

4

Statistical mechanics of complex systems

Ellák Somfai

**To appear in: Complexity Science - The Warwick Master's Course,
Chapter 4, pp. 210-245 (Cambridge University Press, Cambridge, 2013)
ISBN: 9781107640566**

Introduction

In this chapter we introduce statistical mechanics in a very general form, and explore how the tools of statistical mechanics can be used to describe complex systems.

To illustrate what *statistical mechanics* is, let us consider a physical system made of a number of interacting particles. When it is just a single particle in given potential, it is an easy problem: one can write down the solution (even if one could not calculate everything in closed form). Having 2 particles is equally easy, as this so-called “two-body problem” can be reduced to two modified one-body problems (one for the centre of mass, other for the relative position). However, a dramatic change occurs when the number of particles is increased to 3. The study of the three-body problem started with Newton, Lagrange, Laplace and many others, but the general form of the solution is still unknown. Even relatively recently, in 1993 a new type of periodic solution has been found, where 3 equal mass particles interacting gravitationally chase each other in a figure-8 shaped orbit. This and other systems where the degrees of freedom is low belongs to the subject of dynamical systems, and is discussed in detail in Chapter 2 of this volume. When the number of interacting particles increases to very large numbers, like 10^{23} , which is typical for the number of atoms in a macroscopic object, surprisingly it gets simpler again, as long as we are interested only at aggregate quantities. This is the subject of statistical mechanics.

Statistical mechanics is also the microscopic foundation of thermodynamics. It developed a number of powerful tools, which can be used outside of the conventional physics domain, like biology, finance (see

Chapter 6), traffic, and more. It can also be considered as *the science of ignorance*: how to handle a system where we do not know (even prefer not to know) everything. A particular example is renormalisation theory, developed in the second half of the 20th century, which gives a systematic framework to dispense successively with the non-interesting degrees of freedom.

Our approach is that of a physicist: (i) based on models, (ii) fundamentally quantitative, (iii) but not rigorous. Probably the most important concept in science is that it is possible to construct an abstraction (“model”) of the world around us which is admittedly much simpler than the real thing, but nevertheless captures important characteristics, and which enables to make predictions that are not “explicitly put in” into the model. The models are fundamentally quantitative, and we use mathematical language to describe them. We stop here however, and leave rigorous treatment to mathematicians — this approach enables one to progress much quicker, on the expense of losing that what we do is absolutely unshakable.

We start with elements of information theory, which in turn are used to derive the foundations of statistical mechanics based on the maximum entropy principle. Unlike the typical treatment in Physics textbooks, this approach has the advantage that the abstract formalism developed can be used in a more straightforward way for systems outside the typical range of thermal applications. We will follow by considering the effects of fluctuations to provide a link to thermodynamics. One of the most characteristic collective phenomena of complex systems is phase transition, which we approach from the direction of statistical physics. The rest of the chapter will deal with dynamics in some form: we will consider interface growth and collective biological motion like flocking.

4.1 Introduction to information theory

Random variables

While probability theory can (and should) be founded rigorously, in these notes we take a relaxed approach and attempt to define everything without mentioning the probability space. We call a *random variable* an object which can take values when observed, say random variable X can take any values from x_1, x_2, \dots, x_n . When observed many times, x_1 is taken n_1 times, x_2 is taken n_2 times, etc. The *probabilities* of the

outcomes can be defined as relative frequencies in the limit of a large number of observations, for example:

$$\frac{n_1}{\sum n_i} \rightarrow p_1.$$

It immediately follows that the probabilities add up to 1:

$$\sum_{i=1}^n p_i = 1.$$

We then say the the probability of X taking a given value x_i is p_i :

$$P(X = x_i) = p_i.$$

A *function of a random variable* is another random variable: if X takes x_i with probability p_i , then $f(X)$ takes the value $f(x_i)$ with the same probability p_i .

The *expectation* or *average* of a random variable can be considered as observing it many times and taking the average of the observed values. In the statistical mechanics literature the standard notation is angular brackets:

$$\langle X \rangle = \sum_i x_i p_i, \quad \text{or} \quad \langle f(X) \rangle = \sum_i f(x_i) p_i.$$

The above concepts can be easily extended to random variables which can take infinitely many discrete values, and even to ones which take values from a continuum. In the latter case eg. if X can take real values, the probability that X takes any value in $[x, x + dx]$ is $p(x)dx$, where dx is small, and $p(x)$ is called the probability density function. The sums above are replaced with integrals, eg. $\int p(x)dx = 1$. While this naive approach to continuous random variables is sufficient for these notes, in general, especially when dealing with continuous random variables, one needs a rigorous foundation of probability theory.

The above *frequentist* approach to probabilities is not the only one. In a sentence like “Tomorrow we will have a 10% chance for rain”, probabilities are interpreted as a degree of belief or confidence.

The information entropy

Suppose we want to describe the outcome of a sequence of coin tosses (heads or tails): HTHHTHTTTH... This sequence looks very different from that of a lottery play: LLLLL...LLLL...LLWLLLL... The second sequence is much more boring, one can describe it as eg. “trial no. 857,923

was a win, all others were lose”. In the first sequence, however, we cannot get away much better than quoting the whole sequence verbatim.

To quantify this difference, we introduce the *information entropy* H of a random variable. Its intuitive meaning is the amount of uncertainty in an observation of a random variable, or in other words the amount of information we gain when observing a random variable. One can think about it as the “amount of answers” needed on average to learn the outcome of an observation as a response to an optimally crafted question-tree. It is a function of the probabilities only: $H(p_1, p_2, \dots)$.

We require certain regularity properties:

- (i) continuity: $H(p_1, p_2, \dots)$ is a continuous function of its arguments.
- (ii) “sense of direction”: of the random variables that take all outcomes with equal probability, the ones with more outcomes carry more information: the function

$$h(n) := H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) \quad (4.1)$$

is monotonically increasing with n .

- (iii) “consistency”: if H is calculated in two different ways, they should agree. Eg. to calculate the information entropy of a 3-state random variable, we can group the last two states and first obtain the information entropy for the obtained 2-state random variable, and then with probability $p_2 + p_3$ need to resolve the grouped states:

$$H_3(p_1, p_2, p_3) = H_2(p_1, q) + qH_2\left(\frac{p_2}{q}, \frac{p_3}{q}\right)$$

where $q = p_2 + p_3$.

It can be shown that this three requirements restrict the functional form of $H(\cdot)$, see [1]:

$$H(p_1, p_2, \dots, p_r) = -K \sum_{i=1}^r p_i \ln(p_i) \quad (4.2)$$

This is the information entropy of a random variable with probabilities p_1, p_2, \dots, p_r . The constant K sets the units, which can be fused into the logarithm as setting its base. In many of the following formulae we will use the notation

$$H(p_1, p_2, \dots, p_r) = - \sum_{i=1}^r p_i \log(p_i) \quad (4.3)$$

without explicitly specifying the base of the logarithm. Then setting $K = 1/\ln(2)$ in (4.2), or equivalently using \log_2 in (4.3) the information entropy is measured in *bits*. When setting $K = 1$ or using \log_e the units are *nats*, and finally the decimal case is $K = 1/\ln(10)$ or using \log_{10} , when the units are called *bans*.

Multiple random variables

When X and Y are random variables, we can look at the probabilities of (X, Y) pairs. These are called *joint probabilities*:

$$p_{ij} = P(X = x_i, Y = y_j)$$

The probability of one of the random variables (the *marginal probabilities*) are obtained by summing up the joint probabilities on all states of the other random variable:

$$p_i^{(X)} = P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij}$$

and similarly $p_j^{(Y)} = \sum_i p_{ij}$.

Two random variables are called *independent*, if the joint probabilities factorise into marginals for all (i, j) pairs:

$$\text{if } P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \text{for all } i, j$$

or equivalently

$$\text{if } p_{ij} = p_i^{(X)} p_j^{(Y)} \quad \text{for all } i, j$$

The *conditional probabilities* tell the probability of one random variable when we know the value of another:

$$p_{i|j} := P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_j^{(Y)}}$$

The *joint information entropy* is the uncertainty of the (X, Y) pair:

$$H(X, Y) = - \sum_{i,j} p_{ij} \log p_{ij}$$

The *conditional information entropy* gives the uncertainty of X when

Y is known:

$$\begin{aligned} H(X | Y) &:= \langle H(X | Y = y_j) \rangle_Y = \sum_j p_j^{(Y)} (-1) \sum_i p_{i|j} \log p_{i|j} \\ &= - \sum_{i,j} p_{ij} \log \frac{p_{ij}}{p_j^{(Y)}} = H(X, Y) - H(Y) \end{aligned}$$

where in the last step we used $\sum_i p_{ij} = p_j^{(Y)}$.

Finally the *mutual information* is defined as

$$\begin{aligned} I(X; Y) &:= \sum_{i,j} p_{ij} \log \frac{p_{ij}}{p_i^{(X)} p_j^{(Y)}} = H(X) + H(Y) - H(X, Y) \\ &= H(X) - H(X | Y), \end{aligned}$$

so its meaning is the reduction in uncertainty of X due to the knowledge of Y , or in other words how much Y tells about X .

4.2 The maximum entropy framework¹

The maximum entropy principle — an example

Suppose we have a random variable X with known states (values of the observations, x_1, \dots, x_n) but unknown probabilities p_1, \dots, p_n ; plus some extra constrains, eg. $\langle X \rangle$ is known. We are given the task to attempt to have a good guess for the probabilities.

Let's start with one of the simplest examples: X can take 1, 2 or 3 with unknown probabilities, and $\langle X \rangle = \bar{x}$ is known. Fixing $\langle X \rangle$ does not determine the probabilities, for example for $\bar{x} = 2$ any $(p_1, p_2, p_3) = (\frac{1-p_2}{2}, p_2, \frac{1-p_2}{2})$ satisfies the constraint, including eg. $(0, 1, 0)$ or $(\frac{1}{2}, 0, \frac{1}{2})$ or $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Which one is the “best”? According to the *maximum entropy principle*, the best guess is the one which maximises the information entropy under the given constraints.

