

# 1. Introduction

## 1.1 What is Statistical Mechanics?

**Thermal Physics** encompasses all parts of physics where ideas of temperature and entropy come into play. As we shall see this implies the properties of macroscopic systems with a large number of microscopic constituents.

### Macroscopic Approach (Classical Thermodynamics)

- deals with macroscopic variables i.e. variables that do not refer to any microscopic details
- input is phenomenological laws e.g. equation of state
- output is general relations between macroscopic variables
- advantage is the generality of the approach

### Microscopic Approach (Statistical Mechanics)

- starts from a microscopic description and seeks to explain macroscopic properties.
- input is a microscopic model of a given system i.e. what we believe is an adequate microscopic description
- output is predictions for macroscopic properties and behaviour.
- advantage of the approach is that it yields predictions for a given system which can be compared to experiment thus allowing refinement of the microscopic model which in turn deepens our understanding of the system.

The key idea that connects the microscopic and macroscopic approaches is that due to the large number of constituents, macroscopic quantities are precisely defined even though the microscopic specification is not precise.

### Aims of course

- To formulate the ideas of statistical mechanics
- To give microscopic understanding of entropy & second law of thermodynamics which you met in Year 1 Properties of Matter (PoM)
- To derive results you have met in PoM such as equipartition and the Boltzmann factor
- To explore the role of indistinguishability and quantum effects in many body systems

## 1.2 Microscopic approaches

*Extreme philosophy:* If we know microscopic laws e.g. Newtonian mechanics; quantum mechanics etc then we just specify the initial conditions and solve all the equations to determine what happens!

Problems with this approach:

- Analytically intractible
- even on a powerful computer one can only deal with a limited number of constituents say  $N \sim 10^5$ , whereas Avogadro's number is  $N_A = \sim 6 \times 10^{23}$ .
- system may not even be deterministic! i.e. small change in initial conditions leads to completely different final states

Instead we will use the simpler and more powerful *statistical* formulation

- Only aim to make probabilistic statements e.g. the Maxwell probability distribution for the velocities of molecules in an ideal gas (see PoM) doesn't tell us about trajectories of individual gas molecules. However it does give us all we need to know to deduce macroscopic properties such as the pressure of the gas.
- It is a very natural and economical approach i.e. we only introduce as much detail as we need (or want) in the microscopic description and we do not waste time calculating unnecessary details. Therefore we can do the calculations analytically (i.e. on pen and paper).
- It uses the large number  $N$  of microscopic constituents to advantage
- Why the name 'statistical mechanics'?
  - The microscopic model is the 'mechanics' e.g. quantum, Newtonian or even more primitive (but adequate) descriptions
  - 'statistical' refers to the statistical description of the resulting behaviour of a system with a large number of microscopic constituents

Statistical mechanics was founded in the late nineteenth century with the work of Maxwell, Boltzmann, Gibbs ...on classical systems. It was developed hand in hand with quantum mechanics in the twentieth century by Einstein, Bose, Dirac, Fermi . . . It remains an active and constantly developing research area in the 21st century. For example the techniques of statistical mechanics are applied to neural modelling, the study of traffic flow, economic and social systems... in fact any system with a large number of constituents.

More detailed comparison of the two approaches (taking example of a gas):

	Thermodynamics	Statistical Mechanics
Quantity of interest	Macroscopic properties (eg. $P, V, T, C_P, C_V$ )	Microscopic properties (eg: molecular speeds)
Strategy	Avoid microscopic model	Build on microscopic model
Strengths	Generality of results	Provides way of refining microscopic understanding
Weaknesses	No understanding of system-specific features. Conceptually opaque.	Requires additional input (the model). Requires additional techniques (probability theory; classical and quantum mechanics)

### 1.3 Probability

Probabilistic concepts are central to statistical mechanics. You should have a working knowledge of probability theory from your programme so far. We shall not explicitly revise this material in lectures, so refer to your probability notes as needed. For convenience and reference we summarise below some of the main results on which we shall draw.

🔥 Expand to review key concepts of probability

For further reading see chapter 3 of Gould and Tobochnik  
*Definitions*

- frequency definition: the probability  $P$  of an event in a trial is given by

$$P = \lim_{N \rightarrow \infty} \frac{n}{N}$$

where  $n$  is the number of occurrences in  $N$  trials.

- degree of belief: probability is a quantitative measure of our degree of belief that something will occur e.g. if there are  $q$  possible outcomes of a trial (recall that a trial is some procedure where we can measure the outcome), and we have no reason to favour any one outcome over any other, then we would assign probability  $1/q$  to each outcome. Tossing a coin would correspond to  $q = 2$  and rolling a die to  $q = 6$ .

Thankfully both definitions lead to the same numerical values for probability.

*Rules of Probability*

For a trial with  $q$  possible outcomes let the probability of outcome  $i$  be  $P(i)$

- normalisation  $\sum_{i=1}^q P(i) = 1$

- $P(i \text{ or } j) = P(i) + P(j)$  (for mutually exclusive outcomes  $i, j$ ).
- We can also consider ‘compound events’. For example consider two trials then

$$P(i \text{ in trial 1 and } j \text{ in trial 2}) = P(i)P(j)$$

(for outcomes of trial 1 and trial 2 independent)

#### *Probability distributions*

If we associate a numerical value  $x_i$  to event  $i$  then the probability distribution for the ‘random variable’  $x$  is

$$P(i) = \text{Probability that } x = x_i$$

#### *Probability density functions*

In many case the outcome of an event is described by a continuous variable  $x$ . Then the probability density function  $P(x)$  is defined by

$$P(x)dx = \text{Probability that outcome lies in the range } x \text{ to } x + dx$$

and is normalised according to

$$\int_{-\infty}^{\infty} P(x)dx = 1$$

(N.B. often the probability density function is simply called the probability distribution.)

#### *Averages*

The Probability distribution contains the complete information about the trial. However the moments (mean, variance . . . ) can give the important features. They are defined as

$$\begin{aligned} \text{mean } \bar{x} &= \sum_{i=1}^q P(i)x_i \text{ or } \int_{-\infty}^{\infty} P(x)xdx \\ \text{variance } \overline{\Delta x^2} &= \sum_{i=1}^q P(i)(x_i - \bar{x})^2 \text{ or } \int_{-\infty}^{\infty} P(x)(x - \bar{x})^2 \\ \text{i.e. } \overline{\Delta x^2} &= \overline{(x - \bar{x})^2} = \overline{x^2} - \bar{x}^2 \end{aligned}$$

More generally the average of a function  $f(x)$  is given by

$$\overline{f(x)} = \sum_{i=1}^q P(i)f(x_i) \text{ or } \int_{-\infty}^{\infty} P(x)f(x)dx$$

#### **Basic distributions**

For completeness (and for the mathematically inclined) the content below includes some derivations of the functional forms of distributions; you are not expected to know these derivations for this course.

#### *The binomial distribution*

First let us review the concepts of permutations and combinations. If we have  $N$  distinguishable objects—say numbered balls, or a pack of cards—then if we select  $n$  objects from  $N$  (without replacement) how many different possible outcomes are there?

The answer is known as the number of permutations of  $n$  from  $N$  and is equal to

$$\frac{N!}{(N-n)!} = \frac{N(N-1)(N-2)\dots(1)}{(N-n)(N-n-1)\dots(1)} = N(N-1)(N-2)\dots(N-n+1).$$

To see this think of the number of possibilities for the first object ( $= N$ ) then multiply by the number of possibilities for the second ( $N-1$ ) and keep going until the  $n$ th object for which there are ( $N-n+1$ ) possibilities.

Also recall that  $0! = 1$ , so that the number of permutations of  $N$  from  $N$  is  $N!$

In this counting scheme we in fact count as different outcomes the selection of the same objects but in different order e.g. clearly in the number of permutations of  $N$  from  $N$  we always choose the same  $N$  objects but the number of possible orders is  $N!$

Now if we only count as distinct outcomes those events where a different set of objects is selected we are interested in the number of combinations of  $n$  from  $N$ . To obtain the number of combinations we have to divide the number of permutations of  $n$  from  $N$  by  $n!$  (the number of permutations of the  $n$  selected objects). This results in

$$\text{number of combinations of } n \text{ from } N = \frac{N!}{(N-n)!n!} = \binom{N}{n}$$

where  $\binom{N}{n}$  is known as a binomial coefficient and is pronounced ‘ $N$  choose  $n$ ’. (N.B. There are many other symbols that are used for the number of combinations.)

Now we are ready to write down the binomial distribution. We consider  $N$  trials in each of which an event can occur with probability  $p$ . The probability of observing precisely  $n$  events in  $N$  trials is given by the [Binomial distribution](#)

$$P_n = \binom{N}{n} p^n (1-p)^{N-n}$$

The binomial coefficient gives the number of ways of choosing the  $n$  trials where an event occurs out of the total of  $N$  trials.

The factor  $p^n(1-p)^{N-n}$  is the probability that in the  $n$  chosen trials an event occurs and in the rest of the trials an event does not occur. (This factor is the same whatever the  $n$  chosen trials). This argument may seem somewhat obscure at first but should with familiarity become second nature.

Note that for the Binomial distribution

$$\bar{n} = Np \quad \text{and} \quad \Delta n^2 = Np(1-p)$$

It should be pointed out that the factorials involved in combinatorial calculations soon become very large numbers e.g. for  $N = 0, 1, 2 \dots$  one can check that  $N!$  is 1, 1, 2, 6, 24, 120, 720 ... and already  $15! \sim 10^{12}$ .

A very useful approximation for the factorial function is given by the *Stirling approximation*

$$\ln(N!) = N \ln N - N + \frac{1}{2} \ln(2\pi N) + O(1/N)$$

We see that for  $N$  large

$$\boxed{\ln(N!) \simeq N \ln N - N}$$

### Poisson distribution

As an application of Stirling's approximation let us consider the binomial probabilities for  $N \gg n$ . Then we find

$$\begin{aligned} \ln \binom{N}{n} &= \ln N! - \ln(N-n)! - \ln n! \\ &\simeq N \ln N - N - (N-n) \ln(N-n) + (N-n) - \ln n! \\ &\simeq n \ln N - \ln n! \end{aligned}$$

In the calculation we used Stirling's approximation for  $\ln N!$  and  $\ln(N-n)!$  and made use of the fact that for  $N \gg n$ ,  $\ln(N-n) \simeq \ln N - n/N$ . Note that we have only kept the highest order terms. Thus, re-exponentiating the logarithm, we see that

$$\binom{N}{n} \simeq \frac{N^n}{n!}$$

Now consider

$$(1-p)^{N-n} = \exp[(N-n) \ln(1-p)] \simeq \exp(-Np)$$

where we have used  $\ln(1-p) \simeq -p$  for  $p \ll 1$  and  $N-n \simeq N$ . (More strictly we require  $N \rightarrow \infty, p \rightarrow 0$  such that  $Np = \bar{n}$  is finite.) Under these conditions we find when we put the above results together that

$$\boxed{P_n = (Np)^n \frac{\exp(-Np)}{n!} = (\bar{n}^n) \frac{\exp(-\bar{n})}{n!}}$$

This is known as the [Poisson distribution](#) and one can check using the power series representation of the exponential function that  $\sum_{n=1}^N P_n = 1$ . One also finds

$$\boxed{\begin{aligned} \bar{n} &= \sum_{n=1}^N n P_n = Np \\ \overline{\Delta n^2} &= \sum_{n=1}^N (n - \bar{n})^2 P_n = Np \end{aligned}}$$

i.e. the mean and variance are equal.

Basically the Poisson distribution is used when there are a large number of trials in each of which an event is very unlikely but overall one expects a finite number of events.

### *Gaussian Distribution*

Here we consider the limit of the binomial distribution where  $N$  is large and so is  $Np$ . Therefore we expect that for  $P_n$  to be non-vanishingly small, we need  $n$  of the same order as  $Np$ .

Thus we use Stirling's approximation on binomial distribution to show that

$$s(n) \equiv \ln P_n \simeq n \ln p + (N - n) \ln(1 - p) + N \ln N - (N - n) \ln(N - n) - n \ln n$$

for large  $N, n$ . It is straightforward to show that

$$\begin{aligned} s'(n) &= \ln p - \ln(1 - p) + \ln(N - n) - \ln n \\ s''(n) &= -\frac{1}{N - n} - \frac{1}{n} \end{aligned}$$

You should notice that  $s(n)$  is a maximum at  $n = Np$  i.e. at the mean value.

Now expand in powers of  $x = n - Np$ , i.e. make a Taylor expansion around the mean value:

$$s(x) = s(Np) + xs'(Np) + \frac{x^2}{2}s''(Np) \dots$$

Note that since  $s$  is maximised at the mean value the first non-zero term in the expansion is in  $x^2$  and

$$s(x) \simeq s(NP) = \frac{x^2}{2N(1 - p)p}$$

When we return to  $P_n = \exp s(n)$  we find

$$P_n \simeq P_{Np} \exp\left(\frac{-(n - Np)^2}{2N(1 - p)p}\right)$$

In order to determine the constant  $P_{N_p}$  which serves to normalise the distribution we can make the approximation

$$1 = \sum_{n=0}^N P_m \simeq \int_{-\infty}^{\infty} P_{Np} \exp\left(\frac{-(x^2)}{2N(1 - p)p}\right) dx$$

where extending the limits in this way produces no error since the additional contributions are vanishingly small.

We then invoke a standard integral you should be familiar with

$$\int_{-\infty}^{\infty} \exp\left(\frac{-(x^2)}{2\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}.$$

From this we deduce the correct normalisation for the Gaussian approximation to the binomial distribution as

$$P_n \simeq \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(n - Np)^2}{2\sigma^2}\right)$$

where  $\sigma^2 = Np(1 - p)$ .

Note that if  $|n - Np| \gg N^{1/2}$  the probability becomes vanishingly small due to the argument of the exponential becoming very large.

The [Gaussian distribution](#) is often referred to as a ‘Normal distribution’.

## 1.4 Simplicity at large $N$ : a toy example

Suppose that  $N$  coins are tossed; we denote by  $n$  the number of heads obtained. We will determine the probability distribution  $p_n$  (ie the set of probabilities  $p_0, p_1 \dots, p_N$ ) characterising the possible outcomes.

Denote by  $p$  the probability that a head results from a single toss; then  $q \equiv 1 - p$  gives the corresponding probability for a tail.

There are *many distinct ways* of getting  $n$  heads from  $N$  tosses, differing according to which tosses give heads. It should be clear that each one of these distinct ways has the same probability, namely  $p^n q^{N-n}$ , and recalling your probability and statistics lectures, the number of distinct ways in which we can get  $n$  heads from  $N$  tosses is given by the binomial coefficient:

$${N \choose n} \equiv \frac{N!}{(N-n)!n!}$$

It follows that

$$\begin{aligned} p_n &\equiv \text{number of distinct ways of obtaining } n \text{ heads} \\ &\quad \times \text{probability of any specific way of getting } n \text{ heads} \\ &= {N \choose n} p^n q^{N-n} \end{aligned}$$

This is the binomial distribution of probabilities. The distribution is characterised by two key parameters, the *mean*  $\bar{n}$  and the *variance*  $\overline{\Delta n^2}$ . Remind yourself from your probability and statistics lectures that  $\bar{n} \equiv \sum_{n=0}^N np_n = Np$ , and  $\overline{\Delta n^2} \equiv \sum_{n=0}^N (n - \bar{n})^2 p_n = Npq$ .

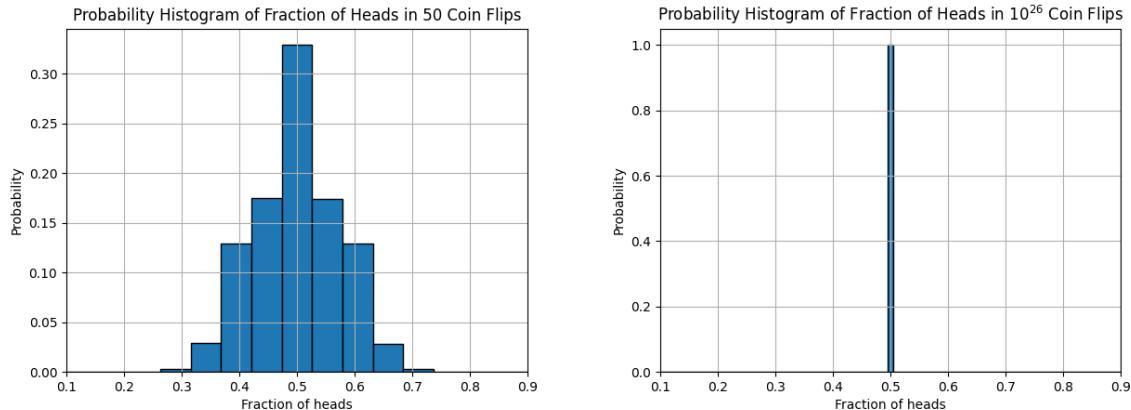
In the present context, setting  $p = q = \frac{1}{2}$  (for an unbiased coin) and defining  $f \equiv n/N$ , the fraction of tosses giving heads, we have

$$\bar{f} = \frac{\bar{n}}{N} = p = \frac{1}{2}$$

and

$$(\overline{\Delta f^2})^{1/2} \equiv \frac{(\overline{\Delta n^2})^{1/2}}{N} = \left(\frac{pq}{N}\right)^{1/2} = \frac{1}{2N^{1/2}} \quad (1)$$

Equation 1 shows that the typical deviation of  $f$  from its mean value is vanishingly small (it is  $O(N^{-1/2})$  for  $N$  large). Thus for large  $N$  we can be very sure that  $f$  will always effectively coincide with its mean (cf Figure 1b). The virtual certainty that comes from dealing with large numbers is one of the distinctive features of statistical physics.



