PHYS11024: Statistical Physics 2025–2026

John Peacock

Room C20, Royal Observatory; jap@roe.ac.uk

http://www.roe.ac.uk/japwww/teaching/statphys.html

Synopsis

This is a course on statistical physics, which aims to understand the bulk properties of physical systems in terms of the dynamics of their microscopic constituents. In part (1) we review the fundamental assumptions of statistical mechanics, focusing on the definition of entropy and its relation to our information about a system. In part (2) we consider time dependent kinetic theory, attempting to understand how entropy can display a monotonic arrow of time in the presence of time-reversible microphysics. In part (3) we consider the statistical mechanics of systems of interacting particles and develop important approximation schemes for their equation of state. In part (4) we examine phase transitions and their unifying phenomenology. We study in detail a simple microscopic model: the Ising model of ferromagnetism. We then consider a general theoretical framework for phases transitions known as Landau Theory.

Textbooks

These notes are intended to be self-contained, but there are many excellent textbooks on the subject. The following are especially recommended for background reading:

- Waldram, J.R. (CUP) *The theory of thermodynamics*. Very clear on the fundamentals of the Gibbs approach and the arrow of time.

- Huang, K. (Wiley) *Statistical Mechanics* A graduate-level text replete with detail, though perhaps at too high a level for some.

- Huang, K. (CRC Press) *Introduction to Statistical Physics, 2nd Edition* A simplified version of Huang's graduate text, aimed at undergraduates, and with a gain in clarity. Avoid the first edition, which suffered from misprints.

- Kennett, M. (CUP) *Essential Statistical Physics*. A good match to the course material, though missing Debye–Hückel theory.

- Chandler, D. (OUP) *Introduction to Modern Statistical Mechanics*. Close to the course in level and content, but aimed at Physical Chemists so has some differences in notation.

- Pathria, R.K. & Beale, P.D. (Academic Press) *Statistical Mechanics*. Graduate-level textbook useful for background.

These notes are based in part on material from Martin Evans, who delivered the course prior to 2022/23. There will be some changes with respect to previous years, but most past exam questions should still be relevant. Note however that the summer exam moved to a format of all compulsory questions in 2023.

Contents

1 Overview

Statistical physics literally means the physics of random processes, where we deal with physical properties of systems that cannot be predicted exactly. Quantum mechanics is the prime example of intrinsic randomness in physics, but the term usually refers to attempts to understand the behaviour of **macroscopic** systems in terms of a finite number of **microscopic** constituents. Thus we seek to understand the experimental properties of matter in terms of processes occurring at the atomic level. Many aspects of this problem can be treated purely classically, even though atoms are quantum objects. The techniques of this field are very general, and can be applied equally well to broader examples such as traffic flow or economics.

The prime concern of this course will naturally be with thermodynamics, since our microscopic understanding of heat is as the random motion of atoms or molecules. We can usefully distinguish two aspects of the subject:

(1) **Statistical mechanics**, which means attempting to understand **classical thermodynamics**, meaning the properties of matter in **thermal equilibrium**.

(2) **Kinetic theory**, meaning the transition between different states, extending classical thermodynamics to the non-equilibrium case.

The principal challenge in statistical mechanics is **entropy**. Compared to more straightforward macroscopic properties of matter, such as total energy, the question of how to compute the entropy from microscopic fundamentals turns out to be difficult and subtle. These difficulties are tied up with the **arrow of time**. Classical thermodynamics presents us with the **second law**, in which entropy for an isolated system can only increase with time. This happens despite the underlying microphysics being invariant under time reversal, so that a video of any physical process that is run backwards should represent another valid process.

This paradox emerged through the work of the Austrian physicist Ludwig Boltzmann (1844-1906), so surely any difficult questions must have been settled long ago? But remarkably, there continues to be a vigorous debate over the arrow of time even today. So be warned: different textbooks approach the subject in different and even inconsistent ways. These lectures will attempt to expose the issues and suggest a resolution; but you will need to be prepared to make up your own mind. Achieving clarity here is not aided by the historical baggage that comes with the subject, which is well-equipped with unhelpful terminology and jargon.

This situation is reminiscent of quantum mechanics, where there is still much discussion about the process of measurement and the origin of randomness; we will touch on some of these issues. But the quantum world has also set out a helpful philosophy: **shut up and calculate** (David Mermin, 1989). This applies equally well to statistical physics, where worrying about deep philosophical issues generates a powerful set of tools, which can be used for important calculations. The latter part of the course therefore concentrates on applications of practical interest, particularly the thermodynamics of imperfect gases and the important general area of phase transitions.

1.1 Elements of classical thermodynamics

Equilibrium and reversibility   We start by revising some of the key elements of classical thermodynamics, as set out in JH *Thermal Physics*. This subject discusses systems in **thermal equilibrium**, characterised by a **temperature**, T. The meaning of this number is that two systems of the same T will be in equilibrium with each other – i.e. no heat will tend to flow if they are placed in thermal contact (the **zeroth law** of thermodynamics). But many physical

processes of interest do not involve equilibrium, so how do we apply thermodynamics to such cases? In practice, we can set up non-equilibrium systems that are inhomogeneous mixtures of different equilibrium systems, e.g. a box of gas where the left and right halves are set at different temperatures. As we know, such a mixture will settle down to a single temperature, and the challenge is to follow what happens during this transition. A key tool in doing so is the concept of a **reversible change**, in which the system and its surrounding can be brought back to their starting point. It is usually convenient to imagine such changes proceeding through a set of intermediate equilibrium stages, i.e. a set of infinitesimal changes. So in our example of a two-temperature gas, we will let a small amount of heat pass slowly from one side of the box to the other, so that each side has plenty of time to reach equilibrium at a new temperature before any more heat flows. This is not to say that all infinitesimal changes will be reversible, but the concept of a infinitesimal reversible change is a key element of thermodynamics.

Functions of state   Beyond the temperature, a physical system will be characterised by other bulk properties, termed **thermodynamic coordinates**: e.g. internal energy, pressure and volume for a gas. An important concept is that these coordinates are **functions of state**, so that their values reflect how the system *is*, rather than history of how it reached that point. This is almost a circular definition, since what we mean by the state is in practice a vector in thermodynamic **state space**, where we specify the values of the coordinates, or **state variables**. Furthermore, the specification of the state seems to be ambiguous, since not all the coordinates are independent. In practice, this means that there is an **equation of state** that relates them (e.g. the **ideal gas law**, $PV = NkT$).

The power of this approach is shown when we consider the **First law** of thermodynamics:

$$dE = \dod Q + \dod W, \tag{1.1}$$

where we write $\dod Q$ and $\dod W$ for the heat flow into and work done on the system rather than dQ and dW: the slashes emphasise that these are not functions of state and so depend on thermodynamic path. For **reversible changes** only, we define the change in **entropy** as

$$\boxed{\dod Q = T\, dS} \tag{1.2}$$

and the work done on the system is $\dod W = -P\, dV$. But then we have the first law in a form that involves only functions of state, so it must be generally true – even when $T\, dS$ no longer arises from a reversible heat flow:

$$\boxed{dE = T\, dS - P\, dV.} \tag{1.3}$$

This relation must reflect the differentiation of a function $E(S, V)$, so that we can identify terms via the chain rule:

$$T = \left.\frac{\partial E}{\partial S}\right|_V; \quad P = -\left.\frac{\partial E}{\partial V}\right|_S. \tag{1.4}$$

But then we can use the symmetry of 2nd partial derivatives to deduce non-trivial **Maxwell relations** between the thermodynamic coordinates, e.g.

$$\left.\frac{\partial T}{\partial V}\right|_S = -\left.\frac{\partial P}{\partial S}\right|_V. \tag{1.5}$$

More generally we can write the first law as

$$dE = T\, dS + \sum_\gamma f_\gamma\, dX_\gamma, \tag{1.6}$$

where f_γ is an applied force and X_γ is a thermodynamic variable (sometimes referred to as a displacement) **conjugate** to that force. In a magnetic system the force would be the external

magnetic field $\underline{B}$ and the displacement would be the magnetisation $\underline{M}$. Note that the forces are **intensive** (i.e. don't depend on the size of the system) but the displacements X_γ are **extensive** (i.e. scale linearly with the size of the system).

The second law   This fundamental fact can be stated in a number of ways. One is to say that

$$\boxed{T\,dS \geq đQ,}\tag{1.7}$$

with equality in the case of reversible changes. Where a change is irreversible, in effect a **frictional** process is operating, and in fact we can write the second law as an equality: $T\,dS = đQ + đW_f$, where the latter term represents the work done by friction. So if we have a box of gas that includes a block of metal sliding with a nonzero velocity on the rough base of the box, this will generate heat in slowing down. The gas will then increase its entropy, exactly as if the block had been absent but that heat equivalent to its kinetic energy had been transferred reversibly.

One normally becomes convinced about the second law via **Clausius's inequality** (Rudolf Clausius 1822-1888), which is based on a **cyclic system** that exchanges heat with an external body at temperature T:

$$\oint \frac{đQ}{T} \leq 0,\tag{1.8}$$

where $đQ$ is the heat given to the system. This is proved in the following stages. First we consider an ideal reversible heat engine that takes in heat Q_1 and produces heat Q_2, yielding work $Q_1 - Q_2$. By considering a succession of such engines, it is possible to argue that $Q_1/Q_2 = T_1/T_2$, in terms of the temperatures of the input and output. Any irreversible engine will be less efficient (work is lost in friction), so it will have a lower value of Q_1/Q_2, so $Q_2/T_2 \geq Q_1/T_1$.

Availability and free energy   We can think more generally about the work that may be extracted from a system. Suppose the system has internal properties P, T etc., but that it is immersed in an enclosure with in general different properties P_0, T_0. The surroundings matter: if $P_0 = P$ and $T_0 = T$, then no work can be extracted since we need a pressure difference to drive a piston and a temperature difference to drive a heat engine. Consider the first law in the following slightly modified form:

$$đW = -dE + đQ,\tag{1.9}$$

where $đW$ is the work done *by* the system (the opposite of the usual convention) and $đQ$ is the heat supplied to the system. But $đW$ will not be useful if it is all expended in pushing back the surroundings. Therefore, we must subtract this environmental term to obtain $đW_a$, the **available work**:

$$đW_a = -dE + đQ - P_0 dV,\tag{1.10}$$

If heat is supplied from the surroundings in a manner that is reversible as far as the surroundings are concerned, then $đQ = -T_0 dS_0$. But by the second law, the total entropy of the universe can only increase, so $dS + dS_0 \geq 0$, implying

$$đW_a \leq -dE + T_0 dS - P_0 dV.\tag{1.11}$$

We can express this neatly by defining the **Availability**, A:

$$\boxed{đW_a \leq -dA; \quad A \equiv E - T_0 S + P_0 V.}\tag{1.12}$$

Now consider the important case where a system is allowed to come into equilibrium with its surroundings, and where we extract no work. If $đW_a = 0$, then $dA \leq 0$, and the system will evolve so as to minimise the availability. As equilibrium is approached, A approaches the **Gibbs Free Energy**:

$$G \equiv E - TS + PV,\tag{1.13}$$

which is therefore minimised in equilibrium. The derivation through the availability justifies the 'free energy' terminology. If we consider a system of fixed volume, which is often a more realistic case, the environmental work term $-P_0 dV$ in equation (1.11) is absent; thus we would define the availability as just $A \equiv E - T_0 S$, so that the quantity to be minimised in this case is the **Helmholtz Free Energy**:

$$\boxed{F \equiv E - TS.} \tag{1.14}$$

This represents a trade-off between minimising the internal energy and maximising the entropy (internal disorder) of a system. The free energy is a key quantity that will feature frequently throughout this course.

The free energy has the virtue that it changes the **natural variables** of the first law from S and V to T and V:

$$\boxed{dF = -S\,dT - P\,dV.} \tag{1.15}$$

This shows directly and simply that F is stationary for changes at constant temperature and volume (and in the same way that G is stationary at constant temperature and pressure). This stationarity can be stated more powerfully in words:

$$\boxed{\text{A system in thermal equilibrium minimises its Free Energy.}} \tag{1.16}$$

The proof of this via the first law is more transparent than the more detailed argument via Availability, which is non-examinable. The first law also shows that the free energy gives a practical means of computing the entropy and pressure:

$$\Rightarrow S = -\left.\frac{\partial F}{\partial T}\right|_V \qquad P = -\left.\frac{\partial F}{\partial V}\right|_T. \tag{1.17}$$

We will take advantage of these relations in what follows.

The chemical potential   All the above applies for **closed** systems that can exchange only heat with their surroundings. But one can generalise further to include the effect on the internal energy of changing the number of particles or indeed the number of each species of particle:

$$\boxed{dE = T\,dS + \sum_\gamma f_\gamma dX_\gamma + \sum_\alpha \mu_\alpha\,dN_\alpha,} \tag{1.18}$$

which defines the **chemical potential** μ_α as the change in internal energy by adding a particle of species α to the system.

Thermodynamic potentials   Using the above ingredients, we can generate useful thermodynamic quantities that are natural functions of different sets of variables;

$$
\begin{array}{llll}
(T, V, N) & F = E - TS & & \text{Helmholtz Free Energy} \\
(S, P, N) & H = E - (-PV) & = E + PV & \text{Enthalpy} \\
(T, P, N) & G = F - (-PV) & = E - TS + PV & \text{Gibbs Free Energy} \\
(T, V, \mu) & \Phi = F - N\mu & = E - TS - \mu N & \text{Grand Potential}
\end{array} \tag{1.19}
$$

All these thermodynamic potentials provide different minimisation principles, based on the first law. For example, $dG = -S\,dT + V\,dP + \mu\,dN$, so we can see that this quantity will be stationary under changes at constant pressure (and T & N), in the same way as F is minimised for changes at constant volume.

Gibbs-Duhem relation   With all these different thermodynamic coordinates in play, we can ask whether they are independent beyond the constraints set by the first law and any equation of

9

state. The answer is no. This can be demonstrated by focusing on variable that are **extensive**, i.e. proportional to the size of the system. For example, the energy and all its natural variables are extensive. Increasing the size of the system by a factor b therefore gives

$$bE(S, \{X_\gamma\}, \{N_\alpha\}) = E(bS, \{bX_\gamma\}, \{bN_\alpha\}). \tag{1.20}$$

Now differentiate both sides with respect to b and then set $b = 1$:

$$E(S, \{X_\gamma\}, \{N_\alpha\}) = \left. \frac{d}{db} E(bS, \{bX_\gamma\}, \{bN_\alpha\}) \right|_{b=1} \tag{1.21}$$

$$= \frac{\partial E}{\partial S} S + \ldots \tag{1.22}$$

$$= TS + \sum_\gamma f_\gamma X_\gamma + \sum_\alpha \mu_\alpha N_\alpha, \tag{1.23}$$

where we have used (1.18) to identify the partial derivatives. Differentiating this relation gives

$$dE = T\,dS + S\,dT + \sum_\gamma [f_\gamma dX_\gamma + X_\gamma df_\gamma] + \sum_\alpha [\mu_\alpha dN_\alpha + N_\alpha d\mu_\alpha], \tag{1.24}$$

and subtracting the first law yields

$$\boxed{0 = S\,dT + \sum_\gamma X_\gamma df_\gamma + \sum_\alpha N_\alpha d\mu_\alpha.} \tag{1.25}$$

This is known as the **Gibbs–Duhem relation** and implies that the intensive variables $T, \{f_\gamma\}$ and $\{\mu_\alpha\}$ are not all independent. For example, in a PVT system with one species, only two of T, P, μ are independent since the Gibbs-Duhem relation becomes

$$0 = S\,dT - V\,dP + N\,d\mu, \tag{1.26}$$

i.e. a change in T, P implies a specified change in μ (see tutorial sheet).

Mathematical aside (non-examinable)   What we have done here is to define another thermodynamic function through what is known as a **Legendre transform**. More generally consider the chain rule applied to a function $f(\{x_i\})$, $i = 1, \ldots, k$:

$$df = \sum_{i=1}^{k} \frac{\partial f}{\partial x_i} dx_i = \sum_{i=1}^{k} u_i\,dx_i. \tag{1.27}$$

Then we can let $g = f - \sum_{i=r+1}^{k} u_i x_i$ and

$$dg = \sum_{i=1}^{r} u_i dx_i - \sum_{i=r+1}^{k} x_i du_i. \tag{1.28}$$

g is a **natural function** of $x_1 \ldots x_r, u_{r+1} \ldots u_k$ and is the Legendre Transform of f. The idea is that the function g should contain the same information as f. As a simple example, consider a function of one variable, $f(x)$, where the transform is $g(u) = f(x) - ux$. Geometrically, g is the intercept of a tangent line on the y axis, as a function of the slope. So our assertion is that the slope as a function of intercept is enough to reconstruct the curve $f(x)$. A more general example of the same philosophy of changing variables from coordinate-like quantities to velocity-like quantities is in dynamics. Here, we know that the equations of motion for a set of coordinates q can be derived either from the Lagrangian $L(q, \dot{q})$, or the Hamiltonian $H(q, p)$.

10

2 Microstates and entropy

The challenge of statistical mechanics is to compute these thermodynamic state functions from first principles. We are dealing with quantities that are experimentally measurable properties of macroscopic materials, and so the thermodynamic state under study is a **macrostate**. But we know that there must be much underlying complexity that is hidden in classical thermodynamics. For example, in an ideal gas the internal energy, E, is the sum of the kinetic energies of the N particles that make up the gas. The exact description of the the system is via its **microstate**, which for a classical system means specifying the positions and velocities of all the constituent particles.

Boltzmann and entropy   But if the origin of the macroscopic energy is clear enough, the microscopic origin of entropy is much less so. Boltzmann's great insight was to realise that there would be a large number, Ω, of different microstates that correspond to a given macrostate, specified by a few global numbers such as total energy, E. Boltzmann further assumed that a system will be equally likely to inhabit any of the possible microstates – a critical assumption, which we will examine in detail. If this is true, then the relative probability of being found in two different macrostates (with different entropies, in particular), will be in the ratios of their **statistical weight** or **multiplicity**, Ω:

$$p(S_1)/p(S_2) = \Omega_1/\Omega_2. \tag{2.1}$$

The second law will then arise **statistically** if Ω is larger for macrostates of larger entropy – i.e. S will probably increase if the system is not placed in an initial state of maximum entropy, although it is apparently not guaranteed to do so.

To make this idea definite, we need to be able to calculate Ω for a given macrostate; but it is not immediately obvious how to go about counting the relevant microstates in general, given just the macroscopic thermodynamic variables. Boltzmann suggested that this dilemma could be solved by proceeding in the opposite direction and *defining* entropy in terms of the statistical weight:

$$\boxed{S = k \ln \Omega,} \tag{2.2}$$

a formula that appears on his gravestone. k is **Boltzmann's constant**, with SI value 1.3806×10^{-23}; we will sometimes write it as k_B, whenever there is a risk of confusion with wavenumber.
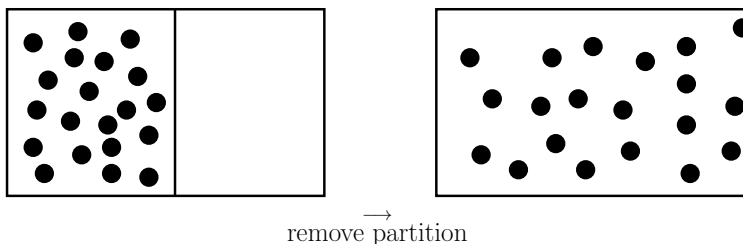
$\xrightarrow{\text{remove partition}}$

Figure 1: Joule expansion: a gas is prepared in an initial macrostate with all the particles in the left hand region of the container. Then the partition is removed. As time goes by the particles become evenly spread out through the container. Macroscopically, this process is irreversible and there is an associated increase of entropy.

We should be clear about the need to prove that this microscopic formula is consistent with the macroscopic **experimental entropy**. One immediately attractive property of Boltzmann's definition is that it is **extensive** (proportional to amount of material). If we make a composite system by adding together independent subsystems, the entropies should add – which they do in this case, because the Ω's multiply. As a more specific example, Boltzmann's formula is often

motivated with the example of free expansion of an ideal gas released from one half of a box by removing a partition (**Joule expansion**). Because the gas is ideal, the expansion does not change the velocity of the particles, so E and T are unchanged in this process. Since $dE = TdS - PdV$, $dE = 0$, and $P = kTN/V$ with T constant, $\Delta S = Nk\Delta \ln V = Nk \ln 2$ in this case. If the microstate is simplified to a choice of L or R, $\Omega = N!/([N/2]!)^2$, which via **Stirling's approximation** gives the Boltzmann ΔS for large N: $\ln \Omega \simeq N \ln N - 2(N/2) \ln(N/2) = N \ln 2$ (at least, for distinguishable particles; we discuss below how to deal with identical particles, although this does not affect the result for ΔS in this case).

A general check on this approach will be performed in detail in section 4, where we show that the classical and statistical entropies are the same for large systems in thermal equilibrium. But for systems of a few particles, the statistical approach and the classical one are not so easily reconciled.

Stirling's approximation   The approximate expression for $N!$ will be needed often; for interest, it arises as follows. $\ln(N!) = \sum_{i=1}^{N} \ln(i)$, and it is easy to see from a sketch that this histogram is close to $\int_1^N \ln x \, dx = N \ln N - N + 1$. An even more accurate approximation is $\ln(N!) \simeq N \ln N - N + \ln(\sqrt{2\pi N})$; but for large enough N, the simple form $\ln(N!) \simeq N \ln N$ is adequate. For $N = 100$, it is too high by a factor 1.27; but for macroscopic applications $N \sim 10^{24}$, and then the error in adopting $\ln(N!) \simeq N \ln N$ is at the % level. The detailed approximation is easily derived by writing $N! = \int_0^\infty x^N \exp[-x] \, dx$ and approximating the integrand as a Gaussian peak around its maximum at $x = N$.

Boltzmann and the 2nd law   The most important aspect of Boltzmann's approach is not simply that it provides a means of computing the macroscopic entropy in terms of microphysics, but that it offers a suggestion for the origin of the 2nd law. Consider Joule expansion, where we know that the gas will expand to fill the whole box and never retreat to the original configuration where all material is on the LHS. There are clearly some microstates (position and velocity of all particles) that are consistent with this macrostate – but there are many more microstates in which the particles are equally distributed between both sides. As we saw above, the ratio of the number of microstates that correspond to the latter configuration is greater than for the initial state by roughly a factor of 2^N. Therefore, if we accept Boltzmann's hypothesis that the probability of a given macrostate is proportional to the number of corresponding macrostates, we would place odds of $2^N : 1$ against the gas being found on the LHS, rather than being spread throughout the box. N does not to be particularly large before this calculation amounts to complete certainty: even for $N = 30$ the odds are a billion to 1, and for a truly macroscopic Avogadro-scale sample with $N \sim 10^{24}$, the colossal magnitude of 2^N is hard to grasp. In this view, then, the 2nd law is statistical: just a result of the balance of probabilities favouring outcomes with a larger value of Ω, and hence of larger entropy. A system that is not in the most probable macrostate will be expected to evolve towards that state, increasing entropy as it goes, until it attains the most probable state of maximum entropy.

Randomisation   Although Boltzmann's argument was and remains a huge breakthrough, it becomes more problematic when considered in detail. Implicitly, the argument that probabilities scale in proportion to Ω rests on the idea that there is some process that will scramble the microstates among the permissible values given macrostate constraints. This is readily seen to be reasonable when a system can interact with a second one. Heat is transferred to the first system (with $dS = dQ/T$ if the transfer is performed reversibly), and our picture of heat is indeed one of randomisation: the microstate will change in some way, over which we lack control or knowledge. The case of an **isolated** system is however more difficult – as we will discuss in detail below. Dynamics can accomplish some of what Boltzmann needs, since the microstate can change with time depending on how it is specified: for example, a particle with a given position and velocity will change its position with time, and eventually will change its velocity if it undergoes collisions. But

as we will see, it not obvious that causal dynamics alone can have the same effect as randomisation of the microstate. Nevertheless, it is common to assume that the process by which microstates change satisfies two general properties. Firstly there is the **ergodic hypothesis**, which states that the system will evolve so that eventually it explores all the microstates that are **accessible** (allowed by macroscopic constraints – e.g. fixed total energy in the case of an isolated system). A stronger assumption is that all accessible microstates are occupied with equal probability (the pompous-sounding **principle of equal a priori probability − PEAPP**). Both of these propositions do indeed apply in some circumstances, but they are insufficiently general. PEAPP might seem reasonable as a description of the system averaged over a long period of time, but what about evolution away from a given initial state? For example, transfer dQ to one end of a box of gas, rather than distributing the heat homogeneously, and you will populate only a subset of the microstates, at least initially. Furthermore, PEAPP certainly does not apply for systems in full thermal equilibrium, where we will see that there are extreme microstates that have a non-zero but small probability of being occupied via thermal fluctuations.

2.1 Gibbs entropy

We therefore need to consider the general situation in which microstates are occupied with different probabilities. In this case, Boltzmann's expression for the entropy generalises to the **Gibbs entropy** (which, following Stigler's law, was first written down by Boltzmann). The Gibbs entropy is derived below in section 3; for now, we simply quote the expression and consider some of its implications:

$$S = -k \sum_i p_i \ln p_i, \tag{2.3}$$

where p_i is the probability that the system is in the ith microstate. This clearly yields the original Boltzmann expression under PEAPP, when $p_i = 1/\Omega$ or zero (because $\ln p_i$ is then a constant that comes outside the sum, and $\sum_i p_i = 1$). If the value of $0 \ln(0)$ bothers you, consider $\ln(x)/x^{-1}$ and use l'Hôpital's rule, where the limit of f/g is f'/g' when either f and g are both zero or both divergent. Thus, with $f = \ln x$ and $g = 1/x$, $\lim_{x\to 0} x \ln x = 0$. As with the original Boltzmann expression we note that the microscopic definition of entropy is $S = 0$ when the microstate is specified exactly. We can therefore see at the outset that the statistical definition of the entropy is a little odd: it is a property of the **ensemble** of possible microstates and is concerned with all the possible states for the system, rather than being a property of the state it actually occupies (unlike other extensive quantities, such as internal energy). This is also true for the original Boltzmann $S = k \ln \Omega$, but the Gibbs form makes this point more explicit.

Additivity of the Gibbs entropy Given a system composed of many independent subsystems, the total Gibbs entropy is the sum of the Gibbs entropies for the subsystems. The probability of being in an overall microstate is $p_1(n_1) \times p_2(n_2) \times \ldots$, where n_1 labels the microstate of system 1 etc., so

$$S_G/k = -\sum_{n_1} \sum_{n_2} \sum_{\ldots} p_1(n_1) p_2(n_2) \ldots (\ln p_1 + \ln p_2 + \ldots). \tag{2.4}$$

Since the p's sum to unity, we get

$$S_G/k = -\sum_{n_1} p_1(n_1) \ln p_1(n_1) - \sum_{n_2} p_2(n_2) \ln p_2(n_2) - \ldots \tag{2.5}$$

Thus the Gibbs definition of entropy is **extensive**, as we would require if it is to correspond to the classical entropy.

Advantages of the Gibbs entropy   The Gibbs entropy has several advantages over the Boltzmann entropy:

(1) It defines entropy directly from the distribution of microstates, thus avoiding the identification of a macrostate.

(2) It defines entropy for systems that are not large (e.g. systems with only one or two states). This is very important since it helps one to 'divide and conquer' by breaking up a system into small subsystems.

(3) It defines entropy for systems that are not in thermal equilibrium but are evolving with time and perhaps undergoing some irreversible process.

Bayesian view   What exactly do we mean by probability here?. Take a given classical system: at any given time, it is undoubtedly in some specific microstate, j, so surely $p_j = 1$ and all other p's are zero, meaning that $S = 0$ at all times? But the point is that we do not *know* what j is, and the use of probabilities represents the fact that our knowledge is restricted to just the range of possibilities. This is exactly what is meant by probability in the Bayesian framework: a **degree of belief**, rather than anything to do with observed frequencies of actual events. In this view, entropy is not an objective property of the system under study, unlike e.g. energy, but is a **subjective** combination of the system and of our detailed knowledge about it. This is not an easy proposition to accept, since one may with justice feel that physical systems ought to obey fundamental laws in a way that is independent of observers: an ice cube will melt, whether or not we look at it. We will try to answer this objection in the following sections.

Frequentist view   We can compare this with the more empirical approach of carrying out N **trials** of some experiment, where N_i of these have the outcome labelled i. Then we can define the probability in terms of relative frequencies:

$$p_i \equiv \lim_{N \to \infty} N_i/N. \tag{2.6}$$

This situation is consistent with the Bayesian view: if we toss a coin a billion times and get 500 million heads, that certainly gives a very strong basis for the belief that $p = 0.5$ for getting heads at the next toss. But the Bayesian framework is more general: the probability of a fair coin toss coming out as heads is 0.5, and we can state this belief before we ever toss it. As we will see below, the frequentist point of view is useful as a practical means of calculation in statistical mechanics.

Disorder and hidden information   It is interesting to compare this with the treatment in *Thermal Physics*. This presented the common terminology in which entropy is equated to **disorder**: the more disordered a macrostate is, the higher the statistical weight and the higher the entropy. For example: water has a higher entropy than ice because the crystal lattice restricts the spatial positions of water molecules, but the molecules have many more possible locations in the liquid phase. This is consistent with the message of the Gibbs expression: $S = 0$ when the system is in a single perfectly known microstate, but S rises if we lose knowledge of exactly which microstate is occupied and have to assign a probability to each possible microstate. Another fruitful way of describing this situation is to say that entropy is a measure of **hidden information** (sometimes 'missing information'): a system will occupy some specific microstate, but the less certain we are about what this is, the more the system has in practice become randomised, the greater the disorder, and the higher the entropy.

2.2   Shannon's information entropy

In 1948, Claude Shannon published a radical article entitled "A Mathematical Theory of Communication", in which he established a result in the theory of information that turns out to be of great importance in understanding statistical thermodynamics.

Suppose we have a message that consists of a stream of different characters: how much computer memory is required to store it? The lazy default is to use **ascii** coding, which assigns a distinct pattern of 8 bits to each character, using $8N$ bits for a message of N characters. But such text can usually be stored much more efficiently than this, by exploiting the fact that some characters are rare and some common, and assigning codes with the fewest bits to the common characters. Further gains are possible if there is repetition of characters: for 100 consecutive occurrences of 'A', we just need to send codes for 'A' and '100'. In general, consider a message of N characters, and suppose there are m different characters, each occurring N_i times, with $i = 1, \ldots, m$. The number of possible distinct messages of this sort is given by the **combinatorial** question of how to choose N_1, N_2 etc. from N. The number of possible combinations is

$$W = \frac{N!}{\prod_{i=1}^{m} N_i!},$$

(2.7)

You may not have seen this **multinomial coefficient** before. It generalises the binomial coefficient, where we choose N_1 from N: the number of ways of doing this is $N!/(N_1!M!)$, where $M = N - N_1$. We can now split M into N_2 and N_3, which gives a further factor of the same binomial form.

For Shannon's communication problem, then, we just need to store a number up to a maximum of W to specify which message was received (assuming the N_i are fixed in advance). This requires **information** – defined as the number of binary digits (bits) – equal to $\mathcal{I} = \log_2 W = \ln W/\ln 2$. For large N we can use **Stirling's approximation**:

$$\ln(N!) \simeq N \ln N,$$

(2.8)

so that the number of messages is

$$\ln W = N \ln N - \sum_{i=1}^{m} N_i \ln N_i = \sum_{i=1}^{m} (N_i \ln N - N_i \ln N_i) = -N \sum_{i=1}^{m} p_i \ln p_i,$$

(2.9)

where $p_i = N_i/N$ is the probability of getting the ith character. The same relation applies with base-2 logs, so that the mean information per character is

$$\mathcal{I}/N = - \sum_{i=1}^{m} p_i \log_2 p_i.$$

(2.10)

Von Neumann advised Shannon to call this the **entropy**, "since no-one knows what entropy really is".

A sanity check on the Shannon formula is provided by the simplest example of only two characters, one of which is rare, $p_2 \ll 1$. The information in the message is then just the lengths of the unbroken runs of character 1, punctuated by a single instance of character 2. The typical run length will be $1/p_2$ so we expect to use $\log_2(1/p_2)$ bits for each run length, of which there will be Np_2 in the message. Thus the mean number of bits needed per character is $-p_2 \log_2 p_2$, which is Shannon's expression since $p_1 \log_2 p_1 \to 0$ as $p_1 \to 1$. We can see that the Shannon formula states that one should be able to code so that on average $\log_2(1/p_i)$ bits are assigned to a character with probability p_i. Rare characters are surprising, and their arrival constitutes much more information than a common character. This all really works. These lecture notes contain approximately 338,000 characters, or 331KB when stored as ascii. But the file size declines to 89KB when passed through the `gzip` algorithm, giving an average of 2.1 bits per character (cf. 4.7 bits if a to z were all used with equal frequency).

It is important to be clear about what the Shannon entropy represents. It does not measure the information in any given message: if we knew in advance what was coming, we could devise a

maximally efficient coding scheme where that particular message was coded with just one bit. So the Shannon entropy is the 'information imprecision' allowing for all the messages that you *might* have received – the amount of storage you need to set aside to cope with any possible message given the frequencies of the various letters. It can therefore be viewed as a measure of **hidden information**, analogous to the physical case of the Gibbs entropy; the hidden information here being which of the many possible messages you are actually about to receive. The fact that the two entropies have the same functional form is therefore more than a coincidence, and the Shannon form illuminates the interpretation of the Gibbs entropy.
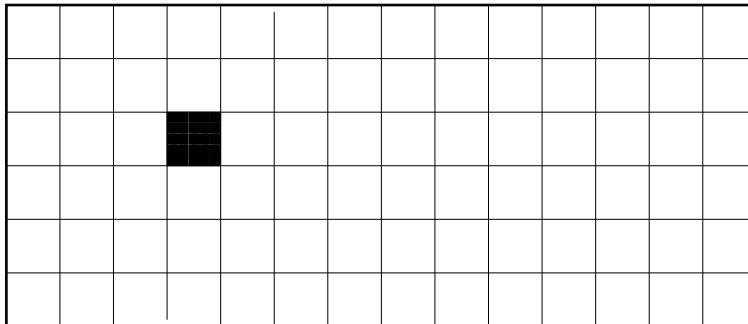
Figure 2: The idea of an ensemble: the 'super-system' is divided up into a large number M of systems e.g. the black square is one system.

3 Ensembles in statistical mechanics

To make use of the appearance of probabilities in the Gibbs entropy, it is convenient to take a frequentist approach and introduce the idea of an **ensemble**. This approach was pioneered by Maxwell, who proposed that we should consider a very large number of copies, M, of the same macroscopic system, but where each copy may occupy a different microstate. In other words, we are making a **realisation** of the random process governed by the probabilities p_i of occupying a given microstate – in effect, tossing some multi-sided coin. This picture aligns with the idea of thermal equilibrium, in which a system is constantly being rearranged as a result of external random influences. Then the members of the ensemble might represent a single system seen at different times – although the concept is more general than this.

In any case, using the frequency definition of probability, the probability of microstate i is related to the number of systems in microstate i divided by M

$$p_i = \lim_{M \to \infty} \frac{m_i}{M}. \tag{3.1}$$

Statistical averages over the probabilities of different microstates can then be carried out in a frequentist manner by simple averaging over members of the ensemble. The macroscopic energy can be thought of as the average over the energy of the energy of each member of the ensemble (each of which has a definite microstate):

$$\boxed{\langle E \rangle \equiv \bar{E} = \sum_i p_i E_i.} \tag{3.2}$$

We will use these notations interchangeably to denote the **ensemble average** of the energy – which in general will be different from the energy of any given member of the ensemble. But note that entropy *cannot* be viewed in this way, since $S = 0$ for any specific microstate.

One can think of this ensemble of systems as a single super-system (see Fig. 2). Practically one can think of a block of material: each system being a small piece of the whole block. Each

of the many pieces can still contain Avogadro size numbers of atoms so we can have both the number M of systems large and the number N of particles in each system large.

Deriving the Gibbs entropy Considering this general ensemble allows the form of the Gibbs entropy to be justified. Consider an ensemble of N versions of a system, with n_i in microstate i. The number of ways of arranging this 'multiverse' of systems is $\Omega = N!/\prod_i n_i!$ (assuming two systems in the same microstate to be indistinguishable). Taking logs and using Stirling's approximation, $S/k = \ln \Omega = N \ln N - \sum_i n_i \ln n_i = -\sum_i n_i \ln(n_i/N)$. Dividing by N to get the **entropy per system**, we get $-\sum_i p_i \ln p_i$, where $p_i = n_i/N$. Note again that the entropy is a property just of the ensemble, so that a single system does not have a meaningful entropy, even though it has distinct value for both the actual energy and its ensemble average.

The key question is whether the different (sub-)systems can interact with each other, and we now distinguish three different kinds of ensembles (with unfortunately rather obscure and unhelpful names).

3.1 Canonical ensemble

The simplest ensemble (which might be better termed the 'thermal equilibrium ensemble') arises when the members of the ensemble can all exchange heat and come into mutual thermal equilibrium. One can think of this as a single member of the ensemble coming into equilibrium with a much larger **heat bath**, which is the source of randomness. Let heat dQ be transferred reversibly to the chosen system, moving it from some initial microstate to a new one and changing its total energy from E to $E + dQ$. Since there is a definite microstate in both cases, there is no change in system entropy; but the heat bath suffers $dS = -dQ/T$, i.e. a change in internal degrees of freedom of $d\ln \Omega = dS/k = -dE/kT$, according to Boltzmann. If we regard the statistical weight Ω as being proportional to probability, we deduce the **Boltzmann factor** for the system being in a given microstate:

$$p(E_i) \propto \exp(-E_i/kT). \tag{3.3}$$

This can be normalized by summing over all states to deduce the **partition function**, Z:

$$\boxed{p(E_i) = \frac{\exp(-E_i/kT)}{Z}; \qquad Z = \sum_i \exp(-E_i/kT).} \tag{3.4}$$

The partition function turns out to be an object of particular practical importance in thermodynamic calculations.