To calculate this solution, we need to find the maximum of $H(p_1, p_2, p_3)$ as a function of p_1, p_2, p_3 , under two constraints: $\langle X \rangle = 1p_1 + 2p_2 + 3p_3 = \bar{x}$ and $p_1 + p_2 + p_3 = 1$. We use the method of Lagrange multipliers: first calculate the unconditional maximum of the original function plus the constraints added with some multiplying factors (the Lagrange multipliers), which give the probabilities in a functional form with the Lagrange

¹ In this section we follow the treatment of [1].

multipliers as parameters.

$$\begin{aligned}
0 &= d \left[H(p_1, p_2, p_3) - \lambda \left(\sum_{i=1}^3 ip_i - \bar{x} \right) - \mu \left(\sum_{i=1}^3 p_i - 1 \right) \right] \\
&= d \left[- \sum_{i=1}^3 p_i \log p_i - \lambda \sum_{i=1}^3 ip_i - \mu \sum_{i=1}^3 p_i \right] \\
&= \sum_{i=1}^3 \{ - \log p_i - 1 - \lambda i - \mu \} dp_i = 0
\end{aligned}$$

Since this has to hold for any dp_i , the curly brackets need to be zero:

$$- \log(p_i) - 1 - \lambda i - \mu = 0, \quad i = 1, 2, 3$$

which with the notation $\lambda_0 = \mu + 1$ gives

$$p_i = e^{-\lambda_0 - \lambda i}.$$

Now we set the Lagrange multipliers by requiring the constraints to be satisfied. The constraint on the sum of probabilities give

$$1 = \sum_{i=1}^3 p_i = e^{-\lambda_0} \sum_{i=1}^3 e^{-\lambda i} \quad \Rightarrow \quad e^{-\lambda_0} = \frac{1}{e^{-\lambda} + e^{-2\lambda} + e^{-3\lambda}}$$

so

$$p_i = \frac{e^{-\lambda i}}{e^{-\lambda} + e^{-2\lambda} + e^{-3\lambda}} = \frac{e^{\lambda(1-i)}}{1 + e^{-\lambda} + e^{-2\lambda}}$$

The other constraint, $\langle X \rangle = \bar{x}$ gives

$$\bar{x} = \sum_{i=1}^3 ip_i = \frac{1 + 2e^{-\lambda} + 3e^{-2\lambda}}{1 + e^{-\lambda} + e^{-2\lambda}} \quad (4.4)$$

Multiplying the equation with the denominator gives a second degree equation for $e^{-\lambda}$, which has the solution

$$e^{-\lambda} = \frac{2 - \bar{x} \pm \sqrt{4 - 3(\bar{x} - 2)^2}}{2(\bar{x} - 3)}$$

Now if we rewrite (4.4) as

$$\bar{x} = \frac{e^{\lambda} + 2 + 3e^{-\lambda}}{e^{\lambda} + 1 + e^{-\lambda}} = 1 + \frac{1 + 2e^{-\lambda}}{e^{\lambda} + 1 + e^{-\lambda}}$$

then p_2 becomes

$$p_2 = \frac{-1 + \sqrt{4 - 3(\bar{x} - 2)^2}}{3}$$

Note that one of the roots have been dropped to keep p_2 non-negative. Finally the other probabilities become

$$p_1 = \frac{3 - \bar{x} - p_2}{2}, \quad p_3 = \frac{\bar{x} - 1 - p_2}{2}$$

This solution has the right behaviour in the limiting cases: when $\bar{x} = 1$, the probabilities $(p_1, p_2, p_3) = (1, 0, 0)$; and when $\bar{x} = 3$, they are $(0, 0, 1)$. For $\bar{x} = 2$, the solution is $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. The maximum entropy solution assigns zero probabilities only when no other possibilities are allowed. This is a very desirable property: it would be a sure failure to propose that a certain state has zero probability, and then find out that a given observation happened to yield that state. The Maximum Entropy solution is guaranteed not to fail there.

Maximum entropy principle — general form

After having this worked out example, we state the maximum entropy principle in a more general form. Suppose we have a random variable X taking known values x_1, \dots, x_n with unknown probabilities p_1, \dots, p_n . In addition, we have m constraint functions $f_k(x)$ with $1 \leq k \leq m < n$, where

$$\langle f_k(X) \rangle = F_k,$$

the F_k s are fixed. Then the maximum entropy principle assigns probabilities in such a way that maximises the information entropy of X under the above constraints. This is the “best guess” in the absence of any further knowledge about the random variable. Since any extra assumption would bring a reduction in uncertainty (see mutual information), we explicitly deny those extra assumptions by maximising the uncertainty.

In the following we calculate various properties of the maximum entropy solution. This may sound dry, but has the advantage that these abstract results can be very easily applied later for concrete examples.

To obtain a formal solution we proceed in a similar way as in the example, maximise the information entropy using Lagrange multipliers:

$$\begin{aligned} 0 &= d \left[H(p_1, \dots, p_n) - \sum_{k=1}^m \lambda_k \left(\sum_{i=1}^n f_k(x_i) p_i - F_k \right) - \underbrace{\mu}_{\lambda_0 - 1} \left(\sum_{i=1}^n p_i - 1 \right) \right] \\ &= \sum_{i=1}^n \left\{ -\log(p_i) - 1 - \sum_{k=1}^m \lambda_k f_k(x_i) - (\lambda_0 - 1) \right\} dp_i \end{aligned}$$

Since this is zero for any dp_i , all n braces have to be zero, giving

$$p_i = \exp\left(-\lambda_0 - \sum_{k=1}^m \lambda_k f_k(x_i)\right) \quad (4.5)$$

Then all the Lagrange multipliers $(\lambda_0, \lambda_1, \dots, \lambda_m)$ are fixed by substituting back into the constraints. The sum of probabilities give

$$1 = \sum_{i=1}^n p_i = e^{-\lambda_0} \sum_{i=1}^n \exp\left(-\sum_{k=1}^m \lambda_k f_k(x_i)\right).$$

The sum after $e^{-\lambda_0}$ appears frequently, so it is useful to consider it separately: we will call it *partition function*

$$Z(\lambda_1, \dots, \lambda_m) := \sum_{i=1}^n \exp\left(-\sum_{k=1}^m \lambda_k f_k(x_i)\right). \quad (4.6)$$

With this notation

$$e^{-\lambda_0} = \frac{1}{Z(\lambda_1, \dots, \lambda_m)}. \quad (4.7)$$

The other constraints are

$$\begin{aligned} F_k &= \sum_{i=1}^n f_k(x_i) p_i = e^{-\lambda_0} \sum_{i=1}^n f_k(x_i) \exp\left(-\sum_{k=1}^m \lambda_k f_k(x_i)\right) \\ &= -\frac{1}{Z} \frac{\partial Z(\lambda_1, \dots, \lambda_m)}{\partial \lambda_k} = -\frac{\partial \log Z(\lambda_1, \dots, \lambda_m)}{\partial \lambda_k}, \end{aligned} \quad (4.8)$$

which is m implicit equations, just enough to determine in principle the m unknowns λ_k . Using (4.7) then the probabilities (4.5) are then fully determined:

$$p_i = \frac{1}{Z(\lambda_1, \dots, \lambda_m)} \exp\left(-\sum_{k=1}^m \lambda_k f_k(x_i)\right) \quad (4.9)$$

Unlike the simple example we had with three states, in practice it is usually not possible to calculate the λ_k s explicitly as a function of F_k s, but as we see later this does not prevent us obtaining lots of useful results.

Consider now the value of the maximised information entropy. It is no longer function of the probabilities, but instead of the constraint values

F_k , and to reflect this we change notation to S :

$$\begin{aligned} \overline{S}(F_1, \dots, F_m) &:= \underbrace{H(p_1, \dots, p_n)}_{\text{from (4.9)}} = - \sum_{i=1}^n p_i \left(-\lambda_0 - \sum_{k=1}^m \lambda_k f_k(x_i) \right) \\ &= \lambda_0 + \sum_{k=1}^m \lambda_k \sum_{i=1}^n f_k(x_i) p_i = \log Z(\lambda_1, \dots, \lambda_m) + \sum_{k=1}^m \lambda_k F_k \end{aligned} \quad (4.10)$$

Now calculate the partial derivatives of S w.r.t. the F_k s, being careful about what is kept constant in the partial derivatives²:

$$\left. \frac{\partial S}{\partial F_k} \right|_{\{F\}} = \sum_{\ell=1}^m \underbrace{\left. \frac{\partial \log Z}{\partial \lambda_\ell} \right|_{\{\lambda\}}}_{F_\ell} \left. \frac{\partial \lambda_\ell}{\partial F_k} \right|_{\{F\}} + \sum_{\ell=1}^m \left. \frac{\partial \lambda_\ell}{\partial F_k} \right|_{\{F\}} F_\ell + \lambda_k = \lambda_k \quad (4.11)$$

Here either $S(F_1, \dots, F_m)$ or $\log Z(\lambda_1, \dots, \lambda_m)$ give a full description of the system, as the other can be calculated using (4.10), and there is a symmetric relation between their partial derivatives: (4.8) and (4.11). We look at this kind of relation between two functions more closely below.

Legendre transform

Consider a convex function $f(x)$, and define the following function

$$f^*(p) := \max_x (px - f(x)). \quad (4.12)$$

We call this³ the *Legendre transform* of $f(x)$. If f is differentiable as well, then the maximum can be calculated as

$$0 = \frac{d}{dx}(px - f(x)) = p - \frac{df(x)}{dx}$$

Its solution for x depends on p , which we call $x(p)$:

$$\left. \frac{df(x)}{dx} \right|_{x=x(p)} = p$$

² In thermodynamics and statistical physics functions of many variables are used extensively, and the notation is not always clear on what the free variables are. When taking partial derivatives, it is essential to be clear on what is kept constant; therefore it is often shown at the bottom of the vertical bar after the partial differential. Eg. the notation $\{\lambda\}$ means all λ_j s are kept fixed except the one we differentiate with.