- (a) For  $N = 50$  tosses, we can be *reasonably* sure that  $f$  will be close to 0.5
- (b) For  $N = 10^{26}$  tosses We can be *absolutely* sure that  $f$  will be indistinguishable from 0.5

Figure 1: Left: a probability histogram of the fraction of heads  $f$  obtained from repeated trials of  $N = 50$  coin tosses. Right: for a vary large number  $N = 10^{26}$  tosses the histogram narrows essentially to a  $\delta$  function.

You can investigate for yourself the effect of changing the number of coin tosses  $N$  in this python simulation (copy it into your favourite Python runtime environment):

```

import random
import matplotlib.pyplot as plt
import numpy as np

def flip_coin(n):
    """Simulate flipping a fair coin n times."""
    outcomes = [random.choice(['H', 'T']) for _ in range(n)]
    heads_count = outcomes.count('H')
    return heads_count / n

def simulate_coin_flips(N, M):
    """Simulate flipping a fair coin N times and repeat M times."""
    fractions = [flip_coin(N) for _ in range(M)]
    return fractions

def plot_probability_histogram(fractions):
    """Plot a histogram of the probabilities of the fractions of heads."""
    weights = np.ones_like(fractions) / len(fractions)
    plt.hist(fractions, bins=np.linspace(0, 1.0, 50), weights=weights, edgecolor='black')
    plt.xlabel('Fraction of heads')
    plt.ylabel('Probability')
    plt.xticks(np.arange(0, 1, 0.1))
    plt.xlim(0.1, 0.9)
    plt.title('Probability Histogram of Fraction of Heads in {} Coin Flips'.format(N))
    plt.grid(True)
    plt.show()

if __name__ == "__main__":
    N = int(input("Enter the number of times to flip the coin each time: "))
    M = int(input("Enter the number of times to repeat the simulation: "))
    fractions = simulate_coin_flips(N, M)
    plot_probability_histogram(fractions)

```

## 2. Foundations: equilibrium of an isolated system

### 2.1 Overview: the aims of this chapter

Consider a macroscopic system (a portion of matter, containing an Avogadro-sized number  $N$  of ‘microscopic constituents’ or ‘particles’ which we shall take to be isolated from the rest of the world so that no energy (either in the form of heat or work) may enter or leave the system. The system thus has a constant total energy which we denote by  $E$ .

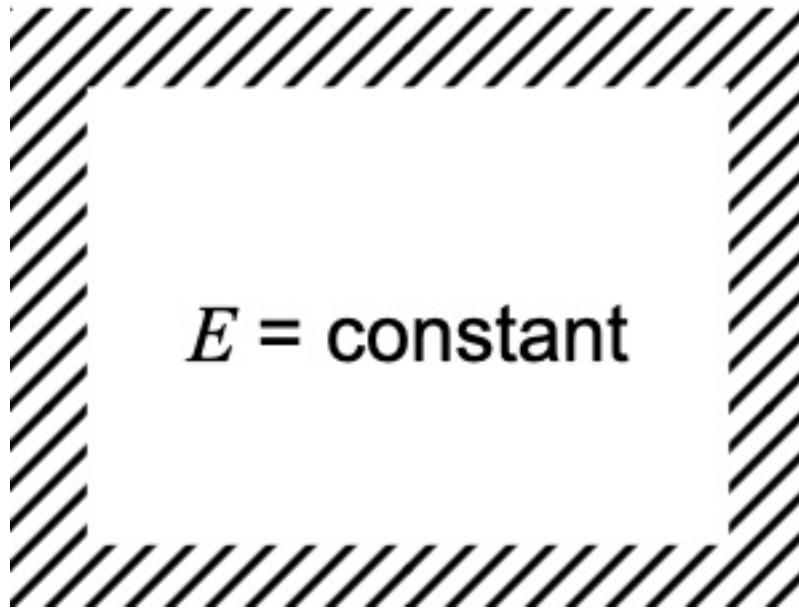


Figure 1: Schematic of a system isolated from its surroundings

We shall develop the statistical mechanics view of such a system. Specifically we shall establish the microscopic significance of entropy and temperature, and begin to see the simplifications due to the large value of  $N$ .

The arguments will be general (applicable irrespective of what the system actually comprises). They will be illustrated with references to two models of real systems:

- the ideal gas model
- a simple model magnet

The ideal gas model is familiar from PoM. The model magnet that we consider here is defined as follows (we will discuss its physical origins more fully in chapters 6 and 7).

- It comprises an array of  $N$  atoms (eg crystal) each with a magnetic dipole moment  $\mathbf{m}$  in an applied magnetic field  $\mathbf{H}$ . The energy of a dipole resides entirely in its interaction with the field which is given by  $\epsilon = -\mathbf{m} \cdot \mathbf{H}$
- In view of quantum mechanics (we shall not discuss the details here),  $\mathbf{m}$  is either
  - parallel to  $\mathbf{H}$  with energy  $\epsilon = -mH$  (ground state)
  - or antiparallel to  $\mathbf{H}$  with energy  $\epsilon = +mH$  (excited state)
- Each magnetic moment then has the energy level diagram shown on the right with two levels.

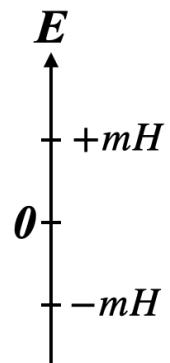


Figure 2: Energy-levels for one magnetic atom in a magnetic field in a simple model magnet

## 2.2 Microstates and macrostates

See Mandl Chapter 2

The terms microstate and macrostate constitute two different levels of description of a macroscopic system; we define them and explore their relationship.

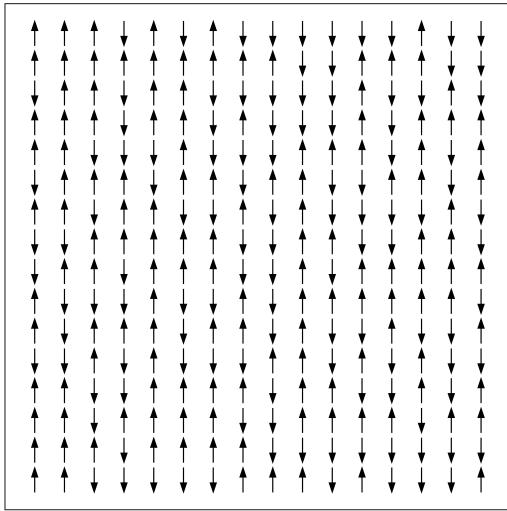
Key point 1

A **microstate** is a complete specification of the state of the system according to the microscopic model.

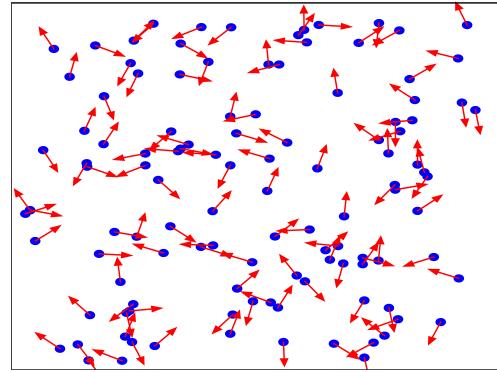
Thus a microstate is the most detailed description of the state of the system we can provide and is dictated by the microscopic model.

**Examples:**

1. Specifying the microstate of the model magnet means specifying the orientation of each of the  $N$  dipoles as in Figure 3a – or equivalently specifying which of the two rungs of its energy level ladder (cf Figure 2) it occupies.
2. Specifying the microstate of the (ideal) gas means (if we choose to use classical language) specifying the positions and velocities of each and every molecule (cf. Figure 3b).



(a) A microstate of a model magnet entails knowledge of the orientation (up or down) of each dipole. This is one possible microstate.



(b) A microstate of a model gas entails knowledge of the positions (blue circles) and velocities (red arrows) of all the molecules. Here is one example.

The microstate will change continually as the particles exchange energy with one another. For instance the molecules of a gas will continually be changing their position and their velocities as they collide with one another; The magnetic dipoles will undergo frequent transitions (hop from rung to rung) under the influence of their mutual interactions.

### 💡 Key point 2

A **macrostate** is a limited description of the state of the system given by the values of macroscopic variables of interest

To understand this, we must first consider macroscopic variables.

- Macroscopic properties are properties reflecting aggregate behaviour of a large number of constituents.
- Some macroscopic properties are fixed by constraints e.g. total energy  $E$ , and number  $N$  are fixed macroscopic properties in the isolated systems we are considering.
- Other macroscopic variables are free e.g. the number of ideal gas molecules in the left hand side of the box in Figure 3b can take on different values

Now we see that a macrostate is a description which depends on what macroscopic properties we are interested in i.e. we have some freedom in choosing what are to be the macrostates. We adopt the notation of denoting the free macroscopic properties of interest (if any)  $\{\alpha\}$  and we

label macrostates by  $N, E, \{\alpha\}$  (note that to lighten the notation  $N$  is often dropped but  $E$  is usually retained, however we shall retain both for the time being.)

### 💡 Key point 3

To any one macrostate there correspond in general very many microstates.

The number of microstates corresponding to a macrostate  $(N, E, \{\alpha\})$  is called the weight of the macrostate and is denoted  $\Omega(N, E, \{\alpha\})$  or more lazily  $\Omega(E, \{\alpha\})$ .

#### Example of model magnet:

Here we illustrate the idea of microstates and macrostates. To begin with we do not consider any free macroscopic variables thus the macrostates are simply labelled by  $N, E$  (which are fixed). We denote by  $n_i (i = 1, 2)$  the number of dipoles in levels 1 ( $\epsilon = -mH$ ) and 2 ( $\epsilon = +mH$ ) (cf Figure 2). In fact  $n_i$  are determined by the constraints:

$$n_1 + n_2 = N, \quad n_2 - n_1 = \frac{E}{mH}$$

so that

$$n_1 = \frac{1}{2}(N - \frac{E}{mH}), \quad n_2 = \frac{1}{2}(N + \frac{E}{mH})$$

As we stated before a microstate is a complete specification of which state every dipole is in. We now calculate the number of microstates which correspond to the values of  $n_2$ ,  $n_1 = N - n_2$

$$\Omega(N, E) = \binom{N}{n_2} = \frac{N!}{(N - n_2)!n_2!} = \frac{N!}{n_1!n_2!}$$

To lighten the notation, set  $n_2 = n$ ,  $n_1 = N - n$ , with  $n \equiv \frac{N}{2}(1 + \frac{E}{NmH})$  (the number of dipoles in the ‘excited’ state) we can then write

$$\Omega(N, E) = \frac{N!}{n!(N - n)!} \tag{1}$$

### 🔥 If you don't follow this

Here is the reasoning, just like that in chapter 1.4, but now couched in the language of dipoles rather than coins. Take a hatful of  $N$  dipoles; pick one which is to be in level  $i = 1$ ; this can be done in  $N$  ways; from the remaining pool of  $N - 1$  pick a second destined for level  $i = 1$ ; this can be done in  $(N - 1)$  ways. Repeating until you have picked all  $n_1$

required for level 1, you have a total of  $N \times (N-1) \times (N-2) \cdots (N-n_1+1) = N!/(N-n_1)!$  different hat-drawing-possibilities. All the rest of the dipoles must be in level 2, so there are no further choices to make. But now we must recognise that many of the ‘hat-drawing-possibilities’ actually lead to the same microstate. The microstate is defined by saying what state each dipole is in; so it doesn’t matter if dipole 1 is the first or the last to be selected for level 1; to allow for this ‘overcounting’ we must divide by  $n_1!$ , the number of hat drawing possibilities which assign  $n_1$  specified dipoles to level 1.

### Explicit example $N = 3$

Here we can label the macrostates by  $n = 0, 1, 2, 3$

Macrostate	Microstates	Weight
$n = 0, E = -3mH$	$\downarrow\downarrow\downarrow$	1
$n = 1, E = -mH$	$\downarrow\downarrow\downarrow$ $\downarrow\uparrow\uparrow$ $\downarrow\downarrow\downarrow$	3
$n = 2, E = +mH$	$\uparrow\uparrow\downarrow$ $\uparrow\downarrow\uparrow$ $\downarrow\uparrow\uparrow$	3
$n = 0, E = +3mH$	$\uparrow\uparrow\uparrow$	1

(Exercise: repeat for the case  $N = 4$ .)

Taking the logarithm of Equation 1 and invoking the Stirling approximation  $\ln N! \approx N(\ln N - 1)$ , good for large  $N$ , we can write

$$\begin{aligned}
 \ln \Omega(N, E) &= \ln N! - \ln n! - \ln(N-n)! \\
 &= N \ln N - N - n \ln n + n - (N-n) \ln(N-n) + (N-n) \\
 &= N \left[ \ln N - \frac{n}{N} \ln n - \left(1 - \frac{n}{N}\right) \ln(N-n) \right] \\
 &= N \left[ -\frac{n}{N} \ln \left(\frac{n}{N}\right) - \left(1 - \frac{n}{N}\right) \ln \left(1 - \frac{n}{N}\right) \right]
 \end{aligned}$$

where we have used a ‘trick’ to write:

$$\ln N = -\frac{n}{N} \ln \frac{1}{N} - \left(1 - \frac{n}{N}\right) \ln \frac{1}{N}$$

Thus

$$\frac{1}{N} \ln \Omega(N, E) = s(n/N)$$

where  $s(x) = -(1-x) \ln(1-x) - x \ln x$

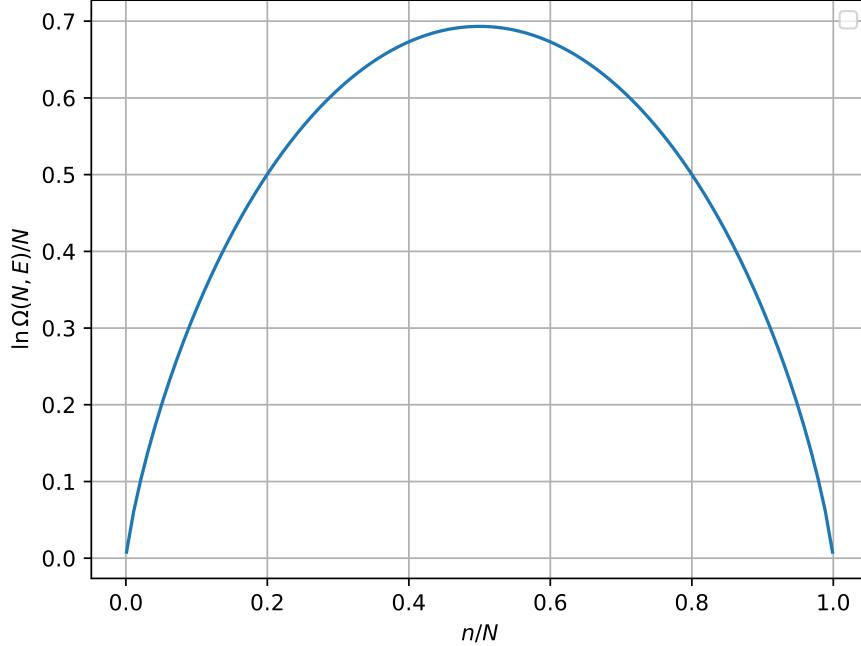


Figure 4: Logarithm of the weight function,  $\ln \Omega(N, E)/N$ , for the simple model magnet, plotted as a function of the fraction  $n/N$  of dipoles in excited states;  $n$  and  $E$  are related by  $n = \frac{N}{2}(1 + E/NmH)$ .

The logarithm of  $\Omega(N, E)$  is displayed in Figure 4. Note the key features:

- When  $E$  has its minimum value,  $E = -NmH$  (ie  $n = 0$ ), then  $\ln \Omega(N, E) = 0$  and  $\Omega(N, E) = 1$ . There is only a single microstate associated with this macrostate.
- As  $E$  increases from its minimum,  $\ln \Omega(N, E)$  increases steeply.
- For  $E = 0$  (ie.  $n = N/2$ ) we have a maximum where  $\ln \Omega(N, E) = N \ln 2$ , and  $\Omega(N, E) = 2^N$ . In general the logarithm of  $\Omega$  is proportional to the size of the system (we say that it is ‘extensive’)
- Since  $\Omega = \exp(Ns(x))$  the weight function is exponentially large in  $N$  but the logarithm is proportional to  $N$ .

The scaling of  $\Omega(N, E)$  with  $N$  is demonstrated by Figure 5 for systems of  $N = 2, 4, 8, 16, 32, 64$  dipoles. We see that the number of microstates for each energy increases strongly with  $N$  and  $\Omega(N, E)$  gets narrower. This trend continues so that for Avogadro's number of dipoles, the plot Figure 6 of  $\Omega(E)$  is practically infinitely high and infinitesimally narrow. This is the key insight of this exercise: for large  $N$  there are overwhelmingly more microstates associated with the ‘equal-shares’ macrostate than there are associated with any significantly different macrostate. Problem 2.2 invites you to explore the properties of the weight function.

