Statistical Mechanics

University of Cambridge Part II Natural Sciences Tripos

Yue Wu

Yusuf Hamied Department of Chemistry
Lensfield Road,
Cambridge, CB2 1EW

yw628@cam.ac.uk

Acknowledgements

Nothing in these lecture notes is original. They are largely based on the notes by Dr. Aleks Reinhardt, who lectured this course in 2024. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

Preface

This course assumes a foundational level of understanding in thermodynamics and statistical mechanics — in particular, the contents that have been lectured in NST Part IA Chemistry *Thermodynamics and Equilibria* and Part IB Chemistry A *Molecular Energy Levels and Thermodynamics*. It may be a little fast for readers without that background. If you are new to thermodynamics/statistical mechanics, but have studied elementary quantum mechanics, an extremely good resource is the course notes on Mathematical Tripos Part II: *Statistical Physics* by Prof. David Tong. You can find it here.

Contents

1 Introduction

> *"Anyone who wants to analyze the properties of matter in a real problem might want to start by writing down the fundamental equations and then try to solve them mathematically. Although there are people who try to use such an approach, these people are the failures in this field..."*
>
> Richard Feynman

Suppose you know all the fundamental laws of physics, if I give you a box containing 10^{23} particles and tell you their mass, their charge, their interactions, and so on, what can you tell me about the stuff in the box?

There's one strategy that definitely won't work: writing down the Schrödinger equation for 10^{23} particles and solving it. That's typically not trivial to do for 2 or 3 particles, and generally impossible for 23 particles, let alone 10^{23}. What's more, even if you could find the wavefunction of the system, what would you do with it? The positions of individual particles are of little interest to anyone. We want answers to much more basic, almost childish, questions about the contents of the box. Is it wet? Is it hot? What colour is it? Is the box in danger of exploding? What happens if we squeeze it, pull it, heat it up? How can we begin to answer these kinds of questions starting from the fundamental laws of physics?

The purpose of this course is to introduce the dictionary that allows you to translate from the microscopic world where the laws of Nature are written to the everyday macroscopic world that we're familiar with.

A large part of this course will be devoted to figuring out the interesting things that happen when you throw 10^{23} particles together. One of the recurring themes will be that

$$10^{23} \neq 1\,. \tag{1.1}$$

More is different[1]: there are key concepts that are not visible in the underlying laws of physics but emerge only when we consider a large collection of particles. One very simple example is temperature. This is not a fundamental concept: it doesn't make sense to talk about the temperature of a single electron. But it would be impossible to talk about physics of the everyday world around us without mentioning temperature. This illustrates the fact that the language needed to describe physics on one scale is very different from that needed on other scales. We'll see several similar emergent quantities in this course, including the phenomenon of phase transitions where the smooth continuous laws of physics conspire to give abrupt, discontinuous changes in the structure of matter.

Statistical mechanics is the art of turning the microscopic laws of physics into a description of Nature on a macroscopic scale.[2]

[1] Anderson, P. W. (1972). More Is Different: Broken symmetry and the nature of the hierarchical structure of science. Science, 177(4047), 393-396.

[2] This well-written introduction is from Prof. David Tong.

2 Fundamentals of Statistical Mechanics

2.1 The Microcanonical Ensemble

We will start by stating the most fundamental assumption in statistical mechanics. It is the idea that we should take the simplest-minded approach possible and treat all microstates the same — since we know nothing else about the system, such a democratic approach seems eminently reasonable.

Postulate (Principle of equal *a priori* probabilities). A system with fixed N, V and E is equally likely to be found in any of its $\Omega(E)$ energy eigenstates[3].

Then the probability that the system with fixed energy E is in a given state $|n\rangle$ with energy E is simply

$$\text{Prob}(|n\rangle) = \frac{1}{\Omega(E)}\,. \tag{2.1}$$

The probability that the system is in a state with some different energy $E' \neq E$ is zero. This type of system with a fixed energy is known as a *microcanonical ensemble*.

2.2 Entropy and the Second Law

Consider a system with total energy E that consists of two weakly interacting subsystems, which means that they can only exchange energy, as shown in figure 2.1. The energy is distributed over the two subsystems such that

$$E_1 + E_2 = E\,. \tag{2.2}$$

For a given E_1, the total number of degenerate states of the system is $\Omega_1(E_1) \times \Omega_2(E_2)$. We choose to take the logarithm of the degeneracy so that the value is extensive. This gives

$$\ln \Omega(E_1, E - E_1) = \ln \Omega_1(E_1) + \ln \Omega_2(E_2)\,. \tag{2.3}$$

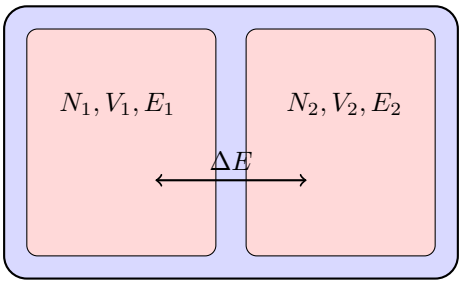


Figure 2.1: Two weakly coupled systems that can exchange energy such that the total energy of the universe is conserved to be E.

By the principle of equal *a priori* probabilities, the most likely value of E_1 is the one that maximises $\ln \Omega(E_1, E - E_1)$, i.e.

$$\left(\frac{\partial \ln \Omega(E_1, E - E_1)}{\partial E_1}\right)_{N_1, V_1} = 0\,. \tag{2.4}$$

Using the fact that $\mathrm{d}E_1 = -\mathrm{d}E_2$, we obtain the condition

$$\left(\frac{\partial \ln \Omega_1(E_1)}{\partial E_1}\right)_{N_1, V_1} = \left(\frac{\partial \ln \Omega_2(E_2)}{\partial E_2}\right)_{N_2, V_2}\,. \tag{2.5}$$

[3]This is not provable. It is the fundamental assumption that the entire Statistical Mechanics is derived upon. Some justifications of this assumption include Boltzmann's H-theorem and Liouville's theorem.

We define

$$\beta(E,V,N) \coloneqq \left(\frac{\partial \ln \Omega(E,V,N)}{\partial E}\right)_{N,V}, \tag{2.6}$$

then the equilibrium condition becomes

$$\beta(E_1,V_1,N_1) = \beta(E_2,V_2,N_2)\,. \tag{2.7}$$

Let's examine how this condition is linked to the three laws in thermodynamics.

2.2.1 Boltzmann Entropy and the Second Law

The thermal equilibrium condition of equal β must be equivalent to the condition that the two subsystems have the same temperature (from classical thermodynamics). Also, just like the thermodynamic entropy S, $\ln\Omega$ is a state function of E, V and N. Moreover, when the thermal equilibrium is reached, $\ln\Omega$ of the total system is maximised, just like the entropy S should be maximised at equilibrium by the second law of thermodynamics. This suggests that $\ln\Omega$ is a monotonically increasing function of entropy S. Since both S and $\ln\Omega$ are extensive, S must be proportional to $\ln\Omega$:

$$S(N,V,E) = k_B \ln\Omega(N,V,E) \tag{2.8}$$

for some constant k_B known as *Boltzmann's constant*.

In the statistical picture, the second law of thermodynamics states that, at thermal equilibrium, the system is most likely to be found in the state that has the largest number of degenerate energy states.

2.2.2 Temperature and the First Law

The thermodynamic definition of temperature is

$$\frac{1}{T} \coloneqq \left(\frac{\partial S}{\partial E}\right)_{V,N}, \tag{2.9}$$

which is obtained from the fundamental equation for the internal energy (the first law of thermodynamics)

$$\mathrm{d}E = T\mathrm{d}S - P\mathrm{d}V + \mu\mathrm{d}N\,. \tag{2.10}$$

Hence β and temperature T is related by

$$\beta = \frac{1}{k_B T}\,. \tag{2.11}$$

The condition of equal β in statistical mechanics translates to the condition of equal temperature in thermodynamics.

2.2.3 The Third Law

The third law of thermodynamics states that at $T = 0$, the entropy of a pure, perfectly crystalline substance is zero, or equally the number of accessible states is equal to one (i.e. $\Omega = 1$, so that $S = k_B \ln\Omega = 0$). In other words: at absolute zero, the system is in its ground state — and this ground state is non-degenerate. However, in reality, residual entropy can exist, meaning that as $T \to 0$, one often observes S approaching a constant.

2.3 The Canonical Ensemble

Now let's consider a system that is at thermal equilibrium with a large heat bath. The total system is isolated so that $E_{\text{tot}} = E_{\text{bath}} + E_{\text{sys}}$ is fixed. The system is weakly coupled to the bath so they can exchange energy. This configuration is commonly called the *canonical ensemble*. Suppose the system is prepared in a specific state $|i\rangle$ with energy E_i, then the bath has an energy $E_{\text{bath}} = E_{\text{tot}} - E_i$ with degeneracy $\Omega_{\text{bath}}(E_{\text{tot}} - E_i)$. The degeneracy of the whole system is $\Omega = \Omega_{\text{sys}}\Omega_{\text{bath}} = \Omega_{\text{bath}}$. The principle of equal *a priori* probabilities applies only to an isolated system, which is the system plus the bath (the universe) in our case, so all the states of the universe with total energy E_{tot} are equally likely to occur. Hence, the probability of finding the system in state $|i\rangle$ is given by

$$\text{Prob}(|i\rangle) = \frac{\Omega_{\text{bath}}(E_{\text{tot}} - E_i)}{\sum_{|j\rangle} \Omega_{\text{bath}}(E_{\text{tot}} - E_j)} \,. \tag{2.12}$$

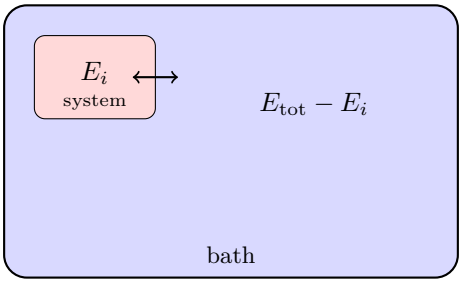


Figure 2.2: The system can exchange energy with the surrounding bath, such that the overall universe has fixed total energy E_{tot}.

Since the bath is very much larger than the system of interest, most of the energy of the universe will be in the bath. To compute $\Omega_{\text{bath}}(E_{\text{tot}} - E_i)$, we expand $\ln \Omega_{\text{bath}}(E_{\text{tot}} - E_i)$ about $E_{\text{bath}} = E_{\text{tot}}$

$$\ln \Omega_{\text{bath}}(E_{\text{tot}} - E_i) = \ln \Omega_{\text{bath}}(E_{\text{tot}}) - E_i \left(\frac{\partial \ln \Omega_{\text{bath}}(E)}{\partial E} \right)_{E=E_{\text{tot}}} + \mathcal{O}\left(\frac{1}{E_{\text{bath}}} \right) , \tag{2.13}$$

or,

$$\ln \Omega_{\text{bath}}(E_{\text{tot}} - E_i) = \ln \Omega_{\text{bath}}(E_{\text{tot}}) - \beta E_i + \mathcal{O}\left(\frac{1}{E_{\text{bath}}} \right) . \tag{2.14}$$

Hence,

$$\text{Prob}(|i\rangle) = \frac{\exp(-\beta E_i)}{\sum_{|j\rangle} \exp(-\beta E_j)} = \frac{\exp(-\beta E_i)}{Q} \,. \tag{2.15}$$

This is the *canonical distribution*, or *Boltzmann Distribution*. It describes the probability of the whole system being found in a particular microstate. The Boltzmann distribution should not be confused with the Maxwell–Boltzmann distribution (the distribution of the particle speeds in ideal gases) or Maxwell–Boltzmann statistics (the distribution of classical particles over energy states in thermal equilibrium).

In equation (2.13), we have ignored higher order terms in the Taylor expansion and claimed that it is $\mathcal{O}(1/E_{\text{bath}})$. This is straightforward to prove. The next term in the expansion is

$$\frac{E_i^2}{2} \left(\frac{\partial^2 \ln \Omega_{\text{bath}}(E)}{\partial E^2} \right)_{E=E_{\text{tot}}} = \frac{E_i^2}{2} \left(\frac{\partial \beta}{\partial E} \right)_{E=E_{\text{tot}}} = -\frac{E_i^2}{2 k_B T^2 C_{V,\text{bath}}} \,, \tag{2.16}$$

where we have used

$$C_V = \left(\frac{\partial E}{\partial T} \right)_{N,V} = \left(\frac{\partial E}{\partial \beta} \right)_{N,V} \left(\frac{\partial \beta}{\partial T} \right)_{N,V} . \tag{2.17}$$

Since both heat capacity and total energy are extensive, we can conclude that this term is $\mathcal{O}(1/E_{\text{bath}})$ as claimed.

2.3.1 Link to Thermodynamics

Knowing the energy distribution (2.15) allows us to compute the average energy of the system at a given temperature

$$\langle E_{\text{sys}} \rangle = \sum_{|i\rangle} \text{Prob}(|i\rangle) E_i = \frac{\sum_{|i\rangle} E_i \exp(-\beta E_i)}{Q} \,. \tag{2.18}$$

Next we note that the numerator looks very similar to the definition of Q. If we differentiate Q with respect to β, we obtain the negative of the numerator, and so we can write

$$\langle E_{\text{sys}} \rangle = -\frac{1}{Q} \left(\frac{\partial Q}{\partial \beta} \right)_{N,V} = - \left(\frac{\partial \ln Q}{\partial \beta} \right)_{N,V} \,. \tag{2.19}$$

We can change the differential variable to T instead

$$\langle E_{\text{sys}} \rangle = - \left(\frac{\partial \ln Q}{\partial T} \right)_{N,V} \frac{\partial T}{\partial \beta} = k_B T^2 \left(\frac{\partial \ln Q}{\partial T} \right)_{N,V} \,. \tag{2.20}$$

We can compare this to the thermodynamic relation

$$E = -T^2 \left(\frac{\partial (A/T)}{\partial T} \right)_{N,V} \,, \tag{2.21}$$

which can be easily derived by differentiating A/T, we can see that the Helmholtz free energy A is related to the partition function through the bridge relation

$$A = -k_B T \ln Q \,. \tag{2.22}$$

2.3.2 Fluctuations

Back to our canonical ensemble — a system of N particles with volume V in the large thermal bath. The probability of finding the system in any one of the $\Omega(N, V, E)$ states with energy E is

$$\text{Prob}(E) = \frac{\Omega(N, V, E) \exp(-\beta E)}{\sum_{|i\rangle} \exp(-\beta E_i)} \,. \tag{2.23}$$

Using the definition of entropy, we may rewrite this as

$$\text{Prob}(E) \sim \exp(-\beta E) \exp \left[\frac{S(N, V, E)}{k_B} \right] \,. \tag{2.24}$$

The most likely energy of the system, E^*, is the one for which

$$\left. \frac{\partial \,\text{Prob}(E)}{\partial E} \right|_{E=E^*} = 0 \quad \Rightarrow \quad \left(\frac{\partial S}{\partial E} \right)_{E=E^*} = \frac{1}{T} \,. \tag{2.25}$$

We can expand S in a Taylor series about E^* and define $\Delta E \coloneqq E - E^*$, we obtain

$$\frac{S(N, V, E)}{k_B} = \frac{S(N, V, E^*)}{k_B} + \frac{1}{k_B T} \Delta E + \frac{1}{2 k_B} \left(\frac{\partial^2 S}{\partial E^2} \right) (\Delta E)^2 + \mathcal{O}((\Delta E)^3) \,. \tag{2.26}$$

Take the logarithm of expression (2.24) and use the expansion (2.26), we can see that

$$\ln \text{Prob}(E^* + \Delta E) = c + \frac{1}{2 k_B} \left(\frac{\partial^2 S}{\partial E^2} \right) (\Delta E)^2 + \mathcal{O}((\Delta E)^3) \,, \tag{2.27}$$

where c is some constant independent of ΔE. The term linear in ΔE vanishes. In the limit of large N, we can ignore terms of order $(\Delta E)^3$ or above and find that

$$\text{Prob}(E^* + \Delta E) = c' \exp \left[\frac{1}{2 k_B} \left(\frac{\partial^2 S}{\partial E^2} \right) (\Delta E)^2 \right] \,, \tag{2.28}$$

where c' is another constant. We now recall the standard thermodynamics relation

$$\left(\frac{\partial^2 S}{\partial E^2}\right) = \left(\frac{\partial(1/T)}{\partial E}\right)_V = -\frac{1}{T^2}\left(\frac{\partial T}{\partial E}\right)_V = -\frac{1}{T^2}\left(\frac{\partial E}{\partial T}\right)_V^{-1} = -\frac{1}{C_V T^2}\,. \tag{2.29}$$

Finally, we normalise the probability density to obtain

$$\mathrm{Prob}(E^* + \Delta E) = (2\pi k_B T^2 C_V)^{-1/2} \exp\left[-\frac{(\Delta E)^2}{2k_B T^2 C_V}\right], \tag{2.30}$$

or

$$\mathrm{Prob}(E) = (2\pi k_B T^2 C_V)^{-1/2} \exp\left[-\frac{(E - E^*)^2}{2k_B T^2 C_V}\right]. \tag{2.31}$$

This is a Gaussian distribution. We find that the mean squared fluctuation in the energy of a system at constant N, V and T is directly related to the heat capacity C_V,

$$\sigma_E^2 = \left\langle (\Delta E)^2 \right\rangle = k_B T^2 C_V\,. \tag{2.32}$$

In fact, whilst the result that the probability distribution function is a Gaussian is important, it is possible to compute the variance of the distribution in a much simpler way if we are only interested in it. The thermal average of the energy is

$$\langle E \rangle = \frac{\sum_{|i\rangle} E_i \mathrm{e}^{-\beta E_i}}{Q} = -\frac{1}{Q}\frac{\partial Q}{\partial \beta}\,, \tag{2.33}$$

and hence the derivative of $\langle E \rangle$ with respect to β is

$$\frac{\partial \langle E \rangle}{\partial \beta} = -\frac{\sum_{|i\rangle} E_i^2 \mathrm{e}^{-\beta E_i}}{Q} + \left(\frac{1}{Q}\frac{\partial Q}{\partial \beta}\right)^2. \tag{2.34}$$

This can be simplified to

$$\frac{\partial \langle E \rangle}{\partial \beta} = -\left\langle E^2 \right\rangle + \langle E \rangle^2 \equiv -\sigma_E^2\,. \tag{2.35}$$

We can relate this to the heat capacity at constant volume

$$C_V = \left(\frac{\partial E}{\partial T}\right)_{N,V} = \left(\frac{\partial E}{\partial \beta}\right)_{N,V}\frac{\partial \beta}{\partial T} = -\frac{1}{k_B T^2}\left(\frac{\partial E}{\partial \beta}\right)_{N,V}. \tag{2.36}$$

Therefore $\sigma_E^2 = k_B T^2 C_V$ as before.

Since both E and C_V are extensive, we can write $\langle E \rangle = N \langle \epsilon \rangle$ and $C_V = N c_V$, where ϵ and c_V are the energy and heat capacity per particle. The relative variance in the energy vanishes in the thermodynamic limit as $N \to \infty$, since

$$\frac{\left\langle (\Delta E)^2 \right\rangle}{\langle E \rangle^2} = k_B T^2 \frac{C_V}{\langle E \rangle^2} = \frac{k_B T^2 c_V}{N \langle \epsilon \rangle^2} \sim \frac{1}{N}\,. \tag{2.37}$$

Similar results apply to other quantities of interest. Fluctuations about the average are negligible in the thermodynamic limit, and the statistical mechanical average of a thermodynamic property will correspond to its measured value.

2.3.3 Helmholtz Energy and Equilibrium

By the second law of thermodynamics, in an isolated system at equilibrium, the entropy of a system is maximised. In the canonical ensemble, a system is in contact with a thermal reservoir. Although

heat can be transferred between them, the total energy is conserved. The system plus the thermal bath is isolated, so we may apply the second law of thermodynamics. The total degeneracy is

$$\Omega_{\text{tot}} = \Omega_{\text{sys}}(E_{\text{sys}})\Omega_{\text{bath}}(E_{\text{tot}} - E_{\text{sys}})\,. \tag{2.38}$$

We can Taylor expand $\ln \Omega_{\text{bath}}(E_{\text{tot}} - E_{\text{sys}})$ up to linear order in E_{sys} and find that

$$\ln \Omega_{\text{tot}} = \ln \Omega_{\text{sys}}(E_{\text{sys}}) + \ln \Omega_{\text{bath}}(E_{\text{tot}}) - \beta E_{\text{sys}}\,. \tag{2.39}$$

The equilibrium condition is that the derivative of $\ln \Omega_{\text{tot}}$ with respect to E_{sys} vanishes. Since $\ln \Omega_{\text{bath}}(E_{\text{tot}})$ is independent of E_{sys}, the equilibrium condition is equivalent to minimising $\beta E_{\text{sys}} - \ln \Omega_{\text{sys}}(E_{\text{sys}}) = \beta(E_{\text{sys}} - TS_{\text{sys}})$. The Helmholtz energy A is defined as

$$A \coloneqq E - TS\,. \tag{2.40}$$

Therefore, at constant temperature and volume, A is at a minimum at equilibrium.

2.4 Pressure

Now consider a system that can exchange volume with a reservoir such that the total volume of the system plus the bath is fixed. The condition for equilibrium is that the total entropy is maximised, so we need to determine the maximum of

$$\ln \Omega(V_{\text{sys}}, V_{\text{tot}} - V_{\text{sys}}) = \ln \Omega_{\text{sys}}(V_{\text{sys}}) + \ln \Omega_{\text{bath}}(V_{\text{tot}} - V_{\text{sys}}) \tag{2.41}$$

with respect to V_{sys}. Using the definition of entropy,

$$\left(\frac{\partial S_{\text{sys}}}{\partial V_{\text{sys}}}\right)_{E,N} + \left(\frac{\partial S_{\text{bath}}}{\partial V_{\text{sys}}}\right)_{E,N} = 0\,. \tag{2.42}$$

Since $\mathrm{d}V_{\text{sys}} = -\mathrm{d}V_{\text{bath}}$,

$$\left(\frac{\partial S_{\text{sys}}}{\partial V_{\text{sys}}}\right)_{E,N} = \left(\frac{\partial S_{\text{bath}}}{\partial V_{\text{bath}}}\right)_{E,N}\,. \tag{2.43}$$

From thermodynamics, we have

$$\mathrm{d}S = \frac{1}{T}\mathrm{d}E + \frac{P}{T}\mathrm{d}V - \frac{\mu}{T}\mathrm{d}N \tag{2.44}$$

and hence $(\partial S/\partial V)_{E,N} = P/T$. Thus the equilibrium condition becomes

$$\frac{P_{\text{sys}}}{T_{\text{sys}}} = \frac{P_{\text{bath}}}{T_{\text{bath}}}\,. \tag{2.45}$$

If the system can exchange both energy and volume with the bath, then the conditions for equilibrium are $T_{\text{sys}} = T_{\text{bath}}$ and $P_{\text{sys}} = P_{\text{bath}}$.

In practice, it is convenient to use the fundamental relation for the Helmholtz energy, $\mathrm{d}A = -S\mathrm{d}T - P\mathrm{d}V + \mu\mathrm{d}N$, and the corresponding expression for the pressure

$$P = -\left(\frac{\partial A}{\partial V}\right)_{T,N} = k_B T \left(\frac{\partial \ln Q}{\partial V}\right)_{T,N}\,, \tag{2.46}$$

where we have used the bridge relation $A = -k_B T \ln Q$. Later, we shall use the relation to obtain a general expression of the pressure in a system of interacting particles. However, we will need to first derive some further results in classical statistical mechanics.

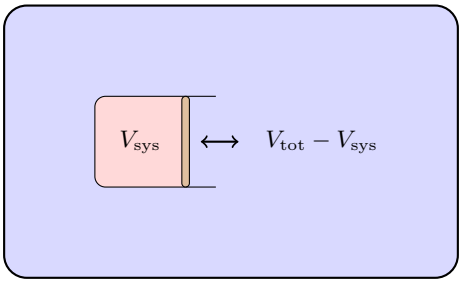


Figure 2.3: A system coupled to the bath such that a piston separating them can change position. The system and the bath can exchange volume.

2.5 Classical Statistical Mechanics

The thermal average of the expectation value of some observable X is

$$\langle X \rangle = \sum_{|i\rangle} \text{Prob}(|i\rangle) \langle i|X|i\rangle = \frac{\sum_{|i\rangle} \exp(-\beta E_i) \langle i|X|i\rangle}{\sum_{|i\rangle} \exp(-\beta E_i)} \,. \tag{2.47}$$

This suggests that in order to compute the thermal average, we first need to solve the Schrödinger equation of the many-body system of interest to obtain the eigenstates $|i\rangle$, then we need to compute the expectation value of the operator X for all those quantum states that have non-negligible statistical weight. This approach is doomed for all but the simplest systems. Fortunately, this can be simplified to a more workable expression in the classical limit $h \to 0$. For a d-dimensional, N-particle system, this is given by[4]

$$\langle X \rangle = \frac{\int \prod_{i=1}^{N} \mathrm{d}^d\mathbf{p}_i \, \mathrm{d}^d\mathbf{r}_i \, \exp\left[-\beta H(\{\mathbf{p}_i\},\{\mathbf{r}_i\})\right] X(\{\mathbf{p}_i\},\{\mathbf{r}_i\})}{\int \prod_{i=1}^{N} \mathrm{d}^d\mathbf{p}_i \, \mathrm{d}^d\mathbf{r}_i \, \exp\left[-\beta H(\{\mathbf{p}_i\},\{\mathbf{r}_i\})\right]} \,, \tag{2.48}$$

where $\mathbf{r}_i$ and $\mathbf{p}_i$ are the coordinate and momentum of the i^{th} particle and

$$H(\{\mathbf{p}_i\},\{\mathbf{r}_i\}) = \sum_{i}^{N} \frac{p_i^2}{2m_i} + U(\{\mathbf{r}_i\}) \tag{2.49}$$

is the classical Hamiltonian. The classical partition function is

$$Q = \frac{1}{h^{dN} N!} \int \prod_{i=1}^{N} \mathrm{d}^d\mathbf{p}_i \, \mathrm{d}^d\mathbf{r}_i \, \exp\left[-\beta H(\{\mathbf{p}_i\},\{\mathbf{r}_i\})\right] \,, \tag{2.50}$$

where the $1/N!$ factor is to account for the indistinguishability of the particles[5] and $1/h^{dN}$ is to make the partition function dimensionless.

2.5.1 Integration over the Momenta

From now on, we assume that the space is three-dimensional and all particles have the same mass m, although both conditions can be easily generalised. Note that due to the form of the Hamiltonian, the integration over momenta and spatial coordinates can be separated

$$Q(N,V,T) = \frac{1}{h^{3N} N!} \int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{p}_i \, \exp\left(-\beta \sum_{i=1}^{N} \frac{p_i^2}{2m}\right) \int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{r}_i \, \exp(-\beta U(\{\mathbf{r}_i\})) \,. \tag{2.51}$$

[4] The derivation is in the appendix section B

[5] The situation is actually a bit subtle… See *Gibbs' Paradox*.

Let's focus on the first integral. We can resolve $\mathbf{p}_i$ into its Cartesian components $p_i^2 = p_{ix}^2 + p_{iy}^2 + p_{iz}^2$, so it can be further factorised as

$$\int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{p}_i \, \exp\left(-\beta \sum_{i=1}^{N} \frac{p_i^2}{2m}\right) = \int \prod_{i=1}^{N} \mathrm{d}p_{ix} \, \mathrm{d}p_{iy} \, \mathrm{d}p_{iz} \, \exp\left(-\frac{\beta p_{ix}^2}{2m}\right) \exp\left(-\frac{\beta p_{iy}^2}{2m}\right) \exp\left(-\frac{\beta p_{iz}^2}{2m}\right) . \tag{2.52}$$

These are just $3N$ identical Gaussian integrals, each evaluated as

$$\int_{-\infty}^{\infty} \mathrm{d}p \, \exp\left(-\frac{\beta p^2}{2m}\right) = \sqrt{\frac{2m\pi}{\beta}} \, . \tag{2.53}$$

Hence, the partition function can be written as

$$\begin{aligned} Q(N,V,T) &= \frac{1}{h^{3N} N!} \left(\frac{2m\pi}{\beta}\right)^{3N/2} \int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{r}_i \, \exp[-\beta U(\{\mathbf{r}_i\})] \\ &=: \frac{1}{\Lambda^{3N} N!} \int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{r}_i \, \exp[-\beta U(\{\mathbf{r}_i\})] \, , \end{aligned} \tag{2.54}$$

where $\Lambda \coloneqq h/\sqrt{2\pi m k_B T}$ is the *de Broglie wavelength*. If the system consists of a mixture of two components A and B, then the classical partition function becomes

$$Q(N_A, N_B, V, T) = \frac{1}{\Lambda_A^{3N_A} \Lambda_B^{3N_B} N_A! N_B!} \int \prod_{i=1}^{N_A} \prod_{j=1}^{N_B} \mathrm{d}^3\mathbf{r}_i \, \mathrm{d}^3\mathbf{r}_j \, \exp[-\beta U(\{\mathbf{r}_i\}, \{\mathbf{r}_j\})] \, . \tag{2.55}$$

In molecular systems, a purely classical description is inappropriate. In particular, the internal energy levels of molecules like electronic energies, vibrations and rotations are not accounted for by the classical Hamiltonian. Then the classical partition function must have the form

$$Q(N,V,T) = \frac{q_{\mathrm{intra}}^N}{\Lambda^{3N} N!} \int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{r}_i \, \exp[-\beta U(\{\mathbf{r}_i\})] \, , \tag{2.56}$$

where q_{intra} is the partition sum over all molecular energy levels

$$q_{\mathrm{intra}} = \sum_i \exp(-\beta \epsilon_i) \, . \tag{2.57}$$

It often has the form

$$q_{\mathrm{intra}} = q_{\mathrm{vib}} q_{\mathrm{rot}} q_{\mathrm{elec}} \, . \tag{2.58}$$

2.5.2 Classical Ideal Gas

Consider a classical ideal gas, where there is no interaction between particles. The Hamiltonian of the system is given only by the kinetic energy contribution

$$H_{\mathrm{ideal}}(\{\mathbf{r}_i\}, \{\mathbf{p}_i\}) = \sum_{i=1}^{N} \frac{p_i^2}{2m} \, . \tag{2.59}$$

Since there is no interaction between particles, $U = 0$, the partition function is given by

$$Q_{\mathrm{ideal}}(N,V,T) = \frac{1}{\Lambda^{3N} N!} \int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{r}_i \, = \frac{V^N}{\Lambda^{3N} N!} \, . \tag{2.60}$$

Hence,

$$A = -k_B T \ln Q_{\text{ideal}} = -k_B T \ln \left(\frac{V^N}{\Lambda^{3N} N!} \right) , \tag{2.61}$$

and from the fundamental equation $\mathrm{d}A = -S\mathrm{d}T - P\mathrm{d}V + \mu \mathrm{d}N$, the expression for the pressure is

$$P_{\text{ideal}} = -\left(\frac{\partial A}{\partial V} \right)_{N,T} = \frac{N k_B T}{V} . \tag{2.62}$$

This is the well-known *ideal gas equation*, which can be rewritten in a perhaps more familiar form

$$PV = nRT . \tag{2.63}$$

2.6 Other Ensembles

There are five common ensembles that we usually consider

(i) *Microcanonical (NVE) ensemble*: an isolated system;

(ii) *Canonical (NVT) ensemble*: can exchange energy with the bath;

(iii) *Isothermal-isobaric (NPT) ensemble*: can exchange both energy and volume with the bath;

(iv) *Isoenthalpic-isobaric (NPH) ensemble*: can exchange volume with the bath;

(v) *Grand canonical (μVT) ensemble*: can exchange particles and energy with the bath.

We have already encountered microcanonical ensemble and canonical ensemble. Now let's consider an isothermal-isobaric ensemble: a system with N particles that can exchange energy and volume with a large reservoir. The probability of finding the system at a given microstate $|i\rangle$ with energy E_i and volume V_i is determined by $\Omega_{\text{bath}}(E_{\text{tot}} - E_i, V_{\text{tot}} - V_i)$. By Taylor expansion as before, and using

$$\left(\frac{\partial S}{\partial E} \right)_{N,V} = \frac{1}{T} , \left(\frac{\partial S}{\partial V} \right)_{E,N} = \frac{P}{T} , \tag{2.64}$$

we get

$$\ln \Omega_{\text{bath}}(E_{\text{tot}} - E_i, V_{\text{tot}} - V_i) = \ln \Omega_{\text{bath}}(E_{\text{tot}}, V_{\text{tot}}) - \frac{E_i + PV_i}{k_B T} , \tag{2.65}$$

where $E_i + PV_i =: H_i$ is the *enthalpy*. The probability of finding the system with volume V_i is therefore given by

$$\text{Prob}(V_i) = \frac{Q(N, V_i, T) \mathrm{e}^{-\beta P V_i}}{\Delta} , \tag{2.66}$$

where Δ is the *isothermal-isobaric partition function* quantum mechanically defined as

$$\Delta \coloneqq \sum_{V_i} Q(N, V_i, T) \mathrm{e}^{-\beta P V_i} . \tag{2.67}$$

Classically, it is defined as

$$\Delta \coloneqq \beta P \int_0^\infty \mathrm{d}V \, Q(N, V, T) \mathrm{e}^{-\beta P V} , \tag{2.68}$$

where βP is to make the quantity dimensionless, and correspondingly the probability density of finding the system with volume V_i is

$$p(V_i) = \frac{\beta P Q(N, V_i, T) \mathrm{e}^{-\beta P V_i}}{\Delta} . \tag{2.69}$$

As for other ensembles, we have the bridge relationship

$$G = -k_B T \ln \Delta . \tag{2.70}$$

Another ensemble of great importance is the grand canonical ensemble. Consider a system of volume V that can exchange particles and energy with the reservoir. The probability of finding the system in a state with N_i particles and energy E_i is determined by $\Omega_{\text{bath}}(E_{\text{tot}} - E_i, N_{\text{tot}} - N_i)$. Using

$$\left(\frac{\partial S}{\partial N}\right)_{E,V} = -\frac{\mu}{T}\,, \tag{2.71}$$

we get

$$\text{Prob}(N_i) = \frac{Q(N_i, V, T)\exp(\beta\mu N_i)}{\Xi}\,, \tag{2.72}$$

where

$$\Xi \coloneqq \sum_{N=0}^{\infty} Q(N, V, T)\exp(\beta\mu N) = \sum_{N=0}^{\infty} Q(N, V, T) z^N, \tag{2.73}$$

is the *grand partition function* and $z \coloneqq \mathrm{e}^{\beta\mu}$ is the *absolute activity* (or sometimes loosely called *fugacity*). The *grand potential* (*Landau potential*), defined as[6]

$$\Phi = E - TS - N\mu = -PV\,, \tag{2.74}$$

is related to the grand partition function by the bridge relation

$$\Phi = -k_B T \ln \Xi\,. \tag{2.75}$$

The average number of particles in the system is

$$\langle N\rangle = \left(\frac{\partial \ln \Xi}{\partial \beta\mu}\right)_{V,T}\,. \tag{2.76}$$

microcanonical ensemble constant N, V, E	*canonical ensemble* constant N, V, T
$S = k_B \ln \Omega(N, V, E)$	$A = E - TS = -k_B T \ln Q(N, V, T)$
$\mathrm{d}E = T\mathrm{d}S - P\mathrm{d}V + \mu\mathrm{d}N$	$\mathrm{d}A = -S\mathrm{d}T - P\mathrm{d}V + \mu\mathrm{d}N$
$(1/T) = (\partial S/\partial E)_{N,V}$	$E = (\partial(\beta A)/\partial\beta)_{N,V}$
$P = T(\partial S/\partial V)_{E,N}$	$P = -(\partial A/\partial V)_{N,T}$
$\mu = -T(\partial S/\partial N)_{E,V}$	$\mu = (\partial A/\partial N)_{V,T}$

isothermal-isobaric ensemble constant N, P, T	*grand canonical ensemble* constant μ, V, T
$G = A + PV = N\mu = -k_B T \ln \Delta(N, P, T)$	$\Phi = A - \mu N = -PV = -k_B T \ln \Xi(\mu, V, T)$
$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}P + \mu\mathrm{d}N$	$\mathrm{d}\Phi = -S\mathrm{d}T - P\mathrm{d}V - N\mathrm{d}\mu$
$H = E + PV = (\partial(\beta G)/\partial\beta)_{N,P}$	$N = -(\partial\Phi/\partial\mu)_{V,T}$
$V = (\partial G/\partial P)_{N,T}$	$P = -(\partial\Phi/\partial V)_{T,\mu}$
$\mu = (\partial G/\partial N)_{P,T}$	$E = N\mu + (\partial(\beta\Phi)/\partial\beta)_{\mu,V}$

Table 1: Common ensembles and related relations.