3.2 Grand canonical ensemble

A simple extension of the canonical ensemble (with an even less useful name) allows also exchange of particles with the heat bath, with a chemical potential (energy per particle transferred) of μ. The first law is $dE = T\,dS + \mu\,dN$; applying this to the bath with $d\ln \Omega = dS/k$ once again, the statistical weight for the (E, N) microstate becomes the **Gibbs factor**,

$$p(E_i, N_i) \propto \exp(-(E_i - \mu N_i)/kT). \tag{3.5}$$

And in analogy with what was done for the canonical ensemble, normalization involves the **grand partition function**, Z_{G}:

$$\boxed{p(E_i, N_i) = \frac{\exp(-(E_i - \mu N_i)/kT)}{Z_{\mathrm{G}}}; \qquad Z_{\mathrm{G}} = \sum_i \exp(-(E_i - \mu N_i)/kT).} \tag{3.6}$$

3.3 Microcanonical ensemble

Here the individual members are **isolated systems**. This is a more difficult case: there is no interaction leading to thermal equilibrium with an external heat bath, and so there is no explicit source of randomness. And yet isolated systems generally do settle into an *internal* thermal equilibrium. Understanding how this happens is at the heart of the second law. Because of the isolation, the total energy is fixed; if the microstates were non-degenerate, then all members of this ensemble would have to be identical. The microcanonical ensemble is then only interesting if there exists a number, N, of microstates with the same energy – in which case we would appeal to PEAPP to say $p_i = 1/N$ for all these states. To cover the non-degenerate case, it is common to define an ensemble where the system energy lies in some small range ΔE, and later taking the limit of $\Delta E \to 0$. But all of this is obscure in several ways. It sounds like we need some random external influence to cause the members of the ensemble to jump between the allowed microstates, but external contact should lead to full thermal equilibrium as in the Canonical ensemble. As mentioned earlier, Boltzmann's vision was that internal dynamical evolution (in particular, collisions between particles) would lead a system to explore different microstates and so possibly result in PEAPP (at least, after a sufficiently long time has elapsed – though certainly this will not hold early in the evolution). But we will see that all of this is more complicated than expected when examined in detail.

3.4 The maximum entropy approach

An illuminating alternative approach to obtaining the Boltzmann and Gibbs factors is to argue that entropy should be maximised. This follows from the free energy, where we know that $F = E - TS$ should reach a minimum in equilibrium (at constant volume). This is a little problematic for a single microstate, since E is fixed and $S = 0$. But apply it to the whole ensemble: the total energy is fixed, but in thermal contact the system will fluctuate between different microstates, for which there are different probabilities and so a well defined entropy over the ensemble. So if we want to find the equilibrium configuration of the ensemble, we must minimise its F – which corresponds to maximising the entropy, since the total energy in the ensemble is fixed. We can express this in terms of expectation values, which must satisfy $\bar{F} = \bar{E} - TS$: we need to maximise S when $\bar{E}$ is given. This is an example of the general idea of constrained maximum entropy:

> $\{p_i\}$ should be such as to maximise S subject to *constraints* imposed by the available information.

We will proceed to apply this principle by maximising the Gibbs entropy, $-k \sum_i p_i \ln p_i$, although we need to be clear that we have yet to prove that this is a good proxy for the classical entropy; this is done in the following section.

Constraints: These are intended to be classical values, so in the statistical context they will typically take the form of **expectation values** of some observables, e.g. $\bar{E} = \sum_i E_i p_i = \text{const}$. The general problem is as follows: maximise $f(x_i)$ subject to the constraint $g(x_i) = \text{const}$, where x_i are various coordinates (the microstate probabilities), f is the function of prime interest (the entropy), and g is a constraint function (which here has the simple form of a linear sum over the coordinates). In fact, there will be more than one constraint: in addition to the value of $\bar{E}$, the probabilities sum to unity:

$$\sum_i p_i = 1 . \tag{3.7}$$

Conditional optimisations of this form can be performed using the method of **Lagrange multipliers** or **undetermined multipliers**. We want f to be stationary for changes in

the x_i that move within the level surface of g:

$$df = \sum_{i=1} \frac{\partial f}{\partial x_i} \, dx_i = \underline{\nabla} f \cdot \underline{dx} = 0 \,, \tag{3.8}$$

where $\underline{\nabla} g \cdot \underline{dx} = 0$. These two conditions can be satisfied by the single criterion

$$\boxed{\underline{\nabla}(f - \lambda g) = 0,} \tag{3.9}$$

where λ is a number that is unknown at this stage. Additional constraints are added in the same way: if we also had $h(x_i) = \text{const}$, we would solve $\underline{\nabla}(f - \lambda_1 g - \lambda_2 h) = 0$ etc.

Mathematical aside (non-examinable)   The justification for this procedure can be seen geometrically. If $\underline{dx}$ is in the level surface of g then $\underline{\nabla} g \cdot \underline{dx}$ is zero by definition. Thus $\underline{\nabla}(f - \lambda g) \cdot \underline{dx}$ is $\underline{\nabla} f \cdot \underline{dx}$ and setting this to zero finds the conditional stationary point of f. But if $\underline{\nabla}(f - \lambda g) = 0$ then we must have $\underline{\nabla}(f - \lambda g) \cdot \underline{dx} = 0$ for *any* $\underline{dx}$, even allowing a component parallel to $\underline{\nabla} g$. But we can arrange for the effect of this component to vanish if we set $\lambda = (\underline{\nabla} f \cdot \underline{\nabla} g)/|\underline{\nabla} g|^2$. We don't yet know what this number is, but we know it exists and so the problem of finding a conditional zero gradient for f can be replaced by the simpler one in which the unconditional gradient of $f - \lambda g$ vanishes.

Example 1   First suppose there are no given constraints other than $\sum_i p_i = 1$. Then we need to make the function h stationary, where

$$h = -k \sum_j p_j \ln p_j - \lambda \sum_j p_j \,. \tag{3.10}$$

Then our conditions become

$$\frac{\partial h}{\partial p_i} = -k \left[\ln p_i + 1 \right] - \lambda = 0 \Rightarrow p_i = \exp[-1 - \lambda/k] \,, \tag{3.11}$$

making p_i constant for all i. But the normalisation constraint requires

$$\boxed{\sum_i p_i = M \exp[-1 - \lambda/k] = 1 \qquad \Rightarrow p_i = \frac{1}{M} \,,} \tag{3.12}$$

where there are M distinct microstates. Thus the maximum entropy state in this case of no constraints is that of maximum ignorance, in which all allowed states are considered equally likely. This is the principle of equal a priori probability (PEAPP). Note that this case covers that of the microcanonical ensemble, where the probabilities cover a range of microstates that have a fixed energy, and so what we have done in effect is maximise the entropy subject to exactly fixed energy.

Example 2   Now suppose we have two constraints in the form of linear averages over quantities y and z that are specified for each state. We now require

$$\frac{\partial h}{\partial p_i} = 0 \,, \quad \text{where} \quad h = -k \sum_j p_j \ln p_j - \lambda \sum_j p_j - \lambda_y \sum_j y_j p_j - \lambda_z \sum_j z_j p_j \,. \tag{3.13}$$

This gives $-k(\ln p_i + 1) - \lambda - \lambda_y y_i - \lambda_z z_i = 0$, so that

$$p_i = \exp(-1 - \lambda/k) \exp(-[\lambda_y y_i + \lambda_z z_i]/k) \,. \tag{3.14}$$

19

The first factor, $\exp(-1 - \lambda/k)$ is a constant (independent of i) and as before can be eliminated via normalization, $\sum_i p_i = 1$:

$$p_i = \frac{1}{Z} \exp(-[\lambda_y y_i + \lambda_z z_i]/k) \quad \text{where} \quad Z = \sum_i \exp(-[\lambda_y y_i + \lambda_z z_i]/k). \tag{3.15}$$

Note also that

$$\frac{\partial^2 h}{\partial p_i^2} = -\frac{k}{p_i} < 0, \qquad \text{as} \quad p_i > 0, \tag{3.16}$$

so it is indeed a maximum.

The last example should look familiar. If we say that y_i is the energy of a microstate, E_i, and z_i is the corresponding number of particles, N_i, then the maximum-entropy solution for the microstate probabilities under the constraint that $\bar{E}$ and $\bar{N}$ are both known is just the Gibbs factor, and we can identify the Lagrange multipliers as

$$\lambda_y = \frac{1}{T}; \quad \lambda_z = -\frac{\mu}{T}. \tag{3.17}$$

Similarly, the simpler problem in which there is only one constraint, $\sum_i p_i y_i = \text{const}$ follows the two-constraint solution, but neglecting the term involving λ_z. Then we get

$$p_i = \frac{1}{Z} \exp(-\lambda_y y_i/k) \quad \text{where} \quad Z = \sum_i \exp(-\lambda_y y_i/k); \tag{3.18}$$

and if we choose y to be the energy, then the maximum-entropy solution with fixed $\bar{E}$ is the Boltzmann factor, with $\lambda_y = 1/T$.

In short, by maximising the Gibbs entropy subject to various constraints we have derived the standard **ensembles** of statistical mechanics. The Lagrange multipliers that enter into the maximisation procedure have been identified with intensive thermodynamics variables: temperature and chemical potential.

In the tutorials, you are invited to show how two systems in equilibrium under exchange of a certain quantity should share the same value of the Lagrange multiplier corresponding to that quantity. For example systems in equilibrium with respect to energy exchange should share the same value of λ_E i.e. the same temperature. This recovers the zeroth law of thermodynamics. Similarly systems in equilibrium with respect to particle exchange should share the same value of λ_N i.e. the same chemical potential.

Alternative extremisation principles   Recall that our original programme was to maximise S subject to constraints. If the constraint is $\bar{E}$ (Canonical ensemble) then we ended up extremising

$$h = S - \lambda_1 \sum_i p_i - \lambda_E \bar{E}, \tag{3.19}$$

which is equivalent to maximising

(i) $h/\lambda_E = -F - \lambda_1/\lambda_E \sum_i p_i$ i.e. *minimising* F subject to the constraint of probability conservation (taking $\lambda_E = 1/T$, as above).

(ii) $h/\lambda_E = -\bar{E} + 1/\lambda_E S - \lambda_1/\lambda_E \sum_i p_i$ i.e. *minimising* $\bar{E}$ subject to the constraint of constant S (i.e. we would have a Lagrange multiplier $\lambda_S = 1/\lambda_E$)

Thus the different extremisation principles (maximisation of entropy, minimisation of energy, minimisation of free energy etc.) are all intimately related and which one applies is determined by which ensemble one is working in, i.e. which constraints apply and which variables are held fixed.

4 Relation to classical thermodynamics

The above discussion has focused on ensemble average quantities, such as the mean energy, $\overline{E}$. For any real system in thermal equilibrium, the exact energy will differ from $\overline{E}$. However, later on we will calculate the size of these fluctuations and show that they become vanishingly small (in fractional terms) as the size of the system tends to infinity. Therefore, in this limit of very large systems (which are the subject of classical experimental thermodynamics), we should expect to be able to use E and $\overline{E}$ interchangeably. Our remaining task is then to check that the ensemble-average statistical quantities do behave in the same way as the corresponding classical quantities, following which we derive some powerful relations that allow the statistical results to be exploited.

Consider first the canonical ensemble. The microscopic definition of the ensemble average energy of a system is

$$\overline{E} = \sum_i p_i E_i \,. \tag{4.1}$$

Now, we know from quantum mechanics that changing the volume of the 'box' for example must change the energy levels. It follows that a change a volume-like coordinate dX_γ must give rise to a change in the mean energy. Therefore a change in the mean energy has two types of contribution, one from changes in p_i and one from changes in E_i due to X_γ

$$
\begin{aligned}
d\overline{E} &= \sum_i \frac{\partial \overline{E}}{\partial p_i} dp_i + \sum_\gamma \frac{\partial \overline{E}}{\partial X_\gamma} dX_\gamma \tag{4.2} \\
&= \sum_i E_i \, dp_i + \sum_\gamma \frac{\partial \overline{E}}{\partial X_\gamma} dX_\gamma \,. \tag{4.3}
\end{aligned}
$$

We now consider a small change in the entropy. Since $S = -k \sum_i p_i \ln p_i$ is just a function of the probabilities p_i, a change in S comes from a change in the probability distribution, and this may be expressed as

$$
\begin{aligned}
dS &= \sum_i \frac{\partial S}{\partial p_i} dp_i \tag{4.4} \\
&= -k \sum_i \left[\ln p_i + 1 \right] dp_i \tag{4.5} \\
&= -k \sum_i \ln p_i \, dp_i \tag{4.6}
\end{aligned}
$$

where the last step uses $\sum_i dp_i = 0$ (conservation of probability). Using the Boltzmann factor for p_i, this becomes

$$dS = \frac{1}{T} \sum_i E_i dp_i \,, \tag{4.7}$$

implying

$$d\overline{E} = T \, dS + \sum_\gamma \frac{\partial \overline{E}}{\partial X_\gamma} dX_\gamma \,. \tag{4.8}$$

Similarly, if we are dealing with systems where particle number can vary, we would use the Gibbs factor for p_i. In this case, we then have

$$dS = \frac{1}{T} \sum_i (E_i - \mu N_i) dp_i \,, \tag{4.9}$$

implying

$$dE = T\,dS + \mu\,d\bar{N} + \sum_{\gamma} \frac{\partial \overline{E}}{\partial X_{\gamma}} dX_{\gamma} \tag{4.10}$$

(because $d\bar{N} = \sum_i N_i dp_i$).

This equation has the same form as the classical first law, provided we make the following identification of the thermodynamic 'generalised forces':

$$f_{\gamma} = \frac{\partial \overline{E}}{\partial X_{\gamma}}. \tag{4.11}$$

For example the pressure is given by

$$-P = \frac{\partial \overline{E}}{\partial V} = \sum_i p_i \frac{\partial E_i}{\partial V}. \tag{4.12}$$

In short, the statistical quantity $\bar{E}$ and the Gibbs entropy can be used consistently as proxies for the classical energy and entropy, and we assume hereafter that we can use them interchangeably for macroscopic systems. But the classical entropy refers specifically to a thermal equilibrium state, whereas Gibbs automatically allows a generalisation to a non-equilibrium situation in which the microstate probabilities are evolving.

4.1 The bridge equation

We can now introduce one of the central results of statistical mechanics. Consider first the canonical ensemble, where we recall the Boltzmann distribution in the form involving the partition function:

$$p_i = \frac{1}{Z}\exp(-E_i/kT) \qquad \text{where} \quad Z = \sum_i \exp(-E_i/kT) \tag{4.13}$$

Substituting this form for p_i into the expression for the Gibbs entropy gives

$$S = -k\sum_i p_i\left[-E_i/kT - \ln Z\right] = \frac{\overline{E}}{T} + k\ln Z. \tag{4.14}$$

Or, recalling the Helmholtz free energy $F = \overline{E} - TS$, we may rewrite the above equation as

$$F = -kT\ln Z \tag{4.15}$$

This remarkable result is sometimes referred to as the **bridge equation**, as it provides a bridge between the microscopic and macroscopic descriptions of a system. The Helmholtz free energy, which is a potential that we introduced on macroscopic grounds, is obtained by using the canonical partition function, which comes from summing over microscopic states.

Similarly, for systems with variable particle number, we follow the same argument but now using the Gibbs factor instead of Boltzmann:

$$S = -k\sum_i p_i\left[-(E_i - \mu N_i)/kT - \ln Z_{\mathrm{G}}\right] = \frac{\overline{E}}{T} - \frac{\mu\overline{N}}{T} + k\ln Z_{\mathrm{G}}, \tag{4.16}$$

where the **grand partition function** is $Z_{\mathrm{G}} = \sum_i \exp[-(E_i + \mu N_i)/kT]$. Hence the **grand potential**, $\Phi = \overline{E} - TS - \mu\overline{N}$, may be expressed as

$$\Phi = -kT\ln Z_{\mathrm{G}} \tag{4.17}$$

Again we have a remarkable **bridge equation** connecting the microscopic and macroscopic descriptions of an open system.

The significance of these results is colossal, because the free energy and the grand potential are the routes by which we can calculate other thermodynamic quantities of importance:

$$S = -\left.\frac{\partial F}{\partial T}\right|_V \qquad P = -\left.\frac{\partial F}{\partial V}\right|_T. \tag{4.18}$$

Thus the evaluation of partition functions turns out to be one of the most important practical tasks in statistical thermodynamics. We can sum up this this section by writing a selection of relations based on partition functions, and contrasting them with their direct definitions in terms of microstate probabilities:

$$\boxed{\begin{array}{ll}
\bar{E} = \sum_i p_i E_i & = -\left.\frac{\partial}{\partial \beta}\right|_V \ln Z \\[2mm]
S = -k \sum_i p_i \ln p_i & = \left.\frac{\partial}{\partial T}\right|_V kT \ln Z \\[2mm]
P = -\sum_i p_i \frac{dE_i}{dV} & = \left.\frac{\partial}{\partial V}\right|_T kT \ln Z \\[2mm]
\bar{N} = \sum_i p_i N_i & = \left.\frac{\partial}{\partial \mu}\right|_{T,V} kT \ln Z_{\mathrm{G}},
\end{array}} \tag{4.19}$$

where we have used the common notation

$$\beta \equiv \frac{1}{kT}. \tag{4.20}$$

5 Fluctuations and response

All the above discussion has concerned thermal equilibrium. But this arises as a result of random influences on a system from the heat bath. We have used the statistics of this process to calculate ensemble average quantities such as the mean energy, $\overline{E}$. But for any single member of the ensemble the actual energy will fluctuate about this value – and this is true for most thermodynamic quantities. The tools we have assembled so far allow us to compute these fluctuations.

5.1 Fluctuations in the energy of a system

Consider the Canonical Ensemble, in which the (internal) energy of a system fluctuates randomly about the fixed mean value $\overline{E}$. First note that the mean energy may be expressed in the Canonical Ensemble as

$$\overline{E} = \sum_i p_i E_i = \frac{\sum_i E_i \exp(-\beta E_i)}{\sum_i \exp(-\beta E_i)} \tag{5.1}$$

$$= -\frac{1}{Z}\frac{\partial Z}{\partial \beta} \tag{5.2}$$

$$= -\frac{\partial \ln Z}{\partial \beta}. \tag{5.3}$$

But this evaluation of $\bar{E}$ is an average over a probability distribution of different microstate energies, so we can use the same reasoning to compute the average of any function of energy, e.g. $\langle E^2 \rangle$. But this is what we need to compute fluctuations. Consider the offset of a given E_i from the ensemble mean:

$$\Delta E_i \equiv E_i - \overline{E}, \tag{5.4}$$

noting that $\langle \Delta E_i \rangle = 0$. So if we square the fluctuation, and take averages, we obtain the usual expression for the **variance** in energy:

$$\langle (\Delta E)^2 \rangle = \langle E^2 \rangle - \langle E \rangle^2. \tag{5.5}$$

And finally, by noting that two derivatives of Z with respect to β brings down E_i^2 inside the sum, we see that the required average of E^2 is also related to a derivative of the the partition function:

$$\boxed{\langle E^2 \rangle = (1/Z) \partial^2 Z / \partial \beta^2.} \tag{5.6}$$

Thus, knowing the partition function, we can compute the fluctuations in energy. Note that this doesn't tell us everything about the fluctuations – e.g. we don't learn how long it takes the energy to change – but we can compute the rms amplitude of any eventual fluctuations.

In practice, this result for the mean-square fluctuation in energy can be conveniently expressed in terms of the heat capacity C_V (at constant volume):

$$C_V \;\; = \;\; \left(\frac{\partial \overline{E}}{\partial T} \right)_V = \frac{d\beta}{dT} \frac{\partial \overline{E}}{\partial \beta} \tag{5.7}$$

$$= \;\; -\frac{1}{kT^2} \left[-\frac{1}{Z} \frac{\partial^2 Z}{\partial \beta^2} + \frac{1}{Z^2} \left(\frac{\partial Z}{\partial \beta} \right)^2 \right], \tag{5.8}$$

where we have first used the chain rule for differentiation. Using the earlier relations between moments of E and derivatives of Z, we see that

$$C_V = \frac{1}{kT^2} [\overline{E^2} - \overline{E}^2], \tag{5.9}$$

and so taking the square-root of both sides leads to an expression for the root-mean-square fluctuation ΔE_{rms}:

$$\Delta E_{\text{rms}} \equiv \left(\overline{(\Delta E)^2} \right)^{1/2} = (C_V/k)^{1/2} \, kT. \tag{5.10}$$

Evidently the relative fluctuation, i.e. the rms fluctuation on the scale of the mean, may be written as

$$\frac{\Delta E_{\text{rms}}}{\overline{E}} = \frac{kT \sqrt{C_V/k}}{\overline{E}} \sim \frac{1}{N^{1/2}}, \tag{5.11}$$

where the last step follows from the fact that both $\overline{E}$ and C_V are **extensive** quantities: $C_V \sim Nk$ and $\bar{E} \sim N kT$. For Avogadro-sized systems, we have $N \sim 10^{24}$ and hence the relative fluctuation in the energy has a root-mean-square value of about $\sim 10^{-12}$. This is less than any experimental precision could detect, so the value of the energy is **sharp**.

This property is often referred to as the **equivalence** of the Canonical and Microcanonical Ensembles i.e. in the Microcanonical Ensemble the energy is strictly fixed and only microstates with the same energy are available, whereas in the Canonical Ensemble microstates of all energy are available but are sampled with the Canonical probabilities, which depend on E_i. The result is that the actual energy fluctuations vanish in the large-N limit and we expect the physical properties of the Microcanonical and Canonical Ensembles to be the same in this limit.

5.2   Magnetisation fluctuations

Let us consider a system in an applied external magnetic field $\underline{H}$ and write the energy of a microstate as

$$E_i(\underline{H}) = E_i(\underline{H} = 0) - \mu_0 \underline{M_i} \cdot \underline{H}, \tag{5.12}$$

where $\underline{M}_i$ is the magnetisation of the system in microstate i and μ_0 is a constant (vacuum permeability). The energy is reduced when dipoles align with field.

Consider the one-dimensional case here. The mean magnetisation may be expressed in the Canonical Ensemble as

$$\overline{M} = \sum_i p_i M_i = \frac{1}{Z} \sum_i M_i \exp(-\beta E_i(0) + \beta \mu_0 M_i H) \tag{5.13}$$

$$= \frac{1}{\beta \mu_0 Z} \frac{\partial Z}{\partial H} = \frac{1}{\beta \mu_0} \frac{\partial \ln Z}{\partial H}. \tag{5.14}$$

Again we see the idea of writing the expectation value of an observable (here M) as the derivative of $\ln Z$ with respect to the conjugate field (here H). In the tutorial you are invited to show that

$$\overline{\Delta M^2} = \frac{kT}{\mu_0} \chi, \tag{5.15}$$

where

$$\chi = \left(\frac{\partial \overline{M}}{\partial H} \right)_{T,V}, \tag{5.16}$$

is the isothermal susceptibility.

5.3 Density fluctuations

Let us now consider a system in the Grand Canonical Ensemble and study the fluctuations in the number of particles – the number of particles in a system in the Grand Canonical Ensemble will fluctuate about the mean value $\overline{N}$. The derivation parallels the one concerning energy fluctuations. We begin by noting how we may express $\overline{N}$:

$$\overline{N} = \sum_{i,N} N p_{i,N} = \frac{\sum_{i,N} N \exp(-\beta(E_{i,N} - \mu N)}{\sum_{i,N} \exp(-\beta(E_{i,N} - \mu N))} \tag{5.17}$$

$$= \frac{1}{\beta Z_{\mathrm{G}}} \frac{\partial Z_{\mathrm{G}}}{\partial \mu} = \frac{1}{\beta} \frac{\partial \ln Z_{\mathrm{G}}}{\partial \mu}. \tag{5.18}$$

This is a relation we have seen previously in the context of the **grand potential** $\Phi = -kT \ln Z_{\mathrm{G}}$, where in general

$$\overline{N} = - \left. \frac{\partial \Phi}{\partial \mu} \right|_{T,V}. \tag{5.19}$$

Now we have

$$\left. \frac{\partial \overline{N}}{\partial \mu} \right|_{T,V} = \frac{1}{\beta} \frac{\partial^2 \ln Z_{\mathrm{G}}}{\partial \mu^2} \tag{5.20}$$

$$= \frac{1}{\beta} \left[\frac{1}{Z_{\mathrm{G}}} \frac{\partial^2 Z_{\mathrm{G}}}{\partial \mu^2} - \frac{1}{Z_{\mathrm{G}}^2} \left(\frac{\partial Z_{\mathrm{G}}}{\partial \mu} \right)^2 \right] = \beta \left[\overline{N^2} - \overline{N}^2 \right], \tag{5.21}$$

giving

$$\overline{\Delta N^2} = kT \left. \frac{\partial \overline{N}}{\partial \mu} \right|_{T,V}. \tag{5.22}$$

25

Since the RHS is extensive $\propto \overline{N}$, we have

$$\frac{\Delta N_{\mathrm{rms}}}{\overline{N}} \propto \frac{1}{\overline{N}^{1/2}} \,. \tag{5.23}$$

$(\Delta N_{\mathrm{rms}} \equiv (\overline{\Delta N^2})^{1/2})$. Thus the typical particle number fluctuation is negligible on the scale of the mean, for sufficiently large systems. This is referred to as the equivalence of the Grand Canonical and Canonical Ensembles in the thermodynamic limit $N \to \infty$.

5.4   General theorem

If we have some macroscopic observable A with conjugate field f we can write the energy of a microstate as

$$E_i = E_i(0) - f A_i \,, \tag{5.24}$$

where A_i is the value of the observable in microstate i and $E_i(0)$ is the energy at $f = 0$ i.e. this term does not involve f. Consider the Canonical Ensemble (a similar general theorem can be deduced for the Grand Canonical Ensemble): following the previous development,

$$\beta \overline{A} = \frac{\partial \ln Z}{\partial f} \,, \tag{5.25}$$

and

$$\chi_{AA} \equiv \frac{\partial \overline{A}}{\partial f} = \beta \, \overline{\Delta A^2} \,. \tag{5.26}$$

Here, χ_{AA} is a **generalised susceptibility** i.e. it is the *response* of observable A to a change in the field conjugate to A. Since we expect $\chi_{AA} \propto N$, we have $\Delta A_{\mathrm{rms}}/N \propto 1/N^{1/2}$, and the the typical fractional fluctuation in A vanishes in the large-N limit.

Although we expect $\chi_{AA} \propto N$, it is possible for the coefficient of proportionality to diverge at some parameter values. In this case the above argument for vanishing fluctuations breaks down and there would be large scale fluctuations. This is realised at certain parameter values where **phase transitions** occur.

6   Counting states

So far, we have assumed that different microstates can be identified without being very specific about how this is to be accomplished. We first give a classical discussion, then deal with the quantum case.

6.1   Classical particles and phase space

The classical Hamiltonian is

$$H = E = \sum_i \frac{|p_i|^2}{2m} + U\left(\{q\}\right) \,, \tag{6.1}$$

where i labels the particles, p_i is the momentum and q_i is the position vector of particle i. Here, p_i and q_i are the **conjugate** or **canonical** coordinates of Lagrangian or Hamiltonian dynamics. Thus, classically the state of a particle i is specified by the six numbers p_x, p_y, p_z, q_x, q_y, q_z. In
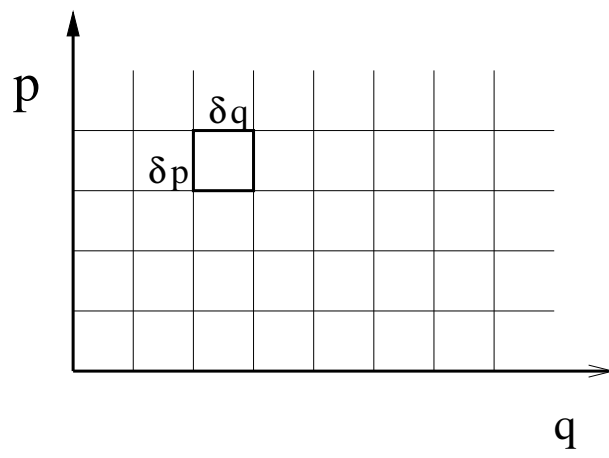
Figure 3: A one-dimensional phase space divided into cells of side δq, δp.

principle, it seems that these need to be specified exactly, but even at the classical level this feels unrealistic. Instead, for a given particle, we consider **cells** defined by a finite range in position and momentum, $\delta q_x \, \delta p_x$ etc. – see Fig. 3 for a one-dimensional illustration. Call the 'volume' of a cell (in the 6 dimensional space)

$$h^3 = (\delta p)^3 (\delta q)^3 \,. \tag{6.2}$$

This 6D space is known as **phase space**, and we can count a finite number of states by integrating over a phase-space volume:

$$N_{\text{states}} = \frac{1}{h^3} \int d^3q \, d^3p. \tag{6.3}$$

The fact that h appears to be arbitrary is clearly unsatisfactory, but:

- Classically one can interpret the cells as the uncertainty in any physical measurement (this was the view of Maxwell).

- In the end the size of the cell turns out to be a 'book-keeping' device that disappears from the final physical results.

- In quantum mechanics the uncertainty principle furnishes a unique definition of the cell size. The notation h turns out to be appropriate.

6.2 Many-particle phase space

We now want to extend this method of describing states to deal with many particles, and there are two distinct approaches, adopted by Boltzmann and by Gibbs:

Distribution function   Here we take the states in the above single-particle phase space, and allow each one to have an **occupation number**, n_i, so that more than one particle can lie in a single state. This is most useful for particles that interact weakly enough that the total energy is dominated by the individual particle kinetic energies – i.e. the particles may interact sufficiently to scatter off each other on occasion, but at any given time the potential energy of such interactions is negligible. In this language, the total particle number and energy are

$$N = \sum_i n_i = \frac{1}{h^3} \int n(q,p) \, d^3q \, d^3p$$

$$E = \sum_i n_i \epsilon_i = \frac{1}{h^3} \int n(q,p) \, \epsilon(q,p) \, d^3q \, d^3p. \tag{6.4}$$

The combination n/h^3 is called the **distribution function**, $f(q, p)$, otherwise known as the **phase-space density**, since $f(q, p)\, d^3q\, d^3p$ is the number of particles existing in a given phase-space volume element. It is worth being clear that the idea of a phase-space density can be used in two different ways. We can consider the exact number of particles in a given cell, n_i, but as usual we might wonder how these numbers could be measured in practice; therefore we might work with the expected number, $\langle n \rangle$, so that the phase-space density provides a multi-particle probability density.

Full phase space But a more general approach is to insist that the specification of a state requires us to list the position and momentum of all particles. Thus the state of the whole system is specified by a single point in a $6N$-dimensional space with coordinates $(q_1, q_2, \ldots, p_1, p_2, \ldots)$. Unfortunately, this space is also commonly referred to as phase space, even though it could benefit from a distinct name to distinguish it from the single-particle form. In this form, the distribution function is a delta-function: the full system is a single point, which is fully contained in a single cell. When we talk about a phase-space density in the Gibbs sense, then, we can only be talking about a probability density.

This is readily supplied in the Canonical Ensemble, for example, where the relative probabilities of different points in phase space are given by a Boltzmann factor. Now the sum over states becomes a 'volume' integral over the $6N$-dimensional space, so that the partition function would be written as

$$Z(T, V, N) = h^{-3N} \int \prod_i d^3q_i\, d^3p_i\, e^{-\beta H(\{\underline{q}\}, \{\underline{p}\})}, \tag{6.5}$$

where the limits of integration for momentum components are $\pm\infty$ and for positions $0, L$.

Independent particles Commonly, the particles that comprise the system will be weakly interacting, so that Hamiltonian (total energy) is a sum of single-particle energies:

$$E = \epsilon_{j_1} + \epsilon_{j_2} + \cdots + \epsilon_{j_N}, \tag{6.6}$$

where the index j_1 labels the different single-particle energy states available to particle 1 etc. The canonical partition function is then

$$Z = \sum_{j_1, j_2 \ldots j_N} \exp -\beta[\epsilon_{j_1} + \epsilon_{j_2} + \cdots + \epsilon_{j_N}]. \tag{6.7}$$

A common trick in dealing with such sums is to realise that the sum over all the indices means a product of single sums:

$$\sum_{j_1, j_2 \ldots j_N} \rightarrow \sum_{j_1} \times \sum_{j_2} \times \cdots \tag{6.8}$$

Furthermore, the sum in the exponential factorises: $\sum_{ij} \exp[A_i + B_j] = (\sum_i \exp[A_i]) \times (\sum_j \exp[B_j])$ etc., leading to

$$Z = \prod_i Z(1)_i \rightarrow [Z(1)]^N, \tag{6.9}$$

where

$$Z(1) = \sum_j \exp(-\beta\epsilon_j), \tag{6.10}$$

is the **single-particle partition function** and the index j runs over the single-particle states – assuming in the final expression that these are the same for all particles, so that $Z(1)$ is

independent of which particle is being considered. The thermodynamic properties of the system now follow via the bridge equation:

$$F = -kT \ln Z = -NkT \ln Z(1) \,. \tag{6.11}$$

So the total free energy is the sum of that from each of the individual non-interacting particles, as expected.

Entropy of identical classical particles It is interesting to contrast the above thermal equilibrium results with the general expression for the entropy of a classical particle system. If the particles are independent, the entropy is $S = NS_1$, where S_1 is the entropy of a single particle. This will generally be of the Gibbs form, $S_1 = -k \sum_i p_i \ln p_i$, where the probability of occupying a single state, which is the probability of lying in a phase-space 'volume' h^3:

$$p_i = \frac{h^3 f(q,p)}{N} . \tag{6.12}$$

The sum over states becomes a phase-space integration:

$$S = -Nk \frac{1}{h^3} \int (h^3 f/N) \ln(h^3 f/N) \, d^3q \, d^3p = -k \int f \ln(h^3 f/N) \, d^3q \, d^3p, \tag{6.13}$$

or

$$S = -k \int f \ln f \, d^3q \, d^3p - Nk \ln h^3 + kN \ln N, \tag{6.14}$$

where we use $N = \int f \, d^3q \, d^3p$. The last term can be recognised as $\ln(N!)$, and its significance arises because we have treated the N particles as distinct (albeit following the same probability distribution). But in most common cases of interest, the particles will be **identical**. The proper treatment of identical particles comes through quantum mechanics, where we deal with spin and wave-function symmetry. But even at the classical level we can note that identical particles are also **indistinguishable**, and this is a problem for counting microstates of a system. If we place particle 1 in state A, and particle 2 in state B, this cannot be distinguished from placing 1 in B and 2 in A. Thus the above procedure overcounts microstates, by a factor of the number of permutations of the N particles. In the original Boltzmann formulation of $S = k \ln \Omega$, we should therefore divide the number of states, Ω, by $N!$. This removes the final term above, yielding

$$\boxed{S = -k \int f \ln f \, d^3q \, d^3p - Nk \ln h^3.} \tag{6.15}$$

Similarly, the overcounting can be corrected by scaling down the partition function:

$$\boxed{Z \to \frac{Z}{N!}.} \tag{6.16}$$

6.3 Identical particles in general

The correct treatment of identical particles becomes even more central when we move beyond the classical limit. The fundamental principle is now that

> the microstate is specified purely by the set of occupation numbers, $\{n_i\}$.

We can see that this formulation implicitly assumes identical particles: if we exchange a pair of particles between two different single-particle states, the occupation numbers do not change. Thus

a description in terms of occupation numbers alone *must* refer to identical particles, and no ad hoc modification of the counting by a factor $N!$ is required. Here, the sum over all microstates is a sum over the allowed occupation numbers for the single-particle states. We write the energy of microstate i as

$$E_i = \sum_j n_j \epsilon_j, \tag{6.17}$$

and the total number of particles in a microstate is

$$N = \sum_j n_j. \tag{6.18}$$

So far so good, but this last equation creates a problem, since the different n_j must add up to give N. This means that summing over the different values of n_j to get the canonical partition function is not easy, because the different n_j are not independent. The solution is to consider instead the grand canonical ensemble, in which N is not fixed, and derive thermodynamic properties from the grand partition function Z_G. This quantity depends on the unknown chemical potential, μ, which sounds like trouble; but as we saw in section 3.4, we can determine μ by applying the softer constraint that on *average* we have $\overline{N}$ particles.

The grand canonical partition function now takes the form

$$Z_G = \sum_{n_1, n_2 \ldots} \exp\{-\beta[n_1\epsilon_1 + n_2\epsilon_2 + \ldots] + \beta\mu[n_1 + n_2 + \ldots]\}, \tag{6.19}$$

where the sum over n_j is over the allowed occupation of state j. Using the same argument as above regarding the factorisation of the exponential, the partition function may be written as

$$\boxed{Z_G = \prod_j Z_j} \tag{6.20}$$

where

$$Z_j = \sum_{n_j} \exp[\beta n_j(\mu - \epsilon_j)]. \tag{6.21}$$

is the **single-state grand partition function** for state j. This result may be compared to equation (6.9) for the partition function of the canonical ensemble of nonconserved particles in the previous subsection. Here the factorisation is over states rather than particles.

The probability of finding the system in the microstate characterised by the set $\{n_j\} \equiv \{n_1, n_2, \ldots\}$, is just given by the **Gibbs factor**, which also factorises over states:

$$p_{\{n_j\}} = \frac{1}{Z_G} \exp\left[\beta\mu \sum_j n_j - \beta \sum_j n_j\epsilon_j\right] = \Pi_j p_{n_j}. \tag{6.22}$$

The last equality is the factorisation over states and p_{n_j} is the probability of finding exactly n_j particles of the system in state j:

$$p_{n_j} = \frac{1}{Z_j} \exp \beta n_j(\mu - \epsilon_j); \quad Z_j = \sum_{n_j} \exp \beta n_j(\mu - \epsilon_j). \tag{6.23}$$

Notice that we are treating the probabilities p_{n_j} as independent; this is consistent with the assumed lack of interaction needed to write the total energy as $n_1\epsilon_1 + n_2\epsilon_2 + \cdots$. It follows that the mean number of particles in a specific state j, with energy ϵ_j, is just

$$\overline{n}_j = \sum_{n_j} n_j p_{n_j} = kT \frac{\partial \ln Z_j}{\partial \mu}, \tag{6.24}$$

where the last step follows because $\partial Z_j/\partial \mu = \sum_{n_j} \beta n_j \exp \beta n_j(\mu - \epsilon_j)$.