For  $N = 10^{23}$  dipoles, Figure 6 shows that the weight function is essentially a  $\delta$  function

In order to understand what we mean by free macroscopic variables  $\{\alpha\}$  we consider as an example the number of dipoles that are in the excited state in the left hand side of our array of  $N$  dipoles. We denote this number by  $n_L$ . This number is not fixed by our constraints; we only have to satisfy  $n_R + n_L = n$  where  $n_R$  is the number of excited dipoles in the right hand region.

We now label our macrostates by  $N, E, n_L$

We can calculate the weight of a macrostate by combinatorics: in the left hand region we can choose the  $n_L$  excited dipoles from  $N/2$  and similarly in the right hand region  $n_R = n - n_L$  are chosen from  $N/2$ . Thus

$$\Omega(N, E, n_L) = \binom{N/2}{n_L} \binom{N/2}{n - n_L}$$

In the special case where  $E = 0$ , so  $n_1 = n_2 = n = N/2$  and  $n_L = N/2 - n_R$ , one can show using Stirling’s approximation (see question 2.3) we have that

$$\frac{1}{N} \ln \Omega(N, 0, n_L) \simeq s(n_L/(N/2))$$

Since  $n_L$  is free, it is in principle possible for the system to move between different macrostates, moreover the different available macrostates have different weights. In particular we see that macrostates with  $n_L \simeq N/4$  have huge weights compared to say  $n_L \simeq N/2$ . Question 2.3 explores this point

### 2.3 Significance of the weight function: Entropy

We have seen that the logarithm of the weight function is the quantity proportional to  $N$ . We now state an important point relating this quantity to the entropy of a macrostate

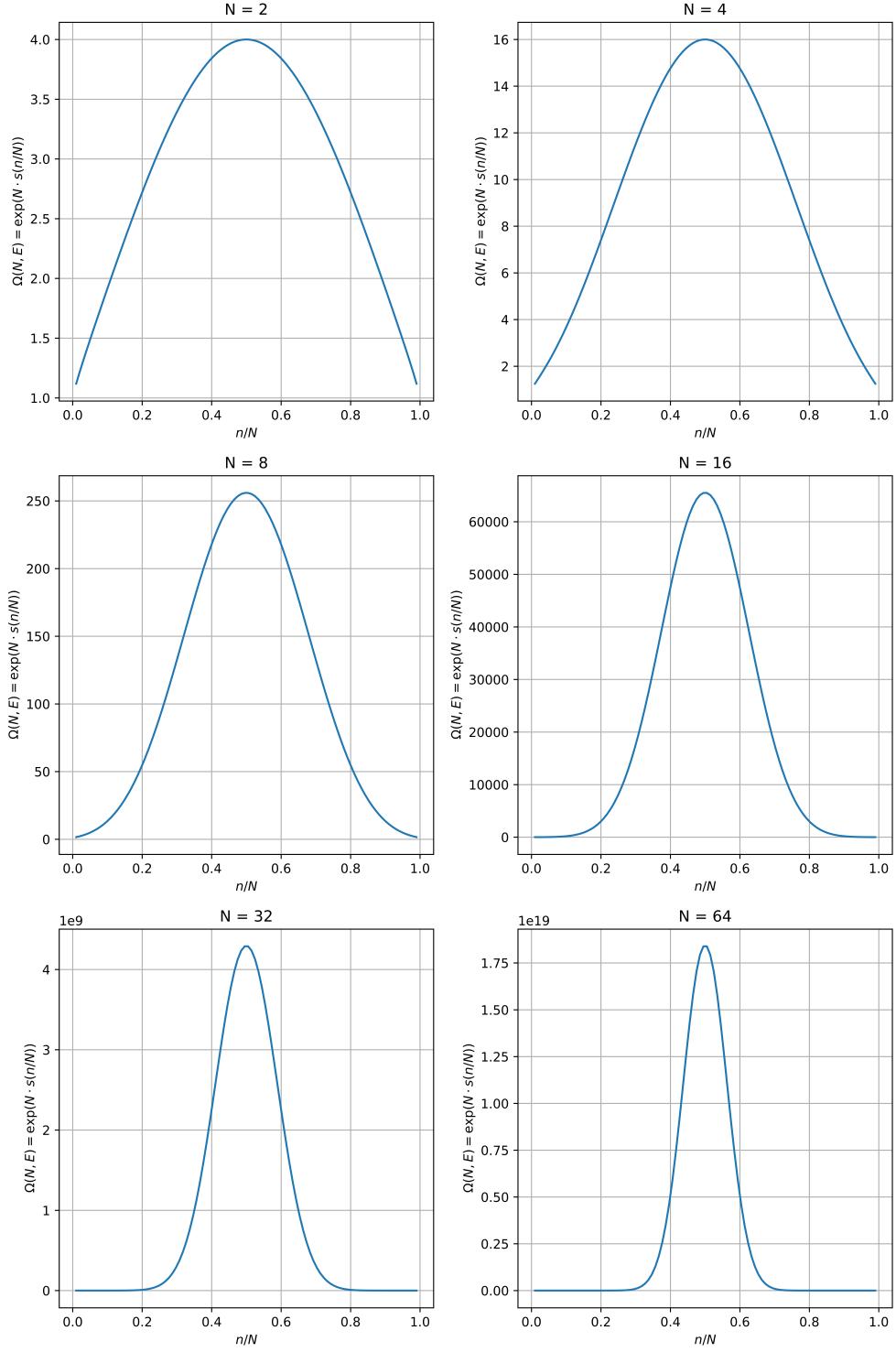


Figure 5: Weight function  $\Omega(N, E)$ , for the simple model magnet comprising  $N = 2, 4, 8, 16, 32, 64$  dipoles, plotted as a function of the fraction  $n/N$  of dipoles in excited states;  $n$  and  $E$  are related by  $n = \frac{N}{2}(1 + E/NmH)$ . Note the y-axis scales and the narrowing of the function with increasing  $N$ .

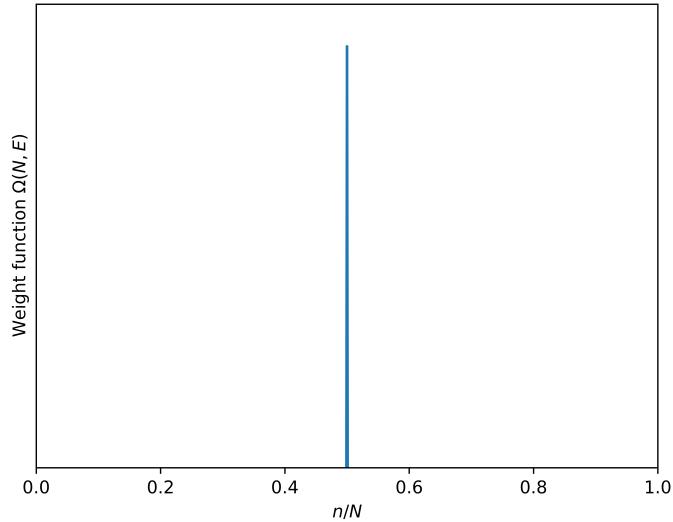


Figure 6: For  $N = 10^{23}$  dipoles the weight function  $\Omega(N, E)$  is essentially a delta function.

Key point 4

$$S(N, E, \{\alpha\}) = k \ln \Omega(N, E, \{\alpha\})$$

where  $k = 1.381 \times 10^{-23} J/K$  is Boltzmann's constant. The entropy  $S(N, E, \{\alpha\})$  of a macrostate is *defined* by this relation which for the moment we consider as a postulate. The logarithmic connection between entropy and probability was formulated by Boltzmann (though the above form of the equation is due to Planck). We shall refer to it as the [Planck equation](#).

### 3. Entropy, equilibrium, and the second law

See Mandl chapter 2

#### 3.1 Principle of equal a-priori probabilities and its consequences

We can now formulate the axiomatic foundation of statistical physics:

💡 Key Point 5: The principle of equal a priori probabilities:

An *isolated* system is equally likely to be found in any one of its allowed microstates.

Although we state it here as a postulate, it can be justified through the degree of belief view of probabilities: our microscopic model prescribes the available microstates which is the most detailed description of the state we can (or are prepared) to offer. Since we have no further information we have to assume that each of these microstates is equally likely. Recalling that the weight of a macrostate is the number of corresponding microstates, a direct consequence of key point 5 is

💡 Key Point 6:

The probability that an isolated system, of energy  $E$ , will be found in a *macrostate*  $N, E, \{\alpha\}$  is proportional to the weight  $\Omega(N, E, \{\alpha\})$ , ie the number of microstates associated with the macrostate  $N, E, \{\alpha\}$

In other words we can think of the microstate as the outcome of a trial (measuring which microstate our system is in) and the number of outcomes corresponding to a macrostate is given by the weight. Thus the probability of being in a macrostate is proportional to the weight. Since different macrostates have different weights they will have different probabilities.

We saw that since weight functions are typically exponentially large in  $N$ , the maximum for the weight function is extremely sharp. This means with macrostates in which the (free) macroscopic variables maximise the weight function are overwhelming more probable than other possible values. This was illustrated in chapter 2 where we plotted the weight function of the model magnet [link to image](#) and [link to image](#).

The probability that an isolated system will be found in macrostate  $N, E, \{\alpha\}$  is proportional to the weight of macrostate  $\Omega(N, E, \{\alpha\})$ . From the sharpening of the weight functions we deduce another key point

 Key Point 7:

The equilibrium values  $\{\alpha^*\}$  of the macroscopic properties  $\{\alpha\}$  of an isolated system are those that maximise the weight function  $\Omega(N, E, \{\alpha\})$

Thus while, in principle, macroscopic properties  $\Omega(N, E, \{\alpha\})$  are free to take up a wide range of values, in practice they are sharply defined so that equilibrium is characterised by macroscopic properties unique to the equilibrium state that is singled out by the constraints.

Saying that a thermodynamic system is ‘in equilibrium’ is saying (essentially) that enough time has elapsed since the last change in constraints for the system to find its way to the most probable macrostate consistent with those constraints.

A consequence of this, together with KP4, is that

 Key Point 8:

The equilibrium entropy of an isolated system of  $N$  constituents with energy  $E$  is  $S(N, E) = k \ln \Omega(N, E, \{\alpha^*\})$

### 3.2 The second law

To illustrate the second law we consider the model magnet of the previous chapter:  $N$  atoms (magnetic dipoles), energy  $E$  and  $n$  atoms in excited states. Let us consider starting from an initial condition where all the excited dipoles are in the left hand half of the system and none in the right. One can think of inserting a thermally insulating wall to ensure that no energy can be transferred between the two sub-systems.

Now when we remove the wall the system finds itself in a macrostate which is far from the most probable. The system will evolve, by more and more of the excited dipoles appearing in the right hand half, through macrostates which have greater and greater weights. Finally it reaches the macrostate where there are equal numbers of excited dipoles in the left and right hand sides and which is the macrostate with the maximal weight. Of course when we say equal we mean equal to within  $O(N^{1/2})$ .

In effect there has been a flow of energy from the left hand side to the right hand side of the system until equilibrium was reached. A related example is of a gas starting with all molecules in the left hand half of the system. Similar considerations apply to that system although now there is a flow of particles or mass from left to right.

We summarise with a statement not quite worthy of being a key point, as we will not pursue dynamics much further, but which is nevertheless worth a box

An isolated system evolves through macrostates with greater and greater weight until it reaches the macrostate with the largest weight and stays there.

Now considering the entropy, which is by Planck's relation proportional to the logarithm of the weight we see that

*The entropy of an isolated system can only increase*

which is a statement of the second law of thermodynamics. Moreover we now understand why the law holds only in a statistical sense. In the example of the ideal gas there is nothing in Newton's laws to prevent all the molecules moving into the left hand side, which would mean a decrease in entropy. However we know statistically that this will basically never happen.

### 3.3 Entropy and disorder

Entropy is often referred to as a measure of disorder. For this statement to be meaningful we have to know what we mean by ordered and disordered.

Basically the idea is that when something is 'disordered' it may be so in many different ways, whereas 'order' places many restrictions on the ways things can exist.

Consider the example of a bedroom. For a bedroom to be tidy i.e. ordered, everything has to be in its correct place, and there are not many ways to place ones possessions and clothes correctly in the room. However in an untidy, i.e. disordered room, there are many, many different ways for things to be scattered about.

Another example is problem question 2.6 where a order/disorder transition is discussed. There copper and gold atoms are arranged in a crystal lattice. In the ordered phase the copper atoms occupy certain positions and the gold atoms occupy other positions. This is an orderly arrangement compared to the disordered phase where any atom can occupy any of the positions. In that question entropy differences are calculated.

Question 2.5 highlights that a liquid (water) is more disordered than a solid (ice).

### 3.4 Additivity of entropy

Consider dividing our system (whether it be the model magnet or ideal gas) in half i.e. into two subsystems 1 and 2, each with number  $N_i$  of constituents and containing energy  $E_i$  where  $i = 1, 2$ . For example, if we divide our array of  $N$  magnetic atoms into two sub-systems of fixed  $N_1 = N_2 = N/2$  which are in thermal contact (they exchange energy). The fixed total energy corresponds to a total of say  $n = N/2$  excited dipoles. The weight for the macrostate where  $n_1$  of the dipoles in the left half are excited is a product of weight functions for the two halves

$$\Omega(N, E, n_1) = \Omega_1(N_1, E_1, n_1)\Omega_2(N_2, E_2, n - n_1)$$

This factorisation of the weight function into two terms from each subsystem is quite general. Thus using Planck's relation we see that for the two sub-systems

$$S(N, E, n_1) = S_1(N_1, E_1, n_1) + S_2(N_2, E_2, n_2).$$

This property of the entropy being the sum of the entropy of the subsystems is quite general and is referred to as additivity of the entropy i.e. the entropy of a composite system is the sum of entropies of macroscopic component subsystems. In particular in an equilibrium system, where the energies of the two sub systems are  $E_1, E_2 = E_1^*, E_2^*$

$$S(N, E) = S(N_1, E_1^*) + S(N_2, E_2^*)$$

### 3.5 Entropy and temperature

We now analyse the condition for equilibrium more mathematically. For our two systems in thermal equilibrium we have

$$S(E) = S_1(E_1) + S_2(E_2) \quad E = E_1 + E_2.$$

where we have been lazy and dropped the  $N$ s since we take them to be fixed.

We wish to consider the effect of changing  $E_1$  to  $E_1 + dE_1$  which implies changing  $E_2$  by  $dE_2 = -dE_1$ , since total  $E$  is fixed.

Now consider the effect on the entropy

$$\begin{aligned} dS &= dS_1 + dS_2 \\ &= \left( \frac{\partial S_1}{\partial E_1} \right) dE_1 + \left( \frac{\partial S_2}{\partial E_2} \right) dE_2 \\ &= \left( \frac{\partial S_1}{\partial E_1} - \frac{\partial S_2}{\partial E_2} \right) dE_1 \end{aligned}$$

Now since the entropy is an extremum (i.e. it should be maximised)  $dS$  must be 0. Therefore we deduce

$$\frac{\partial S_1}{\partial E_1} = \frac{\partial S_2}{\partial E_2}$$

Recalling the zeroth law of thermodynamics, which states that two systems in thermal equilibrium must share the same temperature, and comparing with the condition for equilibrium that we have derived, we deduce that

$$\frac{\partial S_1}{\partial E_1} = \text{some function of } T$$

In fact the correct identification is

 Key Point 9:

$$\frac{1}{T} = \frac{\partial S}{\partial E}$$

(To understand why this particular function of  $T$  is the correct one will need the ideas of free energy minimisation that we shall meet later and in question 3.5)

We now ask what happens when we put two systems at different temperatures  $T_1$  and  $T_2$  in thermal contact. In this case we expect from the second law that  $dS > 0$ .

Using the identification of key point 9 we find

$$dS = \left( \frac{1}{T_1} - \frac{1}{T_2} \right) dE_1$$

So if  $T_1 > T_2$  we must have  $dE_1$  –ve for  $dS$  to be +ve and we see energy (or heat) flows from the hotter to the colder system.

One can generalise the arguments of this subsection to the consideration of systems free to exchange volume (e.g. a balloon) or particles (e.g. a liquid and gas in coexistence).

## 4 The Boltzmann distribution

See Mandl chapter 2.5 and Blundell & Blundell chapter 20

Previously we have considered an isolated system of fixed  $N, E$ . Here we consider systems with fixed  $N$  but instead of  $E$  being fixed, the system is at some fixed temperature  $T$ .

To be specific our system is in equilibrium with a heat bath (or heat reservoir) at constant  $T$ . The zeroth law of thermodynamics then implies our system is at temperature  $T$  and the thermal contact with the reservoir means that energy can be exchanged. Therefore the energy of our system is not fixed and the system can explore microstates of different energy. We expect microstates of the same energy to have the same probabilities (we have no reason to believe otherwise). However since we are at fixed  $T$  we might think that microstates of different energy should have different probabilities. In the following we substantiate this.

### 4.1 Derivation of the Boltzmann distribution

Figure 1 illustrates the setup we shall use.

We will refer to the system of interest as the ‘system’ and the composite system (system plus heat bath) as the ‘composite’. The composite is isolated and has fixed total energy  $E_{TOT}$ . Thus for the composite all microstates are equally likely. But we are interested in microstates of the system.