It is possible to generate a lot more different ensembles by different combinations of Legendre transforms; however, the microcanonical, canonical and grand canonical ensembles are the most important ones (as you can tell from their names).

[6]The second equality below follows from the fact that Φ is extensive, while among the natural variables T, V, N, only V is extensive, so Φ must be proportional to V. If you don't trust this relation, you can check appendix A for a formal proof.

2.6.1 Langmuir Adsorption

Why do we need different ensembles? In fact you will show in one of the exercises that in the thermodynamic limit, all the ensembles are equivalent. However, some calculations may be more difficult in one ensemble than another. The choice of ensemble is primarily a matter of convenience.

As an illustration, consider the adsorption of gas molecules on the metal surface. Assume the metal has M sites, each can accommodate one molecule at most, and if a particle is adsorbed, its energy is lowered by ϵ. The gas has chemical potential μ.

First, let's solve this question in the canonical ensemble. If N particles are adsorbed, the energy of the system is

$$E(N) = -N\epsilon\,, \tag{2.77}$$

and the degeneracy of this energy of the system is

$$\Omega(N) = \frac{M!}{N!(M-N)!}\,. \tag{2.78}$$

The canonical partition function is therefore

$$Q(N, M, T) = \frac{M!}{N!(M-N)!}\exp(\beta N\epsilon)\,. \tag{2.79}$$

The chemical potential is

$$\mu = -k_B T\left(\frac{\partial \ln Q}{\partial N}\right)_{V,T} = -\epsilon + k_B T(\ln N - \ln(M-N)) \tag{2.80}$$

by Stirling's approximation. Define the density of adsorbed particles as $\rho \coloneqq N/M$, then

$$\mu + \epsilon = k_B T \ln\left(\frac{\rho}{1-\rho}\right)\,. \tag{2.81}$$

By rearranging, we get

$$\rho = \frac{1}{1+\exp[-\beta(\mu+\epsilon)]}\,. \tag{2.82}$$

Now let's derive the same result in the grand canonical ensemble.

$$\begin{aligned}\Xi &= \sum_{N=0}^{M}\frac{M!}{N!(M-N)!}\exp[N\beta(\mu+\epsilon)] \\ &= (1+\exp[\beta(\mu+\epsilon)])^M\end{aligned} \tag{2.83}$$

by the binomial theorem. The average number of particles is therefore

$$\langle N\rangle = \left(\frac{\partial \ln \Xi}{\partial \beta\mu}\right)_{M,T} = \frac{M\exp[\beta(\mu+\epsilon)]}{1+\exp[\beta(\mu+\epsilon)]}\,. \tag{2.84}$$

It immediately follows that

$$\langle\rho\rangle = \frac{1}{1+\exp[-\beta(\mu+\epsilon)]} \tag{2.85}$$

as calculated before.

2.7 Pressure of Interacting Particles

2.7.1 Pressure from the Partition Function

Consider a system of interacting gas particles in a cubic box with edge length $L = V^{1/3}$. Define the scaled coordinate

$$\mathbf{s}_i = \frac{\mathbf{r}_i}{L}\,, \tag{2.86}$$

so that each component of $\mathbf{s}_i$ ranges between zero and unity. We can then write

$$Q(N,V,T) = \frac{V^N}{\Lambda^{3N} N!} \int_0^1 \cdots \int_0^1 \prod_{i=1}^N \mathrm{d}^3\mathbf{s}_i \, \exp[-\beta U(\{\mathbf{s}_i\}; L)], \tag{2.87}$$

where we have included L as the parameter of U to indicate that U depends on the real rather than the scaled distances between the particles. The pressure is therefore

$$\begin{aligned} P &= k_B T \left(\frac{\partial \ln V^N}{\partial V}\right)_{T,N} + k_B T \left(\frac{\partial}{\partial V} \ln \int_0^1 \cdots \int_0^1 \prod_{i=1}^N \mathrm{d}^3\mathbf{s}_i \, \exp[-\beta U(\{\mathbf{s}_i\}; L)]\right)_{T,N} \\ &= \frac{N k_B T}{V} - \frac{1}{\int \prod_{i=1}^N \mathrm{d}^3\mathbf{s}_i \, \exp[-\beta U(\{\mathbf{s}_i\}; L)]} \int \prod_{i=1}^N \mathrm{d}^3\mathbf{s}_i \, \exp[-\beta U(\{\mathbf{s}_i\}; L)] \frac{\partial U(\{\mathbf{s}_i\}; L)}{\partial V} \\ &= \underbrace{\frac{N k_B T}{V}}_{\text{ideal}} - \underbrace{\left\langle \frac{\partial U(\{\mathbf{s}_i\}; L)}{\partial V} \right\rangle}_{\text{non-ideal}}\,. \end{aligned} \tag{2.88}$$

We clearly see an ideal contribution and a non-ideal contribution which arise only if the particles are interacting ($U \neq 0$).

We now use the chain rule

$$\frac{\partial U(\{\mathbf{s}_i\}; L)}{\partial V} = \sum_{i=1}^N \sum_{k=1}^3 \frac{\partial U(\{\mathbf{s}_i\}; L)}{\partial r_{ik}} \frac{\partial r_{ik}}{\partial V}\,, \tag{2.89}$$

where $k \in \{1,2,3\}$ labelling the component. We have

$$\frac{\partial r_{ik}}{\partial V} = \frac{\partial L s_{ik}}{\partial V} = s_{ik} \frac{\partial V^{1/3}}{\partial V} = \frac{s_{ik}}{3V^{2/3}} = \frac{r_{ik}}{3V}\,. \tag{2.90}$$

Hence

$$\frac{\partial U(\{\mathbf{s}_i\}; L)}{\partial V} = \sum_{i=1}^N \sum_{k=1}^3 \frac{\partial U(\{\mathbf{s}_i\}; L)}{\partial r_{ik}} \frac{r_{ik}}{3V}\,. \tag{2.91}$$

For conservative forces, $\mathbf{f} = -\boldsymbol{\nabla} U$, so

$$\frac{\partial U(\{\mathbf{r}_i\})}{\partial r_{ik}} = -f_{ik}\,. \tag{2.92}$$

Therefore,

$$P = \frac{N k_B T}{V} + \frac{1}{3V} \left\langle \sum_{i=1}^N \mathbf{f}_i \cdot \mathbf{r}_i \right\rangle\,. \tag{2.93}$$

2.7.2 Pressure for Pairwise-additive Forces

Consider the special case that the intermolecular forces are pairwise-additive such that

$$\mathbf{f}_i = \sum_{j \neq i} \mathbf{f}_{ij}\,, \tag{2.94}$$

where $\mathbf{f}_{ij}$ is the force on particle i exerted by particle j. Then

$$P = \frac{Nk_BT}{V} + \left\langle \sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{1}{3V}\mathbf{r}_i \cdot \mathbf{f}_{ij} \right\rangle . \tag{2.95}$$

If we permute the dummy indices i and j, we get

$$P = \frac{Nk_BT}{V} + \left\langle \sum_{j=1}^{N}\sum_{i\neq j}^{N} \frac{1}{3V}\mathbf{r}_j \cdot \mathbf{f}_{ji} \right\rangle . \tag{2.96}$$

By Newton's third law, we get

$$P = \frac{Nk_BT}{V} + \frac{1}{6V}\left\langle \sum_{i=1}^{N}\sum_{j\neq i}^{N} \mathbf{r}_{ij} \cdot \mathbf{f}_{ij} \right\rangle , \tag{2.97}$$

where $\mathbf{r}_{ij} \coloneqq \mathbf{r}_i - \mathbf{r}_j$. This can be rewritten in the dimensionless form as

$$Z \coloneqq \frac{PV}{Nk_BT} = 1 + \frac{1}{6k_BT}\left\langle \sum_{j\neq i} \mathbf{r}_{ij} \cdot \mathbf{f}_{ij} \right\rangle_i , \tag{2.98}$$

where Z is the *compressibility factor*. Note that the $1/N$ factor and the sum over i combines to yield the average over i. For an ideal gas, $Z = 1$.

2.7.3 Law of Corresponding States

If the interactions between the molecules are pairwise-additive, then the potential energy $U(\{\mathbf{r}_i\})$ can be written as a sum of pair potentials

$$U(\{\mathbf{r}_i\}) = \frac{1}{2}\sum_{j=1}^{N}\sum_{i\neq j} \phi(r_{ij}) . \tag{2.99}$$

Figure 2.4 shows two simple examples of pair potentials. The hard-sphere potential is usually used to describe the interactions between uncharged colloids, and the Lennard-Jones potential is often used to describe the interaction between atoms or simple (nearly spherical) molecules. At distances less than the effective molecular diameter σ, the intermolecular pair potential is harshly repulsive. This is a consequence of the Pauli principle. At large distances $r \gg \sigma$, the London dispersion forces take over, the strength of which decays as $1/r^6$. The Lennard-Jones potential provides a convenient interpolation between the short-range repulsion and the long-range attraction. It is of the form

$$\phi(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] . \tag{2.100}$$

The depth of the potential is ϵ, whilst the range of the potential is determined by the effective diameter σ. Define $r^* \coloneqq r/\sigma$ and $\phi^* \coloneqq \phi/\epsilon$, then

$$\phi^*(r^*) = u_{\mathrm{LJ}}(r^*) \coloneqq 4\left[r^{*-12} - r^{*-6}\right] . \tag{2.101}$$

Therefore, if we express all energies in units of ϵ and distances in units of σ, then all Lennard-Jones potentials are identical.

If in appropriate units, the interactions between any kind of gas molecules are equal, their physics should also be the same. Using this definition, we can rewrite the compressibility in terms of the reduced quantities r^* and $f^*_{ij} \coloneqq f_{ij}\sigma/\epsilon$, giving

$$Z = 1 + \frac{\epsilon}{6k_BT}\left\langle \sum_{j\neq i} \mathbf{r}^*_{ij} \cdot \mathbf{f}^*_{ij} \right\rangle_i . \tag{2.102}$$

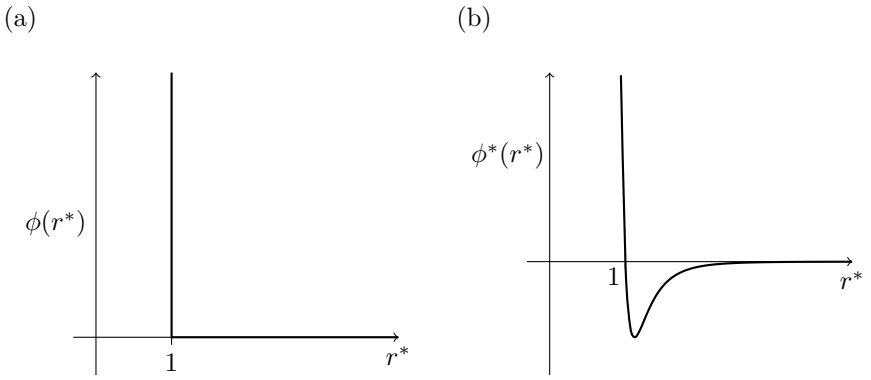


Figure 2.4: Two common pair potentials: (a) Hard-sphere potential (b) Lennard-Jones potential.

Now define the reduced temperature $T^* \coloneqq k_B T/\epsilon$ and the reduced density $\rho^* \coloneqq \sigma^3 N/V$, then

$$Z(\rho^*, T^*) = 1 + \frac{1}{6T^*} \left\langle \sum_{j \neq i} \mathbf{r}_{ij}^* \cdot \mathbf{f}_{ij}^* \right\rangle_i . \tag{2.103}$$

This equation implies that if we express the temperature in units of ϵ/k_B and density in units of σ^{-3}, then the compressibility factors of all substances that can be described by a pair potential in the same functional form collapse onto the same set of curves. This is known as *the principle of corresponding states*.

However, it should be stressed that not all substances obey the same law of corresponding states. For instance, the compressibility factor of dipolar molecules would not collapse onto any of the curves in the figure, because the functional form of the pair potential is too different — you wouldn't expect the interactions in polar species to be dominated by dispersion forces, as in non-polar species.

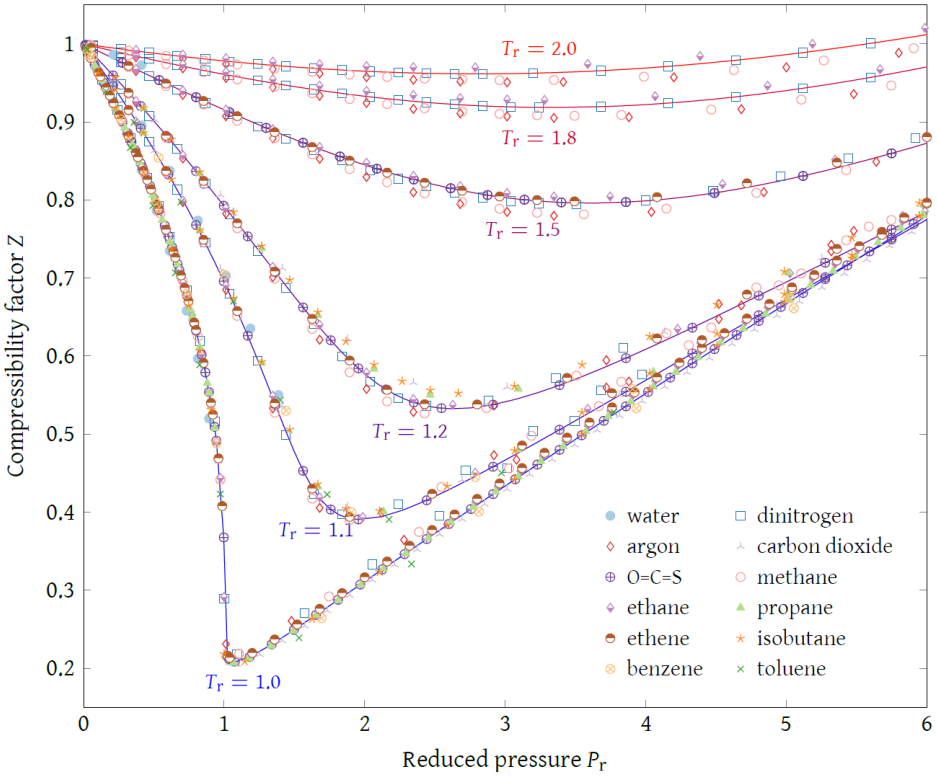


Figure 2.5: This beautiful figure is adapted from the official course notes by Dr. Aleks Reinhardt. The compressibilities of a range of gases are plotted in reduced quantities, and they overlap nicely. This implies that the functional form of their pair potentials must be very similar.

3 Phase Behaviour

3.1 Equilibrium between Phases

Consider an isolated system of N particles in a volume V with a total energy E. The system consists of multiple distinct phases between which particles, volume and energy can be exchanged. We ignore any contributions of the interfaces to the extensive properties, as their contribution scales as $N^{2/3}$, so their ratio compared with bulk contributions scales as $N^{-1/3}$, vanishing in the thermodynamic limit as $N \to \infty$. Then the equilibrium conditions of coexistence of phases are that the pressures of all coexisting phases must be equal ($P_1 = P_2 = \cdots = P$), as must be the temperatures ($T_1 = T_2 = \cdots = T$) and the chemical potentials of all species ($\mu_1^\alpha = \mu_2^\alpha = \cdots = \mu^\alpha$).

3.1.1 Stability

We use the thermodynamic arguments to find conditions for the stability of a single phase. Consider a homogeneous phase in an isolated system with constant N, V and E that is divided into two equal halves. At equilibrium, the entropy of the total system is at a maximum. Now consider transferring a small amount of energy ΔE from one half of the system to the other. At equilibrium, the temperature in the two halves of the system is the same, and hence to linear order of ΔE, the entropy does not change. Consider the second order variation,

$$\Delta S = \frac{1}{2}\left(\frac{\partial(1/T)}{\partial E}(\Delta E)^2\right)_1 + \frac{1}{2}\left(\frac{\partial(1/T)}{\partial E}(\Delta E)^2\right)_2 . \tag{3.1}$$

Since the two halves of the system are identical, these two terms are the same. We can write

$$\Delta S = \frac{\partial(1/T)}{\partial E}(\Delta E)^2 . \tag{3.2}$$

As S is a maximum, $\Delta S \leq 0$, and hence

$$\frac{\partial E}{\partial(1/T)} \leq 0 , \tag{3.3}$$

which implies that

$$-T^2\frac{\partial E}{\partial T} = -T^2 C_V \leq 0 , \tag{3.4}$$

and hence the heat capacity $C_V \geq 0$. Thermodynamic stability implies that the heat capacity is never negative.

A similar argument can be used to show that the compressibility of a system is non-negative. Consider the condition for equilibrium at constant N, V and T. Under these conditions, the Helmholtz energy must be a minimum. We divide the system into two equal parts, both containing $N/2$ particles in a volume $V/2$ at constant T. We now vary the volume of one half of the system by an amount ΔV, and the other by $-\Delta V$. To second order in ΔV, the total variation of the free energy is

$$\Delta A = \frac{\partial^2 A}{\partial V^2}(\Delta V)^2 . \tag{3.5}$$

This can be rewritten as

$$\Delta A = -\frac{\partial P}{\partial V}(\Delta V)^2 = \frac{1}{\kappa V}(\Delta V)^2 , \tag{3.6}$$

where κ is the *isothermal compressibility* of the system. As A is minimised at equilibrium, this implies that $\kappa \geq 0$.

3.1.2 Coexistence

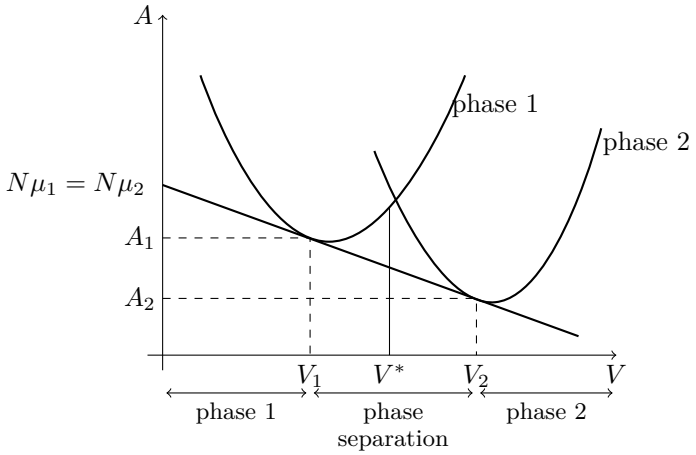
Consider a one-component system that can exist in two phases, with free energies $A_1(N, V, T)$ and $A_2(N, V, T)$. We would like to determine the condition that the two phases coexist. It turns out that there is an easy graphical method to do this. For a given temperature, we plot $A_1(N, V, T)$ and $A_2(N, V, T)$ as a function of V while keeping N and T constant. We now have to check if it is possible to draw a line that is tangent to both curves. Let's suppose this common tangent exists and it touches A_1 at volume V_1 and it touches A_2 at a volume V_2. As the curves A_1 and A_2 have a common tangent, the derivatives of A_1 and A_2 at V_1 and V_2, respectively, are the same. As $(\partial A/\partial V)_T = -P$, this implies that the pressure of phase 1 at V_1 is the same as that of phase 2 at V_2, i.e. $P_1 = P_2$. The tangents also have a common intercept at $V = 0$. This intercept is

$$A_1 - \frac{\partial A_1}{\partial V} V_1 = A_1 + PV_1 = G_1 = N\mu_1 \,, \tag{3.7}$$

and this is equal to

$$A_2 - \frac{\partial A_2}{\partial V} V_2 = A_2 + PV_2 = G_2 = N\mu_2 \,. \tag{3.8}$$

Therefore $\mu_1 = \mu_2$ and the volumes V_1 and V_2 are the volumes of the coexisting phases. A completely analogous analysis can be performed if we plot A/V against N/V, in which case the slope is related to the chemical potential and the intercept is related to the pressure.



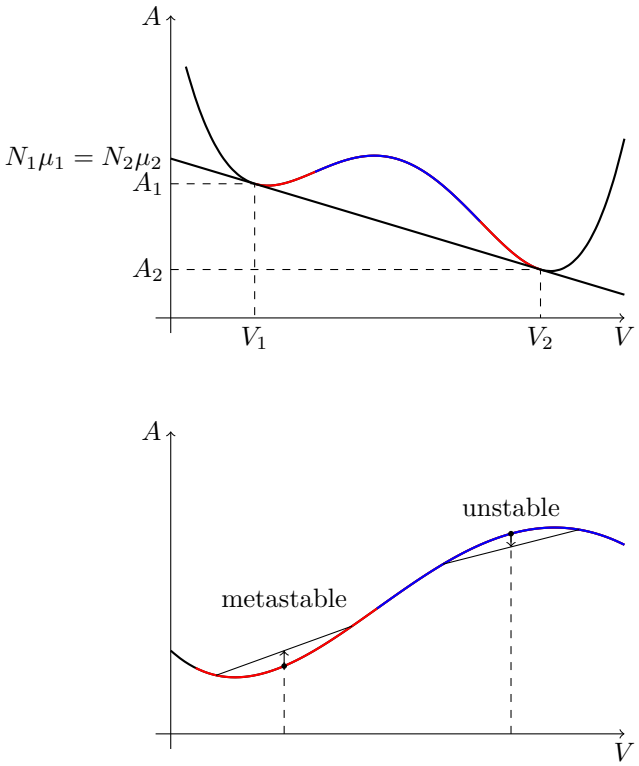
If the system is at some volume V^* with $V_1 \leq V^* \leq V_2$, the system would have lower Helmholtz energy if it is separated into the two phases than if it exists in either of the single phase. Define $v \coloneqq V/N$, then

$$Nv^* = N_1 v_1 + (N - N_1)v_2 \,. \tag{3.9}$$

We can then derive the proportion of particles being in phase 1 at equilibrium

$$x_1 \coloneqq \frac{N_1}{N} = \frac{v^* - v_2}{v_1 - v_2} \,. \tag{3.10}$$

Suppose now we have a free energy curve for some phase. We have shown that the free energy must be a convex function of V because the compressibility κ must be non-negative at equilibrium. Hence, the non-convex part in curve A versus V does not correspond to an equilibrium situation. We can still work out the point of coexistence by common tangent construction. But now we have a region where the phase separation is thermodynamically favourable, but the local second derivative of A with respect to V is still positive, meaning that any perturbation of such system will increase the free energy. Such a system is described as *metastable* as phase separation will involve a free-energy barrier. Between the inflection points, the free energy curve is non-convex and hence in this range a homogeneous phase would be absolutely unstable, with a negative compressibility. Such a system



will phase separate spontaneously under any infinitesimal fluctuation. The boundary between the metastable and unstable regions is called *spinodal*.

It should be stressed that the spinodal, although a useful qualitative concept, is not well defined. It appears naturally when we use approximate expressions for the free energy. However, an exact expression for the equilibrium free energy is necessarily convex and therefore has no spinodal.

3.1.3 Phase Diagrams

A phase diagram can tell us at a glance the conditions under which each phase is thermodynamically stable. For a one-component system, the thermodynamic variables P, T and V, which are usually used to describe the thermodynamic state of the matter, are not independent of one another. For example, if we specify the volume and the temperature of a system, the pressure will be defined as well through the equation of state. We can therefore think of the phase diagram as a surface in the PVT space. Such a three-dimensional phase diagram can be projected onto the $P-T$, $V-T$, $P-V$ or $T-\rho$ planes to give the more familiar two-dimensional phase diagrams.

There are many other parameters that we could vary; for example we could change the composition of a mixture or the strength of an external magnetic field, and so other types of phase diagrams can be constructed, some of which we shall see in the chapters that follow.

3.1.4 Thermodynamic Integration

The second law of thermodynamics states that for an isolated system with energy E, volume V and number of particles N, the entropy S is maximised when the system is at equilibrium. From this formulation of the second law, it is straightforward to derive the corresponding equilibrium conditions for systems that can exchange heat, particles or volume with a reservoir, as we have already shown.

In particular, if a system is in contact with a heat bath, such that its temperature T, volume V and number of particles N are fixed, then the Helmholtz energy $A \coloneqq E - TS$ is at a minimum in equilibrium. Analogously, for a system of N particles at constant pressure P and temperature T, the Gibbs energy $G \coloneqq A + PV$ is at a minimum.

If we wish to know which of two phases, denoted 1 and 2, is stable at a given temperature and density, we simply need to compare the Helmholtz energies A_1 and A_2. Unfortunately, neither the free energy nor the entropy can be measured directly. What we can measure are averages of *mechanical quantities*, i.e. averages of functions of the coordinates and momenta of the molecules in the system, such as the pressure or the dielectric constant. If we denote such a mechanical quantity by $X(\{\mathbf{p}_i\},\{\mathbf{r}_i\})$, then the average of X that can be measured in an experiment at constant N, V and T is

$$\langle X\rangle_{NVT} = \frac{\int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{p}_i\, \mathrm{d}^3\mathbf{r}_i\, X(\{\mathbf{p}_i\},\{\mathbf{r}_i\}) \exp[-\beta H(\{\mathbf{p}_i\},\{\mathbf{r}_i\})]}{\int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{p}_i\, \mathrm{d}^3\mathbf{r}_i\, \exp[-\beta H(\{\mathbf{p}_i\},\{\mathbf{r}_i\})]}\,, \tag{3.11}$$

where $H(\{\mathbf{p}_i\},\{\mathbf{r}_i\})$ is the system's Hamiltonian.

However, the entropy, the free energy and related quantities are not simply averages of functions of the phase-space coordinates of the system. Rather, they are directly related to the volume in phase space that is accessible to a system. For instance, in classical statistical mechanics, the Helmholtz energy A is directly related to the canonical partition function $Q(N,V,T)$ through

$$\begin{aligned} A &= -k_B T \ln Q(N,V,T) \\ &= -k_B T \ln\left[\frac{1}{h^{3N} N!}\int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{p}\, \mathrm{d}^3\mathbf{r}\, \exp[-\beta H(\{\mathbf{p}_i\},\{\mathbf{r}_i\})]\right]. \end{aligned} \tag{3.12}$$

Unlike quantities such as the internal energy, the pressure, or the polarisation, $Q(N,V,T)$ itself is not a canonical average over the phase space. This is why A and, for that matter, S or G, cannot be measured directly. We call quantities that depend directly on the available volume in the phase space *thermal quantities*. However, derivatives of the free energy with respect to the volume V or the temperature T are mechanical quantities and can be measured: namely, they are

$$\left(\frac{\partial A}{\partial V}\right)_{N,T} = -P \text{ and } \left(\frac{\partial (A/T)}{\partial (1/T)}\right)_{N,V} = E\,. \tag{3.13}$$

In order to compute the free energy of a system at a given temperature and density, we must find a reversible path in the volume-temperature plane that links the state under consideration to a state of known free energy. The change in A along that path can then simply be evaluated by integrating (3.13). Only a few thermodynamic states exist whose free energy is known exactly: one such state is the ideal gas; another may be perfectly ordered ground state at $T = 0$ K.

3.2 Gibbs–Bogoliubov Inequality

The aim of thermodynamic perturbation theory is to arrive at an estimate of the free energy (and all derived properties) of a many-body system, using as input information about the free energy and structure of a simpler reference system. We assume that the potential energy function of this reference system is denoted by U_0, while the potential energy function of the system of interest is denoted by U_1. In order to compute the free-energy difference between the known reference system and the system of interest, we use a linear parameterisation of the potential energy function

$$U(\lambda) \coloneqq \lambda U_1 + (1-\lambda) U_0\,. \tag{3.14}$$

The free energy of a system with this generalised potential energy is denoted $A(\lambda)$ and can be computed from $A(\lambda) = -k_B T \ln Q(\lambda)$. In particular, we noted when introducing (3.13) that

derivatives of the free energy are often mechanical observables, and the same applies in this case. The derivative of $A(\lambda)$ with respect to λ is

$$\begin{aligned}\left(\frac{\partial A}{\partial \lambda}\right)_{N,V,T} &= \frac{\int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{r}_i \left(\frac{\partial U(\{\mathbf{r}_i\};\lambda)}{\partial \lambda}\right) \exp[-\beta U(\{\mathbf{r}_i\};\lambda)]}{\int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{r}_i \exp[-\beta U(\{\mathbf{r}_i\};\lambda)]} \\ &= \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \\ &= \langle U_1 - U_0 \rangle_{\lambda} \,, \end{aligned} \tag{3.15}$$

and hence we can express the free-energy difference $A_1 - A_0$ as

$$A_1 - A_0 = \int_0^1 \mathrm{d}\lambda \, \langle U_1 - U_0 \rangle_{\lambda} \,, \tag{3.16}$$

where the subscript λ indicates that we evaluate the thermal average at that value of λ. This expression, first derived by John Kirkwood in 1935, allows us to compute the free energy of a system of interest by relating it to a reference system. This approach has been used extensively in computer simulations since the mid-1980s. It is often referred to as 'artificial' or 'Hamiltonian' thermodynamic integration because it entails changing the underlying interactions in the system's Hamiltonian.

Moreover, it allows us to find bounds on the free energy A_1. It is straightforward to show that

$$\left(\frac{\partial^2 A}{\partial \lambda^2}\right)_{N,V,T} = -\beta \left(\langle (U_1 - U_0)^2 \rangle_{\lambda} - \langle U_1 - U_0 \rangle_{\lambda}^2 \right) \leq 0 \,. \tag{3.17}$$

Since the right-hand side of the equation is the negative of a variance, we can note that the second derivative of A with respect to λ is always non-positive. This implies that

$$\left(\frac{\partial A}{\partial \lambda}\right)_{\lambda=0} \geq \left(\frac{\partial A}{\partial \lambda}\right)_{\lambda>0} \,, \tag{3.18}$$

and hence

$$A_1 \leq A_0 + \langle U_1 - U_0 \rangle_{\lambda=0} \,. \tag{3.19}$$

This variational principle for the free energy is known as the *Gibbs–Bogoliubov inequality*: we compute an upper bound to the free energy of the system of interest from a knowledge of the average of $U_1 - U_0$ evaluated for the reference system. The latter is something that can often be computed because the reference system is by construction a system which we can solve.

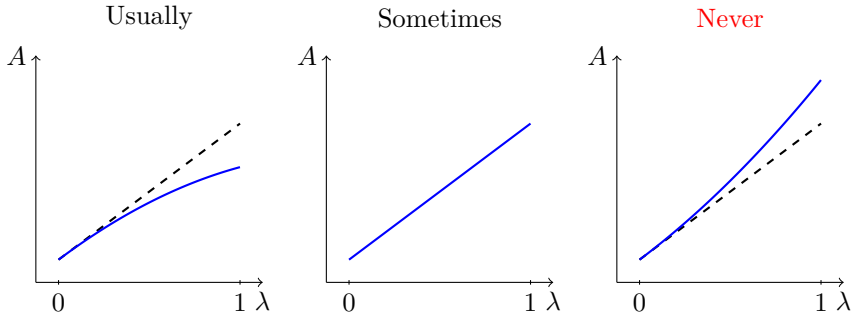


Figure 3.1: The Gibbs–Bogoliubov inequality shows that the first-order perturbation estimation of the free energy is never lower than the true free energy.

We may let our reference system depend on some parameters and use the variational principle to get our best estimation of the upper bound of Helmholtz energy.

It may seem at first glance that we are doing something slightly odd here: what does it actually mean that we compute $\langle U_1 \rangle_{\lambda=0}$ for the reference potential, given that when $\lambda = 0$, the system is

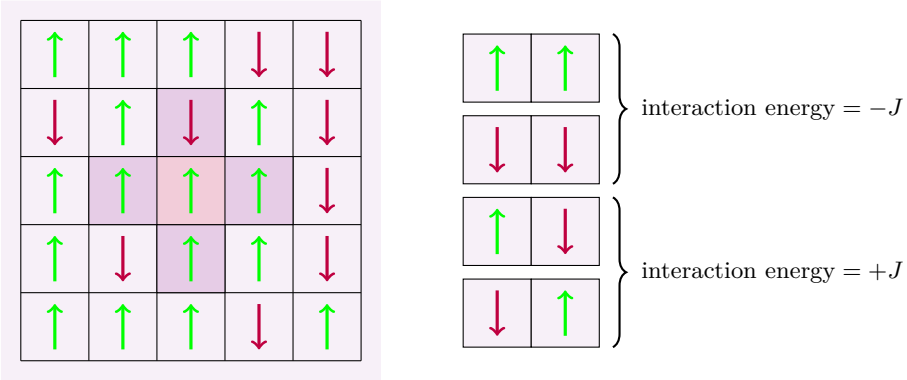


Figure 3.2: A schematic representation of the Ising model in two dimensions.

governed by U_0 only? In this case, U_1 does not contribute to the Boltzmann factor, but that does not mean we are not able to compute what it is for a given microstate. This is similar to computing any other mechanical observable, e.g. the pressure, which also would not appear in the Boltzmann factor. It is evaluated as

$$\langle U_1 - U_0 \rangle_{\lambda=0} = \frac{\int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{r}_i \, (U_1 - U_0) \exp(-\beta U_0)}{\int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{r}_i \, \exp(-\beta U_0)} \,. \tag{3.20}$$

The usefulness of (3.19) depends crucially on the quality of the choice of reference system. A good reference system is not necessarily close in free energy to the system of interest, but one for which the fluctuation in the potential energy difference $U_1 - U_0$ are small. Thermodynamic perturbation theory for simple liquids has been very successfully precise because the structure of the reference system (hard-sphere fluid) and the liquid under consideration (e.g. Lennard-Jones) are very similar. As a result, $\langle U_1 - U_0 \rangle_\lambda$ hardly depends on λ and so its derivative is small.