6.4 Quantum statistics

In order to perform the sums over allowed occupation numbers we need to consider whether the particles are Bosons or Fermions.

Fermions Fermions have half integer spin (e.g. spin $1/2$) and the exclusion principle limits the possible occupation numbers to $n_j = 0$ or 1. Hence the single-state partition function becomes

$$Z_j = \exp 0 + \exp \beta(\mu - \epsilon_j) = 1 + \exp \beta[\mu - \epsilon_j]. \tag{6.25}$$

Bosons Bosons are those particles with integral spin, and the occupation number can take any nonnegative integer value. Thus the single-state partition function now becomes

$$Z_j = \sum_{n_j=0}^{\infty} \exp \beta n_j[\mu - \epsilon_j]. \tag{6.26}$$

The sum is given by the geometric series

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \qquad \text{for} \quad |x| < 1, \tag{6.27}$$

and thus

$$Z_j = \frac{1}{1 - \exp \beta[\mu - \epsilon_j]}. \tag{6.28}$$

Combining both cases Thus we can write the single-particle partition function for both cases as

$$\boxed{Z_j = \{1 \pm \exp \beta[\mu - \epsilon_j]\}^{\pm 1} \qquad \begin{array}{ll} + & \text{Fermions} \\ - & \text{Bosons} \end{array}} \tag{6.29}$$

We now compute the mean number of particles in state j using (6.24)

$$\begin{align}
\overline{n}_j &= \frac{1}{\beta} \frac{\partial \ln Z_j}{\partial \mu} \tag{6.30} \\
&= \pm \frac{1}{\beta} \frac{\partial}{\partial \mu} \ln[1 \pm \exp \beta[\mu - \epsilon_j]] \tag{6.31} \\
&= \frac{\exp \beta[\mu - \epsilon_j]}{1 \pm \exp \beta[\mu - \epsilon_j]}. \tag{6.32}
\end{align}$$

Finally, we can rearrange slightly to yield

$$\boxed{\overline{n}_j = \frac{1}{\exp \beta[\epsilon_j - \mu] \pm 1} \qquad \begin{array}{ll} + & \text{Fermions} \\ - & \text{Bosons} \end{array}} \tag{6.33}$$

and these are the **Fermi–Dirac** and **Bose–Einstein** distributions for the mean number of particles in a given state j. For the Fermi–Dirac case, this expression can be obtained more directly by summing the appropriate series.

Also we obtain for the grand potential

$$\Phi = -kT \ln Z_{\mathrm{G}} = \mp kT \sum_{j} \ln\{1 \pm \exp \beta[\mu - \epsilon_j]\}, \tag{6.34}$$

and observable macroscopic properties can be deduced from this. Recall the differential relation involving Φ: $d\Phi = -P\,dV - S\,dT - N\,d\mu$, so that

$$P = -\left.\frac{\partial \Phi}{\partial V}\right|_{T,\mu} ; \quad S = -\left.\frac{\partial \Phi}{\partial T}\right|_{V,\mu} ; \quad N = -\left.\frac{\partial \Phi}{\partial \mu}\right|_{V,T}, \tag{6.35}$$

all of which are straightforward, given the form of Z_{G}.

31

6.4.1   The classical limit

The classical limit occurs when mean occupation numbers are small, $\bar{n}_i \ll 1$. We will show that this is guaranteed in the limit of

(1) high T

(2) low particle density

or a suitable combination thereof. There is then no distinction between BE and FD statistics. This is achieved by defining the **dilute limit** in which μ is large and negative wrt kT:

$$\boxed{\exp\left[\beta\mu\right] \ll 1}. \tag{6.36}$$

Provided the single-particle energies are positive, this ensures that

$$\exp\beta[\mu - \epsilon_j] \ll 1 \qquad \forall j\,. \tag{6.37}$$

We can now take the expression (6.34) for Φ, $\mp kT \sum_j \ln\{1 \pm \exp\beta[\mu - \epsilon_j]\}$, and use $\ln(1+x) \simeq x$, to obtain

$$\Phi \simeq \mp kT \sum_j (\pm) \exp\beta[\mu - \epsilon_j] = -kT \sum_j \exp\beta[\mu - \epsilon_j]\,, \tag{6.38}$$

or

$$\Phi = -kT \exp\left[\beta\mu\right] Z(1)\,, \qquad Z(1) = \sum_j e^{-\beta\epsilon_j}\,, \tag{6.39}$$

where $Z(1)$ is the **single-particle partition function**. We fix the chemical potential μ as follows. The mean particle number is given as above by

$$\overline{N} = -\left(\frac{\partial\Phi}{\partial\mu}\right)_{T,V} = \exp\left[\beta\mu\right] Z(1)\,; \tag{6.40}$$

alternatively we can get the same result from

$$\overline{N} = \sum_j \bar{n}_j = \sum_j \frac{1}{e^{\beta(\epsilon_j - \mu)} \pm 1} \simeq \sum_j e^{-\beta(\epsilon_j - \mu)} = e^{\beta\mu} Z(1)\,. \tag{6.41}$$

Thus $\mu = kT \ln(\overline{N}/Z(1))$, and using the dilute limit expression for Φ we have

$$\Phi = -kT\overline{N}\,. \tag{6.42}$$

The Helmholtz free energy becomes

$$F \ = \ \Phi + \mu\overline{N} \tag{6.43}$$

$$= \ -kT\overline{N} + kT\overline{N}\ln\left[\frac{\overline{N}}{Z(1)}\right] \tag{6.44}$$

$$\simeq \ -kT\ln\left[\frac{Z(1)^{\overline{N}}}{\overline{N}!}\right] \tag{6.45}$$

where in the last step we have used Stirling's approximation for large N

$$\ln N! \simeq N \ln N - N\,. \tag{6.46}$$

Therefore we deduce that in the dilute limit

$$\boxed{F = -kT\ln Z\,, \qquad Z = \frac{Z(1)^{\overline{N}}}{\overline{N}!}\,.} \tag{6.47}$$

This result was derived by working from the start with indistinguishable particles, and we see that it automatically recovers the overcounting factor, $N!$, which we introduced as a fix to the classical partition function for distinguishable particles.
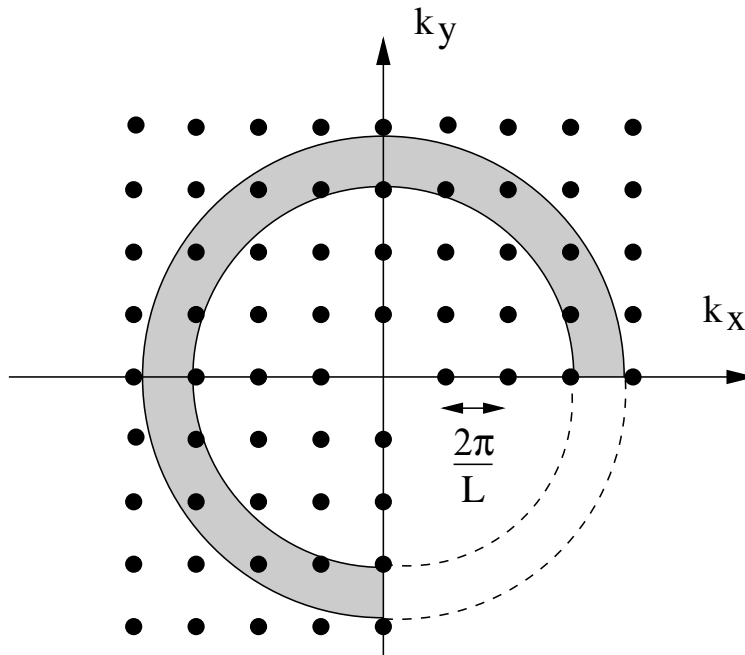
Figure 4: A 2D representation of 'k-space' for periodic boundary conditions. The unit of wavenumber is $2\pi/L$. A shell of radius k and thickness dk is indicated.

6.5 The density of states

All of the above is general, but we now need to consider how the sums over states are actually performed in the quantum case. As with the previous classical discussion of phase space, we will assume that the particles under study are **weakly interacting**, so that a starting point is the quantum mechanics of **free particles**. We will initially assume that these have zero spin. The time independent Schrödinger equation is

$$H\psi = \epsilon\psi \quad \Rightarrow \quad -\frac{\hbar^2}{2m}\nabla^2\psi = \epsilon\psi\,, \tag{6.48}$$

where ϵ is the energy eigenvalue. The natural solutions of this wave equation are plane waves, $\psi \propto \exp[-\underline{k}\cdot\underline{x}]$, each associated with a well-defined momentum, $\underline{p} = \hbar\underline{k}$. We want to discuss the properties of a thermodynamic system consisting of a box of side L, which requires us to introduce boundary conditions. There are two choices: a physical box imposes $\psi = 0$ at the edges, so that the solutions become standing waves and the nice correspondence between wavevector and momentum is broken. For this reason, a more common alternative is **periodic boundary conditions**, so that a particle leaving one wall is reinjected at the other. In effect, we create an infinite system by periodically repeating a single unit of volume L^3. This periodicity requires whole number of oscillations in each direction, so that

$$k_x = n_x\frac{2\pi}{L}, \tag{6.49}$$

etc., where n_x is an integer and $\underline{k} = (k_x, k_y, k_z)$. Allowed modes therefore lie on a mesh in k-space, with a spacing of the fundamental $2\pi/L$, as shown for 2D in Fig. 4. We can therefore count microstates very simply, by considering a region in k-space and asking how many mesh points lie within it. As L is made larger and the mesh spacing becomes finer, this is obviously well approximated by the volume of the region divided by that of the fundamental cell:

$$\boxed{N_{\text{states}} = \left(\frac{L}{2\pi}\right)^3 \int d^3k,} \tag{6.50}$$

where $(L/2\pi)^3$ is the **density of states**; in n dimensions, this would be $(L/2\pi)^n$.

In particular, for a spherical shell of 'radius' k and thickness dk,

$$dN_{\text{states}} = \left(\frac{L}{2\pi}\right)^3 4\pi k^2 \, dk = \frac{L^3}{2\pi^2} k^2 \, dk. \tag{6.51}$$

Now, the energy depends only on the magnitude k of the vector in k-space:

$$\epsilon = \frac{p^2}{2m} = \frac{\hbar^2}{2m}k^2 \quad \Rightarrow \quad d\epsilon = \frac{\hbar^2}{m} k \, dk \quad \Rightarrow \quad \epsilon^{1/2}d\epsilon = \frac{\hbar^3}{2^{1/2}m^{3/2}}k^2 \, dk. \tag{6.52}$$

We can thus rewrite the number of states in terms of energy:

$$\boxed{dN_{\text{states}}/d\epsilon \equiv g(\epsilon) = \left(\frac{2m}{\hbar^2}\right)^{3/2} \frac{V}{4\pi^2} \epsilon^{1/2}.} \tag{6.53}$$

The important features to note here are that the density of states increases with energy as $\epsilon^{1/2}$ and increases linearly with the volume $V = L^3$. It is easy to generalise this to particles with spin, since each wave mode would have a spin degeneracy, g_s, and we simply multiply the density of states by this. For fermions (spin 1/2), this is $g_s = 2$. Photons have spin 1, but because they are massless, they only have two spin states, $g_s = 2$ again, corresponding to left- and right-handed circular polarizations.

It is worth noting that the same result for $g(\epsilon)$ would have been obtained if we had worked with a box with standing-wave boundary conditions. Then we need half-wavelengths across the box, so that the density of states would be 8 times larger, V/π^3. But then k_x etc. would be purely positive and so we would integrate over only the positive octant of k-space. But the periodic approach allows a direct relation to the classical formulation: we have $p = \hbar k$, so that

$$N_{\text{states}} = \frac{V}{(2\pi)^3} \int d^3k = \frac{V}{(2\pi\hbar)^3} \int d^3p. \tag{6.54}$$

Thus the density of states in phase space is $1/(2\pi\hbar)^3$, a result that was anticipated when we wrote the classical result as $1/h^3$. At the classical level, the number h was arbitrary – but we now see that it is indeed the unrationalized Planck's constant, as the notation anticipated.

6.5.1 Ideal quantum gases

We can use the density of states to calculate the thermodynamics of ideal quantum gases: especially black body radiation and the corresponding system for massive particles. As you know, the massive Bose gas is of particular interest in showing the radical phenomenon of Bose-Einstein condensation, which is an example of a phase transition. All of this was covered in some detail in Junior Honours *Statistical Mechanics*, but it is useful to recapitulate some of the main results; we will also extend the treatment to consider massive quanta and to the general **relativistic** relation for $\epsilon(p)$:

$$\epsilon = \sqrt{m^2c^4 + p^2c^2}, \tag{6.55}$$

rather than either the $m = 0$ photon case, or the non-relativistic $\epsilon = p^2/2m$.

The key feature of black-body radiation is that it is not a system of fixed photon number: in reaching equilibrium with the walls of an enclosing box at temperature T, photons can be emitted or absorbed by the walls until equilibrium is reached. The consequence of this is that the **chemical potential** evolves to reach $\mu = 0$. This makes sense, as we saw that the chemical potential only enters the equilibrium maximum-entropy solution when the expected number of

particles is held constant as a constraint; without this, the microstate probabilities are Boltzmann factors, rather than Gibbs factors. When we look at the first law written in terms of free energy, $dF = -SdT - PdV + \mu dN$, we see that F will be stationary at constant T and V if $\mu = 0$. Thus the quantum gas will find its equilibrium by minimising the Helmholtz free energy.

Since the number of particles is proportional to the density of states, which is proportional to V, we see firstly that the number of particles is **extensive** ($\propto V$), and also that the number density, $n = N/V$ (not to be confused with an occupation number) is *independent* of V:

$$n = g \frac{1}{(2\pi\hbar)^3} \int_0^\infty \frac{4\pi\, p^2 dp}{e^{\epsilon(p)/kT} \pm 1}. \tag{6.56}$$

Similar reasoning gives the energy density, u, since it is only necessary to multiply the integrand by a factor $\epsilon(p)$ for the energy in each mode:

$$u = \rho\, c^2 = g \frac{1}{(2\pi\hbar)^3} \int_0^\infty \frac{4\pi\, p^2\, dp}{e^{\epsilon(p)/kT} \pm 1}\, \epsilon(p). \tag{6.57}$$

The pressure and the entropy density can be obtained as follows, using the grand partition function and $\Phi = -kT \ln Z_{\mathrm{G}}$:

$$P = -\left.\frac{\partial \Phi}{\partial V}\right|_{\mu,T} ; \quad S = -\left.\frac{\partial \Phi}{\partial T}\right|_{\mu,V}. \tag{6.58}$$

The grand partition function is

$$Z_{\mathrm{G}} = \prod_{\text{states}} \sum_{n=0}^{n_{\max}} e^{-(\epsilon-\mu)n/kT}$$

$$= \prod_{\text{states}} Z_{\mathrm{G}}(1); \tag{6.59}$$

$$Z_{\mathrm{G}}(1) = \left[1 \pm e^{-(\epsilon-\mu)/kT}\right]^{\pm 1} \quad (+ : \mathrm{F} \quad - : \mathrm{B}).$$

This gives

$$P = -\frac{\Phi}{V} = g \frac{4\pi kT}{(2\pi\hbar)^3} \int p^2\, dp\, \ln Z_{\mathrm{G}}(1)$$

$$s = S/V = g \frac{4\pi k}{(2\pi\hbar)^3} \left[\int p^2\, dp\, \ln Z_{\mathrm{G}}(1) + \int p^2\, dp\, \frac{(\epsilon-\mu)/kT}{e^{(\epsilon-\mu)/kT} \pm 1}\right], \tag{6.60}$$

although in this case we will set $\mu = 0$ in these general expressions.

Black-body radiation For $kT \gg mc^2$ the particles behave as if they were massless, so that $\epsilon = pc$. It is often convenient to work in terms of frequency, so that $h\nu = pc$ and the energy density per unit bandwidth is

$$\boxed{\frac{du}{d\nu} = \frac{4\pi g\, h\, \nu^3/c^3}{\exp[h\nu/kT] - 1},} \tag{6.61}$$

which is the familiar **Planck function**. This can be integrated over frequency, using

$$\int_0^\infty dy\, y^3/(\exp[y] - 1) = \pi^4/15, \tag{6.62}$$

to obtain

$$u = \frac{g\pi^2 k^4}{30 c^3 \hbar^3}\, T^4; \tag{6.63}$$

the famous aT^4 law. Similarly, the total number density scales as T^3:

$$n = \left(\frac{kT}{c}\right)^3 \frac{4\pi g}{(2\pi\hbar)^3} \int_0^\infty \frac{y^2 dy}{e^y \pm 1}. \tag{6.64}$$

35

Given the energy density $u(T)$, we can obtain the entropy density directly from the first law, considering a box of constant volume that is heated up from $T = 0$: $T\,ds = du$. Since $u \propto T^4$, it is easy to perform the integral and obtain

$$s = \frac{4u}{3T}. \tag{6.65}$$

Thus the entropy of black-body radiation also scales as T^3, meaning that the total entropy of the radiation field is, in suitable units, just the total number of photons. We should check that this simple direct derivation of the entropy density agrees with our general expression in equation (6.60). This expression looks rather more complicated, but can be simplified by integrating the first term by parts: $\int dx\, x^2 \ln(1 - \exp[-x]) = (1/3) \int dx\, x^2\, x/(\exp[x] - 1)$. We can then see that both terms are proportional to the one we need for the energy density, and we recover the desired $s = (4/3)u/T$.

Cosmic background radiation   Perhaps the most significant application of these formulae relates to the fact that the universe as a whole is permeated by black-body radiation with a temperature of $2.725\,\mathrm{K}$. This background radiation lies in the microwave band, with the spectrum peaking at wavelengths around $1\,\mathrm{mm}$, and was discovered in 1965. The significance of this radiation is that it shows that the universe was once in a hot and dense state – the **Hot Big Bang**. This conclusion arises because we have known since the work of Vesto Slipher around 1917 that the universe is expanding, with all galaxies moving away from each other. What happens to thermal radiation in this process? If material in the universe is uniform, there can be no heat flow in any direction and the process should be **adiabatic**, so that the entropy remains constant. If we define a **cosmic scale factor**, $a(t)$, which is proportional to the size of the universe and is taken to be unity today, then the entropy density must scale as $s \propto a^{-3}$. But since $s \propto T^3$, we see that the radiation cools as the universe expands, and conversely was extremely hot at early times when the universe was small:

$$T \propto \frac{1}{a(t)}. \tag{6.66}$$

Non-relativistic limit   Here we can neglect the ± 1 in the occupation number, in which case

$$n = e^{-mc^2/kT}(2mkT)^{3/2}\frac{4\pi g}{(2\pi\hbar)^3}\int_0^\infty e^{-y^2}y^2 dy. \tag{6.67}$$

The dominant factor here is the suppression by the exponential factor $e^{-mc^2/kT}$. What is going on here demands a little further explanation, however. We have assumed that the particle numbers can adjust automatically to reach equilibrium – but what physical process permits this? Usually there will be an additional symmetry that requires the existence of antiparticles, which can be created or destroyed in pairs (particularly obvious with charged particles such as $e^\pm$). At high enough T (as in the earliest phases of the expanding universe, near to the big bang), typical thermal energies can create a particle-antiparticle pair, so that the early phases of the universe naturally contain nearly equal numbers of all particle species. But as the universe expands and cools, it becomes thermodynamically preferred for particle-antiparticle pairs to annihilate. But even so there will remain some non-zero **relic** density of such particle pairs. The most natural explanation for the infamous **dark matter** that dominates the matter density of the universe is that it is an incompletely annihilated relic of exotic massive particles, which were once as common as photons.

Gravitational thermodynamics   The total entropy of the universe seems to be dominated by that of the cosmic background. We have seen that this evolves **adiabatically**, with the numbers of photons being conserved. In practice there are $\sim 10^9$ background photons for each proton in the universe, so the entropy of ordinary matter turns out to be rather unimportant.

However, this argument turns out to have neglected a much more important source of entropy: the contribution of **black holes**. For a non-rotating black hole of mass M, there is an **event horizon** of radius $2GM/c^2$. In 1972, Jacob Bekenstein noted that merging a pair of black holes will lead to an area for the event horizon that is larger for the final object than for the sum of the initial areas. He therefore proposed that the area of the event horizon must be a measure of the entropy of the black hole, writing the relation

$$\boxed{\frac{S}{k} = \alpha \frac{A}{L_{\mathrm{P}}^2},}$$
(6.68)

where A is the area of the horizon ($16\pi[GM/c^2]^2$ in the non-rotating case). The coefficient α is an unknown dimensionless number, as must be the case because S/k is dimensionless and because the area has been normalized to the square of the **Planck length**:

$$L_{\mathrm{P}} \equiv \sqrt{\frac{\hbar G}{c^3}}.$$
(6.69)

This is a unit of distance (1.6×10^{-35}m) derived from the fundamental physical constants, and it represents the onset of quantum gravity (the point at which the horizon size $2GM/c^2$ is of order the Compton wavelength of a black hole, $\hbar/Mc$). It therefore seems reasonable that a horizon area of one Planck unit could not have distinct microstates, although the Bekenstein formula requires that the number of microstates is roughly

$$\Omega \sim \exp[A/L_{\mathrm{P}}^2],$$
(6.70)

and the derivation of this formula is an open challenge in theoretical physics. Claims have been made that the corresponding calculation for rapidly rotating black holes can be accomplished in string theory, but the non-rotating case remains unsolved.

If a black hole has an entropy, it also has a temperature:

$$\frac{1}{kT} = \frac{\partial S/k}{\partial E} = \frac{\partial \, 4\pi\alpha(2GM/c^2 L_{\mathrm{P}})^2}{\partial \, Mc^2} = \frac{32\pi\alpha GM}{c^3\hbar}.$$
(6.71)

In 1974, Hawking solved the equations of quantum field theory in the curved spacetime of the black hole to deduce that black holes emitted thermal radiation at the **Hawking temperature**, determining $\alpha = 1/4$ and hence fixing the entropy of a black hole. This entropy is colossal in practical terms. Every galaxy has a black hole at its centre, amounting to perhaps 0.1% of the baryonic material; for the Milky Way, the mass is $4\times10^6 M_\odot = 8\times10^{36}$kg. Our galaxy is reasonably typical, and the spacing between such galaxies is roughly 20 million light years. So we have an entropy of $S/k = 1.7 \times 10^{90}$, and an entropy density of $\rho_S/k = 2.6 \times 10^{20}$m^{-3}. We can compare this with the entropy density from 2.725 K black-body radiation (equation 6.65), which is lower by a factor of about 2×10^{11}. Therefore, to a very good approximation, practical responsibility for the cosmic operation of the second law of thermodynamics has been subcontracted to the population of black holes and their growth through gravitational mergers.

7 Non-equilibrium processes and the arrow of time

7.1 The reversibility paradox

Having set up the microscopic formulation of entropy, it is time to ask if our ideas can account for the second law of thermodynamics: that entropy for an isolated system will never decrease, but rather evolve in the direction of increasing S, until settling down in a maximum-entropy

state of internal thermal equilibrium. As before, we can use the convenient example of **Joule expansion** to illustrate this (Fig. 1): when particles are released (slowly) from the LH side of a box, they spread and eventually become evenly distributed throughout the container. The growing disorder of the particle distribution seems to exemplify the second law and define a clear **arrow of time**, a term due to Eddington.

The problem with this familiar view is that the microscopic physics at work is invariant under time reversal. Classically, each particle obeys Newton's law $m\ddot{\underline{x}} = \underline{F}$, where the acceleration is unchanged under $t \to -t$. The force will also unchanged under this transformation, as can most easily be seen by assuming that it arises from a potential, $F = -\underline{\nabla}\Phi$. Thus if we take a video of the evolution and reverse time (so that all velocities are reversed), what we observe will be a valid physical process that obeys Newton's laws. But then the particles will return to their initial positions and entropy will seem to have decreased. Why is such a violation of the second law not observed?

Less obviously, the same invariance under time reversal also applies to quantum mechanics. Schrödinger's equation is first order in time, but is unchanged under the more complicated transformation

$$t \to -t \quad \text{and} \quad \psi \to \psi^*. \tag{7.1}$$

But replacing ψ by its complex conjugate has no effect on observables (most simply, the probability density $|\psi|^2$ is unaltered).

The need to reconcile this microscopic reversibility with the thermodynamic irreversibility seen empirically has generated a vigorous debate among physicists, which has continued from the time of Boltzmann to the present day. There are two schools of thought, and they propose very different solutions to the problem:

(1) The large number view   states that the second law is not absolute and that it will be routinely violated for small systems. For example, place $N = 2$ particles in the LHS of the Joule box and let them go: you won't have to wait too long before both return to the LHS once again. But for such a coincidence to happen with $N = 10^{24}$ particles spread throughout the full box requires either exquisitely adjusted initial conditions or a period of time much longer than the age of the universe.

(2) The information view   states that the second law applies exactly, but that we have to be more careful with how we calculate the total entropy of system plus observer, and to ask how if or how an empirical violation of the second law could be accomplished (conversion of heat into work).

This divergence of views comes down to the difference between the Boltzmann and Gibbs treatments of entropy, and the following sections contrast these approaches.

7.2   Boltzmann entropy and the H-theorem

H theorem   An immediate puzzle with the second law is to ask how an isolated system can reach thermal equilibrium, which normally we think of in terms of energy exchange with a heat bath. If we initially place a system in some given microstate, what is the origin of the randomness that produces a non-zero probability of occupying a range of microstates at some later time? The answer has to be that the system is still capable of reaching **internal equilibrium** through interactions between its constituent particles. We therefore need to consider how the state of the system evolves under **scattering**.

Boltzmann approached the question of the evolution of entropy by defining the following quantity in terms of the classical distribution function:

$$H = \int f \ln f \, d^3 p. \tag{7.2}$$

There is no integration over space, but Boltzmann was interested in the homogeneous case and focused on the evolution of the velocity distribution. As we saw earlier (equation 6.15), this quantity is equal to the entropy for a classical gas, to within a (negative) multiplicative factor and an additive constant. Because of the change in sign, we can see that the second law will be proved if we can show that H *decreases* monotonically with time for an isolated system. Boltzmann claimed to be able to prove this, together with the statement that $dH/dt = 0$ when f is the Maxwellian (i.e. in thermal equilibrium).

The Boltzmann equation First consider how $f(q, p)$ evolves in the absence of collisions. All particles at a given point in phase space are moving in the same direction, so a cell in phase space transforms to a new cell at some later time, and all the particles in the initial cell remain within the new one. In fact, the volume of these two cells are the same, so that the phase space density in the new cell does not change with time. This is trivial to see in the case of zero applied force: a square in $x - p_x$ space changes to a parallelogram, with no change in area. In the presence of forces, we need to be more careful. When conserving particles, we must be dealing with the continuity equation, which in normal 3D space looks as follows:

$$\frac{\partial}{\partial t} \rho = -\underline{\nabla} \cdot \underline{j} = -\underline{\nabla} \cdot \rho \underline{v}. \tag{7.3}$$

Here, $\rho \to f$, and our 6D velocity is $(\underline{\dot{q}}, \underline{\dot{p}})$, so that

$$\frac{\partial}{\partial t} f + \underline{\nabla}_q \cdot (f\underline{\dot{q}}) + \underline{\nabla}_p \cdot (f\underline{\dot{p}}) = 0. \tag{7.4}$$

But when we expand the derivatives, two terms cancel provided we assume that the dynamics are described in terms of a **Hamiltonian**. In that case, we have **Hamilton's equations**

$$\dot{q} = \frac{\partial H}{\partial p}; \quad \dot{p} = -\frac{\partial H}{\partial q}, \tag{7.5}$$

and $\partial\dot{q}/\partial q + \partial\dot{p}/\partial p$ vanishes through the symmetry of 2nd partial derivatives. We therefore obtain the **collisionless Boltzmann equation**,

$$\frac{\partial}{\partial t} f + \underline{\dot{q}} \cdot \underline{\nabla}_q f + \underline{\dot{p}} \cdot \underline{\nabla}_p f = 0. \tag{7.6}$$

We recognise the LHS here as just the total time derivative, df/dt, familiar in 3D in the form of the **convective derivative**, $\partial/\partial t + \underline{v} \cdot \underline{\nabla} = 0$ (which just says that we see things change as we move either because they really change with time, or because we move to a new location where conditions are different). Finally, the full Boltzmann equation has to allow for collisions, which can scatter particles out of a given moving 6D volume element and change f:

$$\frac{\partial}{\partial t} f + \underline{\dot{q}} \cdot \underline{\nabla}_q f + \underline{\dot{p}} \cdot \underline{\nabla}_p f = \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}. \tag{7.7}$$

Moments and fluid equations The Boltzmann equation is an important general tool for discussing the evolution of statistical systems of classical or quantum particles: e.g. the neutron flux in atomic bombs and reactors, or the evolution of photons scattered by free electrons, as manifested in the Cosmic Microwave Background. A description in terms of a distribution function

goes beyond the familiar case of a fluid treatment, in which material in a small volume element has a single common velocity. At first sight, it would therefore seem incorrect to treat even an ideal gas as a fluid, because molecules at a given location have a distribution of velocities, travelling in different directions. But it turns out that appropriate averaging of the Boltzmann equation yields the equations of fluid mechanics even in such cases. To do this, we will write the momentum c in terms of a particle velocity and a mass: $\underline{p} = m\underline{\dot{q}} = m\underline{u}$. It is also convenient to write the Boltzmann equation in its original form, (7.4), with a collision term on the RHS:

$$\frac{\partial}{\partial t} f + \underline{\nabla}_q \cdot (f\underline{u}) + \underline{\nabla}_p \cdot (f\underline{\dot{p}}) = \dot{f}_{\text{coll}}. \tag{7.8}$$

Now integrate this equation over momentum, and define the number density, n, and the mean velocity, $\underline{v} \equiv \langle \underline{u} \rangle$:

$$n = \int f \, d^3p; \qquad \underline{v} = \frac{1}{n} \int f \, \underline{u} \, d^3p. \tag{7.9}$$

The first two terms in the Boltzmann equation become $\dot{n}$ and $\underline{\nabla} \cdot (n\underline{v})$. The third term vanishes through the divergence theorem, and assuming $f \to 0$ at infinite momentum. If the collision term on the RHS represents only scattering, then it does not create or destroy particles and so must average to zero (although this would not be so if the collisions involved annihilations). We then get the fluid **continuity equation**

$$\boxed{\dot{n} + \underline{\nabla} \cdot (n\underline{v}) = 0.} \tag{7.10}$$

This is exactly what we would expect if all particles travelled at the same mean velocity, so any dispersion in velocities doesn't enter. Other equations of interest can be obtained by taking different moments of the Boltzmann equation. In particular, suppose we multiply the equation by $\underline{u}$ before integrating over momentum. The first term gives $\partial(n\underline{v})/\partial t$. For the second, it is better to consider the ith component, which has the tensor form $\nabla_j n\langle u_i u_j \rangle$. For the third term, we have $p_i(\partial/\partial p_j)(f a_j)$, where $\underline{a}$ is the acceleration. Now, we can rewrite the derivative using

$$p_i \frac{\partial}{\partial p_j} f = \frac{\partial}{\partial p_j}(f p_i) - f \delta_{ij}. \tag{7.11}$$

The first term here integrates to zero, as before, leaving $-n a_i$. Lastly, the collision term conserves momentum and integrates to zero, so the velocity moment of the Boltzmann equation is

$$\frac{\partial(n v_i)}{\partial t} + \nabla_j n\langle u_i u_j \rangle = n a_i. \tag{7.12}$$

If we write $u_i = v_i + w_i$, then $\langle u_i u_j \rangle = v_i v_j + \langle w_i w_j \rangle$. In principle the latter term has off-diagonal terms, but we can consider the simplest case of isotropic velocities, for which $m\langle w_i w_j \rangle = P\delta_{ij}$: pressure just reflects an internal dispersion of velocities. Then we have

$$\frac{\partial(\rho v_i)}{\partial t} + \nabla_j(\rho v_i v_j) + \nabla_i P = \rho a_i \tag{7.13}$$

(multiplying through by m and writing the mass density as $\rho = mn$). Finally, we can use continuity to eliminate derivatives of the density, and write everything back in vector notation. The result is the **Euler equation**, showing how the change of mean velocity is driven by pressure gradients and external forces:

$$\boxed{\frac{\partial \underline{v}}{\partial t} + (\underline{v}\cdot\underline{\nabla})\,\underline{v} = -\frac{1}{\rho}\underline{\nabla}P + \underline{a}.} \tag{7.14}$$

This equation can be given a more direct and intuitive derivation but it is nevertheless pleasing to see that it emerges directly from the Boltzmann formalism.

Collisions and the H theorem Boltzmann approached the collision term by enunciating the hypothesis of **Stosszahlansatz** or **molecular chaos**, in which the velocities of colliding particles are uncorrelated, and statistically independent of position. From this, Boltzmann was able to prove the **H theorem**, which states that the statistical entropy will indeed increase. The proof is regrettably fairly long and detailed: see e.g. chapters 3 & 4 of Huang's *Statistical Mechanics*. In the end, skipping this derivation is not a catastrophe as we will develop better methods for understanding entropy at the microscopic level – although it is important for understanding the development of the subject to know about Boltzmann's approach. And indeed the H theorem initially does seem satisfying, as we know that the macroscopic **experimental entropy** will increase, so can the statistical Boltzmann entropy really be considered a theoretical calculation of this empirical quantity?

Arrow of time The H theorem has been severely criticised, starting in Boltzmann's own time. As shown earlier, the classical particle dynamics that he considered are invariant under time reversal, so it is apparently impossible to have a systematic increase (or decrease) of entropy (**Loschmidt's paradox**). The weak point seems to be the Stosszahlansatz, which in effect smuggles randomness into the initial conditions, thus forbidding the sort of fine-tuning that would allow something like the Joule expansion to be reversed.

The modern view of Boltzmann's work is therefore that causal microscopic processes alone cannot assure an entropic arrow of time, and that the experimental increase of entropy in isolated systems must be traced to the nature of the initial conditions. These are not only below the maximum possible entropy, but are also presumed not to be chosen specially. The latter point is important: otherwise, we could set up the Joule expansion with all particles moving to the left, so that the gas would spontaneously compress and reduce its classical entropy. But in reality the initial condition for the gas in the LH side of the box will be thermal equilibrium, and such an initial fine-tuned microstate will be unlikely. Similarly, if we pick a set of time-reversed particles in the full box, these will indeed end up in the left hand region. But again this occurs for that *particular* microstate we have prepared. For the overwhelming majority of microstates associated with the initial macrostate, the time evolution will be as expected i.e. there will be only small fluctuations in the number of particles in the left hand side. Nevertheless, from this point of view it does seem inevitable that the Boltzmann entropy could decrease. This can be proved more formally via the **Poincaré recurrence** theorem: after a sufficient time, any dynamical system will reach a point in phase space that is arbitrarily close to its initial state (see e.g. Huang Chapter 4). As a simple estimate, imagine a box of gas with sufficiently high density that the mean free path is small compared to the side of the box, L. In that case, by the time the particle has travelled a distance L (made 'one crossing'), it will be effectively impossible to predict which side of the box it lies in. The probability that all particles are in one side is therefore $(1/2)^N$, which is impressively small for $N \sim 10^{24}$. But after sufficient crossings, a rare event will happen:

$$\text{Poincaré recurrence time} \sim 1/\text{prob} \sim 2^{10^{24}} \text{ crossings} \gg \text{age of universe.} \qquad (7.15)$$

So for practical purposes, this can be neglected. Nevertheless, in this picture one apparently has to accept that the second law is purely statistical: it will routinely be violated in small-N systems, and we are saved from seeing this in the laboratory only because we work closer to $N \sim 10^{24}$.

7.3 The Gibbs view on the arrow of time

The Gibbsian view of all this is very different. The distribution function used by Boltzmann gives only limited microscopic information: the phase-space density is the *probability* of finding a particle in a given phase-space cell, so it reflects a lack of knowledge of precisely where the particles making up a gas actually lie, or of what their velocities are. But if we use Gibbs's $6N$-dimensional phase space, then the system is represented by a single point: that corresponds

to a single microstate, which will then have $S = 0$, independent of how that point evolves under scattering.

In the Gibbs picture, then, we are forced to be more explicit about our lack of knowledge of the microstate, even though this aspect is also implicit in the Boltzmann approach. This is clear from the appearance of probabilities in the general expression for the Gibbs entropy, $S = -k \sum_i p_i \ln p_i$. In the context of phase space, the microstate is specified by the $6N$-dimensional vector

$$\underline{X} = (q_1, \ldots, q_{3N}, p_1, \ldots, p_{3N}), \tag{7.16}$$

and so we need to consider a **probability density** in this space:

$$d\text{prob} = \rho(\underline{X}, t) \, d\Gamma; \qquad d\Gamma \equiv \prod_{i=1}^{3N} dq_i \, dp_i \equiv d^{3N}q \, d^{3N}p. \tag{7.17}$$

Thus ρ keeps track of the **ensemble** of possible independent systems, each moving separately through phase space. The 'mass' in this fluid of points is conserved, since the total probability, $\int \rho(\underline{X}, t) \, d\Gamma$, is unity. The Gibbs entropy can then be written in terms of ρ:

$$S = -k \int \rho \ln \rho \, d\Gamma + \text{constant}. \tag{7.18}$$

The Gibbs entropy is problematic for an isolated system. Suppose we place the system in one specific initial microstate, so that $S = 0$, and ask how it will evolve. At least at the classical level, the particles obey causal dynamics: it is in principle possible to calculate the effects of collisions. Thus the system always stays in a single microstate and so the entropy always remains zero: the Gibbs entropy is a **constant of the motion** and apparently does not allow for a second law. Choosing a single initial microstate is unrealistic, and instead we should have a set of $\{p_i\}$ conditional on some macroscopic information, but this doesn't change the argument: each initial microstate evolves to a unique new one, with the same p_i, so the sum for the entropy is unchanged. Things would be different if the non-equilibrium system was placed in contact with a heat bath, as then the $p = 1$ for a single microstate would be progressively reduced and the probability of occupying other states would grow, evolving towards the maximum-entropy state that we know gives probabilities based on Boltzmann factors. But without this external randomness, how can the p_i change?