³ The Legendre transform is sometimes defined with a sign difference: $f^*(p) = \max(f(x) - px)$. The advantage of our notation is that the inverse, as we soon see, is completely symmetric.

which plugged into (4.12) gives

$$f^*(p) = px(p) - f(x(p)).$$

Now let's calculate the Legendre transform of f^* :

$$(f^*)^*(y) = \max_p (yp - f^*(p)).$$

Again, if f^* is differentiable then

$$\left. \frac{df^*(p)}{dp} \right|_{p=p(y)} = y.$$

However,

$$\frac{df^*(p)}{dp} = \frac{px(p) - f(x(p))}{dp} = x(p) + p \frac{dx(p)}{dp} - \underbrace{\frac{df(x)}{dx}}_p \bigg|_{x(p)} \frac{dx(p)}{dp} = x(p)$$

so

$$y = \left. \frac{df^*(p)}{dp} \right|_{p=p(y)} = x(p(y)),$$

thus

$$f^{**}(y) = yp(y) - f^*(p(y)) = yp(y) - p(y)x(p(y)) + f(x(p(y))) = f(y).$$

We just obtained that the functions $f^{**}(\cdot)$ and $f(\cdot)$ are equal, or equivalently the Legendre transform is its own inverse.

The Legendre transform can be easily generalised to concave functions: in the definition max needs to be replaced by min.

The other generalisation applies to functions of multiple variables: the Legendre transform of $f(x_1, \dots, x_m)$ is

$$f^*(p_1, \dots, p_k) = \sum_{k=1}^m x_k p_k - f(x_1, \dots, x_m), \quad \text{where } p_k = \frac{\partial f}{\partial x_k}$$

Now looking back to the maximum entropy solution, (4.10) and (4.8) establish that $S(F_1, \dots, F_m)$ and $-\log Z(\lambda_1, \dots, \lambda_m)$ are Legendre transforms of each other. Having seen the symmetric structure of the Legendre transform, (4.11) is no longer surprising. The only remaining bit is to show that $-\log Z$ is indeed either convex or concave so that the Legendre transform is defined, to which we come back soon.

Reciprocity laws and covariances

We can easily derive relationships between partial derivatives of the constraints F_k and Lagrange multipliers λ_k . By changing the order of partial differentiations we obtain

$$\left. \frac{\partial F_k}{\partial \lambda_j} \right|_{\{\lambda\}} = \left. \frac{\partial^2 - \log Z}{\partial \lambda_j \partial \lambda_k} \right|_{\{\lambda\}} = \left. \frac{\partial^2 - \log Z}{\partial \lambda_k \partial \lambda_j} \right|_{\{\lambda\}} = \left. \frac{\partial F_j}{\partial \lambda_k} \right|_{\{\lambda\}}. \quad (4.13)$$

Similarly

$$\left. \frac{\partial \lambda_k}{\partial F_j} \right|_{\{F\}} = \left. \frac{\partial^2 S}{\partial F_j \partial F_k} \right|_{\{F\}} = \left. \frac{\partial^2 S}{\partial F_k \partial F_j} \right|_{\{F\}} = \left. \frac{\partial \lambda_j}{\partial F_k} \right|_{\{F\}}.$$

By cursory observation one might say the second equation is just the reciprocal of the first one, so it is not telling anything new. This is wrong, as the quantities that are kept fixed at differentiation are not the same. However, the naive notion of inverse holds in a more intricate way: the matrices with elements $A_{jk} = \partial F_j / \partial \lambda_k$ and $B_{jk} = \partial \lambda_j / \partial F_k$ are inverses of each other: $A = B^{-1}$.

When we set $\langle f_k(X) \rangle = F_k$, we required that *the expectation of $f_k(X)$* is what is prescribed, but still it varies from observation to observation. Now we look at how large these fluctuations are.

The *covariance* of two random variables is defined as

$$\text{Cov}(X, Y) := \langle [X - \langle X \rangle][Y - \langle Y \rangle] \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle,$$

which is a measure of “how much Y is above its average at the same time when X is above its average”. A covariance of a random variable with itself is called *variance*:

$$\text{Var}(X) := \text{Cov}(X, X) = \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2,$$

with the convenient meaning that its square root (the *standard deviation* σ) measures how much a random variable differs from its average, suitably weighted. (The variance is always non-negative, as it is the average of a non-negative quantity: a square.)

So we can calculate the covariance of $f_k(X)$ and $f_j(X)$:

$$\text{Cov}(f_j(X), f_k(X)) = \langle f_j(X) f_k(X) \rangle - \langle f_j(X) \rangle \langle f_k(X) \rangle$$

The first term using (4.9) is

$$\begin{aligned}\langle f_j(X)f_k(X) \rangle &= \frac{1}{Z} \sum_{i=1}^n f_j(x_i)f_k(x_i) \exp\left(-\sum_{\ell=1}^m \lambda_\ell f_\ell(x_i)\right) \\ &= \frac{1}{Z} \frac{\partial^2 Z(\lambda_1, \dots, \lambda_m)}{\partial \lambda_j \partial \lambda_k}.\end{aligned}$$

As a side remark, the above calculation easily generalises to averages of arbitrary products of f_k s:

$$\langle f_j^{m_j}(X)f_k^{m_k}(X) \dots \rangle = \frac{1}{Z} \left(\frac{\partial^{m_j}}{\partial \lambda_j^{m_j}} \frac{\partial^{m_k}}{\partial \lambda_k^{m_k}} \dots \right) Z.$$

Coming back to the covariance

$$\begin{aligned}\text{Cov}(f_j(X), f_k(X)) &= \frac{1}{Z} \frac{\partial^2 Z}{\partial \lambda_j \partial \lambda_k} - \frac{1}{Z^2} \frac{\partial Z}{\partial \lambda_j} \frac{\partial Z}{\partial \lambda_k} = \frac{\partial^2 \log Z}{\partial \lambda_j \partial \lambda_k} \\ &= -\frac{\partial F_k}{\partial \lambda_j} = -\frac{\partial F_j}{\partial \lambda_k},\end{aligned}$$

where we have seen the last steps already in (4.13). Similarly for variance

$$0 \leq \text{Var}(f_k(X)) = \frac{\partial^2 \log Z}{\partial \lambda_k^2} = -\frac{\partial F_k}{\partial \lambda_k} \quad (4.14)$$

This confirms that the second derivative of $\log Z$ is non-negative, ie. $\log Z$ is a convex function, which we implicitly assumed when mentioned that $-\log Z$ and S are Legendre transforms of each other.

Suppose now that the constraint functions f_k depend on an external parameter: $f_k(X; \alpha)$. Everything, including Z and S become dependent on α . To see its effect we calculate partial derivatives:

$$\begin{aligned}-\frac{\partial \log Z}{\partial \alpha} \Big|_{\{\lambda\}} &= -\frac{1}{Z} \sum_{i=1}^n \exp\left(-\sum_{k=1}^m \lambda_k f_k(x_i; \alpha)\right) \sum_{k=1}^m -\lambda_k \frac{\partial f_k(x_i; \alpha)}{\partial \alpha} \\ &= \sum_{k=1}^m \lambda_k \left\langle \frac{\partial f_k}{\partial \alpha} \right\rangle.\end{aligned} \quad (4.15)$$

Similarly, using $S = \log Z + \sum_k \lambda_k F_k$:

$$\begin{aligned}\frac{\partial S(F_1, \dots, F_n; \alpha)}{\partial \alpha} \Big|_{\{F\}} &= \sum_{k=1}^m \underbrace{\frac{\partial \log Z}{\partial \lambda_k} \Big|_{\{\lambda\}}}_{-F_k} \frac{\partial \lambda_k}{\partial \alpha} \Big|_{\{F\}} + \frac{\partial \log Z}{\partial \alpha} \Big|_{\{\lambda\}} + \\ &+ \sum_{k=1}^m \frac{\partial \lambda_k}{\partial \alpha} \Big|_{\{F\}} F_k = \frac{\partial \log Z}{\partial \alpha} \Big|_{\{\lambda\}}.\end{aligned}$$

So the partial derivatives of $\log Z$ and S with respect to α are equal, though one should note that the variables kept fixed are the natural variables in each case.

4.3 Applications of the maximum entropy framework

The microcanonical ensemble

The simplest system to consider is the isolated one, with no interaction with its environment. A physical example can be a thermally and mechanically isolated box containing some gas, conventionally these are called *microcanonical ensembles*. With no way to communicate, we have no information about the current state of the system. To put it in the maximum entropy framework, we do not have any constraint to apply.

The maximum entropy solution for such a system is

$$Z = \sum_{i=1}^n 1, \quad p_i = \frac{1}{Z}, \quad S = \log Z.$$

Using the conventions of statistical physics the number of states is denoted by Ω , and the unit of entropy is k_B : recall this sets the prefactor and/or the base of the logarithm in (4.2)-(4.3). Using this notation (the MC subscript denotes microcanonical):

$$Z = \Omega, \quad p_i = \frac{1}{Z} = \frac{1}{\Omega}, \quad S_{\text{MC}} = k_B \ln \Omega.$$

In this most simple system all internal states have equal probability.

The canonical ensemble

In the next level of increasing complexity, we allow the exchange of one conserved quantity with the external environment. The physical example is a system which is thermally coupled (allowing energy exchange) with its environment; conventionally these are called *canonical ensembles*. Using this terminology we label the internal states with their energy. By having the ability to interact with the system, we can control eg. the average energy of the system by changing the condition of the environment, corresponding to having one constraint in the maximum entropy formalism.