3.3 Mean-Field Theory and the Ising Model

The Gibbs–Bogoliubov inequality can be used as a starting point from which we can derive mean-field theory, which provides a systematic approximation to the free energy of a many-body system. Let's start with a seemingly simple yet interesting and arguably the most important model in statistical mechanics and related fields — the Ising model. The Ising model describes the magnetism of materials by considering spins on a lattice with Hamiltonian

$$U_1 = -\frac{J}{2} \sum_{i=1}^{N} \sum_{\langle i,j \rangle} s_i s_j - B \sum_{i=1}^{N} s_i \,, \tag{3.21}$$

where $\langle i, j \rangle$ indicates that the particle j is one of the nearest neighbours of i and the factor of $1/2$ ensures that we do not double count the interactions. B is the external magnetic field, which we will now set to be zero. The Ising model is not just relevant for magnetic systems; in appendix section D, we show that a lattice-gas model that can be used to describe the liquid-vapour transition is actually equivalent to the Ising model.

We wish to approximate this model system using a reference system with a much simpler Hamiltonian, namely one that consists only of a sum of one-particle contributions. For example,

$$U_0 = -\sum_{i=1}^{N} h s_i \,, \tag{3.22}$$

where h denotes the effective field that replaces the interaction with the other particles. The molecular partition function counting all states of a single spin (with energy $-hs$) in the reference system is

$$q_0 = \int \mathrm{d}s \, \exp(\beta h s) \tag{3.23}$$

and the free energy per spin is

$$a_0(h) = -k_B T \ln q_0 = -k_B T \ln \int \mathrm{d}s \, \exp(\beta h s) \, . \tag{3.24}$$

Here, we use the shorthand a for A/N. If the spins are quantised and we only have two spin states ($+1$ and -1), the partition function reduces to

$$q_0 = \exp(\beta h) + \exp(-\beta h) = 2\cosh(\beta h) \tag{3.25}$$

and the free energy becomes

$$a_0(h) = -k_B T \ln(2\cosh(\beta h)) \, . \tag{3.26}$$

We can easily compute the average value of s in the reference system,

$$\langle s \rangle_0 = \frac{\int \mathrm{d}s \, s \exp(\beta h s)}{q_0} = \frac{1}{q_0} \frac{\partial}{\partial \beta h} \int \mathrm{d}s \, \exp(\beta h s) = k_B T \frac{\partial \ln q_0}{\partial h} \, , \tag{3.27}$$

or simply

$$\langle s \rangle_0 = -\frac{\partial a_0(h)}{\partial h} \, . \tag{3.28}$$

In the case of spins ± 1,

$$\langle s \rangle_0 = \tanh(\beta h) \, , \tag{3.29}$$

which we could have obtained directly from the definition of the mean magnetisation, but the expression involving the derivative of a_0 will come in useful shortly. Now, we consider the Gibbs–Bogoliubov inequality, where

$$\begin{aligned} a_{\mathrm{MF}} &= a_0 + \left\langle -\frac{J}{2} \sum_{\langle i,j \rangle} s_i s_j + h s_i \right\rangle_0 \\ &= a_0 - \frac{J}{2} z \langle s \rangle_0^2 + h \langle s \rangle_0 \, . \end{aligned} \tag{3.30}$$

In the last line, we have introduced z, the coordination number of particle i. Moreover, we have used the fact that, in the reference system, different spins are uncorrelated, $\langle s_i s_j \rangle_0 = \langle s_i \rangle_0 \langle s_j \rangle_0$. We now look for the optimum value of h, i.e. the one that minimises our estimate of a_{MF}. Carrying out the differentiation with respect to h, we find that

$$\begin{aligned} 0 &= \frac{\partial}{\partial h} \left(a_0 - \frac{J}{2} z \langle s \rangle_0^2 + h \langle s \rangle_0 \right) \\ &= -\langle s \rangle_0 - (Jz \langle s \rangle_0 - h) \frac{\partial \langle s \rangle_0}{\partial h} + \langle s \rangle_0 \\ &= -(Jz \langle s \rangle_0 - h) \frac{\partial \langle s \rangle_0}{\partial h} \, . \end{aligned} \tag{3.31}$$

Since $\frac{\partial \langle s \rangle_0}{\partial h} \neq 0$, we can conclude that

$$h = Jz \langle s \rangle_0 \, . \tag{3.32}$$

Finally, we can insert this expression for h into (3.29) to obtain an implicit equation for $\langle s \rangle_0$,

$$\langle s \rangle_0 = \tanh(\beta J z \langle s \rangle_0) \, , \tag{3.33}$$

which can be solved to yield $\langle s \rangle_0$ as a function of T.

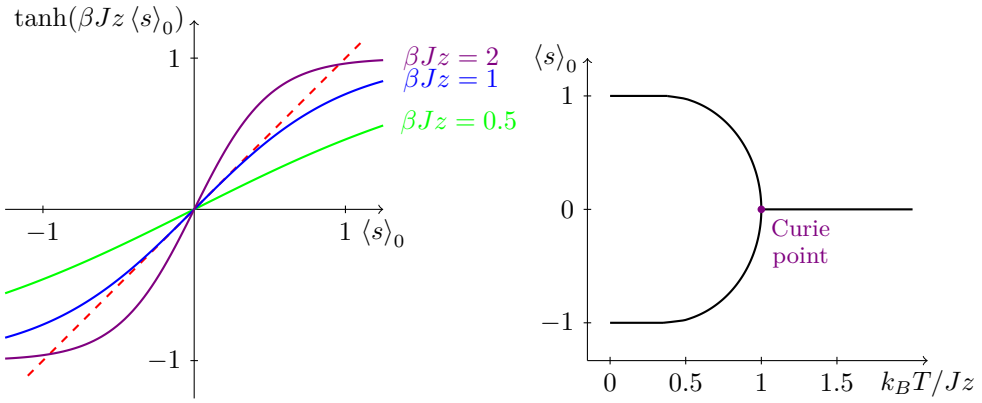


Figure 3.3: (a) A graphical method for the solution of the mean magnetisation in Ising model. (b) The phase diagram of an Ising model system.

Above the *critical* (or *Curie*) *temperature* $k_BT = Jz$, there is only one solution $\langle s\rangle_0 = 0$: at high temperatures, the entropy dominates the free energy and so the spins are randomly distributed. The significance of the entropy as a driving force is diminished as the temperature is decreased, and below the Curie temperature, the net alignment of spins, which is energetically favourable, becomes dominant, and we get two extra solutions with non-zero mean magnetisation. Amongst all the three solutions, the $\langle s\rangle_0 = 0$ solution is actually a maximum rather than a minimum so we will ignore it. In the absence of an external magnetic field, the system is equally likely to be magnetised in the up or the down spin configurations, but the fluctuations causing the transitions between the two are extremely unlikely. The ergodic hypothesis[7] is thus formally violated, and this phenomenon is sometimes described as *spontaneous symmetry breaking*.

The free energy estimate that we obtain when inserting (3.32) into (3.30) is

$$a_{\mathrm{MF}} = a_0 + \frac{J}{2} z \langle s\rangle_0^2 \,. \tag{3.34}$$

The subscript 'MF' in this expression stands for the mean field approximation. It is very important to bear in mind that the free energy that results from the mean-field approximation is in general not simply the free energy of the reference system with the effective field, but has an additional term that depends on the difference between the two potentials measured for the system with the effective field.

3.3.1 Validity of Mean Field Theory

Having solved the Ising model using mean field theory, a question that arises is: are our results correct?

There is actually a version of the Ising model for which the mean field theory is exact: it is the $d = \infty$ dimensional lattice. This is unphysical. Roughly speaking, mean field theory works for large d because each spin has a large number of neighbours and so indeed sees something close to the average spin.

But what about dimensions of interest? Mean field theory gets things most dramatically wrong in $d = 1$. In that case, no phase transition actually occurs. We will solve the 1D Ising model exactly in the appendix section F. There is a general lesson here: in low dimensions, both thermal and quantum fluctuations are more important and invariably stop systems forming ordered phases.

[7] Being *ergodic* means that a point (the state of the system in this case) will eventually visit all parts of the space (the phase space) as it evolves over time, in a uniform and random sense.

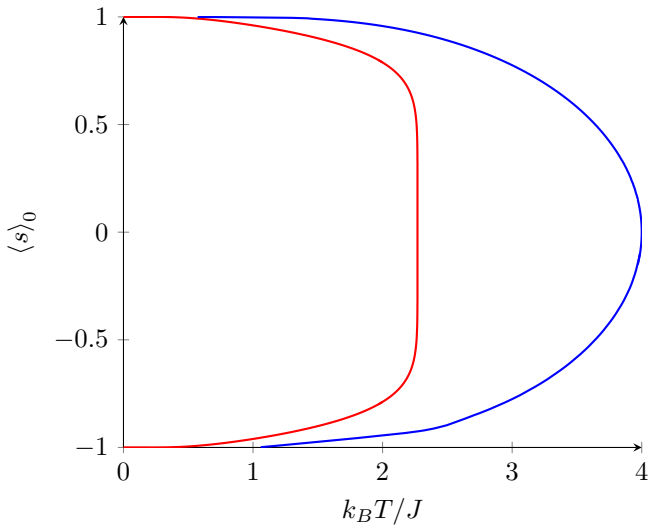


Figure 3.4: Mean magnetisation of the exact (red) and the mean field (blue) 2D square lattice Ising model.

In higher dimensions, $d \geq 2$, the crude features of the phase diagram, including the existence of a phase transition, given by mean field theory are essentially correct. In fact, the very existence of a phase transition is already worthy of comment. The defining feature of a phase transition is behaviour that jumps discontinuously as we vary β. Mathematically, the functions must be non-analytic. Yet all properties of the theory can be extracted from the partition function Q which is a sum of smooth, analytic exponential functions

$$Q = \sum_{|i\rangle} \mathrm{e}^{-\beta E_i} \,. \tag{3.35}$$

How can we get a phase transition? The loophole is that Q is only necessarily analytic if the sum is finite. But there is no such guarantee when the number of lattice sites $N \to \infty$. We reach a conclusion: phase transitions only strictly happen in the thermodynamic limit. There are no phase transitions in finite systems.

The 2D Ising model in a square lattice has been solved exactly, first by Onsager — his method was famously complicated. In appendix section F, we will use a clever trick to obtain the exact result of critical temperature, which is

$$k_B T = \frac{2J}{\ln(\sqrt{2}+1)} \approx 2.269 J \,. \tag{3.36}$$

This is actually not hugely different from the mean field approximation, which predicts

$$k_B T = Jz = 4J \,. \tag{3.37}$$

3.4 Liquid-Gas Phase Transition

Now we will investigate probably the most common phase transition — liquid-gas phase transition. We will take a hard-sphere gas as a reference system and add a very weak, attractive interaction $-\epsilon$ that extends to a very large distance R_c. This crude model leads to the famous van der Waals equation of state.

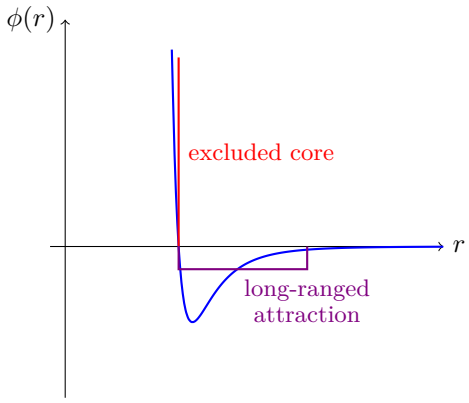


Figure 3.5: A typical interatomic potential (shown in blue) can be approximated by a hard-core repulsion and a very weak attractive interaction $-\epsilon$.

3.4.1 Hard-Sphere Gas

The hard-sphere potential is given by

$$U_0 = \sum_{i<j} \phi(r_{ij}) \,, \tag{3.38}$$

where

$$\phi(r_{ij}) = \begin{cases} 0 & \text{if } r_{ij} > \sigma \\ \infty & \text{if } r_{ij} \leq \sigma \,, \end{cases} \tag{3.39}$$

where σ is the diameter of the particle.

Provided the hard-sphere gas is sufficiently dilute, we can estimate that the addition of each additional hard sphere reduces the volume available to the next one by $v_{\text{ex}} = 4\pi\sigma^3/3$, the volume excluded by a single hard sphere. In the dilute limit, we estimate that no two excluded volumes will overlap. An illustration of this set-up is given in the figure below. In this case, we can simply modify the perfect gas partition function to give

$$\begin{aligned} Q_{\text{HS}} &\approx \frac{1}{\Lambda^{3N} N!} V(V - v_{\text{ex}})(V - 2v_{\text{ex}}) \dots (V - (N-1)v_{\text{ex}}) \\ &= \frac{V^N}{\Lambda^{3N} N!} \left(1 - \frac{v_{\text{ex}}}{V}\right) \left(1 - 2\frac{v_{\text{ex}}}{V}\right) \dots \left(1 - (N-1)\frac{v_{\text{ex}}}{V}\right) . \end{aligned} \tag{3.40}$$

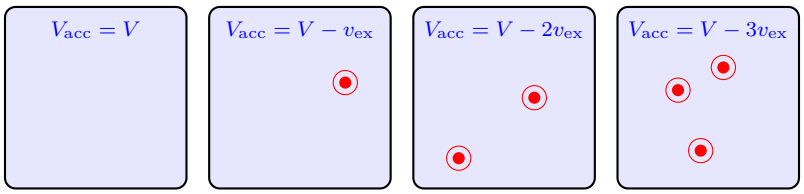


Figure 3.6: Ignoring any volume excluded by the walls and any excluded volume overlaps, each particle inserted into the system excludes a volume v_{ex}, reducing the accessible volume.

In the dilute regime, $(N-1)v_{\mathrm{ex}}/V \ll 1$, so we can use the power series $\exp(-x) = 1 - x$, giving

$$\begin{aligned} Q_{\mathrm{HS}} &= \frac{V^N}{\Lambda^{3N} N!} \exp\left(-\frac{v_{\mathrm{ex}}}{V}\right) \exp\left(-\frac{2v_{\mathrm{ex}}}{V}\right) \dots \exp\left(-\frac{(N-1)v_{\mathrm{ex}}}{V}\right) \\ &= \frac{V^N}{\Lambda^{3N} N!} \exp\left(-\frac{v_{\mathrm{ex}}}{V} \sum_{i=1}^{N-1} i\right) . \end{aligned} \tag{3.41}$$

The summation is an arithmetic series that we can evaluate as $N(N-1)/2$. Since $Nv_{\mathrm{ex}}/2V \ll 1$ by construction, we can ignore the term linear in N in the exponent and write

$$Q_{\mathrm{HS}} \approx \frac{V^N}{\Lambda^{3N} N!} \exp\left(-\frac{N^2 v_{\mathrm{ex}}}{2V}\right) = \frac{V^N}{\Lambda^{3N} N!} \left[\exp\left(-\frac{N v_{\mathrm{ex}}}{2V}\right)\right]^N . \tag{3.42}$$

If we use the same expansion again and if we note that $v_{\mathrm{ex}} = 8v_0$, where v_0 is the volume of a single hard sphere particle, we obtain

$$Q_{\mathrm{HS}} \approx \frac{(V - 4Nv_0)^N}{\Lambda^{3N} N!} . \tag{3.43}$$

Using $A = -k_B T \ln Q$, Stirling's approximation and $P = -(\frac{\partial A}{\partial V})_{N,T}$, we can write

$$P_{\mathrm{HS}} \approx \frac{N k_B T}{V - 4Nv_0} = \frac{k_B T}{v_0} \frac{\phi}{1 - 4\phi} , \tag{3.44}$$

where in the last step we rewrote the volume in terms of the overall fraction of volume occupied, $\phi \coloneqq Nv_0/V$. We note that this function diverges when $\phi \to 1/4$. In reality, hard spheres can be compressed to the close-packed volume fraction $\phi = \pi/\sqrt{18} \approx 0.7405$, and so the approximate partition function above is not appropriate at higher densities. This hard-sphere equation of state, and the corresponding van der Waals equation of state that we will derive from it, are thus approximations that are only reasonable at low density.

3.4.2 Mean-Field Attractions

We have accounted for, in an approximated way, the hard-core repulsion between particles. We would now like to add the weak pairwise attraction $-\epsilon$. In particular, we consider the limit of very weak attractions ($\epsilon \to 0$) that are very long-range ($R_c \to \infty$). In this limit, the potential energy of the perturbed fluid can be computed directly. Within a shell of radius R_c, there will be, on average, $N_c = (4/3)\pi R_c^3 \rho$ other particles, where ρ is the number density of the fluid. All these particles contribute $-\epsilon/2$ to the potential energy of the fluid, with the factor of $1/2$ accounting for the double-counting of pairwise interaction. For $R_c \to \infty$, the number of neighbours N_c also tends to infinity and hence the relative fluctuation in N_c becomes negligible, as it is of order $1/\sqrt{N_c}$. It then follows that the fluctuation in the perturbation (i.e. $U_1 - U_0$) also becomes negligible and, since we have shown that the second derivative of the free energy is proportional to the variance in $U_1 - U_0$, the first-order Maclaurin expansion of the free energy in λ becomes exact. Therefore the Gibbs–Bogoliubov inequality becomes an identity. The free energy per particle of the van der Waals fluid is therefore

$$\left(\frac{A}{N}\right)_{\mathrm{vdW}} (\rho, T) = \left(\frac{A}{N}\right)_{\mathrm{HS}} (\rho, T) - \rho a , \tag{3.45}$$

where we have defined $a = (2\pi/3) R_c^3 \epsilon$. The corresponding pressure is

$$P_{\mathrm{vdW}} = -\left(\frac{\partial A_{\mathrm{vdW}}}{\partial V}\right)_{N,T} = \rho^2 \frac{\partial (A/N)_{\mathrm{vdW}}}{\partial \rho} = P_{\mathrm{HS}}(\rho, T) - a\rho^2 , \tag{3.46}$$

where P_{HS} denotes the pressure of the hard-sphere reference system. We therefore have the well-known van der Waals equation

$$(P_{\mathrm{vdW}} + a\rho^2)(V - Nb) = N k_B T , \tag{3.47}$$

where $b = 4v_0$. This equation is very interesting to study because it is probably the simplest model system which exhibits vapour-liquid coexistence. A simple way of computing the phase diagram involves the use of Maxwell's equal-area construction.

3.4.3 Maxwell Equal-Area Construction

We can sketch the behaviour of the pressure with volume above and below the critical point. Above the critical point, a single fluid phase exists; below the critical point, a vapour (low-density fluid) and a liquid (high-density fluid) are predicted. Below the critical point, there are regions where the equation of state describes a mechanically unstable state where the isothermal compressibility is negative, and where a single pressure corresponds to multiple possible volumes. Such regions are known as van der Waals loops. The critical point itself occurs at the point at which the minimum and the maximum of the van der Waals loop merge, i.e. the first and second derivatives of the pressure with respect to volume vanish.

In regions where there are van der Waals loops, we can immediately find the points corresponding to limits of mechanical stability: these spinodals correspond to the points at which $(\partial P/\partial V)$ changes sign. Finding the points of coexistence (the *binodals*) is slightly more challenging. Various approaches can be used; here, we will use a method called Maxwell construction.

We have already shown that the condition for phase equilibrium — in this case between the liquid and the vapour — is the equality of pressures, temperatures and chemical potentials. From our discussion of the common tangent construction approach, we know that the condition that the chemical potentials are equal is given by

$$N\mu_1 = N\mu_2 = A_1 + P_{\text{comm}}V_1 = A_2 + P_{\text{comm}}V_2\,, \tag{3.48}$$

where P_{comm} is the common pressure. This is fulfilled if $P_{\text{comm}}(V_1 - V_2) = A_1 - A_2$.

We can also express the pressure as $P = -(\partial A/\partial V)_{N,T}$, and therefore, if we use thermodynamic integration, we find that

$$A_1 - A_2 = -\int_0^{V_1} \mathrm{d}V\, P + \int_0^{V_2} \mathrm{d}V\, P\,. \tag{3.49}$$

Hence the condition for coexistence is

$$P_{\text{comm}}(V_2 - V_1) = \int_{V_1}^{V_2} \mathrm{d}V\, P\,. \tag{3.50}$$

This condition means that the area of the rectangle spanned by V_1 and V_2 along the volume axis and 0 and P_{comm} along the pressure axis is equal to the area under the P curve between V_1 and V_2. As can be seen in the figure, the only difference between the two areas is the region where area 1 carves an area into the rectangle and area 2 protrudes from the rectangle: to satisfy (3.50), we simply require that area 1 should equal area 2. In practice, we need to vary P_{comm} until the two areas match. A phase diagram corresponding to the van der Waals equation of state is shown in the figure.

We used an approximate equation of state for a hard-sphere gas in our derivation, and so, as we argued above, the van der Waals equation of state is also only reasonable at low density. If, on the other hand, we use the exact equation of state of hard spheres as deduced from computer simulations, then we can compute the 'exact' equation of state of the van der Waals model (i.e. the equation of state that van der Waals would have given an arm and a leg for). Using this approach, Longuet–Higgins and Widom were the first to compute the true phase diagram of the van der Waals model.

3.4.4 The Critical Point

Let's now return to discuss what happened at the critical point. Mathematically, a critical point corresponds to where

$$\frac{\partial P}{\partial V} = \frac{\partial^2 P}{\partial V^2} = 0\,. \tag{3.51}$$

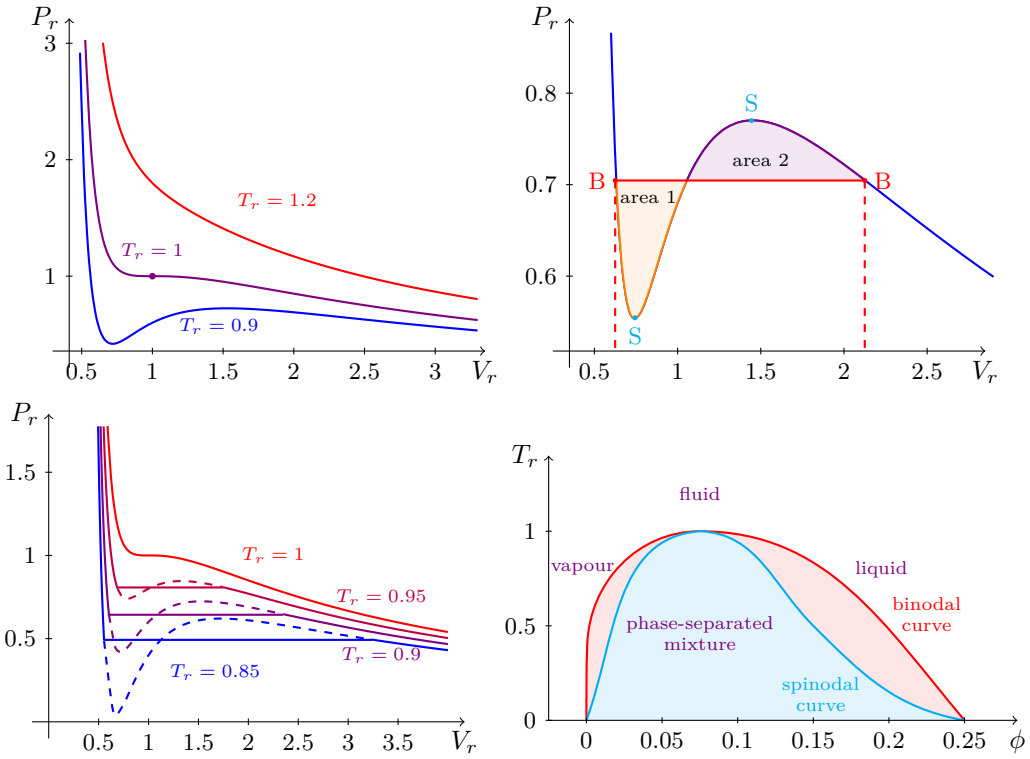


Figure 3.7: (a) The pressure as a function of the volume above, at and below the critical temperature. All quantities are scaled by their values at the critical point. (b) Below the critical point, the pressure has a 'van der Waals loop'. A Maxwell equal-area construction, shown here, is used to find the point of coexistence (labelled 'B'). Spinodals, where the system becomes mechanically unstable, are labelled 'S'. (c) The coexistence region results in constant pressure region. (d) Phase diagram for the van der Waals model determined using Maxwell constructions.

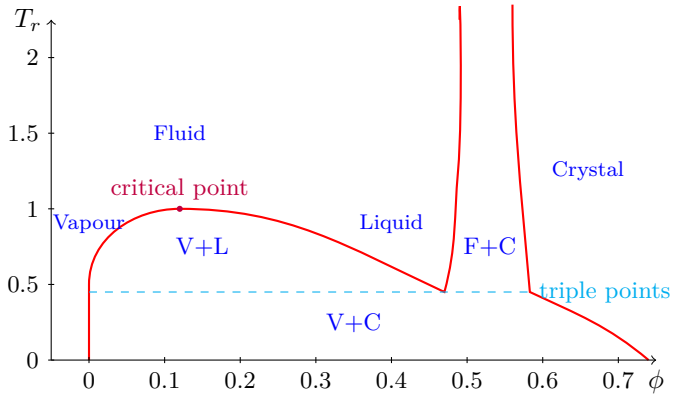


Figure 3.8: Longuet–Higgins–Widom-style phase diagram for the 'exact' van der Waals model. This phase diagram was computed by using the hard-sphere system as a reference system and adding a weak, long-ranged attractive interaction.

Having an expression of the pressure, we can of course solve for this condition by brute force, but there is a slightly more elegant way to find the critical point. We will again use the notation $v = V/N$ to denote the volume per particle to reduce some notational cluttering. We rearrange the van der Waals equation to get a cubic

$$Pv^3 - (Pb + k_B T)v^2 + av - ab = 0\,. \tag{3.52}$$

For $T < T_c$, the equation has three real roots, while for $T > T_c$, there is just one. This means that precisely at the $T = T_c$, the three roots must therefore coincide before two of them move off onto the complex plane. This means that at the critical point, the curve can be rewritten as

$$P_c(v - v_c)^3 = 0\,. \tag{3.53}$$

By comparing the coefficients, we get

$$k_B T_c = \frac{8a}{27b} \quad , \quad v_c = 3b \quad , \quad P_c = \frac{a}{27b^2}\,. \tag{3.54}$$

The Law of Corresponding States Revisited

This time, we can express the temperature, pressure and volume in reduced quantities relative to the critical values

$$T_r \coloneqq \frac{T}{T_c} \quad , \quad v_r \coloneqq \frac{v}{v_c} \quad , \quad p_r \coloneqq \frac{P}{P_c} \tag{3.55}$$

and rewrite the van der Waals equation of state into a form universal to all gases

$$P_r = \frac{8T_r}{3v_r - 1} - \frac{3}{v_r^2}\,. \tag{3.56}$$

This again illustrates the law of corresponding states. Moreover, since the three critical quantities T_c, P_c and V_c are written in just two variables a and b, we can construct a combination of them which is independent of a and b and therefore should be the same for all gases. This is the universal compressibility ratio

$$\frac{P_c v_c}{k_B T_c} = \frac{3}{8}\,. \tag{3.57}$$

Comparing to real gases, for which this value ranges from 0.28 to 0.30, our proposed value is a little high. We shouldn't be too discouraged by this; after all, we knew from the beginning that the van der Waals equation is unlikely to be accurate in the liquid regime where the particle number density is high.

Critical Exponents

We now turn to ask a different question. How do various quantities change as we approach the critical point?

First, we can ask what happens to the difference in (inverse) densities $v_{\rm gas} - v_{\rm liquid}$ as we approach the critical point along the coexistence curve. For $T < T_c$, or equivalently $T_r < 1$, the reduced van der Waals equation has two stable solutions

$$P_r = \frac{8T_r}{3v_{r,\rm gas} - 1} - \frac{3}{v_{r,\rm gas}^2} = \frac{8T_r}{3v_{r,\rm liquid} - 1} - \frac{3}{v_{r,\rm liquid}^2} \,. \tag{3.58}$$

Solving this for T_r, we get

$$T_r = \frac{(3v_{r,\rm liquid} - 1)(3v_{r,\rm gas} - 1)(v_{r,\rm liquid} + v_{r,\rm gas})}{8v_{r,\rm liquid}^2 v_{r,\rm gas}^2} \,. \tag{3.59}$$

Notice that as we approach the critical point, $v_{r,\rm liquid}, v_{r,\rm gas} \to 1$ and the equation above tells us that $T_r \to 1$ as expected. We can see exactly how we approach $T_r = 1$ by expanding the right-hand side for small $\epsilon = v_{r,\rm gas} - v_{r,\rm liquid}$. To do this quickly, it's best to notice that the equation is symmetric in $v_{r,\rm gas}$ and $v_{r,\rm liquid}$, so close to the critical point we can write $v_{r,\rm gas} = 1 + \epsilon/2$ and $v_{r,\rm liquid} = 1 - \epsilon/2$. Substituting this into the equation above and keeping just the leading order term, we find

$$T_r \approx 1 - \frac{1}{16}(v_{r,\rm gas} - v_{r,\rm liquid})^2 \,, \tag{3.60}$$

or rearranging, as we approach T_c,

$$v_{\rm gas} - v_{\rm liquid} \sim (T_c - T)^{1/2} \,. \tag{3.61}$$

This answers our first question.

Our second question is: how does the volume change with pressure as we move along the critical isotherm? It turns out that we can answer this question without doing any work. We know that at the critical point, $\partial P/\partial v = \partial^2 P/\partial v^2 = 0$, so a Taylor expansion around the critical point must start with the cubic term,

$$P - P_c \sim (v - v_c)^3 \,. \tag{3.62}$$

This answers our second question.

Our final question concerns the compressibility, defined as

$$\kappa = -\frac{1}{v}\left(\frac{\partial v}{\partial P}\right)_T \,. \tag{3.63}$$

We want to understand how changes as we approach $T \to T_c$ from above. We already know that at the critical point $\partial P/\partial v = 0$. So expanding for temperatures close to T_c, we expect

$$\left(\frac{\partial P}{\partial v}\right)_{T=T_c} = -a(T - T_c) + \ldots \tag{3.64}$$

This tells us the compressibility should diverge at the critical point, scaling as

$$\kappa \sim (T - T_c)^{-1} \,. \tag{3.65}$$

We now have three answers to three questions:

$$v_{\rm gas} - v_{\rm liquid} \sim (T_c - T)^{\beta} \qquad \beta = \frac{1}{2} \tag{3.66}$$

$$P - P_c \sim (v - v_c)^{\delta} \qquad \delta = 3 \tag{3.67}$$

$$\kappa \sim (T - T_c)^{-\gamma} \qquad \gamma = 1 \,. \tag{3.68}$$

Do they agree with experiment? Remember that we're not sure that we can trust the van der Waals equation at the critical point so we should be nervous. However, there is also reason for some confidence. Notice, in particular, that when computing (3.62) and (3.65), we didn't actually need any details of the van der Waals equation. We simply needed to assume the existence of the critical point and an analytic Taylor expansion of various quantities in the neighbourhood. Given that the answers follow from such general grounds, one may hope that they provide the correct answers for a gas in the neighbourhood of the critical point even though we know that the approximations that went into the van der Waals equation aren't valid there. Fortunately, that isn't the case: the physics is much more interesting than that!

The experimental results for a gas in the neighbourhood of the critical point do share one feature in common with the discussion above: they are completely independent of the atomic make-up of the gas — again illustrating the law of corresponding states. However, the scaling that we computed using the van der Waals equation is not fully accurate. The correct results are as follows:

$$v_{\text{gas}} - v_{\text{liquid}} \sim (T_c - T)^{\beta} \qquad \beta \approx 0.32 \tag{3.69}$$

$$P - P_c \sim (v - v_c)^{\delta} \qquad \delta \approx 4.8 \tag{3.70}$$

$$\kappa \sim (T - T_c)^{-\gamma} \qquad \gamma \approx 1.2\,. \tag{3.71}$$

The quantities β, γ and δ are examples of *critical exponents*. The van der Waals equation provides only a crude first approximation to the critical exponents.

Fluctuations

We see that the van der Waals equation didn't do too badly in capturing the dynamics of an interacting gas. It gets the qualitative behaviour right, but fails on precise quantitative tests. So what went wrong? We mentioned during the derivation of the van der Waals equation that we made certain approximations that are valid only at low density. So perhaps it is not surprising that it fails to get the numbers right near the critical point where $v = 3b$. But there's actually a deeper reason that the van der Waals equation fails: fluctuations.

This is simplest to see in the grand canonical ensemble. You will show in one of the exercises that $\Delta N/N \sim 1/\sqrt{N}$, which essentially allows us to work in a grand canonical ensemble even when we have a fixed particle number under normal conditions. In the context of the liquid-gas transition, fluctuating particle number is the same thing as fluctuating density $\rho = N/V$. You will also show in one of the exercises that

$$\frac{\Delta N^2}{N} = -\frac{1}{\beta V}\left(\frac{\partial \langle N\rangle}{\partial V}\right)_{p,T}\left(\frac{\partial V}{\partial p}\right)_{N,T}\,. \tag{3.72}$$

This relates the fluctuations in the particle number to the compressibility, which is diverging at the critical point. This means that there are large fluctuations in the density of the fluid at this point. The result is that any simple equation of state, like the van der Waals equation, which works only with the average volume, pressure and density will miss this key aspect of the physics.

Understanding how to correctly account for these fluctuations is the subject of critical phenomena, which is unfortunately a far deeper subject. It has close links with the renormalization group and conformal field theory which also arise in fancy places like particle physics and string theory. We will briefly account for this in the appendix section H using Landau–Ginzburg theory.

Universality

Something magical happens if you try to compute the critical exponents of the Ising model under our mean field treatment using the method analogous to above. They are given by

$$m_0 \sim (T_c - T)^{\beta} \qquad \beta = \frac{1}{2} \tag{3.73}$$

$$B \sim m^{\delta} \qquad \delta = 3 \tag{3.74}$$

$$\chi \sim (T - T_c)^{-\gamma} \qquad \gamma = 1\,, \tag{3.75}$$

where $m = \langle s \rangle$ is the mean magnetisation, m_0 is the mean magnetisation at $B = 0$ and κ is the magnetic susceptibility defined as

$$\chi = N \left(\frac{\partial m}{\partial B} \right)_T . \tag{3.76}$$

These critical exponents are exactly the ones that we have calculated for the van der Waals equation of state!