Pseudorandomness and causality   It is at this point that the Bayesian view of probability as a degree of belief or knowledge can be useful, since we can ask how well the prediction of the future microstate would work in practice. Even for the simplest case of a single particle bouncing off elastic walls in x alone, guaranteeing that it will be in the LH side of the box at a time t in the future requires a precision in velocity $\delta v < L_{\text{box}}/t$, which in due course becomes arbitrarily small. But things will be much worse with collisions, where the particle dynamics will be **chaotic**, so that a small uncertainty in initial conditions becomes magnified exponentially with time (as is appreciated by anyone who has ever mis-hit a snooker ball very slightly). Thus the future evolution of the system rapidly becomes in effect unpredictable – and this is really what we mean by randomness. Every computer includes a random number generator, which is actually a causal algorithm. But in practice these **pseudorandoms** are indistinguishable from the real thing. We can therefore see that Boltzmann was perhaps being reasonable in arguing that collisions can lead to random thermal equilibrium, with an associated increase of entropy.

This gives us a Bayesian interpretation of the second law in terms of information. The Gibbs entropy is of the Shannon form and represents the **hidden information** about the system. It should be clear that our knowledge about the system can only decline with time, if we make no further measurements. As our confidence in predicting the future microstate erodes, the entropy
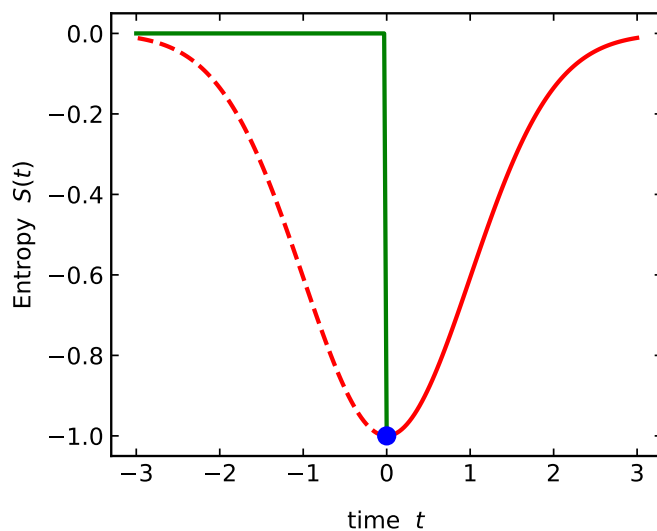
Figure 5: Suppose a system is in a state of below-maximum entropy at $t = 0$, represented by the blue point. The 2nd law states that the entropy will increase until it reaches the maximum possible value, represented by the red curve. One might argue that time-reversibility of the equations of motion would imply the reflection of this behaviour, so that S declined with increasing time at $t < 0$. But (at least for large systems), the low-entropy point is an initial condition that we impose by interfering with the system, so the actual $S(t)$ up to this point follows the green curve. The breaking of time symmetry then reflects causality.

increases. Here, the origin of the arrow of time is very clear: it is **causality**. We generate some information about the system by imposing an initial condition at time t_i, and the effects of this propagate to $t > t_i$, albeit with decreasing certainty: see Fig. 5. But the state of the system at $t < t_i$ cannot be influenced at all by what we learn about the system in the future.

Demons and information We can illuminate this further by considering **Maxwell's demon** of 1867. Maxwell had the idea of a microscopic creature that could open or close a flap to ensure that particles only passed in one direction through the middle of the box, allowing entropy to decrease. Actually, Maxwell's original argument was based on segregating fast and slowly-moving particles, so that the two sides of the box would attain different temperatures; but in either case, there is the potential to turn an equilibrium situation into one of lower experimental entropy. In the case of temperature segregation, we could certainly run a heat engine to extract work, thus turning some of the initial heat in the box purely into work – which is an explicit violation of one form of the second law. The resolution of the demon paradox is now agreed: an entropy increase has to occur in the brain of the demon corresponding to the information they process. As argued by Landauer in 1961, taking one bit of computer memory with a definite value and then erasing it ready for the next measurement involves an entropy change $k \ln 2$ (so that there is an irreducible heat generation of $kT \ln 2$ for each binary operation). Allowing for this computational entropy generation, the total universe of (box + demon) does not undergo an entropy decrease. But if we wait long enough, the effect produced by the demon will occur spontaneously: so does entropy decrease in that case? That would be so if we could guarantee to extract work from the system. However any L-R temperature difference will fluctuate, so we would need to monitor the system continuously in order to know when L is hotter than R and that our hypothetical heat engine should be connected. This change in the information stored in our brains cannot be neglected: we are Maxwell's demon.

This approach based on information is utterly different from the Boltzmann view, in which the statistical definition of entropy is not guaranteed to increase, and where the 2nd law is something

that tends to hold approximately in the limit of large numbers, but where the entropy can fluctuate up and down. But the Gibbsian approach delivers a much more definitive conclusion:

> The second law is a true law of physics, which applies without exception, not just 'most of the time'.

We now explore the evolution of the Gibbs entropy in more detail, firstly in the classical regime, and then via a quantum treatment, in order to understand more fully how this promise is fulfilled.

7.4   Entropy and phase-space dynamics

We will discuss the Gibbs entropy in the context of its $6N$-dimensional full **phase space**, which is spanned by the p_i and q_i co-ordinates of all particles in the system. As in our discussion of the Boltzmann equation, the dynamics in full phase space is most naturally described by **Hamiltonian dynamics**, since this concerns the changes in the coordinates, q_i and their **conjugate momenta**, p_i. Hamilton's equations read

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \tag{7.19}$$

For a system of point particles of mass m interacting via a potential $U(\{q\}) = U(\{\underline{r}\})$, the **generalised** momentum and position coordinates can be chosen as ordinary Cartesian ones. The **Hamiltonian** is then the total energy, written in terms of the p's and q's as

$$H = \sum_i \frac{p_i^2}{2m} + U(\{q_i\}) = E, \tag{7.20}$$

so that Hamilton's equations have the familiar Newtonian form:

$$\dot{q}_i = \frac{p_i}{m}, \qquad \dot{p}_i = -\frac{\partial U}{\partial q_i}. \tag{7.21}$$

7.4.1   Liouville's theorem

Liouville's theorem states the important result that the phase-space fluid is **incompressible**. We have previously seen the key elements of the proof in discussing the 6D Boltzmann equation. First define the $6N$-dimensional **phase space velocity** of a point $\underline{X}$ as

$$\underline{V} = \underline{\dot{X}} = (\dot{q}_1, \ldots, \dot{q}_{3N}, \dot{p}_1, \ldots, \dot{p}_{3N}). \tag{7.22}$$

In these terms, the conserved nature of total probability must require a **continuity equation** in phase space. The flux density of probability will be just $\rho \underline{V}$, so the continuity equation is

$$\frac{\partial \rho}{\partial t} = -\underline{\nabla} \cdot (\underline{V}\rho) \equiv -\sum_{i=1}^{3N} \left[\frac{\partial}{\partial q_i}(\dot{q}_i \rho) + \frac{\partial}{\partial p_i}(\dot{p}_i \rho) \right]. \tag{7.23}$$

Expanding the derivatives gives

$$-\frac{\partial \rho}{\partial t} = \underline{V} \cdot \underline{\nabla}\rho + \rho \underline{\nabla} \cdot \underline{V} \equiv \sum_{i=1}^{3N} \left[\dot{q}_i \frac{\partial \rho}{\partial q_i} + \dot{p}_i \frac{\partial \rho}{\partial p_i} \right] + \rho \sum_{i=1}^{3N} \left[\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right]. \tag{7.24}$$

But from Hamilton's equations the second pair of terms on the RHS cancel, so that

$$\underline{\nabla} \cdot \underline{V} = \sum_{i=1}^{3N} \left[\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right] = \sum_{i=1}^{3N} \left[\frac{\partial}{\partial q_i} \left(\frac{\partial H}{\partial p_i} \right) + \frac{\partial}{\partial p_i} \left(-\frac{\partial H}{\partial q_i} \right) \right] = 0 \,, \tag{7.25}$$

and hence we have

$$-\frac{\partial \rho}{\partial t} = \underline{V} \cdot \underline{\nabla} \rho = \sum_{i=1}^{3N} \left[\dot{q}_i \frac{\partial \rho}{\partial q_i} + \dot{p}_i \frac{\partial \rho}{\partial p_i} \right] \,, \tag{7.26}$$

As d/dt, the time derivative, is given by

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \sum_{i=1}^{3N} \left[\frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right] = \frac{\partial \rho}{\partial t} + \underline{V} \cdot \underline{\nabla} \rho \,, \tag{7.27}$$

then we have

$$\boxed{\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \underline{V} \cdot \underline{\nabla} \rho = 0 \,.} \tag{7.28}$$

It is apparent that d/dt, the total time derivative, is a **convective** time derivative that moves with the fluid of phase-space points. Thus we have proved that the fluid is **incompressible**:

> if we follow the trajectory of any phase-space point, the density at its location, ρ, will not change with time.

More generally (see tutorial) we can show that any function u of the canonical variables $\{q_i, p_i\}$ satisfies

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + [u, H] \,, \tag{7.29}$$

and so we can write Liouville's equation in the form

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + [\rho, H] = 0 \,, \tag{7.30}$$

where we define the **Poisson bracket** of functions u and v of the canonical variables $\{q_i, p_i\}$ as

$$[u, v] = \sum_i \left(\frac{\partial u}{\partial q_i} \frac{\partial v}{\partial p_i} - \frac{\partial u}{\partial p_i} \frac{\partial v}{\partial q_i} \right) \,, \tag{7.31}$$

(see tutorial sheet). Finally we note that to have a *stationary* distribution

$$\frac{\partial \rho}{\partial t} = 0 \,, \tag{7.32}$$

which is a necessary condition for equilibrium, we require that $[\rho, H] = 0$. Choosing $\rho = $ const. or $\rho = \rho(H)$ recovers the microcanonical and canonical ensembles respectively.

7.4.2 Entropy and coarse graining

We now write the Gibbs entropy in the language of phase-space density. We need $S_{\text{Gibbs}} = -k \sum p_i \ln p_i$, where now the probability is to be thought of as the probability of occupying a cell of volume $\delta\Gamma$: $p_i \to \rho\,\delta\Gamma$. For small $\delta\Gamma$, the sum approaches an integral:

$$S_{\text{Gibbs}} = -k \int d\Gamma\, \rho \ln \rho + \text{const.}, \tag{7.33}$$

where the integral is over all phase space. The constant is not well defined, but it is independent of ρ. If we take S_{Gibbs} as our definition of S, then it follows from Liouville's theorem that

$$\frac{dS_{\text{Gibbs}}}{dt} = -k \frac{d}{dt} \int d\Gamma\, \rho \ln \rho = 0 \tag{7.34}$$

(see tutorial). Thus the classical Gibbs entropy is constant in time, rather than increasing, as we argued more directly above, using an argument based on discrete microstates. We should however note the subtlety that we have implicitly assumed that the dynamical trajectories move one initial phase-space point to another in a **1-to-1 mapping**: phase-space volume elements distort their shape, but do not overlap. This may not seem obvious, but assume that two distinct phase-space points do evolve to the same point at some later time. Then the time-reversibility of the dynamics would be violated, as there can only be a single time-reversed trajectory from that point.

This constancy of S_{Gibbs} applies for any initial ρ, and not just a single microstate, or very close to one. A particular example would be a system in thermal equilibrium, where the microstate probabilities are given by Boltzmann factors. We know that this is a maximum entropy state, which will not change under dynamical evolution. So a fixed Gibbs entropy is not a problem there. But our main interest is is in **non-equilibrium initial conditions**, and we need to know the initial microstate probabilities **conditional** on the macroscopic constraints we have applied. In practice, we can see that this problem only has a clear solution if we set up initial macroscopic systems that are **inhomogeneous** mixtures of subsystems that are locally in equilibrium: Joule expansion is such an example, as is the case of a box of gas with equal mass at different temperatures, T_1 and T_2, in each half. In the latter case, all of phase space has a non-zero probability, but the Boltzmann factors are different in each half (see Fig. 6). Classically, this system will evolve towards equilibrium at $T = (T_1 + T_2)/2$, with microstate probabilities that are Boltzmann factors corresponding to T. But since this has a higher entropy that the original state, such an end point is forbidden by Liouville's theorem. In the end, the failure of the Gibbs entropy to increase is less about that entropy definition, and more a reflection of the fundamental puzzle that Hamiltonian-based interactions in an isolated system **cannot create true thermal equilibrium**, with Boltzmann microstate probabilities. But if we can solve this general problem, then the Gibbs entropy will naturally increase.

Coarse graining One way of understanding the behaviour here is to focus on the *shape* of elements in phase space. As a consequence of Liouville's theorem, these must maintain their volume (the density doesn't change, and total probability must remain at unity). But at the same time, different points in phase space will tend to move apart. The way in which these requirements are reconciled is that volume elements distort into complicated regions of extreme 'threadiness', as depicted in Fig. 7. This is very much like stirring ink into water: in principle, there is always ink in one place and water in another, but after a while the ink appears to have a uniform density, lower than the original one. Thus in practice the Gibbs entropy will be redefined as

$$\bar{S} = -k \int d\Gamma\, \bar{\rho} \ln \bar{\rho}, \tag{7.35}$$

where $\bar{\rho}$ is a **coarse grained** probability density, defined by averaging ρ over some fixed local scale Λ in phase space.

Concavity of entropy   The coarse-graining procedure smooths out the local probability density, and we can prove that this can only increase $\bar{S}$. To see this, note that the function $s(\rho) = -k\rho \ln \rho$ is **concave**, so that the function of the average is $\geq$ the average of the function: $s(\lambda a + (1-\lambda)b) \geq \lambda s(a) + (1-\lambda)s(b)$ for $\lambda \in [1,0]$. For a convex function the opposite is true. The truth of this statement can be grasped geometrically, as shown in Fig. 8.

This is all very well, but what sets the coarse-graining scale? Surely we can always look with sufficient care than we can recover the true fine-grained distribution function? This is true, but recall that all of this effort was invested in order to find a microscopic quantity that would serve as a proxy for the experimental entropy measured in a laboratory. This latter quantity uses only macroscopic properties of a system: from this limited information, we would have no reason to believe that there were fine-scale features in the phase-space distribution. And however much information we have, the distortion of volume elements becomes more extreme with time, and the scale of fine structure in the distribution will eventually fall below any phase-space resolution that we impose. Ultimately, then, the final state of an evolved isolated system can no longer be distinguished from true thermal equilibrium.

The linkage between coarse-graining and entropy increase is an illustration of Shannon's view relating entropy to hidden information. In the coarse-graining process, we do explicitly lose information: only by doing so is the Gibbs entropy allowed to increase. Therefore **the increase of entropy is linked to our knowledge about the system**, rather than anything it is doing internally, in a manner that may appear questionable. Can it really be that macroscopic and reproducible phenomena, such as heat flow, depend on how we handle information? Perhaps yes, since the division between work and heat is somewhat arbitrary. Were we able to track all the particle positions, there would be no need to talk about heat energy, or heat flow.

Coarse-graining and time asymmetry   This lack of detailed microscopic knowledge manifests itself in the breaking of the time symmetry that should be present in the underlying dynamics. The result is some familiar equations of physics that break time-reversal symmetry. One of these is the diffusion equation, discussed later in the course. As a further example of coarse-graining leading to irreversibility, consider Newton's law with a viscous term:

$$m\ddot{\underline{r}}_i + \kappa\dot{\underline{r}}_i = -\underline{\nabla}_i U(\underline{r}) \,. \tag{7.36}$$

The drag term dissipates energy in an irreversible way, and the presence of the first-order time derivative makes it mathematically clear that the system does not have time-reversal symmetry. But the viscous drag on a particle is a result of complicated processes involving molecular collisions and in principle we could write out all these other processes using microscopic time-reversible equations (and keeping energy conserved). The viscous term in the equation approximates microscopic reversible processes by a coarse-grained description.

8   Quantum statistical mechanics and mixed states

All of the above discussion of the arrow of time was classical, and we might wonder if a quantum treatment can help resolve some of the puzzles associated with the 2nd Law. In this section we therefore discuss some general aspects of how quantum mechanics has to be adapted to deal with statistical ensembles.

8.1 Many-particle Schrödinger equation

Consider a system of N particles of mass m. (For simplicity we suppress any spin index.) The time-independent Schrödinger equation reads

$$H\Psi(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N) = E\Psi(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N), \tag{8.1}$$

where E is the energy of the eigenstate $\Psi(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N)$. In general, the **Hamiltonian** operator is

$$H = -\sum_{k=1}^{N} \frac{\hbar^2}{2m} \nabla_k^2 + U(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N), \tag{8.2}$$

where k labels the particles. The first term is kinetic energy and the second term, U, is the **interaction potential** which depends on the relative positions of all the particles e.g. we could have electrostatic Coulomb interaction between particles with charge e. That would be an example of a '2-body' interaction i.e. pairwise interactions between particles.

Weakly interacting case   One case where we can solve (8.1) is when the potential energy term contains no interaction terms, i.e. when

$$U(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N) = \sum_{k=1}^{N} V(\underline{r}_k). \tag{8.3}$$

Then we can write the Hamiltonian as

$$H(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N) = \sum_{k=1}^{N} h(\underline{r}_k), \qquad h(\underline{r}) = -\frac{\hbar^2}{2m} \nabla^2 + V(\underline{r}). \tag{8.4}$$

The Hamiltonian h has wavefunctions ψ_1, ψ_2, ... and corresponding energies ϵ_1, ϵ_2, ... common for all the particles. Thus the many-body eigenfunctions are factorised

$$\Psi(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N) = \prod_{k=1}^{N} \psi_{\alpha_k}(\underline{r}_k) \qquad \text{with} \quad E = \sum_{k=1}^{N} \epsilon_{\alpha_k}. \tag{8.5}$$

where $\alpha_1, \alpha_2, \ldots$ are a collection of (integer) labels for the wavefunctions and energies.

Spin and symmetry   This factorised wavefunction (8.5) applies for **distinguishable** particles. In order to impose **indistinguishability** we should take suitably symmetrised combinations of these eigenfunctions, depending on whether the spin is integral (Bosons) or half-integral (Fermions). For Fermions the wavefunction is antisymmetric

$$\Psi(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N) \propto \sum_{P} \epsilon(P) \psi_{\alpha_{P_1}}(\underline{r}_1) \ldots \psi_{\alpha_{P_N}}(\underline{r}_N), \tag{8.6}$$

where P is a permutation of $1, \ldots, N$; $\epsilon(P) = \pm$ depending on whether we have an even or odd permutation (not to be confused with energy). For example for two particles we have

$$\Psi = \frac{1}{\sqrt{2}} \left(\psi_{\alpha_1}(\underline{r}_1)\psi_{\alpha_2}(\underline{r}_2) - \psi_{\alpha_2}(\underline{r}_1)\psi_{\alpha_1}(\underline{r}_2) \right). \tag{8.7}$$

If two particles try to occupy the same state, then $\psi_{\alpha_1} = \psi_{\alpha_2}$ and Ψ vanishes. For Bosons the wavefunction is similarly a symmetric combination of all permutations P

$$\Psi(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N) \propto \sum_{P} \psi_{\alpha_{P_1}}(\underline{r}_1) \ldots \psi_{\alpha_{P_N}}(\underline{r}_N). \tag{8.8}$$

Dirac notation Here and below it is convenient to use Dirac's **bra-ket notation**. This is a shorthand motivated by the matrix element for an operator, and emphasising the similarity to an inner product:

$$\int \psi_1^*(\mathbf{x}) \, A \, \psi_2(\mathbf{x}) \, d^3x \qquad \longrightarrow \qquad \langle 1|A|2 \rangle. \tag{8.9}$$

i.e. a **bracket** is made from a **bra** $\langle \psi |$ and a **ket** $|\psi\rangle$. This is like a scalar product involving a row vector and a column vector: the bra and ket are alternative and equally valid representations of the same state vector. We write eigenstates of the Hamiltonian (or any Hermitian operator) as $H|i\rangle = E_i|i\rangle$; it can be shown that the eigenstates are orthonormal:

$$\langle i|j \rangle = \delta_{i,j} \ . \tag{8.10}$$

They also form a complete set of states so that any state function can be expanded in eigenstates:

$$|\psi\rangle = \sum_i c_i |i\rangle = \sum_i |i\rangle \langle i|\psi\rangle \tag{8.11}$$

(using orthonormality). We can write this equation using the general **completeness relation**:

$$\boxed{\sum_i |i\rangle \langle i| = 1.} \tag{8.12}$$

To compute the expectation value of some observable corresponding to a Hermitian operator A we use

$$\langle A \rangle = \langle \psi|A|\psi\rangle = \sum_i \langle i| \, c_i^* \, A \sum_j c_j|j\rangle = \sum_{ij} c_i^* c_j A_{ij} \tag{8.13}$$

where A_{ij} are the matrix elements of operator A in the energy eigenbasis. So for example if A is H, H will be diagonal in the energy eigenbasis

$$\langle i|H|j \rangle = E_j \delta_{ij} \tag{8.14}$$

and

$$\langle H \rangle = \sum_i |c_i|^2 E_i \tag{8.15}$$

8.2 Quantum mechanics of non-isolated systems: the density matrix

In the above, we have assumed that a wave function description is always applicable – *but this is not always true*. An isolated system may be described by a wave function, but a sub-part of a single system will not be so described unless the overall wave function factorises. If x and y are the coordinates of two particles, A & B, and $\psi(x, y) = \psi_A(x)\psi_B(y)$, then there is a well-defined wave function for A, independent of what B is doing – but this is not so if the wave function fails to factorise. In other words, the overall system of Schrödinger's cat plus killing apparatus may be described by a wave function, but the cat is not so easily disentangled from its environment. Similarly, a system in thermal contact with a heat bath cannot be described by a wave function. In such cases, we have an **ensemble of states**. Simple quantum mechanics assumes that the state of the system is $|\psi\rangle$, a single Hilbert-space vector which we refer to as a **pure state**. A pure state can be in the standard form of a **linear superposition**, $|\psi\rangle = \sum_i a_i|i\rangle$, where the probability of an experiment finding the system in an eigenstate $|i\rangle$ is $p_i = |a_i|^2$. But in an ensemble, we consider something else: a **mixed state** where all properties of the system reflect the system being **prepared** in state $|i\rangle$, with **classical probability** p_i. These may sound like the same thing, but a pure state can display **quantum interference**, whereas a mixed

state will not. The difference can be seen when we take the expectation value of an operator, A, that is not diagonal in the eigenstates:

$$\text{pure state}: \quad \langle A \rangle = \sum_i \sum_j a_i^* a_j \langle i|A|j \rangle$$

$$\text{mixed state}: \quad \langle A \rangle = \sum_i p_i \langle i|A|i \rangle. \tag{8.16}$$

We need to satisfy the latter relation for *any* operator A, and this is not possible with a wave-function description.

The general concept that handles the quantum mechanics of **composite systems** is the **density matrix** (which could do with a better and more informative name), invented by von Neumann in 1927. To describe simply what this is, suppose that a given system has eigenstates $|i\rangle$ corresponding to some operator. If this system is described by a wave function, this will be expandable in terms of the eigenstates: $|\psi\rangle = \sum_i a_i |i\rangle$, where $a_i = \langle i|\psi \rangle$. We can therefore introduce the **projection operator** $|i\rangle\langle i|$, which operates on $|\psi\rangle$ to yield its ith 'component': $|i\rangle\langle i| \, |\psi\rangle = a_i |i\rangle$. Now, when a non-isolated system is observed, the resulting measurement will still be one of the eigenvalues of the operator, and there will be some probability p_i that the system will be found in the corresponding eigenstate. The density operator is the weighted sum of the projectors over all possible states:

$$\rho = \sum_i p_i \, |i\rangle\langle i|. \tag{8.17}$$

Traces and expectations Now, ρ can be represented by a matrix,

$$\rho_{ij} \equiv \langle i|\rho|j \rangle, \tag{8.18}$$

and the way to find the probability of the system being found in a state $|j\rangle$ is use the **trace** (where the **trace** of any matrix is the sum of the diagonal elements, $\text{Tr} \, M = M_{ii}$): $p_j = \text{Tr} \big(|j\rangle\langle j|\rho \big)$. The proof is as follows:

$$\begin{aligned}
\text{Tr} \big(|j\rangle\langle j|\rho \big) &= \sum_n \langle n|j \rangle \, \langle j|\rho \, |n\rangle \\
&= \sum_{n,i} p_i \langle n|j \rangle \, \langle j|i \rangle \, \langle i|n \rangle \\
&= \sum_{n,i} p_i \langle i|n \rangle \, \langle n|j \rangle \, \langle j|i \rangle \\
&= \sum_i p_i \langle i|j \rangle \, \langle j|i \rangle = p_j,
\end{aligned} \tag{8.19}$$

where some reordering of terms and completeness on n has been used.

Similarly, the density operator gives a general means of obtaining expectation values of any operator, A:

$$\langle A \rangle = \text{Tr} \, (A\rho) = \text{Tr} \, (\rho A). \tag{8.20}$$

First consider what the expectation of an operator looks like for a single state $|\psi\rangle$: $\langle A \rangle \equiv \langle \psi|A|\psi \rangle$. If the state is expanded in some basis $|\psi\rangle = \sum_i a_i |i\rangle$, then

$$\langle A \rangle = \sum_{i,j} a_i a_j^* \, \langle j|A|i \rangle = \sum_{i,j} \rho_{ij} A_{ji} = \text{Tr} \, [\rho A]. \tag{8.21}$$

The name 'density matrix' is clearly applicable here, since $\rho = |\psi\rangle\langle\psi|$, and $\rho_{ij} = \langle i|\rho|j\rangle = a_i a_j^*$. In the general case,

$$\begin{aligned}
\mathrm{Tr}\,(A\rho) &= \sum_n \langle n|A\rho|n\rangle = \sum_{n,i} p_i\langle n|A|i\rangle\,\langle i|n\rangle \\
&= \sum_{n,i} p_i\langle i|n\rangle\,\langle n|A|i\rangle = \sum_i p_i\langle i|A|i\rangle,
\end{aligned} \tag{8.22}$$

where the last step uses completeness, $\sum_n |n\rangle\langle n| = 1$; thus $\langle A\rangle$ is just the weighted sum of $\langle A\rangle$ over each of the possible pure states. Note in the above that the order of A and ρ is irrelevant. The trace of a matrix M is M_{ii}, so the trace of a product is independent of order: $A_{ij}B_{ji} = B_{ji}A_{ij}$.

Thus a density matrix is needed in general in order to understand how the relation $\langle A\rangle = \sum_i p_i\langle i|A|i\rangle$ can arise. We could try constructing a wave function

$$|\psi\rangle = \sum_i p_i^{1/2}|i\rangle, \tag{8.23}$$

but then the expectation value would be

$$\langle A\rangle = \langle\psi|A|\psi\rangle = \sum_{ij} p_i^{1/2} p_j^{1/2} A_{ij}. \tag{8.24}$$

This contains extra 'off diagonal terms' (involving A_{ij} $i \neq j$) that are absent from the desired expression $\sum_i p_i\langle i|A|i\rangle$. The classical probabilities appearing in this latter expression thus cannot be accommodated within a single wave function.

Density matrix, measurement, and decoherence   Note that this formalism does also include the simpler case of a **pure state**, where the system is described by a single wave function. Consider for example the case where this is a superposition of two other states: $|\psi\rangle = a|1\rangle + b|2\rangle$, where $\langle 1|2\rangle = 0$. The density matrix is

$$\rho = |\psi\rangle\langle\psi| = |a|^2|1\rangle\langle 1| + |b|^2|2\rangle\langle 2| + ab^*|1\rangle\langle 2| + a^*b|2\rangle\langle 1|. \tag{8.25}$$

The indication of the possibility of quantum interference here is the cross terms $|1\rangle\langle 2|$ and $|2\rangle\langle 1|$. For a 'classical' system, which would definitely be either in state $|1\rangle$ or $|2\rangle$, these off-diagonal terms would be zero, so that

$$\rho = |a|^2|1\rangle\langle 1| + |b|^2|2\rangle\langle 2|. \tag{8.26}$$

Now, the key point is that in both classical and non-classical cases the probability of a measurement finding the system in state $|1\rangle$ is $|a|^2$, independently of the off-diagonal terms: $\mathrm{Tr}\,(|1\rangle\langle 1|\rho) = |a|^2$. This gives a possible solution to our intuitive worries about macroscopic objects possibly being in a quantum superposition. Suppose we set up Schrödinger's cat in a pure state that is a mixture of the $|\mathrm{alive}\rangle$ and $|\mathrm{dead}\rangle$ states. If the cat is an isolated system, that mixture will of course be preserved, and interference will be possible at some later time. In practice, however, the cat is a complex system of many sub-parts all of which interact with its environment – and these interactions are the very agency that will cause the balance of probabilities to tilt from life to death. In these circumstances, the rapid phase changes of the microscopic components of the cat will mean that the wave function at later times will not stay coherent with its initial value, and the opportunity for macroscopic interference will be lost. The idea of **decoherence** is thus that these microscopic interactions very rapidly cause the off-diagonal components of the density matrix to decay to zero. When we open the box, we find a classical system in which the cat is definitely dead or alive, not a superposition of these states.

Canonical density matrix   The canonical ensemble is a classic example of a **mixed state** where a density matrix is required. For a thermal ensemble, the occupation probabilities are Boltzmann factors, so

$$\rho = \frac{1}{Z}\sum_i \exp(-E_i/kT)\,|i\rangle\langle i|. \tag{8.27}$$

Now, $\exp(-E_i/kT)\,|i\rangle = \exp(-H/kT)\,|i\rangle$ (expand the exponential term by term) and completeness says $\sum_i |i\rangle\langle i| = 1$, so for a system in thermal equilibrium

$$\rho = \frac{1}{Z}\,\exp(-H/kT). \tag{8.28}$$

Finally, the Canonical partition function can be written as

$$Z = \mathrm{Tr}\left[e^{-\beta H}\right] \tag{8.29}$$

because

$$\mathrm{Tr}\left[e^{-\beta H}\right] = \sum_{i=0}^{\infty}\langle i|e^{-\beta H}|i\rangle = \sum_{i=0}^{\infty} e^{-\beta E_i}\langle i|i\rangle = \sum_{i=0}^{\infty} e^{-\beta E_i}. \tag{8.30}$$

Von Neumann entropy Finally we state the form of the entropy using the density matrix, which is known as the von Neumann entropy:

$$S = -k\,\mathrm{Tr}\left[\rho\ln\rho\right]. \tag{8.31}$$

This is extends the Shannon-Gibbs entropy from statistical mechanics and classical information theory to the field of quantum mechanics and quantum information theory. In a tutorial, we will show that this general formula indeed corresponds to the Gibbs entropy as we have met it earlier.

9 The master equation and detailed balance

9.1 The quantum entropy paradox

We now look again at the issue of reconciling macroscopic irreversibility with microscopic reversibility, but this time from a quantum perspective. Since any increase of the Gibbs entropy requires us to move from the certainty of occupying a single microstate to having a range of possible states in play, we might initially have some vague notion that 'quantum uncertainty' could accomplish the necessary blurring. But any such hope is quickly seen to be incorrect. Suppose we choose as the microstates the eigenstates of the system's quantum Hamiltonian H. But each $|i\rangle$ undergoes independent **unitary** evolution under H, so there is no mechanism for an isolated system to jump from one eigenstate to another (this is the fundamental problem of measurement in quantum mechanics). If we write the state of the system as

$$|\Psi\rangle = \sum_i c_i|i\rangle, \tag{9.1}$$

then the evolution becomes

$$i\hbar\frac{\partial|\Psi\rangle}{\partial t} = H|\Psi\rangle = \sum_i c_i E_i|i\rangle, \tag{9.2}$$

with solution (assuming time-independent H)

$$|\Psi(t)\rangle = \sum_i c_i e^{-iE_i t/\hbar}|i\rangle, \tag{9.3}$$

which simply applies a phase factor to the eigenstates. That is, Schrödinger's equation applied to energy eigenstates as microstates *cannot change* their classical weights $p_i = |c_i|^2$. Accordingly, the entropy $S = -k\sum |c_i|^2\ln|c_i|^2$ is trivially constant: $dS/dt = 0$, just as in classical mechanics.

We can generalise this argument to the case where the microstates $|i\rangle$ are not eigenstates of H, which requires the **density matrix**:

$$\rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i| \,, \tag{9.4}$$

where the $|\Psi_i\rangle$ are some state vectors (not necessarily energy eigenstates) and p_i are classical probabilities. In this language, we saw that the entropy is

$$S = -k \operatorname{Tr} [\rho \ln \rho] \,. \tag{9.5}$$

Now use the Schrödinger equation and its adjoint (and the fact that H is Hermitian):

$$i\hbar \frac{\partial |\Psi\rangle}{\partial t} = H|\Psi\rangle \qquad -i\hbar \frac{\partial \langle\Psi|}{\partial t} = \langle\Psi|H \,.$$

This allows us to differentiate the product in ρ, to obtain its time derivative:

$$\frac{\partial \rho}{\partial t} = -\frac{i}{\hbar} [H, \rho] \,, \tag{9.6}$$

which is the **von Neumann equation**. The **unitary** time evolution of $|\Psi_i\rangle$ and ρ becomes

$$|\Psi_i(t)\rangle = e^{-iHt/\hbar}|\Psi_i(0)\rangle \quad ; \quad \rho(t) = e^{-iHt/\hbar}\rho(0)e^{iHt/\hbar} \,. \tag{9.7}$$

But this unitary transformation is a time-dependent change of basis, which is not something that can affect the trace of a matrix, so that again we have

$$\frac{\mathrm{d}S}{\mathrm{d}t} = 0 \,. \tag{9.8}$$

9.2   Fermi's master equation

The conflict between this constancy of quantum entropy under unitary evolution and the need for entropy to increase in practice can be resolved by appealing to the concept of **decoherence** introduced above. Consider again our earlier example of the density matrix for a pure state, which we have seen degenerates through decoherence to $\rho = |a|^2|1\rangle\langle1| + |b|^2|2\rangle\langle2|$, so that there are classical probabilities $|a|^2$ and $|b|^2$ of being in states 1 or 2. The entropy for this situation is $S/k = -|a|^2 \ln |a|^2 - |b|^2 \ln |b|^2$, which is positive. But if the initial pure state undergoes unitary evolution then there can be no possibility for its entropy to change from zero. The loss of the phase information through decoherence is therefore the cause of irreversibility and the arrow of time. Decoherence reduces our knowledge about the system, so there is hidden information – which as we have seen is required in order for entropy to increase.

To get anything resembling the second law, we must therefore break unitary evolution and adopt a **coarse graining** approach to quantum mechanics, in which information is discarded. One way of doing this is to acknowledge that we cannot write down the true Hamiltonian, containing all the microscopic detail of a real many particle system; instead, we generally adopt an approximate Hamiltonian H_0 and describe the system in terms of a set of eigenstates of H_0, $|i\rangle$. The true Hamiltonian is given by H_0 plus a small perturbation:

$$H = H_0 + h \,. \tag{9.9}$$

In terms of this true Hamiltonian, the states $|i\rangle$ are *approximate* energy eigenstates. The matrix elements of h in this basis are

$$h_{ij} = \langle i|h|j\rangle = h_{ji}^* \,, \tag{9.10}$$

where * denotes complex conjugate. This perturbation induces **quantum jumps** between the approximate states $|i\rangle$. These states are treated in an explicitly nonunitary way: we assign to them **evolving classical probabilities** p_i, instead of keeping track of their quantum amplitudes. This can be viewed as a form of coarse graining in which the quantum coherence between these different states is wiped out.

The transition rates induced by a perturbation that oscillates at frequency ω can be calculated using time-dependent perturbation theory, and the standard result is **Fermi's Golden Rule**:

$$\nu_{ij} = \frac{2\pi}{\hbar} |h_{ij}|^2 \, \delta(E_i - E_j \pm \hbar\omega) \,, \tag{9.11}$$

where the difference signs in $\pm\hbar\omega$ correspond to emission or absorption of energy from the oscillating perturber. You may be more familiar with this in the form where there is a continuum of final states, so that the delta-function in energy is replaced by a density of final states, $\rho(E)$, but the above form is the more fundamental result, which applies for an explicit pair of initial and final states. Note that because $h_{ij} = h_{ji}^*$, the jump rates are **symmetric** in either direction

$$\nu_{ij} = \nu_{ji} \,. \tag{9.12}$$

In a small enough time interval dt, the probability of making a transition is $p = \nu \, dt$, and the probabilities p_i change according to

$$dp_i = \left(-p_i \sum_j \nu_{ij} + \sum_j p_j \, \nu_{ji} \right) dt \,, \tag{9.13}$$

where the first term is the probability that the system is in state $|i\rangle$ initially but jumps out of it into some other state, and the second is the probability that the system is initially elsewhere but jumps into state $|i\rangle$. Thus

$$\frac{dp_i}{dt} = \sum_j (p_j \nu_{ji} - p_i \nu_{ij}) = \sum_j \nu_{ij} (p_j - p_i) \,, \tag{9.14}$$

where the last step uses equality of the two rates. This is called the **Master equation**: it is *first order in time* and clearly does not have time reversal symmetry.

9.3 Quantum H theorem

Consider transitions just between a particular pair of states, $|1\rangle$ and $|2\rangle$, so that $p_1 + p_2 = \text{const.}$ and in a small interval dt, $dp_1 = -dp_2$. The master equation says $dp_1 = -\nu_{12}(p_1 - p_2)dt$, and we can add this to itself to get

$$d(p_1 - p_2) = -2\nu_{12}(p_1 - p_2)dt \,. \tag{9.15}$$

Since ν_{ij} is always positive, we see that $|p_1 - p_2|$ always decreases: transitions will always tend to equalise occupation probabilities. The entropy change which results from this is always positive. The proof is again based on the **concavity** of the function $s(p) = -kp \ln p$, which we considered previously in the classical case. If p_1 and p_2 evolve to p_1' and p_2', we can see from Fig. 9 that the entropy must increase. Since the master equation describes the overall effect of many pairwise contributions of this kind, each of which can only increase the entropy, we have proved a quantum version of the H theorem:

$$\frac{dS}{dt} \geq 0 \,, \tag{9.16}$$

which is the second law.