The maximum entropy solution for one constraint reads

$$Z(\lambda) = \sum_{i=1}^n e^{-\lambda f(x_i)}, \quad p_i = \frac{1}{Z} e^{-\lambda f(x_i)}, \quad S(F) = \log Z(\lambda) + \lambda F.$$

The conventional units for entropy is k_B for canonical ensembles as well, and as we mentioned the states are labelled with energy: $f(x_i) = E_i$ with average energy (the value of the constraint) $F = E$. Finally, the Lagrange multiplier λ is called $\beta = 1/(k_B T)$ in statistical physics, where T is temperature (measured in Kelvins), and k_B is the Boltzmann constant. Thus we have

$$Z(\beta) = \sum_{i=1}^n e^{-\beta E_i}, \quad p_i = \frac{e^{-\beta E_i}}{Z} = \frac{e^{-\frac{E_i}{k_B T}}}{Z}, \quad S_C(\langle E \rangle) = k_B \ln Z + \frac{\langle E \rangle}{T}$$

In p_i the exponential factor $e^{-\beta E_i}$ is called Boltzmann factor, while Z provides the normalisation.

Having established this connection, we can easily translate the results of the maximum entropy formalism. Eqs. (4.8) and (4.11) become

$$\langle E \rangle = -\frac{\partial \ln Z}{\partial \beta} \quad \text{and} \quad \frac{1}{T} = \frac{\partial S_C}{\partial \langle E \rangle}$$

Eq. (4.14) gives the energy fluctuation:

$$\sigma_E^2 = \text{Var}(E) = \frac{\partial^2 \ln Z}{\partial \beta^2} = -\frac{\partial \langle E \rangle}{\partial \beta} = \underbrace{\frac{\partial \langle E \rangle}{\partial T}}_{C_V} k_B T^2$$

where C_V is the *heat capacity* of the system. This is an interesting relation, connecting microscopic fluctuations with macroscopic thermodynamic quantities.

In practice it is useful to define the following quantity, called *Helmholtz free energy*:

$$A := -k_B T \ln Z = \langle E \rangle - T S_C,$$

If we consider it as a function of temperature, $A(T)$, its derivative is

$$\frac{\partial A}{\partial T} = -k_B \ln Z - k_B T \frac{\partial \ln Z}{\partial \beta} \frac{1}{k_B T^2} = -S_C$$

This leads to a relation with the energy. Our approach so far determined the entropy S_C as a function of average energy $\langle E \rangle$. Considering its inverse function $\langle E \rangle(S_C)$, we see that its Legendre transform is $-A(T)$.

It is interesting to note that

$$\sum_i \exp\left(-\frac{E_i}{k_B T}\right) = Z = \exp\left(-\frac{A}{k_B T}\right),$$

so the sum of Boltzmann factors equals to a single Boltzmann factor with energy replaced with the Helmholtz free energy. We will see its implications later in the grand canonical ensemble.

Next we consider a system made of two subsystems, which are sufficiently uncoupled. The joint partition function can be written as [labeling the left and right subsystem with (L) and (R)]:

$$Z = \sum_i \sum_j e^{-\beta(E_i^{(L)} + E_j^{(R)})} = \left(\sum_i e^{-\beta E_i^{(L)}}\right) \left(\sum_j e^{-\beta E_j^{(R)}}\right) = Z^{(L)} Z^{(R)}$$

This means that $\ln Z$ is additive: $A = -k_B T \ln Z = A^{(L)} + A^{(R)}$. Other quantities, like the entropy or the energy have the similar additive property, and we call these *extensive* quantities.

Physical examples for canonical ensembles

We have seen that to calculate any statistical mechanics quantity for a given system, the partition function is calculated first, and then any other quantity is easily expressed. We will consider physical systems, like a particle at position x , momentum $p = mv$, and energy $E = p^2/(2m) + U(x)$, where U is the potential. In systems made of discrete states the formula involves a sum over the states. For continuous systems, however, the sum needs to be replaced by integration:

$$\sum_i (\cdot) \leftrightarrow \frac{1}{h} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp (\cdot) \quad (4.16)$$

This is a *semiclassical* formula: not quantum mechanical, as x and p are independent variables and not non-commuting operators; but not purely classical either as the Planck constant h is involved. Instead of fully understanding, we just rationalise this formula as (i) a constant needs to appear in front of the integrals to make the full expression dimensionless, as Z should be, and (ii) in quantities involving $\log Z$ the prefactor $1/h$ becomes an additive constant, and in particular for the entropy it sets its zero level.

The simplest example is a one-dimensional box of length L . The po-

tential can be taken as zero within the box and infinity outside, giving

$$\begin{aligned} Z &= \frac{1}{h} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp \exp\left(-\beta \left[\frac{p^2}{2m} + U(x)\right]\right) \\ &= \frac{1}{h} \int_0^L dx \int_{-\infty}^{\infty} dp \exp\left(-\beta \frac{p^2}{2m}\right) = \frac{L}{h} \sqrt{\frac{2\pi m}{\beta}} = \frac{L}{\lambda} \end{aligned} \quad (4.17)$$

where we used the Gaussian integral $1 = \int_{-\infty}^{\infty} \exp(-x^2/2\sigma^2)/\sqrt{2\pi\sigma^2}$. The factors other than L are collected into a quantity of dimension length: $\lambda = h/\sqrt{2\pi mk_B T}$, called *thermal de Broglie wavelength*. When it is small compared to characteristic length scales, in our case $\lambda \ll L$, the system can be considered as classical; while if $\lambda \gtrsim L$, proper quantum mechanics needs to be used. Interestingly, this does not only involve size, but also mass and temperature. This is the reason why typically electrons are always quantum mechanical, but full atoms can be considered as classical (as is done in molecular dynamics simulations). The exception is very light atoms at very low temperature, when inherently quantum effects like superfluidity of helium can be observed.

The ideal gas is a model of gases where gas atoms or molecules are point particles which do not interact. Since in the energy the x , y , and z components are decoupled, the coordinates of all N particles can be considered as independent, which using (4.16) and (4.17) leads to

$$Z = \frac{1}{N!} \left(\frac{L}{h} \sqrt{\frac{2\pi m}{\beta}}\right)^{3N} = \frac{1}{N!} \left(\frac{V}{\lambda^3}\right)^N \quad (4.18)$$

The $1/N!$ comes from the fact that the particles are indistinguishable: states where eg. particle 1 has position and momentum $\mathbf{r}_a, \mathbf{p}_a$ and particle 2 has $\mathbf{r}_b, \mathbf{p}_b$ is identical to the state where particle 1 of $\mathbf{r}_b, \mathbf{p}_b$ and particle 2 of $\mathbf{r}_a, \mathbf{p}_a$; the factor corrects the double counting in the integrals. Having Z , it is easy to show that the average energy $\langle E \rangle = (3/2)Nk_B T$, the Helmholtz free energy $A = Nk_B T(\log(\rho\lambda^3) - 1)$, and the entropy $S = Nk_B(5/2 - \log(\rho\lambda^3))$, where $\rho = N/V$ is the number density.

It is interesting to see that considering V as a parameter of the system, we can apply (4.15) to obtain a new relation. Plugging in $\alpha = V$ and $\partial f/\partial\alpha = \partial E/\partial V = -p$ (the latter can be considered as a definition of pressure):

$$-\frac{\partial \log Z}{\partial V} = -\beta \left\langle \frac{\partial E}{\partial V} \right\rangle$$

which using (4.18) gives

$$Nk_B T = \langle p \rangle V$$

This is called *equation of state*, as it provides a relation between state variables like pressure, volume and temperature.

Next we consider another fundamental system, the *harmonic oscillator*. One can think about it as a point mass m moving in one dimension, connected to a spring of stiffness k , of which the other end is kept fixed. If the position x is measured from the equilibrium position (unstretched spring), then the force acting on the the point mass is $-kx$, yielding Newton's equation $md^2x/dt^2 = -kx$. This has a solution $x = A \sin(\omega t + \phi)$, where the amplitude A and phase ϕ are parameters set by the initial condition, and the frequency is $\omega = \sqrt{k/m}$. The energy stored in the spring can be written as $kx^2/2 = m\omega^2 x^2/2$, so the total energy is

$$E = \frac{p^2}{2m} + \frac{m\omega^2}{2}x^2 \quad (4.19)$$

Using the standard recipe we first calculate the partition function:

$$Z = \frac{1}{h} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp e^{-\beta \frac{p^2}{2m}} e^{-\beta \frac{m\omega^2}{2}x^2} = \frac{1}{h} \sqrt{\frac{2\pi m}{\beta}} \sqrt{\frac{2\pi}{m\omega^2\beta}} = \frac{1}{\hbar\omega\beta}$$

where we introduced $\hbar = h/(2\pi)$. Then

$$\langle E \rangle = -\frac{\partial \ln Z}{\partial \beta} = -\frac{\partial}{\partial \beta} \ln \frac{1}{\beta} = k_B T \quad \text{and} \quad C_V = \frac{\partial \langle E \rangle}{\partial T} = k_B$$

This last result is a realisation of the principle of *equipartition*: each quadratic half-degree of freedom [like x and p in (4.19)] contributes $k_B T/2$ to the average energy, and consequently $k_B/2$ to the heat capacity.

We will now apply these results to calculate the heat capacity of solids. Far away from the the melting temperature the many-body potential of the atoms in a crystal can be considered quadratic. Collecting all $3N$ coordinates of the N atoms into a vector $\mathbf{x} = (x_1, x_2, \dots, x_{3N})$, the potential is

$$U(\mathbf{x}) = U_0 + \sum_{i=1}^{3N} \frac{\partial U}{\partial x_i} (x_i - x_i^0) + \frac{1}{2} \sum_{i,j=1}^{3N} \frac{\partial^2 U}{\partial x_i \partial x_j} (x_i - x_i^0)(x_j - x_j^0) + \dots$$

where the series expansion is truncated at the quadratic term. The equation of motion involves the $3N \times 3N$ dynamical matrix $\partial^2 U / \partial x_i \partial x_j$, which separates into $3N$ independent one-dimensional harmonic oscillators corresponding to the normal modes and eigenfrequencies. This leads to $C = 3Nk_B$, known as Dulong-Petit law, which turns out to be correct at high temperatures.