We saw that the mean field approach to the Ising model gave the same critical exponents as the van der Waals equation. They are both wrong, and they are both wrong in the same, complicated, way with regard to the same true answer! Why on earth would a system of spins on a lattice have anything to do with the phase transition between a liquid and gas? It is as if all memory of the microscopic physics — the type of particles, the nature of the interactions — has been lost at the critical point. And that's exactly what happens.

What we're seeing here is evidence for *universality*. There is a single theory which describes the physics at the critical point of the liquid gas transition, the 3D Ising model and many other systems. This is a theoretician's dream! We spend a great deal of time trying to throw away the messy details of a system to focus on the elegant essentials. But, at a critical point, Nature does this for us! Although critical points in two dimensions are well understood, there is still much that we don't know about critical points in three dimensions.

3.5 Widom's Particle Insertion Method

We all have an intuitive idea of what most thermodynamic variables are, like temperature, pressure, volume *etc.*, but the chemical potential μ still seems a little bit mysterious. Here, we will introduce an idea proposed by Benjamin Widom, which would help us to understand chemical potential in an intuitive way, and calculate it easily.

Consider a one-component system of N particles, with Helmholtz energy $A = -k_B T \ln Q(N, V, T)$. Then for sufficiently large N, we can effectively treat the number of particles as a continuum. The chemical potential is given by

$$\begin{aligned} \mu = \left(\frac{\partial A}{\partial N} \right)_{V,T} &= -k_B T \lim_{\Delta N \to 0} \frac{\ln Q(N + \Delta N, V, T) - \ln Q(N, V, T)}{\Delta N} \\ &\approx -k_B T \ln \left(\frac{Q(N+1, V, T)}{Q(N, V, T)} \right) . \end{aligned} \tag{3.77}$$

$\Delta N = 1$ is the smallest variation we can do. This should hold when N is large and we can effectively treat $\Delta N = 1$ as an "infinitesimal" change.

For an ideal gas, $Q(N, V, T) = V^N / \Lambda^{3N} N!$, and so using (3.77), the chemical potential is

$$\mu^{\mathrm{id}} = -k_B T \ln \left(\frac{V}{\Lambda^3 (N+1)} \right) . \tag{3.78}$$

In the $N \to \infty$ limit, this agrees with a perhaps more familiar form of the chemical potential of an ideal gas

$$\mu^{\mathrm{id}} = k_B T \ln(\Lambda^3 \rho)\,, \tag{3.79}$$

where $\rho = N/V$ is the particle number density. Then for a non-ideal gas system, we may define the *excess chemical potential* as

$$\begin{aligned}\mu^{\mathrm{ex}}(N/V,T) &\coloneqq \mu(N/V,T) - \mu^{\mathrm{id}}(N/V,T) \\ &= -k_B T \ln\left(\frac{\int \prod_{j=1}^{N+1} \mathrm{d}^3\mathbf{r}_j \, \exp[-\beta U(\{\mathbf{r}_i\}_{i=1}^{N+1})]}{V \int \prod_{j=1}^{N} \mathrm{d}^3\mathbf{r}_j \, \exp[-\beta U(\{\mathbf{r}_i\}_{i=1}^{N})]}\right).\end{aligned} \tag{3.80}$$

We separate the potential energy of the $N+1$-particle system as a sum of the N-particle system and the interaction energy of the $N+1^{\mathrm{th}}$ particle with the rest of the system, so that $U(\{\mathbf{r}_i\}_{i=1}^{N+1}) = U(\{\mathbf{r}_i\}_{i=1}^{N}) + \Delta U_{N,N+1}$. We can then rewrite

$$\begin{aligned}\mu^{\mathrm{ex}}(N/V,T) &= -k_B T \ln\left(\frac{\int \mathrm{d}^3\mathbf{r}_{N+1} \int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{r}_i \, \exp[-\beta U(\{\mathbf{r}_i\}_{i=1}^{N})] \exp[-\beta \Delta U_{N,N+1}]}{V \int \prod_{j=1}^{N} \mathrm{d}^3\mathbf{r}_j \, \exp[-\beta U(\{\mathbf{r}_i\}_{i=1}^{N})]}\right) \\ &= -k_B T \ln\left(\frac{1}{V}\int \mathrm{d}^3\mathbf{r}_{N+1} \, \langle \exp(-\beta \Delta U_{N,N+1})\rangle_N\right).\end{aligned} \tag{3.81}$$

Here $\langle - \rangle_N$ denote the canonical ensemble average over the configuration space of the N-particles. In other words, the excess chemical potential is related to the average of $\langle \exp(-\beta \Delta U_{N,N+1})\rangle_N$ over all possible positions of the particle $N+1$. In a translationally invariant system, such as a liquid or a gas, this quantity should not depend $\mathbf{r}_{N+1}$, so we may write

$$\mu^{\mathrm{ex}}(N/V,T) = -k_B T \ln \langle \exp(-\beta \Delta U_{N,N+1})\rangle_N\,. \tag{3.82}$$

This is often known as the particle insertion method because it relates the excess chemical potential to the average of the Boltzmann factor $\exp(-\beta \Delta U_{N,N+1})$ associated with the random insertion of an additional particle in the system where N particles are already present.

3.5.1 Excess Chemical Potential of a Hard-Sphere Gas

As an example, let's calculate the excess chemical potential of a hard-sphere gas system. A trial particle is inserted into the system — it has a probability P to not overlap with the already-existing N particles, and probability $1-P$ to overlap with some of the particles. By the definition of a hard sphere potential, if there is no overlapping of particles, $\Delta U_{N,N+1} = 0$ so $\exp(-\beta \Delta U_{N,N+1}) = 1$; if the particles overlap, $\Delta U_{N,N+1} = \infty$ so $\exp(-\beta \Delta U_{N,N+1}) = 0$. Hence we have the average

$$\langle \exp(-\beta \Delta U_{N,N+1})\rangle = P \times 1 + (1-P) \times 0 = P\,. \tag{3.83}$$

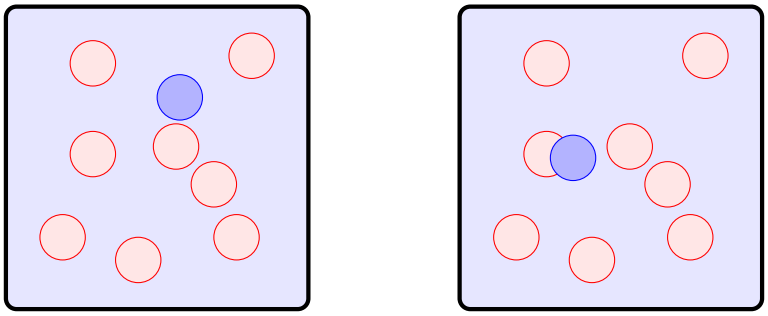
The excess chemical potential is then

$$\mu^{\mathrm{ex}} = -k_B T \ln P\,. \tag{3.84}$$

3.6 Virial Expansion for Classical Imperfect Gases

In the limit of dilute gas, the mean distance between the nearest-neighbour atoms greatly exceeds the range of the interatomic potential $\phi(r)$. Under these conditions, the particles do not interact with each other except during very rare binary collisions, so we can treat them as ideal:

$$Z \coloneqq \frac{PV}{N k_B T} = 1 \tag{3.85}$$



$\Delta U_{N,N+1} = 0$
$\exp(-\beta \Delta U_{N,N+1}) = 1$

$\Delta U_{N,N+1} = \infty$
$\exp(-\beta \Delta U_{N,N+1}) = 0$

Figure 3.9: A sample configuration of a hard sphere system.

when $\rho^* = (N\sigma^3/V) \ll 1$. The probability of finding two atoms within a volume of the order of the atomic volume σ^3 is of the order of $(\rho^*)^2$, and hence we expect deviation from ideal gas compressibility (3.85) as ρ^* increases. It therefore seems that we can express the compressibility factor of a non-ideal gas as a power series in the particle number density

$$Z \coloneqq \frac{P}{\rho k_B T} = 1 + B_2(T)\rho + B_3(T)\rho^2 + \dots \tag{3.86}$$

This is known as the *virial expansion* of the compressibility factor, and the coefficients $B_n(T)$ are known as the n^{th} *virial coefficient*.[8]

There is a traditional way in statistical mechanics to calculate these virial coefficients known as the *cluster expansion*. This method is a bit mathematically challenging, and it gets more and more complicated when calculating higher orders of virial coefficients — but it is a really elegant method, and it has important implications in some other related fields like perturbation theory or even quantum field theory. It is a shame not to mention it, so I put it in the appendix section E. However, Widom's particle insertion method provides us with an alternative easier way to calculate the second virial coefficient.[9]

First, we want to get an expression for μ^{ex}. From the Gibbs–Duhem relation $V\mathrm{d}P = N\mathrm{d}\mu$, we get

$$\frac{\partial P}{\partial \rho} = \rho \frac{\partial \mu}{\partial \rho}\,. \tag{3.87}$$

As $\mu^{\text{id}} = k_B T \ln(\rho\Lambda^3)$,

$$\frac{\rho}{k_B T}\frac{\partial \mu^{\text{id}}}{\partial \rho} = 1\,. \tag{3.88}$$

If we differentiate the virial expansion with respect to ρ, we get

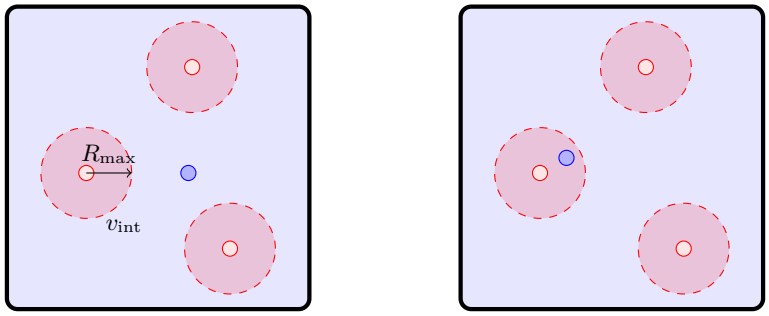
$$\frac{1}{k_B T}\frac{\partial P}{\partial \rho} = 1 + 2B_2(T)\rho + 3B_3(T)\rho^2 + \dots \tag{3.89}$$

From (3.87), we can rewrite

$$\begin{aligned}\frac{\rho}{k_B T}\frac{\partial \mu}{\partial \rho} &= 1 + 2B_2(T)\rho + 3B_3(T)\rho^2 + \dots \\ &= \frac{\rho}{k_B T}\frac{\partial \mu^{\text{id}}}{\partial \rho} + 2B_2(T)\rho + 3B_3(T)\rho^2 + \dots\end{aligned} \tag{3.90}$$

[8] Technically, this expansion is only valid if the interatomic potential decays sufficiently quickly. We will show this in appendix section E.

[9] This method will encounter a whole lot of complications when we try to calculate higher virial coefficients, such as overlap of interaction zones *etc.* If we are interested in them, the systematic cluster expansion is a much better choice.



Probability $= \dfrac{V - V_{\text{int}}}{V}$ Probability $= \dfrac{V_{\text{int}}}{V}$

Figure 3.10: Widom insertion method to determine the excess chemical potential of a dilute non-ideal gas system.

where we observed that 1 corresponds to the derivative of the ideal gas chemical potential (3.88), so we can naturally relate the rest of the virial expansion with the derivative of the excess chemical potential

$$\frac{\rho}{k_B T}\frac{\partial \mu^{\text{ex}}}{\partial \rho} = 2B_2(T)\rho + 3B_3(T)\rho^2 + \dots \tag{3.91}$$

Dividing this equation by ρ and integrating from 0 to ρ, we obtain

$$\mu^{\text{ex}} = k_B T \sum_{n=2}^{\infty} \frac{n}{n-1} B_n(T)\rho^{n-1} = k_B T[2B_2\rho + \dots]\,. \tag{3.92}$$

We can work out the expansion of the excess chemical potential to obtain the second virial coefficient.

We then use Widom method (3.82) to work out μ^{ex}. In the regime of extremely dilute gas ($\rho \to 0$), we may assume an inserted particle will either have significant interaction with only one of the already-present N particles, or it will interact significantly with none of them, where we assumed that the interaction between particles will be negligible beyond some finite distance $R_{\max}$. This corresponds to a particle number density $\rho \ll R_{\max}^{-3}$, and our end result should be independent of $R_{\max}$. We may also assume that the interaction volumes $v_{\text{int}} = \frac{4}{3}\pi R_{\max}^3$ of the particles do not overlap. Then the total interaction volume of the N particles is

$$V_{\text{int}} = N v_{\text{int}} = \frac{4}{3} N \pi R_{\max}^3\,. \tag{3.93}$$

Now, if we insert a particle randomly, it has probability $(V - V_{\text{int}})/V$ to not fall into the interaction zone with any of the other particles, giving $\Delta U = 0$. With a probability V_{int}/V, the insertion will happen in one of the interaction zones, and the average value of $\exp(-\beta\Delta U)$ is

$$\frac{1}{v_{\text{int}}}\int_{v_{\text{int}}} \mathrm{d}^3\mathbf{r}\, \exp[-\beta\phi(\mathbf{r})]\,. \tag{3.94}$$

Therefore, we have the average

$$\begin{aligned}\langle \exp(-\beta\Delta U_{N,N+1})\rangle_N &= \left(\frac{V - V_{\text{int}}}{V}\right) + \left(\frac{V_{\text{int}}}{V}\right) \times \frac{1}{v_{\text{int}}}\int_{v_{\text{int}}} \mathrm{d}^3\mathbf{r}\, \exp[-\beta\phi(\mathbf{r})] \\ &= 1 + \rho\left(-v_{\text{int}} + \int_{v_{\text{int}}} \mathrm{d}^3\mathbf{r}\, \exp[-\beta\phi(\mathbf{r})]\right)\,.\end{aligned} \tag{3.95}$$

Since $v_{\text{int}} = \int_{v_{\text{int}}} \mathrm{d}^3\mathbf{r}$, we can write this as

$$\langle \exp(-\beta\Delta U_{N,N+1})\rangle_N = 1 + \rho\int_{v_{\text{int}}} \mathrm{d}^3\mathbf{r}\, [\exp(-\beta\phi(\mathbf{r})) - 1]\,. \tag{3.96}$$

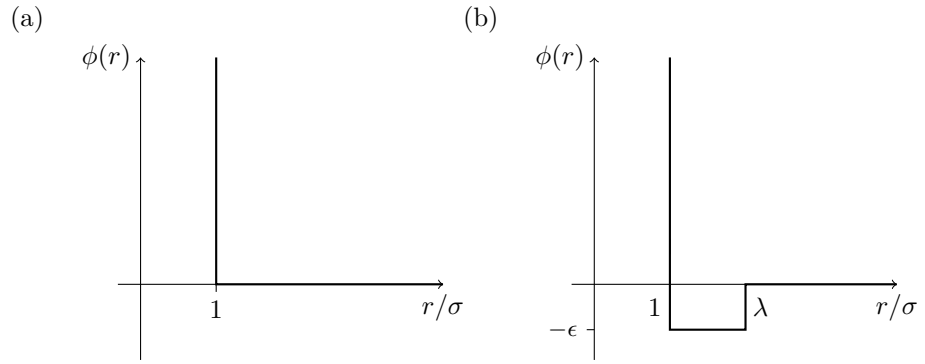


Figure 3.11: Two pair potentials: (a) Hard-sphere potential (b) Square-well potential.

By the Widom formula (3.82), we have

$$\mu^{\text{ex}} = -k_B T \ln\left(1 + \rho \int_{v_{\text{int}}} \mathrm{d}^3\mathbf{r}\,[\exp(-\beta\phi(\mathbf{r})) - 1]\right). \tag{3.97}$$

By construction, $\phi(\mathbf{r}) \approx 0$ outside the interaction zone, so we can change the limits of the integral to the whole space. Finally, as $\rho v_{\text{int}} \ll 1$, we can use Taylor expansion $\ln(1+x) \approx x$ to expand it to the first order

$$\mu^{\text{ex}} \approx k_B T \rho \int \mathrm{d}^3\mathbf{r}\,[1 - \exp(-\beta\phi(\mathbf{r}))]\,. \tag{3.98}$$

This is the term linear in ρ in the density expansion of μ^{ex}. Comparing with our expansion of μ^{ex}, we identify the second virial coefficient

$$B_2(T) = \frac{1}{2}\int \mathrm{d}^3\mathbf{r}\,[1 - \exp(-\beta\phi(\mathbf{r}))]\,. \tag{3.99}$$

If the potential is spherically symmetric, $\phi(\mathbf{r}) = \phi(r)$, we may write

$$B_2(T) = 2\pi \int \mathrm{d}r\, r^2 [1 - \exp(-\beta\phi(r))]\,. \tag{3.100}$$

The quantity $1 - \exp(-\beta\phi(\mathbf{r}))$ occurring repeatedly in our expression is called the *Mayer f function* and is usually denoted as $f(r)$. It seems to just somehow randomly emerge in our derivation, but it actually has a significant importance in the canonical derivation of virial coefficients — see appendix section E.

Let's calculate $B_2(T)$ for two simple interatomic potentials. First, let's consider the hard-sphere potential

$$\phi_{\text{HS}}(r) = \begin{cases} \infty & \text{if } r < \sigma \\ 0 & \text{if } r \geq \sigma\,, \end{cases} \tag{3.101}$$

and so $f(r) = -1$ if $r < \sigma$ and $f(r) = 0$ otherwise. Thus

$$B_2^{\text{HS}}(T) = \frac{2\pi\sigma^3}{3}\,. \tag{3.102}$$

It is a half of the excluded volume around a sphere. In this case, B_2 is independent of temperature, and hence such a system is called *athermal*. It reflects the fact that there is no energy scale in the hard-sphere potential — no matter how high the temperature, the entropy never wins the infinite energy when particles overlap. The corresponding correction in Z (or equivalently P) from ideal gas law is positive, as the hard-sphere potential is purely repulsive.

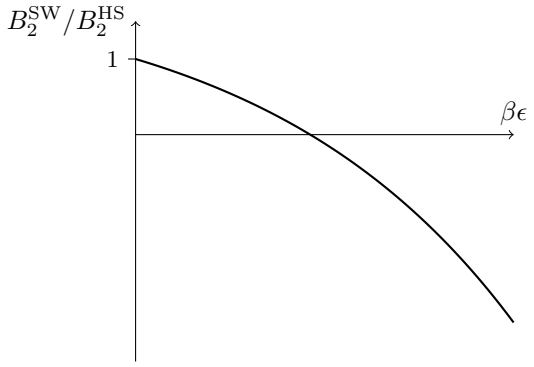


Figure 3.12: A plot of $B_2(T)$ against $\beta\epsilon$. The temperature at which $B_2(T) = 0$ is the Boyle temperature.

Next, consider the square-well potential

$$\phi_{\mathrm{SW}}(r) = \begin{cases} \infty & \text{if } r < \sigma \\ -\epsilon & \text{if } \sigma \leq r < \lambda\sigma \\ 0 & \text{if } r \geq \lambda\sigma\,. \end{cases} \tag{3.103}$$

Here, ϵ parameterises the depth of the well and λ parameterises its width. In this case,

$$B_2^{\mathrm{SW}}(T) = \frac{2\pi\sigma^3}{3}\left[1 - (\lambda^3 - 1)(\exp(\beta\epsilon) - 1)\right]\,. \tag{3.104}$$

This is plotted in the figure below. $B_2(T)$ can be both positive and negative — in low T, $B_2 < 0$ because attraction dominates, whilst at high T, $k_BT \gg \epsilon$ so the attractive part of the potential is unimportant, while the thermal energy still can't break the infinitely high potential barrier of the hard particle core, so the hard-sphere limit is recovered. The temperature at which $B_2(T) = 0$ is called the *Boyle temperature*, because at this temperature, the system behaves very close to being ideal (at least to the first order in ρ).

3.7 Flory–Huggins Theory of Polymer Solutions

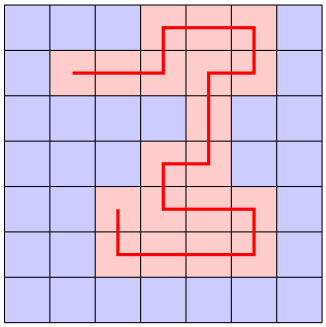
Let's consider polymers living on a lattice. Suppose we have a cubic lattice of M sites in total, and we have N monodisperse (of the same size) polymer molecules, each comprises r segments. Each lattice site can either be occupied by a single polymer segment, or it will be taken up by the solvent. The polymer chain is connected by covalent bonds, which we are never going to break in any cases, so we may safely ignore their contributions to the energy of our system — it is just an additive constant. We further assume that neighbouring polymer segments that are not covalently bonded has an interaction energy ϵ_{PP}; neighbouring solvent molecules interact with energy ϵ_{SS}; neighbouring solvent-polymer segment pair interact with ϵ_{SP}.

We will again use mean field approximation and calculate the excess chemical potential using Widom's insertion method. If we replace r solvent sites by a polymer chain of length r, then

$$\mu^{\mathrm{ex}} = -k_BT \ln\langle\exp(-\beta\Delta U)\rangle\,. \tag{3.105}$$

First, let's calculate the probability that the inserted polymer does not overlap with the already-existing N polymer. If we denote the fractional occupation by

$$\eta = \frac{Nr}{M}\,, \tag{3.106}$$



then the probability of the successful insertion of the first segment without overlapping is given by $1-\eta$. We then encounter some problems. The insertion of successive segments in a polymer chain is correlated — if the first segment is inserted near a polymer chain, then the probability of overlapping when inserting the second segment is very high, since the second segment must be in a neighbouring site of the first segment. On the other hand, if the first segment inserted is far from any other polymers, then the second segment has zero probability of overlapping. To make some progress, we will make a drastic simplification that the probabilities for the successful insertion of subsequent segments are uncorrelated, i.e. the inserted segments are not connected.[10] Then the probability that a single chain can be inserted without overlaps is $(1-\eta)^r$.

Next, we compute the average energy change of a successful insertion. Again, to simplify our calculation, we have to assume that the inserted segments are decoupled. Moreover, we further assume that the existing polymers are also distributed randomly in the lattice sites. Therefore, for any particular inserted segment, the probability of any particular neighbour being occupied by an existing polymer segment is η, and the probability of it being a solvent is $1-\eta$. Therefore, the average change in energy of any nearest neighbour pair is

$$\begin{aligned}\Delta\epsilon_{\text{nn}} &= \eta(\epsilon_{\text{PP}} - \epsilon_{\text{SP}}) + (1-\eta)(\epsilon_{\text{SP}} - \epsilon_{\text{SS}}) \\ &= \eta(\epsilon_{\text{PP}} + \epsilon_{\text{SS}} - 2\epsilon_{\text{SP}}) + (\epsilon_{\text{SP}} - \epsilon_{\text{SS}})\,. \end{aligned} \tag{3.107}$$

If we denote

$$\Delta\epsilon \coloneqq \epsilon_{\text{SP}} - \frac{\epsilon_{\text{PP}} + \epsilon_{\text{SS}}}{2} \tag{3.108}$$

$$\delta\epsilon \coloneqq \epsilon_{\text{SP}} - \epsilon_{\text{SS}}\,, \tag{3.109}$$

then the average total change in energy of inserting a polymer of r segments is

$$\begin{aligned}\Delta U &= rz\Delta\epsilon_{\text{nn}} \\ &= -2rz\eta\Delta\epsilon + rz\delta\epsilon\,, \end{aligned} \tag{3.110}$$

where z is the average number of non-bonded nearest neighbours of a segment.

Now we can estimate μ^{ex} using Widom's formula, which gives

$$\mu^{\text{ex}} = rk_BT\ln(1-\eta) - 2rz\eta\Delta\epsilon + rz\delta\epsilon\,. \tag{3.111}$$

The ideal contribution to the chemical potential of N polymers is

$$\mu^{\text{id}} = k_BT\ln\frac{N}{M} = k_BT\ln\eta - k_BT\ln r\,. \tag{3.112}$$

We add these two terms together and throw away all the constant terms independent of η (since they don't affect the phase behaviour), then the mean field chemical potential of the polymer solution is

$$\mu = k_BT\ln\eta - rk_BT\ln(1-\eta) - 2rz\eta\Delta\epsilon + \text{const.} \tag{3.113}$$

[10] As you can tell, this is a horribly drastic assumption. If our model goes wrong compared with experiments, this is likely to be the main culprit (but rather surprisingly, our model actually agrees with the experiments astonishingly well!).

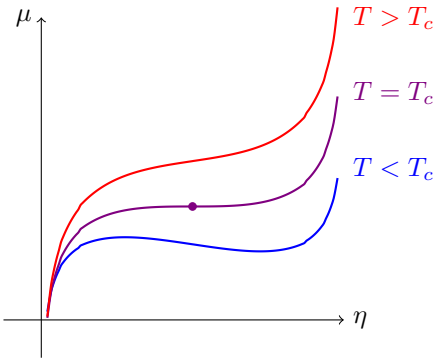


Figure 3.13: The Flory–Huggins chemical potential. Phase separation may occur when $T < T_c$.

It is common to define the *Flory–Huggins parameter* χ by

$$\chi \coloneqq \frac{z\Delta\epsilon}{k_B T}\,, \tag{3.114}$$

then we get

$$\mu = k_B T(\ln\eta - r\ln(1-\eta) - 2r\chi\eta)\,, \tag{3.115}$$

where we have omitted the constant term. This is the *Flory–Huggins expression* for the chemical potential of a polymer solution.

3.7.1 Flory–Huggins Critical Point

It is experimentally observed that a polymer solution shows phase behaviour, where below a certain critical temperature, a polymer solution with some η values will separate into phases. Let's rationalise this behaviour using the chemical potential we calculated.

In order for a phase separation to occur, there must be two phases with different η values with the same chemical potential.[11] Therefore, a plot of μ versus η must have a van der Waals loop (see figure 3.7). As we increase the temperature, this loop will disappear when the maximum and the minimum of the μ-η curve merge. Hence, to find the critical point, we only need to find the point where the equation

$$\begin{aligned} \frac{\partial\beta\mu}{\partial\eta} &= \frac{1}{\eta} + \frac{r}{1-\eta} - 2r\chi = 0 \\ \implies \quad 2r\chi\eta^2 - 2r\left(\chi - \frac{1}{2} + \frac{1}{2r}\right)\eta + 1 &= 0 \end{aligned} \tag{3.116}$$

has a double root.[12] This happens at the point

$$\eta_{\text{crit}} = \frac{r\left(\chi_{\text{crit}} - \frac{1}{2} + \frac{1}{2r}\right)}{2r\chi_{\text{crit}}} \tag{3.118}$$

when the discriminant vanishes:

$$r^2\left(\chi_{\text{crit}} - \frac{1}{2} + \frac{1}{2r}\right)^2 = 2r\chi_{\text{crit}}\,. \tag{3.119}$$

[11] This is a necessary condition, but it is not sufficient — the pressure must also be equal.

[12] When dealing with the van der Waals gas, we identified the critical point to be the point at which both the first and the second derivative of A with respect to V vanishes. One can show that this is equivalent to the condition here once we identify V with M as the lattice analogue, by showing

$$\left(\frac{\partial P}{\partial M}\right)_T \propto \left(\frac{\partial\mu}{\partial\eta}\right)_T \tag{3.117}$$

using Gibbs–Duhem relation.

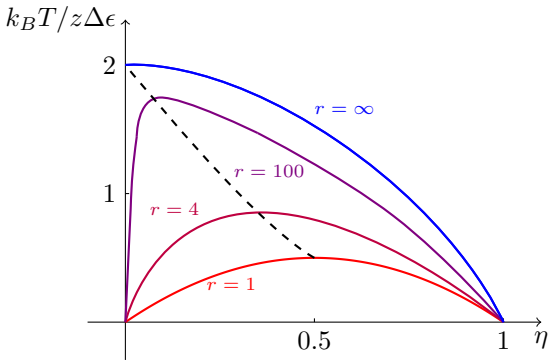


Figure 3.14: The Flory–Huggins phase diagram.

Solving these conditions gives

$$\chi_{\text{crit}} = \frac{(1+\sqrt{r})^2}{2r} \tag{3.120}$$

$$\eta_{\text{crit}} = \frac{1}{1+\sqrt{r}}\,. \tag{3.121}$$

For $r = 1$, $\eta_{\text{crit}} = 0.5$ and $\chi_{\text{crit}} = 2$. The μ-η diagram is symmetric, and this is in fact identical to the regular solution model which you will investigate in one of the exercises. As r increases, η_{crit} decreases as the μ-η diagram becomes more and more asymmetric, and χ_{crit} also decreases, so the critical temperature T_{crit} increases. In the limit of large polymers, $\lim_{r\to\infty}\eta_{\text{crit}} = 0$ and $\lim_{r\to\infty}\chi_{\text{crit}} = \frac{1}{2}$.

What leads to the asymmetry of the phase diagram at non-unity r? We can investigate the energy and entropic change of mixing explicitly. Consider a fully mixed system with Nr polymer segments and $M - Nr$ solvent particles. The potential energy contributed by each polymer segment is $U_{\text{P}} = z\eta\epsilon_{\text{PP}} + z(1-\eta)\epsilon_{\text{SP}}$, and the energy contributed by each solvent particle is $U_{\text{S}} = z\eta\epsilon_{\text{SP}} + z(1-\eta)\epsilon_{\text{SS}}$. If the polymers and the solvent are completely separated, then $U_{\text{P}}^{\text{pre}} = z\epsilon_{\text{PP}}$ and $U_{\text{S}}^{\text{pre}} = z\epsilon_{\text{SS}}$ respectively. Hence, the energy change of mixing is

$$\begin{aligned}\Delta_{\text{mix}}U &= \frac{1}{2}[Nr(U_{\text{P}} - U_{\text{P}}^{\text{pre}}) + (M - Nr)(U_{\text{S}} - U_{\text{S}}^{\text{pre}})] \\ &= Mz\Delta\epsilon\eta(1-\eta) = Mk_BT\chi\eta(1-\eta)\,.\end{aligned} \tag{3.122}$$

This is symmetric in η. If N_1 particles of volume V_1 and N_2 particles of volume V_2 are mixed, forming a mixture of volume $V = V_1 + V_2$, then the entropy change of mixing is

$$\Delta_{\text{mix}}S/k_B = N_1 \ln\left(\frac{V}{V_1}\right) + N_2 \ln\left(\frac{V}{V_2}\right)\,. \tag{3.123}$$

In our case, we have $N_1 = V_1 = M - Nr$ solvent particles and $N_2 = V_2 = Nr$ polymer molecules. This gives

$$\begin{aligned}\Delta_{\text{mix}}S/k_B &= (M - Nr)\ln\frac{M}{M-Nr} + N\ln\frac{M}{Nr} \\ &= -M\left[(1-\eta)\ln(1-\eta) + \frac{\eta}{r}\ln\eta\right]\,.\end{aligned} \tag{3.124}$$

We can see that this is asymmetric in η.

Poor solvent $\chi > \frac{1}{2}$ | θ solvent $\chi = \frac{1}{2}$ | Good solvent $\chi < \frac{1}{2}$

Figure 3.15: The effect of solvent quality on the behaviour of polymers in solution.

3.7.2 Flory–Huggins Parameter

The Flory–Huggins parameter χ actually has more physical meanings: it tells us whether a solvent is good or not for a given polymer. Recall the virial series for the excess chemical potential

$$\beta\mu^{\text{ex}} = 2B_2\rho + \frac{3}{2}B_3\rho^2 + \dots \tag{3.125}$$

Using $\rho = N/M$ and $\eta = Nr/M$, we can rewrite the expansion in terms of η

$$\beta\mu^{\text{ex}} = \frac{2B_2}{r}\eta + \dots \tag{3.126}$$

The virial expansion applies in the dilute regime, i.e. at small η, and so we can use Taylor expansion $\ln(1-x) \approx -x$ in the excess chemical potential (3.111) to write

$$\beta\mu^{\text{ex}} \approx r\eta + 2r\eta\chi\,. \tag{3.127}$$

Comparing the coefficients of the two expansions, we can get the second virial coefficient

$$B_2 = r^2\left[\frac{1}{2} - \chi\right]\,. \tag{3.128}$$

We find that when $\chi = \frac{1}{2}$, $B_2 = 0$ and the ideal behaviour is recovered (at least to the leading order). At this point, the energetic and entropic contributions cancel out exactly, and the polymer behaves like a freely joined chain (see later chapters). The conditions under which this is achieved are known as the θ conditions (just like the Boyle temperature for non-ideal gas). By contrast, if the energetic penalty is smaller than the entropic gain, $\chi < \frac{1}{2}$, the entropic term dominates and the second virial coefficient is positive. The polymer segments repel each other and maximise the mixing. This increases the osmotic pressure of the polymer solution, and we say that the solvent is a *good solvent*. By contrast, if $\chi > \frac{1}{2}$, then $B_2 < 0$. The polymer segments attract each other, and the solvent is said to be a *poor solvent*.

3.8 Landau Theory of Phase Transitions

We have spent a long time investigating phase transitions for different systems case by case. But we have seen that the van der Waals equation and mean field Ising model gave the same wrong answers for the critical exponents with regard to the same true answer. This universality suggests that there should be a unified way to look at phase transitions. Such a method was developed by Lev Landau. It is worth stressing that the Landau approach to phase transitions often only gives qualitatively correct results. However, its advantage is that it is extremely straightforward and easy.

3.8.1 Order of Phase Transitions

Let's first introduce a classification of phase transitions. As we can see in previous sections, free energy plays an important role in phase transition, and hence a natural classification, introduced by Paul Ehrenfest, is based on how the free energy behaves at the critical point. The free energy itself of the two phases must be equal at the phase transition, so the free energy must be continuous at the critical point as a function of temperature or volume/pressure *etc.* — but its derivatives do not have to. If the first derivative of the free energy with respect to pressure/volume and temperature are discontinuous, then the transition is called a *first order* phase transition. If the first derivatives are continuous, but the second ones are not, then the phase transition is called a *second order* phase transition, and so on. There are rarely phase transitions above third order.[13]

The reason we are interested in this is that, a first order phase transition behaves very differently from a second order one. For example, if the first derivatives of the Gibbs free energy of some system are discontinuous, then there will be a volume change and a non-zero enthalpy of phase transition (which is the latent heat that you are familiar with)

$$\left(\frac{\partial G_1}{\partial P}\right)_T - \left(\frac{\partial G_2}{\partial P}\right)_T = V_1 - V_2 = \Delta_{\text{trs}} V \neq 0 \tag{3.129}$$

$$\left(\frac{\partial G_1}{\partial T}\right)_P - \left(\frac{\partial G_2}{\partial T}\right)_P = -S_1 + S_2 = -\Delta_{\text{trs}} S = -\frac{\Delta_{\text{trs}} H}{T_{\text{trs}}} \neq 0\,. \tag{3.130}$$

Moreover, in a first order phase transition, a phase can remain metastable over a significant region. A famous example is the supercooling of water. In such metastable regions, the state of the system depends on its history. This phenomenon is known as *hysteresis*. By contrast, in a continuous phase transition, the system smoothly passes to the new phase without needing to overcome any free energy barrier, making such transition very fast.