The coarse graining implicit in the master equation succeeds in reconciling microscopic quantum dynamics with the observed thermodynamic properties of large systems. But the actual nature of the coarse graining is less clear than in the classical case; it involves discarding quantum information (phase). Also, this raises the question of whether the phenomenon of **wavefunction collapse** (implicit in the concept of a quantum jump) is somehow required to fix the arrow of time. Yet again, we see that the concept of the observer has entered the discussion. As we saw with the discussion in terms of the density matrix, an isolated system will undergo unitary evolution: there will be no quantum jumps and no change in entropy. The contrary conclusion from the master equation in then a statement about what changes as a result of us performing measurements on the system and acquiring (or not) information. This is perhaps the ultimate illustration of the fact that the Gibbs formulation of entropy is a subjective quantity.

9.4 Summing up the second law

This concludes our survey of the attempts to understand thermodynamics (and entropy in particular) via a fundamental microscopic approach. Along the way, we have developed a number of interesting tools and concepts, and for the remainder of the course it is time to 'shut up and calculate' as we look at the plentiful applications of these ideas. But the journey has been long and complicated, so it is worth a final attempt to take stock: did we succeed in understanding the second law and the arrow of time? The most important elements of any such understanding are the following:

(1) A general definition of entropy is the Gibbs expression, $S_G = -k \sum_i p_i \ln p_i$, involving the probability of finding a system in the ith microstate. In thermal equilibrium, the p_i are given by Boltzmann or Gibbs factors and S_G is then maximised and is equal to the classical experimental entropy – a quantity that is meaningful only for large systems in equilibrium.

(2) The central difficulty in moving beyond equilibrium thermodynamics and proving $\Delta S > 0$ is to understand how an isolated system can attain thermal equilibrium, when there is no external random influence. This is apparently impossible, because S_G cannot change in an isolated system.

(3) Nevertheless, a small range of initial microstates in phase space will spread to cover a large range of states at late times, in a way that cannot be distinguished from true thermal equilibrium. S_G will then increase, as a result of 'coarse graining': our lack of precise knowledge of the exact microstate occupation probabilities.

(4) Even in thermal equilibrium, however, rare microstates will sometimes be populated – and some of these can lead to evolution towards a temporary state of lower classical entropy. But these transient states can only be used to extract work from heat (and thus violate the second law) if the system is monitored continuously. This information processing by the observer generates additional entropy, so that the total entropy of the universe still always increases.

(5) In the end, the arrow of time derives from causality: any non-equilibrium initial conditions that we impose affect the universe – and our knowledge about it – in the future, but not in the past.

10 Applications of the master equation: random walk and diffusion

The basic structure of the master equation underlies most irreversible processes in physics, whether quantum or classical. As a ubiquitous example, we consider the case of diffusion on a lattice in one dimension, i.e. a continuous time random walk. This could be the motion of a vacancy (or of an impurity atom) in a crystal that moves by swapping places with (other) atoms. We label the lattice sites of the system by positive and negative integers $i = -L/a \ldots + L/a$ where a is the lattice spacing and L is very large (effectively infinite so that we avoid discussion of boundary conditions). The state of the system is then given by i, the spatial position of the particle. In small time interval dt there is probability $\nu \, dt$ for our particle to move to the right in time interval dt and the same to the left. Then

$$\nu_{ij} = \begin{cases} \nu & j = i - 1 \\ \nu & j = i + 1 \\ 0 & \text{otherwise} \end{cases} \tag{10.1}$$

so that

$$\frac{dp_i}{dt} = \sum_j \nu_{ij}(p_j - p_i) \tag{10.2}$$

$$= \nu \, (p_{i-1} - p_i) + \nu \, (p_{i+1} - p_i) \, . \tag{10.3}$$

This is the master equation for a random walk – specifically, for a *continuous time* random walk on a **discrete lattice**. When the time variable is also discrete, we end up with the binomial distribution for p_i if we start at the origin (say). The continuous time case is a bit messier (the solution for $p_i(t)$ involves modified Bessel functions rather than Binomial coefficients).

But things simplify if we assume that p_i varies smoothly with i (in fact this will always be true eventually even if not to begin with). In that case we can rewrite the master equation as continuous also in space by introducing the probability density $p(x,t)dx$ of finding our object in a small region centred on $x = ia$. Then the idea is to replace $p_i(t) \to p(x,t)$ and expand $p_{i\pm1} = p(x \pm a)$ for a small. i.e.

$$p_{i\pm1} = p(x,t) \pm a \frac{\partial p(x,t)}{\partial x} + \frac{a^2}{2} \frac{\partial^2 p(x,t)}{\partial x^2} + \ldots \tag{10.4}$$

Then keeping terms up to $O(a^2)$ we obtain

$$\boxed{\frac{\partial p}{\partial t} = a^2 \nu \frac{\partial^2 p}{\partial x^2} \, ,} \tag{10.5}$$

which is the **diffusion equation** for the probability density of a particle with **diffusion coefficient** $D = a^2\nu$. The dimensions of D are $[L^2/T]$. The diffusion equation can also describe many non-interacting particles diffusing: we then replace p by ρ, the density or concentration of particles. The normalisation is $\int \rho \, dx = N$ where N is the number of particles. The diffusion equation, like the master equation from which it was derived, explicitly violates time reversal symmetry and allows entropy to increase.

The solution of the diffusion equation for an initial condition in which the particle is localised at the origin (formally, $p(x,0) = \delta(x)$) is the Gaussian

$$\boxed{p(x,t) = (4\pi Dt)^{-1/2} \exp[-x^2/4Dt] \, .} \tag{10.6}$$

This formula may be checked directly and can also be derived by taking the Fourier transform of the diffusion equation. As t increases the Gaussian 'bell-shaped' curve spreads out, becoming wider and lower at the peak (Fig. (10)). The width grows as $\langle x^2 \rangle^{1/2} = (2Dt)^{1/2}$, so that a particle will take time $\propto L^2$ to explore a region of size L.

10.1   Detailed balance

As we have seen, by introducing a type of coarse-graining, the master equation violates time reversal symmetry and leads to the second law. Remarkably, however, the fact that the underlying microphysics is actually time reversal-symmetric has deep consequences that survive the coarse-graining procedure and provide the cornerstones of nonequilibrium thermodynamics, as follows.

Isolated system   For an isolated system in equilibrium, the Principle of equal a priori probability (PEAPP) holds ($p_i^{\mathrm{eq}} = \mathrm{const.} = p^{\mathrm{eq}}$) and we have transition rate symmetry, so that

$$\boxed{p_i^{\mathrm{eq}} \nu_{ij} = p_j^{\mathrm{eq}} \nu_{ji} \,.} \tag{10.7}$$

This is the condition of **detailed balance**: the rate of quantum jumps from i to j (which is the left hand side) is the same as from j to i. In the other words the probability flux from i to j is exactly balanced by the probability flux from j to i. This is a stronger statement than the master equation, which only states that to have a stationary distribution ($\dot{p}_i = 0$) there should be *overall* balance between rate of jumping into and out of state i. But here we see that equilibrium requires balance between *any* two pairs of states.

The result is very powerful, because it applies not only to individual states but any grouping of them: consider two groups of states A and B. Summing the detailed balance condition over i a member of group A and j a member of group B yields

$$\sum_{i \in A} \sum_{j \in B} p_i^{\mathrm{eq}} \nu_{ij} = \sum_{i \in A} \sum_{j \in B} p_j^{\mathrm{eq}} \nu_{ji} \,. \tag{10.8}$$

Now consider the mean rate of transition from group A to group B

$$p_A^{\mathrm{eq}} \nu_{AB} = \sum_{i \in A} p_i^{\mathrm{eq}} \sum_{j \in B} \nu_{ij} \,. \tag{10.9}$$

The left hand side involves the total probability of being in group A:

$$p_A^{\mathrm{eq}} = \sum_{i \in A} p_i^{\mathrm{eq}} = p^{\mathrm{eq}} \, \Omega_A \,, \tag{10.10}$$

where Ω_A is the number of microstates in A. This probability is multiplied by the mean rate, ν_{AB}, of transitions from group A to group B *given* that the system is in group A:

$$\nu_{AB} = \frac{\sum_{i \in A} p_i^{\mathrm{eq}} \sum_{j \in B} \nu_{ij}}{\sum_{i \in A} p_i^{\mathrm{eq}}} = \frac{1}{\Omega_A} \sum_{i \in A} \sum_{j \in B} \nu_{ij} \,, \tag{10.11}$$

and similarly for the right hand side.

To understand the right hand side note that we must start in some state i in group A. Thus we consider the total rate from state i into group B given by $\sum_{j \in B} \nu_{ij}$ then average over the states i in group A each with their equilibrium probability p_i^{eq}. Then the equation becomes

$$p_A^{\mathrm{eq}} \nu_{AB} = p_B^{\mathrm{eq}} \nu_{BA} \,, \tag{10.12}$$

i.e. the transitions from A to B are exactly balanced by the transitions from B to A and we have detailed balance between the two groups. It is important here to note since $p_A \neq p_B$ for different size groups the transitions rates between the two groups are generally not symmetric:

$$\nu_{AB} \neq \nu_{BA} \,. \tag{10.13}$$

Canonical ensemble Now we can use this result to describe dynamics in the canonical ensemble. Recall the setup that can be used to derive the canonical ensemble: a microstate of the 'composite' of system + bath specifies the state of both. But if we just specify the microstate of the system, then this corresponds to many possible states (with the allowed energy) of the bath and thus to a group of states of the composite. Therefore transitions between states (say α and β) of the system in the canonical ensemble are between groups of states in the composite and obey detailed balance

$$p_\alpha^{\text{eq}} \nu_{\alpha\beta} = p_\beta^{\text{eq}} \nu_{\beta\alpha} \, . \tag{10.14}$$

Further since we know in the canonical ensemble that $p_\alpha^{\text{eq}} \propto e^{-\beta E_\alpha}$ we find

$$\frac{\nu_{\alpha\beta}}{\nu_{\beta\alpha}} = e^{\beta(E_\alpha - E_\beta)} \, . \tag{10.15}$$

Thus the transition rates are asymmetric and the ratio is a function of the energy difference between the states.

Spontaneous emission Detailed balance was used to great effect by Einstein. Consider for simplicity two states of different energies: transitions between them must conserve energy and usually this will involve the emission or absorption of a photon, where $h\nu$ is equal to the energy difference. The problem is that the golden rule, applied to some perturbing radiation field, predicts that the transition rates are symmetric: $\nu_\uparrow = \nu_\downarrow = BU$, where U is the electromagnetic energy density and B is some calculable coefficient. But detailed balance says that the rates must be unequal. Einstein guessed that the correct equation involved **spontaneous emission**: $\nu_\downarrow = A + BU$. Since we have

$$\frac{\nu_\downarrow}{\nu_\uparrow} = \frac{A + BU}{BU} = \frac{P_{\text{lower}}}{P_{\text{upper}}} = \exp[\beta h\nu], \tag{10.16}$$

we learn that

$$A = BU(\exp[-\beta h\nu] - 1). \tag{10.17}$$

But we saw that $U \propto \nu^3(\exp[-\beta h\nu] - 1)^{-1}$, so we get an expression for A, independent of the temperature used in the argument. A more transparent way to express the answer, though, is to recognise that $(\exp[-\beta h\nu] - 1)^{-1} = \bar{n}$, the photon occupation number. Hence

$$\boxed{\frac{\nu_\downarrow}{\nu_\uparrow} = \frac{\bar{n} + 1}{\bar{n}} .} \tag{10.18}$$

It is as if downwards transitions are stimulated by one more photon than is actually present. This can be understood when the electromagnetic field is treated quantum mechanically.

Currents One general consequence of detailed balance is that, in equilibrium, there are no microscopic probability currents or, more generally, macroscopic currents of physical quantities. As an example consider the one-dimensional diffusion equation in the continuity equation form

$$\frac{\partial \rho}{\partial t} + \frac{\partial J}{\partial x} = 0 \, , \tag{10.19}$$

where

$$J = -D\frac{\partial \rho}{\partial x} \, , \tag{10.20}$$

is the diffusive particle current (diffusion causes a net flow down a concentration gradient). The current is of the globally conserved particle concentration ρ. A stationary solution $\partial\rho/\partial t = 0$ requires $J = \text{const}$. An equilibrium solution goes further and fixes

$$J^{\text{eq}} = 0 \, , \tag{10.21}$$

implying that ρ^{eq} is spatially constant. A non-equilibrium stationary solution would have $J \neq 0$ thus implying

$$\rho(x) = \text{const.} - \frac{J}{D}x . \tag{10.22}$$

Such a linear density profile would have to be maintained in a stationary state by forcing particles through the system from an external reservoir for example. The system is then held out of equilibrium in a non-equilibrium steady state.

Monte Carlo simulation Detailed balance can be used as a practical tool in computing. Suppose we want to simulate the interactions of many particles in order to follow them into their equilibrium state. Solving exact equations of motion in the **molecular dynamics** method can be highly time consuming. But to obtain the equilibrium properties of the system, it may be much faster to use a dynamics which does not look anything like the actual equations of motion. If in equilibrium the artificial dynamics obey the principle of detailed balance, then it is (almost) guaranteed that the steady state found by simulation is the true equilibrium state, namely

$$p_\alpha = \frac{1}{Z} e^{-\beta E_\alpha} . \tag{10.23}$$

The best known example is the **Monte Carlo method**, in which the dynamical algorithm consists of random jumps, according to the **Metropolis algorithm**. Here, the jump rates $\nu_{\alpha\beta}$ for all pairs of states (α, β) take the form

$$\nu_{\alpha\beta} = \nu_0 \begin{cases} 1 & E_\beta < E_\alpha \\ \exp(-\beta(E_\beta - E_\alpha)) & E_\beta \geq E_\alpha \end{cases} \tag{10.24}$$

$$\tag{10.25}$$

$$\equiv \nu_0 \min\left(1, \exp(-\beta(E_\beta - E_\alpha))\right) , \tag{10.26}$$

where ν_0 is an irrelevant constant. This satisfies detailed balance as we have

$$p_\alpha^{\text{eq}} \nu_{\alpha\beta} = \frac{1}{Z} e^{-\beta E_\alpha} \times \nu_0 \min\left(1, e^{-\beta(E_\beta - E_\alpha)}\right) \tag{10.27}$$

$$= \frac{1}{Z} \nu_0 \min\left(e^{-\beta E_\alpha}, e^{-\beta E_\beta}\right) \tag{10.28}$$

$$= \frac{1}{Z} e^{-\beta E_\beta} \times \nu_0 \min\left(1, e^{-\beta(E_\alpha - E_\beta)}\right) \tag{10.29}$$

$$= p_\beta^{\text{eq}} \nu_{\beta\alpha} . \tag{10.30}$$

This is useful in simulations as it dynamically generates a realization of the canonical distribution p_α^{eq}, from which we can directly find various averages $\langle O \rangle = \sum_\alpha O_\alpha p_\alpha^{\text{eq}}$.

Note that the Monte Carlo method is a stochastic algorithm, i.e. is non-deterministic. It thus represents a form of coarse-graining over the deterministic microscopic dynamics. (The epithet Monte Carlo was inspired by the spinning of wheels in a Casino.)

11   The Langevin approach and the dynamics of fluctuations

11.1   The random walk

We have seen that the approach to thermal equilibrium ends up involving random influences – either exactly so if there is external thermal contact, or effectively so in isolated systems despite

the operation of causal dynamics. It is therefore interesting to study particle dynamics by means of **stochastic differential equations**, which include explicitly random terms. To illustrate what is involved, let us revisit the random walk and the diffusion limit of section 10:

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2}. \tag{11.1}$$

This is the equation for the time evolution of the **probability distribution** of the position of a particle, x.

The **Langevin equation**, on the other hand is a stochastic equation that tries to describe the time dependence of x directly. Consider a small time interval Δt with

$$x(t + \Delta t) = x(t) + \Delta x(t), \tag{11.2}$$

where the step $\Delta x(t)$ is a random variable. When the lattice spacing is $|\Delta x| = a$,

$$\Delta x(t) = \begin{cases} +a & \text{probability} \quad \nu \Delta t \\ -a & \text{probability} \quad \nu \Delta t \\ 0 & \text{probability} \quad (1 - 2\nu \Delta t) \,. \end{cases} \tag{11.3}$$

Note that $\langle \Delta x \rangle = 0$ and $\langle (\Delta x)^2 \rangle = 2a^2 \nu \Delta t = 2D\Delta t$. Also note that $\Delta x(t)$ are uncorrelated for different t. What we want to do now is take the 'continuum limit' i.e. let Δx and Δt go to zero in some prescribed way. There are some mathematical subtleties involved, and to make things work the lattice spacing a should in fact scale as

$$a \propto (\Delta t)^{1/2} \tag{11.4}$$

when we take the limit $\Delta t \to 0$. But the continuum form of the equation is intuitively reasonable:

$$\boxed{\frac{dx}{dt} = \eta(t)}, \tag{11.5}$$

where

$$\langle \eta(t) \rangle = 0 \tag{11.6}$$
$$\text{and} \quad \langle \eta(t)\eta(t') \rangle = \Lambda \, \delta(t - t') \,. \tag{11.7}$$

A random variable $\eta(t)$ with these statistics is known as **white noise**.

We thus have **stochastic processes** $x(t)$ and η, related through a stochastic differential equation, commonly referred to as a **Langevin equation**. We can formally integrate (11.5) to obtain

$$x(t) - x_0 = \int_0^t dt' \eta(t') \tag{11.8}$$

To extract useful information from this solution we should take averages. Using the properties (11.6,11.7)

$$\langle x(t) - x_0 \rangle = 0 \tag{11.9}$$
$$\langle (x(t) - x_0)^2 \rangle = \int_0^t dt' \int_0^t dt'' \langle \eta(t')\eta(t'') \rangle = \Lambda \int_0^t dt' \int_0^t dt'' \delta(t' - t'') = \Lambda \, t. \tag{11.10}$$

But the solution of the diffusion equation gives $2Dt$ as the variance in x, so we identify the diffusion coefficient D as

$$\boxed{D = \Lambda/2.} \tag{11.11}$$

These two descriptions, diffusion equation and Langevin equation, are in fact equivalent and generate the same Gaussian probability distribution $p(x)$. A detailed proof is given in the *Advanced*

60

Statistical Physics course, but informally we can appeal to the **central limit theorem**: (11.8) is a sum of uncorrelated random variables, which under general conditions has a Gaussian probability distribution. In summary, the Langevin approach then describes a **Gaussian Markov stochastic process**: one whose state at some future time t has a Gaussian probability distribution that is purely conditional on its state at time t_0, not on the history by which the system reached that state (independence of history is what we mean by a Markov process).

11.2 Brownian motion

We now apply the Langevin approach to the motion of a single colloidal particle in a fluid. This particle is constantly being bombarded by collisions with the (smaller) fluid molecules; this **Brownian motion** was first reported by Reverend Robert Brown in 1827. It was analysed in 1905 by Einstein (in his most cited research paper) who realised that, among other things, it proved that the suspending fluid was indeed made of discrete molecules, an idea which up to then had remained controversial. Today, when laboratory experiments on individual atoms are commonplace, it seems shocking that the reality of atoms could have been doubted as recently as the early 20th Century. In this uncertain context, we can only admire Boltzmann's courage in seeking microscopic explanations of thermodynamics, and we can understand more readily why his efforts met with such resistance.

Approaching the phenomenon of Brownian motion via a Langevin equation is a form of coarse-graining, which once again introduces irreversibility into macroscopic physics. That is, in principle we could keep track of all the (reversible) molecular dynamics which gives rise to the collision but we cannot practically do this. Instead we postulate the existence of **random forces** that are not derived from any Hamiltonian, as first done by Paul Langevin in 1908.

Consider a particle of mass m immersed in a fluid. According to Langevin, its equation of motion is (in one dimension)

$$m\frac{d^2x}{dt^2} = -\gamma\frac{dx}{dt} + f(t)\,. \tag{11.12}$$

The first term on the right is a damping force caused by friction between the particle and the fluid. Note that often $\mu \equiv 1/\gamma$ is called the **mobility** of such a sphere (not to be confused with chemical potential). The final term $f(t)$ is a random force. This stochastic process should be correlated in time over some period set by the duration of molecular conditions, t_c, but we assume that this is short enough that in practice we can model the correlations with a delta-function so that, as in the random walk,

$$\langle f(t)\rangle = 0 \tag{11.13}$$
$$\langle f(t_1)f(t_2)\rangle = \Gamma\delta(t_1 - t_2)\,. \tag{11.14}$$

Now let us integrate the equation. For simplicity take $m = 1$. Introducing an integrating factor gives

$$\frac{d}{dt}\left[v(t)\mathrm{e}^{\gamma t}\right] = \mathrm{e}^{\gamma t}f(t)\,, \tag{11.15}$$

which we can integrate up to obtain

$$v(t) - v_0\mathrm{e}^{-\gamma t} = \int_0^t dt'\mathrm{e}^{-\gamma(t-t')}f(t')\,. \tag{11.16}$$

Thus the expectation of the velocity damps away through friction:

$$\langle v(t)\rangle = v_0\mathrm{e}^{-\gamma t}\,. \tag{11.17}$$

Now consider the mean squared velocity

$$\langle v(t)^2 \rangle = v_0^2 e^{-2\gamma t} + \int_0^t dt' \int_0^t dt'' e^{-\gamma(t-t')-\gamma(t-t'')} \langle f(t')f(t'') \rangle \tag{11.18}$$

$$= v_0^2 e^{-2\gamma t} + \Gamma \int_0^t dt' e^{-2\gamma(t-t')} \tag{11.19}$$

$$= v_0^2 e^{-2\gamma t} + \frac{\Gamma}{2\gamma} \left[1 - e^{-2\gamma t} \right] . \tag{11.20}$$

So initially $\langle v^2 \rangle \simeq v_0^2$, but eventually the particle forgets its initial velocity and comes into thermal equilibrium with the fluid:

$$\langle v^2 \rangle \to \frac{\Gamma}{2\gamma} \quad \text{for} \quad \gamma t \gg 1 . \tag{11.21}$$

By equipartition, a particle in equilibrium should have kinetic energy ($m = 1$ here)

$$\frac{1}{2}\langle v^2 \rangle = \frac{1}{2} kT . \tag{11.22}$$

Therefore we deduce

$$\boxed{\Gamma = 2\gamma kT} . \tag{11.23}$$

The magnitude of the noise fluctuations Γ are thus related to the viscous or dissipative term controlled by γ – reasonably enough, as both effects result from the same microscopic dynamics involving molecular collisions.

Einstein relation   Integrating again gives (see tutorial)

$$x(t) - x_0 = \frac{v_0}{\gamma}(1 - e^{-\gamma t}) + \int_0^t dt' \int_0^{t'} dt'' e^{-\gamma(t'-t'')} f(t''), \tag{11.24}$$

leading to

$$\langle (x(t) - x_0)^2 \rangle = \frac{\Gamma}{\gamma^2}t - \frac{\Gamma}{\gamma^3}\left[1 - e^{-\gamma t} \right] + \frac{1}{\gamma^2}\left(v_0^2 - \frac{\Gamma}{2\gamma} \right)\left[1 - e^{-\gamma t} \right]^2 , \tag{11.25}$$

and giving the short and long time limits

$$\langle (x(t) - x_0)^2 \rangle \simeq \begin{cases} v_0^2 t^2 & \gamma t \ll 1 \\ \frac{\Gamma}{\gamma^2}t = \frac{2kT}{\gamma}t & \gamma t \gg 1 . \end{cases} \tag{11.26}$$

Thus initially the Brownian particle moves ballistically with the initial velocity v_0; but after a sufficiently long time, it performs diffusive motion (where the acceleration term in equation [11.12] can be ignored). Comparing with equation (10.6) with diffusion coefficient D, i.e. $\langle (x(t)-x_0)^2 \rangle = 2Dt$, we obtain

$$\boxed{D = \frac{\Gamma}{2\gamma^2} = \frac{kT}{\gamma}} \tag{11.27}$$

This equation is known as the **Einstein relation** and relates the diffusion coefficient (the **fluctuation**) of the particle to its **dissipation**.

Nernst–Einstein relation   Finally we can consider a particle with charge q and allow for the effect of an external field E. This field will produce a systematic force qE to add into (11.12) and the equation for the mean velocity becomes

$$m\left\langle \frac{dv}{dt} \right\rangle = -\gamma\langle v \rangle + qE . \tag{11.28}$$

62

The terminal **drift velocity** (when $\langle dv/dt \rangle = 0$) is then

$$\langle v \rangle = \frac{qE}{\gamma}. \tag{11.29}$$

This idea of a drift velocity provides a direct shortcut to the Einstein relation. If we consider equilibrium under the external field, then we would expect the density of particles to be governed by a Boltzmann factor: $\rho(x) \propto \exp[qEx/kT]$ (because $-qEx$ is a potential energy if the force points in the $+x$ direction). The drift velocity produces a flux density of particles in the $+x$ direction of $j = \rho\langle v \rangle$, but in equilibrium this will be cancelled by the flux density of diffusion in the direction of lower density:

$$\rho\langle v \rangle = D d\rho/dx = D(qE/kT)\rho \quad \Rightarrow 1/\gamma = D/kT. \tag{11.30}$$

If we define the **mobility** μ as

$$\mu = \frac{\text{terminal drift velocity}}{\text{applied force}}, \tag{11.31}$$

then as here $\mu = \langle v \rangle/(qE) = 1/\gamma$, we have

$$\boxed{\mu = \frac{D}{kT}} \tag{11.32}$$

which is known as the **Nernst-Einstein relation**. Note that it relates the **response** of the drift velocity to the external electric field) to **fluctuations** in the equilibrium state, governed by D. We now develop this idea a little further.

11.3 Correlation, response, and the fluctuation-dissipation theorem

We now show how this idea of relating response and dissipation to statistical fluctuations can be made more general. Consider some variable X (usually thermodynamic) of mean zero, that can fluctuate in time (e.g. a local magnetisation, m, or the local density in a fluid). X is measured as a departure from the global mean value i.e. it is a fluctuation about the mean value. We are interested in the **autocorrelation function** of these fluctuations in an equilibrium state, defined as

$$\langle X(\tau)X(\tau + t) \rangle. \tag{11.33}$$

In equilibrium (steady state) this must be independent of the initial time τ, hence

$$\langle X(\tau)X(\tau + t) \rangle = M_{XX}(t), \tag{11.34}$$

i.e. M_{XX} is only a function of the time difference t. A typical $M_{XX}(t)$ is sketched in Fig. 11. The correlations will tend to zero at large time lag, in a way that is often modelled with an exponential behaviour involving a **correlation time**, t_c:

$$M_{XX}(t) \propto \exp(-t/t_c). \tag{11.35}$$

Different fluctuating variables can also be correlated in time with each other, for example the magnetisation at two nearby places is correlated. To study this, we can define similarly the **c**ross-correlation function:

$$\langle X(\tau)Y(\tau + t) \rangle = M_{XY}(t), \tag{11.36}$$

for any pair of variables, X and Y.

Now the principle of detailed balance (which as we saw arises from microscopic reversibility) implies that there are no currents flowing in equilibrium. Once equilibrium is reached there is thus no direction of time, since there is no current that could distinguish the forward and backward time directions. Detailed balance therefore implies that the fluctuations arising in equilibrium are time-reversal symmetric

$$M_{XY}(t) = \langle X(\tau)Y(\tau + t) \rangle = M_{XY}(-t). \tag{11.37}$$

Moreover, the cross-correlation obeys

$$M_{YX}(t) = \langle Y(\tau)X(\tau + t) \rangle = \langle X(\tau')Y(\tau' - t) \rangle = M_{XY}(-t) \tag{11.38}$$

(defining $\tau' \equiv \tau + t$ as a new time to average over). Combining this with the previous result, we find that *the cross-correlation function is symmetric in the indices X and Y*:

$$\boxed{M_{XY}(t) = M_{YX}(t)}. \tag{11.39}$$

This is a rather nontrivial consequence of microscopic time reversal symmetry.

Response function Suppose we now perturb our system by applying a small 'thermodynamic force' f_X associated with a quantity X: as in section 5, this adds a perturbation $-f_X X$ to the Hamiltonian. Most simply, X would be position and $f_X = F_{\text{ext}}$. But many general alternatives exist: for example, if X a local magnetisation, m, then $f_X = H$, the applied magnetic field. There are various experimental 'protocols' for studying the effect of such perturbations:

- Consider a perturbation that remains turned on for a long time (say from $t = -\infty$) and is then switched off at (say) $t = 0$. For $t > 0$ the resulting average response of Y decays away.

- Or we can give the system a sharp 'kick' at $t = 0$ and see how it relaxes. This gives the Green's function of the system.

In either case, we can define the **response function**, $R_{YX}(t)$ (an effective susceptibility):

$$\boxed{\langle Y(t) \rangle \equiv R_{YX}(t)\, f_X}. \tag{11.40}$$

Now, if the perturbation is small enough, it will (at least in each local neighbourhood) produce a change so weak that *this could have arisen by a spontaneous fluctuation*. If that is the case, the perturbation $Y(t)$ will on average decay in time just as if it were such a fluctuation, which happened by chance to be present at $t = 0$.

Fluctuation-dissipation theorem This idea that the perturbation can be treated as an 'honorary fluctuation' is of deep importance, since it means that the way in which spontaneous fluctuations are correlated in time is something that can be approached by studying the response to imposed external forces. The explicit connection between the two is given by the following *fluctuation-dissipation theorem*, in which the correlation function $M_{XY}(t)$ ('fluctuation') is directly proportional to the response function $R_{XY}(t)$ ('dissipation'):

$$\boxed{M_{YX}(t) = kT\, R_{YX}(t)}. \tag{11.41}$$

We have quoted this relation without proof. The factor kT can be motivated to within a numerical factor by dimensional analysis: $[Y] = [R][f]$, so $[Y][X] = [R][fX]$. But $[M] = [Y][X]$, so $[M] = [fX][R]$: thus the coefficient of proportionality must have dimensions of energy, which is satisfied by kT. We do not attempt a proof here as it would be too lengthy (see the final chapter of the textbook by Chandler for a relatively simple exposition). In the following section we will

however look at a particular example in Brownian motion, for which is straightforward to verify the theorem.

Onsager theorem   Finally, a remarkable theorem can be obtained by combining two previous results given above the fluctuation-dissipation theorem and symmetry with respect to its indices of the correlation function.

$$\boxed{R_{XY}(t) = R_{YX}(t)} \tag{11.42}$$

that is, *the response function is itself symmetric in the X, Y indices.* This, in essence, is Onsager's theorem, which states that the mean response of a variable X to a small thermodynamic force f_Y acting on some other variable Y, is entirely determined by the mean response of Y to f_X.

Onsager's theorem is a remarkably subtle result, in which the *time reversibility* of microphysics strongly constrains the *time irreversible* relaxation of a macroscopic system perturbed away from equilibrium. Onsager's achievement was not just to prove the result, but to realise that such a connection could possibly exist in the first place. This is the same Lars Onsager who solved the 2D Ising Model (see below) and he was awarded the Nobel prize (in Chemistry) in 1968.

Application: thermoelectricity   Suppose you have a system consisting of two blocks of metal in contact. Let the left hand block be hotter than the right one. In that case, not only heat *but also particles* (in this case, electrons) will cross from left to right. This is easy to understand: hotter electrons, on the left, are more likely to travel to the right than the cold ones on the right are, to travel to the left. This is the *thermoelectric effect.*

Now take a different experiment where the temperatures are the same but the chemical potentials of the electrons are not. That can be arranged by having a voltage difference between the two blocks. Onsager's theorem states that *not only particles, but also heat energy* will flow from one block to the other. That is not so easy to understand. It is called the **Peltier effect**, and is widely used to make solid-state heat pumps. Onsager's theorem correctly predicts the magnitude of the Peltier effect from measurements of the thermoelectric effect. There are many similar, and equally unintuitive, applications of the theorem.

11.4   Over-damped Brownian motion

Let us return to the Brownian motion of a colloidal particle suspended in a fluid acted upon by a systematic external force F_{ext}. For convenience we shall also add a harmonic restoring force (think of the particle tethered by a weak spring to the origin) with spring constant κ. This is basically a trick to keep the displacement of the particle bounded as $t \to \infty$. The Langevin equation now reads

$$m\frac{d^2 x}{dt^2} = -\gamma\frac{dx}{dt} + F_{\text{ext}} - \kappa x + f(t) , \tag{11.43}$$

where the statistics of the random force f is as before.

We consider the case of the external force switched on up to $t = 0$ then switched off. To simplify things we consider the *overdamped* regime where we can ignore the acceleration term on the LHS of the equation; this is justifiable if the viscosity γ is large enough. Then

$$\frac{dx}{dt} = \mu\left[F_{\text{ext}} - \kappa x + f(t)\right] \quad \text{where} \quad \mu = \frac{1}{\gamma} , \tag{11.44}$$

(i.e. as terminal velocity $\langle dx/dt \rangle = 1/\gamma\, F_{\text{ext}}$, as $\langle x \rangle$ then vanishes, so we can take the mobility as $\mu = 1/\gamma$). Integrating the equation gives

$$\frac{d}{dt}\left[x e^{\mu\kappa t}\right] = \mu e^{\mu\kappa t}\left(F_{\text{ext}} + f(t)\right) , \tag{11.45}$$

or

$$x(t) = \mu \int_{-\infty}^{0} dt' e^{-\mu\kappa(t-t')} \left(F_{\text{ext}} + f(t') \right) + \mu \int_{0}^{t} dt' e^{-\mu\kappa(t-t')} f(t') . \tag{11.46}$$

The average displacement is

$$\langle x(t) \rangle_{F_{\text{ext}}} = \frac{1}{\kappa} e^{-\mu\kappa t} F_{\text{ext}} . \tag{11.47}$$

[So for $\mu\kappa t \ll 1$ then $\langle x(t) \rangle_{F_{\text{ext}}} = F_{\text{ext}}/\kappa - F_{\text{ext}}\mu t$ i.e. the particle moves back from F_{ext}/κ to zero.]
Comparing with $\langle x \rangle_f = Rf$ gives the response function:

$$\boxed{R_{xx} = \frac{1}{\kappa} e^{-\mu\kappa t} .} \tag{11.48}$$

Next we consider the correlation function

$$M_{xx}(t) = \langle x(\tau)x(\tau + t) \rangle , \tag{11.49}$$

when there is no external force present. In this case $x(t)$ becomes

$$x(t) = \mu \int_{-\infty}^{t} dt' e^{-\mu\kappa(t-t')} f(t') , \tag{11.50}$$

and the correlation function can be computed

$$M_{xx}(t) = \mu^2 \int_{-\infty}^{\tau} dt' e^{-\mu\kappa(\tau-t')} \int_{-\infty}^{\tau+t} dt'' e^{-\mu\kappa(\tau+t-t'')} \langle f(t')f(t'') \rangle \tag{11.51}$$

$$= \mu^2 \Gamma \int_{-\infty}^{\tau} dt' e^{-\mu\kappa(t+2\tau-2t')} \tag{11.52}$$

$$= \frac{\mu\Gamma}{2\kappa} e^{-\mu\kappa t} = \frac{\mu\Gamma}{2} R_{xx} . \tag{11.53}$$

In section 11.2 we derived the relation

$$\Gamma = 2\gamma kT = \frac{2kT}{\mu} \Rightarrow \mu\Gamma/2 = kT , \tag{11.54}$$

which we can now see recovers the fluctuation-dissipation theorem, $M_{xx} = kTR_{xx}$.

12 Many-body problem: example of vibrations in solids

In this section the many-body problem is discussed with reference to a particular context: vibrations of a crystalline solid. We will see that many of the general statistical tools developed in earlier sections find fruitful applications here.

12.1 Recap of many-particle Schrödinger equation

Recall the time independent many-particle Schrödinger equations

$$H\Psi(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N) = E\Psi(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N) , \tag{12.1}$$

where

$$H = -\sum_{k=1}^{N} \frac{\hbar^2}{2m} \nabla_k^2 + U(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N) , \tag{12.2}$$

where k labels the particles and U is the **interaction potential**, which generally depends on the relative positions of all the particles. In general U contains interactions between particles, and then a direct solution of the Schrödinger equation is not feasible. But if there are no interaction terms, then we have seen that the equation can be solved (12.1). In this case, the Hamiltonian is

$$H = \sum_{k=1}^{N} \left[-\frac{\hbar^2}{2m} \nabla_k^2 + V(r_k) \right] = \sum_{k=1}^{N} h_k \qquad (12.3)$$

and the many-body eigenfunctions are factorised

$$\Psi(r_1, r_2, \ldots, r_N) = \prod_{k=1}^{N} \psi_{\alpha_k}(r_k) \qquad \text{with} \quad E = \sum_{k=1}^{N} \epsilon_{\alpha_k} . \qquad (12.4)$$

We refer to this as a **diagonal** form of the Hamiltonian since the particles are not coupled by any 'off-diagonal' terms. Crudely speaking, much of many-body physics is concerned with making transformations and/or approximations that render the Hamiltonian into a diagonal form.

12.2 Einstein's theory of specific heat

Einstein replaced the full complicated Hamiltonian of a crystalline solid with a simple weakly interacting H involving only single-particle potentials. Consider N atoms situated near the sites of a regular lattice, and let r_k denote the displacement vector of atom k from its lattice site. Suppose each atom sits in its own quadratic potential well and oscillates independently of the others. The Hamiltonian is then

$$H = \sum_{k=1}^{N} \left[-\frac{\hbar^2}{2m} \nabla_k^2 + +\frac{1}{2} m\omega^2 |r_k|^2 \right] , \qquad (12.5)$$

where ω is a free parameter that we can be fixed by comparing to experimental data. The Hamiltonian is thus a sum of N 3D **quantum harmonic oscillators**, all having the same **angular frequency**, ω.