At low temperatures quantum mechanical effects have to be taken into account, which we do simply by replacing the classical harmonic oscillators with quantum harmonic oscillators. For our purposes the quantum harmonic oscillator is a system with discrete energy levels: in the i^{th} state $E_i = (i + \frac{1}{2})\hbar\omega$, where $i = 0, 1, \dots$. Being a discrete system the partition function involves just a sum, which here is a geometric sum:

$$Z = \sum_{i=0}^{\infty} e^{-\beta(i+\frac{1}{2})\hbar\omega} = \frac{e^{-\frac{1}{2}\beta\hbar\omega}}{1 - e^{-\beta\hbar\omega}} = \frac{1}{2 \sinh\left(\frac{\beta\hbar\omega}{2}\right)}$$

The average energy and heat capacity are

$$\langle E \rangle = -\frac{\partial \ln Z}{\partial \beta} = \frac{\hbar\omega}{2} \coth \frac{\hbar\omega}{2k_B T}$$

$$C = \frac{\partial \langle E \rangle}{\partial T} = k_B \left(\frac{\hbar\omega}{2k_B T} \right)^2 \frac{1}{\sinh^2 \frac{\hbar\omega}{2k_B T}}$$

At high temperature (small β) the argument of \sinh is small, which expands to $\sinh x \sim x$. This leads to $C \rightarrow k_B$, which is the classical result.

At low temperature (large β) however, the argument of \sinh is large, expanding to $\sinh x \sim \frac{1}{2}e^x$. This gives $C \approx k_B \left(\frac{\hbar\omega}{k_B T} \right)^2 e^{-\frac{\hbar\omega}{k_B T}}$, resulting in exponential suppression at low temperatures. Naively applying this result to crystals leads to the *Einstein model* of solids, which at low temperatures simply gives $C = 3Nk_B \left(\frac{\hbar\omega}{2k_B T} \right)^2 / \sinh^2 \frac{\hbar\omega}{2k_B T}$.

This is still incorrect, however, since all quantum harmonic oscillators are assumed to have the same frequency. In the *Debye model* of solids the proper spectrum of frequencies is used, which indeed reproduces experimental measurements at low temperatures as well. The Reader is referred to standard solid state physics textbooks for details.

The grand canonical ensemble

We now allow the exchange of two conserved quantities with the external environment: to follow the physical example of *grand canonical ensembles*, these are the energy and the particle number. In the maximum entropy formalism this corresponds to constraining the average energy and the average particle number. As before the units of entropy is k_B , and the i^{th} state has energy E_i and particle number N_i . The Lagrange multiplier conjugate to energy is $\beta = 1/(k_B T)$ as in the canonical ensemble. The other one, however, is conventionally denoted by $-\mu\beta = -\frac{\mu}{k_B T}$.

Accordingly the grand canonical partition function (denoted by Ξ) and the probabilities of the states are

$$\Xi(\beta, \mu) = \sum_i e^{-\beta(E_i - \mu N_i)}$$

$$p_i = \frac{1}{\Xi} e^{-\frac{1}{k_B T}(E_i - \mu N_i)},$$

while the entropy, now function of the average energy and average particle number, using (4.10) becomes

$$S_{GC}(\langle E \rangle, \langle N \rangle) = k_B \ln \Xi + \frac{\langle E \rangle}{T} - \frac{\mu \langle N \rangle}{T}.$$

The simple relations (4.8) and (4.11) become more complicated due to the fact that the physical variables, especially μ , are not simply the Lagrange multipliers but functions of them:

$$\langle E \rangle = - \left. \frac{\partial \ln \Xi}{\partial \beta} \right|_{-\mu, \beta} = - \left. \frac{\partial \ln \Xi}{\partial \beta} \right|_{\mu} + \left. \frac{\partial \ln \Xi}{\partial \mu} \right|_{\beta} \mu k_B T$$

$$\langle N \rangle = - \left. \frac{\partial \ln \Xi}{\partial -\mu\beta} \right|_{\beta} = k_B T \left. \frac{\partial \ln \Xi}{\partial \mu} \right|_{\beta}$$

$$\frac{1}{T} = \left. \frac{\partial S_{GC}}{\partial \langle E \rangle} \right|_{\langle N \rangle}$$

$$-\frac{\mu}{T} = \left. \frac{\partial S_{GC}}{\partial \langle N \rangle} \right|_{\langle E \rangle}$$

In the grand canonical ensemble not only the energy fluctuates, but also the particle number:

$$\sigma_N^2 = \text{Var}(N) = \left. \frac{\partial^2 \ln \Xi}{\partial (-\mu\beta)^2} \right|_{\beta} = k_B T \left. \frac{\partial \langle N \rangle}{\partial \mu} \right|_{\beta}$$

The reciprocity relations also become more complicated, for example

$$\left. \frac{\partial \langle E \rangle}{\partial -\mu\beta} \right|_{\beta} = \left. \frac{\partial \langle N \rangle}{\partial \beta} \right|_{-\mu, \beta}$$

becomes

$$-k_B T \left. \frac{\partial \langle E \rangle}{\partial \mu} \right|_{\beta} = \left. \frac{\partial \langle N \rangle}{\partial \beta} \right|_{\mu} - \mu k_B T \left. \frac{\partial \langle N \rangle}{\partial \mu} \right|_{\beta}$$

An important quantity is the grand free energy (we will see soon the relevance of the free energies), which is defined as

$$\Phi(T, \mu) := -k_B T \ln \Xi = \langle E \rangle - \mu \langle N \rangle - T S_{GC}$$

It is interesting to note that the partition function can be written as

$$\begin{aligned} e^{-\beta\Phi} = \Xi &= \sum_i e^{-\beta E_i} e^{\beta\mu N_i} = \sum_{N=0}^{\infty} e^{\beta\mu N} \sum_j e^{-\beta E_{j,N}} \\ &= \sum_N e^{-\beta(A(T;N) - \mu N)} \end{aligned}$$

In this expression microscopic states with the same particle number N are lumped together into a macroscopic state, and the sum of their Boltzmann factors is replaced by a single Boltzmann factor where the role of the energy is played by an appropriate free energy. This manipulation is called *partial trace*, a terminology borrowed from the quantum formalism of statistical mechanics.

4.4 Fluctuations and thermodynamics

In the previous section we calculated the energy and particle number fluctuations in the canonical and grand canonical ensembles. Considering how the *relative* fluctuations depend on the system size, we obtain

$$\begin{aligned} \left. \frac{\sigma_E}{\langle E \rangle} \right|_{\text{canonical ensemble}} &= \frac{\sqrt{k_B T^2 \frac{\partial \langle E \rangle}{\partial T}}}{\langle E \rangle} \sim \frac{1}{\sqrt{\langle E \rangle}} \\ \left. \frac{\sigma_N}{\langle N \rangle} \right|_{\text{grand canonical ens.}} &= \frac{\sqrt{k_B T \frac{\partial \langle N \rangle}{\partial \mu}}}{\langle N \rangle} \sim \frac{1}{\sqrt{\langle N \rangle}} \end{aligned}$$

In both cases the relative fluctuations decay as the $-\frac{1}{2}$ power of the system size. In the $N \rightarrow \infty$ limit, called *thermodynamic limit*, the fluctuating quantities (when rescaling with the system size) become definite, not random. Thus we can replace $\langle E \rangle$ with E etc. This is why statistical mechanics is the microscopic foundation of thermodynamics.

In many cases fluctuations are the aggregate effect of many independent contributions. To consider this case more rigorously, suppose X_i are *iid* (independent, identically distributed) random variables, with $\langle X_i \rangle = \mu$ and $\text{Var}(X_i) = \sigma^2$. Then the *Central limit theorem* states that

$$Z_n := \frac{\overbrace{X_1 + \dots + X_n}^{S_n} - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} \mathcal{N}(0, 1)$$

Here $\mathcal{N}(0, 1)$ is the distribution of standard normal (Gaussian) random

variables, ie. with zero mean and unit variance. The notation \xrightarrow{D} means convergence in distribution:

$$\lim_{n \rightarrow \infty} P(Z_n < z) = P(\zeta < z)$$

where ζ is a standard normal random variable. Note that this is pointwise convergence of the cumulative distribution function, which is weaker than the convergence of the probability density function.

The Central limit theorem is behind the fact that the normal distribution is so prevalent: for macroscopic fluctuations often the microscopic contributions are sufficiently independent. As we have seen before the relative fluctuations of the sum decrease as $1/\sqrt{n}$:

$$\frac{\sigma_{S_n}}{\langle S_n \rangle} \rightarrow \frac{\sqrt{n}\sigma}{n\mu} \sim \frac{1}{\sqrt{n}}$$

A simple application is the one-dimensional random walk: X_i takes values ± 1 each with probability $1/2$. The resulting trajectory, $S_n = X_1 + \dots + X_n$ is *like* a Gaussian variable with mean zero and standard deviation \sqrt{n} , when sufficiently coarse grained to remove the discreteness.

Certain important cases fall outside the applicability of the Central limit theorem, like distributions where the variance (or the mean as well) is undefined. One such example is the Cauchy (or Lorentz) distribution, defined by the probability density function

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{or} \quad f(x) = \frac{1}{\pi\gamma \left(1 + \left(\frac{x-x_0}{\gamma}\right)^2\right)}$$

Surprisingly the average of n iid Cauchy random variables has the same distribution as just one, which means that if one deals with such quantities, taking averages is useless.