Our aim is then to arrive at a unified theory to look at phase transitions and to determine whether a phase transition is first order or not in a simple way.

3.8.2 Free Energy Expansion and Order Parameter

What makes two phases different from each other? A plausible answer is symmetry. As Landau points out, a symmetry is either there or not there — there isn't something like a partial symmetry. Just like a crystal is either cubic or it is not — it cannot be 'slightly cubic'. The symmetry of a system cannot change continuously.

To quantify this, we introduce a concept called the *order parameter*. It is some suitably chosen parameter in our system of interest that is zero in one of the phases and non-zero in another so that we hope when the phase transition occurs, the order parameter will jump from zero to a non-zero value. Sometimes it is obvious what to take as the order parameter; other times less so. For magnetic or electric systems, the order parameter is typically some form of magnetization (as for the Ising model) or the polarization. For the liquid-gas transition, the relevant order parameter is the difference in densities between the two phases, $v_{\text{gas}} - v_{\text{liquid}}$. In something like a liquid crystal, as shown in figure 3.16, a suitable order parameter would be some sort of measurement of the average orientation of the molecules along some axis.

Landau theory is based around the free energy, so the next thing we do is to write the free energy in terms of the order parameter. Let's be clear of what we are doing here. We have certainly worked out some expressions like this before. For example, in the Ising model, where the mean magnetisation

[13] However, it should be pointed out that under modern theory of critical phenomena, this classification can no longer be strictly maintained — but it is still a useful one.

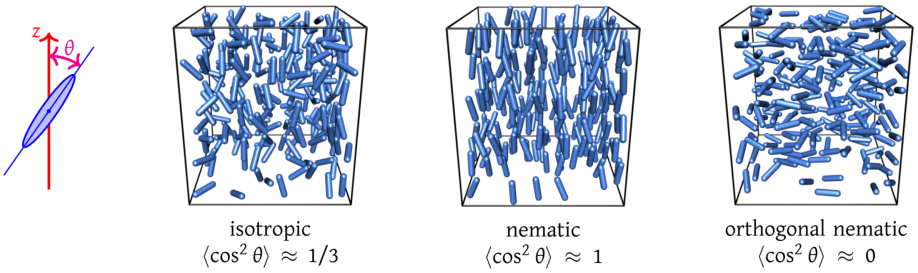


Figure 3.16: A schematic diagram of the isotropic and nematic phases of a liquid crystal. The order parameter can be chosen to be $M = \left\langle 3\cos^2\theta - 1 \right\rangle /2$, which is 0 in the isotropic case, 1 if all particles are perfectly aligned along z axis, and $-1/2$ if all particles are orthogonal to the z axis. Figure adapted from official course notes.

m is the order parameter, we have worked out the free energy expression (3.34)

$$A = \frac{1}{2}NJzm^2 - Nk_BT\ln[2\cosh(\beta Jzm)]\,. \tag{3.131}$$

But so far in this course, we've considered only systems in equilibrium. The free energy, like all other thermodynamic potentials, has only been defined on equilibrium states. Yet the equation above can be thought of as an expression for A as a function of m, where m can be any number, including the non-equilibrium ones. Of course, what we should do is to substitute in the equilibrium value of m given by (3.33) to give the true equilibrium free energy with solid physical meaning, but it seems a shame to throw out $A(m)$ when it is such a nice function. Surely we can put it to some use!

The key in Landau theory is to treat the function $A = A(N, V, T; m)$ seriously. This means that we are extending our viewpoint away from equilibrium states to a whole class of states which have a constant average value of m. You could imagine some external magical power that holds m fixed. The free energy $A(N, V, T; m)$ is then telling us the equilibrium properties in the presence of this magical power. Perhaps more convincing is what we do with the free energy in the absence of any magical constraint. We have seen that equilibrium is guaranteed if we sit at the minimum of A. This means that what we want to do is to solve

$$\frac{\partial A}{\partial m} = 0\,, \tag{3.132}$$

and find where A is the minimum. If this occurs at $m = 0$, then at this condition the disordered phase with zero value of order parameter is stable; if this occurs at non-zero value of m, then the ordered phase is stable.

However, finding an expression of the free energy A seems like a difficult thing to do. This is exactly what we have done in this whole chapter, spending a lot of efforts writing down the free energy of different system, which we have found is not an easy task even for simple models! Moreover, we promised that we will derive a general theory of phase transition, but the details of the free energy clearly depends on the physics of the system. How do we just write down the free energy in the general case? The trick is to assume that we can expand the free energy in an analytic power series in the order parameter. For this to be true, the order parameter must be small which is guaranteed if we are close to a critical point (since we assume that for $T > T_c$ we have the disordered phase with $m = 0$). This allows us to write

$$A(N, V, T, M) = A(N, V, T, 0) + aM + bM^2 + cM^3 + dM^4 + \ldots \tag{3.133}$$

Let's look at a couple of simple examples.

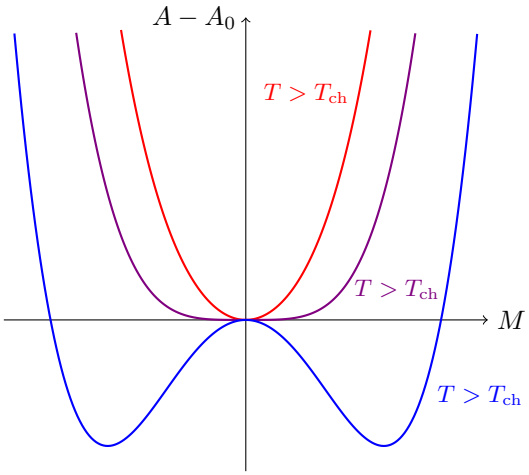


Figure 3.17: A sketch of the Landau free energy A as a function of the order parameter M for a system that is an even function of M. The phase transition occurs at $T = T_{\text{ch}}$, where the quadratic coefficient changes the sign if the quartic coefficient is assumed positive.

3.8.3 Second Order Phase Transitions

We first look at a system in which there is a symmetry in the order parameter by $M \leftrightarrow -M$. An example of such a system is Ising model without external fields, where the order parameter is the magnetisation m — a system with magnetisation $+m$ is completely equivalent to a system with magnetisation $-m$, so the free energy must be an even function of m. We are investigating the phase transitions caused by changes in temperatures only, so for this kind of a system, we will write

$$A(T, M) = A(T, 0) + b(T)M^2 + d(T)M^4 + f(T)M^6 + \dots \tag{3.134}$$

We will focus on the temperature regime where the quadratic coefficient changes its sign, since this is the lowest order term and should cause the greatest effect on the behaviour of A near $M = 0$. We will assume this sign change occurs at $T = T_{\text{ch}}$ so that we can write, to the linear order, $b(T) = \alpha(T - T_{\text{ch}})$. We will approximate higher coefficients to be constant near $T = T_{\text{ch}}$. Let's see what happens if $\alpha > 0$ and $d > 0$, so that

$$A(T, M) \approx A(T, 0) + \alpha(T - T_{\text{ch}})M^2 + dM^4 \,. \tag{3.135}$$

A plot of $A(T, M)$ at different T is shown in figure 3.17.

It can be seen that, above T_{ch}, the minimum of the free energy occurs at $M = 0$, so the disordered phase is more stable. Below T_{ch}, the point $M = 0$ changes from being a minimum in free energy to a local maximum. Instead the minimum in the free energy is easily found to be

$$m = \pm\sqrt{\frac{\alpha(T_{\text{ch}} - T)}{2d}} \,. \tag{3.136}$$

We can substitute this expression back into our free energy expansion, so that we know the free energy of the most stable state is

$$A_{\text{min}} = \begin{cases} A(T, 0) & T \geq T_{\text{ch}} \\ A(T, 0) - \frac{\alpha^2(T - T_{\text{ch}}^2)}{4d} & T < T_{\text{ch}} \,. \end{cases} \tag{3.137}$$

Assuming $A(T, 0)$ is a smooth function of T, we can see that both the free energy and its first derivative are continuous at the phase transition. However, the second derivative, which is related to the heat capacity

$$C_V = -T\left(\frac{\partial^2 A}{\partial T^2}\right)_V \,, \tag{3.138}$$

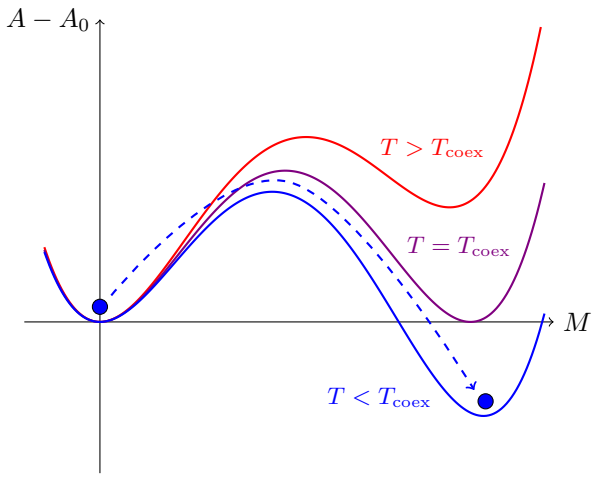


Figure 3.18: A sketch of the Landau free energy A as a function of the order parameter M when cubic terms are included in the expansion. The free energy is now not the same for positive and negative M. The phase transition is now first order, and in order for that to occur, a large free energy barrier is needed to overcome.

is discontinuous. This is characteristic of a second-order phase transition.

You will show in one of the exercises that the Ising model with no external field is exactly the situation depicted above by explicitly expanding the free energy.

3.8.4 First Order Phase Transition

If we relax the condition of the quartic coefficient d being positive, and instead assume that the M^6 coefficient is positive, then we have the free energy expansion

$$A(T, M) \approx A(T, 0) + \alpha(T - T_{\text{ch}})M^2 + dM^4 + fM^6 \tag{3.139}$$

with $\alpha, f > 0$ while $d < 0$. This leads to a more interesting tricritical point, which you will investigate in one of the exercises. Crucially, this time, the first derivative of the free energy is discontinuous at the critical temperature

$$T = T_{\text{ch}} + \frac{d^2}{4\alpha f}\,, \tag{3.140}$$

so this results in a first order phase transition.

Now, let's instead look at an important cases when A includes odd powers of M.

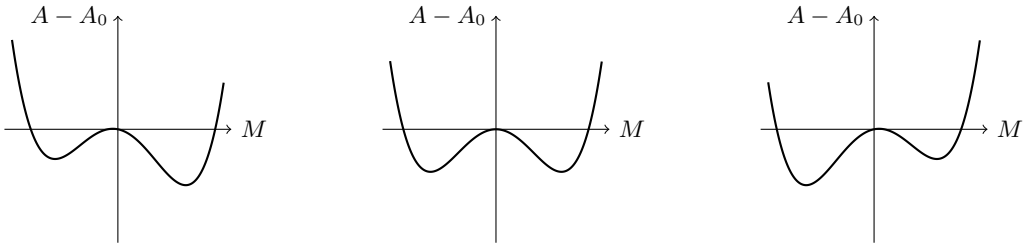
$$A(T, M) = A(T, 0) + \alpha(T - T_{\text{ch}})M^2 + cM^3 + dM^4 + \dots \tag{3.141}$$

We set $\alpha, d > 0$ while $c < 0$. A plot of $A(M)$ at different T is shown in figure 3.18.

The phase transition now occurs when A has two double roots. Hence, it can be solved that the critical temperature is

$$T - T_{\text{ch}} = \frac{c^2}{4da}\,. \tag{3.142}$$

The minimum jumps to $M = -c/2d$ and a first order phase transition takes place. To prove this, we would like to work out how $A_{\text{min}}(T) = A(M_{\text{min}}, T)$ changes with temperature. Since M_{min} also depends on the temperature, we can't differentiate A directly. Although we can substitute the expression of M_{min} into the expression of A to obtain an explicit expression of A_{min} and differentiate



(a) $B > 0$ (b) $B = 0$ (c) $B < 0$

Figure 3.19: The Landau free energy of Ising model with non-zero magnetic fields at low temperatures for different magnetic fields.

it, we will instead work it in a more clever way. Consider the total differential

$$\mathrm{d}A = \left(\frac{\partial A}{\partial M}\right)_{V,T} \mathrm{d}M + \left(\frac{\partial A}{\partial T}\right)_{V,M} \mathrm{d}T\,, \tag{3.143}$$

which gives the total derivative

$$\left(\frac{\mathrm{d}A}{\mathrm{d}T}\right)_V = \left(\frac{\partial A}{\partial M}\right)_{V,T} \left(\frac{\partial M}{\partial T}\right)_V + \left(\frac{\partial A}{\partial T}\right)_{M,V}\,. \tag{3.144}$$

Now since at the free energy minimum, $(\partial A/\partial M)_{V,T} = 0$, we have

$$\left(\frac{\mathrm{d}A_{\min}}{\mathrm{d}T}\right)_V = \left(\frac{\partial A}{\partial T}\right)_{M=M_{\min},V} = \alpha M^2\,. \tag{3.145}$$

Since there is a jump of M at phase transition, the derivative of free energy minimum against the temperature is also discontinuous, giving a first order phase transition.

Note also that in this situation, in order for a phase transition to occur, the system needs to overcome a large free energy barrier. This is known as *hysteresis*. This type of phase transitions occur when there is no symmetry between states with positive and negative order parameters, like freezing, evaporation and the isotropic-nematic transitions. They are all first order.

Finally, let's look at a somehow complicated system. Let's consider the Ising model with a non-zero external magnetic field. It is not difficult to show that

$$A(T, B, M) = A(T, B, 0) + a(T, B)M + b(T, B)M^2 + d(T, B)M^4 \ldots \tag{3.146}$$

This is essentially the same expression as for the $B = 0$ case, just with an extra linear term. The magnetic field creates a preferred direction in the system, making it no longer symmetric about $M \leftrightarrow -M$. Also, the equilibrium state at high temperature no longer has $M = 0$.

The free energy once again has a double well, except now slightly skewed. The local maximum is still an unstable point. But this time around, the minima with the lower free energy is preferred over the other one. This is the true ground state of the system. In contrast, the point which is locally, but not globally, a minimum corresponds to a metastable state of the system. In order for the system to leave this state, it must first fluctuate up and over the energy barrier separating the two. Thus, when we switch the magnetic field from $B > 0$ to $B < 0$, the phase transition is first order.

At very high temperature, the double well potential is lost in favour of a single minimum. There is a unique ground state, albeit shifted from $M = 0$ by the presence of the linear term $a(T, B)$ in the expansion above (which translates into the magnetic field B in the Ising model). The temperature at which the metastable ground state of the system is lost corresponds to the spinodal point in our discussion of the liquid-gas transition. This time, when we switch the direction of the magnetic field,

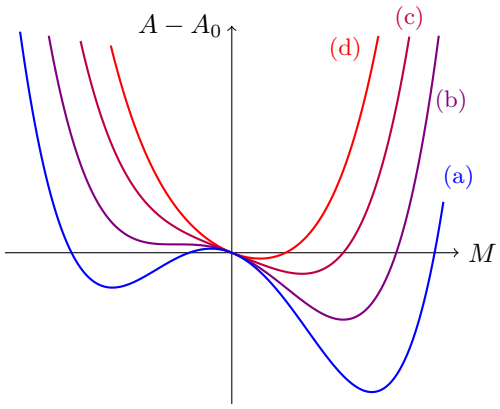


Figure 3.20: The Landau free energy of an Ising model with non-zero magnetic field as temperature changes. (a) The temperature is low and the free energy curve shows a double well. (b) This is the spinodal where the metastable state disappears. (c) This is the phase transition temperature when $B = 0$. Above this temperature, there will be no phase transition as we change B. (d) This is the high temperature case.

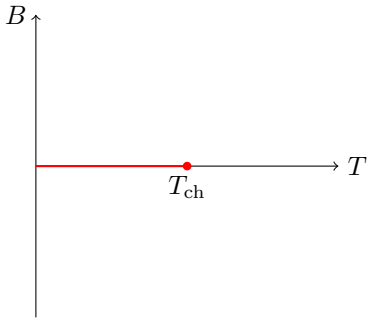


Figure 3.21: The phase diagram of the Ising model.

no phase transition will occur, and similarly when the magnetic field is non-zero, there will be no phase transition as we change the temperature.

It is worth noticing that although Landau theory is a simple and useful way of analysing phase transitions, it is not rigorous. In particular, it makes the assumption that the state of a system is completely determined by the minimum of the Landau free energy. However, as we have seen before, fluctuations about the minimum are crucial in phase transitions. This problem can be fixed by a more advanced theory called Landau–Ginzburg theory, which you can find in appendix section H.

4 Transport Phenomena

4.1 Diffusion and Einstein Relation

At non-zero temperatures, molecules are always in motion. If they are in liquid or gas phase where there are enough translational degrees of freedom, they will move and collide frequently and as a result, follow highly irregular trajectories that are often called *random walks*. It is difficult to observe the thermal motion of molecules directly, but large particles also undergo thermal motion. The most famous example is the Brownian motion, which was discovered by Robert Brown and was first rationalised by Albert Einstein. This random motion of molecules is responsible for the process that we call *diffusion*. If we carefully introduce a small droplet of ink into a glass of water, then even without stirring, the droplet will become increasingly fuzzy until the colour is uniform in the water. The initial inhomogeneities are washed out over time.

In what follows, we will consider the simplest type of diffusion known as *self-diffusion*. This is when the diffusing molecules move in a medium consisting of the same species. In experiment, we can study self-diffusion if we can label some of the molecules without affecting their interactions with the others by controlling their nuclear spins or isotopic composition.

The macroscopic law that describes diffusion is *Fick's first law*. It states that whenever there is a local concentration gradient of some species, then there will be a flux $\mathbf{j}$ proportional to the negative of the concentration gradient. In one dimension, if the concentration of some species at a point x at time t is $c(x,t)$, then there will be a flux of the species given by

$$j(x,t) = -D\frac{\partial c}{\partial x}\,, \tag{4.1}$$

where the constant of proportionality D is known as the *diffusion coefficient*.

Now consider a small region between x and $x+\delta x$. The change in the number of the species of interest over a small time interval δt is

$$\delta N = N_{\text{in}} - N_{\text{out}} = [j(x) - j(x+\delta x)]\delta t\,. \tag{4.2}$$

By the definition of concentration in one dimension, we have

$$\begin{aligned}\frac{\partial c}{\partial t} &= \lim_{\delta t\to 0}\frac{\delta c}{\delta t} = \lim_{\delta t\to 0}\lim_{\delta x\to 0}\frac{\delta N}{\delta x \delta t}\\ &= -\lim_{\delta x\to 0}\frac{j(x+\delta x)-j(x)}{\delta x}\\ &= -\frac{\partial j}{\partial x}\,.\end{aligned} \tag{4.3}$$

From this, we get *Fick's second law*

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2}\,, \tag{4.4}$$

which is also known as the *diffusion equation* in one dimension.

This is a second-order linear partial differential equation. It is not difficult to solve. A particularly important case is when the initial condition is the Dirac delta function $c(x,0) = \delta(x)$. Then the solution is given by

$$c(x,t) = \frac{1}{(4\pi D t)^{1/2}}\exp\left(-\frac{x^2}{4Dt}\right)\,. \tag{4.5}$$

This is a Gaussian of width increasing with time. However, in this course, we will not need to solve this diffusion equation directly. Instead, we just want to find out time dependence of the mean squared width of a concentration profile,

$$\left\langle x^2(t)\right\rangle = \int \mathrm{d}x\, x^2 c(x,t)\,, \tag{4.6}$$

where we have assumed the concentration is normalised

$$\int \mathrm{d}x\, c(x,t) = 1 \tag{4.7}$$

so that the total particle number of the diffusion species we are interested in is always unity. We can multiply the diffusion equation by x^2 and integrate to get

$$\int \mathrm{d}x\, x^2 \frac{\partial c}{\partial t} = D \int \mathrm{d}x\, x^2 \frac{\partial^2 c}{\partial x^2}\,. \tag{4.8}$$

We can move the differentiation with respect to t outside of the integration by the Leibniz integral rule, giving

$$\frac{\partial}{\partial t} \int \mathrm{d}x\, x^2 c \equiv \frac{\partial}{\partial t} \left\langle x^2(t) \right\rangle = D \int \mathrm{d}x\, x^2 \frac{\partial^2 c}{\partial x^2}\,. \tag{4.9}$$

We can apply integration by parts twice to the right-hand side to get

$$\frac{\partial \left\langle x^2 \right\rangle}{\partial t} = D \left[x^2 \frac{\partial c}{\partial x} \right]_{-\infty}^{\infty} - 2D \left[xc \right]_{-\infty}^{\infty} + 2D \int \mathrm{d}x\, c(x,t)\,. \tag{4.10}$$

If the concentration and its derivative decay fast enough,[14] like in the delta function case where c decays as a Gaussian, then the previous two boundary terms vanish, and we get

$$\frac{\partial \left\langle x^2(t) \right\rangle}{\partial t} = 2D\,, \tag{4.11}$$

or equivalently

$$\left\langle x^2(t) \right\rangle = 2Dt\,. \tag{4.12}$$

This is the *Einstein relation*, relating the diffusion coefficient D to the width of the concentration profile. This can easily be generalised to 3D systems to give

$$\left\langle \mathbf{x}^2(t) \right\rangle = 6Dt\,. \tag{4.13}$$

It should be realised that, whilst D is a macroscopic transport coefficient, $\left\langle \mathbf{x}^2 \right\rangle$ has a microscopic interpretation: it is the mean squared distance over which the labelled molecules have moved in a time interval t. If we can follow the motion of individual molecules, we can then determine D from a knowledge of the particle trajectories.

4.2 Random Walk Model of Diffusion

Let's consider a simplified model of diffusion, in which both space and time are discretised. We assume that the particles live on an infinite 1D array of points, and they jump to a neighbouring point at a rate Γ per unit time. At every jump, the particle will move a distance ℓ, either to the left or to the right, with equal probability, and the direction of each jump is independent. Then the average distance of each jump is

$$\langle \Delta x_1 \rangle = 0.5 \times \ell - 0.5 \times \ell = 0 \tag{4.14}$$

as one might expect from symmetry: the particles do not move on average. However, the mean squared displacement of the particle after one jump is non-zero,

$$\left\langle (\Delta x_1)^2 \right\rangle = 0.5 \times \ell^2 + 0.5 \times (-\ell)^2 = \ell^2\,. \tag{4.15}$$

[14] The concentration should decay faster than x^{-1} in order for the $\int \mathrm{d}x\, c$ to converge, and so these boundary terms should vanish.

Now let's consider what will happen after time t. By then, the molecule will have attempted $M = \Gamma t$ jumps on average, and the mean displacement will be

$$\langle \Delta x \rangle = \left\langle \sum_{i=1}^{M} \Delta x_i \right\rangle = \sum_{i=1}^{M} \langle \Delta x_i \rangle = 0 \,, \tag{4.16}$$

where Δx_i is the displacement of the i^{th} jump. Next, we have

$$\left\langle x^2(t) \right\rangle = \left\langle \left(\sum_{i=1}^{M} \Delta x_i \right)^2 \right\rangle = \left\langle \sum_{i=1}^{M} \sum_{j=1}^{M} \Delta x_i \Delta x_j \right\rangle = \sum_{i=1}^{M} \sum_{j=1}^{M} \langle \Delta x_i \Delta x_j \rangle \,. \tag{4.17}$$

As different jumps are uncorrelated, we have

$$\langle \Delta x_i \Delta x_j \rangle = \frac{1}{4}[\ell \times \ell + \ell \times (-\ell) + (-\ell) \times \ell + (-\ell) \times (-\ell)] = 0 \tag{4.18}$$

for $i \neq j$, while if $i = j$, $\langle \Delta x_i \Delta x_j \rangle = \left\langle (\Delta x_i)^2 \right\rangle = \ell^2$. Therefore,

$$\left\langle x^2(t) \right\rangle = \sum_{i=1}^{M} \left\langle (\Delta x_i)^2 \right\rangle = M\ell^2 = \Gamma t \ell^2 \,. \tag{4.19}$$

We can relate this to the Einstein relation

$$\frac{\partial \left\langle x^2(t) \right\rangle}{\partial t} = 2D \,, \tag{4.20}$$

and so we get the diffusion coefficient for a random walk model

$$D = \frac{1}{2} \Gamma \ell^2 \,. \tag{4.21}$$

This relates the macroscopic diffusion coefficient D to the microscopic jump frequency Γ and jump distance ℓ.

We can use this to model the diffusion of gas molecules in 3D. If the molecules move at a typical thermal speed v between collisions, and the mean free path between collisions is λ (both of which can be obtained from kinetic theory), then we can model the movement of a molecule between each collision as a single jump in random walk, with mean jump distance in any direction given by[15]

$$\ell^2 \approx \frac{\lambda^2}{3} \tag{4.22}$$

and mean jumping rate

$$\Gamma \approx \frac{v}{\lambda} \,. \tag{4.23}$$

Then the diffusion coefficient can be estimated to be

$$D = \frac{1}{2} \Gamma \ell^2 \approx \frac{1}{6} v \lambda \,. \tag{4.24}$$

4.3 The Green–Kubo Relation

The above random walk is a highly simplified model. Molecules in liquid and gas do not move by jumps. Rather, their velocity changes continuously in time. If a molecule has an initial x-velocity $v_x(0)$ at time $t = 0$, then its velocity will still be close to $v_x(0)$ at a short time afterwards before

[15] We have a factor $\frac{1}{3}$ because the walk is in three dimensions, which probably does not matter for an order-of-magnitude estimate.

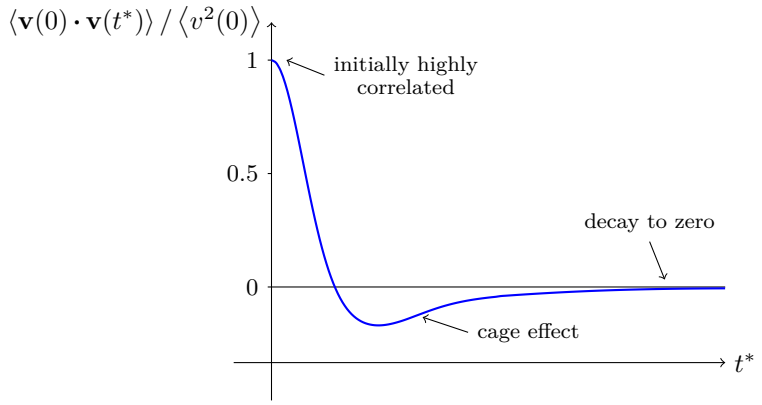


Figure 4.1: A schematic plot of velocity autocorrelation functions for liquid. For short time t^*, the velocity is highly correlated to its initial value since the velocity of a particle does not change significantly before colliding with another particle. In a liquid, particles are often surrounded by a cage of other particles, and so we often see an anti-correlation after a short time, where the particle has bounced back after hitting the cage and is travelling in the opposite direction from the one it started with. In the long time limit, the particle has collided with many other particles, which has effectively randomised its velocity vector, so it is no longer correlated with the initial value. The velocity autocorrelation function decays to zero.

it collides with another molecule. 'Short' in this context depends on the conditions. For a water molecule at room temperature, it takes around 10^{-13} s before a molecule forgets about its initial velocity.

The displacement $\Delta\mathbf{r}(t)$ that we are interested in is simply the time integral of the velocity of the labelled particle,

$$\Delta\mathbf{r}(t) = \int_0^t \mathrm{d}t' \, \mathbf{v}(t') \,. \tag{4.25}$$

The Einstein relation (4.11) allows us to relate the diffusion coefficient in terms of the particle velocities, so we start from

$$D = \frac{1}{2} \lim_{t\to\infty} \frac{\partial \left\langle x^2(t) \right\rangle}{\partial t} \,, \tag{4.26}$$

where we have only considered the x component of the mean squared displacement for convenience.[16] We can write $x(t)$ as the time integral of the x-component of the labelled particle velocity, so

$$\left\langle x^2(t) \right\rangle = \left\langle \left(\int_0^t \mathrm{d}t \, v_x(t') \right)^2 \right\rangle = \int_0^t \mathrm{d}t' \int_0^t \mathrm{d}t'' \, \langle v_x(t') v_x(t'') \rangle \,. \tag{4.27}$$

The quantity $\langle v_x(t')v_x(t'')\rangle$ is called the *velocity autocorrelation function*. It measures the average correlation between the velocity of a particle at times t' and t''. A typical velocity autocorrelation function is shown in figure 4.1.

We can simplify this integral

$$\begin{aligned}
\left\langle x^2(t) \right\rangle &= \int_0^t \mathrm{d}t' \int_0^{t'} \mathrm{d}t'' \, \langle v_x(t') v_x(t'') \rangle + \int_0^t \mathrm{d}t' \int_{t'}^t \mathrm{d}t'' \, \langle v_x(t') v_x(t'') \rangle \\
&= \int_0^t \mathrm{d}t' \int_0^{t'} \mathrm{d}t'' \, \langle v_x(t') v_x(t'') \rangle + \int_0^t \mathrm{d}t'' \int_0^{t''} \mathrm{d}t' \, \langle v_x(t') v_x(t'') \rangle \\
&= 2 \int_0^t \mathrm{d}t' \int_0^{t'} \mathrm{d}t'' \, \langle v_x(t') v_x(t'') \rangle \,.
\end{aligned} \tag{4.28}$$

[16] We set the limit $t \to \infty$ to ensure the motion of the molecule is diffusive rather than ballistic. You can see more on this discussion in one of the exercises.

The velocity autocorrelation function is an equilibrium property of the system, since it describes the average correlation between velocities at different times along an equilibrium trajectory. As equilibrium properties are time invariant, we can shift the time origin by an arbitrary amount Δt, so that

$$\langle v_x(t')v_x(t'')\rangle = \langle v_x(t'+\Delta t)v_x(t''+\Delta t)\rangle \ . \tag{4.29}$$

We therefore have

$$\langle v_x(t')v_x(t'')\rangle = \langle v_x(t'-t'')v_x(0)\rangle \ , \tag{4.30}$$

so

$$\left\langle x^2(t)\right\rangle = 2\int_0^t \mathrm{d}t' \int_0^{t'} \mathrm{d}t'' \ \langle v_x(t'-t'')v_x(0)\rangle \ . \tag{4.31}$$

Finally using the Einstein relation, we can write

$$\begin{aligned} D &= \lim_{t\to\infty} \frac{\partial}{\partial t} \int_0^t \mathrm{d}t' \int_0^{t'} \mathrm{d}t'' \ \langle v_x(t'-t'')v_x(0)\rangle \\ &= \lim_{t\to\infty} \int_0^t \mathrm{d}t'' \ \langle v_x(t-t'')v_x(0)\rangle \ . \end{aligned} \tag{4.32}$$

By introducing the variable $\tau = t - t''$, we have

$$D = \int_0^\infty \mathrm{d}\tau \ \langle v_x(\tau)v_x(0)\rangle \tag{4.33}$$

known as the *Green–Kubo relation*, which relates the diffusion coefficient D to the velocity autocorrelation function. In an isotropic 3D system, this becomes

$$D = \frac{1}{3}\int_0^\infty \mathrm{d}\tau \ \langle \mathbf{v}(\tau)\cdot\mathbf{v}(0)\rangle \ . \tag{4.34}$$

More generally, a Green–Kubo relation is any equation that relates an integral of a time-correlation function to a transport coefficient, such as shear viscosity, thermal conductivity and electrical conductivity *etc.* It should be emphasised that for classical systems, the Green–Kubo relation for D and the Einstein relation $\left\langle x^2\right\rangle = 2Dt$ are strictly equivalent.

4.4 Langevin Equation and Stokes–Einstein Relation

The Green–Kubo expression for the diffusion coefficient provides a way of estimating the diffusion coefficient of particles dissolved in viscous fluid, since in such cases, we can make a few reasonable assumptions to compute the velocity autocorrelation function. In particular, we will assume the following:

- When a particle is subjected to a constant external force $\mathbf{F}$, the average drift velocity will be given by
$$\langle \mathbf{v}\rangle = \frac{\mathbf{F}}{\gamma} \, , \tag{4.35}$$
where γ is the *friction coefficient*, and a related quantity is $\mu = 1/\gamma$ known as the *mobility*.
- If you put some fluid between two parallel plates of area A separated by h and move the two plates relative to each other with constant relative velocity v, then usually there will be a constant velocity gradient
$$\frac{\partial u}{\partial z} = \frac{v}{h} \tag{4.36}$$

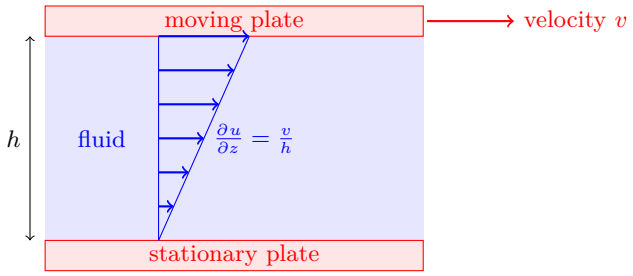


Figure 4.2: When two parallel plates filled with fluid are moving with relative velocity v, the retarding force will be proportional to both the area of the plates and the velocity gradient, with proportionality constant known as the viscosity. To avoid boundary effects in experiments, this is usually done using rotating discs.

within the fluid. The magnitude of the force F that has to be exerted on the top plate to maintain the relative motion is given by

$$F = \eta A \frac{v}{h}\,, \tag{4.37}$$

where the constant of proportionality η is known as the *viscosity*. We will assume that our fluid has a constant viscosity η.

- For a spherical particle with a radius a moving in a fluid with viscosity η, we can approximate the friction coefficient γ by the *Stokes expression*

$$\gamma = 6\pi\eta a\,. \tag{4.38}$$

This assumes no slippage of the fluid at the surface of the sphere.

- The particles experience random collisions with their neighbours. The effect of these collisions is to exert a random force $F^R(t)$ on our particle. We will assume that $F^R(t)$ for $t > 0$ is not correlated with the velocity of the particle at time $t = 0$.

For simplicity, we again only consider the x component of the motion. The above assumptions allow us to write, by Newton's second law,

$$m\dot{v}_x(t) + \gamma v_x(t) = F_x^R(t)\,. \tag{4.39}$$

This is known as the *Langevin equation*.

We can multiply both sides of the Langevin equation (4.39) by $v(0)$ and use the assumption $\langle v_x(0)F_x^R(t)\rangle = 0$ to obtain the condition of the velocity correlation function

$$m\,\langle v_x(0)\dot{v}_x(t)\rangle + \gamma\,\langle v_x(0)v_x(t)\rangle = 0\,. \tag{4.40}$$

This is an easy ordinary differential equation with solution

$$\langle v_x(0)v_x(t)\rangle = \langle v_x(0)^2\rangle \exp\left(-\frac{\gamma t}{m}\right)\,. \tag{4.41}$$

At equilibrium, the velocities of the particles are distributed according to the Maxwell distribution, with $\langle v_x^2(0)\rangle = k_BT/m$.[17] We then have

$$D = \int_0^\infty \mathrm{d}t\,\frac{k_BT}{m}\exp\left(-\frac{\gamma t}{m}\right) = \frac{k_BT}{\gamma}\,. \tag{4.42}$$

[17] Or, equally well, by the equipartition of energy.