A further simplification is that a three-dimensional harmonic oscillator is the same as 3 one-dimensional harmonic oscillators. Therefore we can consider the system as $3N$ one-dimensional harmonic oscillators which we label by k

$$E = \sum_{k=1}^{3N} \epsilon_k \qquad \text{where} \quad \epsilon_k = \hbar\omega(n_k + 1/2) . \qquad (12.6)$$

Where n_k gives the energy level of oscillator k. Recall from quantum mechanics that n_k is an integer taking values $0, 1, \ldots \infty$ The partition function is then

$$Z = [Z(1)]^{3N} , \qquad (12.7)$$

where

$$Z(1) = \sum_{n=0}^{\infty} \exp[-\beta\hbar\omega(n + 1/2)] \qquad (12.8)$$

$$= \frac{e^{-\alpha/2}}{1 - e^{-\alpha}} \qquad \text{where} \quad \alpha = \beta\hbar\omega , \qquad (12.9)$$

(summing the geometric series). Up to an unimportant factor $e^{-\alpha/2}$, our single oscillator partition function $Z(1)$ is the same as a single state partition function for Bosons (with chemical potential $\mu = 0$). So we can think of the quanta of energy in the oscillator as Bosons.

We now proceed to calculate the average system energy

$$\overline{E} = -\frac{\partial \ln Z}{\partial \beta} \tag{12.10}$$

$$= -3N\frac{\partial}{\partial \beta}\left[-\frac{\alpha}{2} - \ln\left(1 - e^{-\alpha}\right)\right] \tag{12.11}$$

$$= \text{const.} + \frac{3N\hbar\omega}{e^{\beta\hbar\omega} - 1}. \tag{12.12}$$

The heat capacity is given by

$$C_V = \left(\frac{\partial \overline{E}}{\partial T}\right)_V = -k\beta^2\frac{\partial}{\partial \beta}\left(\text{const.} + \frac{3N\hbar\omega}{e^{\beta\hbar\omega} - 1}\right), \tag{12.13}$$

or

$$\boxed{C_V = 3Nk\left(\hbar\omega\beta\right)^2 \frac{e^{\beta\hbar\omega}}{(e^{\beta\hbar\omega} - 1)^2}.} \tag{12.14}$$

This has the following limits. At high temperature, expanding the exponentials gives

$$C_V = 3Nk, \tag{12.15}$$

which is the classical **equipartition** result: energy kT per quadratic degree of freedom. At low temperatures, $\beta \to \infty$:

$$C_V = 3Nk\left(\hbar\omega\beta\right)^2 e^{-\beta\hbar\omega}. \tag{12.16}$$

This tends to zero as $\beta \to \infty$, as observed. But in detail the specific heat at low temperatures is seen to scale as T^3, which is not the dependence predicted by the Einstein model. The reason for this discrepancy is that co-operative effects become important at low temperatures, so that we need to consider collective oscillations of the atoms rather than the independent oscillations of the Einstein theory.

12.3 Harmonic approximation, phonons and the Debye model

In general the potential energy U could be a very complicated function of the atomic displacements, $U(\underline{r}_1, \underline{r}_2, \ldots \underline{r}_N)$. But assuming the displacements to be small, we can make a Taylor series expansion (in $3N$ variables $x_i^{(k)}$):

$$U = U_0 + \sum_{ki}\left.\frac{\partial U}{\partial x_{ki}}\right|_0 x_{ki} + \frac{1}{2}\sum_{ki,lj}\left.\frac{\partial^2 U}{\partial x_{ki}\partial x_{lj}}\right|_0 x_{ki}x_{lj} + \cdots, \tag{12.17}$$

Since U must be a minimum at the equilibrium positions the second term (linear in the x's) vanishes. Retaining only the final term on the RHS is known as the **harmonic approximation**. Abbreviating the second partial derivatives to $A_{ki;lj}$ we may write the total Hamiltonian as

$$H = -\frac{\hbar^2}{2m}\sum_{k=1}^{N}\nabla_k^2 + \frac{1}{2}\sum_{ki,lj}A_{ki,lj}x_{ki}x_{lj}. \tag{12.18}$$

Now, because of the structure of the harmonic approximation we can diagonalise this Hamiltonian by transforming to appropriate coordinates. This is because it is simply the Hamiltonian of a system of $3N$ **coupled harmonic oscillators**, which you will have studied in dynamics courses. Systems of coupled oscillators have **normal modes** in which all the displacements oscillate with the same frequency. Classically, this arises because the equation of motion for one

coordinate of one particle, q is $\ddot{q} = -\partial V/\partial q$, where V is the total potential energy of the system. In the harmonic approximation, this is $(1/2)(\partial^2 V \partial q_i \partial q_j) q_i q_j$, so that differentiation gives $\ddot{\mathbf{q}} = -\mathbf{V}'' \cdot \mathbf{q}$, where $\mathbf{V}''$ is the matrix of 2nd derivatives. This has oscillatory solutions, where ω^2 is the eigenvalue of $\mathbf{V}''$ and the eigenvectors are the normal modes. The motion of the system can then be expressed in terms of a superposition of normal mode excitations and the energy can be expressed as the sum of the energy in each normal mode. Since there are $3N$ co-ordinates there are $3N$ normal modes. The diagonalisation involves some linear transformation to **normal coordinates**:

$$q_r = \sum_{ki} L_{r,ki} x_{ki} \,, \qquad r = 1, \ldots, 3N \,, \tag{12.19}$$

where $L_{r,ki}$ are the components of the transformation matrix. By design, this will convert the Hamiltonian into the form

$$H = \sum_{r=1}^{3N} \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial^2 q_r} + \frac{1}{2} m \omega_r^2 q_r^2 \right] \,. \tag{12.20}$$

The normal modes are quantum oscillators, and the quanta of energy are known as **phonons**. Phonons really do exist and are much studied experimentally e.g. via the scattering neutrons or X-rays by a solid. The energy of the system is then of the usual oscillator form:

$$E = \sum_{r=1}^{3N} \hbar \omega_r (n_r + 1/2) \qquad n_r = \quad \text{number of phonons in mode } r \,. \tag{12.21}$$

We can compute the canonical partition function my summing over modes

$$Z = \prod_{r=1}^{3N} Z_r \qquad \text{where} \quad Z_r = \frac{e^{-\alpha_r/2}}{1 - e^{-\alpha_r}} \quad \text{with} \quad \alpha_r = \beta \hbar \omega_r \,. \tag{12.22}$$

The (average) energy follows as

$$\overline{E} = -\frac{\partial \ln Z}{\partial \beta} \tag{12.23}$$

$$= \text{const.} + \sum_r \frac{\hbar \omega_r}{e^{\beta \hbar \omega_r} - 1} \,. \tag{12.24}$$

It is normal to ignore the first term, as it is just a constant. The second can be compared to the energy of an ideal Bose gas at $\mu = 0$. The phonons behave with Bose statistics and $1/(\exp(\beta \hbar \omega_r) - 1)$ is the mean number of phonons in mode r. Thus we have a gas of **free phonons** (the chemical potential $\mu = 0$ since the number of phonons is not a conserved quantity).

The algebra here is identical to that in the case of the Einstein model, except that the frequencies of the different oscillators need not be the same (and they are not, as we will shortly demonstrate). If the number of modes is large, we may suspect that they will be finely spaced in frequency, so that the total energy can be written using a **density of modes**, $g(\omega)$:

$$\boxed{\overline{E} = \int \frac{\hbar \omega}{e^{\beta \hbar \omega} - 1} g(\omega) \, d\omega \,.} \tag{12.25}$$

Calculating this density exactly for a discrete crystal lattice is not straightforward, but the problem can be solved with the following useful approximation.

12.3.1 Debye theory

In the Debye theory we neglect the fine structure of the crystal lattice, and treat the solid instead as a continuous **elastic** body. For this, the normal modes will be sound waves, and we can argue that these will be insensitive to the existence of a lattice as long as the wavelength is $\gg$ the lattice spacing. But equally, this approach will not be exact for phonons of the shortest wavelengths. Previously we derived the density of states for wave modes based on the density in k-space:

$$g(\epsilon) = \frac{\Gamma(k)}{d\epsilon/dk} \; ; \qquad \Gamma(k) = g_s \frac{V}{2\pi^2} k^2 \,, \tag{12.26}$$

where we have included a degeneracy factor g_s. Now, for wave modes we have the standard relation between wavenumber and angular frequency

$$k = \frac{\omega}{c_s} \,, \tag{12.27}$$

where c_s is the speed of sound (taken here as constant). There will be two transverse and one longitudinal wave modes, so $g_s = 3$. Changing variables to ω using $g(\omega)d\omega = \Gamma(k)dk$, we get

$$g(\omega) = 3\frac{V}{2\pi^2} \frac{\omega^2}{c_s^3} \equiv AV\omega^2 \,. \tag{12.28}$$

This gives an **ultraviolet catastrophe**: a divergent number of modes if we integrate to $\omega = \infty$; but there should be only $3N$ normal modes in total. To respect this we introduce a 'cut-off' frequency ω_{max}:

$$\int_0^{\omega_{\text{max}}} AV\omega^2 d\omega = \frac{1}{3} AV\omega_{\text{max}}^3 = 3N \,, \tag{12.29}$$

from which we find

$$\omega_{\text{max}} = \left(\frac{9}{A}\frac{N}{V}\right)^{1/3} = \left(6\pi^2 \frac{N}{V}\right)^{1/3} c_s \,. \tag{12.30}$$

This maximum frequency defines a characteristic temperature, Θ_D, through

$$k\Theta_D = \hbar\omega_{\text{max}} \,. \tag{12.31}$$

We now proceed to calculate the energy through

$$
\begin{aligned}
\overline{E} &= \int_0^\infty \hbar\omega \, g(\omega) \, \overline{n(\omega)} \, d\omega \\
&= AV\hbar \int_0^{\omega_{\text{max}}} \frac{\omega^3}{\exp(\beta\hbar\omega) - 1} \, d\omega \,.
\end{aligned}
$$

First consider the high T limit $T \gg \Theta_D (= \hbar\omega_{\text{max}}/k_B)$ which implies $\hbar\omega/k_B T \ll 1$. We expand $\exp(\beta\hbar\omega) \simeq 1 + \beta\hbar\omega$ to obtain

$$\overline{E} \simeq AV\hbar \int_0^{\omega_{\text{max}}} \frac{w^3}{\beta\hbar\omega} \, d\omega = 3Nk_B T \tag{12.32}$$

where we have used (12.29). As with the Einstein theory, this is just the classical equipartition result.

Now consider the low T limit $T \ll \Theta_D$. In this case we change variables to $x = \beta\hbar\omega$ to express the result in terms of a dimensionless integral:

$$\overline{E} = \frac{AV\hbar}{(\beta\hbar)^4} \int_0^{\Theta_D/T} \frac{x^3}{\exp(x) - 1} \, dx$$

But for $T \ll \Theta_D$ we can replace the upper limit of the integral by infinity, which yields a standard integral whose value is $\pi^4/15$. Thus at low temperatures $\overline{E} \propto T^4$ and the heat capacity

$$\boxed{C_V \propto T^3.}$$
(12.33)

This is the key result of the Debye theory and it improves greatly upon the Einstein model prediction. It is well verified experimentally for e.g. for copper or solid Argon (see Baierlein Fig. 6.9). The reason this works is that at low T only low frequency modes will be excited and it is precisely these long-wavelength modes that are correctly described by approximating the solid as an elastic medium. Thus the Debye theory is correct at both low and high T, but remains an approximation in between.

13 Interactions in classical gases

We now turn our attention to another many-particle interacting system – classical gases in which the particles interact weakly (or interact strongly, but only on rare occasions).

13.1 Recovery of classical ideal gas

As a starting point consider again the ideal gas, where a particle's energy is simply its kinetic energy,

$$\epsilon_i(\underline{q}_i, \underline{p}_i) = \frac{p_i^2}{2m},$$
(13.1)

and the partition function is just

$$
\begin{aligned}
Z_{\text{ideal}} &= \frac{V^N}{N!\,h^{3N}} \prod_{i=1}^{N} \left[\int d^3 p_i\, e^{-\beta p_i^2/2m} \right] = \frac{V^N}{N!\,h^{3N}} \left[\int dp\, e^{-\beta p^2/2m} \right]^{3N}
\end{aligned}
$$
(13.2)

$$
= \frac{V^N}{N!\,h^{3N}} \left[\left(\frac{2m\pi}{\beta} \right)^{1/2} \right]^{3N}
$$
(13.3)

$$
= \frac{1}{N!} \left[\frac{V}{\lambda_T^3} \right]^N .
$$
(13.4)

(cf. equation 6.5). We have used the formula for the Gaussian integral

$$\int_{-\infty}^{\infty} dz\, e^{-\alpha z^2} = \left(\frac{\pi}{\alpha} \right)^{1/2}$$
(13.5)

and the definition of the thermal de Broglie wavelength

$$\lambda_T = \left(\frac{h^2}{2\pi m k T} \right)^{1/2} .$$
(13.6)

Note that h is present in Z via λ_T. Thus, quantities stemming from $\ln Z$ such as the free energy $F = -kT \ln Z$ or entropy S will retain a dependence on h, as an additive constant. However it should be noted that it is only free energy and entropy *differences* that can be measured experimentally (absolute values of free energy are not measured) and in the expression for e.g. a free energy difference the h dependence cancels – as we would hope for a classical system.

Explicitly we have (using Stirling's formula)

$$F_{\text{ideal}} = -kT \ln Z_{\text{ideal}} \quad = \quad kTN \left[\ln N - 1 - \ln V + N \ln \lambda_T^3 \right] \tag{13.7}$$

$$= \quad -kTN \ln \frac{Ve}{N\lambda_T^3} \tag{13.8}$$

so for example we have

$$P_{\text{ideal}} \quad = \quad -\left. \frac{\partial F_{\text{ideal}}}{\partial V} \right|_{T,N} = \frac{NkT}{V}, \tag{13.9}$$

$$S_{\text{ideal}} \quad = \quad -\left. \frac{\partial F_{\text{ideal}}}{\partial T} \right|_{V,N} = kN \left[\ln \frac{V}{N} \left(\frac{2\pi mkT}{h^2} \right)^{3/2} + \frac{5}{2} \right] \tag{13.10}$$

as expected. The latter expression is the **Sackur–Tetrode entropy**.

13.2 Interactions as a perturbation of the ideal gas

A classical gas comprises a system with N particles each of mass m, in a volume V. We assume that the density

$$\rho = \frac{N}{V}, \tag{13.11}$$

is low enough that we can

- treat interactions as perturbations about the ideal gas

- (mostly) neglect quantum effects and use classical statistical mechanics.

In the interacting case an important simplification arises from the fact that the interaction potential $U(\underline{q}_1, \ldots, \underline{q}_N)$ does not depend on the particle momenta. Then

$$H(\{\underline{q}\}, \{\underline{p}\}) = \sum_{i=1}^{N} \frac{p_i^2}{2m} + U(\underline{q}_1, \ldots, \underline{q}_N), \tag{13.12}$$

and Z factors into two pieces:

$$Z(T, V, N) = \frac{1}{N! h^{3N}} \int \prod_i d^3 p_i \, e^{-\beta p_i^2 / 2m} \int \prod_i d^3 q_i \, e^{-\beta U(\{\underline{q}\})}, \tag{13.13}$$

corresponding to integrals over positions, and integrals over momenta. These factors can conveniently be separated (in dimensionless form) as follows:

$$Z(T, V, N) = Z_{\text{ideal}} \, Q, \tag{13.14}$$

where the first factor is the partition function for the ideal gas,

$$Z_{\text{ideal}} = \frac{[V/\lambda_T^3]^N}{N!}. \tag{13.15}$$

The second factor is normalised so that it takes value unity for ideal non-interacting gases:

$$\boxed{Q = V^{-N} \int \prod_{i=1}^{N} d^3 q_i \, e^{-\beta U(\underline{q}_1, \ldots, \underline{q}_N)}}; \tag{13.16}$$

this is known as the **configurational integral**. Accordingly, the interactions between our particles enter only via Q, and if we can evaluate this we have all we need to compute $Z(T, V, N)$.

From the factorisation of Z there follows the decomposition of the free energy

$$F(T, V, N) = F_{\text{ideal}}(T, V, N) - kT \ln Q, \tag{13.17}$$

and of the equation of state,

$$P(T, V, N) = -\left.\frac{\partial F}{\partial V}\right|_{T,N} = \frac{NkT}{V} + P_{\text{conf}}. \tag{13.18}$$

Thus the **configurational pressure** is

$$P_{\text{conf}} = kT \frac{\partial \ln Q}{\partial V}, \tag{13.19}$$

which gives the correction to the ideal gas equation of state, i.e. the correction to the pressure due to interactions.

13.2.1   Virial expansion

In the following we take the interaction potential to be of the form

$$U(\{\underline{q}\}) = \frac{1}{2} \sum_{i \neq j} \phi(|\underline{q}_i - \underline{q}_j|) \equiv \sum_{i<j} \phi_{ij}, \tag{13.20}$$

Note that this implies

- 2 body interactions

- a central potential (depends only on distance between particles)

Examples of interatomic potentials are as in Fig. 13.

- The hard-sphere potential

$$\phi(r) = \begin{cases} \infty & r \leq a \\ 0 & r > a \end{cases} \tag{13.21}$$

- the Lennard–Jones potential

$$\phi(r) = 4\epsilon[(r_0/r)^{12} - (r_0/r)^6]. \tag{13.22}$$

Here ϵ is related to the depth of the attractive well and r_0 the hard-sphere radius.

For the Lennard–Jones case note the **generic features** of repulsion at short distance, attraction at intermediate distance and interaction $\to 0$ at long range. In the generic case, there is no intrinsic limit to the number of particles that can be interacting simultaneously with each other. Consequently new **collective phenomena**, such as the phase transition from a vapour to a liquid, become possible. The hard-sphere case, on the other hand, is not generic and is a special case.

Our task is to calculate the configurational integral (13.16) which we write in the form

$$Q = \frac{1}{V^N} \int \prod_i d^3 q_i \prod_{i<j} F_{ij} , \qquad \text{where} \quad F_{ij} = e^{-\beta \phi_{ij}} . \tag{13.23}$$

Just to be clear, let us write out the products

$$Q = \frac{1}{V^N} \int d^3 q_1 \dots d^3 q_N \, F_{12} F_{13} \dots F_{1N} \times F_{23} \dots F_{2N} \times \dots \times F_{N-1 \, N} . \tag{13.24}$$

Note that there are $(N-1) + (N-2) + \dots + 1 = N(N-1)/2$ F_{ij}'s in the product. This is just the number of ways of choosing 2 sites from N. We can think of Q as being a spatial average i.e. it is the integral of something (the product of F_{ij}) over the spatial co-ordinates divided by the volume V^N. Therefore we can write

$$Q = \left\langle \prod_{i<j} F_{ij} \right\rangle , \tag{13.25}$$

where the angle brackets indicate the spatial average. Now this is very difficult to calculate because, for example F_{ij} is **correlated** with F_{ik}. But if we ignore the correlations, then we can replace the average of the product by the product of the averages:

$$Q \simeq \prod_{i<j} \langle F_{ij} \rangle \tag{13.26}$$

$$= \langle F \rangle^{N(N-1)/2} , \tag{13.27}$$

i.e. we replace the average of the product by the product of the averages. Note that $\langle F_{ij} \rangle = \langle F \rangle$ does not depend on the indices i,j. Furthermore, we can write

$$F_{ij} = 1 + f_{ij} , \tag{13.28}$$

where $f_{ij} = \exp[-\beta \phi_{ij}] - 1$ will typically be small in a dilute gas, where the interparticle separations are large and ϕ_{ij} is small.

Let us take $i, j = 1, 2$ as representative

$$\langle F \rangle \equiv \langle F_{12} \rangle = 1 + \frac{1}{V^N} \int d^3 q_1 \dots d^3 q_N \, f_{12} \tag{13.29}$$

$$= 1 + \frac{1}{V^2} \int d^3 q_1 \, d^3 q_2 \, f_{12} , \tag{13.30}$$

where we have integrated out the coordinates not involved in f_{12} and used $\int d^3 q = V$. Now we change variables to centre of mass and relative coordinates

$$\underline{r} = \underline{q}_1 - \underline{q}_2 , \qquad \underline{R} = \frac{1}{2}(\underline{q}_1 + \underline{q}_2) . \tag{13.31}$$

[Consider each component separately: each of the three Jacobians is 1, so $d^3 r \, d^3 R = d^3 q_1 \, d^3 q_2$; see tutorial sheet.] Noting that there is only dependence on the relative coordinate $\underline{r}$ so that $\int d^3 R = V$, we obtain

$$\langle F_{12} \rangle = 1 + \frac{1}{V} \int d^3 r \, [\exp -\beta \phi(r) - 1] , \tag{13.32}$$

which we write as

$$\langle F_{12} \rangle = 1 - \frac{2 B_2}{V} , \tag{13.33}$$

where $\hspace{4cm}$ [note: $d^3r = 4\pi r^2 dr$]

$$B_2 = -\frac{1}{2}\int d^3r\left[e^{-\beta\phi(r)} - 1\right] \tag{13.34}$$

The constant B_2 is known as the **second virial coefficient**, and it depends on the form of the interaction potential ϕ. The term 'virial' was introduced by Clausius and literally means 'pertaining to interactions between molecules'.

So within our crude approximation we have

$$Q = \left(1 - \frac{2B_2}{V}\right)^{N(N-1)/2}, \tag{13.35}$$

and following (13.17) we find

$$F = F_{\text{ideal}} - kT\ln Q \simeq F_{\text{ideal}} + \frac{N^2 kT}{V}B_2, \tag{13.36}$$

where we have approximated $N(N-1) \simeq N^2$ and $\ln(1 - 2B_2/V) \simeq -2B_2/V$ as N, V are large. Then, from (13.18, 13.19) we obtain

$$\frac{P}{kT} = n + B_2 n^2 \tag{13.37}$$

where $n = N/V$. Finally we note

$$S = -\left.\frac{\partial F}{\partial T}\right|_V = S_{\text{ideal}} - n^2 kV\frac{\partial}{\partial T}(TB_2). \tag{13.38}$$

Notes

- This is the start of the **virial expansion**. This is a **perturbation expansion** in the number density n about the ideal gas limit. That is, we see the start of a series

$$\frac{P}{kT} = n + B_2 n^2 + B_3 n^3 \cdots \tag{13.39}$$

- A simple approximation for a generic ϕ_{ij} (see tutorial sheet) yields

$$B_2 = b_0 - a_0/kT, \tag{13.40}$$

where b_0 and a_0 are positive constants (see tutorial sheet). This recovers the expansion of the **van der Waals equation of state**, which is usually written as

$$\left(P + n^2 a_0\right) = \frac{nkT}{1 - nb_0}. \tag{13.41}$$

Also one can show the entropy is reduced below that of the ideal gas. This is to be expected from information theory: the attractive interactions should make the positions of the atoms correlated. Correlations in the positions then reduce the hidden information about the microscopic state of the system. *This is an important general principle*

75

13.2.2 Extension to higher order (non-examinable)

To improve upon our crude approximation, and in order to calculate the higher order virial coefficients let us write

$$F_{ij} = \langle F \rangle + \lambda_{ij}, \tag{13.42}$$

where λ_{ij} is the deviation of F_{ij} from its average spatial value. Then

$$Q = \langle (\langle F \rangle + \lambda_{12}) (\langle F \rangle + \lambda_{13}) \cdots \rangle, \tag{13.43}$$

which we wish to expand in powers of λ_{ij}. We do not explicitly carry out this procedure here, but it results in

$$Q = \langle F \rangle^{N(N-1)/2} + \binom{N}{3} \langle F \rangle^{N(N-1)/2 - 3} \langle \lambda_{12} \lambda_{23} \lambda_{13} \rangle + \dots \tag{13.44}$$

Notes

- Because $\langle \lambda_{ij} \rangle = 0$, an average of a product of λ's in which a subscript appears only once is zero. Therefore there are no linear or quadratic terms in the above expansion.

- The first term is the approximation derived above. The second term involves a 'cluster' of three sites say 123 and $\langle \lambda_{12} \lambda_{23} \lambda_{13} \rangle$ is representative of such terms. The binomial coefficient is the number of ways of choosing three sites from N.

- Physically the first term, which involved averaging F_{ij}, represents the interaction of two particles i.e. a 'two particle molecular cluster'. The second term, as noted above, involves 'three particle particle clusters', hence this is known as a **cluster expansion** or alternatively a **Mayer expansion**.

- To extend to higher orders diagrammatic methods are required to enumerate the various clusters and this works well for low densities This was actually the first instance of diagrammatic methods see e.g. the books by Huang or Pathria for details.

- But at higher densities the expansion at best converges slowly and fails to converge at all near the phase transition to the liquid phase the This is because near the phase transition larger and larger clusters of particles become important. Also perturbation theory about the ideal gas can hardly be expected to describe the liquid phase, which is a different state of matter.

14 Reduced density distributions and Debye–Hückel theory

This section treats classical fluids from the point of view of the two-point reduced density distribution, and is an alternative way to look at the virial expansion.

14.1 Correlations and distribution functions

A distribution function in the Gibbs sense is the probability density $\rho(q_1, \dots, q_{3N}, p_1, \dots, p_{3N})$ in $6N$-dimensional phase space, giving the simultaneous probability of finding each of the particles near their position and momentum. This can be **reduced** in a number of ways to depend on fewer variables. First, we can integrate over momentum coordinates to obtain a purely spatial PDF,

$\rho(q_1, \ldots, q_{3N})$; then, we can integrate over the positions of some of the particles to leave a PDF that governs the remaining m particles, In statistics, this process would be termed **marginalization**.

One-particle reduced density   This is simply

$$\rho_1(\underline{r}_1) \equiv \rho = N/V , \tag{14.1}$$

where N is the number of particles, V is the volume and the last relation assumes that the fluid is **homogeneous** so that its statistical properties do not depend on position.

Two-particle reduced density   The two-particle reduced density $\rho_2(\underline{r}_1, \underline{r}_2)$ is the probability density for finding a pair of particles at given positions $\underline{r}_1, \underline{r}_2$ – i.e. $\rho_2(\underline{r}_1, \underline{r}_2)dV_1dV_2$ is the probability of finding a particle in the volume dV_1 around $\underline{r}_1$ *and* finding a particle in dV_2 around $\underline{r}_2$. If the particles are independent this will be just $\rho_2 = \rho^2$, but in general there will be some degree of correlation:

$$\boxed{\rho_2(\underline{r}_1, \underline{r}_2) = \rho^2 g(|\underline{r}_1 - \underline{r}_2|)} \tag{14.2}$$

Here $g(r)$ is called the **radial distribution function** or **pair distribution**. It is sometimes also called the **pair correlation function**, but this is poor terminology, since uncorrelated particles will have $g = 1$. A better approach is to write $g = 1 + \xi$, where ξ is what should be called the correlation function (alternatively **auto-correlation function** or **2-point correlation function**).

Note that $g(r)$ depend only on distance between the two positions, reflecting the assumption that we are dealing with a **homogeneous** and **isotropic** fluid. To understand $g(r)$ more precisely recall that a joint probability, $\rho_2(\underline{r}_1, \underline{r}_2)$ may be written as a product

$$\rho_2(\underline{r}_1, \underline{r}_2) = \rho_2(\underline{r}_1|\underline{r}_2)\rho_1(\underline{r}_2), \tag{14.3}$$

where $\rho_2(\underline{r}_1|\underline{r}_2)$ is the **conditional probability** of a particle being at $\underline{r}_1$ *given* that there is a particle at $\underline{r}_2$ and the probability of there being a particle at $\underline{r}_1$ is just the density $\rho_1(\underline{r}_2) = \rho$. Thus comparing with (14.2) we deduce that $\rho g(r)$ is the conditional probability that there is a particle at a point a distance r from the origin, given that there is a particle at the origin.

It is worth making a careful distinction between the terms uncorrelated and independent. Independence always implies no correlation, but the converse is not strictly true, although the distinction is often ignored (can you envisage a 2D distribution that is uncorrelated but not independent?). In a system of non-interacting particles, the particles are independent and uncorrelated – but interactions between the particles will introduce correlations. Generally correlations between particles imply a gain in information i.e. the correlation means that if you know something about one particle you are likely to know something about other particles. Thus interactions generally generate correlations which reduce the hidden information or entropy.

- General features of $g(r)$ for a real (interacting) gas are:

    - $g(r) \to 1$ as $r \to \infty$ so that the density far away from a given particle is just the overall density ρ
    - $g(0) = 0$, so that with hard-sphere interactions it is impossible to have particles overlapping when $\underline{r}_1 = \underline{r}_2$
    - $g(r) > 1$ means the particles at $\underline{r}_1$ and $\underline{r}_2$ are correlated and $0 < g(r) < 1$ means the particles are anticorrelated.

For a typical dense fluid g(r) looks like the middle figure in Fig. 14. The radial distribution

function contains useful information about how the particles in the fluid are correlated, and can be measured rather directly by diffraction experiments. (see e.g. the book by Chandler, Ch.7).

Generally $g(r)$ is difficult to calculate – actually it contains the same information as the equation of state (see below). A simple approximation for $g(r)$ is to take

$$g(r) = e^{-\beta\phi(r)},\tag{14.4}$$

where ϕ is the two-particle potential. This is intuitively reasonable, as $g(r)$ then has the form of a Boltzmann weight: the probability of finding a particle at r given that there is one at the origin) is evaluated as if all the others were absent. This expression obeys the correct limits at $r \to \infty$ ($\phi \to 0$) and $r \to 0$ ($\phi \to \infty$).

14.2 The virial equation of state   (proof is non-examinable)

As before we assume that the interaction energy $U(\{r\})$ is pairwise additive: $U = \sum_{i<j} \phi(r_{ij})$. In this case a very useful formula for the pressure P can be found

$$P = P_{\text{ideal}} + P_{\text{conf}} = \rho kT - \frac{\rho^2}{6} \int_0^\infty \left(r\frac{d\phi}{dr}\right) g(r)\, 4\pi r^2 dr\tag{14.5}$$

This is a formula for the **virial equation of state** which is more general than the previous virial expansion in ρ and is exact for pairwise interactions.

A proof is as follows: introduce reduced coordinates $\tilde{r} = r/L$, where $L = V^{1/3}$. Then we can write the **configurational integral** of section 13.2 as

$$Q = \frac{1}{V^N} \int \prod_i d^3r_i\, e^{-\beta U(\{r\})} = \int \prod_i d^3\tilde{r}_i\, e^{-\beta \sum_{i<j} \phi(L\tilde{r}_{ij})},\tag{14.6}$$

where the limits of integration for each coordinate are now L-independent (previously they were all from 0 to L). We have $F_{\text{conf}} = -kT \ln Q$ and so

$$
\begin{aligned}
P_{\text{conf}} &= -\left.\frac{\partial F_{\text{conf}}}{\partial V}\right|_T = kT\frac{\partial}{\partial V}\ln Q &\text{(14.7)}\\[2mm]
&= \frac{kT}{3L^2}\frac{1}{Q}\frac{\partial}{\partial L}\int \prod_i d^3\tilde{r}_i\, e^{-\beta \sum_{i<j}\phi(L\tilde{r}_{ij})} &\text{(14.8)}\\[2mm]
&= -\frac{1}{3V}\frac{1}{V^N Q}\int \prod_i d^3r_i \sum_{i<j} r_{ij}\phi'(r_{ij})\, e^{-\beta \sum_{i<j}\phi(r_{ij})} &\text{(14.9)}\\[2mm]
&= -\frac{1}{3V}\int \prod_i d^3r_i \sum_{i<j} r_{ij}\phi'(r_{ij})\, \rho(r_1,\ldots,r_N), &\text{(14.10)}
\end{aligned}
$$

(where $\phi' = d\phi/dr$). Integrate over everything except the i, j^{th} particles

$$
\begin{aligned}
P_{\text{conf}} &= -\frac{1}{3V}\sum_{i<j}\int d^3r_i d^3r_j\, r_{ij}\phi'(r_{ij})\frac{\rho_2(r_i, r_j)}{N(N-1)} &\text{(14.11)}\\[2mm]
&= -\frac{1}{6V}\int d^3r_1 d^3r_2\, r_{12}\phi'(r_{12})\rho_2(r_1, r_2) &\text{(14.12)}\\[2mm]
&= -\frac{\rho^2}{6V}\int d^3r_1 d^3r_2\, r_{12}\phi'(r_{12})g(r_{12}). &\text{(14.13)}
\end{aligned}
$$

As before let $r = r_1 - r_2$, $R = (r_1 + r_2)/2$, $\int d^3R = V$ which finally proves the result.

Recovering the virial expansion   Expand $g(r)$ in powers of ρ:

$$g(r) = g_0(r) + \rho g_1(r) + O(\rho^2) . \tag{14.14}$$

From the previous virial expansion in section 13 we see that the leading correction to the pressure is obtained from $g_0(r)$. From this expansion

$$
\begin{aligned}
P_{\text{conf}} &= B_2 \rho^2 kT && (14.15) \\
&= -\frac{\rho^2}{2} \int_0^\infty \left[e^{-\beta\phi(r)} - 1 \right] 4\pi r^2 dr \times kT && (14.16) \\
&= -\frac{\rho^2}{2} \left[\left[[e^{-\beta\phi(r)} - 1] \tfrac{4}{3}\pi r^3 \right]_0^\infty - \int_0^\infty \frac{d}{dr} [e^{-\beta\phi(r)} - 1] \tfrac{4}{3}\pi r^3 dr \right] \times kT && (14.17) \\
&= -\frac{\rho^2}{6} \int_0^\infty r\phi'(r) e^{-\beta\phi(r)} 4\pi r^2 dr , && (14.18)
\end{aligned}
$$

after integrating by parts and noting that for large r, B_2 requires $\phi(r) \sim 1/r^{3+\epsilon}$ where $\epsilon > 0$ then $r^3(e^{-\beta\phi(r)} - 1) \sim 1/r^\epsilon$ and vanishes at large r. This recovers the previous expression; comparing these expressions then gives to leading order $g(r) = \exp[-\beta\phi(r)]$, as was justified informally above.

14.3   Debye–Hückel theory

An interesting application of these ideas is to a **plasma**: a fluid of charged particles. This may seem problematic at first glance, since the pairwise interaction is the $1/r$ **Coulomb potential** and the above expressions for the virial expansion diverge (tutorial sheet). We need a different approach, which will lead us to conclude that collective effects in the plasma result in **screening** of the long-range force – so that in practice the particles do nevertheless participate in effective weak short-range interactions.

N.B. In this section we change notation to use $n(r)$ for (number) density at r and $\rho(r)$ for charge density. We also use $\phi(r)$ for electrostatic potential and $U = e\phi$ for potential energy. For simplicity we consider a 'one-component' plasma, in which a gas of point particles of charge q at average number density n_∞ resides in a static uniform background of charge density $-qn_\infty$. This is a good model for classical electrons in a semiconductor, where the background is the fixed ionic charges, and the electron density is low enough for the classical limit to hold. And even in a free plasma, the huge difference in inertia between electrons and protons means that we make little error in treating the latter as stationary (see tutorial sheet).

Poisson–Boltzmann equation   In the previous section, we showed that the radial distribution function $g(r) = n(r)/n_\infty$ was given by a Boltzmann factor. The derivation given there does not apply for a slowly-declining Coulomb potential, but the result feels general (it is no different to e.g. the variation of gas density with height in an isothermal atmosphere). We therefore adopt this assumption and write

$$n(r) = n_\infty g(r) = n_\infty e^{-q\phi(r)/kT} . \tag{14.19}$$

In the end we will see that this is in fact consistent with the previous analysis.

The potential ϕ obeys Poisson's equation

$$\nabla^2 \phi(r) = -\rho(r)/\epsilon , \tag{14.20}$$

where ϵ is the dielectric constant of the surrounding medium. The charge density $\rho(r)$ consists of three contributions:

(i) That of the point charge q at the origin

$$\rho_0 = q\delta(\underline{r}). \tag{14.21}$$

(ii) That of the fixed background charge density

$$\rho_{\text{fixed}}(r) = -qn_\infty \tag{14.22}$$

(iii) That of all the other free charges in the plasma, which are themselves arranged around the central charge according to some yet unknown charge density

$$\rho_{\text{free}}(r) = qn_\infty g(r) = qn_\infty e^{-q\phi(r)/kT}. \tag{14.23}$$

We therefore have $\rho = \rho_0 + \rho_{\text{free}} + \rho_{\text{fixed}}$ and hence

$$\nabla^2 \phi = -\frac{n_\infty q}{\epsilon}[e^{-\beta q\phi} - 1] - \frac{q}{\epsilon}\,\delta(\underline{r}). \tag{14.24}$$

This nonlinear **Poisson–Boltzmann equation** can be solved numerically to give $\phi(r)$. To see what sort of physics is involved, consider the limit where $\phi \ll kT$ (low charge densities and/or high temperatures):

$$e^{-\beta q\phi(r)} \simeq 1 - \beta q\phi. \tag{14.25}$$

In that case, the Poisson–Boltzmann equation becomes *linearised* to

$$\nabla^2 \phi = \frac{n_\infty q^2}{\epsilon}\beta\phi - \frac{q}{\epsilon}\,\delta(\underline{r}) \tag{14.26}$$

or

$$\boxed{\nabla^2 \phi - \frac{\phi}{\lambda_{\text{D}}^2} = -\frac{q}{\epsilon}\,\delta(\underline{r})} \tag{14.27}$$

which is called the **Debye–Hückel equation**; the constant

$$\boxed{\lambda_{\text{D}} = \left(\frac{kT\epsilon}{q^2 n_\infty}\right)^{1/2}} \tag{14.28}$$

has dimensions of length and is called the **Debye screening length**.

In the Debye–Hückel equation $\phi(r)$ is the Green's function for a charge q at the origin. Formally, this can be derived by Fourier transform techniques, but a direct approach is straightforward in this case. For spherical symmetry,

$$\nabla^2 \phi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial \phi}{\partial r}\right) = \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\phi), \tag{14.29}$$

so that a solution of the Debye–Hückel equation away from $r = 0$ is that $r\phi$ is in the form of a sum of exponentials. Choosing the decaying solution, we have

$$\boxed{\phi(r) = A\frac{e^{-r/\lambda_{\text{D}}}}{r} = \frac{q}{4\pi\epsilon}\frac{e^{-r/\lambda_{\text{D}}}}{r}.} \tag{14.30}$$

Happily, this vacuum solution tends to a $1/r$ potential at small r, which will provide a delta-function on differentiation. This is analogous to the solution of the ordinary Poisson equation, where the same reasoning yields $\phi = A/r$. Obviously, we need the same $A = q/4\pi\epsilon$ in this case,

since the $\phi/\lambda_\mathrm{D}^2$ term is negligible at $r = 0$ compared to the δ-function. We derive this constant by integration of Poisson's equation over a small sphere about the origin. The RHS integrates to $-q/\epsilon$ (the volume of the δ-function is unity), and the LHS can be rewritten using the divergence theorem:

$$\int \nabla^2 \phi \, dV = \int \underline{\nabla}\phi \cdot d\underline{S}, \tag{14.31}$$

where the latter integral is $-4\pi A$.