Let the probability that system is in a given microstate labelled  $i$ , of energy  $E_i$ , be  $P_i$ . Now if the system is in this microstate there is energy  $E_{TOT} - E_i$  left for the bath, and this corresponds to many possible microstates for the bath. Since for the composite all microstates are equally likely, we deduce

$$P_i = \text{constant} \times \Omega_b(E_{TOT} - E_i)$$

where the subscript  $b$  refers to the bath.

 If you don't follow this

Here is the reasoning. For fixed total energy  $E_{TOT}$ , the total number of microstates of the composite of system and bath is  $\Omega = \Omega_s \Omega_b$ , a constant. If we consider only the microstate of the system with energy  $E$ , then their number is  $\Omega_s(E) = \Omega / \Omega_b(E_{TOT} - E)$ . Now as all microstates of the system having a given energy are equally likely, the probability  $P_i(E)$  of the system being in a specific microstate  $i$  with energy  $E$  is  $P_i(E) = 1 / \Omega_s(E) = \Omega^{-1} \Omega_b(E_{TOT} - E)$ .

Now using the Planck relation (key point 4)

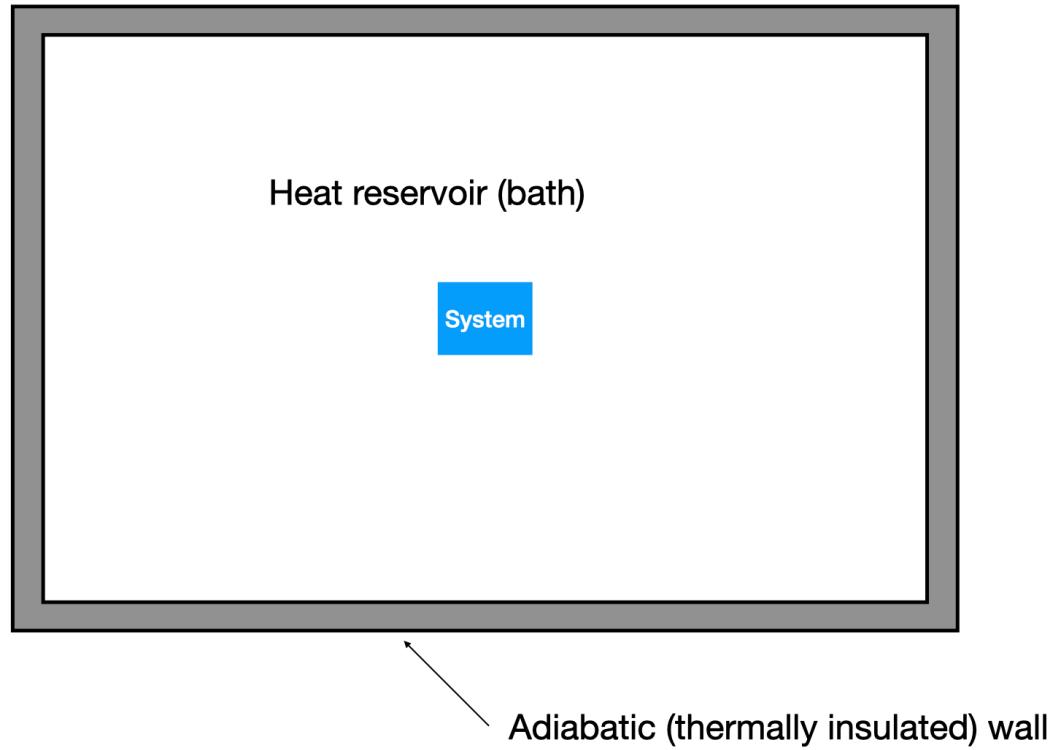


Figure 1: Composite system of a large heat reservoir within which is the system of interest

$$\Omega_b(E_{TOT} - E_i) = \exp\left(\frac{S_b(E_{TOT} - E_i)}{k_B}\right)$$

Since  $E_i \ll E_{TOT}$  we can perform a Taylor expansion

$$S_b(E_{TOT} - E_i) = S_b(E_{TOT}) - E_i \frac{\partial S_b}{\partial E_b} + \frac{E_i^2}{2} \frac{\partial^2 S_b}{\partial E_b^2} + \dots$$

Only keeping the first two terms (it turns out that the third term is small) and using key point 9 to identify

$$\frac{1}{T_b} = \frac{\partial S_b}{\partial E_b}$$

we find

$$S_b(E_{TOT} - E_i) \simeq S_b(E_{TOT}) - \frac{E_i}{T}$$

The first term on the right hand side is a constant therefore when we insert this in the expression for  $P_i$  (or  $\Omega_b(E_{TOT} - E_i)$ ) above we obtain

$$P_i \propto \exp\left(-\frac{E_i}{kT}\right)$$

In order that the probability be normalised we require

$$\sum_i P_i = 1$$

where the sum is over all microstates of the system which implies

$$P_i = \frac{\exp\left(-\frac{E_i}{kT}\right)}{\sum_j \exp\left(-\frac{E_j}{kT}\right)}$$

Thus the Boltzmann distribution for the probability that a system in equilibrium with a heat bath at temperature  $T$  is in microstate  $i$  is

 Key Point 10:

$$P_i = \frac{1}{Z} \exp(-\beta E_i) \quad \text{where } Z = \sum_j \exp(-\beta E_j) \quad \text{and } \beta = \frac{1}{kT}$$

You have met this distribution in year 1 properties of matter. for example in considering an isothermal atmosphere where particles of mass  $m$  have height probability distribution  $P(h) \propto$

$\exp(-mgh)$ . To see this is a Boltzmann distribution recall that the potential energy of  $kT$  such a particle is  $mgh + \text{constant}$ .

The Boltzmann distribution is also known as the canonical distribution (canonical simply means standard).

The quantity  $Z$  is known as the partition function or partition sum. Although introduced quietly here as a normalisation we shall see that it holds the key to calculating all properties of the system!

## 4.2 Boltzmann distribution for a single particle system

As a simple application of the Boltzmann distribution, let us consider a single magnetic dipole in equilibrium with a heat bath. (This is discussed in Baierlein 5.3 but with some differences in notation.)

Recall that a single dipole has two microstates  $\downarrow, \uparrow$  with energies  $-mH, +mH$ .

A simple application of the Boltzmann distribution gives:

$$P(\downarrow) = \frac{\exp\left(\frac{mH}{kT}\right)}{Z}, \quad P(\uparrow) = \frac{\exp\left(-\frac{mH}{kT}\right)}{Z}$$

where the partition function  $Z$  is given by:

$$Z = \exp\left(\frac{mH}{kT}\right) + \exp\left(-\frac{mH}{kT}\right) = 2 \cosh\left(\frac{mH}{kT}\right)$$

The average energy of the dipole is given by:

$$\bar{E} = \sum_i E_i P_i = \frac{1}{Z}[-mHP(\downarrow) + mHP(\uparrow)] = -mH \frac{\sinh\left(\frac{mH}{kT}\right)}{\cosh\left(\frac{mH}{kT}\right)} = -mH \tanh\left(\frac{mH}{kT}\right)$$

To see the temperature dependence of  $\bar{E}$ , we plot it as a function of  $y = \frac{kT}{mH}$ :

$$\bar{E} = -mH \tanh\left(\frac{1}{y}\right)$$

You should be able to sketch this curve by recalling the properties of the tanh function:

$$\tanh x \approx x \quad \text{for small } x, \quad \tanh x \approx 1 \quad \text{for } x \gg 1$$

The curve is plotted in Figure 2.

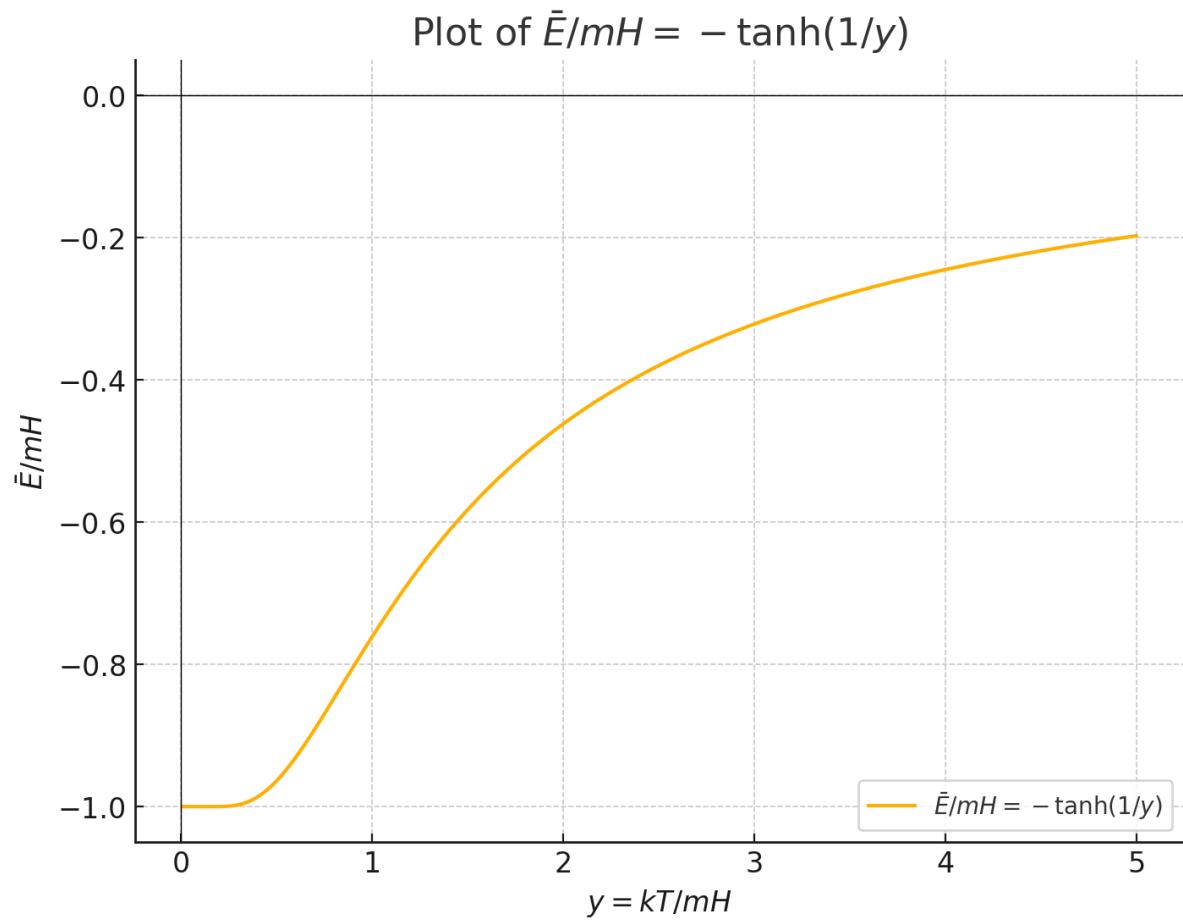


Figure 2: Mean energy  $E/mH$  versus scaled temperature for a single dipole

### 4.3. Single constituent in an isolated system for large $N$

In order to gain more insight into the generality of the Boltzmann distribution, we return briefly to the scenario of chapters 3 and 4 where we considered an isolated system with a large number  $N$  of constituents. We saw we could define an entropy  $S(E)$  through the Planck relation and a temperature through key point 9.

In particular, for the model magnet, we used the Planck relation to find the entropy:

$$S(E) = -Nk[x \ln x + (1-x) \ln(1-x)], \quad x = \frac{n}{N} = \frac{1}{2} \left( 1 + \frac{E}{NmH} \right)$$

where  $n$  is the number of excited dipoles and  $N$  is the total number of dipoles.

Using key point 9, we find:

$$\frac{1}{T} = \frac{\partial S}{\partial E} = \frac{dx}{dE} \frac{\partial S}{\partial x} = -\frac{Nk}{2NmH} [\ln x + 1 - \ln(1-x) - 1] = -\frac{k}{2mH} \ln \left( \frac{x}{1-x} \right)$$

Inverting this, we find:

$$\exp \left( -\frac{2mH}{kT} \right) = \frac{x}{1-x}$$

so that:

$$x = \frac{n}{N} = \frac{\exp \left( -\frac{2mH}{kT} \right)}{1 + \exp \left( -\frac{2mH}{kT} \right)} = \frac{\exp(-\beta mH)}{\exp(\beta mH) + \exp(-\beta mH)}$$

The frequency definition of probability implies that  $\frac{n}{N}$  is the probability that a single dipole is in the excited state. We see we recover the Boltzmann distribution of the previous subsection for a single dipole!

At first, this seems very confusing: for an isolated system, we have all microstates equally likely—so where did the Boltzmann distribution suddenly appear from?

The answer is to realize that a single dipole exchanges energy with the  $N - 1$  other dipoles. Therefore, for large  $N$ , the other dipoles act as a heat bath for the single dipole, which is why we recover the Boltzmann distribution for a single dipole.

Another way of thinking about this is to remember that although all microstates of the isolated system are equally likely, we have the constraint of total fixed energy. Therefore, for a single dipole, its two states are not equally likely since the state it is in dictates how much energy is left to share among the other dipoles. Therefore, the global constraint of fixed energy induces a Boltzmann distribution for a single dipole.

## 5. Free energy minimisation

See Mandl 2.6

### 5.1. Energy and its fluctuations

We have seen that the energy of a system connected to a heat bath exhibits thermal fluctuations controlled by the Boltzmann distribution. Here we calculate the mean and variance of these fluctuations. The mean energy is defined as usual as:

$$\bar{E} = \sum_i E_i P_i$$

where  $P_i$  is the probability of the system being in microstate  $i$ . Inserting the Boltzmann distribution yields:

$$\bar{E} = \frac{1}{Z} \sum_i E_i \exp(-\beta E_i) = -\frac{1}{Z} \sum_i \frac{\partial}{\partial \beta} \exp(-\beta E_i) = -\frac{1}{Z} \frac{\partial Z}{\partial \beta}$$

Note the trick of using the derivative to bring down  $E_i$  from the argument of the exponential. A further manipulation to note is:

$$\frac{\partial \ln Z}{\partial \beta} = \frac{1}{Z} \frac{\partial Z}{\partial \beta}$$

and we find:

$$\bar{E} = -\frac{\partial \ln Z}{\partial \beta} = -\frac{dT}{d\beta} \frac{\partial \ln Z}{\partial T} = kT^2 \frac{\partial \ln Z}{\partial T}$$

Now we consider fluctuations of the energy. The variance is:

$$\overline{(\Delta E)^2} = \overline{(E - \bar{E})^2} = \bar{E}^2 - \bar{E}^2$$

It turns out that this variance is related to the heat capacity  $C$  defined by

$$C \equiv \frac{\partial \bar{E}}{\partial T} = \frac{d\beta}{dT} \frac{\partial \bar{E}}{\partial \beta} = -\frac{1}{kT^2} \frac{\partial \bar{E}}{\partial \beta}$$

The relationship (derived below) is

$$C = -\frac{1}{kT^2} [\overline{E^2} - \overline{E}^2] = \frac{(\overline{\Delta E})^2}{kT^2}$$

 Open to see derivation

In order to demonstrate the relation of the energy fluctuations to the heat capacity, we write out the final expression for  $C$  above using the Boltzmann distribution:

$$C = -\frac{1}{kT^2} \frac{\partial}{\partial \beta} \left( \frac{\sum_i E_i \exp(-\beta E_i)}{Z} \right) = -\frac{1}{kT^2} \left[ -\frac{\sum_i E_i \exp(-\beta E_i)}{Z^2} \frac{\partial Z}{\partial \beta} - \frac{\sum_i E_i^2 \exp(-\beta E_i)}{Z} \right]$$

which recalling:

$$\begin{aligned} \sum_i E_i \exp(-\beta E_i) &= \overline{E} = -\frac{1}{Z} \frac{\partial Z}{\partial \beta} \\ \sum_i E_i^2 \exp(-\beta E_i) &= \overline{E^2} \end{aligned}$$

yields:

$$C = -\frac{1}{kT^2} [\overline{E^2} - \overline{E}^2] = \frac{(\overline{\Delta E})^2}{kT^2}$$

It is interesting to pause and think what this means. The heat capacity can be thought of as the response of a macroscopic variable (the energy) to changing an external variable (the temperature). Thus, the response is related to the fluctuations of the macroscopic variable.

Since  $E$  is proportional to  $N$ , so will be  $C = \partial \overline{E} / \partial T$ . Therefore:

$$\overline{(\Delta E)^2} \propto N$$

and the fractional deviation:

$$\frac{(\overline{\Delta E})^{1/2}}{\overline{E}} \propto \frac{1}{\sqrt{N}}$$

From this, we see that for a system in equilibrium with a heat bath, although the energy is a free macroscopic variable, it takes on a sharply defined value  $\overline{E}$ . A consequence is that a large system in equilibrium with a heat bath should have essentially the same behavior as a large isolated system with fixed energy  $\overline{E}$ .

## 5.2. Energy distribution, entropy, and free energy

Since, in the Boltzmann distribution, the energy is a free macroscopic variable, we can label the macrostates of the system by their energy, and the weight of the macrostate of energy  $E$  is  $\Omega(E)$ , the number of microstates with energy  $E$ .