If we use the Stokes approximation for γ, we get

$$D = \frac{k_B T}{6\pi\eta a}\,, \tag{4.43}$$

known as the *Stokes–Einstein relation*. It is extremely useful and it correctly reproduces the experimental fact that for many liquids, the product $D\eta$ is very nearly proportional to the temperature, even when D and η each vary orders of magnitude. It is also a good way to determine Boltzmann's constant experimentally. The experiment was done in 1909 by Jean-Baptiste Perrin and won him the 1926 Nobel prize.

The Stokes–Einstein relation actually has profound implications. It tells us diffusion and viscosity are closely related — their microscopic origins are both the random bombardment of molecules. It is also an important example of the fluctuation-dissipation theorem. The fluctuations of the particle as it undergoes its random walk are related to the drag force (or dissipation of momentum) that the particle feels as it moves through the fluid.

4.5 The Einstein–Smoluchowski Relation

There is another more general way of deriving the relation between the mobility and the diffusion coefficient D. Consider a solution in a closed volume V. If there is a concentration gradient of the dissolved species, then this will result in a diffusion flux $\mathbf{j}$ as predicted by Fick's first law. The diffusive flux equals the product of the number density and the average drift velocity of these particles,

$$\mathbf{j} = \rho \langle \mathbf{v} \rangle \,. \tag{4.44}$$

Now suppose that the dissolved particles are subject to an external potential $U(x)$. If this potential is not constant, then there will be a net force acting on the particles, given by

$$f_x = -\frac{\partial U(x)}{\partial x}\,, \tag{4.45}$$

where we have again reduced the problem to one dimension for simplicity. The average drift velocity of a particle due to this force is

$$\langle v_x \rangle = \mu f_x = -\mu \frac{\partial U(x)}{\partial x}\,. \tag{4.46}$$

As the particles move under the influence of this force, the density becomes inhomogeneous. But if the particle number density is not constant, there is also a diffusion flux. When the system reaches equilibrium, the flux due to diffusion and the flux due to external force will balance

$$0 = j_{\text{tot}} = \rho(x) \langle v_x \rangle - D\frac{\partial \rho(x)}{\partial x} = -\rho(x)\mu \frac{\partial U(x)}{\partial x} - D\frac{\partial \rho(x)}{\partial x}\,. \tag{4.47}$$

At equilibrium, the probability of finding a particle at a position x for an inhomogeneous system is proportional to the Boltzmann factor,

$$\rho(x) = \rho_0 \exp\left(-\frac{U(x)}{k_B T}\right)\,. \tag{4.48}$$

Therefore, we have

$$-\rho(x)\mu \frac{\partial U(x)}{\partial x} + \frac{D}{k_B T}\rho(x)\frac{\partial U(x)}{\partial x} = 0\,. \tag{4.49}$$

This equation is satisfied for all x if

$$\mu = \frac{D}{k_B T}\,, \tag{4.50}$$

which is known as the *Einstein–Smoluchowski relation*. We have recovered the same expression as obtained from the Langevin equation, but we have made fewer assumptions along the way. Moreover, this approach is more easily generalisable to include external fields.

A more systematic, formal approach to fluctuation-dissipation relations and the link between transport coefficients and equilibrium statistical mechanics is explored in appendix section I

5 Polymers

In the last chapter, we found that the distribution of displacements of particles under diffusion is given by a Gaussian function. Rather interestingly, the same distribution describes the end-to-end distance of a polymer chain consisting of non-interacting segments. The similarity is more than superficial: it arises from the random walk nature of both phenomena.

5.1 Size of an Ideal Polymer

Let's consider a polymer that consists of N segments of length ℓ. When the polymer is fully stretched into a straight line, the total length of the polymer chain is

$$R_c = N\ell\,, \tag{5.1}$$

known as the *contour length* of the polymer. However, a real polymer molecule is rarely completely straight: when it coils up, its size will be less than R_c. We will model the polymer as fully flexible, so that every segment can rotate freely around its connecting point to the previous segment. Of course this is a huge simplification, but in practice, many polymers do behave like a freely joined chain, although the length ℓ is larger than the size of a monomer — it is the length at which the orientation of the polymer chain becomes uncorrelated. In addition, we will make the assumption that the polymer is *ideal*, such that different segments do not interact. We have seen that under θ conditions, the polymer segments do behave ideally, and this can be used as a rough estimate of the behaviour of non-ideal polymers.

We want to find out the average *end-to-end* distance of an ideal polymer. If we denote the vector pointing from the beginning to the end point of the i^{th} segment by $\boldsymbol{\ell}_i$ such that $\|\boldsymbol{\ell}_i\| = \ell$, then the total end-to-end vector is

$$\mathbf{R}_{\text{ee}} = \sum_{i=1}^{N} \boldsymbol{\ell}_i\,. \tag{5.2}$$

We would like to compute its average. However, by symmetry, a polymer segment has equal probability to point in any orientation, so $\langle \boldsymbol{\ell}_i \rangle = \mathbf{0}$. As a result, the average end-to-end distance of a polymer is

$$\langle \mathbf{R}_{\text{ee}} \rangle = \left\langle \sum_{i=1}^{N} \boldsymbol{\ell}_i \right\rangle = \sum_{i=1}^{N} \langle \boldsymbol{\ell}_i \rangle = \mathbf{0} \tag{5.3}$$

as one might predict by symmetry. Just like in the case of diffusion, a better measure for the size of a coiled ideal polymer is the *mean squared end-to-end distance* $\langle \mathbf{R}_{\text{ee}}^2 \rangle$,

$$\begin{aligned} \langle \mathbf{R}_{\text{ee}}^2 \rangle &= \left\langle \left(\sum_{i=1}^{N} \boldsymbol{\ell}_i \right)^2 \right\rangle = \left\langle \left(\sum_{i=1}^{N} \boldsymbol{\ell}_i \right) \cdot \left(\sum_{i=1}^{N} \boldsymbol{\ell}_i \right) \right\rangle \\ &= \left\langle \sum_{i,j=1}^{N} \boldsymbol{\ell}_i \cdot \boldsymbol{\ell}_j \right\rangle = \sum_{i,j=1}^{N} \langle \boldsymbol{\ell}_i \cdot \boldsymbol{\ell}_j \rangle = \ell^2 \sum_{i,j=1}^{N} \langle \cos\theta_{ij} \rangle\,. \end{aligned} \tag{5.4}$$

For different segments $i \neq j$, the orientations are uncorrelated, so $\langle \cos\theta_{ij} \rangle = 0$. However, for $i = j$, $\theta_{ij} = 0$ and hence $\cos\theta_{ii} = 1$. As a consequence, all the cross-terms with $i \neq j$ vanish and we have

$$\langle \mathbf{R}_{\text{ee}}^2 \rangle = N\ell^2\,. \tag{5.5}$$

The root mean squared end-to-end distance is therefore

$$\sqrt{\langle \mathbf{R}_{\text{ee}}^2 \rangle} = \sqrt{N}\ell\,. \tag{5.6}$$

5.2 Probability Distribution for the End-to-End Distance

We can actually do better than simply calculating the mean squared end-to-end distance of a polymer: we can actually compute the probability distribution of $\mathbf{R}_{\text{ee}}$. However, to simplify matters, we will only consider the one-dimensional case. Again, consider N freely joined segments of length ℓ. We denote the number of segments pointing right by N_R, and the number pointing to the left by N_L. We denote their difference $N_R - N_L =: D$ and clearly the end-to-end distance is $X_{\text{ee}} = D\ell$. We can compute the probability of finding the chain in a conformation with a given value of D. The total number of conformations of N polymer segments is 2^N, and the number of ways in which we can have N_R segments pointing to the right and N_L segments pointing to the left is

$$\Omega(N_R, N_L) = \frac{N!}{N_R! N_L!} = \frac{N!}{[\frac{1}{2}(N+D)]![\frac{1}{2}(N-D)]!}\,. \tag{5.7}$$

Hence, the probability of having a conformation with some D is

$$\text{Prob}(N, D) = \frac{N!}{[\frac{1}{2}(N+D)]![\frac{1}{2}(N-D)]!}\left(\frac{1}{2}\right)^N\,. \tag{5.8}$$

We can expand this using Stirling's approximation

$$\ln x! = x\ln x - x + \ln\sqrt{2\pi x} + O\left(\frac{1}{x}\right) \tag{5.9}$$

to get

$$\begin{aligned}\ln \text{Prob}(N, D) = &-\frac{1}{2}N\ln\left(1 - \frac{D^2}{N^2}\right) + \ln\sqrt{2\pi N} \\ &- \ln\sqrt{\pi(N+D)} - \ln\sqrt{\pi(N-D)} - \frac{1}{2}D\ln\left(\frac{1+\frac{D}{N}}{1-\frac{D}{N}}\right)\end{aligned} \tag{5.10}$$

for large N. For $D/N \ll 1$, we can use Taylor expansion $\ln(1 \pm x) = \pm x + O(x^2)$ when $|x| \ll 1$ to write

$$\ln \text{Prob}(N, D) \approx -\frac{D^2}{2N} - \ln\sqrt{\pi N/2}\,, \tag{5.11}$$

from which we have

$$\text{Prob}(N, D) = \frac{1}{\sqrt{\pi N/2}}\exp\left(-\frac{D^2}{2N}\right)\,. \tag{5.12}$$

However, if you try to integrate this expression, you will find that this is not properly normalised:

$$\int_{-\infty}^{\infty} \mathrm{d}D\ \text{Prob}(D, N) = 2\,. \tag{5.13}$$

What is wrong with this expression? Notice that if in order to keep $N_R + N_L = N$ fixed, if N_R increases by 1, then N_L should decrease by 1 — this leads to a change in D by 2. When N is even, then D can only be even, and if N is odd, then D can only be odd. This means that half of the probability distribution of D should actually be 0.

$$\text{Prob}(N, D) = \begin{cases} \frac{1}{\sqrt{\pi N/2}}\exp\left(-\frac{D^2}{2N}\right) & \text{if } N \equiv D \,(\text{mod}\, 2) \\ 0 & \text{otherwise}\,. \end{cases} \tag{5.14}$$

However, in practice, it is more often to just halve the probability density and write

$$\text{Prob}(D) = \frac{1}{\sqrt{2\pi N}}\exp\left(-\frac{D^2}{2N}\right) \tag{5.15}$$

so that it is properly normalised and gives the probability of D being in some interval from D to $D + \Delta D$ correctly when N is large so that we can effectively treat D as a continuum. From this Gaussian distribution, we can also deduce the mean squared end-to-end distance of the one-dimensional chain to be

$$\left\langle X_{\mathrm{ee}}^2 \right\rangle = \ell^2 \left\langle D^2 \right\rangle = N\ell^2 \tag{5.16}$$

as obtained previously.

5.3 Thermodynamics of Ideal Polymers

Finally we can make a link to the thermodynamics. From the statistical mechanical definition of the entropy, we have

$$S(D) = -\frac{D^2 k_B}{2N} + \text{const.} \tag{5.17}$$

where const. accounts for all terms that are independent of D. This expression is valid for small D. The end-to-end elongation of the polymer is $X_{\mathrm{ee}} = D\ell$, so the entropy of a polymer chain with end-to-end distance X_{ee} is

$$S(X_{\mathrm{ee}}) = -\frac{X_{\mathrm{ee}}^2 k_B}{2N\ell^2} + \text{const.} \tag{5.18}$$

Since the polymer is ideal, there is no energetic preference for any value of X_{ee}, and the internal energy E only depends on the temperature, just as for an ideal gas. By the first law of thermodynamics,

$$\mathrm{d}E = đw + đq\,. \tag{5.19}$$

The heat transfer is given by the Clausius expression $đq = T\mathrm{d}S$ and the work done to change the length of a polymer is $đw = F\mathrm{d}X_{\mathrm{ee}}$, where F is the external force applied to stretch the polymer. We then have

$$\mathrm{d}E = T\mathrm{d}S + F\mathrm{d}X_{\mathrm{ee}}\,, \tag{5.20}$$

and so at constant temperature,

$$\left(\frac{\partial E}{\partial X_{\mathrm{ee}}}\right)_T = 0 = T\left(\frac{\partial S}{\partial X_{\mathrm{ee}}}\right)_T + F\,, \tag{5.21}$$

and hence

$$F = -T\left(\frac{\partial S}{\partial X_{\mathrm{ee}}}\right)_T\,. \tag{5.22}$$

By our expression of the entropy, we get

$$F = \frac{k_B T}{N\ell^2} X_{\mathrm{ee}}\,. \tag{5.23}$$

The restoring force $-F$ is linear in the extension X_{ee} when the extension is small: $X_{\mathrm{ee}} \ll \sqrt{N}\ell$ — just like a normal spring. But unlike a metal spring, the spring constant is proportional to temperature. This is exactly the force you feel when stretching a polymer, like a rubber band, and rather surprisingly, this force is purely entropic!

We can combine (5.16) and (5.23) to get an expression of the spring constant of an ideal polymer

$$\kappa = \frac{k_B T}{\langle X_{\mathrm{ee}}^2 \rangle}\,. \tag{5.24}$$

In the post-lecture question, you will show that this result is actually also valid for non-ideal polymers.

Finally, let's consider what will happen if we stretch a polymer adiabatically, i.e. without heat exchange to the environment. For a reversible process, this means that the entropy is held constant,

as $\mathrm{d}q_{\mathrm{rev}} = T\mathrm{d}S$. The internal energy is clearly a function of T and X_{ee}, so by chain rule, the total derivative of energy against the extension is

$$\left(\frac{\partial E}{\partial X_{\mathrm{ee}}}\right)_S = \left(\frac{\partial E}{\partial T}\right)_{X_{\mathrm{ee}}} \left(\frac{\partial T}{\partial X_{\mathrm{ee}}}\right)_S + \left(\frac{\partial E}{\partial X_{\mathrm{ee}}}\right)_T . \tag{5.25}$$

By comparing to the fundamental equation

$$\mathrm{d}E = T\mathrm{d}S + F\mathrm{d}X_{\mathrm{ee}} , \tag{5.26}$$

we have

$$\left(\frac{\partial E}{\partial X_{\mathrm{ee}}}\right)_S = F , \tag{5.27}$$

and we also know that by definition,

$$\left(\frac{\partial E}{\partial T}\right)_{X_{\mathrm{ee}}} = C_{X_{\mathrm{ee}}} , \tag{5.28}$$

the heat capacity at constant extension. For an ideal polymer, $(\partial E/\partial X_{\mathrm{ee}})_T = 0$, and so we have

$$\left(\frac{\partial T}{\partial X_{\mathrm{ee}}}\right)_S = \frac{F}{C_{X_{\mathrm{ee}}}} = \frac{k_B T}{C_{X_{\mathrm{ee}}} N\ell^2} X_{\mathrm{ee}} . \tag{5.29}$$

As the heat capacity is always positive, it follows that the temperature increases when the polymer is stretched adiabatically, and decreases when the polymer is released.

Appendices

A Thermodynamics

Thermodynamics is a remarkable discipline. It provides us with relations between measurable quantities such as $(\partial S/\partial V)_T = (\partial P/\partial T)_V$. These relations are valid for any substance. But, precisely for this reason, thermodynamics contains no information whatsoever about the underlying microscopic structure of a substance. Thermodynamics preceded statistical mechanics and quantum mechanics, yet not a single thermodynamic relation had to be modified in light of these later developments. The reason is simple: thermodynamics is a phenomenological science. It is based on the properties of matter as we observe them, not upon any theoretical ideas that we may have about matter.

A.1 The First Law

The first law of thermodynamics expresses the empirical observation that energy is conserved, even though it can be converted into various forms. The internal energy of a system can be changed by either performing work on the system or by transferring an amount of heat.

Law (First law of thermodynamics). Let đq be the heat transferred to a system and let đw be the work done to the system. Then the change in the internal energy of the system is given by

$$\Delta E = đq + đw\,. \tag{A.1}$$

A.2 The Second Law

The second law is based on experimental observations.

Law (Kelvin's second law). It is impossible to make an engine that converts heat from a single heat bath (i.e. a large reservoir at equilibrium) into work.

This observation is equivalent to another equally empirical observation.

Law (Clausius' second law). Heat can never flow spontaneously (i.e. without work being performed) from a cold reservoir to a warmer reservoir.

This statement is actually a bit more subtle than it seems because, before we have defined temperature, we can only distinguish hotter and colder by looking at the direction of heat flow. What the second law says is that it is never possible to make heat flow spontaneously in the "wrong" direction. How do we get from such a seemingly trivial statement to something as abstract as entropy? This is most easily achieved by introducing the concept of a reversible heat engine.

A.2.1 Heat Engines

A reversible engine is, as the word suggests, an engine that can be operated in reverse. During one cycle (i.e. a sequence of steps that is completed when the engine is returned into its original state), this engine takes in an amount of heat q_1 from a hot reservoir, converts part of it into work w and delivers the remaining amount of heat q_2 to a cold reservoir. The reverse process is that, by performing an amount of work w, we can take an amount of heat q_2 from the cold reservoir and deliver an amount of heat q_1 to the hot reservoir. Reversible engines are an idealisation because in any real engine, there

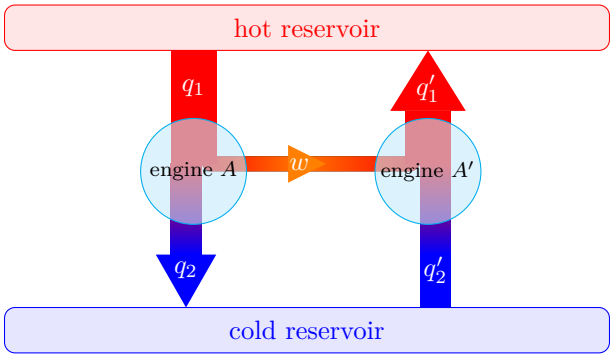


Figure A.1: If the efficiencies of the two heat engines are not equal, then this setup violates the second law of thermodynamics.

will be additional heat losses. However, the ideal reversible engine can be approximated arbitrarily closely by a real engine if, at every stage, the real engine is sufficiently close to equilibrium. As the engine is returned to its original state at the end of one cycle, its internal energy E has not changed. Hence, the first law tells us that $\Delta E = q_1 - (w + q_2) = 0$ or $q_1 = w + q_2$.

Now consider the efficiency of the engine $\eta \coloneqq w/q_1$, i.e. the amount of work delivered per amount of heat taken in. At first, one might think that η depends on the precise design of our reversible engine, but this turns out not to be the case: η is the same for all reversible engines operating between the same two reservoirs. To demonstrate this, suppose that we have another reversible engine with efficiency η' that takes in an amount of heat q_1' from the hot reservoir, delivers the same amount of work w, and then delivers an amount of heat q_2' to the cold reservoir, as shown in figure A.1. Suppose $\eta \neq \eta'$, then we use the work generated by the engine with the higher efficiency (say η) to drive the second engine (with efficiency η') in reverse. The amount of heat delivered to the hot reservoir by the second engine is

$$q_1' = \frac{w}{\eta'} = \frac{q_1 \eta}{\eta'} \,. \tag{A.2}$$

As $\eta' < \eta$ by construction, it follows that $q_1' > q_1$. Hence there is a net heat flow from the cold reservoir into the hot reservoir, which contradicts the second law of thermodynamics (in the form "heat can never spontaneously flow from a cold to a hot reservoir"). We have therefore proved by contradiction that the efficiency of all reversible heat engines operating between the same reservoirs is identical.

A.2.2 Absolute Temperatures

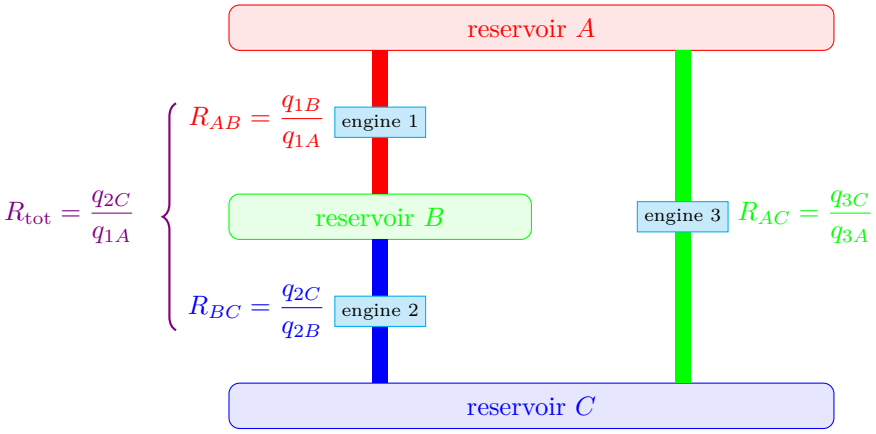
Since η does not depend on the specific design of a reversible engine, the only variables it can depend on are the temperatures t_1 and t_2 of the reservoirs it connects, i.e. $\eta = \eta(t_1, t_2)$. We are denoting these temperatures with a lowercase t for the time being because we have not yet defined precisely what we mean by the absolute temperature; the "temperatures" defined here merely need to be well-ordered (i.e. monotonic) and serve to identify which reservoir is colder and which is hotter.

For further convenience, we will also consider the quantity

$$R_{12}(t_1, t_2) \coloneqq 1 - \eta(t_1, t_2) = \frac{q_1 - w}{q_1} = \frac{q_2}{q_1} \,. \tag{A.3}$$

Suppose now that we have a reversible engine that consists of two stages: engine 1 working between reservoirs A and B, and engine 2 working between reservoirs B and C, as shown in the figure. The three reservoirs are at temperatures t_A, t_B and t_C, respectively. We have R_{AB} and R_{BC}

$$R_{AB} = \frac{q_{1B}}{q_{1A}} \,, \quad \text{and} \quad R_{BC} = \frac{q_{2C}}{q_{2B}} \,.$$



Since by construction, the combined engine $1+2$ is a reversible engine working in two stages, reservoir B must be neither a source nor a sink of energy, otherwise the overall efficiency of the engine would depend on how many such stages it is made up of, but we saw earlier that the efficiency of every reversible engine is the same. We must therefore conclude that $q_{1B} = q_{2B}$. The overall value of $R_{\text{tot}} = 1 - \eta_{\text{tot}}$ for the combined reversible engine is thus given by

$$R_{\text{tot}} = \frac{q_{2C}}{q_{1A}} = R_{AB} R_{BC} \,. \tag{A.4}$$

Suppose that in addition, we have another reversible engine, engine 3, that works directly between reservoirs A and C, with

$$R_{AC} = \frac{q_{3C}}{q_{3A}} \,. \tag{A.5}$$

Since the engines are reversible, engine 3 and the combined engine $1+2$ must be equally efficient, as we showed above. This means that

$$R_{AB} R_{BC} = R_{AC} \,. \tag{A.6}$$

However, recall that these ratios are functions solely of the temperatures of the reservoirs. The right-hand side does not depend on t_B, and so the product of R_{AB} and R_{BC} must therefore result in the cancellation of this functional dependence. The above equality can only hold in general provided that

$$R_{AB} = \frac{f(t_B)}{f(t_A)} \,, \qquad R_{BC} = \frac{f(t_C)}{f(t_B)} \quad \text{and} \quad R_{AC} = \frac{f(t_C)}{f(t_A)} \,, \tag{A.7}$$

where $f(t)$ is some function of our previously defined temperature t.

What we do next is to introduce an *absolute* (or *thermodynamic*) *temperature* T given by $T = f(t)$ so that

$$R_{\text{AB}} = \frac{q_B}{q_A} = \frac{T_B}{T_A} \,. \tag{A.8}$$

This well-defines the temperatures! We can first claim a certain object has a temperature T_0. Then to measure the temperature of another object, we only need to construct a reversible heat engine between them and measure the efficiency η. Then the temperature ratio of the two objects would be $R = 1 - \eta$.

The absolute temperature is defined up to an overall scaling constant. In practice, this is fixed such that 1 unit in the absolute (Kelvin) scale is equal to 1 degree Celsius. But that choice is of course purely historical.

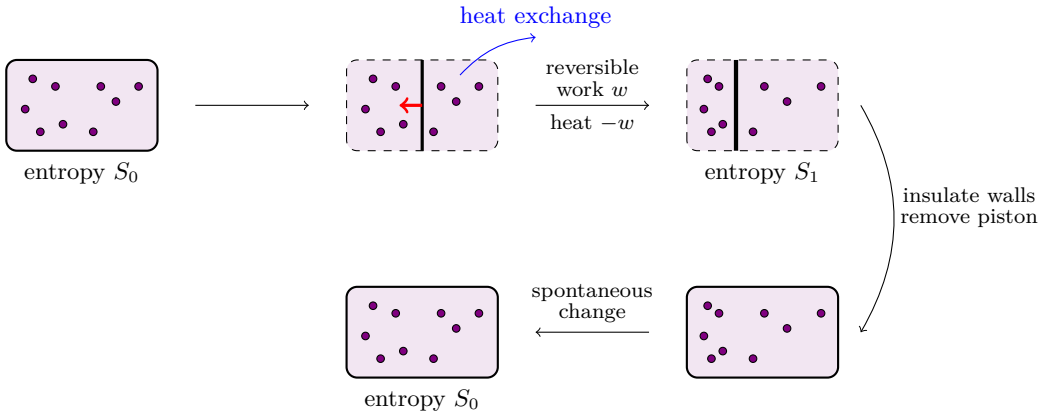


Figure A.2: To drive a system away from equilibrium, e.g. by changing the density, we need to do work. If we do this work reversibly and maintain the total energy of the system, this entails a loss of heat from the system. Hence $S_1 < S_0$. When we remove the constraint, the system spontaneously returns to equilibrium and its entropy is once again S_0. For spontaneous changes, $\Delta S > 0$.

A.2.3 Entropy

The reason we derived everything above is to introduce the entropy. To do so, we rewrite (A.8) as

$$\frac{q_A}{T_A} = \frac{q_B}{T_B}\,, \tag{A.9}$$

where q_A is the heat that flows in reversibly at the high temperature T_A, and q_B is the heat that flows out reversibly at the low temperature T_B. We see therefore that, during a complete cycle, the difference between q_A/T_A and q_B/T_B is zero. Recall that, at the end of a cycle, the internal energy of the system has not changed. Now (A.9) tells us that there is also another quantity that we call *entropy* and that we denote by S that is unchanged when we restore the system to its original state. In the language of thermodynamics, we call S a *state function*. We do not know what S is, but we do know how to compute its change. In the above example, the change in S was given by $\Delta S = (q_A/T_A) - (q_B/T_B) = 0$. In general, the change in entropy of a system due to the reversible addition of an infinitesimal amount of heat δq_{rev} from a reservoir at temperature T is

$$\delta S = \frac{\delta q_{\text{rev}}}{T}\,. \tag{A.10}$$

We also note that S is *extensive*. That means that the total entropy of two non-interacting systems is equal to the sum of the entropies of the individual systems. Consider a system with a fixed number of particles N and a fixed volume V. If we transfer an infinitesimal amount of heat δq to this system, then the change in the internal energy of the system, $\mathrm{d}E$, is equal to δq. Therefore

$$\left(\frac{\partial S}{\partial E}\right)_{V,N} = \frac{1}{T}\,. \tag{A.11}$$

From these, we can get the most famous, albeit not most intuitively obvious, statement of the second law of thermodynamics:

Law (Second law of thermodynamics). Any spontaneous change in a system that exchanges neither heat nor particles with its environment can never lead to a decrease of the entropy.

Hence, at equilibrium, the entropy of an isolated system is at a maximum.

As an example, let us consider a system with an energy E, volume V and number of particles N that is at equilibrium. The entropy of this system is $S_0(E, V, N)$. Suppose we wish to change

something in this system; for instance, we increase the density in one half and decrease it in the other. As the system was at equilibrium, this change does not occur spontaneously; in order to effect this change, we must perform a certain amount of work w, for instance by placing a piston in the system and moving it. Let us perform this work reversibly in such a way that E, the total energy of the system, stays constant, as do V and N. The first law tells us that we can only keep E constant if, while we do the work, we allow an amount of heat q to flow out of the system, such that $-q = w$. However, when an amount of heat $-q$ flows out of the system, (A.10) tells us that the entropy S of the system must decrease: $S_1(E, V, N) < S_0(E, V, N)$.

Having completed the change in the system, we insulate the system thermally from the rest of the world, and we remove the constraint that kept the system in its special constrained state. In our piston example, this might entail making a hole in the piston separating the two parts of the system. Now the system goes back spontaneously and irreversibly to equilibrium. However, since the volume is fixed, no work is done, and since the system is thermally insulated, no heat is transferred. Hence the final energy E is equal to the original energy, and V and N are also unchanged. This means that the system is now back in its original equilibrium state and its entropy is once more equal to $S_0(E, V, N)$.

The entropy change during this spontaneous change is equal to $\Delta S = S_0 - S_1$. But, as $S_1 < S_0$, it follows that $\Delta S > 0$.

A.3 The Third Law

From this point on, we can derive all of thermodynamics, except one law — the so-called third law of thermodynamics.

Law (Third law of thermodynamics). At $T = 0$, the entropy of the equilibrium state of a pure, perfectly crystalline substance equals zero.

The third law is not nearly as fundamental as the first two, and in any case, as discussed in the main text, statistical mechanics gives us a more direct interpretation of its meaning.

A.4 Fundamental Equations

From the first law of thermodynamics, we write, in infinitesimal form

$$\mathrm{d}E_{\mathrm{rev}} = đq_{\mathrm{rev}} + đw_{\mathrm{rev}}\,. \tag{A.12}$$

If the system only does reversible PV work, then $đw_{\mathrm{rev}} = -P\mathrm{d}V$. For reversible processes, the *Clausius inequality*, $\mathrm{d}S \geq đq/T$ is an equality (A.10); this is a statement of the second law of thermodynamics. Thus we can write $\mathrm{d}E_{\mathrm{rev}} = T\mathrm{d}S - P\mathrm{d}V$. Since the internal energy is a state function, this equation applies between any equilibrium state, not just along reversible pathways. The equation is known as the *fundamental equation* for the internal energy.

A.4.1 Legendre Transforms

Consider a general function $f(x, y)$, whose total differential is

$$\mathrm{d}f = \left(\frac{\partial f}{\partial x}\right)_y \mathrm{d}x + \left(\frac{\partial f}{\partial y}\right)_x \mathrm{d}y = u\mathrm{d}x + v\mathrm{d}y\,. \tag{A.13}$$

Since u is the derivative of f with respect to x, we say that u and x are *conjugate variables*. Then, consider the function $F = f - (ux)$, whose differential is

$$\mathrm{d}F = \mathrm{d}f - \mathrm{d}(ux) = v\mathrm{d}y - x\mathrm{d}u\,. \tag{A.14}$$

The function F is known as the *Legendre transform* of f. Evidently, F remains a function of y, but is now a function of u, the conjugate pair of x, rather than of x itself.

To see how this formalism can be applied to thermodynamics, consider the *enthalpy*. When you were first introduced to it in Part IA, we said that we wanted a state function corresponding to the heat flow at constant pressure. But how do we know how to define it? The enthalpy is the appropriate measure of "energy" for a system at constant pressure rather than constant volume, so we would like to convert $E = E(S, V)$ into $H = H(S, P)$ so that P is a natural variable of our function H that we can control, and V is a derived property. The enthalpy is the Legendre transform of E, defined as

$$H = E + PV \tag{A.15}$$

so that

$$\mathrm{d}H = T\mathrm{d}S + V\mathrm{d}P\,. \tag{A.16}$$

Similarly, we may change the natural variable from S to T for internal energy and enthalpy, giving *Helmholtz free energy* $A = A(T, V)$ and *Gibbs free energy* $G = G(T, P)$ defined by Legendre transforms

$$A = E - TS\,, \tag{A.17}$$

$$G = H - TS\,, \tag{A.18}$$

so that we have the fundamental equations

$$\mathrm{d}A = -S\mathrm{d}T - P\mathrm{d}V \tag{A.19}$$

$$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}P\,. \tag{A.20}$$

The above relations are for the cases where the number of particles N is fixed. If the number of particles can also change, we need to add a further term $+\mu\mathrm{d}N$ to the above fundamental equations, where μ is the *chemical potential*. This allows us to define further Legendre transforms to change the natural variables from N to μ. The one we are interested in is the *Grand potential*

$$\Phi = E - TS - N\mu\,, \tag{A.21}$$

so that

$$\mathrm{d}\Phi = -S\mathrm{d}T - P\mathrm{d}V - N\mathrm{d}\mu\,. \tag{A.22}$$

This completes the definition of the five most common *thermodynamic potentials*: E, H, A, G, Φ. More thermodynamic potentials can be defined by doing further Legendre transforms, but they are of less interest.

A.4.2 Maxwell Relations

Schwarz's theorem guarantees that if $f = f(x, y, \dots)$ have continuous second partial derivatives, then

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\,. \tag{A.23}$$

Since the fundamental equations are exact differentials, we can apply the Schwarz's theorem to our thermodynamic potentials. For example, from the fundamental equation of the internal energy

$$\mathrm{d}E = T\mathrm{d}S - P\mathrm{d}V + \mu\mathrm{d}N\,, \tag{A.24}$$

we have

$$\left[\frac{\partial}{\partial V}\left(\frac{\partial E}{\partial S}\right)_V\right]_S = \left[\frac{\partial}{\partial S}\left(\frac{\partial E}{\partial V}\right)_S\right]_V\,, \tag{A.25}$$

which gives the relation

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial P}{\partial S}\right)_V\,. \tag{A.26}$$

We can get a lot more relations like this from the fundamental equations of different thermodynamic potentials, using the symmetry of partial derivatives with respect to different variables. They are called the *Maxwell relations*. They are not worth memorising as they can be easily derived.