The solution (14.30) is called the **screened Coulomb potential**, and is the net potential around a point charge, once the other charges have arranged themselves so as to minimise their free energy. At distances large compared to λ_D, the charge at the centre is cancelled by the collective response of the other charges, leaving an effective potential that is well localised – see Fig. 15. The screened coulomb potential is sometimes called the **Yukawa potential**, after Yukawa who first proposed this form for the strong force mediated by exchange of (massive) pions. The result is the same as you would predict if the photon (which mediates the Coulomb force) were a massive particle. Accordingly it is sometimes said that, because of the many-body interactions between charges, photons in a plasma 'acquire mass'.

Debye–Hückel theory is an example of a **mean-field** theory. What this term generally means is that correlations are ignored at some level: here we have ignored correlations except two-point correlations. We shall explore the meaning of mean-field theories more fully in later sections. We also note that Debye–Hückel theory is a self-consistent approximation, i.e. the density $n(r)$ depends on the potential $\phi(r)$, which is itself determined from the density.

15 Phase transitions and the Ising model

The Debye–Hückel theory shows how the properties of a system can be affected by correlated behaviour induced by interactions between molecules. A more familiar and more radical example of this sort is the case of phase transitions, in which the properties of matter can change **discontinuously** when physical conditions are altered. This is undoubtedly one of the most fascinating parts of physics, where we encounter phenomena that are **emergent**: effects that are not obviously to be expected from a first inspection of the microscopic laws that are at work. For this reason, much effort has been invested in simple models where the connection to phase transitions can be understood, and the remaining lectures focus on one particular case: the Ising model.

15.1 Review of basic phenomenology

15.1.1 Critical point of liquid–gas system

- Along the vapour-pressure (boiling) co-existence curve in Fig. 16a, the gas and liquid coexist i.e. the fluid can exist in two different forms or phases characterised by different densities.

- This co-existence curve in Fig. 16a terminates at **the critical point** which has unique thermodynamic co-ordinates T_c, P_c, ρ_c where ρ is the density.

- In Fig. 16b the critical isotherm has zero slope at ρ_c, which means that the isothermal compressibility given by

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T, \tag{15.1}$$

diverges at the critical point. This implies that there are large scale fluctuations in the volume/density.

- In the coexistence region in Fig. 16b (or line in Fig. 16a) the liquid and gas coexist and the fluid separates into a mixture of gas and liquid, which have densities ρ_l and ρ_g, with the required overall density ρ.

15.1.2 Critical point of a magnet

- A magnetic solid, made up of atoms with dipole moments, exhibits no global magnetisation at high T (in zero applied field). This is known as the **paramagnetic** phase.

- But in this phase an applied field will produce a mean magnetisation. This scales linearly with field for small fields but saturates at large field (Fig. 17).

- Interactions between dipoles, namely the quantum exchange interactions that tend to align the dipoles, become important at low T.

- For $T < T_c$ a global magnetisation emerges even in zero applied field i.e. the dipoles tend to line up in the same direction without the aid of an applied magnetic field. This is known as a **ferromagnetic** phase.

Furthermore

- For $T < T_c$ note the discontinuity in M as we cross the coexistence line i.e.

$$\begin{cases} H = 0^+ & M > 0 \\ H = 0^- & M < 0 \end{cases} \tag{15.2}$$

- At T_c we have $M_c = H_c = 0$ which is due to the symmetry between the two ferromagnetic phases $\pm |M|$.

- The critical isotherm $H(M)$ has zero slope at $M = 0$, which implies that at T_c the response function diverges:

$$\chi = \left. \frac{\partial M}{\partial H} \right|_{H=0} \to \infty, \tag{15.3}$$

and that there are large scale fluctuations in the magnetisation.

15.1.3 Critical point of a binary alloy

Finally a brief mention is made here of a less familiar system exhibiting a phase transition.

- There are equal concentrations of A and B type atoms arranged on a regular lattice.

- For $T > T_c$ we have the disordered phase where the atoms are arranged randomly on the lattice sites.

- For $T < T_c$ an ordered state emerges where A and B atoms are concentrated on their own separate sublattices.

15.1.4   Common features

There are countless other examples of phase transitions, but our aim here is to unify the common features:

- **Co-existence Curve**: line separating two phases differing by macroscopic properties.

- **Critical Point**: end of co-existence curve.

- **Order Parameter** O: characterises the difference between the two phases.

    For example

    - Fluid: $O \equiv \rho_l - \rho_g$ the density difference between the two phases.
    - Magnet: $O \equiv M$ the global magnetisation.

- **Phase Transition**: qualitative change in macroscopic properties as some parameter e.g. T is varied.

    Generally we have two types of phase transition:

    - a *discontinuous* transition (often referred to as *'first order'*) exhibits a jump in O e.g. on crossing the co-existence curve
    - a *continuous* transition (or *'second order'*, ...) has the change in $O \to$ zero (but in a nonanalytic way – see later) e.g. in the passage through the critical point. The transition is accompanied by divergence of response functions and accompanying large-scale fluctuations.

15.2   Basic model: the Ising model

We now introduce the most widely studied model system in statistical physics, written down by Ernst Ising in his 1925 PhD thesis.

- We have N spins on a lattice, which we can take for simplicity to be a simple cubic lattice. The spins each occupy one lattice site i where $i = 1 \ldots N$ and take values $S_i = \pm 1$ 'up' or 'down'.

- The **Configurational Energy**, or Hamiltonian of the system, is given by

$$E(\{S_i\}) = -h \sum_i S_i - J \sum_{<ij>} S_i S_j \tag{15.4}$$

    - First term: h is the 'field' i.e. the applied magnetic field. If S_i is aligned to the field it gives a lower energy contribution.
    - Second term: here, $<>$ means nearest neighbour (nn) pairs. The number of nn of a site is z, the **co-ordination number** of the lattice; for example, on a simple cubic lattice $z = 6$. But, as will frequently be the case, we need to make sure that the energy from each pair is only counted once, independent of the order of the sites that constitute the pair: thus the total number of independent nn pairs is $Nz/2$, not Nz.
    - J is the **coupling constant**, here assuming **isotropy**, so that that all adjacent spins couple with equal strength. If $J > 0$ when the neighbouring spins S_i and S_j are aligned, they give a lower contribution to the energy.

– The sign may seem puzzling here, given that we are used to classical magnetic dipoles having a lower energy when they are *antiparallel*, and this sign difference is key to the phenomenon of ferromagnetism. Its origin is via quantum symmetry. Consider two Fermions, whose wave function must be antisymmetric overall. If both are spin up, the spin wave function must be symmetric, so that the spatial wave function is antisymmetric: the particles avoid each other and there is less Coulomb repulsion.

- The Partition Function

$$Z = \sum_{\{S_i=\pm1\}} e^{-\beta E(\{S_i\})} \qquad \beta = 1/kT \,. \tag{15.5}$$

The **configurational sum** (often referred to as the **trace**) is explicitly

$$\sum_{\{S_i=\pm1\}} = \sum_{S_1=\pm1} \sum_{S_2=\pm1} \cdots \sum_{S_N=\pm1} \,. \tag{15.6}$$

Why might we expect a phase transition in this model?

$$F(E) = E - TS \,, \tag{15.7}$$

is the Helmholtz free energy. The equilibrium state is given by minimising F with respect to the S_i, which in turn determine E and S. As usual, T sets the balance between minimising E and maximising S:

At low T minimise E $\Rightarrow$ ground states $\uparrow\uparrow \ldots \uparrow\uparrow$ and $\downarrow\downarrow \ldots \downarrow\downarrow$ dominate

At high T maximise S $\Rightarrow$ disordered states $\uparrow\downarrow\downarrow \ldots \uparrow\downarrow\uparrow$ dominate

But we need to show that there is a phase transition between the two regimes rather than a smooth crossover.

16 Mean-Field theory of the Ising model

Even though it is an extremely simple idealisation, it is a sad fact that there is no exact solution for the Ising model (or many other interesting models) on a 3D lattice. Therefore one has to resort to approximations. This section will go through in detail a mean-field approximation, which is always the first recourse in trying to construct a theory.

16.1 The mean-field approximation

16.1.1 The Weiss approximation

Recall that the Ising configurational energy is

$$E(\{S_i\}) = -h \sum_i S_i - J \sum_{<ij>} S_i S_j \,. \tag{16.1}$$

Consider all contributions involving spin S_j

$$\epsilon(S_j) = -h S_j - J S_j \sum_k^{nn} S_k \,, \tag{16.2}$$

where the sum is over nearest neighbours (nn) k of site j. We now *approximate* the sum by replacing the S_k with their mean value

$$\epsilon_{\mathrm{mf}}(S_j) = -hS_j - JS_j \sum_{k}^{nn} \langle S_k \rangle \equiv -h_{\mathrm{mf}} S_j , \tag{16.3}$$

where z is the co-ordination number, the number of nn sites,

$$h_{\mathrm{mf}} = h + Jzm \tag{16.4}$$

and m, the magnetisation per spin, is just the mean value of any given spin:

$$m = \frac{1}{N} \sum_{i} \langle S_i \rangle = \langle S_k \rangle \ \text{ for all } k . \tag{16.5}$$

Thus the mean-field approximation is to replace the configurational energy by the energy of a non-interacting system of spins each experiencing a field h_{mf}. Note that here we are calculating the energy of a single spin, so we count the interactions with the full number of nearest neighbours, z, and do not divide by 2. For this problem we can immediately write down the single-spin Boltzmann distribution:

$$p(S_j) = \frac{e^{-\beta \epsilon_{\mathrm{mf}}(S_j)}}{\sum_{S_j=\pm 1} e^{-\beta \epsilon_{\mathrm{mf}}(S_j)}} = \frac{e^{\beta h_{\mathrm{mf}} S_j}}{e^{\beta h_{\mathrm{mf}}} + e^{-\beta h_{\mathrm{mf}}}} . \tag{16.6}$$

Now, we have a *consistency* condition to fulfil: the value of the magnetisation m predicted by the Boltzmann distribution should be equal to the value of m used in the expression for h_{mf}. Thus we require

$$m = \sum_{S_j=\pm 1} p(S_j) S_j \tag{16.7}$$

$$= \frac{e^{\beta h_{\mathrm{mf}}} - e^{-\beta h_{\mathrm{mf}}}}{e^{\beta h_{\mathrm{mf}}} + e^{-\beta h_{\mathrm{mf}}}} = \tanh(\beta h_{\mathrm{mf}}) , \tag{16.8}$$

and we arrive at the mean-field equation for the magnetisation:

$$\boxed{m = \tanh(\beta h + \beta Jzm)} \tag{16.9}$$

Note that, although not necessary here, an alternative approach is to use the usual type of trick based on the partition function:

$$m = \frac{1}{\beta} \frac{\partial}{\partial h_{\mathrm{mf}}} \ln \left(\sum_{S_j=\pm 1} e^{\beta h_{\mathrm{mf}} S_j} \right) = \frac{1}{\beta} \frac{\partial}{\partial h_{\mathrm{mf}}} \ln \left(2 \cosh(\beta h_{\mathrm{mf}}) \right) . \tag{16.10}$$

16.1.2 Solving the consistency condition

First we will consider the case $h = 0$ (zero applied field). The solutions of

$$m = \tanh(\beta Jzm) , \tag{16.11}$$

can be understood graphically, as shown in Fig. 18. We see that for low β (high T) the only solution is $m = 0$ whereas for high β (low T) there are three possible solutions $m = 0$ and

$m = \pm|m|$. The solutions with $|m| > 0$ appear when the slope of the tanh function at the origin is greater than 1 (i.e. the gradient of the straight line m)

$$\frac{d}{dm}\tanh(\beta Jzm)\Big|_{m=0} > 1.\tag{16.12}$$

Using the expansion of tanh for small argument $\tanh x \simeq x - x^3/3$ (we only need the first term at this point), the mean-field equation becomes

$$m = \beta Jzm + O(m^3).\tag{16.13}$$

Thus a nonzero solution emerges when the gradient of the RHS reaches unity; the critical temperature at which this happens is given by

$$\boxed{\beta_c Jz = 1 \quad \text{or} \quad T_c = \frac{zJ}{k}}.\tag{16.14}$$

Thus for $T > T_c$ only the paramagnetic $m = 0$ solution is available, whereas for $T < T_c$ we also have the ferromagnetic solutions $\pm|m|$. These are the physical solutions for $T < T_c$, as we shall see in the next section.

16.2   Critical behaviour

Consider again equation (16.11), which in terms of T_c is

$$m = \tanh\left(m\frac{T_c}{T}\right).\tag{16.15}$$

We wish to analyse the emergence of the ferromagnetic solutions when T is near T_c, that is, in the *critical regime* where $T \simeq T_c$ and $|m| \ll 1$. Now expanding $\tanh(x)$ to $O(x^3)$ we obtain

$$m = m\frac{T_c}{T} - \frac{m^3}{3}\left(\frac{T_c}{T}\right)^3.\tag{16.16}$$

Thus $m = 0$ or

$$m^2 = 3\left(\frac{T}{T_c}\right)^3\left(\frac{T_c}{T} - 1\right).\tag{16.17}$$

We now define the *reduced temperature t* by

$$\boxed{t = \frac{T - T_c}{T_c}}\tag{16.18}$$

which gives

$$\frac{T}{T_c} = 1 + t \qquad \frac{T_c}{T} = \frac{1}{1+t}.\tag{16.19}$$

The reduced temperature t measures the proximity to the critical point. When t is small (16.17) becomes

$$m^2 = 3(1+t)^3\left(\frac{1}{1+t} - 1\right) = -3t + O(t^2).\tag{16.20}$$

Thus

$$\boxed{\begin{array}{ll} T > T_c & m = 0 \\ T < T_c & m \simeq \pm(3|t|)^{1/2} + O(|t|^{3/2}). \end{array}}\tag{16.21}$$

86

We now proceed to compute the susceptibility

$$\chi = \frac{\partial m}{\partial h}\bigg|_{h=0} , \tag{16.22}$$

and we should now expand (16.9). To capture the leading order behaviour in t, h it suffices to expand to first order in h:

$$m = m\frac{T_c}{T} + \beta h - \frac{m^3}{3}\left(\frac{T_c}{T}\right)^3 . \tag{16.23}$$

There is also a term $\propto hm^2$ but since near the critical point $m = 0$ or $m \propto t^{1/2}$, this term will not contribute or will give a correction higher order in t. You should check this once you've understood how to obtain (16.25). Taking the derivative wrt h yields

$$\chi = \chi\frac{T_c}{T} + \beta - \chi m^2 \left(\frac{T_c}{T}\right)^3 , \tag{16.24}$$

or

$$\chi = \frac{\beta}{\left(1 - \frac{T_c}{T} + m^2 \left(\frac{T_c}{T}\right)^3\right)} . \tag{16.25}$$

Then we find for t small and using the appropriate expression for m ($m = 0$ or (16.17))

$$T > T_c : \qquad \chi = \frac{\beta}{1 - \frac{T_c}{T}} \simeq \frac{\beta_c}{t} \tag{16.26}$$

$$T < T_c : \qquad \chi = \frac{\beta}{1 - \frac{T_c}{T} + 3(\frac{T_c}{T} - 1)} \simeq \frac{\beta_c}{2|t|} . \tag{16.27}$$

Again we have just identified the leading behaviour as $t \to 0$; if we had taken the $m = 0$ solution below T_c in (16.25) we would have obtained a negative response function which is unphysical. The critical behaviour is sketched in Fig. 19. Note how $|m| > 0$ emerges in a non-analytic way since $\partial|m|/\partial t$ diverges at $t = 0$. Also note the divergence in χ (the response function to the applied field) at $t = 0$.

16.3   Limitations of mean-field theory

The essence of the mean-field assumption is the neglect of *correlations* between spins i.e. we effectively replace

$$\langle S_i S_j \rangle \simeq \langle S_i \rangle \langle S_j \rangle \qquad i \neq j . \tag{16.28}$$

Note we can write the energy in the form

$$E(\{S_i\}) = E_0 - h \sum_j S_j , \tag{16.29}$$

where $E_0 = -J \sum_{\langle ij \rangle} S_i S_j$, which is the same form as that which we used to discuss the fluctuation response theorem in section 5 if we identify $f = h$ and $A = \sum_j S_j$. Then we know that

$$\chi_{AA} = \frac{\partial \langle A \rangle}{\partial f} = \beta \left[\langle A^2 \rangle - \langle A \rangle^2 \right] . \tag{16.30}$$

In our case (χ here is susceptibility per site)

$$\chi = \frac{\partial m}{\partial f} = \frac{1}{N}\chi_{AA} = \frac{\beta}{N} \sum_{jk} [\langle S_j S_k \rangle - \langle S_j \rangle \langle S_k \rangle] . \tag{16.31}$$

This equation is exact. But if we now insert the mean field approximation then all terms in the sum with $j \neq k$ will vanish, leaving

$$\chi = \beta[1 - m^2]. \tag{16.32}$$

Clearly this does not diverge at T_c so the mean-field approximation is inconsistent with regard to χ. The root of the problem lies in the neglect of correlations, which become important as the critical point is approached.

16.4 The correlation length

Let us define the 'connected correlation function'

$$\begin{align} G_{ij} &= \langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle \tag{16.33} \\ &= \langle (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle) \rangle. \tag{16.34} \end{align}$$

Clearly G_{ij} measures the correlations between the *fluctuations* $\Delta S_i = S_i - \langle S_i \rangle$: a positive G_{ij} implies that the fluctuations of the spins about their mean value are correlated. Let R_{ij} be the distance between spins i and j. For large R_{ij} we expect

$$G_{ij} \simeq C(R_{ij}) e^{-R_{ij}/\xi}, \tag{16.35}$$

where C is some slowly varying function but $e^{-R_{ij}/\xi}$ 'cuts off' the correlation function at the correlation length ξ. Thus ξ is the scale at which correlations decrease significantly.

When $T > T_c$ so that $\langle S_i \rangle = 0$ a typical microstate will consist of clusters of up and down spins with the overall magnetisation being zero. Intuitively the correlation length gives a measure of the (linear) size of the largest clusters of correlated spins. See Fig. 20 for a one-dimensional illustration. As the temperature is decreased to the critical temperature the size of the clusters diverge and we expect ξ to diverge.

One can refine mean-field theory to include the calculation of the 'two-point' correlation function G_{ij}. This is similar in spirit to the Debye–Hückel theory but is rather more technical. The result is that a correlation length is predicted which grows like

$$\xi \propto |t|^{-1/2}, \tag{16.36}$$

diverging at T_c. However the mean-field theory is still inconsistent due to the neglect of three-point and higher order correlations. Correlations and fluctuations on all scales and all orders become important as criticality is approached.

16.5 Summary of mean-field picture and comparison with experiment

The mean-field theory predicts

- The critical temperature $T_c = zJ/k$ at $h = 0$

- The singular, critical behaviour is described by

$$\begin{align} m &\sim m_- |t|^{\tilde{\beta}} \quad (T < T_c) \tag{16.37} \\ \chi &\sim \chi_\pm |t|^{-\gamma} \tag{16.38} \\ \xi &\sim \xi_\pm |t|^{-\nu} \tag{16.39} \end{align}$$

$\tilde{\beta}, \gamma, \nu$ are known as **critical exponents** and take mean-field values $\tilde{\beta} = 1/2, \gamma = 1, \nu = 1/2$.

(The standard notation for the order parameter (magnetisation) exponent is β and there is an obvious clash with inverse temperature so we use a tilde to be clear.)

The precise meaning of, for example, $m \propto |t|^{\tilde{\beta}}$ is

$$\lim_{|t| \to 0} \frac{\ln m}{\ln |t|} = \tilde{\beta}. \tag{16.40}$$

Actually there are even more critical exponents. For example,

- one can define the zero-field heat capacity

$$C_h = \left. \frac{\partial \overline{E}}{\partial T} \right|_{h=0} \propto |t|^{-\alpha}, \tag{16.41}$$

- near and below T_c one can characterise the discontinuity in the m order parameter across the coexistence line by

$$h \propto |m|^\delta \operatorname{sgn}(m). \tag{16.42}$$

Mean-field values are $\alpha = 0$ and $\delta = 3$ (see tutorial sheet).

The set α, $\tilde{\beta}$, γ, δ, ν characterise the critical point. It turns that some of these exponents are implied by the others and in fact there are only three independent exponents which we can take as $\tilde{\beta}$, γ, ν.

Now experimental data reveal that

- Systems do exhibit such singularities

- But the critical exponents *differ* from the mean-field values

- However critical exponents are system independent e.g. for fluids, binary alloys and many magnets it has been found that $\tilde{\beta} \simeq 0.31$, $\gamma \simeq 1.25$, $\nu \simeq 0.64$ in three dimensions. Thus apparently unrelated systems share the same set of critical exponents. This is referred to as **Universality** and remained a mystery for many years.

To summarise, mean-field theory is successful in that it qualitatively describes the critical behaviour but is quantitatively incorrect. Moreover as we shall find next section it is qualitatively incorrect in one dimension. On the other hand as we shall discuss later mean field theory does in fact give the correct critical exponents in high enough space dimension ($d \geq 4$ for the Ising model).

17   Exact results for the Ising model

In one dimension the total Ising energy becomes

$$E = -h \sum_{i=1}^{N} S_i - J \sum_{i=1}^{N} S_i S_{i+1}. \tag{17.1}$$

Note that we have assumed periodic boundary conditions which mean we take $S_{N+1} = S_1$ i.e. the spins are on a ring. Although at first this seems completely bizarre, it is a standard device to make things simple and is harmless – one can consider other boundary conditions but it won't alter the physics. Note that we only consider pairs connecting i to $i + 1$, which deals with the overcounting problem.

17.1 Solution in 1D for $h = 0$

$$\uparrow\uparrow \cdots \uparrow\uparrow\downarrow\downarrow \cdots \downarrow\downarrow\uparrow\uparrow \cdots \uparrow\uparrow \tag{17.2}$$

Example of two domain walls in a one-dimensional system of N Ising spins.

As seen in the last section, a microstate consists of clusters or **domains** of spins of the same sign separated by **domain walls**. In the absence of an external field there are two possible ground states that contain no domain walls (the all up and all down configurations of spins). The number of domain walls in a microstate specifies the energy relative to the ground state i.e. each domain wall costs energy $2J$. We can thus view the properties of the system more fruitfully in terms of the statistics of domain walls. Let

$$n_i = \frac{1}{2}(1 - S_i S_{i+1}) = \begin{cases} 1 & \text{if domain wall present} \\ 0 & \text{if no domain wall present} \end{cases} \tag{17.3}$$

Then

$$S_i S_{i+1} = 1 - 2n_i \tag{17.4}$$

and

$$E = -NJ + 2J \sum_{i=1}^{N} n_i \, . \tag{17.5}$$

The first term is the energy in the case that all spins are aligned (no walls); this is just a constant which can be ignored. But each wall changes the interaction energy from $-J$ to $+J$ at that site, and the system thus becomes a *non-interacting* set of domain walls with states $n_i = 1$ (domain wall) with energy $\epsilon = 2J$ and $n_i = 0$ (no domain wall) with zero energy.

As the system is non-interacting we can immediately write down the Boltzmann distribution for the *single* domain wall problem, which is given by

$$p(n_i) = \frac{e^{-2\beta J n_i}}{1 + e^{-2\beta J}} \, . \tag{17.6}$$

The probability $p(1) = p$ that a domain wall is present then has the behaviour in the high and low temperature limits

$$p \simeq \begin{cases} e^{-2\beta J} + \dots & \to 0 \text{ as } \beta \to \infty \ (T \to 0) \\ \frac{1}{2} - \frac{1}{2}\beta J + \dots & \to \frac{1}{2} \text{ as } \beta \to 0 \ (T \to \infty) \end{cases} \tag{17.7}$$

So we get the expected limits of no domain walls at $T = 0$ and a disordered state where spins are randomly up or down as $T \to \infty$. However already we can see that $p(1)$ goes smoothly between these two extremes and there is no phase transition. Thus for $T > 0$ we are always in the paramagnetic phase.

To understand better what is happening physically we should consider the two-point correlation function and examine the correlation length. Let j and k be two sites separated by distance $\ell = |j - k|$ and let m be the number of domain walls between the two sites. Then

$$S_j S_k = \begin{cases} +1 & \text{if } m \text{ even} \\ -1 & \text{if } m \text{ odd} \end{cases} = (-1)^m \tag{17.8}$$

and we can average this to obtain

$$\langle S_j S_k \rangle = \sum_m p_m (-1)^m \,, \tag{17.9}$$

where p_m is the probability that there are precisely m domain walls between the two sites. p_m is given by the binomial distribution for having m domain walls in the ℓ possible locations between the two sites with a domain wall present with probability p so

$$p_m = \binom{\ell}{m} p^m (1-p)^{\ell-m} \,, \qquad \sum_{m=0}^{\ell} p_m = (p + 1 - p)^{\ell} = 1 \,. \tag{17.10}$$

Thus

$$\langle S_j S_k \rangle = \sum_{m=0}^{\ell} \binom{\ell}{m} p^m (1-p)^{\ell-m} (-1)^m = (-p + 1 - p)^{\ell} = (1 - 2p)^{\ell} \,. \tag{17.11}$$

We can write this as

$$\boxed{\langle S_j S_k \rangle = e^{\ell \ln(1-2p)} = e^{-\ell/\xi} \,,} \tag{17.12}$$

where the **correlation length** is

$$\xi = \frac{1}{|\ln(1-2p)|} = \begin{cases} \frac{1}{2} e^{2\beta J} & \to \infty \text{ as } \beta \to \infty \ (T \to 0) \\ \frac{1}{|\ln(\beta J)|} & \to 0 \text{ as } \beta \to 0 \ (T \to \infty) \end{cases} \tag{17.13}$$

So the correlation length gets longer as T decreases but only diverges as $T \to 0$. From this we conclude that

$$\lim_{|j-k| \to \infty} \langle S_j S_k \rangle = 0 \quad \forall T > 0 \,, \tag{17.14}$$

which implies that we are always in the paramagnetic phase. For a ferromagnetic phase we would obtain a non-zero limit and there would be **long-range order**.

Thus we have shown that there is no long-range order (ferromagnetic phase) and no phase transition in the 1D Ising model. This directly contradicts our mean-field prediction of a phase transition (for $z = 2$) at $T_c = 2J/k$. Mean-field theory is disastrously wrong in one dimension.

17.1.1 Absence of long-range order in 1D

Let us consider two phases of a one-dimensional system separated by a domain wall.

$$AA \cdots AABB \cdots BB \tag{17.15}$$

Domain wall between two phases A and B

and assume that the two phases have equal free energies. In the Ising model the two phases are the ferromagnetic up and down phases. We wish to determine whether a domain wall is favoured thermodynamically. The energy cost ΔE of the domain wall will be *finite* if the interactions are short-ranged. For example the Ising model we have studied has nearest neighbour interactions and the energy cost is $\Delta E = 2J$. We could also consider an Ising model with next nearest neighbour interactions etc. The important point is that the energy cost of a domain wall does *not* depend on the system size N.

The entropy gain due to the creation of a domain wall can be computed from the Boltzmann entropy

$$\Delta S = k \ln \Omega, \tag{17.16}$$

where here $\Omega = N$ is the number of possible positions for the domain wall. Thus the free energy difference of a state of two domains divided by a domain wall over the ordered state of one phase is

$$\Delta F = \Delta E - T\Delta S = \Delta E - kT \ln N, \tag{17.17}$$

which for all $T > 0$ will be negative for sufficiently large N. Thus entropy wins, domain walls are always created and long range order cannot be maintained in one dimension.

17.2   General solution of 1D Ising model

For the general case ($h \neq 0$) we wish to calculate the partition function

$$Z = \sum_{\{S_i=\pm1\}} e^{-\beta E(\{S_i\})} = \sum_{S_1=\pm1} \sum_{S_2=\pm1} \cdots \sum_{S_N=\pm1} e^{-\beta E(\{S_i\})}. \tag{17.18}$$

Now let us write the energy in the following way

$$E(\{S_i\}) = -\frac{h}{2}\sum_i S_i - \frac{h}{2}\sum_i S_{i+1} - J\sum_i S_i S_{i+1}. \tag{17.19}$$

Then

$$\exp(-\beta E) = \prod_i \exp\left[\frac{\beta h}{2}(S_i + S_{i+1}) + \beta J S_i S_{i+1}\right], \tag{17.20}$$

and we have

$$Z = \sum_{\{S_i=\pm1\}} \prod_i T(S_i, S_{i+1}), \tag{17.21}$$

where

$$T(S_i, S_{i+1}) = \exp\left[\frac{\beta h}{2}(S_i + S_{i+1}) + \beta J S_i S_{i+1}\right]. \tag{17.22}$$

Write the values $T(S_i, S_{i+1})$ as a 2×2 symmetric matrix – the **Transfer Matrix**

$$T = \begin{pmatrix} e^{\beta(J+h)} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta(J-h)} \end{pmatrix}, \tag{17.23}$$

where the first (second) row corresponds to $S_i = +1$ (-1) and the first (second) column corresponds to $S_{i+1} = +1$ (-1).

Now writing out Z, and recalling we are using periodic boundary conditions, we have

$$\begin{aligned} Z &= \sum_{\{S_i=\pm1\}} T(S_1, S_2)T(S_2, S_3)\dots T(S_{N-1}, S_N)T(S_N, S_1) \tag{17.24} \\ &= \text{Tr}\left[T^N\right] = \lambda_+^N + \lambda_-^N. \tag{17.25} \end{aligned}$$

In going from the first to second line we have used the usual rules of matrix multiplication and the definition of the trace as the sum of the diagonal elements; in going from the second to third

line we have used the usual properties of the eigenvalues and trace of a symmetric matrix (see tutorial). Direct calculation (see tutorial sheet) gives $\lambda_\pm$, the eigenvalues of T, as

$$\lambda_\pm = e^{\beta J} \cosh \beta h \pm \sqrt{e^{2\beta J} \sinh^2 \beta h + e^{-2\beta J}} \, . \tag{17.26}$$

Then the free energy per spin is

$$f = -\frac{kT}{N} \ln Z \to -kT \ln \lambda_+ \tag{17.27}$$

for large N (as $\lambda_+ - \lambda_- > 0$, so $\lambda_-/\lambda_+ < 1$). The thermodynamic properties can be obtained by taking the various derivatives of f with respect to h, β.

17.3 Existence of a phase transition in 2D

Now consider the Ising model on a square lattice. A domain wall between the up and down phases becomes an extended object: a chain of links as shown in Fig. 21. Let the length of the chain be $\tilde{N}$: then the energy cost of the whole chain is

$$\Delta E = 2J\tilde{N} \, , \tag{17.28}$$

since we have $\tilde{N}$ nearest neighbour pairs of spins of opposite signs.

To evaluate the entropy gain due to a single domain wall forming in the system, we have to estimate Ω, the number of possible paths for the domain wall. If we start at the left hand side and the size of the lattice is $N = L \times L$ then there are L starting positions. At each step the domain wall can move to the right, move up or move down. This suggests that the number of domain walls is approximately

$$\Omega \sim L \times 3^{\tilde{N}} \, . \tag{17.29}$$

This is a crude estimate (in fact it is an upper bound), but it is adequate because we take the logarithm:

$$\Delta S = k\tilde{N} \ln 3 + k \ln L \simeq k\tilde{N} \ln 3 \, . \tag{17.30}$$

Then

$$\Delta F \simeq \tilde{N} \left[2J - kT \ln 3\right] \, . \tag{17.31}$$

Clearly for sufficiently low T, namely

$$kT < \frac{2J}{\ln 3} \, , \tag{17.32}$$

$\Delta F > 0$ and a domain wall is not thermodynamically favoured. Thus the ordered phase is stable for low enough T (but $T > 0$) and long range order is maintained.

However we know that at large T the domain wall must be favoured and we should have a disordered phase. This will occur at a critical temperature T_c. The above domain wall argument can be sharpened (see chapter 14 of Huang) so that it actually gives a rigorous lower bound

$$kT_c > \frac{2J}{\ln 3} \, , \tag{17.33}$$

which compares favourably to the exact value determined from the exact solution of the 2D Ising model.

17.4 The exact Ising model in 2D and 3D

The Ising model in 2D was solved exactly by Lars Onsager in 1944. This is a heroic calculation, which is alas far too long and detailed for this course: chapter 15 of Huang takes 24 pages over the derivation. Onsager's result for T_c was

$$kT_c = \frac{2J}{\ln(\sqrt{2}+1)} . \tag{17.34}$$

What the crude argument misses is the existence of small 'bubbles' of the minority phase inside the majority phase at finite (but low) T. The magnetisation of the system for zero applied field is zero for $T > T_c$, but for lower T it takes the wonderful form

$$m = \left[1 - (\sinh 2\beta J)^{-4}\right]^{1/8} . \tag{17.35}$$

The expression for the specific heat is complicated, but near T_c it is dominated by a logarithmic singularity. If $T = T_c + \Delta T$, then

$$\frac{C_h}{k} \simeq \frac{2}{\pi} \left(\frac{2J}{kT_c}\right)^2 \ln \left|\frac{T_c}{\Delta T}\right| . \tag{17.36}$$

This singularity is sufficiently weak that the internal energy is continuous at T_c, so there is no latent heat and the transition is second order.

There has been a long history of searches for a corresponding analytic solution of the 3D Ising model. So far these efforts have been without success – although tantalisingly there is no proof that such a solution is impossible. But at present, the properties of the 3D Ising model have only be derived by numerical Monte-Carlo studies. Qualitatively, the behaviour is similar to that in 2D, with spontaneous magnetisation below

$$kT_c^{3D}/J \simeq 4.5, \tag{17.37}$$

as opposed to 2.3 in 2D.

17.5 Mapping to a lattice gas

The Ising model is isomorphic to a variety of other physical models of interest, including phenomena as broad as neural networks or even economics. In the lattice gas model each site of a lattice is either occupied by a particle or is empty. Thus each lattice site has associated with it a variable, c_i with occupation number

$$c_i = \begin{cases} 1 \\ 0 \end{cases} . \tag{17.38}$$

The overall number concentration of the gas is

$$c = \text{fraction of particles with occupation number unity} = \frac{\sum_i c_i}{N} . \tag{17.39}$$

A hard-sphere repulsion between lattice gas particles is wired in due to the maximum occupancy $c_i = 1$. A short-range attractive potential is introduced by an energy $-\epsilon < 0$ associated with a pair of neighbouring particles.

$$E = -\epsilon \sum_{<ij>} c_i c_j . \tag{17.40}$$

Now the canonical partition function for the lattice gas should respect the fact that only microstates with precisely Nc particles are allowed

$$Z = \sum_{\{c_i=1,0\}} e^{-\beta E} \, \delta \left(\sum_i c_i - Nc \right) , \tag{17.41}$$

where the δ–function restricts the sum to the allowed microstates. However it is easier to work with the Grand Canonical Ensemble where we allow the particle number to fluctuate but introduce a chemical potential μ to tune the average number of particles:

$$Z_{\text{LG}} = \sum_{\{c_i=1,0\}} e^{-\beta(E-\mu\sum_i c_i)} = \sum_{\{c_i=1,0\}} \exp\left[\beta\epsilon \sum_{<ij>} c_i c_j + \beta\mu \sum_i c_i \right] , \tag{17.42}$$

where the sum over $c_i = 1, 0$ is now unrestricted.

The effective energy $E - \mu\sum_i c_i$ can be mapped onto the Ising energy (plus a constant) by the identification

$$S_i = 2c_i - 1 = \pm 1 , \qquad J = \frac{1}{4}\epsilon , \qquad h = \frac{1}{4}(\epsilon z + 2\mu) . \tag{17.43}$$

(z is the co-ordination number; the number of nn's of a site in a 3D cubic lattice is 6). Therefore

$$Z_{\text{LG}} = \text{const.} \times Z_{\text{Ising}} , \tag{17.44}$$

and on taking logs of this equation the grand potential of the lattice gas and free energy of the Ising model are the same (up to an unimportant constant)

$$\Phi_{\text{LG}}(T, \mu) = F_{\text{Ising}}(T, h) + \text{const.} . \tag{17.45}$$

Now consider the phase diagram for magnetic system in the h-T plane (Fig. 22). The coexistence curve of the lattice gas is now given by

$$\mu_c = -\tfrac{1}{2}\epsilon z , \tag{17.46}$$

(which corresponds to $h = 0$) and the critical point will be given by the Ising value $T_c(J) = T_c(\epsilon/4)$. Thus for $T < T_c$ we have a discontinuous transition as μ is decreased through μ_c from a 'liquid' phase to a 'gas' phase:

$$c_{\text{liq}} = \tfrac{1}{2}(1 + |m|) \quad \rightarrow \quad c_{\text{gas}} = \tfrac{1}{2}(1 - |m|) , \tag{17.47}$$

where $\pm|m|$ is the magnetisation in the ferromagnetic phase of the Ising model.

18 Landau theory of phase transitions

In the last section we saw that the ferromagnetic transition and the liquid-gas transition are related in the sense that the Ising model can describe them both. Here a deeper understanding will be developed with a model-independent theory of why the critical points of different systems share the same properties i.e. why we have **Universality**.