As in chapter 3, we can use the Planck relation to determine the entropy of a macrostate:

$$S(E) = k \ln \Omega(E)$$

We now return to the question of the sharpness of  $E$ . Let us consider again:

$$\bar{E} = \frac{1}{Z} \sum_i E_i \exp(-\beta E_i)$$

Generally, for a system of large  $N$ , we expect many of the microstates to have the same energy. Therefore, rather than sum over all microstates, it is more convenient to sum over the possible values of the energy:

$$\bar{E} = \frac{1}{Z} \sum_E \Omega(E) E \exp(-\beta E)$$

To understand this, realize that we are simply grouping together microstates with the same energy  $E$  of which there are  $\Omega(E)$ .

The same idea of grouping microstates with the same energy gives the probability of the system having energy  $E$ :

$$P(E) = \frac{1}{Z} \Omega(E) \exp(-\beta E)$$

Since we expect  $E \sim N$ ,  $\exp(-\beta E)$  is a sharply decreasing function. Also,  $\Omega(E)$  is exponentially large in  $N$  and  $\Omega$  is a sharply increasing function of  $E$ . The two conspire to give as a product a sharply peaked probability distribution around  $\bar{E}$ . This is plotted in Fig. X

To go further, we rewrite  $P(E)$  and use the Planck relation:

$$P(E) = \frac{1}{Z} \exp\left(-\beta\left(E - \frac{\ln \Omega}{\beta}\right)\right) = \frac{1}{Z} \exp(-\beta(E - TS)) = \frac{1}{Z} \exp(-\beta F)$$

where:

$$F(E) \equiv E - TS(E)$$

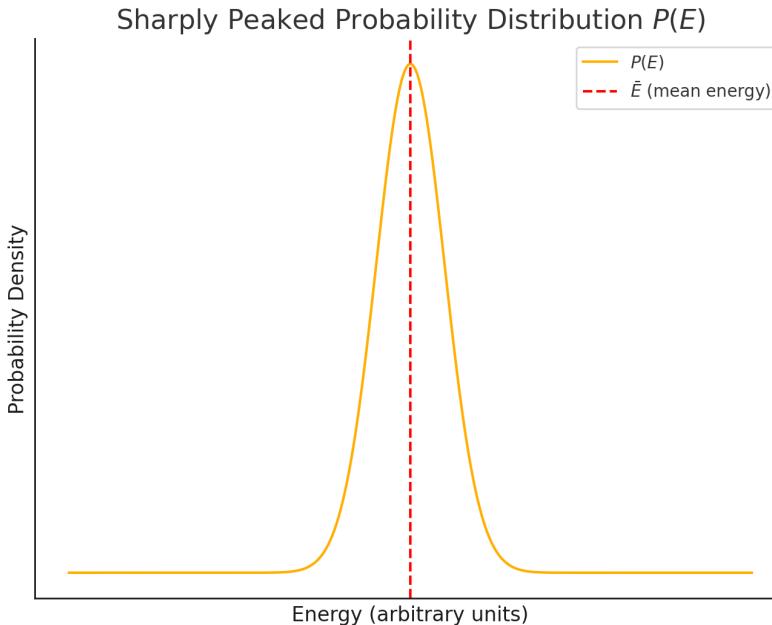


Figure 1: Sharply peaked probability distribution of the energy

$F$  is the Helmholtz free energy, which is often referred to in statistical mechanics simply as the free energy.

Now the peak of  $P(E)$  occurs when the argument of the exponential is maximized. Thus, we see that the equilibrium value of the energy  $E$  is determined by minimizing the free energy. This is a profound and general concept:

💡 Key Point 11:

The equilibrium values of the macroscopic properties  $E, \{\alpha\}$  of a system in equilibrium are such as to minimize the free energy  $F(E, \{\alpha\}) \equiv E - TS(E, \{\alpha\})$ .

We write equilibrium values as  $S(T)$ ,  $F(T)$ , etc.

### 5.3. Free energy: The energy–entropy competition

First, consider low  $T$ : here  $F$  is dominated by  $E$ , therefore minimizing  $F$  corresponds to minimizing  $E$  (exactly so for  $T = 0$ ).

Now consider high  $T$ : here  $F$  is dominated by  $S$ , therefore minimizing  $F$  corresponds to maximizing  $S$  (exactly so for  $T = \infty$ ).

Thus, one has a competition between energy and entropy, the arbiter of which is temperature.

However, it should always be borne in mind that equilibrium is such as to maximize the entropy of the universe under the constraint that energy is conserved. For example, in an isolated system, the energy is constant, therefore we just maximize the entropy. For a system in equilibrium with its environment, the energy is free to vary. Thus, the minimization of the free energy for a system in equilibrium tells us the best way to distribute energy between the system and the universe such that the entropy of the universe is maximized.

We now begin to understand the different phases in question 2.5:

- At low temperatures, ice forms. The crystal structure is energetically favorable and releases energy (latent heat of fusion) to the environment, which maximizes the entropy of the universe, i.e., from the viewpoint of the system, energy considerations win.
- At higher temperatures, storing more energy in the system gives the liquid phase (water), which has a high entropy and allows the entropy of the universe to be maximized. From the point of view of the system, entropy wins.

#### 5.4. Utility of the partition function

Let us first write  $Z$  as

$$Z = \sum_E \Omega(E) \exp(-\beta E)$$

In the same way as  $P(E)$  is an extremely sharp function, the summand is very sharply peaked at  $\bar{E}$  and of width  $N^{1/2}$  about the peak. Therefore, we can write

$$Z \approx \Omega(\bar{E}) \exp(-\beta \bar{E}) \times O(N^{1/2})$$

$$= \exp(-\beta F(T)) \times O(N^{1/2})$$

(since  $F(T) = F(\bar{E})$  for a large system in equilibrium). Thus

$$\ln Z = -\beta F(T) + O(\ln N^{1/2})$$

We see the last term can be ignored since we are in the ‘large  $N$ ’ regime where  $N \gg \ln N$ . Thus, the equilibrium value of the free energy

$$F(T) = \bar{E} - TS(T) = -kT \ln Z$$

This is a very important formula.

Also, the expression can be written in terms of the entropy

$$S(T) = k \ln Z + \frac{\bar{E}}{T}$$

Therefore, for a system in equilibrium at temperature  $T$ , all the thermodynamic properties can be obtained from  $Z$  from the following formulae:

💡 Key Point 12:

$$F = -kT \ln Z, \quad \bar{E} = kT^2 \frac{\partial}{\partial T} \ln Z, \quad S = k \ln Z + \frac{\bar{E}}{T}$$

## 7: More on magnetism; review

In this chapter we tie up some loose ends and take stock.

### 7.1 Paramagnetism and ferromagnetism

*See Mandl 3.1*

So far we have considered the ‘model magnet’ without really explaining why such a simple model is a good caricature of the real physics. This needs a knowledge of quantum mechanics that you have acquired in the quantum mechanics course.

The magnetic behaviour of solids is due to the behaviour of the electrons and/or nucleus. The simplest case is where the magnetic behaviour is due to a single ion of the solid molecule that has a single electron which acts as a ‘spin 1/2’. Quantum mechanics tells us that when we apply a magnetic field (which implies a ‘quantisation axis’) then the dipole moment is quantised parallel to the field as  $\pm m$ . This is the basis of the ‘model magnet’ which is a model of paramagnetism, i.e., magnetic ordering in response to an external field.

In the weakly interacting system, we assumed that energy was only stored in interaction between dipoles and the external field. In reality, the dipoles interact amongst themselves. The ‘model magnet’ is a good model when the molecules of the solid are large so that the single ions carrying the dipole moment are well separated (See Mandl 3.1).

Interactions between dipoles have two origins. Firstly, we have the classical electromagnetic interaction where each dipole produces a field that its neighbors sit in. This tends to align dipoles. Secondly, we have a quantum mechanical interaction which occurs when the wavefunctions of neighboring atoms overlap. This produces what is known as an ‘exchange interaction’, which generally tends to align dipoles and is stronger than the classical interaction. If the interaction between dipoles is strong enough, it can lead to the phenomenon of ferromagnetism, where the dipoles tend to align without the aid of an external field.

### 7.2 Simple model of ferromagnetism

As a very basic model of ferromagnetic interactions, let us restrict ourselves to a one-dimensional array of dipoles.

Here there is no external field, and the only energy comes from the ferromagnetic interactions between particles: if two neighboring dipoles are in the same direction, this is favorable, but if two neighboring dipoles are in opposite directions, this costs energy.

Thus, each ‘domain wall’ where neighboring dipoles are of opposite directions costs energy  $J$ , say.

$\uparrow\uparrow \cdots \uparrow\uparrow\downarrow\downarrow \cdots \downarrow\downarrow\uparrow\uparrow \cdots \uparrow\uparrow$

Example of two domain walls (three domains) in a one-dimensional array of  $N$  dipoles.

To estimate the dependence of the number of such domain walls on temperature, we can use a minimization of free energy argument.

Let us consider  $n$  domain walls. The energy cost is  $nJ$ .

The number of ways of arranging the  $n$  domain walls on the lattice is the number of ways of choosing  $n$  from the  $N - 1$  possible places the domain walls can be. (Do you see why it is  $N - 1$ ? but actually  $N - 1 \approx N$  for large  $N$ .) Thus:

$$S(n) = k \ln \binom{N-1}{n} \approx -kN[x \ln x + (1-x) \ln(1-x)]$$

where  $x = \frac{n}{N}$ .

Thus, the free energy as a function of  $x = \frac{n}{N}$  is:

$$F(x) = E(x) - TS(x) = N \{ Jx + kT[x \ln x + (1-x) \ln(1-x)] \}$$

Minimising with respect to  $x$  yields:

$$\begin{aligned} J + kT[\ln x - \ln(1-x)] &= 0 \\ \Rightarrow \frac{x}{1-x} &= \exp\left(\frac{-J}{kT}\right) \\ \Rightarrow \bar{n} &= N \frac{\exp\left(\frac{-J}{kT}\right)}{1 + \exp\left(\frac{-J}{kT}\right)} \end{aligned}$$

The average number of domains  $n + 1$  tends to  $N/2$  as  $T \rightarrow \infty$  and to 1 as  $T \rightarrow 0$  (in this limit  $n$  is not large so strictly the analysis isn't valid, but the results give some insight).

As usual, there is a competition between energy, here favoring ferromagnetic order, and entropy: at very low  $T$  ( $kT \ll J$ ), energy wins, and we have very large ferromagnetically ordered domains; at high  $T$  ( $kT \gg J$ ), entropy dominates and the domains are very small.

### 7.3 Pause for breath

We have now covered all the fundamental concepts of statistical mechanics. Before proceeding it is worthwhile to try and recap on the basic ideas. You should go through the key points (13 of them) of chapters 3–7. Here we try and pull some ideas together.

- First we met microstates and macrostates. One thing to remember is that for a system of  $N$  ‘particles’ the number of microstates is typically exponentially large in  $N$  e.g.  $2^N$  for the model magnet. Also to each macrostate there corresponds there typically corresponds a large number of microstates, exponentially large in  $N$ .
- For an isolated system all microstates are equally likely and the system evolves to the macrostate with the largest weight. This corresponds to the second law i.e. maximising the entropy.
- For a system in equilibrium with a heat bath the entropy of the composite (system + heat bath) is maximised and this gives the Boltzmann distribution for the system. From the point of view of the system, its free energy is minimised.
- The concept of equilibrium when energy exchange is present gives a statistical mechanics definition of temperature.
- Always, contact is made with thermodynamics through the fact that macroscopic variables are sharply defined for large  $N$  and correspond to thermodynamic functions of state.
- Calculations are most easily done for weakly interacting systems using the partition function technique. This is the usual approach. It implies each particle (of the system of weakly interacting particles) has a single particle Boltzmann distribution.
- Minimisation of the free energy is an alternative way of obtaining macroscopic variables that gives good intuition about energy–entropy competition and can be used for interacting systems (see e.g. chapter 7.2).

We have also seen some equivalences. First, a single particle in a large isolated system takes on a Boltzmann distribution since the rest of the particles act as a heat bath. Second, for a large system in equilibrium with a heat bath the energy fluctuations are vanishingly small therefore the system becomes basically equivalent to a large system in isolation.

## 6. Systems of weakly interacting constituents

### 6.1. Factorisation of the partition function

We have already discussed in chapter 3 the concept of weakly interacting particles (or constituents)—basically, we can write the energy of the system as a sum of contributions from each particle since no energy is stored in any interaction potential. The Boltzmann distribution is quite general and is not restricted to weakly interacting systems; the derivation of chapter 5 only required that the system interact weakly with its environment. However, it is weakly interacting systems that are most easily treated.

For a weakly interacting system, we can write the energy of a microstate labelled  $r$  as:

$$E_r = \epsilon_{i_1} + \epsilon_{i_2} + \epsilon_{i_3} + \cdots + \epsilon_{i_N}$$

where  $\epsilon_{i_n}$  is the energy of particle  $n$  which is in state  $i_n$ .

For example, in the model magnet, the microstate is given by the states of all  $N$  dipoles; the state of each dipole  $n$  is either the ground state ( $i_n = 1; \epsilon_1 = -mH$ ) or the excited state ( $i_n = 2; \epsilon_2 = +mH$ ). The total energy is the sum of the energies of each dipole.

Recalling the definition of  $Z$  from key point 10, we have:

$$\begin{aligned} Z &= \sum_r \exp(-\beta E_r) \\ &= \sum_{i_1 \cdots i_N} \exp(-\beta [\epsilon_{i_1} + \epsilon_{i_2} + \cdots + \epsilon_{i_N}]) \\ &= \left[ \sum_{i_1} \exp(-\beta \epsilon_{i_1}) \right] \cdots \left[ \sum_{i_N} \exp(-\beta \epsilon_{i_N}) \right] \\ &= [Z(1)]^N \end{aligned}$$

where  $Z(1)$  is the partition function for a single particle, c.f. the single dipole calculation of chapter 5.2. You should make sure you follow each step in the above development.

The significance of this is that the partition function for the  $N$  particle system factorises into a product of  $N$  partition functions, one for each particle. This produces an immense simplification, basically because:

$$\ln Z = \ln [Z(1)]^N = N \ln Z(1)$$

and  $Z(1)$  should be simple to calculate.

As an example, consider the mean energy which from chapter 6.1 or key point 12 is given by:

$$\overline{E} = -\frac{\partial}{\partial \beta} \ln Z = -N \frac{\partial}{\partial \beta} \ln Z(1) = N\bar{\epsilon}$$

where  $\bar{\epsilon}$  is the average energy of a single dipole. The fact that the mean energy is the sum of the mean energies of each dipole is, of course, expected in the absence of (strong) interactions between particles.

Similarly,

$$F(T) = -kT \ln Z = -NkT \ln Z(1)$$

Another consequence of the factorisation of  $Z$  is that if we are interested in the state of, say, particle 1, then we can ‘sum out’ the states of all the other particles 2 to  $N$ :

$$\begin{aligned} P_{i_1} &= \sum_{i_2 \dots i_N} \exp(-\beta [\epsilon_{i_1} + \epsilon_{i_2} + \dots + \epsilon_{i_N}]) \times Z^{-1} \\ &= \frac{\exp(-\beta \epsilon_{i_1}) Z(1)^{N-1}}{Z(1)^N} \\ &= \frac{\exp(-\beta \epsilon_{i_1})}{Z(1)} \end{aligned}$$

This is identical to the probability distribution for a single particle discussed in 4.2.

### Key Point 13:

In a system of  $N$  weakly interacting, distinguishable particles, the system partition function is simply  $Z = [Z(1)]^N$  and the single particle probability distribution is  $P_i = \frac{\exp(-\beta \epsilon_i)}{Z(1)}$ .

Thus we see that the problem is reduced to that of the single particle problem. The key thing to remember though is that for the single particle problem the energy, etc., fluctuate, whereas in the  $N$  particle system for large  $N$ , the macroscopic variables are sharp and have well-defined thermodynamic values.

**N.B.** the reason for the qualification ‘distinguishable’ will become apparent later when we address quantum particles.

## 6.2. The model magnet

See Mandl 3.1-3.3

In 5.2, we examined the single particle partition function, i.e., the partition function for a single dipole, so effectively we have already done the hard work. We saw:

$$Z(1) = 2 \cosh(x) \quad \text{where} \quad x = \frac{mH}{kT}$$

Using the general results for weakly interacting systems (previous subsection), we have:

$$\bar{E} = N\bar{\epsilon} = -N \frac{\partial}{\partial \beta} \ln Z(1) = -NmH \frac{\partial}{\partial x} \ln Z(1) = -NmH \tanh(x)$$

$$\begin{aligned} S(T) &= k \ln Z + \frac{\bar{E}}{T} \\ &= Nk \ln Z(1) + \frac{N\bar{\epsilon}}{T} \\ &= Nk [\ln(\exp(x) + \exp(-x)) - x \tanh(x)] \end{aligned}$$

For the magnetisation, we have:

$$\bar{M} = Nm \tanh(x)$$

we see that for low fields, which means  $mH \ll kT$  so that  $x$  is small:

$$\bar{M} = \frac{Nm^2 H}{kT}$$

and for high fields  $mH \gg kT$  so that  $x \gg 1$ :

$$\bar{M} \approx Nm$$

Let us now consider the zero-field magnetic susceptibility  $\chi(H = 0)$  defined as:

$$\chi(H = 0) \equiv \left( \frac{\partial \bar{M}}{\partial H} \right) \Big|_{H=0} = \frac{Nm^2}{kT}$$

You should understand that the susceptibility measures the response of the magnetisation of the system to a small externally applied field. It turns out that this response is actually related

to the magnetisation fluctuations at zero field. This is analogous to the results for the specific heat in terms of the variance of the energy fluctuations.