A.4.3 Euler Relations

Since all natural variables of the internal energy E are extensive quantities,

$$\alpha E(N,V,S) = E(\alpha N, \alpha V, \alpha S) \tag{A.27}$$

for $\alpha \in \mathbb{R}$. The total differential is now

$$\mathrm{d}(\alpha E) = \left(\frac{\partial(\alpha E)}{\partial(\alpha N)}\right)_{\alpha V,\alpha S} \mathrm{d}(\alpha N) + \left(\frac{\partial(\alpha E)}{\partial(\alpha V)}\right)_{\alpha N,\alpha S} \mathrm{d}(\alpha V) + \left(\frac{\partial(\alpha E)}{\partial(\alpha S)}\right)_{\alpha N,\alpha V} \mathrm{d}(\alpha S)\,. \tag{A.28}$$

Differentiating both sides with respect to α gives

$$\begin{aligned}\left(\frac{\partial(\alpha E)}{\partial \alpha}\right)_{N,V,S} &= \left(\frac{\partial(\alpha E)}{\partial(\alpha N)}\right)_{\alpha V,\alpha S}\left(\frac{\partial(\alpha N)}{\partial \alpha}\right)_{N,V,S} \\ &\quad + \left(\frac{\partial(\alpha E)}{\partial(\alpha V)}\right)_{\alpha N,\alpha S}\left(\frac{\partial(\alpha V)}{\partial \alpha}\right)_{N,V,S} \\ &\quad + \left(\frac{\partial(\alpha E)}{\partial(\alpha S)}\right)_{\alpha N,\alpha V}\left(\frac{\partial(\alpha S)}{\partial \alpha}\right)_{N,V,S}\,.\end{aligned} \tag{A.29}$$

Simplify this derivative and set $\alpha = 1$ gives

$$\begin{aligned}E &= \left(\frac{\partial E}{\partial N}\right)_{V,S} N + \left(\frac{\partial E}{\partial V}\right)_{N,S} V + \left(\frac{\partial E}{\partial S}\right)_{N,V} S \\ &= \mu N - PV + TS\,.\end{aligned} \tag{A.30}$$

This is the *Euler relation*.[18]

Substituting the Euler relation into the expressions for the other main potentials gives

$$A = -PV + \mu N \tag{A.31}$$
$$H = TS + \mu N \tag{A.32}$$
$$G = \mu N \tag{A.33}$$
$$\Phi = -PV\,. \tag{A.34}$$

This is essentially why we said the chemical potential μ is equal to the Gibbs energy per particle.

We may differentiate $G = \mu N$ to get $\mathrm{d}G = N\mathrm{d}\mu + \mu \mathrm{d}N$. From the fundamental equation $\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}P + \mu \mathrm{d}N$, we can get the *Gibbs–Duhem relation*

$$N\mathrm{d}\mu = V\mathrm{d}P - S\mathrm{d}T\,. \tag{A.35}$$

B From Quantum to Classical

It is possible to derive the classical partition function directly from the quantum partition function without resorting to hand-waving. It will also show us why the factor of $1/h$ sits outside the partition function. The derivation is a little tedious, but worth seeing. (Similar techniques are useful in later courses when you first meet the path integral). To make life easier, let's consider a single particle moving in one spatial dimension. It has position operator $\hat{x}$, momentum operator $\hat{p}$ and Hamiltonian

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{x})\,. \tag{B.1}$$

[18] It was not discovered by Euler in an investigation of thermodynamics, which did not exist in his day. It is called the Euler relation just because Euler's theorem on homogeneous functions leads to it.

If $|n\rangle$ is the energy eigenstate with energy E_n, the quantum partition function is

$$Q = \sum_n \mathrm{e}^{-\beta E_n} = \sum_n \left\langle n \middle| \mathrm{e}^{-\beta \hat{H}} \middle| n \right\rangle , \tag{B.2}$$

where the exponential of an operator $\hat{X}$ is defined via the expansion

$$\mathrm{e}^{\hat{X}} = \sum_{n=0}^{\infty} \frac{\hat{X}^n}{n!} . \tag{B.3}$$

In what follows, we'll make liberal use of the fact that we can insert the identity operator anywhere in this expression. We have the resolution of the identities in the position eigenstates $|x\rangle$ and the momentum eigenstates $|p\rangle$:

$$1 = \int \mathrm{d}x \; |x\rangle \langle x| \;\text{ and }\; 1 = \int \mathrm{d}p \; |p\rangle \langle p| . \tag{B.4}$$

We insert two copies of the identity built from position eigenstates into our partition function to get

$$\begin{aligned} Q &= \sum_n \left\langle n \middle| \int \mathrm{d}x \, \middle| x \right\rangle \left\langle x \middle| \mathrm{e}^{-\beta \hat{H}} \int \mathrm{d}x' \, \middle| x' \right\rangle \langle x'|n\rangle \\ &= \int \mathrm{d}x \, \mathrm{d}x' \left\langle x \middle| \mathrm{e}^{-\beta \hat{H}} \middle| x' \right\rangle \sum_n \langle x'|n\rangle \langle n|x\rangle . \end{aligned} \tag{B.5}$$

And now $\sum_n |n\rangle \langle n|$ is the identity, while $|x'\rangle \langle x| = \delta(x - x')$, so

$$Q = \int \mathrm{d}x \left\langle x \middle| \mathrm{e}^{-\beta \hat{H}} \middle| x \right\rangle . \tag{B.6}$$

We see that the result is to replace the sum over energy eigenstates with a sum (or integral) over position eigenstates. If you wanted, you could play the same game and get the sum over any complete basis of eigenstates of your choosing. This means that we can write the partition function in a basis independent fashion as

$$Q = \operatorname{tr} \mathrm{e}^{-\beta \hat{H}} . \tag{B.7}$$

So far, our manipulations could have been done for any quantum system. Now we want to use the fact that we are taking the classical limit. This comes about when we try to factorize $\mathrm{e}^{-\beta H}$ into a momentum term and a position term — the trouble is that this isn't always possible when there are matrices (or operators) in the exponent. The Hamiltonian $\hat{H}$ is the sum of a kinetic part $\hat{T}$ and a potential part $\hat{V}$. The kinetic energy eigenstates are also the momentum eigenstates, but potential energy eigenstates are the position eigenstates, so $\hat{H} = \hat{T} + \hat{V}$ is not diagonal in either basis i.e. $\hat{T}$ and $\hat{V}$ does not commute. In fact, we have

$$\mathrm{e}^{\hat{A}} \mathrm{e}^{\hat{B}} = \mathrm{e}^{\hat{A} + \hat{B} + \frac{1}{2}[\hat{A}, \hat{B}] + \cdots} . \tag{B.8}$$

For us $[x, p] = i\hbar$. This means that if we're willing to neglect terms of order $\hbar$ — which is the meaning of taking the classical limit — then we can write

$$\mathrm{e}^{-\beta \hat{H}} = \mathrm{e}^{-\beta \hat{T}} \mathrm{e}^{-\beta \hat{V}} = \mathrm{e}^{-\beta \hat{p}^2/2m} \mathrm{e}^{-\beta V(\hat{x})} . \tag{B.9}$$

We can now start to replace some of the operators in the exponent, like $V(\hat{x})$, with functions $V(x)$. (Note the subtle but important notational difference!)

$$\begin{aligned} Q &= \int \mathrm{d}x \left\langle x \middle| \mathrm{e}^{-\beta V(\hat{x})} \mathrm{e}^{-\beta \hat{p}^2/2m} \middle| x \right\rangle \\ &= \int \mathrm{d}x \, \mathrm{d}x' \, \mathrm{d}r \, \mathrm{d}r' \left\langle x \middle| \mathrm{e}^{-\beta \hat{V}(\hat{x})} \middle| x' \right\rangle \langle x'|p\rangle \left\langle p \middle| \mathrm{e}^{-\beta \hat{p}^2/2m} \middle| p' \right\rangle \langle p'|x\rangle \\ &= \int \mathrm{d}x \, \mathrm{d}p \left\langle x \middle| \mathrm{e}^{-\beta \hat{V}(\hat{x})} \middle| x \right\rangle \langle x|p\rangle \left\langle p \middle| \mathrm{e}^{-\beta \hat{p}^2/2m} \middle| p \right\rangle \langle p|x\rangle \\ &= \int \mathrm{d}x \, \mathrm{d}p \, |\langle x|p\rangle|^2 \, \mathrm{e}^{-\beta p^2/2m} \mathrm{e}^{-\beta V(x)} , \end{aligned} \tag{B.10}$$

where we have used the fact that all matrix elements are diagonal. We have that

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}} \mathrm{e}^{\mathrm{i}px/\hbar} \tag{B.11}$$

is the wavefunction of the momentum eigenstate, so

$$Q = \frac{1}{h} \int \mathrm{d}x \, \mathrm{d}p \, \mathrm{e}^{-\beta H(x,p)} \, . \tag{B.12}$$

C Pressure from Clausius Virial

Very often, there is not just one way of deriving a particular result. In the main text, we derived an expression for the pressure of interacting particles in terms of the inter-particle forces directly from the partition function by using a scaling argument. We show here a derivation using a completely different line of reasoning, originally proposed by Clausius.

Consider a system of N particles in a volume V at total energy E. Define the *virial* of the system to be

$$\mathcal{V} \coloneqq \sum_{i=1}^{N} \mathbf{r}_i \cdot \mathbf{p}_i \, . \tag{C.1}$$

This is one half of the time derivative of the moment of inertia I of the system since

$$\frac{1}{2}\frac{\partial I}{\partial t} = \frac{1}{2} \sum_{i=1}^{N} m_i \frac{\partial}{\partial t} \mathbf{r}_i \cdot \mathbf{r}_i = \sum_{i=1}^{N} m_i \mathbf{r}_i \cdot \dot{\mathbf{r}}_i = \mathcal{V} \, . \tag{C.2}$$

Since the particles are constrained in a container, and the total kinetic energy is bounded by E, $\mathcal{V}$ is also bounded both above and below, by $\mathcal{V}_{\max}$ and $\mathcal{V}_{\min}$. At long times $\tau \to \infty$, the time average of the time derivative of $\mathcal{V}$, denoted by $\dot{\mathcal{V}}$, vanishes:

$$\overline{\dot{\mathcal{V}}} = \frac{1}{\tau} \int_0^\tau \dot{\mathcal{V}}(t) \mathrm{d}t \, = \frac{\mathcal{V}(\tau) - \mathcal{V}(0)}{\tau} \to 0 \text{ as } \tau \to \infty \, . \tag{C.3}$$

We can also evaluate this time derivative explicitly:

$$\overline{\dot{\mathcal{V}}} = \sum_{i=1}^{N} \overline{\dot{\mathbf{r}}_i \cdot \mathbf{p}_i} + \overline{\mathbf{r}_i \cdot \dot{\mathbf{p}}_i} = 0 \, . \tag{C.4}$$

We have $\dot{\mathbf{r}}_i = \mathbf{p}_i / m_i$ and $\dot{\mathbf{p}}_i = \mathbf{f}_i$ by Newton's second law, where $\mathbf{f}_i$ is the force on the particle i, so we can write

$$\sum_{i=1}^{N} \frac{\overline{p_i^2}}{m_i} + \overline{\mathbf{r}_i \cdot \mathbf{f}_i} = 0 \, . \tag{C.5}$$

The first term on the left is simply twice the kinetic energy of the system. Using the equipartition theorem, we can write this as $3Nk_BT$. We thus find that

$$3Nk_BT + \sum_{i=1}^{N} \overline{\mathbf{r}_i \cdot \mathbf{f}_i} = 0 \, . \tag{C.6}$$

The force $\mathbf{f}_i$ acting on a particle can originate either from the intermolecular interactions or interactions with the wall of the container. Let us consider the contribution to the virial due to the interaction with a surface element $\mathrm{d}\mathbf{S}$ of the wall of the container. We denote the position of this surface element by $\mathbf{r}_S$. Whenever a particle collides with this surface element, its coordinate must be

given by $\mathbf{r}_S$. The contribution to the force virial arising from collisions with this surface element is thus given by

$$\overline{\left(\sum_{i=1}^{N}\mathbf{r}_i\cdot\mathbf{f}_i\right)^{(\mathrm{d}\mathbf{S})}} = \sum_{i=1}^{N}\overline{\mathbf{r}_S\cdot\mathbf{f}_i^{(\mathrm{d}\mathbf{S})}} = \mathbf{r}_S\cdot\sum_{i=1}^{N}\overline{\mathbf{f}_i^{(\mathrm{d}\mathbf{S})}} = \mathbf{r}_S\cdot\overline{\mathbf{F}_{\mathrm{tot}}^{\mathrm{d}\mathbf{S}}}\,. \tag{C.7}$$

$\overline{\mathbf{F}_{\mathrm{tot}}^{\mathrm{d}\mathbf{S}}}$ is the average total force exerted on the particles by the wall of the container — by Newton's third law, this is also the force exerted on the container wall by the particles. This is exactly the force causing the pressure! Therefore we have

$$\overline{\left(\sum_{i=1}^{N}\mathbf{r}_i\cdot\mathbf{f}_i\right)^{(\mathrm{d}\mathbf{S})}} = \mathbf{r}_S\cdot(-P)\mathrm{d}\mathbf{S}\,. \tag{C.8}$$

We can now integrate over the entire surface of the container to obtain

$$\sum_{i=1}^{N}\overline{\mathbf{r}_i\cdot\mathbf{f}_i} = -P\int\mathrm{d}\mathbf{S}\cdot\mathbf{r}_S\,. \tag{C.9}$$

Using the divergence theorem, we can write

$$-P\int\mathrm{d}\mathbf{S}\cdot\mathbf{r}_S = -P\int\mathrm{d}V\,(\boldsymbol{\nabla}\cdot\mathbf{r}) = -3PV\,. \tag{C.10}$$

Therefore,

$$3Nk_BT + \sum_{i=1}^{N}\overline{\mathbf{r}_i\cdot\mathbf{f}_i^{\mathrm{inter}}} - 3PV = 0\,. \tag{C.11}$$

We get the same result

$$P = \frac{Nk_BT}{V} + \frac{1}{3V}\left\langle\sum_{i=1}^{N}\mathbf{r}_i\cdot\mathbf{f}_i\right\rangle\,, \tag{C.12}$$

where notice that we have replaced the time average by a Boltzmann-weighted average over states at constant N, V and T.

D Ising and Lattice-Gas Model Equivalence

The Ising model for a lattice of interacting magnetic spins is amongst the most widely used models in physics because many other models can be mapped onto it. As an example, we consider the relation between the Ising model and a lattice gas. The lattice gas model is a simple model used to describe the liquid-vapour transition. Let us first consider the Ising model, in which neighbouring parallel spins have an interaction energy $-J$, whilst anti-parallel spins have an interaction energy $+J$. We denote the value of the spin on lattice site i by $s_i \in \{\pm 1\}$. The Ising Hamiltonian in the presence of a magnetic field B is given by

$$H_{\mathrm{Ising}} = -\frac{J}{2}\sum_{i=1}^{N}\sum_{\langle i,j\rangle}s_is_j - B\sum_{i=1}^{N}s_i\,. \tag{D.1}$$

The partition function of the Ising model can be written as

$$Q_{\mathrm{Ising}} = \sum_{(s_k)_{k=1}^{N}}\exp\left[-\beta\left(-\frac{J}{2}\sum_{i=1}^{N}\sum_{\langle i,j\rangle}s_is_j - B\sum_{i=1}^{N}s_i\right)\right]\,, \tag{D.2}$$

where $(s_k)_{k=1}^{N}$ denote all possible arrangements of spin values $(s_1,\ldots,s_N)$.

Now consider a gas on a lattice with N sites, where there can be at most one particle per lattice site. Neighbouring particles have an interaction energy $-\epsilon$. The Hamiltonian is

$$H_{\mathrm{LG}} = -\frac{\epsilon}{2}\sum_{i=1}^{N}\sum_{\langle i,j\rangle} n_i n_j \,, \tag{D.3}$$

where $n_i \in \{0,1\}$ denotes the number of particles on site i. The partition function is

$$Q_{\mathrm{LG}}(M) = \sum_{(n_k)_{k=1}^N|_M} \exp\left[-\beta\left(-\frac{\epsilon}{2}\sum_{i=1}^{N}\sum_{\langle i,j\rangle} n_i n_j\right)\right] , \tag{D.4}$$

where M is the number of particles and $(n_k)_{k=1}^N|_M$ is the arrangement of occupation numbers that satisfy $\sum_i n_i = M$. It turns out to be more convenient to consider the grand partition function Ξ instead to remove this awkward constraint.

$$\begin{aligned}\Xi &= \sum_{M=0}^{N}\sum_{(n_k)_{k=1}^N|_M} \exp\left[-\beta\left(-\frac{\epsilon}{2}\sum_{i=1}^{N}\sum_{\langle i,j\rangle} n_i n_j\right)\right] \mathrm{e}^{\beta\mu M} \\ &= \sum_{(n_k)_{k=1}^N} \exp\left[-\beta\left(-\frac{\epsilon}{2}\sum_{i=1}^{N}\sum_{\langle i,j\rangle} n_i n_j - \mu\sum_{i=1}^{N} n_i\right)\right] . \end{aligned} \tag{D.5}$$

This looks remarkably similar to the canonical partition function of the Ising model (D.2) — it seems that we only need to change the letters and (D.5) will become (D.2). But there is actually an important difference: s_i takes value in $\{\pm 1\}$, while n_i takes value in $\{0,+1\}$. We have to identify $s = 2n - 1$ to make the exact mapping between the two cases. On doing so, we also need to identify

$$J = \frac{\epsilon}{4} \text{ and } B = \frac{\mu}{2} + \frac{q\epsilon}{4} \,. \tag{D.6}$$

After doing so, the grand partition function of this lattice gas will be exactly the same as the canonical partition function of an Ising model in a magnetic field. This probably makes the universality we introduced in the main text less shocking than it first seems.

E More on Interacting Gases

In this section, we will re-do some of our previous calculations in a more unified approach, and extend the results to the beautiful and important cluster expansions.[19]

E.1 Mayer f Functions and Second Virial Coefficients

The Hamiltonian of the interacting gas is

$$H = \sum_{i=1}^{N}\frac{p_i^2}{2m} + \sum_{i>j} U(r_{ij}) \,. \tag{E.1}$$

We would like to find a general method calculating the coefficients for the *virial expansion*

$$\frac{p}{k_B T} = \frac{N}{V} + B_2(T)\frac{N^2}{V^2} + B_3(T)\frac{N^3}{V^3} + \dots \tag{E.2}$$

[19] This section is extremely mathematically demanding. This is included in my notes because of my personal interests. This is not in the official notes.

As we've seen before, the partition function is

$$Q(N,V,T) = \frac{1}{h^{3N}N!}\int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{p}_i\,\mathrm{d}^3\mathbf{r}_i\,\mathrm{e}^{-\beta H} = \frac{1}{\Lambda^{3N}N!}\int \prod_{i} \mathrm{d}^3\mathbf{r}_i\,\mathrm{e}^{-\beta\sum_{j<k}U(r_{jk})}\,. \tag{E.3}$$

We still need to do the integral over positions, which looks hard. The interactions mean that the integrals don't factor in any obvious way.

An obvious thing to try is the Taylor expansion, which should better be called a *cumulant expansion* in this context.

$$\mathrm{e}^{-\beta\sum_{j<k}U(r_{jk})} = 1 - \beta\sum_{j<k}U(r_{jk}) + \frac{\beta^2}{2}\sum_{j<k,l<m}U(r_{jk})U(r_{lm}) + \dots \tag{E.4}$$

Unfortunately, this is not very useful. We want each term to be smaller than the preceding one. But as $r_{ij}\to 0$, the potential $U(r_{ij})\to\infty$, which does not look promising for an expansion parameter.

We will instead work with the following quantity called the *Mayer f function*.

$$f(r) \coloneqq \mathrm{e}^{-\beta U(r)} - 1\,. \tag{E.5}$$

This is a nicer expansion parameter. When the particles are far, $f(r)\to 0$ and when particles are close, $f(r)\to -1$.

Let us define

$$f_{ij} = f(r_{ij})\,, \tag{E.6}$$

then we can write the partition function as

$$Q(N,V,T) = \frac{1}{\Lambda^{3N}N!}\int \prod_{i} \mathrm{d}^3\mathbf{r}_i \prod_{j>k}(1+f_{jk}) = \frac{1}{\Lambda^{3N}N!}\int \prod_{i} \mathrm{d}^3\mathbf{r}_i \left(1 + \sum_{j>k} f_{jk} + \sum_{j>k,l>m} f_{jk}f_{lm} + \dots\right). \tag{E.7}$$

The first term simply gives a factor of the volume V for each integral, so we get V^N. The second term has a sum, each element of which is the same. They all look like

$$\int \prod_{i=1}^{N} \mathrm{d}^3\mathbf{r}_i\, f_{12} = V^{N-2}\int \mathrm{d}^3\mathbf{r}_1\,\mathrm{d}^3\mathbf{r}_2\, f(r_{12}) = V^{N-1}\int \mathrm{d}^3\mathbf{r}\, f(r)\,, \tag{E.8}$$

where, in the last equality, we have changed the integration variables from $\mathbf{r}_1$ and $\mathbf{r}_2$ to the centre of mass $\mathbf{R} = \frac{1}{2}(\mathbf{r}_1+\mathbf{r}_2)$ and the separation $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. You might worry that the limits of integration change in the integral over $\mathbf{r}$, but the integral over $f(r)$ only picks up contributions from atomic size distances and this is only actually a problem close to the boundaries of the system where it is negligible.

There is a term like this for each pair of particles, so there are in total $\frac{1}{2}N(N-1)\approx\frac{1}{2}N^2$ terms. Then ignoring terms $O(f^2)$, the partition function is given approximately by

$$Q(N,V,T) = \frac{V^N}{\Lambda^{3N}N!}\left(1 + \frac{N^2}{2V}\int \mathrm{d}^3\mathbf{r}\, f(r) + \dots\right) = Q_{\text{ideal}}\left(1 + \frac{N}{2V}\int \mathrm{d}^3\mathbf{r}\, f(r) + \dots\right)^N, \tag{E.9}$$

where we promoted one power of N from the front of the integral to an overall exponent. Massaging the expression in this way ensures that the free energy is proportional to the number of particles as one would expect:

$$A = -k_B T \ln Q = A_{\text{ideal}} - N k_B T \ln\left(1 + \frac{N}{2V}\int \mathrm{d}^3\mathbf{r}\, f(r)\right) . \tag{E.10}$$

From this expression for the free energy, it is clear that we are indeed performing an expansion in density of the gas since the correction term is proportional to N/V. This form of the free energy will give us the second virial coefficient $B_2(T)$.

We can be somewhat more precise about what it means to be at low density in which our expression is valid. The exact form of the integral $\int \mathrm{d}^3\mathbf{r}\, f(r)$ depends on the potential, but for both the Lennard Jones potential and the hard-core repulsion the integral is approximately $\int \mathrm{d}^3\mathbf{r}\, f(r) \sim r_0^3$, where r_0 some measure of the length scale of the potential (We will compute this later). For the expansion to be valid, we want each term with an extra power of f to be smaller than the preceding one. That means that the second term in the argument of the log should be smaller than 1. In other words,

$$\frac{N}{V} \ll \frac{1}{r_0^3} . \tag{E.11}$$

We have a name for the substance packed as closely as in the right hand side — we call them *liquids*. Our expansion is valid for densities of the gas that are much lower than that of the liquid state.

E.2 van der Waals Equation of State

We can use the free energy equation (E.10) to compute the pressure of the gas. Expanding the log gives

$$P = -\frac{\partial A}{\partial V} = \frac{N k_B T}{V}\left(1 - \frac{N}{2V}\int \mathrm{d}^3\mathbf{r}\, f(r) + \dots\right) . \tag{E.12}$$

As expected, the pressure deviates from that of an ideal gas. We can characterise this by writing

$$\frac{PV}{N k_B T} = 1 - \frac{N}{2V}\int \mathrm{d}^3\mathbf{r}\, f(r) . \tag{E.13}$$

To understand this, we need to calculate the integral. But first, let us look at two trivial examples.

- Repulsion. Suppose that $U(r) > 0$ for all r with $U(\infty) = 0$. Then $f < 0$ and so the pressure will increase.
- Attraction. If $U(r) < 0$ then $f > 0$ and so the pressure will decrease.

Let's try this with some different potentials. We can define a hard core potential with van der Waals attraction

$$U(r) = \begin{cases} \infty & r < r_0 \\ -U_0\left(\frac{r_0}{r}\right)^6 & r \geq r_0 . \end{cases} \tag{E.14}$$

The integral of the Mayer f function is

$$\int \mathrm{d}^3\mathbf{r}\, f(r) = \int_0^{r_0} \mathrm{d}^3\, r(-1) + \int_{r_0}^{\infty} \mathrm{d}^3\mathbf{r}\, \left(\mathrm{e}^{\beta U_0 (r_0/r)^6} - 1\right) . \tag{E.15}$$

We will approximate the second integral in the high temperature limit with $\beta U_0 \ll 1$, where $\mathrm{e}^{\beta U_0 (r_0/r)^6} \approx 1 + \beta U_0 (r_0/r)^6$. Then

$$\begin{aligned} \int \mathrm{d}^3\mathbf{r}\, f(r) &= -4\pi \int_0^{r_0} \mathrm{d}r\, r^2 + \frac{4\pi U_0}{k_B T}\int_{r_0}^{\infty} \mathrm{d}r\, \frac{r_0^6}{r^4} \\ &= \frac{4\pi r_0^3}{3}\left(\frac{U_0}{k_B T} - 1\right) . \end{aligned} \tag{E.16}$$

Insert this into equation (E.13), we have

$$\frac{PV}{Nk_BT} = 1 - \frac{N}{V}\left(\frac{a}{k_BT} - b\right) , \tag{E.17}$$

where a and b are constants defined by

$$a = \frac{2\pi r_0^3 U_0}{3} \quad , \quad b = \frac{2\pi r_0^3}{3} . \tag{E.18}$$

We can recognise that this is indeed the second virial coefficient.

We can rearrange this equation to a more useful form

$$k_BT = \frac{V}{N}\left(P + \frac{N^2}{V^2}a\right)\left(1 + \frac{N}{V}b\right)^{-1} . \tag{E.19}$$

As we are working in an expansion in density, we can do another Taylor expansion to get

$$k_BT = \left(P + \frac{N^2}{V^2}a\right)\left(\frac{V}{N} - b\right) . \tag{E.20}$$

This again gives the famous van der Waals equation of state. This equation is valid at low densities and high temperatures.

It seems that the equation we obtained (E.13) can be used to compute the equation of state for any potential between atoms. However, there are limitations. The integral only converges for the long-range potentials in the form $1/r^n$ when $n \geq 4$. This means that the techniques described above do not work for long-range potentials with fall-off $1/r^3$ or slower. This includes the important case of $1/r$ Coulomb interactions.

E.3 The Cluster Expansion

We will now develop the full expansion and explain how to compute the higher virial coefficients.

Let us go back to our first expression of the partition function in terms of f.

$$\begin{aligned} Q(N,V,T) &= \frac{1}{\Lambda^{3N}N!}\int \prod_i \mathrm{d}^3\mathbf{r}_i \prod_{j>k}(1 + f_{jk}) \\ &= \frac{1}{\Lambda^{3N}N!}\int \prod_i \mathrm{d}^3\mathbf{r}_i \left(1 + \sum_{j>k} f_{jk} + \sum_{j>k,l>m} f_{jk}f_{lm} + \dots\right) . \end{aligned} \tag{E.21}$$

Above we effectively related the second virial coefficient to the term linear in f: this is the essence of the equation of state equation (E.13). One might think that terms quadratic in f give rise to the third virial coefficient and so on. But we will see that the expansion is more subtle than that.

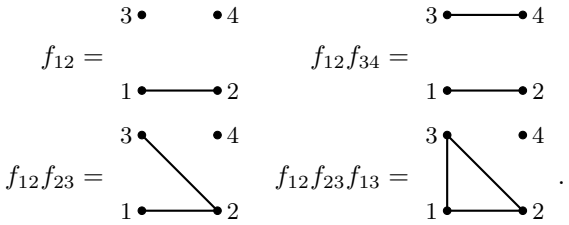
The expansion in equation (E.21) includes terms of the form $f_{ij}f_{kl}f_{mn}\dots$, where the indices denote pairs of atoms, (i,j) and (k,l) and so on. These pairs may have atoms in common or they may all be different. However, the same pair never appears twice in a given term as you may check by going back to the first line in equation (E.21).

We will introduce a diagrammatic method to keep track of all the terms in the sum. To each term of the form $f_{ij}f_{kl}f_{mn}\dots$ we associate a picture using the following rules:

- Draw N atoms (We actually need pictures with much less than $N \sim 10^{23}$ atoms).
- Draw a line to connect each pair of atoms that appear as indices.

This allows us to never repeat the same pair of indices.

For example, if $N = 4$, we may have the following pictures for different terms in the expansion:



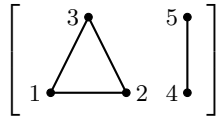
(E.22)

We call these diagrams *graphs*. Each possible graph appears exactly once in the partition function equation (E.21). In other words, the partition function is a sum over all graphs. We still have to do the integrals over all positions r_i. We will denote the integral over graph G to be $W[G]$. Then the partition function is

$$Q(N, V, T) = \frac{1}{\Lambda^{3N} N!} \sum_G W[G] \,. \tag{E.23}$$

Nearly all the graphs that we can draw will have disconnected components. For example, those graphs that correspond to just a single f_{ij} will have two atoms connected and the remaining $N - 2$ sitting alone. Those graphs that correspond to $f_{ij}f_{kl}$ fall into two categories: either they consist of two pairs of atoms (like the second example above) or, if (i, j) shares an atom with (k, l), there are three linked atoms (like the third example above). Importantly, the integral over positions $\mathbf{r}_i$ then factorises into a product of integrals over the positions of atoms in disconnected components. This is illustrated by an example with $N = 5$ atoms

$$W\left[\;\right] = \left(\int \mathrm{d}^3\mathbf{r}_1\, \mathrm{d}^3\mathbf{r}_2\, \mathrm{d}^3\mathbf{r}_3\, f_{12}f_{23}f_{31}\right)\left(\int \mathrm{d}^3\mathbf{r}_4\, \mathrm{d}^3\mathbf{r}_5\, f_{45}\right) . \tag{E.24}$$



We call the disconnected components of the graph *clusters*. If a cluster has ℓ atoms, we will call it an ℓ-cluster. The $N = 5$ example above has a single 3-cluster and a single 2-cluster. In general, a graph G will split into m_ℓ ℓ-clusters. Clearly, we must have

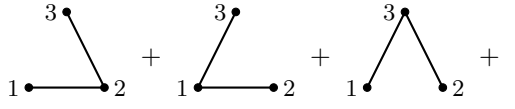
$$\sum_{\ell=1}^{N} m_\ell \ell = N \,. \tag{E.25}$$

The key idea is that we will organise the expansion in such a way that the $(\ell + 1)$-clusters are much less important than the ℓ-clusters. Then we do not need to draw all $N \sim 10^{23}$ atoms in the graph.

To see how this works, let us focus on the 3-clusters for now. There are 4 different ways that we can have a 3-cluster:

Each of these 3-clusters will appear in a graph with any other combination of clusters among the remaining $N - 3$ atoms. But since clusters factorise in the partition function, we know that Q must include a factor

$$U_3 = \int \mathrm{d}^3\mathbf{r}_1\, \mathrm{d}^3\mathbf{r}_2\, \mathrm{d}^3\mathbf{r}_3 \left(+ + + \right) . \tag{E.26}$$



U_3 contains terms of order f^2 and f^3. It turns out that this is the correct way to arrange the expansion: not in terms of the number of lines in the diagram, which is equal to the power of f, but instead the number of atoms that they connect. The partition function will similarly contain factors associated to all other ℓ-clusters. We define the corresponding integrals as

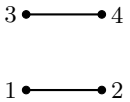
$$U_\ell \equiv \int \prod_{i=1}^{\ell} \mathrm{d}^3\mathbf{r}_i \sum_{G\in\{\ell\text{-cluster}\}} G\,. \tag{E.27}$$

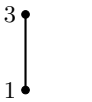
Notice that U_1 is simply the integral over space, so $U_1 = V$. The full partition function must be a product of U_ℓ's. The tricky part is to get all the combinatoric factors right to make sure that you count each graph exactly once. The sum over graphs G that appears in the partition function turns out to be

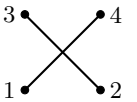
$$\sum_G W[G] = N! \sum_{(m_\ell)} \prod_\ell \frac{U_\ell^{m_\ell}}{(\ell!)^{m_\ell} m_\ell!}\,. \tag{E.28}$$

Where the sum over (m_ℓ) means to take the sum over all possible $(m_1, m_2, \ldots, m_N)$ under the constraint (E.25), and the product $N!/\prod_\ell m_\ell!(\ell!)^{m_\ell}$ counts the number of ways to split the particles into m_ℓ ℓ-clusters, while ignoring the different ways to internally connect each cluster: this is taken into account in the integral U_ℓ.

Let us make a couple checks to make sure that we have got the right answer. First consider $N = 4$ atoms splitting into two 2-clusters (i.e. $m_2 = 2$). There are three such diagrams.

$f_{12}f_{34} =$ 

, $f_{13}f_{24} =$ 

, $f_{14}f_{23} =$ 

. (E.29)

Each of these gives the same answer when integrated, namely U_2^2 so the final result should be $3U_2^2$.We can check this against the relevant terms in equation (E.28), which are $4!U_2^2/2!^2 2! = 3U_2^2$ as expected.

Another check: $N = 5$ atoms with $m_2 = m_3 = 1$. All diagrams come in the combinations

$$U_3U_2 = \int \prod_{i=1}^{5} \mathrm{d}^3\mathbf{r}_i \left(\right. \tag{E.30}$$

$$\left. \right)$$

together with graphs that are related by permutations. The permutations are fully determined by the choice of the two atoms that sit in the pair: there are 10 such choices. The answer should therefore be $10U_3U_2$. Comparing to equation (E.28), we have $5!U_3U_2/3!2! = 10U_3U_2$ as required.

The end result for the partition function is therefore

$$Q(N,V,T) = \frac{1}{\Lambda^{3N}} \sum_{(m_\ell)} \prod_\ell \frac{U_\ell^{m_\ell}}{(\ell!)^{m_\ell} m_\ell!}\,. \tag{E.31}$$

The problem with computing this sum is that we still have to work out the different ways that we can split N atoms into different clusters. In other words, we still have to obey the constraint equation (E.25). It will be much easier if we do not have to worry about this. Then we could just sum over any m_ℓ. This is exactly what we can do if we work in the grand canonical ensemble where N is not fixed. The grand partition function is

$$\Xi(\mu,V,T) = \sum_N z^N Q(N,V,T)\,, \tag{E.32}$$

where $z \coloneqq \mathrm{e}^{\beta\mu}$ is the *fugacity*. Then we can write

$$\begin{aligned}\Xi(\mu,V,T) &= \sum_N z^N Q(N,V,T) \\ &= \sum_{m_\ell=0}^{\infty} \prod_{\ell=1}^{\infty} \left(\frac{z}{\Lambda^3}\right)^{m_\ell \ell} \frac{1}{m_\ell!} \left(\frac{U_\ell}{\ell!}\right)^{m_1} \\ &= \prod_{\ell=1}^{\infty} \exp\left(\frac{U_\ell z^\ell}{\Lambda^{3\ell}\ell!}\right) . \end{aligned} \tag{E.33}$$

Define

$$b_\ell = \frac{\Lambda^3}{V}\frac{U_\ell}{\ell!\Lambda^{3\ell}} . \tag{E.34}$$

Notice in particular that $U_1 = V$ so this definition gives $b_1 = 1$. Then we can write the grand partition function as

$$\Xi(\mu,V,T) = \prod_{\ell=1}^{\infty} \exp\left(\frac{V}{\Lambda^3} b_\ell z^\ell\right) = \exp\left(\frac{V}{\Lambda^3}\sum_{\ell=1}^{\infty} b_\ell z^\ell\right) . \tag{E.35}$$

Something interesting happened here. The sum over all the diagrams got rewritten as the exponential over the sum of all the clusters. This is a general lesson which also carries over to many places — like when we are dealing with the Ising model later, or even in fancy stuff like quantum field theory when dealing with the Feynman diagrams.