When generalising this phenomena one arrives at the concept of *stable distributions*: these are families of distributions where the sum of such random variables is from the same family. More formally, let $\text{Fam}(\Theta)$ represent a family of distributions where Θ denotes all the parameters. Suppose X_1 and X_2 are from this family. If their linear combination is also from this family:

$$X_1 \sim \text{Fam}(\Theta_1), \quad X_2 \sim \text{Fam}(\Theta_2) \quad \Rightarrow \quad aX_1 + bX_2 \sim \text{Fam}(\Theta_3) + c$$

then we call Fam a stable distribution.

We have seen that both the normal and the Cauchy are stable distributions. One more where the probability density function can be given in closed form is the Levy distribution:

$$f(x) = \sqrt{\frac{c}{2\pi}} \frac{e^{-c/(2x)}}{x^{3/2}}$$

which can be generalised to the 4-parameter Levy-skew- α -stable family.

This distribution underpins the *Levy flight*, which is similar to a random walk, but the increments are taken from a heavy tailed distribution,

$$f(x) \sim 1/|x|^{\alpha+1}, \quad \text{where } 0 < \alpha < 2.$$

Thermodynamics

As we mentioned, thermodynamics is founded by statistical mechanics. However, it can also be considered as a self-standing axiomatic theory, based on the following axioms:

- (0) There exist a relation between thermodynamic systems. This relation is called thermodynamic equilibrium, and it is transitive (equivalence relation):
if $A \sim B$ and $B \sim C$, then $A \sim C$.
Here A , B and C label different systems. For example in thermal equilibrium this means a transitive relation between the temperatures of the three systems.
- (1) Energy conservation: the total energy of an isolated system is fixed.
Thus if during some process a system absorbs heat ΔQ , as well as work $\Delta W = -p \Delta V + \dots$ is made on it, then its energy changes by $\Delta E = \Delta Q + \Delta W$.
- (2) In an isolated system the entropy does not decrease.
Thus if during some process a system absorbs heat ΔQ , then its entropy changes by $\Delta S = \Delta Q/T + \Delta S_{\text{internal}} \geq \Delta Q/T$.
- (3) The entropy at absolute zero temperature is zero, or it is independent of other parameters, so can be set to zero. Another form is that the ground state of a quantum system has finite multiplicity (or at least not exponential with N).
A simple consequence is that the heat capacity vanishes at absolute zero temperature:

$$C_X = \left. \frac{\partial Q}{\partial T} \right|_X = T \left. \frac{\partial S}{\partial T} \right|_X \rightarrow 0 \quad \text{if } T \rightarrow 0$$

where the subscript X corresponds to the quantity held fixed, it can be anything suitable, eg. volume or pressure.

Let us suppose now that we bring a thermodynamic system from one state to another. Certain thermodynamic quantities, like work or heat do depend on which path is taken, while others like total energy, entropy, free energy, temperature, pressure etc. are path-independent. We call the latter ones *state variables*.

Now we return to free energies. Consider a system kept at fixed temperature T , which undergoes some change. If it is otherwise isolated from its environment (canonical ensemble), then its energy changes only by the absorbed heat: $\Delta E = \Delta Q$. The change in entropy is

$$\Delta S = \frac{\Delta Q}{T} + \Delta S_{\text{int}} \geq \frac{\Delta Q}{T}$$

since the second law of thermodynamics states that $\Delta S_{\text{int}} \geq 0$. Introducing the Helmholtz free energy as $A(T) = E - TS$, its change is

$$\Delta A = \Delta E - T\Delta S \leq 0,$$

so the free energy never increases during any change or transition. Then it follows that at stable equilibrium it must be minimal. This is a very important point, which makes free energies central in thermodynamics. The corresponding observation in statistical mechanics, where there are always fluctuations, is that the probability of a macroscopic state is proportional to $\exp\left(-\frac{A(T)}{k_B T}\right)$. For a large system the difference between free energies of different states are large: $\Delta A \gg k_B T$, so only the state with the lowest free energy is observed.

The above observations hold for any ensemble, when the appropriate free energy is used. To illustrate this we will have another example. Consider now a system kept at constant temperature T and pressure p , so exchange of heat and volume with its environment is allowed. When undergoing a change, its energy changes by $\Delta E = \Delta Q - p\Delta V$. The change in entropy is

$$\Delta S = \frac{\Delta Q}{T} + \Delta S_{\text{int}} \geq \frac{\Delta Q}{T}$$

If we now introduce the Gibbs free energy: $G(T, p) = E - TS + pV$, its change is

$$\Delta G = \Delta E - T\Delta S + p\Delta V \leq 0$$

non-positive again. The Gibbs free energy is the relevant quantity for constant T , p environments, like in biochemical reactions.

4.5 Phase transitions

As we mentioned in the beginning of this chapter, complex systems display a nontrivial collective behaviour of the constituents. Phase transitions are the most spectacular of these emergent phenomena. They correspond to sudden change in behaviour by a small change in the controlling parameters.

Probably the first example to come to mind is the melting of ice, and boiling of water – those in the domain of Physics are the best understood, since the microscopic interactions are completely known. There are many other, potentially more contested examples as well, like the fall of communism in East Europe around 1989.

Coming back to Physics, the transition occurs between different phases. The possible phases include the grade school examples of solids, liquids and gases, but there are many more, like magnets, a whole zoo of liquid crystals, superconductors, superfluids, different crystal structures of the same material etc.

When considering phases, *symmetry* is an important concept. It can be thought of as a collection of operations that map a system back to itself. If two states have different symmetry, these are necessarily different phases. However, it is possible to have two different phases with the same symmetry, like both liquids and gases are homogeneous (translation invariant in a statistical sense) and isotropic (all directions are equivalent). When state A and state B corresponds to phases of different symmetry, any path connecting A and B has at some point a jump in symmetry, thus it necessarily involves a phase transition. However, when the phases are of the same symmetry, it is possible, as in Fig. 4.1, to move from A to B without a phase transition.

Order parameter field

The first step to describe a phase transition is to define an *order parameter field*. This is an extraction of the relevant quantities from a large number of degrees of freedom. This is not straightforward to do, and represents a very important step in understanding the phase transition. A few examples:

liquids and gases: the order parameter field is the density ρ : it is a *scalar* assigned to each point in space.

magnets: the order parameter field is the locally averaged magnetisation: it is a *vector* \mathbf{v} at each space point.

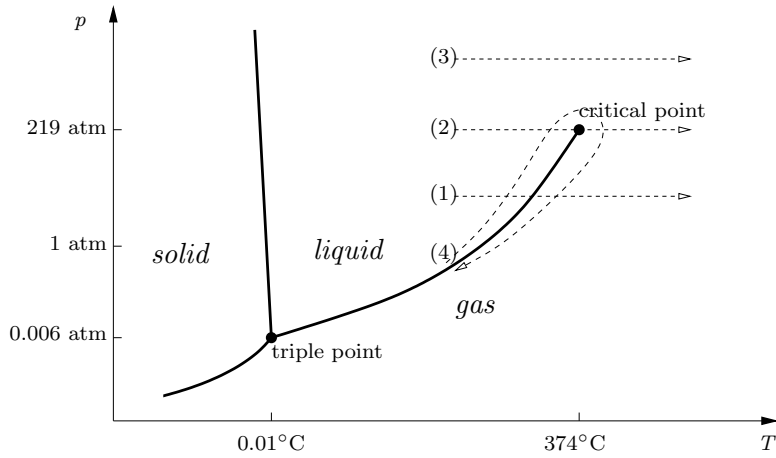


Figure 4.1 Simplified phase diagram of water. Path (1) corresponds to an abrupt phase transition between the liquid and gas phases, path (2) is a continuous phase transition, while path (3) involves no phase transition as it crosses neither a coexistence line nor a critical point. Since liquid and gas has the same symmetry, it is possible, as in path (4), to move from one side of the coexistence line to the other without a phase transition.

With the extra restriction of fixed temperature, $|\mathbf{v}|$ becomes fixed: corresponding to a point on the surface of a *sphere*.

nematic liquid crystals: these consist of thin rod molecules, which have an orientational order. The order parameter field is the locally averaged orientation \mathbf{d} . Since the two ends of the molecules are equivalent, a 180° rotation does not change the orientation. Mathematically this property can be expressed as $\mathbf{d} = -\mathbf{d}$: these objects are called *directors*. With the extra restriction of fixed temperature, $|\mathbf{d}|$ becomes fixed, corresponding to a point on a *hemisphere*.

crystals: the order parameter field is the local translation needed to return an atom to a perfect lattice site. Since all perfect lattice sites are equivalent, this means that our object does not change if we add integer times a lattice vector. Eg. in two dimensions this means $\mathbf{u} = \mathbf{u} + ma_x\hat{x} + na_y\hat{y}$, where a_x and a_y are the lattice constants and \hat{x} and \hat{y} are the lattice directions. These objects are often called *wrapped vectors*. They can be represented by a rectangle, where the pairs of opposite sides are identified, which is topologically equivalent to the surface of a torus.

superconductors: the order parameter field is a complex number Ψ ,

corresponding to the quantum mechanical phase of the condensed state. At fixed temperature $|\Psi|$ is fixed, ie. a point on a *circle*.

Topological defects

A phase is often not perfect: the order parameter field is not uniform everywhere. Often these are small fluctuations that can return easily to the perfect state, but sometimes this is not the case. We call *defect* a “tear” (some sort of singularity or discontinuity) in the order parameter field, and *topological defect* is a “tear which cannot be patched”. This means that no continuous local deformation of the order parameter field can remove the defect.

The simplest example is a vortex line superconductor at fixed T . If we consider a closed path encircling the line, the order parameter goes around the circle in the order parameter space. The lack of defect would correspond to the order parameter being constant as we follow the closed path. Since it is not possible to deform continuously the order parameter field from the initial state where following the closed path we go around the circle in the order parameter space to a final state where we don't move in the order parameter space, this is a topological defect. A rigorous treatment of this phenomena belongs to the *topology* branch of Mathematics. It follows that the vortex line cannot end (in reality it can only end at the boundary of the sample).