The  $\frac{1}{T}$  dependence of  $\chi$  is known as the Curie law. It implies that the dipoles become more ‘susceptible’ to an external magnetic field at lower temperature. See Mandl 3.1 for a comparison with real experiments.

Finally, let’s consider the heat capacity (at constant field):

$$C_H = \left( \frac{\partial \bar{E}}{\partial T} \right)_H = \left( \frac{\partial x}{\partial T} \right)_H \left( \frac{\partial \bar{E}}{\partial x} \right)_H = Nkx^2 \operatorname{sech}^2(x)$$

Note the low-temperature (large  $x$ ) behaviour. Recalling:

$$\operatorname{sech}(x) = \frac{2}{\exp(x) + \exp(-x)} \sim 2 \exp(-x) \text{ for large } x$$

we see that the heat capacity vanishes for small  $T$ .

This is a general feature. Basically, at zero temperature, all particles are in the ground state. One has to raise the temperature until  $kT$  is comparable with  $2mH$  (the energy difference to the excited state) before a significant number of dipoles are excited. Thus near  $T = 0$ , the derivative of internal energy with respect to  $T$  is zero.

$kT$  is often referred to as the ‘thermal energy’.

### 6.3. More general factorisation

We have seen in this chapter how the fact that for weakly interacting particles the energy is a sum of terms allows factorisation of  $Z$ .

A similar mechanism occurs in systems which are not necessarily weakly interacting, but whose energy may be expressed as a sum of contributions associated with different ‘aspects’ of its behaviour say  $\alpha, \beta$ .

$$E = E_\alpha + E_\beta + \dots$$

Then if the microscopic ‘degrees of freedom’ which contribute to the different aspects are independent, we find that the partition function factorises:

$$Z = Z_\alpha \times Z_\beta \times \dots$$

A consequence of this is that the free energy is a sum of contributions from each aspect.

For example, in a real magnetic solid, contributions to the energy will come from vibrations of the atoms about their average positions (see chapter 8) as well

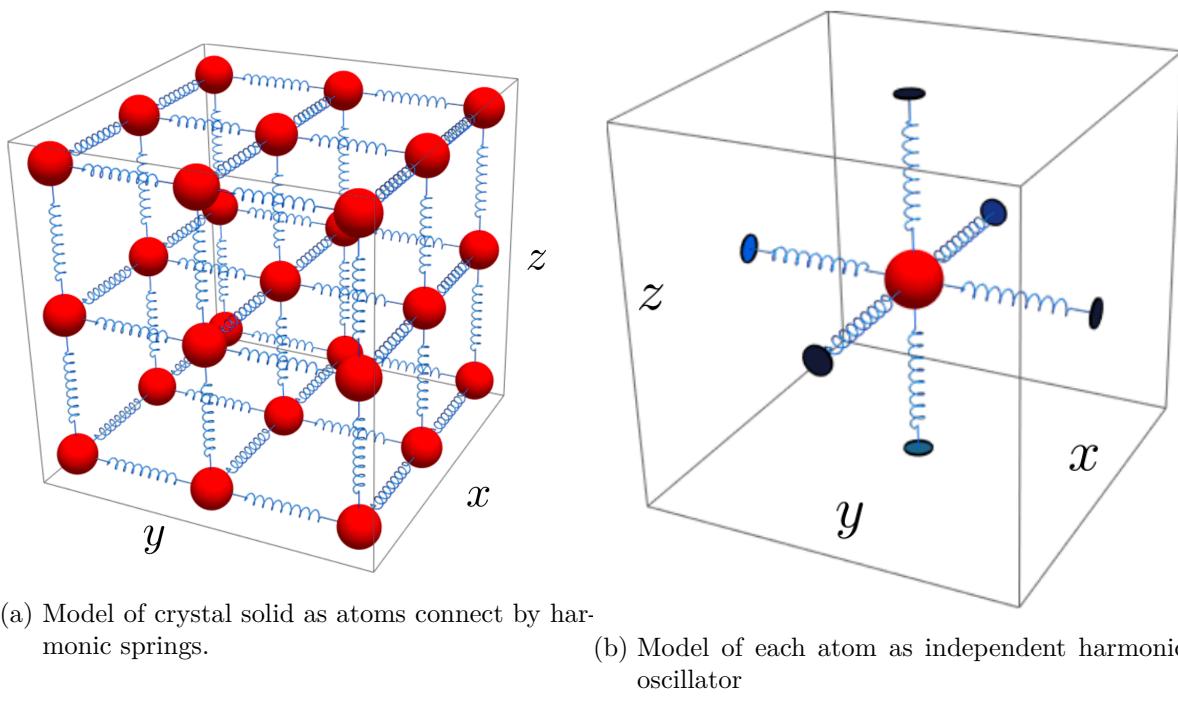
## 8 Einstein's model of a simple solid

Here we apply the Boltzmann distribution to the vibrational energy of atoms in a solid. In so doing we use a fundamental model of quantum physics—the quantum harmonic oscillator.

### 8.1 Simple model of a solid

*See Mandl 6.2*

Recall from your first year Properties of Matter course that in a crystalline solid, atoms sit on the sites of a regular array (lattice). A model for the vibrational motion of the atoms about their equilibrium positions (lattice sites) is to consider the atoms attached to each other by springs (see Figure 1a). However this is a (strongly) interacting system since energy is stored in interaction potentials (the springs) between atoms.



We seek a weakly interacting system that models the behaviour. A simple approximation is illustrated in Figure 1b. Each atom sits in its own harmonic potential and the motion of atoms independent of the others. This is the Einstein model.

Let us first consider the system classically. Each oscillator has energy

$$\varepsilon = \frac{1}{2}\kappa x^2 + \frac{1}{2}mv^2$$

where  $x$  is the displacement from equilibrium position. Thus in 3d the system has  $N \times 3 \times 2$  quadratic contributions ( $N$  particles, 3 dimensions, P.E. + K.E.) to the energy. The equipartition principle then implies that the energy and heat capacity of the system should be

$$\bar{E} = 6N \times \frac{1}{2}kT = 3NkT$$

$$C_V = \frac{\partial \bar{E}}{\partial T} = 3Nk$$

If you have forgotten the equipartition principle you can find it in your first year Properties of Matter notes. However we will understand it at a deeper level in a few chapters' time.

The prediction for the heat capacity was well borne out experimentally for many monoatomic solids and is known as the Dulong–Petit law. However diamond has a smaller heat capacity than the Dulong–Petit law predicts. Einstein showed that this is a quantum effect.

To understand why at a qualitative level quantum effects are important in diamond. Recall that diamond is very hard. This means that the ‘spring constant’ of the classical oscillator modelling each atom in Figure 1 is very large. Therefore the frequency of the oscillators is very large and the typical displacements very small. This last fact implies that quantum effects have to be taken into account—roughly speaking, from the uncertainty principle if the displacements are very small then we know the atoms’ positions with high accuracy, therefore there must be uncertainty in the momenta (velocities).

## 8.2 Statistical mechanics of the quantum harmonic oscillator

From quantum mechanics recall that a one dimensional harmonic oscillator has energy levels

$$\varepsilon = \left(n + \frac{1}{2}\right) \hbar\omega$$

where  $n = 0, 1, 2 \dots$ . The ground state energy  $n = 0$  is  $\frac{1}{2}\hbar\omega$ . The generalisation to a three-dimensional oscillator is easy

$$\varepsilon_{3d} = \left(n_x + n_y + n_z + \frac{3}{2}\right) \hbar\omega$$

Einstein’s model is a system of  $N$  3d quantum oscillators all with the same frequency  $\omega$  in thermal equilibrium;  $\omega$  is chosen to fit the experimental data.

From the statistical mechanics of weakly interacting systems we have  $Z = [Z(1)]^N$ . Now since the energy of an oscillator is a sum of three similar contributions we will have a further factorisation (see chapter 6.3) and

$$Z = [Z(1)]^N = [Z_{1d}(1)]^{3N}$$

where  $Z_{1d}(1)$  is the partition function for a single 1d oscillator. The task reduces to the calculation of  $Z_{1d}(1)$

$$Z_{1d}(1) = \sum_{n=0}^{\infty} \exp\left(-\beta\hbar\omega\left[n + \frac{1}{2}\right]\right)$$

To evaluate the sum recall the geometric series

$$\sum_{n=0}^{\infty} a^n = \frac{1}{1-a}$$

Thus

$$Z_{1d}(1) = \frac{\exp\left(-\frac{x}{2}\right)}{1 - \exp(-x)}$$

where  $x = \beta\hbar\omega$ .

### 8.3 Thermodynamic properties

We now turn the handle to crank out the thermodynamics variables. First note

$$\overline{E} = 3N\overline{\varepsilon} = 3N\hbar\omega\left(\overline{n} + \frac{1}{2}\right)$$

It is easiest to calculate  $\varepsilon$  directly from the relation

$$\begin{aligned} \overline{\varepsilon} &= -\frac{\partial}{\partial\beta} \ln Z_{1d}(1) = -\frac{dx}{d\beta} \frac{\partial}{\partial x} \ln Z_{1d}(1) \\ &= -\hbar\omega \frac{\partial}{\partial x} \left[ -\ln(1 - \exp(-x)) - \frac{x}{2} \right] \end{aligned}$$

$$= \hbar\omega \left[ \frac{\exp(-x)}{1 - \exp(-x)} + \frac{1}{2} \right]$$

and we see

$$\bar{n} = \frac{\exp(-x)}{1 - \exp(-x)} = \frac{1}{\exp(x) - 1}$$

 Expand to see an alternative derivation

Aside: An alternative derivation is to use the Boltzmann distribution  $P_n$  for the levels  $n$  of a 1d oscillator explicitly

$$\bar{n} = \sum_{n=0}^{\infty} P_n n = \frac{1}{Z_{1d}(1)} \sum_{n=0}^{\infty} n \exp \left[ -x \left( n + \frac{1}{2} \right) \right]$$

then use the identity

$$\sum_{n=0}^{\infty} n a^n = a \frac{d}{da} \sum_{n=0}^{\infty} a^n = a \frac{d}{da} \frac{1}{1-a} = \frac{a}{(1-a)^2}$$

The mean total energy is given by

$$\bar{E} = 3N\hbar\omega \left[ \frac{\exp(-x)}{1 - \exp(-x)} + \frac{1}{2} \right]$$

and the heat capacity is given by

$$C_V = \left( \frac{\partial \bar{E}}{\partial T} \right)_V = \left( \frac{\partial x}{\partial T} \right)_\omega \left( \frac{\partial \bar{E}}{\partial x} \right)_\omega$$

Note that here the ‘constant volume’ constraint on the partial derivative is the same as the ‘constant  $\omega$ ’ constraint. This is because the only way volume (or any other variable except temperature) can enter into the model is through the parameter  $\omega$ .

We find

$$C_V = -3N \frac{\hbar\omega}{kT^2} \frac{d}{dx} \left[ \frac{1}{\exp(x) - 1} \right] = 3Nk \frac{x^2 \exp(x)}{(\exp(x) - 1)^2}$$

## 8.4 High and low temperature behaviour

To define the high temperature and low temperature regimes we define a characteristic temperature  $T^*$  defined by  $x = 1$  i.e. when  $kT$  equals the excitation energy  $\hbar\omega$

$$T^* = \frac{\hbar\omega}{k}$$

Thus in the high temperature regime  $T \gg T^*$  ( $x \ll 1$ ) we find

$$\bar{n} \approx \frac{1}{1 + x \dots - 1} \approx \frac{1}{x} = \frac{kT}{\hbar\omega}$$

$$\bar{E} \approx 3NkT + \frac{3}{2}N\hbar\omega$$

we see that  $\bar{n}$  and hence  $\bar{\varepsilon}$  and  $\bar{E}$  increase linearly with temperature. Moreover the heat capacity becomes

$$C_V = \left( \frac{\partial \bar{E}}{\partial T} \right)_V \approx 3Nk$$

and we recover the Dulong–Petit law.

In the low temperature regime  $T \ll T^*$  ( $x \gg 1$ ) we have

$$\bar{n} \approx \exp(-x)$$

$$\frac{C_V}{3Nk} \approx x^2 \exp(-x)$$

The mean energy level  $\bar{n}$  tends to zero meaning most oscillators are in the ground state; the heat capacity is very much less than  $3Nk$ .

Thus when  $x = \frac{\hbar\omega}{kT}$  is large, quantum effects, in particular the effect of a discrete gap between the ground state and first excited state, become important. The Einstein model quite successfully explained the experimentally observed heat capacity of diamond (see Mandl 6.2). However the very low temperature behaviour of the heat capacity was still not quite right. In fact a  $T^3$  dependence appeared to emerge. This can be explained by an elaborated version of the Einstein theory known as the Debye theory.

When  $x = \frac{\hbar\omega}{kT}$  is small we recover the ‘classical’ results i.e. results where Planck’s constant does not appear in the thermodynamic quantities (except as an arbitrary constant in the energy). Quite generally high temperature is the classical limit (if indeed one exists).

## 9. Ideal gas and indistinguishability

*Blundell and Blundell chapter 21*

So far we have applied the Boltzmann theory for weakly interacting systems to solids and their magnetic and vibrational properties. We now turn to the other example of a weakly interacting system mentioned earlier- the ideal gas. In fact we have to modify the theory a little to cope with this system due to the property of indistinguishability.

### 9.1 Indistinguishability

Since atoms of some element or molecules of the same type all have the same chemical make-up they are truly identical.

In solids we can still tell which atom is which because they are localised at the lattice sites of the crystal structure. So we can label the atom nearest to site 1, as atom number 1 and know that basically it will always be the same atom.

However in gases, or liquids, the identity of the particles is problematic. The problem stems from quantum mechanics. In quantum mechanics to correctly describe a many particle system one has to consider a many-particle wavefunction. We will not delve into the issues here, but will just argue that we cannot treat particles individually because we cannot keep track of their identities.

#### 💡 Key point 14

Identical non-localised particles must be treated as indistinguishable.

### 9.2 Semi-classical treatment

We seek to modify in a simple way the Boltzmann theory to take into account indistinguishability.

Roughly speaking, the number of microstates of a system of indistinguishable particles is less by a factor of  $1/N!$  compared to the case where the particles are treated distinguishably. Thus to correct for the overcounting due to indistinguishability, we divide the partition function by  $N!:$

$$Z_{\text{indist.}} = \frac{1}{N!} \sum_{i_1=i_2=\dots=i_N} \exp(-\beta[\epsilon_{i_1} + \epsilon_{i_2} + \dots + \epsilon_{i_N}]) = \frac{1}{N!} [Z(1)]^N$$

where  $i_n$  is the state of particle  $n$ .

### 9.3 Calculation of $Z(1)$

We now have to evaluate  $Z(1)$  the single particle partition function. Here the single particle is simply a free particle in a box (container for the ideal gas) which we take to be a cube of side  $L$ .

The time independent Schrödinger's equation for the free particle ( $V = 0$ ) reduces to the equation for standing waves:

$$\left(-\frac{\hbar^2}{2M}\nabla^2 - \epsilon\right)\psi = 0$$

Consider first the one-dimensional case which becomes:

$$\psi'' = -k^2\psi \quad \text{where} \quad k^2 = \frac{2M\epsilon}{\hbar^2} = 0$$

We have to fit the boundary conditions that  $\psi$  vanishes at the boundaries  $x = 0, L$ . Thus:

$$\psi = A \sin kx \quad \text{with} \quad k = \frac{n\pi}{L} \quad \text{and} \quad n = 1, 2, 3, \dots$$

The generalisation to three dimensions with the boundary conditions that  $\psi$  vanishes at  $x, y, z = 0, L$  is straightforward:

$$\begin{aligned} \psi &= A \sin k_x x \sin k_y y \sin k_z z \quad \text{with} \quad k_x = \frac{n_x \pi}{L}, \quad k_y = \frac{n_y \pi}{L}, \quad k_z = \frac{n_z \pi}{L} \\ \epsilon &= \frac{\hbar^2}{2M}(k_x^2 + k_y^2 + k_z^2) = \frac{\hbar^2 \pi^2}{2ML^2}(n_x^2 + n_y^2 + n_z^2) \end{aligned}$$

The task now is to sum over all allowed values of  $n_x, n_y, n_z$ :

$$Z(1) = \sum_{n_x, n_y, n_z} \exp(-\beta\epsilon(n_x, n_y, n_z)) = \exp(-\beta\epsilon(1, 1, 1)) + \exp(-\beta\epsilon(2, 1, 1)) + \dots$$

In principle this looks quite difficult but the way to simplify matters is to convert the sum into an integral.