18.1 Mean-field theory of Ising model revisited

First we repeat the mean-field theory of the Ising model by constructing an approximate free energy. In the mean-field approximation the mean total Ising energy becomes

$$\overline{E} = -h\sum_{i=1}^{N}\langle S_i\rangle - J\sum_{\langle ij\rangle}\langle S_i S_j\rangle \simeq -h\sum_{i=1}^{N}\langle S_i\rangle - J\sum_{\langle ij\rangle}\langle S_i\rangle\langle S_j\rangle \tag{18.1}$$

$$= -hNm - \frac{1}{2}JzNm^2, \tag{18.2}$$

where $\sum_{\langle ij\rangle}$ counts $Nz/2$ independent nearest-neighbour pairs and m is the magnetisation. Further we can write

$$m = \langle S_i\rangle = \{\text{prob. up spin}\} - \{\text{prob. down spin}\} \tag{18.3}$$

$$= c - (1-c), \tag{18.4}$$

where c is the mean-field probability of a spin being up. This gives the Gibbs entropy of a system of N spins as

$$S = -Nk\left[c\ln c + (1-c)\ln(1-c)\right]. \tag{18.5}$$

The Helmholtz free energy as a function of m is then $F(m) = E(m) - TS(m)$, and the free energy per spin is

$$f(m) = \frac{F(m)}{N} = -\frac{1}{2}Jzm^2 - hm + kT\left[c\ln c + (1-c)\ln(1-c)\right]. \tag{18.6}$$

To find the equilibrium state we should minimise f with respect to m (at constant h), giving the familiar mean-field equation for the magnetisation:

$$m = \tanh\beta(Jzm + h). \tag{18.7}$$

To summarise, the procedure is to approximate the free energy (in this case by using a non-interacting, mean-field energy) as a function of the order parameter, then minimise.

18.2 Spontaneous symmetry breaking

Consider the free energy (with $h = 0$) as a function of m for different T:

$$\frac{f(m)}{kT} = -\frac{T_c}{T}\frac{m^2}{2} + \left[\left(\frac{1+m}{2}\right)\ln\left(\frac{1+m}{2}\right) + \left(\frac{1-m}{2}\right)\ln\left(\frac{1-m}{2}\right)\right]. \tag{18.8}$$

This function is plotted in Fig. 23, normalised so that $f(0) = 0$.

- $f(m)$ is symmetric in m i.e. $f(m) = f(-m)$.

- For $T > T_c$ there is a single minimum at $m = 0$, which is the equilibrium state.

- At the critical point $T = T_c$

$$\left.\frac{\partial^2 f}{\partial m^2}\right|_{m=0} = 0 \quad \text{at} \quad T = T_c. \tag{18.9}$$

- For $T < T_c$ two symmetric minima emerge; $m = 0$ becomes a maximum, which is thus unstable.

- For $T < T_c$ the symmetry $m \to -m$ is **spontaneously broken** since the system must select one of the two minima for its equilibrium state.

The structure of $f(m)$ is easier to see if we expand in m around $m = 0$. Then we have (see tutorial sheet)

$$\boxed{\frac{f(m)}{kT_c} = \frac{m^2}{2}\left(\frac{T - T_c}{T}\right) + \frac{m^4}{12} + O(m^6).}$$
(18.10)

The first two non-zero terms are all that is required to give the characteristic structure of $f(m)$ above and below T_c as sketched in the figure. Indeed one sees clearly that at the critical point the coefficient of m^2 vanishes and the ferromagnetic phases emerge when the coefficient is negative $(T < T_c)$.

18.3   Landau free energy

We now develop a theory where we write down an expression such as $f(m)$ directly, without going through detailed calculations. The idea is to forget the details of the microscopic model and consider just the symmetries. The recipe is as follows:

- Identify the order parameter, m say. This should be
  – zero in the high temperature disordered phase.
  – non-zero in the ordered phase.

- We aim just to analyse the behaviour near to the critical point where m is small.

- We define the free energy density as a power series in the order parameter m.

- The series must only contain terms that respect the symmetry of the order parameter.

- The series is truncated as soon as the physics is captured, i.e. minimisation of free energy is possible.

As an example consider a magnetic system with the symmetry $m \leftrightarrow -m$ (in zero applied field). The order parameter is just the magnetisation m and we write down the free energy as

$$f(m) = \text{const.} + am^2 + bm^4 + O(m^6).$$
(18.11)

Note that the symmetry $m \to -m$ excludes any odd powers of m. The constant is unimportant and we can set it to zero. As we shall see, we need only consider the m^2 and m^4 terms. The coefficients a, b are unknown smooth functions of temperature. Minimising f wrt m, $\partial f/\partial m = 0$, yields

$$2am + 4bm^3 = 0 \quad \Rightarrow \quad m = 0 \quad \text{or} \quad m^2 = -\frac{a}{2b}.$$
(18.12)

The transition occurs when a changes sign. Therefore we identify the leading-order temperature dependence of a as

$$a(T) \simeq (T - T_c) \times \text{const.} = t\,a_0 \quad \text{where} \quad t = \frac{T - T_c}{T_c}$$
(18.13)

In the presence of a small applied field h (which breaks the symmetry) we add a linear term in m to $f(m)$

$$f(m) = -hm + am^2 + bm^4.$$
(18.14)

We can then understand the discontinuous transition below T_c by sketching the form of $f(m)$ when h passes through zero – the global minimum of $f(m)$ changes discontinuously from $|m|$ (for $h > 0$) to $-|m|$ (for $h < 0$); see Fig. 24.

97

18.4   Critical exponents

We can now easily recover the mean-field critical exponents.

- $m \propto \pm|t|^{\tilde{\beta}}$ $(T < T_c$ with $h = 0)$

  Minimising $f(m)$ in the low temperature ordered phase gives

  $$m = \left(\frac{a_0|t|}{2b}\right)^{1/2} \quad \Rightarrow \quad \tilde{\beta} = 1/2 \; . \tag{18.15}$$

- $h \propto |m|^\delta \mathrm{sgn}(m)$ (with $t = 0$)

  At T_c, $t = 0$ and we have

  $$f(m) = -hm + bm^4 \, . \tag{18.16}$$

  Therefore minimising f yields

  $$h = 4bm^3 \quad \Rightarrow \quad \delta = 3 \, . \tag{18.17}$$

- $\chi \propto \chi_\pm |t|^{-\gamma}$ (with $h = 0$)

  To work out the susceptibility χ near criticality we first minimise $f(m)$ w.r.t. m, which yields

  $$0 = -h + 2am + 4bm^3 \, . \tag{18.18}$$

  Then take the derivative wrt h to obtain an expression for χ:

  $$0 = -1 + 2a\chi + 12bm^2\chi \, . \tag{18.19}$$

  Using the relevant expressions for m at $h = 0$, we find

  $$\chi \sim \begin{cases} \frac{1}{2a_0 t} & \text{for} \quad t < 0 \quad (T > T_c) \\ \frac{1}{4a_0|t|} & \text{for} \quad t > 0 \quad (T < T_c) \end{cases} \tag{18.20}$$

  Thus $\gamma = 1$.

- $c_h \propto |t|^{-\alpha}$ (with $h = 0$), or $f \propto |t|^{2-\alpha}$ (see tutorial sheet)

- $\xi \propto \xi_\pm |t|^{-\nu}$

  In order to obtain ν (the correlation length exponent) we need to generalise to a spatially-varying order parameter $m(\underline{r})$. This is something like Debye–Hückel theory, where the result is $\nu = 1/2$ as mentioned previously.

Universality   The critical exponents at a continuous (second order) phase transition depend only on

- symmetry of the order parameter

- range of interactions

- dimension of space

Systems that are equivalent in this sense are said to be in the same **universality class**.

An example of two physically dissimilar transitions that lie in the same universality class is the ferromagnetic transition in the Ising model and the liquid-gas transition in a fluid. This is may not seem surprising, since the lattice gas model maps onto the Ising Model. But there is more to it than this, because the lattice gas model has a particle-hole symmetry ($c \to 1 - c$) that is not apparently shared by real fluids. Despite this, the two do indeed lie in the same universality class.

Unfortunately, the critical exponent values actually predicted from the Landau expansion are, in most cases, not correct. For example, we can compare the Landau predictions for Ising magnets, which are the same in both two and three dimensions, with those of Onsager's exact 2D solution and the best estimates in 3D from experiment and simulation:

$$\begin{array}{lllll} \text{Landau} & \beta = 1/2 & \delta = 3 & \alpha = 0 & \gamma = 1 \\ \text{Ising 2D} & \beta = 1/8 & \delta = 15 & \alpha = 0 & \gamma = 7/4 \\ \text{Ising 3D} & \beta = 0.31 & \delta = 5.2 & \alpha = 0.12 & \gamma = 1.24 \end{array} \tag{18.21}$$

The reason for the failure of Landau theory is the fact that it is a *mean-field* theory and thus neglects correlations and therefore *fluctuations* near the critical point.

Consider, for example, the heat capacity which in three dimensions behaves near the critical point as

$$C_h = \left.\frac{\partial E}{\partial T}\right|_{h=0} \propto |T - T_c|^{-0.12} . \tag{18.22}$$

Referring back to our discussion in section 5 this means that even for a thermodynamically large system, the fluctuations in the energy

$$\langle (\Delta E)^2 \rangle = kT^2 C_h , \tag{18.23}$$

become formally divergent at the critical point (note that for magnets we have C_h instead of C_V here). Similar remarks hold for other thermodynamic quantities, all of whose fluctuations would normally be negligible in a large system. This includes the mean magnetisation m. So the idea of minimising the free energy density $f(m)$ with respect to a *single*, well-defined value of m is questionable to say the least. Note that the fluctuations are cooperative: they involve many spins working in consort and cannot be viewed as the sum of many independent fluctuations of individual spins.

Let us note here that Landau theory can be extended to include the effect of space dimension and becomes **Landau–Ginzburg theory**. Above an **upper critical dimension** (ucd), which is $d = 4$ for the Ising Universality Class, the fluctuations are not important and Landau exponents become exact.

18.5   Other Landau free energies

One might have the impression from the discussion so far that Landau theory always leads to the same expansion for $f(m)$. Let us illustrate that this is not the case with a quick look at some more complicated examples.

- Vector order parameter:

  Consider a vector magnetic moment $\underline{m}$. Where $\underline{m}$ has components $\alpha = 1, 2, \ldots$. Since the free energy f is a scalar it involves scalar invariants of $\underline{m}$ i.e. $|\underline{m}|^2$

$$f = a_0 t |\underline{m}|^2 + b|\underline{m}|^4 . \tag{18.24}$$

In Fig. 25 we see for $\underline{m} = (m_1, m_2)$ that f has a 'Mexican hat' form and the ground state has infinite degeneracy. Thus the system can be moved around the ground state 'manifold' with zero free energy cost. These are known as Goldstone excitations.

- Tensor order parameter:

  An example of a tensorial order parameter occurs in the study of liquid crystals. These are rod-like molecules but with no head or tail. The orientation of a molecule is described by a 'head less' vector $\underline{n}$ and the order parameter is given by the correlations between the $\underline{n}$

  $$Q_{ij} = \langle n_i n_j \rangle - \frac{1}{3}\delta_{ij} \tag{18.25}$$

  which is a tensor. If the orientations are random then since $\underline{n}$ is normalised ($\sum_i n_i^2 = 1$), $\langle n_i n_j \rangle = \frac{1}{3}\delta_{ij}$ and the order parameter vanishes in the high temperature 'isotropic phase'. However at low temperatures there are nematic phases where the molecules line up.

  In the case of the tensor order parameter there is actually a cubic term $\text{Tr}(Q^3)$ which comes into play. This term is a scalar invariant and there is no symmetry to exclude it from the Landau expansion.

- Discontinuous Phase Transitions:

  An interesting point is that the presence a cubic term in the Landau expansion

  $$f(m) = am^2 + cm^3 + bm^4 \tag{18.26}$$

  can cause a discontinuous or first order phase transition, and this indeed happens in liquid crystals. You should try sketching the cubic form of f to see how this can occur.

  Similarly a free energy of the form

  $$f(m) = am^2 + bm^4 + cm^6 \tag{18.27}$$

  will lead to a discontinuous phase transition if the coefficient b is negative. Again sketch $f(m)$ to see how this happens.

18.6   Broader applications

The symmetry-based approach of Landau theory is extremely powerful, and need not be confined to its initial applications in condensed matter. In subjects as different as particle physics and cosmology, heavy use is made of **effective field theory**, in which again a starting point is to write down the most general form for quantities of interest, subject to known symmetries. Spontaneous symmetry breaking is central to modern particle physics, where space is permeated by a complex **Higgs field**, ϕ, with a potential of the same form as Fig. 25 (where the two coordinates are the real and imaginary parts of ϕ). Such a potential was well known in condensed matter (where it also manifests itself in superconductivity), and it was Philip Anderson who suggested in 1962 that the phenomenon might also apply in particle physics. It then took until 1964 for Higgs and others to realise that the idea both allowed mass to be given to otherwise massless particles such as the weak-interaction vector boson, W – and that a new massive particle was predicted, corresponding to quantized oscillations of the Higgs field itself. It is a tribute to the power of Landau's ideas that they were able to reach so far across different areas of physics.
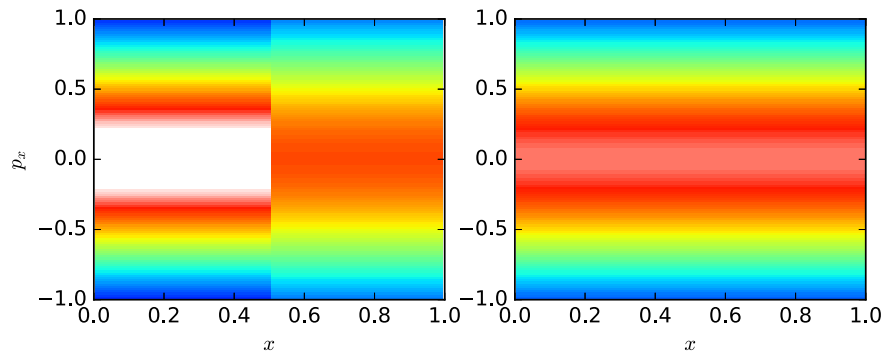
Figure 6: An example of the phase-space density before and after an irreversible change in a 1D 'box' of gas. Initially, the two halves are at different temperatures, T_1 and T_2 (left). After evolution, the distribution is Boltzmann with $T = (T_1 + T_2)/2$.
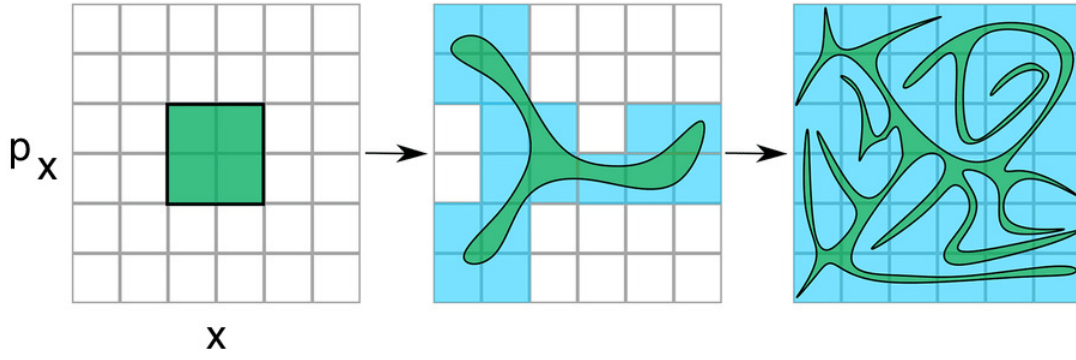
Figure 7: Coarse graining at work. Phase-space density is constant, so the total occupied phase-space volume must be constant in order for probability to be conserved. But the shape of the volume elements can become highly distorted and 'thready': any practical pixellization will then include much empty space, so that the average phase-space density is reduced.
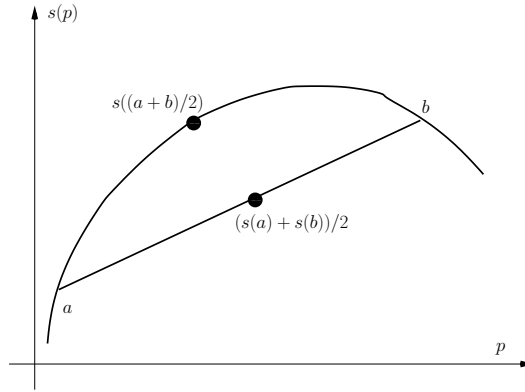
Figure 8: Property of a concave function, which always lies above the straight line joining any two points. This is guaranteed if the second derivative is negative.
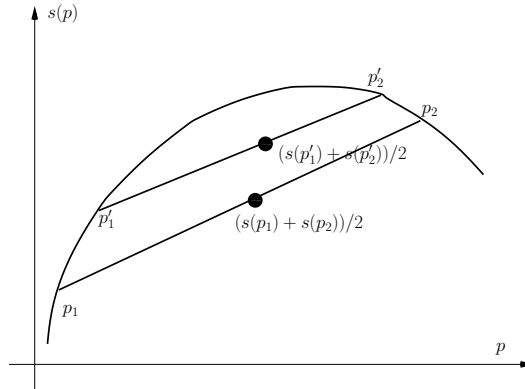
Figure 9: Due to concavity the entropy always increases under the master equation, which brings the probabilities p_1 and p_2 progressively closer.
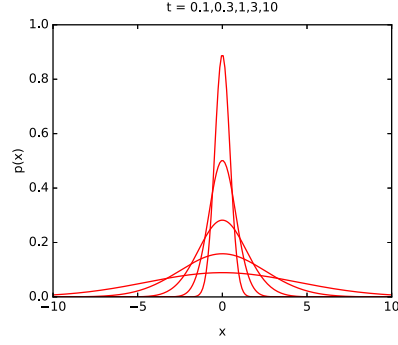
Figure 10: The solution of the diffusion equation (Gaussian distribution) with $D = 1$ as t increases.
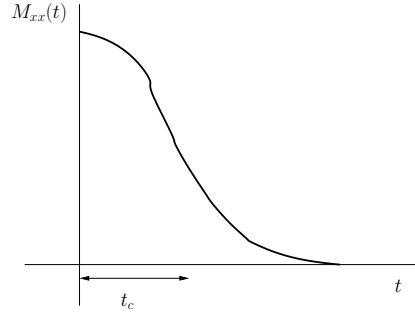
Figure 11: Sketch of the autocorrelation function of some zero-mean thermodynamic variable X against time lag: $M_{XX}(t)$.
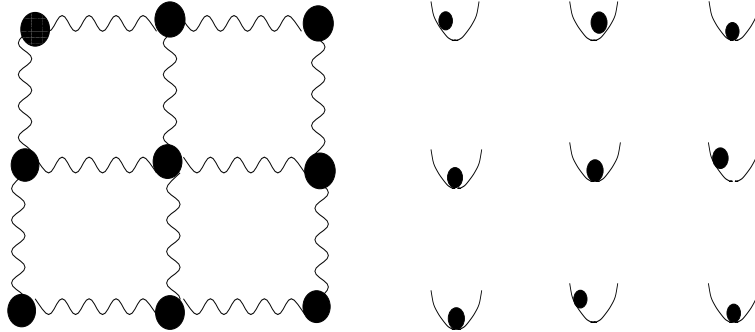
Figure 12: Left: model of crystal solid as atoms connected by harmonic springs for simplicity just to the nearest neighbour (this is the harmonic approximation). Right: model of each atom as an independent harmonic oscillator (Einstein Model).
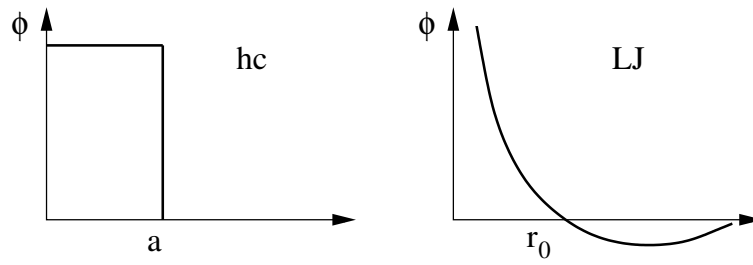
Figure 13: Sketch of the hard-sphere potential and the Lennard–Jones potential.
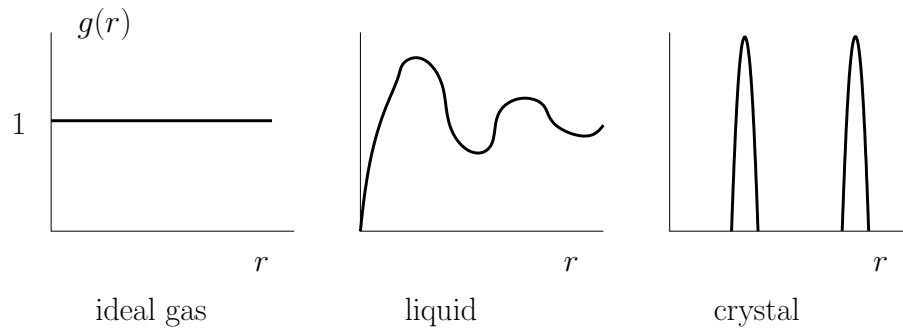
Figure 14: Sketch of $g(r)$ for an ideal gas, a dense gas and for a crystalline solid.
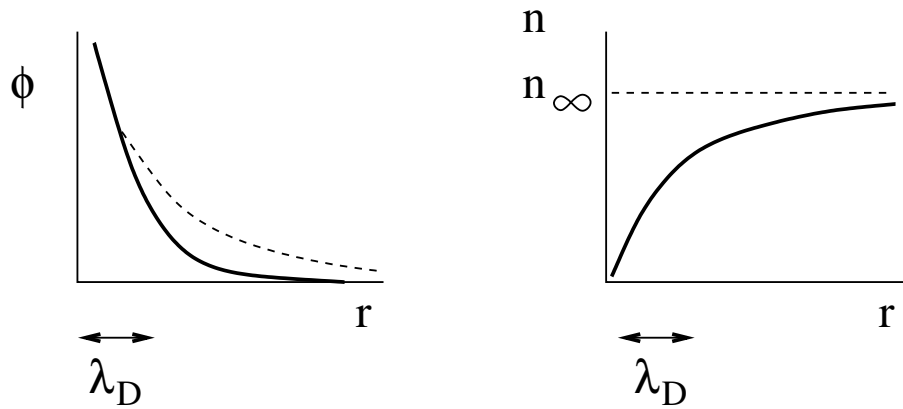
Figure 15: Sketch of the potential $\phi(r)$ and the density $n(r)$, given a charge (of same sign) at the origin. Note in the first sketch how the Coulomb potential is screened out at distances of order λ_D and in the second sketch how there is a depletion zone size of order λ_D.
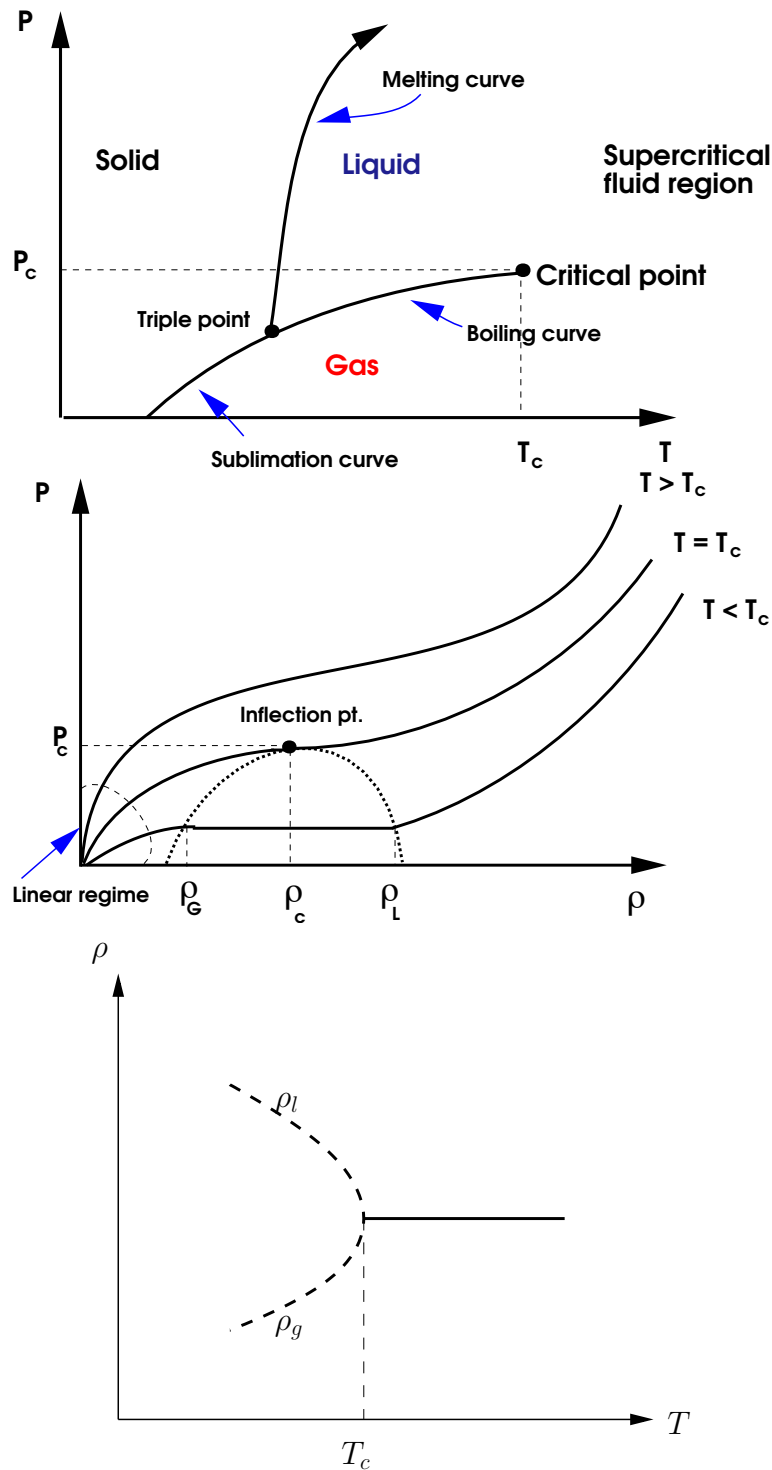
Figure 16: (a) Phase diagram for a fluid in the P–T plane. Note the 'vapour-pressure' (boiling) curve which separates the liquid and gas phases and terminates at the critical point (b) Isotherms in the P–ρ plane. Note the emergence of a flat piece in the co-existence region when $T < T_c$ (c) Plot of $\rho(T)$ as we move along the co-existence curve. Note the emergence of two values ρ_l and ρ_g for $T < T_c$. [Pictures (a), (b) from Tuckerman.]
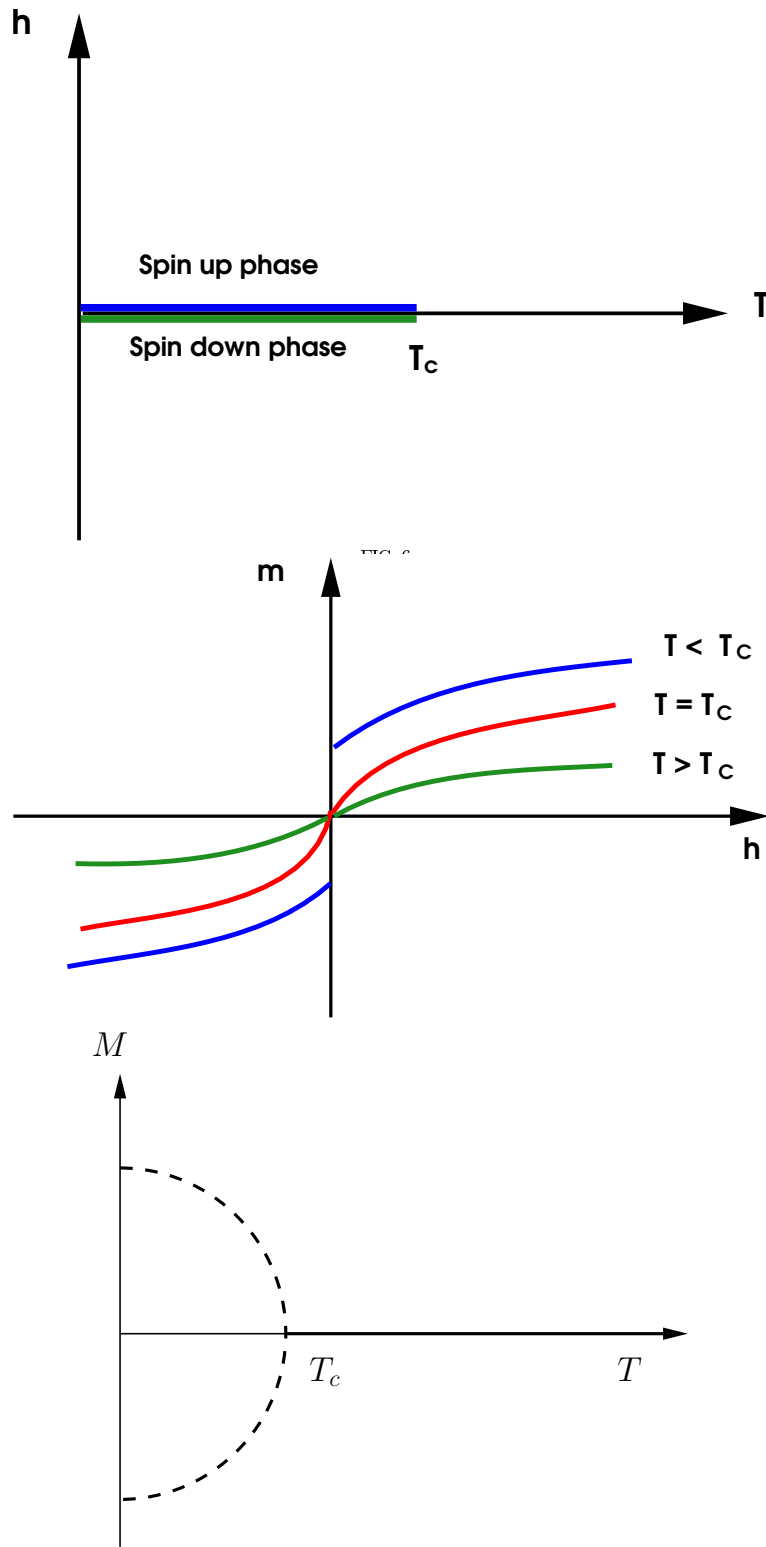
Figure 17: (a) Phase diagram for a magnetic system in the H–T plane where H is the applied magnetic field. The coexistence curve is along the $H = 0$ axis (b) Isotherms in the M–H plane where M is the global magnetisation. Note the emergence of a flat piece when $T < T_c$ (c) Plot of $M(T)$. Note the emergence of two non-zero values $\pm|M|$ for $T < T_c$.
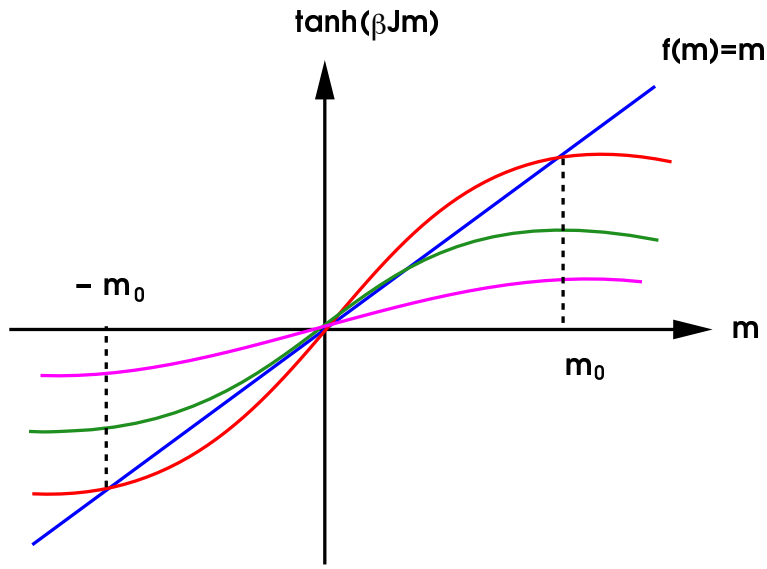
Figure 18: Picture of $\tanh(\beta Jzm)$ versus m and the straight line m versus m. The intersections give the solutions of (16.11).
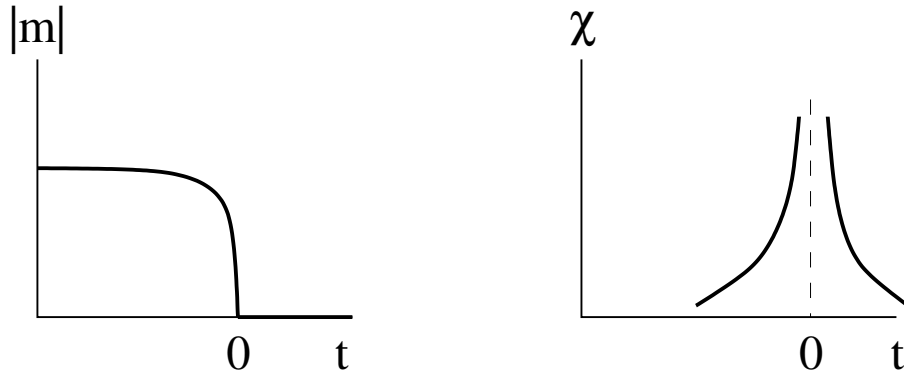
Figure 19: Sketch of the critical behaviour of $|m|$ and χ as functions of the reduced temperature t.
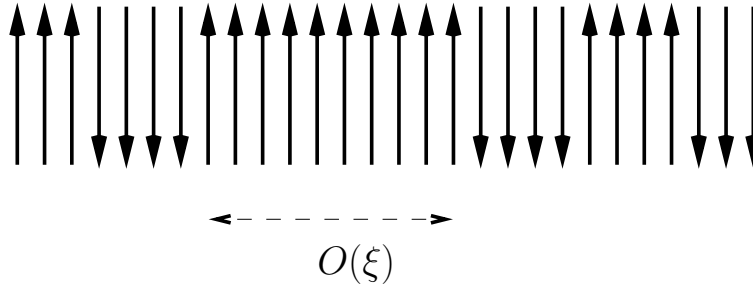

Figure 20: Sketch of a typical configuration of spins above T_c which consists of clusters of correlated spins but zero overall magnetisation. ξ is the length of the largest clusters.
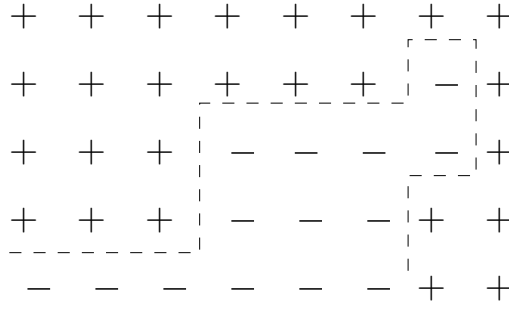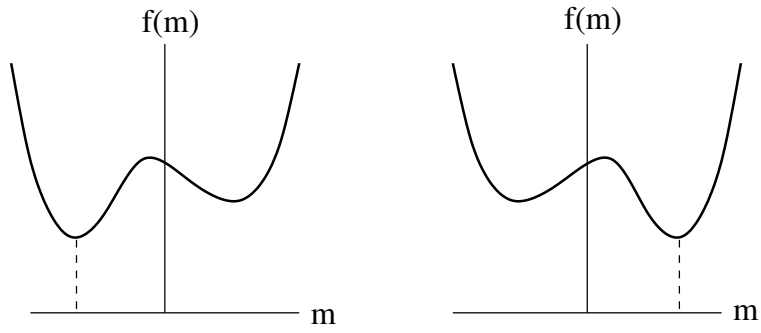
Figure 21: In 2D a domain wall between up and down domains is an extended object.
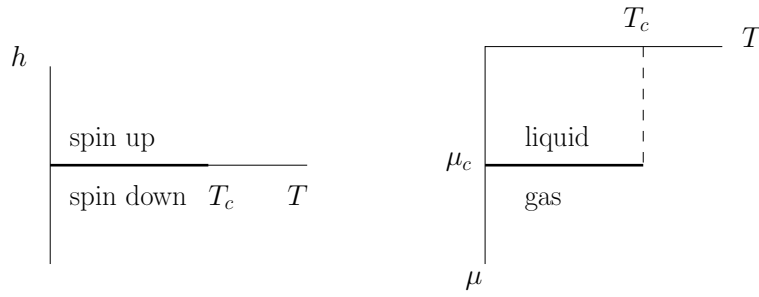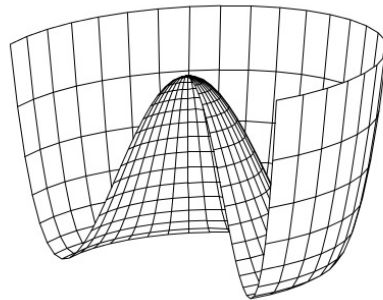

Figure 22: Phase diagram for magnetic system in the h–T plane (left sketch) and of the lattice gas in the μ–T plane (right sketch).
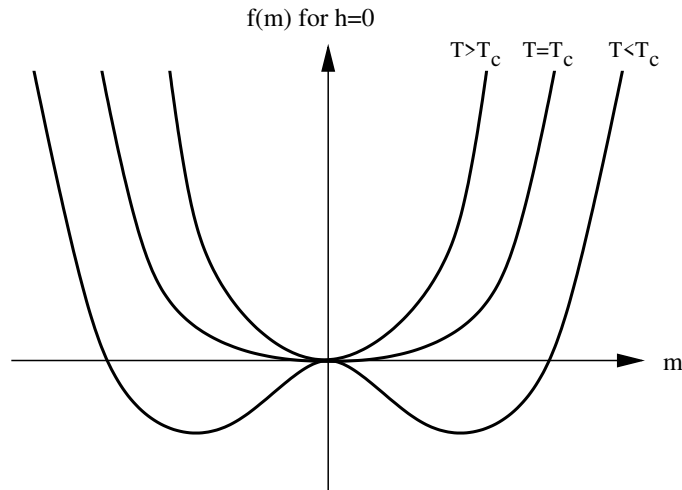
Figure 23: Sketches of $f(m)$ for $T > T_c$, $T = T_c$ and $T < T_c$.

Figure 24: Sketch of Landau free energy $f(m) = -hm + am^2 + bm^4$ for $a < 0$ and $b > 0$, as a function of the applied field: $h < 0$ (left) and $h > 0$ (right).

Figure 25: Sketch of Landau free energy in low T phase for $\underline{m} = (m_1, m_2)$.