Another example is a crystal, where in half of the domain there is an extra layer of atoms. The edge of this layer is a line defect: encircling it corresponds to moving on a closed loop on the torus of the order parameter space, which cannot be contracted continuously to a single point.

Our last example is nematic liquid crystals at fixed T . The order parameter space is a hemisphere, where one can select a closed loop which cannot be contracted continuously to a single point. This means that nematic liquid crystals can have line defects. This in contrast with three-dimensional magnets, which do not have line defects: there the order parameter space is a full sphere: there any circle can be contracted to a point, just imagine to move them to the north pole.

Abrupt phase transitions

Phase transitions can be grouped into two classes: *abrupt phase transitions*, where the order parameter has a discontinuity, and *continuous*

phase transitions, where the order parameter has a singularity but still continuous.

Abrupt phase transitions are sometimes called first order, and the continuous one as second order. This terminology originates from Ehrenfest's classification, which considered the lowest derivative of the free energy that is discontinuous at the transition. Ehrenfest's approach is no longer used, as it turned out that different higher order phase transitions are not fundamentally different, some even do not fit in (eg. divergent derivatives).

First let us consider abrupt phase transition, and our example will be water held at fixed T and p . As illustrated on Fig. 4.1, if the parameters (T and p) are varied on a path which crosses a phase coexistence curve, a sudden jump in properties occur. This can be captured by the order parameter, which is in this case the density ρ .

What underlies an abrupt phase transition is that when the external parameters are varied, the current phase, which used to be stable (lowest free energy), becomes metastable, as another phase, which might be very far in configuration space, becomes lower in free energy. In the phase diagram (representing the lowest free energy phase as a function of parameters like T and p) one crosses from a domain A where the free energy of phase A is lower to domain B where that of phase B is lower.

When the coexistence curve (of equal free energies) is crossed, the system does not necessarily recognise immediately that some other state would be lower in free energy: this happens via fluctuations. Suppose that by fluctuations a small sphere of radius R of the lower free energy phase is formed. To fix notation let us consider cooling down a gas to form liquid at fixed pressure. Then the Gibbs free energy of the small liquid droplet is

$$G_{\text{droplet}}(R) = c_1 R^2 - c_2 R^3 \Delta T ,$$

see Fig. 4.2. The first term corresponds to surface tension: it costs free energy to create an interface between the two phases. This is positive and quadratic in R , dominating the expression for small R . The origin of the second term is the difference between the bulk free energies: this is proportional to the volume of the droplet ($\sim R^3$). The prefactor is proportional to ΔT (the difference between the coexistence temperature and the actual temperature), as the bulk free energies are smooth functions of the temperature, therefore their difference can be approximated to be linear in ΔT . This term is negative, and dominates at large R . The function $G_{\text{droplet}}(R)$ has a maximum, which is easy to obtain by differen-

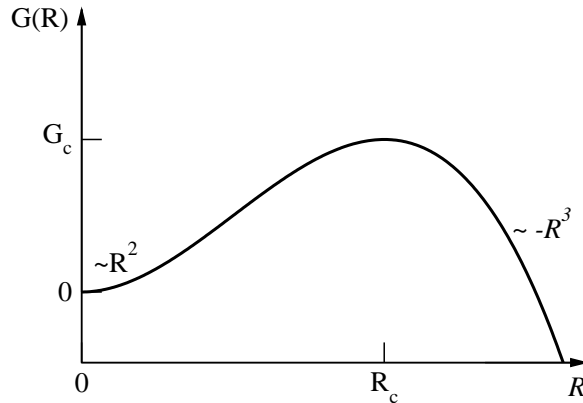


Figure 4.2 The Gibbs free energy of a droplet as a function of its radius.

tiation: the maximum can be denoted by G_c , which is taken at finite R_c . Due to the temperature dependence of the volume term, $G_c \sim \Delta T^{-2}$.

The free energy of a large droplet is negative, therefore this is the thermodynamically stable state. However, the system needs to discover that there is indeed a phase with lower free energy than the current one, which it does via fluctuations. In the simplest picture the system attempts to overcome the free energy barrier many times, and each attempt succeeds with probability $\exp(-\beta G_c)$. The rate by which the attempts are made (denoted by “prefactor” below) is determined by fundamental frequencies of the system, and is often very hard to predict, but the strong temperature dependence is contained mostly in the Boltzmann factor. So the rate by which the droplets form spontaneously, often called nucleation rate, is given by

$$\Gamma = (\text{prefactor}) e^{-\frac{G_c}{k_B T}} = (\text{prefactor}) e^{-\frac{c_3}{k_B T} \frac{1}{\Delta T^2}} .$$

This means *very* small rate for small ΔT , that is why a metastable phase has a macroscopic lifetime for small enough undercooling.

All the above relates to spontaneously forming droplets in free space, or “bulk nucleation”. In practice there might be other channels to form droplets, like on the surface of the container, around an impurity etc., which all have their individual barrier height; the smallest of these providing the dominant contribution to the overall nucleation rate.

Continuous phase transitions

When a system is brought through a path in parameter space which goes through a critical point [see path (2) in Fig. 4.1], the order parameter stays continuous, but at this point has a singularity: the derivative is infinite similarly to the square root function. In fact in this type of phase transition typically a large number of physical quantities have singularities, or in other words have non-trivial power law dependencies.

One such quantity is the parameter dependence (eg. temperature dependence) of the jump of the order parameter between the two phases near the critical point. For the liquid-gas transition of water, the order parameter can be approximated by

$$\frac{\rho}{\rho_c} = 1 + s \left(1 - \frac{T}{T_c}\right) \pm \rho_0 \left(1 - \frac{T}{T_c}\right)^\beta \quad (4.20)$$

Another example is the uniaxial magnet, where the order parameter, the magnetisation, has a nonzero value below the Curie temperature, which can be either positive or negative, and its absolute value scales as

$$M \propto \left(1 - \frac{T}{T_c}\right)^\beta .$$

Surprisingly the value of the exponent β for both the liquid-gas transition of water and the uniaxial magnet is the same, an often quoted value is $\beta = 0.325 \pm 0.005$. This is not coincidence: the scaling exponents (the exponents of the power laws for the singular quantities) are the same for a number of other physical quantities as well⁴.

The above examples illustrate the concept of *universality*: the behaviour near a continuous phase transition point is independent of the microscopic details. The various systems can be classified into a small number of groups. These groups are called *universality classes*: collections of systems with the same singular properties at the critical point. For example in thermal systems the scaling behaviour (the scaling exponent and the shape of the scaling function) of the susceptibility, specific heat, correlation length etc. are identical within a universality class, while other quantities like the parameters (eg. temperature) corresponding to the critical point, and also the prefactors are different. The uni-

⁴ In some sense the uniaxial magnet is simpler, as the phase coexistence curve falls on the $T = 0$ line. For the liquid-gas transition of water the coexistence curve is parallel to neither the T nor the p axis in the phase diagram, giving rise to the linear term in (4.20). The singular term, which is the important one, however, is the same.

versality class typically depends only on fundamental properties, like the symmetry of the order parameter, and the dimensionality.

As we have seen, the liquid-gas transition of water (and other materials!) and uniaxial magnets are in the same universality class. This is quite surprising at first sight, but on closer inspection one can see that the order parameter is of the same type (scalar), the dimensionality (3 dimensions) is the same, and in fact it is possible to find a mapping between the two systems.

Universality is a very powerful concept, since one member of a universality class fully represents all other members. Thus to learn everything about a particular universality class it is enough to find a simple enough member which is tractable theoretically or numerically. This is the reason why physicists often study very simple “toy” models, which genuinely represent the more complex members of their universality class in all important aspects.

Fluctuations, which were exponentially suppressed in abrupt phase transitions, are very prevalent in continuous phase transitions. The characteristic spatial size of the fluctuations (the correlation length) diverges at the critical point.

4.6 Surface growth⁵

A large class of processes in nature can be described as growth processes, where one phase grows and invades a region previously occupied by another phase. These processes are often not in equilibrium (in fact far from equilibrium), where a persistent driving force brings the process forward, resulting in a propagating rough front. The growth front is not progressing uniformly, it is not smooth, because of some noise or inherent disorder in the system.

To give some concrete examples, we can think of the progress of the wetting of a paper or table cloth, or the burning front of a slowly burning paper. An example from the nanosciences is molecular beam epitaxy, where atoms are deposited in vacuum onto the surface of a crystal, which due to the shot noise of the deposited atoms creates a rough surface after a number of deposited layers. The growth of bacterial colonies can also be in this class of processes, especially when the growth substrate is relatively dry so the bacteria don’t swim, and the nutrients are abundant.

⁵ A detailed treatment of the statistical mechanics aspects of surface growth can be found in [3].

(It must be mentioned that the bacterial colony growth is much more complex phenomena, including sometimes even genetic shifts. And even in more simple cases a different physical phenomenon can dominate, like in the scarce nutrients regime, where the constant driving is replaced by growth dominated by the diffusion of nutrients. In the latter case instead of the rough growth front a fine branching fractal pattern emerges.)

As usual, we only would like to know aggregate information about the process, like how “non-smooth” the front is. We consider the surface growth problem on an initially flat substrate; denote the coordinates along the substrate as x , and the distance of the interface from the substrate is h (height). We can define the average height as

$$\bar{h}(t) = \langle h(x, t) \rangle_x, \quad \bar{h}(t) \sim t$$

which typically grows linearly in time. The *roughness* or *width* of the surface is defined as the root-mean-square deviation of h from \bar{h} :

$$w(L, t) = \sqrt{\langle (h(x, t) - \bar{h}(t))^2 \rangle_x}$$

where L is the linear size (length) of the substrate.