Referring to Figure 2 we see that  $n$ -space contains points, which represent the allowed quantum states of the particle, distributed with unit density. Since the energy depends only on the magnitude  $n$  of the position vector of a point in  $n$ -space:

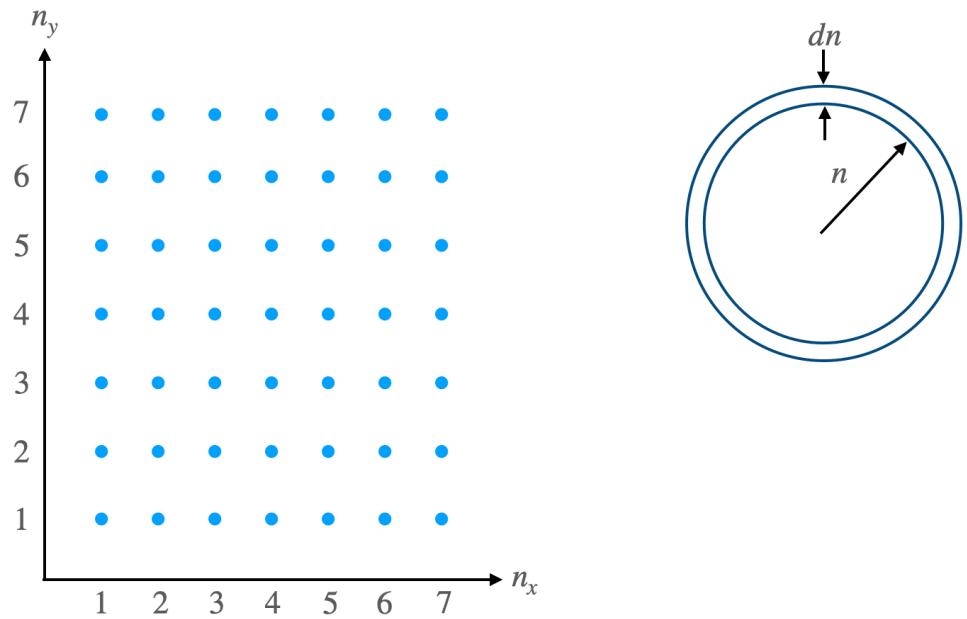


Figure 1: A 2-d representation of ‘n-space’. The density of points is one. An integration shell of radius  $n$  and thickness  $dn$  is indicated.

$$n = (n_x^2 + n_y^2 + n_z^2)^{1/2}$$

we convert to an integral with respect to  $n$ . We think of adding up shells of constant  $n$  whose volume are  $\frac{1}{8}4\pi n^2 dn$  and which, since the density of points is one, contain on average  $\frac{1}{8}4\pi n^2 dn$  points. The volume is of a spherical shell of radius  $n$  and thickness  $dn$ ; the division by 8 is because we are restricted to the octant where all components of  $n$  are positive.

$$\sum_{n_x, n_y, n_z} [\dots] \rightarrow \frac{1}{8} \int [\dots] 4\pi n^2 dn$$

In particular:

$$Z(1) \rightarrow \frac{\pi}{2} \int_0^\infty \exp\left(-\beta \frac{\hbar^2 \pi^2 n^2}{2ML^2}\right) n^2 dn$$

This is of course an approximation. It is a good approximation when:

$$\frac{\hbar^2 \pi^2}{2ML^2} \ll kT$$

## 10. Ideal gas in the low density limit

*Blundell and Blundell chapter 21*

In the last chapter, the calculation of  $Z(1)$  for the ideal gas was begun. In this chapter, we finish the calculation and discuss when the semi-classical treatment is valid. The concept of a density of states is also introduced.

### 10.1 Calculation of $Z(1)$

We first finish the job begun in the last chapter:

$$Z(1) = \pi \int_0^\infty \exp(-an^2) n^2 dn \quad \text{where} \quad a = \frac{\hbar^2 \pi^2}{2kT M L^2}$$

The Gaussian integral that should be familiar is:

$$\int_{-\infty}^\infty \exp(-an^2) dn = \left(\frac{\pi}{a}\right)^{1/2}$$

Thus, we see:

$$Z(1) = -\frac{1}{2} \frac{d}{da} \left( \int_{-\infty}^\infty \exp(-an^2) dn \right) = -\frac{1}{2} \frac{d}{da} \left( \frac{\pi}{a} \right)^{1/2} = \frac{\sqrt{\pi}}{4a^{3/2}}$$

The factor  $\frac{1}{2}$  comes from the limit 0 to  $\infty$  in the integral for  $Z(1)$ . Inserting the expression for  $a$  gives:

$$Z(1) = \frac{\pi^{3/2}}{8} \left( \frac{2ML^2}{\beta\hbar^2\pi^2} \right)^{3/2} = V \left( \frac{2\pi M k T}{h^2} \right)^{3/2}$$

where  $h = 2\pi\hbar$  has been used.

## 10.2 Density of states

Consider again the approximation of a sum by an integral:

$$\sum_{n_x, n_y, n_z} A(n) \rightarrow \frac{1}{8} \int_0^\infty A(n) 4\pi n^2 dn$$

where  $A(n)$  is any function of  $n$ . Now change variables to  $k$  (see 9.3):

$$n = \frac{L}{\pi} k \quad dn = \frac{L}{\pi} dk$$

and we find:

$$\sum_{n_x, n_y, n_z} A(n) \rightarrow \frac{1}{8} \int A(k) \frac{4\pi L^2 k^2}{\pi^2} \frac{L}{\pi} dk = \int A(k) \Gamma(k) dk$$

where:

$$\Gamma(k) dk = \frac{k^2}{2\pi^2} V dk$$

is the number of states with  $k$  between  $k$  and  $k + dk$ .  $\Gamma(k)$  is known as the “density of states” (here, in  $k$  space).

Similarly, we can change variables to  $\epsilon$ :

$$\epsilon = \frac{\hbar^2 k^2}{2M} \quad k = \left( \frac{2M\epsilon}{\hbar^2} \right)^{1/2} \quad dk = \frac{1}{2} \left( \frac{2M}{\hbar^2} \right)^{1/2} \epsilon^{-1/2} d\epsilon$$

Now, to obtain the density of states in energy space, denoted here by  $g(\epsilon)$ , equate:

$$\Gamma(k) dk = g(\epsilon) d\epsilon$$

and change variables:

$$g(\epsilon) d\epsilon = \left( \frac{2M}{\hbar^2} \right)^{3/2} \frac{V}{4\pi^2} \epsilon^{1/2} d\epsilon$$

The meaning of the density of states  $g(\epsilon)$  is that  $g(\epsilon) d\epsilon$  is the number of states with energy between  $\epsilon$  and  $\epsilon + d\epsilon$ .

If you're unsure of this change of variables from  $\Gamma(k)$  to  $g(\epsilon)$ , you can carry it out as follows:  
Consider for any function  $A(k)$ :

$$\int A(k)\Gamma(k)dk = \int A(k)\Gamma(k)\frac{dk}{d\epsilon}d\epsilon = \int A(\epsilon)g(\epsilon)d\epsilon$$

Using the density of states, one can write, for example:

$$Z(1) = \int_0^\infty \exp(-\beta\epsilon)g(\epsilon)d\epsilon$$

Notice that the density of states increases with energy, meaning that at higher energies, more states are available to the particle.

### 10.3 Thermodynamic variables

In the semi-classical treatment developed over this and the previous chapter:

$$Z = \frac{Z(1)^N}{N!} = \frac{V^N}{N!} \left( \frac{2\pi M k T}{h^2} \right)^{3N/2}$$

We now proceed to calculate the usual thermodynamic variables:

$$F = -kT \ln Z = NkT \left[ \ln \left( \frac{N}{V} \right) - 1 - \frac{3}{2} \ln \left( \frac{2\pi M k T}{h^2} \right) \right]$$

where Stirling's approximation for  $\ln N!$  was used.

Energy:

$$E = kT^2 \frac{\partial \ln Z}{\partial T} = \frac{3}{2} NkT$$

Entropy:

$$S = \frac{E - F}{T} = Nk \left[ \ln \left( \frac{V}{N} \left( \frac{2\pi M k T}{h^2} \right)^{3/2} \right) + \frac{5}{2} \right]$$

Pressure:

$$P = - \left( \frac{\partial F}{\partial V} \right)_T = \frac{NkT}{V} \quad (\text{ideal gas law})$$

Heat capacity:

$$C_V = \frac{3}{2}Nk \quad (\text{equipartition})$$

Note that

- the ideal gas law and equipartition of energy are recovered.
- The recovery of the ideal gas law finally identifies our statistical mechanical definition of temperature (see key point 9) with the thermodynamic temperature.

## 10.4 Validity of semi-classical treatment

The semi-classical treatment developed here is valid when the particles are partly distinguishable, i.e., when  $d_{typ}$  the typical distance between particles is much larger than the typical thermal de Broglie wavelength:

$$d_{typ} \gg \lambda_{typ}$$

For an ideal gas, this is equivalent to requiring the number density  $n = N/V$  be low, specifically:

$$n\lambda_{typ}^3 \ll 1$$

Using  $\lambda_{typ} = \frac{h}{\sqrt{3MkT}}$ , this condition becomes:

$$n \left( \frac{h}{\sqrt{3MkT}} \right)^3 \ll 1 \quad \Rightarrow \quad n \left( \frac{h^2}{2\pi M k T} \right)^{3/2} \ll 1$$

This inequality is often expressed as  $n\lambda_{typ}^3 \ll 1$ , known as the **classical limit** or **dilute gas limit**. When this condition is violated, quantum effects become important, and we must turn to quantum statistics (Bose-Einstein or Fermi-Dirac statistics).

The semi-classical treatment breaks down when the temperature is very low or the density is very high because, under these conditions, the wavefunctions of individual particles overlap, and quantum indistinguishability becomes essential.

## 10.5 Summary

In this chapter we completed the calculation of  $Z(1)$  for an ideal gas and introduced the concept of a density of states. We discussed the validity of the semi-classical treatment, noting that it is applicable in the low-density limit where the wave packets of individual particles do not overlap. When the typical particle separation becomes comparable to or smaller than the de Broglie wavelength, quantum effects must be considered, and the semi-classical treatment breaks down.

Next we will explore quantum statistics and examine how the behavior of ideal gases changes in the high-density regime where quantum effects become significant.

# 11. Systems with varying particle number

*See Mandl 8.1*

So far we have developed the Boltzmann distribution where the number of particles  $N$  is fixed and the energy is a free macroscopic variable. In this chapter, we consider systems where the particle number is also free to fluctuate.

## 11.1 Motivations

There are two reasons for allowing the particle number to fluctuate:

1. We may wish to consider systems free to exchange particles. An important example is phase co-existence whereby, e.g., a liquid and its vapor are in equilibrium, and a molecule may either be part of the liquid phase or of the gas phase.
2. The second motivation is that since the particle number is a macroscopic variable, for a large system it is sharply defined at some mean value. Thus, a large system with varying particle number is expected to have the same behavior as a system of fixed particle number.

We then use the system with varying particle number as ‘a means to an end,’ the end being the study of quantum gases for which the calculations turn out to be easier when the particle number is not fixed.

## 11.2 The chemical potential

Consider first a system as in Figure 1, where two halves (at equilibrium at the same temperature) are free to exchange particles. Since the total number  $N$  of particles is conserved:

$$dN_1 = -dN_2$$

As the free energy is extensive, it can be written as a sum of contributions from each half of the system:

$$F = F_1(N_1) + F_2(N_2)$$

$$dF = \frac{\partial F_1(N_1)}{\partial N_1} dN_1 + \frac{\partial F_2(N_2)}{\partial N_2} dN_2 = \left[ \frac{\partial F_1(N_1)}{\partial N_1} - \frac{\partial F_2(N_2)}{\partial N_2} \right] dN_1$$

At equilibrium, the free energy should be minimized, and  $dF = 0$ , therefore we must have:

$$\frac{\partial F_1(N_1)}{\partial N_1} = \frac{\partial F_2(N_2)}{\partial N_2}$$

?

### Definition

$$\mu \equiv \left( \frac{\partial F}{\partial N} \right)_{T,V}$$

is the chemical potential which is a quantity which is common to two systems that can exchange particles.

A statement that is often made is: “the chemical potential is like a temperature for particle number”. If a system is not in equilibrium so that there is a chemical potential gradient, then particles will diffuse down then gradient. This is similar to heat diffusing down a temperature gradient.

### 11.3 Example

A specific example may be helpful. Consider an ideal gas confined to two connected volumes (see Figure 3). In the semi-classical treatment, the partition function factorizes into the contribution from the upper volume  $Z_u$  and from the lower volume  $Z_l$  :

$$Z(N_l, N_u) = Z_l(N_l) \times Z_u(N_u) = \left( \frac{V}{\lambda^3} \right)^{N_l} \frac{1}{N_l!} \times \left( \frac{V}{\lambda^3} \right)^{N_u} e^{-\beta mgHN_u} \frac{1}{N_u!}$$

where  $\lambda = \frac{h}{\sqrt{2\pi MkT}}$ . Using semi-classical results from 10.3:

$$F_l = -kT \ln Z_l = N_l kT \left[ \ln \left( \frac{N_l \lambda^3}{V} \right) - 1 \right]$$

$$F_u = N_u kT \left[ \ln \left( \frac{N_u \lambda^3}{V} \right) - 1 \right] + N_u MgH$$

$$\mu_l = \frac{\partial F_l}{\partial N_l} = kT \ln \left( \frac{N_l}{V \lambda^3} \right)$$

$$\mu_u = \frac{\partial F_u}{\partial N_u} = kT \ln \left( \frac{N_u}{V \lambda^3} \right) + mgH$$

Equating the chemical potentials  $\mu_l = \mu_u$  yields:

$$kT \ln \left( \frac{N_l}{N_u} \right) = MgH \Rightarrow N_u = N_l e^{-MgH/kT}$$

This gives the density dependence on height in the isothermal atmosphere (tutorial 8.1). Note that  $\mu$  increases with density, and  $\mu_u$  increases with  $H$ , which supports the claim that the chemical potential measures the tendency for particles to diffuse.

 Expand to read about an equivalent definition of  $\mu$

Consider the scenario of Figure 4, where a system isolated from the rest of the universe comprises two subsystems free to exchange energy and particles. Since the total system is isolated, we must maximize the entropy (see chapter 4, Tutorial 3.1):

$$S = S_1(E_1, N_1) + S_2(E_2, N_2)$$

$$dS = \left[ \left( \frac{\partial S_1}{\partial E_1} \right) - \left( \frac{\partial S_2}{\partial E_2} \right) \right] dE_1 + \left[ \left( \frac{\partial S_1}{\partial N_1} \right) - \left( \frac{\partial S_2}{\partial N_2} \right) \right] dN_1$$

Since  $dE_1$  and  $dN_1$  are independent, to have  $dS = 0$ , we must have common values of:

$$\left( \frac{\partial S}{\partial E} \right)_{N,V} = \frac{1}{T}, \quad \left( \frac{\partial S}{\partial N} \right)_{E,V} = -\frac{\mu}{T}$$

where we have used the fact that for an ideal gas  $E = 0$  so from  $F = E - TS$ ,  $S = -F/T$ . Thus, we identify:

$$\mu = -T \left( \frac{\partial S}{\partial N} \right)_{E,V}$$

This definition is equivalent to the previous boxed definition of  $\mu$ .

## 11.4 Grand canonical distribution

Consider the scenario of Figure 5, where, in analogy with chapter 4, the ‘composite’ comprises the system and the bath, which this time serves as a reservoir of energy and particles at constant  $T$  and  $\mu$ .

A microstate  $r$  of the system has energy  $E$  and  $N$  particles. In analogy with chapter 4, using the principle of equal a priori probabilities, the probability of microstate  $r$  obeys:

$$P_r \propto \Omega_b(E_{\text{tot}} - E, N_{\text{tot}} - N) = \exp \left( \frac{S_b(E_{\text{tot}} - E, N_{\text{tot}} - N)}{k} \right)$$

where the Planck relation has been used. We proceed as in chapter 4 and Taylor expand the entropy (since  $N, E \ll N_{\text{tot}}, E_{\text{tot}}$ ), keeping only the first few terms:

$$\begin{aligned} S_b(E_{\text{tot}} - E, N_{\text{tot}} - N) &= S_b(E_{\text{tot}}, N_{\text{tot}}) - E \frac{\partial S_b(E_{\text{tot}}, N_{\text{tot}})}{\partial E} - N \frac{\partial S_b(E_{\text{tot}}, N_{\text{tot}})}{\partial N} + \dots \\ &= \text{const} - \frac{E}{T} + \frac{N\mu}{T} \end{aligned}$$

and

$$P_r \propto \exp\left(\frac{1}{kT}(N\mu - E)\right)$$

We have now derived the ‘grand canonical distribution’ or Gibbs-Boltzmann distribution (as opposed to the canonical distribution, which is the Boltzmann distribution). The probability that a system in equilibrium with a reservoir of energy and particles at temperature  $T$  and chemical potential  $\mu$  is in microstate  $r$ , which has energy  $E_r$  and particle number  $N_r$ , is:

 Key Point 15:

$$P_r = \frac{1}{Z} \exp(-\beta E_r + \beta \mu N_r), \quad Z = \sum_j \exp(-\beta E_j + \beta \mu N_j), \quad \beta = \frac{1}{kT}$$

## 12. Quantum gases

In the previous chapter we derived the grand canonical distribution. It applies to a system in equilibrium with a reservoir of energy and particles. This kind of system is referred to as an open system. The grand canonical distribution is given by:

$$Pr = \frac{1}{Z} \exp(-\beta E_r + \beta \mu N_r)$$

where

$$Z = \sum_r \exp(-\beta E_r + \beta \mu N_r)$$

Here, we use  $r$  to label the microstates to avoid a later clash of notation. Microstate  $r$  contains  $N_r$  particles and has energy  $E_r$ .

### 12.1. $N$ as a function of $\mu$

In the grand canonical distribution, microstates of the system with all numbers of particles are possible. However, we expect that for large  $N$ , the distribution of the particle number will become sharp.

