (Pergamon Press, 1980), §§1-7.
2. F Reif, *Fundamentals of Thermal and Statistical Physics*, (McGraw-Hill, 1965), Chapter 2.
3. R K Pathria, *Statistical Mechanics*, (Pergamon Press, 1985), §1.2, §§3.1-6.