In typical surface growth processes the width initially grows as a power of time, then at some crossover time (which depends on L) it saturates:

$$\begin{array}{lll} \text{early times, } t \ll t_\times: & w(L, t) \sim t^\beta & \beta: \text{ growth exponent} \\ \text{late times, } t \gg t_\times: & w(L, t) \sim w_{\text{sat}}(L) \sim L^\alpha & \alpha: \text{ roughness exponent} \\ \text{crossover time:} & t_\times \sim L^z & z: \text{ dynamic exponent} \end{array}$$

When plotting width against time, we can achieve data collapse if the width is rescaled by the saturation width, and time by the crossover time. This way the surface width can be expressed by a single-argument scaling function:

$$w(L, t) \sim L^\alpha f\left(\frac{t}{L^z}\right) \quad f(u) \sim \begin{cases} u^\beta, & \text{if } u \ll 1 \\ \text{const}, & \text{if } u \gg 1 \end{cases}$$

which is called the Family-Vicsek scaling relation.

Evaluating w at the crossover time where approximately both the early and the late time behaviour holds, we can obtain a relation between the exponents:

$$\left. \begin{array}{l} w(L, t_\times) \sim L^\alpha \\ \sim t_\times^\beta \sim L^{z\beta} \end{array} \right\} \quad z = \frac{\alpha}{\beta}$$

The surface generated by the growth process has no intermediate characteristic length scales, thus displays some kind of self-similarity. In this case the x and the h directions are not equivalent, so different magnification in the x and h direction yields statistically similar objects:

$$h(x) \quad \text{similar to} \quad b^\alpha h\left(\frac{x}{b}\right)$$

Such functions are called self-similar with self-similarity exponent α . One such familiar function is the graph of random walk (note here t plays the role of x):

$$\langle (y(t + \Delta t) - y(t))^2 \rangle_t \sim \Delta t \quad \Rightarrow \quad \alpha = \frac{1}{2}$$

These surface growth processes can be illustrated by simple models, for example:

- random deposition: unit square blocks are released above integer positions of a substrate, and they just land on top of previously dropped blocks. For this model α is undefined, and $\beta = 1/2$.
- random deposition with surface relaxation: as for random deposition, but the blocks are allowed to jump to the nearest neighbour substrate position to achieve lowest position.
- restricted solid-on-solid (RSOS) model: only those growth events are allowed which keep local slope bounded: maintain $|h(x) - h(x+1)| \leq 1$

Another approach to understand the surface growth processes is to consider continuum equations:

$$\frac{\partial h}{\partial t} = G[h(\cdot), x, t] + \eta$$

where the noise term is often unbiased and delta-correlated:

$$\text{eg. } \eta(x, t) : \quad \langle \eta(x, t) \rangle = 0, \quad \langle \eta(x, t) \eta(x', t') \rangle = 2D \delta(x - x') \delta(t - t')$$

These equations have the following desired symmetries:

$$t \rightarrow t + \Delta t$$

$$h \rightarrow h + \Delta h$$

$$x \rightarrow x + \Delta x$$

$$x \rightarrow -x, \text{ or rotation}$$

$$h \rightarrow -h \quad (\text{in certain cases, eg. if in equilibrium})$$

The simplest such equation is the Edwards-Wilkinson equation:

$$\frac{\partial h}{\partial t} = \nabla^2 h + \eta(x, t) \quad \alpha = 1 - \frac{d}{2}, \quad \beta = \frac{1}{2} - \frac{d}{4}, \quad z = 2$$

The simplest one breaking the $h \rightarrow -h$ symmetry is the Kardar-Parisi-Zhang (KPZ) equation:

$$\frac{\partial h}{\partial t} = \nabla^2 h + \frac{\lambda}{2} (\nabla h)^2 + \eta(x, t) \quad \text{for } d = 1 : \quad \alpha = \frac{1}{2}, \quad \beta = \frac{1}{3}, \quad z = \frac{3}{2}$$

which describes eg. the RSOS model. The continuum equations and discrete models can be classified into universality classes (as we have seen in continuous phase transitions), where the scaling exponents are the same across a class.

4.7 Collective biological motion: Flocking

In this lecture we will consider a model, which is a first step towards understanding the motion of groups of animals: for example schools of fish, herds of quadruples, flocks of birds (of which flocks of starlings is particularly impressive), as well as cooperative motion in bacterial colonies.

The model consists of a collection of self-propelled particles (representing e.g. the individual birds), which move in continuous space with constant speed but varying direction[4]. This arrangement is quite unlike typical physical systems made of passive particles, as momentum and energy is not conserved; but much more typical in biological systems where the individuals are more advanced: active, having own energy supply.

The particles move for time Δt in a ballistic motion, then select a new direction for their velocity. For simplicity we restrict the particles in two dimensions, where the direction is given by an angle θ . The update rules of the i th particle's position, velocity and direction is

$$\begin{aligned} \mathbf{x}_i(t + \Delta t) &= \mathbf{x}_i(t) + \mathbf{v}_i(t) \Delta t \\ \mathbf{v}_i(t) &= v(\cos \theta_i(t) \hat{x} + \sin \theta_i(t) \hat{y}) \\ \theta_i(t + \Delta t) &= \langle \theta_j(t) \rangle_{j: |\mathbf{x}_j - \mathbf{x}_i| < R} + \underbrace{\Delta \theta}_{\text{uniform in } [\eta/2, \eta/2]} \end{aligned}$$

So at the end of each Δt timestep each particle takes the average direction of all particles within radius R , and adds a noise of controlled amplitude.

What are the relevant parameters of this model? For any model we should take care to reduce the number of parameters as much as possible (but no further!) in order not to be lost in a high dimensional parameter space.

Without loss of generality we can set $\Delta t = 1$ and $R = 1$, i.e. these will be our units of time and space. For the speed v of the particles we already fixed the units, so need to consider its actual value. The $v \rightarrow 0$ limit would mean stationary objects (which would correspond to the XY model of magnets), while $v \rightarrow \infty$ limit would mean that each timestep a given particle meets completely new ones, corresponding to complete mixing or mean-field models. For our flocking model in the relevant, intermediate range, say $0.003 < v < 0.3$, the actual value of v turns out not to affect the behaviour. So as long as the value of v is moderate, its value does not matter: it is not a relevant parameter either. There is one more symbol above, the amplitude of the noise, $0 < \eta < 2\pi$, which is indeed one of the relevant parameters. The second relevant parameter is a bit hidden: it is the average density of particles, ρ . Finally we have the system size: either the number of particles N or the side L of the 2D box (the two are related by $\rho = N/L^2$), which does matter but we will strive to see everything in the thermodynamic limit $N \rightarrow \infty$.

We see that for some values of the parameters the particles align the direction of their velocities globally, creating an ordered phase, while for other values there is no global order either because each particle moves randomly or there are small coherent groups but these are uncorrelated with each other. To quantify this order, we define the order parameter as the average velocity over the whole system normalised by v , or to follow [4] we take the absolute value:

$$\Phi = \frac{1}{Nv} \left| \sum_i \mathbf{v}_i \right|$$

As the parameters are varied, in the $N \rightarrow \infty$ limit Φ changes from nonzero to zero value. On the ordered (nonzero) side we see power law scaling:

$$\begin{aligned} \text{fixed } \rho: \quad \Phi &\sim (\eta_c(\rho) - \eta)^\beta \\ \text{fixed } \eta: \quad \Phi &\sim (\rho - \rho_c(\eta))^{\beta'}, \end{aligned}$$

where numerically it was found that $\beta \approx 0.45$ and $\beta' \approx 0.35$ [4].

This system is not an equilibrium system, the phase transition is purely due to kinetic effects; this is an example of kinetic phase transitions.

It turns out that like for many other continuous phase transitions, the order parameter, which is now function of two parameters $\Phi(\eta, \rho)$, can

in fact be written as a scaling function of as single variable:

$$\Phi(\eta, \rho) = \tilde{\Phi}\left(\frac{\eta}{\eta_c(\rho)}\right), \quad \tilde{\Phi}(u) \sim \begin{cases} (1-u)^\beta, & \text{if } u < 1 \\ 0 & \text{if } u > 1 \end{cases}$$

With the help of the scaling function we can now look at the scaling exponents β and β' . For fixed η and varying ρ , we defined β' as $\Phi \sim (\rho - \rho_c(\eta))^{\beta'}$. Denoting $\epsilon = \rho - \rho_c(\eta)$:

$$\begin{aligned} \Phi(\eta, \rho_c(\eta) + \epsilon) &= \tilde{\Phi}\left(\frac{\eta}{\eta_c(\rho_c(\eta) + \epsilon)}\right) \approx \tilde{\Phi}\left(\frac{\eta}{\eta + \epsilon \left.\frac{d\eta_c}{d\rho}\right|_{\rho=\rho_c(\eta)}}\right) \\ &\approx \tilde{\Phi}\left(1 - \frac{\epsilon}{\eta} \frac{d\eta_c}{d\rho}\right) \sim \epsilon^\beta \end{aligned} \quad (4.21)$$

so this shows [5] that $\beta = \beta'$, despite the original different numerical estimates.

There is recent ongoing research (as of 2010-2011) on the dynamics of flocks of starlings in Italy. The current understanding is that while the above model is a good starting point, real birds don't really consider a fixed radius neighbourhood, but instead watch a fixed number of neighbours and keep them oriented by trying to maintain relative angles between them fixed.

References

- [1] E. T. Jaynes, *Probability theory: the logic of science*. Cambridge University Press (2003).
- [2] J. P. Sethna, *Statistical mechanics: entropy, order parameters, and complexity*. Oxford University Press (2006).
- [3] A.-L. Barabási, H. E. Stanley, *Fractal concepts in surface growth*. Cambridge University Press (1995).
- [4] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, O. Shocket; *Phys. Rev. Lett.* **75**, 1226 (1995)
- [5] A. Czirók, H. E. Stanley, T. Vicsek; *J. Phys. A: Math. Gen.* **30**, 1375 (1997)