The mean number of particles is calculated as follows:

$$N = \sum_r N_r P_r = \frac{1}{Z} \sum_r N_r \exp(\beta [N_r \mu - E_r])$$

$$N = \frac{1}{\beta} \frac{\partial \ln Z}{\partial \mu}$$

where a similar idea to chapter 6 — where the mean energy was written as a derivative with respect to  $\beta$  — has been used.

By a similar argument to Q1.5 we find

$$\frac{(\Delta N^2)^{1/2}}{N} \sim \frac{1}{N^{1/2}}$$

Therefore, we see that  $N$  is sharp about  $\bar{N}$  (i.e., fluctuations are small on the scale of the mean).

Thus, we

 Key Point 16:

choose  $\mu$  to fix  $N = N(\mu)$

in the same way that in the canonical distribution for a large system  $T$  fixes  $E = E(T)$ .

Remember also that  $\mu$  depends on temperature.

## 12.2. Indistinguishable particles

Recall the definition of a microstate (chapter 2):

- **Distinguishable particles:** A microstate is specified by  $i_1, i_2, \dots, i_N$ , i.e., the state of each particle.
- **Indistinguishable particles:** A microstate is specified by  $n_1, n_2, \dots$ , where  $n_i$  is the occupation number, i.e., the number of particles in single-particle state  $i$ .

Thus, for indistinguishable particles in microstate  $r$  specified by the set  $\{n_i\}$ :

$$N_r = \sum_i n_i \quad \text{and} \quad E_r = \sum_i n_i \epsilon_i$$

where  $\epsilon_i$  is the energy of single-particle state  $i$ .

Note: In the following, the term ‘quantum state’ will be used interchangeably with ‘single-particle state.’

We can write:

$$\beta(N_r \mu - E_r) = \beta \sum_i n_i (\mu - \epsilon_i)$$

and the sum over all possible microstates  $r$  becomes:

$$\sum_r \rightarrow \sum_{n_1} \sum_{n_2} \dots$$

i.e., a sum over all possible occupation numbers  $n_i$  of all quantum states.

It should be noted that in the canonical distribution, where  $N$  is fixed, a sum over all  $n_i$  is non-trivial because, for example, the value of occupation number  $n_1$  affects what values the other occupation numbers are allowed to take. However, in the grand canonical distribution, the sums ‘decouple,’ and the problem of the constraint on  $N$  is replaced by the problem of choosing  $\mu$  to give the desired  $N(\mu)$ .

Now consider the grand canonical partition function:

$$\begin{aligned}
Z &= \sum_r \exp(\beta [N_r \mu - E_r]) \\
&= \left[ \sum_{n_1} \sum_{n_2} \dots \right] \exp \left( \beta \sum_i n_i (\mu - \epsilon_i) \right) \\
&= \left[ \sum_{n_1} \exp(\beta n_1 (\mu - \epsilon_1)) \right] \times \left[ \sum_{n_2} \exp(\beta n_2 (\mu - \epsilon_2)) \right] \times \dots \\
&= Z_1 \times Z_2 \times \dots = \prod_i Z_i
\end{aligned}$$

where  $Z_i$  is the partition function for quantum state  $i$ .

Thus, a factorization into single-state partition functions occurs. This should be contrasted with the factorization into single-particle partition functions that occurred in the canonical (Boltzmann) distribution.

Using the Factorization of  $Z$  consider the probability of a microstate  $r = \{n_1, n_2, n_3, \dots\}$ :

$$\begin{aligned}
P_r = P(n_1, n_2, \dots) &= \frac{\exp(\beta n_1 (\mu - \epsilon_1))}{Z_1} \times \frac{\exp(\beta n_2 (\mu - \epsilon_2))}{Z_2} \times \dots \\
&= P(n_1) P(n_2) \dots
\end{aligned}$$

where the single state distribution is:

$$P(n_i) = \frac{\exp(\beta n_i (\mu - \epsilon_i))}{Z_i}$$

To understand this result, one can think of the quantum state  $i$  as being free to exchange particles with the rest of the quantum states, which therefore act as a reservoir of particles and energy. Thus, the quantum state  $i$  is itself an open system.

### 12.3. Fermions and Bosons

In quantum mechanics, you have met the concepts of:

- Spin
- Fermions and bosons

Fermions have spin equal to a half-integral multiple of  $\hbar$ , e.g., the magnetic dipoles we considered in the model magnet have spin  $s = 1/2$ , and therefore, there are  $2s + 1 = 2$  spin states (up or down for the dipole).

Examples of fermions are electrons, neutrons, protons, and composite particles made of an odd number of fermions, e.g.,  ${}^3\text{He}$ , whose nucleus contains two protons and a neutron and is therefore a fermion; also, the whole atom is a fermion.

Bosons have spin equal to an integral multiple of  $\hbar$  (note that they can have spin zero). Examples are photons and composite particles made up of an even number of fermions, e.g.,  ${}^4\text{He}$ .

The most important thing for our purposes is the **Pauli exclusion principle**:

**There can be at most one fermion in any quantum state.**

In quantum mechanics, you will see how this comes from the antisymmetry of the many-particle wavefunction for fermions, but the boxed fact is all you need to know here.

Now, consider the single state partition function. Due to the exclusion principle for fermions, an occupation number  $n_i$  can only take the values 0 or 1. Therefore, for fermions:

$$Z_i = \sum_{n_i=0,1} \exp(\beta n_i(\mu - \epsilon_i)) = 1 + \exp(\beta(\mu - \epsilon_i))$$

On the other hand, there is no such restriction for bosons, for which  $n_i$  can take all values from 0 to  $\infty$ .

For bosons:

$$Z_i = \sum_{n_i=0}^{\infty} \exp(\beta n_i(\mu - \epsilon_i)) = \frac{1}{1 - \exp(\beta(\mu - \epsilon_i))}$$

for  $\exp(\beta(\mu - \epsilon_i)) < 1$ .

To understand this result for bosons, recall that:

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \quad \text{for } |x| < 1$$

We can now calculate  $n_i$ , the average number of particles in quantum state  $i$ . To do the calculation at the same time for fermions and bosons, write:

$$Z_i = [1 \pm \exp(\beta(\mu - \epsilon_i))]^{\pm 1}$$

where + refers to bosons, and - refers to fermions.

Now:

$$\bar{n}_i = \sum_{n_i} n_i P(n_i) = \frac{1}{\beta} \frac{\partial \ln Z_i}{\partial \mu}.$$

Thus:

$$\begin{aligned} \bar{n}_i &= \pm \frac{1}{\beta} \frac{\partial}{\partial \mu} \ln [1 \pm \exp(\beta(\mu - \epsilon_i))] \\ &= \pm \frac{1}{\beta} (\pm \beta) \frac{\exp(\beta(\mu - \epsilon_i))}{1 \pm \exp(\beta(\mu - \epsilon_i))} \\ &= \frac{\exp(\beta(\mu - \epsilon_i))}{1 \pm \exp(\beta(\mu - \epsilon_i))} \\ &= \frac{1}{\exp(\beta(\epsilon_i - \mu)) \pm 1} \end{aligned}$$

The final results for the mean number of particles in quantum state  $i$  (that has energy  $\epsilon_i$ ) are known as the **Fermi-Dirac** and **Bose-Einstein** distributions:

 Key Point 17:

$$\bar{n}_i = f_{\pm}(\epsilon_i) = \frac{1}{\exp(\beta(\epsilon_i - \mu)) \pm 1}$$

where + refers to fermions, and - refers to bosons.

$N(\mu)$  (or  $\mu(N)$ ) is determined by the equation:

$$N=\sum_i \overline{n_i}=\sum_i f_\pm(\epsilon_i)$$

## 13. The Ideal Fermi gas

*Blundell and Blundell chapter 30*

### Low density limit

Consider the limit  $e^{\mu/kT} \ll 1$ , i.e.,  $\mu$  large and negative. Then:

$$\overline{n}_i = \frac{1}{\exp(\frac{\epsilon_i - \mu}{kT}) \pm 1} \approx e^{-\frac{\epsilon_i - \mu}{kT}}$$

for both Fermi-Dirac (F-D) and Bose-Einstein (B-E) distributions.

Now fix  $\mu(N, T)$  through the constraint

$$N = \sum_i \overline{n}_i \approx e^{\mu/kT} \sum_i e^{-\epsilon_i/kT} = e^{\mu/kT} Z(1)$$

where we have identified the canonical single particle partition function  $Z(1)$ . Thus we see that in the limit  $e^{\mu/kT} \ll 1$ :

$$Z(1)N \ll 1$$

which, referring to **chapters 11/12**, is the low density/high temperature limit where a semi-classical treatment is adequate.

### Ideal Fermi gas

We return now to the Fermi-Dirac distribution from key point 17 (chapter 12). The function  $f_+$  is often referred to simply as the **Fermi function**.

Consider the limit  $T \rightarrow 0$  ( $\beta \rightarrow \infty$ ):

$$f_+(\epsilon) = \frac{1}{\exp(\frac{\epsilon - \mu}{kT}) + 1} \rightarrow \begin{cases} 1 & \text{if } \epsilon < \epsilon_f \\ 0 & \text{if } \epsilon > \epsilon_f \end{cases}$$

where  $\epsilon_f$  is the **Fermi energy** defined by:

 Key Point 18

$$\epsilon_f = \lim_{T \rightarrow 0} \mu(T)$$

You should convince yourself that the Fermi-Dirac distribution is equivalent to the probability that a quantum state of energy  $\epsilon$  is occupied.

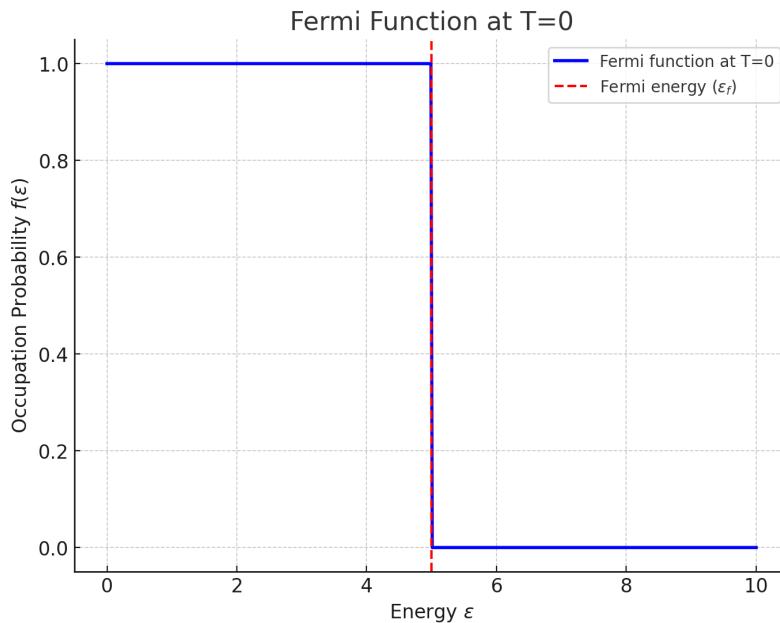


Figure 1: Fermi function at zero temperature

In Figure 1, we observe the following features:

- All states up to  $\epsilon_f$  are filled with probability 1.
- All states above  $\epsilon_f$  are empty.
- This is very different from a classical gas where at zero temperature, all gas molecules would have zero energy.
- It is a direct result of the **exclusion principle**, which leads to an “effective repulsion” between fermions.

We now calculate  $\epsilon_f$ . First, turn the sum over quantum states into an integral:

$$N = \sum_i f_+(\epsilon_i) \approx \int_0^\infty d\epsilon g(\epsilon) f_+(\epsilon)$$

where  $g(\epsilon)$  is the density of states, a concept first introduced in **chapter 10**. As a reminder:

$$g(\epsilon)d\epsilon = \text{number of states in energy range } [\epsilon, \epsilon + d\epsilon]$$

We saw in **chapter 11** that for spinless particles in a box:

$$g(\epsilon) = D\epsilon^{1/2}, \quad D = \left(\frac{2m}{h^2}\right)^{3/2} \frac{1}{4\pi^2}$$

Incorporating the idea of spin, each translational (“standing wave”) state corresponds to  $2s+1$  states since the particle has  $2s+1$  possible spin states. Therefore:

$$g(\epsilon) = \tilde{D}\epsilon^{1/2}, \quad \tilde{D} = (2s+1)D$$

Thus:

$$N = \int_0^{\epsilon_f} \tilde{D}\epsilon^{1/2}d\epsilon = \frac{2}{3}\tilde{D}\epsilon_f^{3/2}$$

which implies:

$$\epsilon_f = \left(\frac{3N}{2\tilde{D}V}\right)^{2/3}$$

or equivalently:

$$\epsilon_f = \frac{h^2}{2m} \left(\frac{6\pi^2 N}{(2s+1)V}\right)^{2/3}$$

In **Question 4.5**, it is shown that:

$$E = \int_0^{\epsilon_f} g(\epsilon)\epsilon d\epsilon = \frac{3}{5}N\epsilon_f$$

### Important Points:

- $\epsilon_f$  decreases with the mass  $M$  of the fermion.
- $\epsilon_f$  increases with the number density  $N/V$ .
- $\epsilon_f$  defines a characteristic temperature through  $\epsilon_f = kT_f$ .
- At zero temperature, there is a finite energy per particle  $\epsilon = \frac{3}{5}\epsilon_f$ .

## Low temperature behaviour

Now consider the Fermi function at low but finite  $T$ . The meaning of “low” will be specified shortly. Note that:

$$f_+(\epsilon) = \frac{1}{\exp(\frac{\epsilon-\mu}{kT}) + 1}$$

approaches 1 if  $(\epsilon - \mu)/kT \ll -1$  and 0 if  $(\epsilon - \mu)/kT \gg 1$ , with  $f_+(\epsilon) = 1/2$  when  $\epsilon = \mu$ . The Fermi function is a sigmoid shape illustrated in Figure 2. It differs from the zero-temperature step-function only when  $|\epsilon - \mu| \sim O(kT)$ .

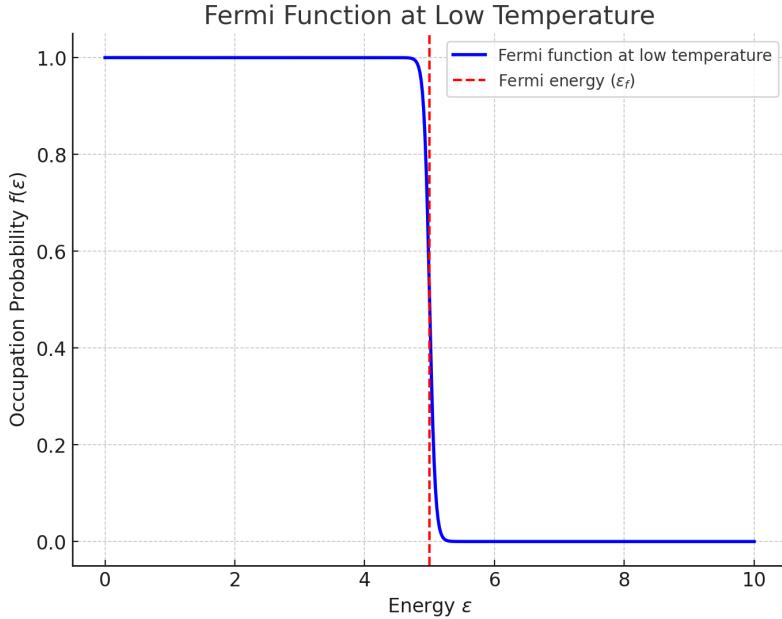


Figure 2: Fermi function at zero temperature

For the Fermi function to retain its characteristic shape, we must have  $kT \ll \mu(T) \approx \epsilon_f$ , implying  $T \ll T_f$ . We assume that for low  $T$ :

$$\mu(T) \approx \epsilon_f \left( 1 - O\left(\frac{T^2}{T_f^2}\right) \right)$$

Thus, the change in  $\mu$  is second-order in  $T/T_f$  and can safely be ignored in the regime  $T \ll T_f$ .

An intuitive interpretation of Figure 2 is that at low temperatures, the general scenario is similar to that at zero temperature, where all states up to energy  $\epsilon_f$  are filled. The difference

is that states within energy  $O(kT)$  below  $\epsilon_f$  are vacated with some probability, and previously empty states within  $O(kT)$  above  $\epsilon_f$  are filled with some probability. In other words, some fermions are thermally excited above the Fermi energy.

We investigate the result of this thermal excitation by calculating the heat capacity. To avoid a complicated calculation required to get the exact result, we instead make a rough estimate (for a more careful argument, see Baierlein 9.1). We expect:

$$E(T) - E(0) \sim N \cdot \frac{kT}{\epsilon_f} \cdot kT$$

i.e., the change in energy is  $N$  times the fraction of fermions excited (roughly  $kT/\epsilon_f$ ) times the typical excitation (roughly  $kT$ ).

Therefore, the heat capacity is approximately:

$$C_V \sim \frac{E(T) - E(0)}{T} \sim \frac{Nk^2T}{\epsilon_f}$$

The important point is that this is **linear in  $T$** . This contrasts with the classical gas, where  $C_V$  is a constant (equal to  $3Nk/2$ ).

### Summary of Ideal Fermi Gas

1.  $E/N \approx E_{\text{classical}}/N$ , i.e., the gas has a large kinetic energy per molecule, even at low  $T$ .
2.  $C_V$  is small.