Back to the main plot of the story, we can now compute the pressure

$$\frac{PV}{k_B T} = \ln \Xi = \frac{V}{\Lambda^3}\sum_{\ell=1}^{\infty} b_\ell z^\ell , \tag{E.36}$$

and the number of particles

$$\frac{N}{V} = \frac{z}{V}\frac{\partial}{\partial z}\ln \Xi = \frac{1}{\Lambda^3}\sum_{\ell=1}^{\infty} \ell b_\ell z^\ell . \tag{E.37}$$

Dividing the two gives us the equation of state

$$\frac{PV}{N k_B T} = \frac{\sum_\ell b_\ell z^\ell}{\sum_\ell \ell b_\ell z^\ell} . \tag{E.38}$$

The only downside is that the equation of state is expressed in terms of z. To massage it into the form of the virial expansion, we need to invert equation (E.37) to get z in terms of the particle density N/V. Equating equation (E.38) with the virial equation (and defining $B_1 = 1$), we have

$$\begin{aligned}\sum_{\ell=1}^{\infty} b_\ell z^\ell &= \sum_{\ell=1}^{\infty} B_\ell \left(\frac{N}{V}\right)^{\ell-1} \sum_{m=1}^{\infty} m b_m z^m \\ &= \sum_{\ell=1}^{\infty} \frac{B_\ell}{\Lambda^{3(\ell-1)}} \left(\sum_{n=1}^{\infty} n b_n z^n\right)^{\ell-1} \sum_{m=1}^{\infty} m b_m z^m \\ &= \left[1 + \frac{B_2}{\Lambda^3}(z + 2b_2 z^2 + 3b_3 z^3 + \dots) + \frac{B_3}{\Lambda^6}(z + 2b_2 z^2 + 3b_3 z^3 + \dots) + \dots\right] \\ &\quad \times [z + 2b_2 z^2 + 3b_3 z^3 + \dots] , \end{aligned} \tag{E.39}$$

where we have used both $B_1 = 1$ and $b_1 = 1$. Expanding out the left- and right-hand sides to order z^3 gives

$$z + b_2 z^2 + b_3 z^3 = z + \left(\frac{B_2}{\Lambda^3} + 2b_2\right) z^2 + \left(3b_3 + \frac{4b_2 B_2}{\Lambda^3} + \frac{B_3}{\Lambda^6}\right) z^3 + \dots \tag{E.40}$$

Comparing the terms and recollecting the definitions of b_ℓ in terms of U_ℓ gives the second virial coefficient

$$B_2 = -\Lambda^3 b_2 = -\frac{U_2}{2V} = -\frac{1}{2V}\int \mathrm{d}^3\mathbf{r}_1\, \mathrm{d}^3\mathbf{r}_2\, f(\mathbf{r}_1 - \mathbf{r}_2) = -\frac{1}{2}\int \mathrm{d}^3\mathbf{r}\, f(r) . \tag{E.41}$$

This reproduces the result (E.13) we found before.

We now also have an expression for the third coefficient

$$B_3 = \Lambda^6(4b_2^2 - 2b_3) \tag{E.42}$$

where we still have nasty integrals to do before we have a concrete result…

F Some Exact Results on the Ising Model

In this section, we will describe some exact results for the Ising model using techniques that do not rely on the mean field approximation.[20] Many of the results that we derive have broader implications for systems beyond the Ising model.

There is an exact solution for the Ising model in $d = 1$ dimension and, when $B = 0$, in $d = 2$ dimensions. Here we will describe the $d = 1$ solution but not the full $d = 2$ solution because it is infamously complicated. We will, however, derive a number of results for the $d = 2$ Ising model which, while falling short of the full solution, nonetheless provides important insights into the physics.

F.1 The Ising Model in $d = 1$ Dimension

We start with the *Ising chain*, the Ising model on a one-dimensional line. Here we will see that the mean field approximation fails miserably, giving qualitatively incorrect results: the exact results show that there are no phase transitions in the Ising chain. We will write the energy in a slightly unusual form

$$E = -J\sum_{i=1}^{N} s_i s_{i+1} - \frac{B}{2}\sum_{i=1}^{N}(s_i + s_{i+1})\,. \tag{F.1}$$

We will impose periodic boundary conditions, so the spins live on a circular lattice with $s_{N+1} \equiv s_1$. The partition function is then

$$Q = \sum_{s_1=\pm1} \cdots \sum_{s_N=\pm1} \prod_{i=1}^{N} \exp\left(\beta J s_i s_{i+1} + \frac{\beta B}{2}(s_i + s_{i+1})\right)\,. \tag{F.2}$$

The crucial observation that allows us to solve the problem is that this partition function can be written as a product of matrices. We adopt notation from quantum mechanics and define the 2×2 matrix,

$$\langle s_i|T|s_{i+1}\rangle \coloneqq \exp\left(\beta J s_i s_{i+1} + \frac{\beta B}{2}(s_i + s_{i+1})\right)\,. \tag{F.3}$$

The row of the matrix is specified by the value of $s_i = \pm 1$ and the column by $s_{i+1} = \pm 1$. T is known as the *transfer matrix* and, in more conventional notation, is given by

$$\mathsf{T} = \begin{pmatrix} \mathrm{e}^{\beta J + \beta B} & \mathrm{e}^{-\beta J} \\ \mathrm{e}^{-\beta J} & \mathrm{e}^{\beta J - \beta B} \end{pmatrix}\,. \tag{F.4}$$

The sums over the spins $\sum_{s_i}$ and product over lattice sites $\prod_i$ in the partition function (F.2) simply tell us to multiply the matrices defined in (F.3) and the partition function becomes

$$Q = \langle s_1|T|s_2\rangle\,\langle s_2|T|s_3\rangle \ldots \langle s_N|T|s_1\rangle = \operatorname{tr} T^N\,, \tag{F.5}$$

[20] Again, this section is extremely mathematically demanding. It is included in my notes because of my personal interests. It is not in the official notes.

where the trace arises because we have imposed periodic boundary conditions. To complete the story, we need only compute the eigenvalues of T to determine the partition function. A quick calculation shows that the two eigenvalues of T are

$$\lambda_\pm = \mathrm{e}^{\beta J}\cosh\beta B \pm \sqrt{\mathrm{e}^{2\beta J}\cosh^2\beta B - 2\sinh 2\beta J}\,, \tag{F.6}$$

where, clearly, $\lambda_- < \lambda_+$. The partition function is then

$$Q = \lambda_+^N + \lambda_-^N = \lambda_+^N\left(1 + \frac{\lambda_-^N}{\lambda_+^N}\right) \approx \lambda_+^N \tag{F.7}$$

for very large N.

The partition function Q contains many quantities of interest. In particular, we can use it to compute the magnetisation as a function of temperature when $B = 0$. This, recall, is the quantity which is predicted to undergo a phase transition in the mean field approximation, going abruptly to zero at some critical temperature. In the $d = 1$ Ising model, the magnetisation is given by

$$\langle s\rangle_0 = \frac{1}{N\beta}\left.\frac{\partial \ln Q}{\partial B}\right|_{B=0} = \frac{1}{\lambda_+\beta}\left.\frac{\partial\lambda_+}{\partial B}\right|_{B=0} = 0\,. \tag{F.8}$$

We see that the true physics for $d = 1$ is very different than that suggested by the mean field approximation. When $B = 0$, there is no magnetisation! While the J term in the energy encourages the spins to align, this is completely overwhelmed by thermal fluctuations for any value of the temperature.

There is a general lesson in this calculation: thermal fluctuations always win in one-dimensional systems. They never exhibit ordered phases and, for this reason, never exhibit phase transitions. The mean field approximation is bad in one dimension.

F.2 2D Ising Model

Let's now turn to the Ising model in $d = 2$ dimensions. We'll work on a square lattice and set $B = 0$. Rather than trying to solve the model exactly, we'll have more modest goals. We will compute the partition function in two different limits: high temperature and low temperature.

F.2.1 Low Temperature Limit and Peierls Droplets

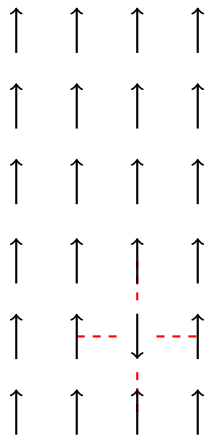
The partition function is given by the sum over all states, weighted by $\mathrm{e}^{-\beta E}$. At low temperatures, this is always dominated by the lowest lying states. For the Ising model, we have[21]

$$Q = \sum_{(s_i)} \exp\left(\beta J\sum_{\langle ij\rangle} s_i s_j\right)\,. \tag{F.9}$$

The low temperature limit is $\beta J \to \infty$, where the partition function can be approximated by the sum over the first few lowest energy states. All we need to do is list these states.

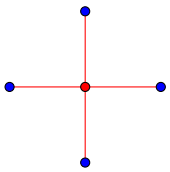
The ground states are easy. There are two of them: spins all up or spins all down. For example, the ground state with spins all up looks like
Each of these ground states has energy $E = E_0 = -2NJ$.

The first excited states arise by flipping a single spin. Each spin has $q = 4$ nearest neighbours — denoted by red lines in the example below — each of which leads to an energy cost of $J - (-J) = 2J$.



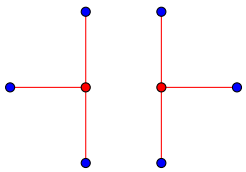
The energy of the first excited states is therefore $E_1 = E_0 + 8J$.
There are, of course, N different spins that we can flip and, correspondingly, the first energy level has a degeneracy of N.

To proceed, we introduce a diagrammatic method to list the different states. We draw only the broken bonds which connect two spins with opposite orientation and, as in the diagram above, denote these by red lines. We further draw the flipped spins as red dots, the unflipped spins as blue dots. The energy of the state is determined simply by the number of red lines in the diagram. Pictorially, we write the first excited state as



$$E_1 = E_0 + 8J$$
$$\text{Degeneracy} = N$$

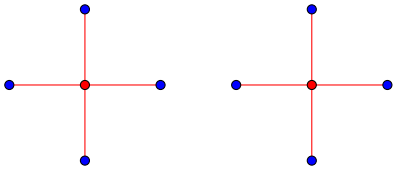
The next lowest state has six broken bonds. It takes the form where the extra factor of 2 in the



$$E_2 = E_0 + 12J$$
$$\text{Degeneracy} = 2N$$

degeneracy comes from the two possible orientations (vertical and horizontal) of the graph.

Things are more interesting for the states which sit at the third excited level. These have 8 broken bonds. The simplest configuration consists of two, disconnected, flipped spins

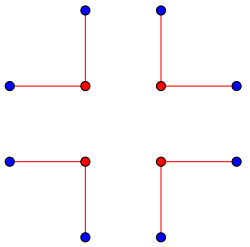


$$E_3 = E_0 + 16J$$
$$\text{Degeneracy} = \tfrac{1}{2}N(N-5) \tag{F.10}$$

The factor of N in the degeneracy comes from placing the first graph; the factor of $N - 5$ arises because the flipped spin in the second graph can sit anywhere apart from on the five vertices used in the first graph. Finally, the factor of $1/2$ arises from the interchange of the two vertices.

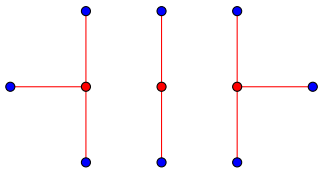
[21] From now on, we will directly use $\sum_{\langle ij \rangle}$ (without the comma) to denote the sum over all neighbouring pairs which we used to denote as $\sum_i \sum_{\langle i,j \rangle}$ to reduce the cluttering of notations.

There are also three further graphs with the same energy E_3. These are and where the degeneracy



$E_3 = E_0 + 16J$

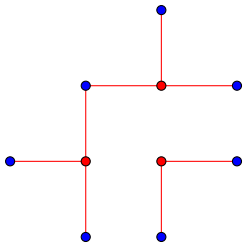
Degeneracy $= N$



$E_3 = E_0 + 16J$

Degeneracy $= 2N$

comes from the two orientations, and finally where the degeneracy comes from the four orientations.



$E_3 = E_0 + 16J$

Degeneracy $= 4N$

We can of course calculate more excited states but we will stop here. Adding all the graphs above gives us an expansion of the partition function in power of $\mathrm{e}^{-\beta J} \ll 1$. This is

$$Q = 2\mathrm{e}^{2N\beta J}\left(1 + N\mathrm{e}^{-8\beta J} + 2N\mathrm{e}^{-12\beta J} + \frac{1}{2}(N^2+9N)\mathrm{e}^{-16\beta J} + \ldots\right), \tag{F.11}$$

where the overall factor of 2 originates from the two ground states of the system. We'll make use of the specific coefficients in this expansion in section F.2.3. Before we focus on the physics hiding in the low temperature expansion, it's worth making a quick comment that something quite nice happens if we take the log of the partition function and expand $\ln(1+x) = x - \frac{1}{2}x^2 + \ldots$, we get

$$\ln Q = \ln 2 + 2N\beta J + N\mathrm{e}^{-8\beta J} + 2N\mathrm{e}^{-12\beta J} + \frac{9}{2}N\mathrm{e}^{-16\beta J} + \ldots \tag{F.12}$$

The thing to notice is that the N^2 term in the partition function has precisely cancelled out and $\ln Q$ is proportional to N, which is to be expected since the free energy of the system is extensive. Looking back, we see that the N^2 term was associated to the disconnected diagrams in (F.10). There is actually a general lesson hiding here: the partition function can be written as the exponential of the sum of connected diagrams. We saw exactly the same issue arise in the cluster expansion in (E.35).

Peierls Droplets

Continuing the low temperature expansion provides a heuristic, but physically intuitive, explanation for why phase transitions happen in $d \geq 2$ dimensions but not in $d = 1$. As we flip more and more spins, the low energy states become *droplets*, consisting of a region of space in which all the spins are flipped, surrounded by a larger sea in which the spins have their original alignment. The energy cost of such a droplet is roughly

$$E \sim 2JL\,, \tag{F.13}$$

where L is the perimeter of the droplet. Notice that the energy does not scale as the area of the droplet since all spins inside are aligned with their neighbours. It is only those on the edge which are misaligned and this is the reason for the perimeter scaling. To understand how these droplets contribute to the partition function, we also need to know their degeneracy. We will now argue that the degeneracy of droplets scales as

$$\text{Degeneracy} \sim \mathrm{e}^{\alpha L} \tag{F.14}$$

for some value of α. To see this, consider firstly the problem of a random walk on a 2D square lattice. At each step, we can move in one of four directions. So the number of paths of length L is

$$\#\text{paths} \sim 4^L = \mathrm{e}^{L \ln 4}\,. \tag{F.15}$$

Of course, the perimeter of a droplet is more constrained than a random walk. Firstly, the perimeter can't go back on itself, so it really only has three directions that it can move in at each step. Secondly, the perimeter must return to its starting point after L steps. And, finally, the perimeter cannot self-intersect. One can show that the number of paths that obey these conditions is

$$\#\text{paths} \sim \mathrm{e}^{\alpha L}\,, \tag{F.16}$$

where $\ln 2 < \alpha < \ln 3$. Since the degeneracy scales as $\mathrm{e}^{\alpha L}$, the entropy of the droplets is proportional to L.

The fact that both energy and entropy scale with L means that there is an interesting competition between them. At temperatures where the droplets are important, the partition function is schematically of the form

$$Q \sim \sum_L \mathrm{e}^{\alpha L}\mathrm{e}^{-2\beta J L}\,. \tag{F.17}$$

For large β (i.e. low temperature) the partition function converges. However, as the temperature increases, one reaches the critical temperature

$$k_B T_c \approx \frac{2J}{\alpha} \tag{F.18}$$

where the partition function no longer converges. At this point, the entropy wins over the energy cost and it is favourable to populate the system with droplets of arbitrary sizes. This is how one sees the phase transition in the partition function. For temperature above T_c, the low-temperature expansion breaks down and the ordered magnetic phase is destroyed.

We can also use the droplet argument to see why phase transitions don't occur in $d = 1$ dimension. On a line, the boundary of any droplet always consists of just two points. This means that the energy cost to forming a droplet is always $E = 2J$, regardless of the size of the droplet. But, since the droplet can exist anywhere along the line, its degeneracy is N. The net result is that the free energy associated to creating a droplet scales as

$$A = 2J - k_B T \ln N \tag{F.19}$$

and, as $N \to \infty$, the free energy is negative for any $T > 0$. This means that the system will prefer to create droplets of arbitrary length, randomizing the spins. This is the intuitive reason why there is no magnetic ordered phase in the $d = 1$ Ising model.

F.2.2 High Temperature Limit

We now turn to the 2D Ising model in the opposite limit of high temperature. Here we expect the partition function to be dominated by the completely random, disordered configurations of maximum entropy. Our goal is to find a way to expand the partition function in $\beta J \ll 1$.

We again work with zero magnetic field, $B = 0$ and write the partition function as

$$Q = \sum_{(s_i)} \exp\left(\beta J \sum_{\langle ij\rangle} s_i s_j\right) = \sum_{(s_i)} \prod_{\langle ij\rangle} \mathrm{e}^{\beta J s_i s_j} \,. \tag{F.20}$$

There is a useful way to rewrite $\mathrm{e}^{\beta J s_i s_j}$ which relies on the fact that the product $s_i s_j$ only takes ± 1. It doesn't take long to check the following identity:

$$\begin{aligned} \mathrm{e}^{\beta J s_i s_j} &= \cosh\beta J + s_i s_j \sinh\beta J \\ &= \cosh\beta J(1 + s_i s_j \tanh\beta J) \,. \end{aligned} \tag{F.21}$$

Using this, the partition function becomes

$$\begin{aligned} Q &= \sum_{(s_i)} \prod_{\langle ij\rangle} \cosh\beta J(1 + s_i s_j \tanh\beta J) \\ &= (\cosh\beta J)^{qN/2} \sum_{(s_i)} \prod_{\langle ij\rangle} (1 + s_i s_j \tanh\beta J) \,, \end{aligned} \tag{F.22}$$

where the number of nearest neighbours is $q = 4$ for the 2D square lattice.

With the partition function in this form, there is a natural expansion which suggests itself. At high temperatures $\beta J \ll 1$ which, of course, means that $\tanh\beta J \ll 1$. But the partition function is now naturally a product of powers of $\tanh\beta J$. This is again somewhat analogous to the cluster expansion for the interacting gas. As in the cluster expansion, we will represent the expansion graphically.

We need no graphics for the leading order term. It has no factors of $\tanh\beta J$ and is simply

$$Q \approx (\cosh\beta J)^{2N} \sum_{(s_i)} 1 = 2^N (\cosh\beta J)^{2N} \,. \tag{F.23}$$

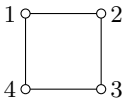
Let's now turn to the leading correction. Expanding the partition function (F.22), each power of $\tanh\beta J$ is associated to a nearest neighbour pair $\langle i, j\rangle$. We'll represent this by drawing a line on the lattice:

$$= s_i s_j \tanh\beta J \,. \tag{F.24}$$

But a sum over graphs of these kinds vanishes: each factor of $\tanh\beta J$ in (F.22) also comes with a sum over all spins s_i and s_j. And these are $+1$ and -1 which means that they simply sum to zero

$$\sum_{s_i, s_j} s_i s_j = +1 - 1 - 1 + 1 = 0 \,. \tag{F.25}$$

How can we avoid this? The only way is to make sure that we're summing over an even number of spins on each site, since then we get factors of $s_i^2 = 1$ and no cancellations. Graphically, this means that every site must have an even number of lines attached to it. The first correction is then of the form



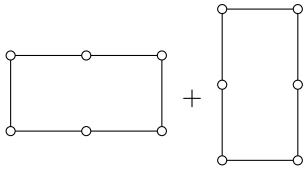
$$= (\tanh\beta J)^4 \sum_{(s_i)} s_1 s_2\, s_2 s_3\, s_3 s_4\, s_4 s_1 = 2^4 (\tanh\beta J)^4 \,. \tag{F.26}$$

There are N such terms since the upper left corner of the square can be on any one of the N lattice sites (assuming periodic boundary conditions for the lattice). So including the leading term and first correction, we have

$$Q = 2^N (\cosh\beta J)^{2N} (1 + N(\tanh\beta J)^4 + \dots) \,. \tag{F.27}$$

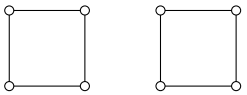
We can go further. The next terms arise from graphs of length 6 and the only possibilities are rectangles, oriented as either landscape or portrait. Each of them can sit on one of N sites, giving a

contribution



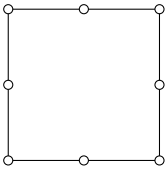
$$= 2N(\tanh \beta J)^6 \,. \tag{F.28}$$

Things get more interesting when we look at graphs of length 8. We have four different types of graphs. Firstly, there are the trivial, disconnected pair of squares

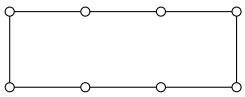


$$= \frac{1}{2}N(N-5)(\tanh \beta J)^8 \,. \tag{F.29}$$

Here the first factor of N is the possible positions of the first square; the factor of $N-5$ arises because there are 5 locations the second square can't be: overlapping or neighbouring with the first square. Finally, the factor of $1/2$ comes because the two squares are identical. The other graphs of length 8 are a large square, a rectangle and a corner.

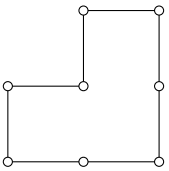


$$= N(\tanh \beta J)^8 \,, \tag{F.30}$$



$$= 2N(\tanh \beta J)^8 \tag{F.31}$$

and



$$= 4N(\tanh \beta J)^8 \,. \tag{F.32}$$

Adding all contributions together gives us the first few terms in high temperature expansion of the partition function

$$Q = 2^N(\cosh \beta J)^{2N}\Big(1 + N(\tanh \beta J)^4 + 2N(\tanh \beta J)^6 + \frac{1}{2}(N^2+9N)(\tanh \beta J)^8 + \dots\Big) \,. \tag{F.33}$$

You might have already noticed, everything above are weirdly familiar to everything in the low temperature expansion.

F.2.3 Kramers–Wannier Duality

We have computed the partition function perturbatively in two extreme regimes of low temperature and high temperature. The physics in the two cases is, of course, very different. At low temperatures, the partition function is dominated by the lowest energy states; at high temperatures it is dominated by maximally disordered states. Yet comparing the partition functions at low temperature (F.11) and high temperature (F.33) reveals an extraordinary symmetry between them! The two series agree if we exchange

$$\mathrm{e}^{-2\beta J} \longleftrightarrow \tanh \beta J \,. \tag{F.34}$$

Of course, we've only checked the agreement to the first few orders in perturbation. Below we shall prove that this miracle continues to all orders in perturbation theory. The symmetry of the partition

function under the interchange (F.34) is known as *Kramers–Wannier duality*. Before we prove this duality, we will first just assume that it is true and extract some consequences.

We can express the statement of the duality more clearly. The Ising model at temperature β is related to the same model at temperature $\tilde{\beta}$, defined as

$$e^{-2\tilde{\beta}J} = \tanh \beta J \,. \tag{F.35}$$

This way of writing things hides the symmetry of the transformation. A little algebra shows that this is equivalent to

$$\sinh 2\tilde{\beta}J = \frac{1}{\sinh 2\beta J} \,. \tag{F.36}$$

This is a hot-cold duality — when β is large, $\tilde{\beta}$ is small. Kramers–Wannier duality is the statement that, when $B = 0$, the partition functions of the Ising model at two temperatures are related by

$$\begin{aligned} Q(\beta) &= \frac{2^N (\cosh \beta J)^{2N}}{2e^{2N\tilde{\beta}J}} Q(\tilde{\beta}) \\ &= 2^{N-1} (\cosh \beta J \sinh \beta J)^N Q(\tilde{\beta}) \,. \end{aligned} \tag{F.37}$$

This means that if you know the thermodynamics of the Ising model at one temperature, then you also know the thermodynamics at the other temperature. Notice however, that it does not say that all the physics of the two models is equivalent. In particular, when one system is in the ordered phase, the other typically lies in the disordered phase.

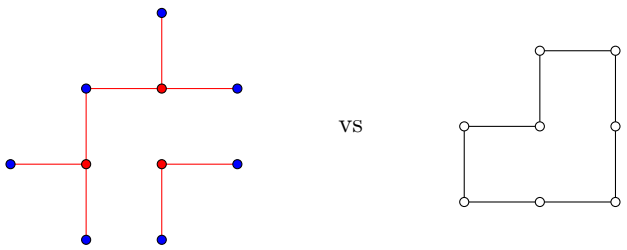
One immediate consequence of the duality is that we can use it to compute the exact critical temperature T_c. This is the temperature at which the partition function is singular in the $N \to \infty$ limit. If we further assume that there is just a single phase transition as we vary the temperature, then it must happen at the special self-dual point $\beta = \tilde{\beta}$. This is

$$k_B T = \frac{2J}{\ln(\sqrt{2}+1)} \approx 2.269 J \,. \tag{F.38}$$

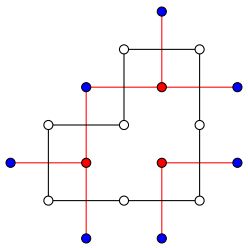
The exact solution of Onsager confirms that this is indeed the transition temperature. It's also worth noting that it's fully consistent with the more heuristic Peierls droplet argument (F.18) since $\ln 2 < \ln(1+\sqrt{2}) < \ln 3$.

Proving the Kramers–Wannier Duality

The key idea that we need can actually be found in the various graphs that we have drawn. You may have realised that the graphs we have drawn in the two expansions are actually the same, albeit drawn differently. For example, consider the two "corner" diagrams



The two graphs are *dual*. The red lines in the first graph intersect the black lines in the second as can be seen by placing them on top of each other:
The same pattern occurs more generally: the graphs appearing in the low temperature expansion are in one-to-one correspondence with the dual graphs of the high temperature expansion. Here we will show how this occurs and how one can map the partition functions onto each other.



Let's start by writing the partition function in the form (F.22) that we met in the high temperature expansion and presenting it in a slightly different way,

$$\begin{aligned} Q(\beta) &= \sum_{(s_i)} \prod_{\langle ij \rangle} (\cosh \beta J + s_i s_j \sinh \beta J) \\ &= \sum_{(s_i)} \prod_{\langle ij \rangle} \sum_{k_{ij}=0,1} C_{k_{ij}}(\beta J)(s_i s_j)^{k_{ij}} \,, \end{aligned} \tag{F.39}$$

where we have introduced the rather strange variable k_{ij} associated to each nearest neighbour pair that takes values 0 and 1, together with the functions

$$C_0(\beta J) = \cosh \beta J \qquad \text{and} \qquad C_1(\beta J) = \sinh \beta J \,. \tag{F.40}$$

The variables in the original Ising model were spins on the lattice sites. The observation that the graphs which appear in the two expansions are dual suggests that it might be profitable to focus attention on the links between lattice sites. Clearly, we have one link for every nearest neighbour pair. If we label these links by ℓ, we can trivially rewrite the partition function as

$$Q = \sum_{k_\ell=0,1} \prod_{\ell} \sum_{(s_i)} C_{k_\ell}(\beta J)(s_i s_j)^{k_\ell} \,. \tag{F.41}$$

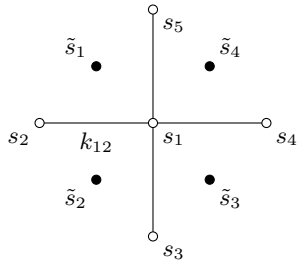
Notice that the strange label k_{ij} has now become a variable that lives on the links ℓ rather than the original lattice sites i.

At this stage, we do the sum over the spins s_i. We've already seen that if a given spin, say s_i, appears in a term an odd number of times, then that term will vanish when we sum over the spin. Alternatively, if the spin s_i appears an even number of times, then the sum will give 2. We'll say that a given link ℓ is turned on in configurations with $k_\ell = 1$ and turned off when $k_\ell = 0$. In this language, a term in the sum over spin s_i contributes only if an even number of links attached to site i are turned on. The partition function then becomes

$$Q = 2^N \sum_{k_\ell} \prod_{\ell} C_{k_\ell}(\beta J) \Big|_{\text{constrained}} \,. \tag{F.42}$$

Now we have something interesting. Rather than summing over spins on lattice sites, we're now summing over the new variables k_ℓ living on links. This looks like the partition function of a totally different physical system, where the degrees of freedom live on the links of the original lattice. But there's a catch — that "constrained" label on the sum. This is there to remind us that we don't sum over all k_ℓ configurations; only those for which an even number of links are turned on for every lattice site. And that's annoying. It's telling us that the k_ℓ aren't really independent variables. There are some constraints that must be imposed.

Fortunately, for the 2D square lattice, there is a simple way to solve the constraint. We introduce yet more variables, $\tilde{s}_i$ which, like the original spin variables, take values ± 1. However, the $\tilde{s}_i$ do not live on the original lattice sites. Instead, they live on the vertices of the dual lattice. For the 2D square lattice, the dual vertices are drawn in the figure. The original lattice sites are in white; the dual lattice sites in black.



The link variables k_ℓ are related to the two nearest (dual) spin variables $\tilde{s}_i$ as follows:

$$k_{12} = \frac{1}{2}(1 - \tilde{s}_1\tilde{s}_2) \tag{F.43}$$

$$k_{13} = \frac{1}{2}(1 - \tilde{s}_2\tilde{s}_3) \tag{F.44}$$

$$k_{14} = \frac{1}{2}(1 - \tilde{s}_3\tilde{s}_4) \tag{F.45}$$

$$k_{15} = \frac{1}{2}(1 - \tilde{s}_1\tilde{s}_4) \tag{F.46}$$

Notice that we've replaced four variables k_ℓ taking values $0, 1$ with four variables $\tilde{s}_i$ taking values ± 1. Each set of variables gives 2^4 possibilities. However, the map is not one-to-one. It is not possible to construct for all values of k_ℓ using the parameterization in terms of $\tilde{s}_i$. To see this, we need only look at

$$\begin{aligned} k_{12} + k_{13} + k_{14} + k_{15} &= 2 - \frac{1}{2}(\tilde{s}_1\tilde{s}_2 + \tilde{s}_2\tilde{s}_3 + \tilde{s}_3\tilde{s}_4 + \tilde{s}_1\tilde{s}_4) \\ &= 2 - \frac{1}{2}(\tilde{s}_1 + \tilde{s}_3)(\tilde{s}_2 + \tilde{s}_4) \\ &= 0, 2 \text{ or } 4\,. \end{aligned} \tag{F.47}$$

In other words, the number of links that are turned on must be even. But that's exactly what we want! Writing the k_ℓ in terms of the auxiliary spins $\tilde{s}_i$ automatically solves the constraint that is imposed on the sum in (F.42). Moreover, it is simple to check that for every configuration $\{k_\ell\}$ obeying the constraint, there are two configurations of $\{\tilde{s}_i\}$. This means that we can replace the constrained sum over $\{k_\ell\}$ with an unconstrained sum over $\{\tilde{s}_i\}$. The only price we pay is an additional factor of $1/2$.

$$Q(\beta) = \frac{1}{2}2^N \sum_{(\tilde{s}_i)} \prod_{\langle ij \rangle} C_{\frac{1}{2}(1-\tilde{s}_i\tilde{s}_j)}(\beta J)\,. \tag{F.48}$$

Finally, we'd like to find a simple expression for C_0 and C_1 in terms of $\tilde{s}_i$. That's easy enough. We can write

$$\begin{aligned} C_k(\beta J) &= \cosh\beta J \exp(k \ln\tanh\beta J) \\ &= (\sinh\beta J \cosh\beta J)^{1/2} \exp\left(-\frac{1}{2}\tilde{s}_i\tilde{s}_j \ln\tanh\beta J\right)\,. \end{aligned} \tag{F.49}$$

Substituting this into our newly re-written partition function gives

$$\begin{aligned} Q(\beta) &= 2^{N-1} \sum_{(\tilde{s}_i)} \prod_{\langle ij \rangle} (\sinh\beta J \cosh\beta J)^{1/2} \exp\left(-\frac{1}{2}\tilde{s}_i\tilde{s}_j \ln\tanh\beta J\right) \\ &= 2^{N-1}(\sinh\beta J \cosh\beta J)^N \sum_{(\tilde{s}_i)} \exp\left(-\frac{1}{2}\ln(\tanh\beta J) \sum_{\langle ij \rangle} \tilde{s}_i\tilde{s}_j\right)\,. \end{aligned} \tag{F.50}$$

This final form of the partition function in terms of the dual spins $\tilde{s}_i$ has exactly the same functional form as the original partition function in terms of the spins s_i! More precisely, we can write

$$Q(\beta) = 2^{N-1}(\sinh 2\beta J)^N Q(\tilde{\beta})\,, \tag{F.51}$$

where

$$e^{-2\tilde{\beta}J} = \tanh \beta J \tag{F.52}$$

as advertised. This is the Kramers–Wannier duality.

This seems to be too out of a tangent from our main theme, but the concept of *duality* is a major feature in theoretical research. The key idea is that when the temperature gets large there may be a different set of variables in which a theory can be written where it appears to live at low temperature. The same idea often holds in quantum theories, where duality maps strong coupling problems to weak coupling problems.

The duality in the Ising model is special for two reasons: firstly, the new variables $\tilde{s}_i$ are governed by the same Hamiltonian as the original variables s_i. We say that the Ising model is *self-dual*. In general, this need not be the case — the high temperature limit of one system could look like the low-temperature limit of a very different system. Secondly, the duality in the Ising model can be proven explicitly. For most systems, we have no such luck. Nonetheless, the idea that there may be dual variables in other, more difficult theories, is compelling. Commonly studied examples include the exchange particles and vortices in two dimensions, and electrons and magnetic monopoles in three dimensions.

G Lee–Yang Theorem

You may have noticed that the flavour of our discussion on phase transitions is a little different from the previous chapter. Our previous philosophy was to regard the partition function as the most important quantity that we can use to derive everything, but when doing phase transitions, we dumped the partition function as soon as we could, preferring instead to work with the macroscopic variable like the free energy. Why didn't we stick with the partition function and examine phase transitions directly?

The reason is that the approach using the partition function is hard! In this short appendix section, which is somewhat tangential to our main discussion, we will describe how phase transitions manifest themselves in the partition function.

For concreteness, let's go back to the classical non-ideal gas with pairwise-additive interactions, although the results we derive will be more general. We'll work in the grand canonical ensemble, with the partition function

$$\Xi = \sum_N z^N Q(N,V,T) = \sum_N \frac{z^N}{N!\Lambda^{3N}} \int \prod_i \mathrm{d}^3\mathbf{r}_i \, e^{-\beta \sum_{j<k} U(r_{jk})} \,. \tag{G.1}$$

To regulate any potential difficulties with short distances, it is useful to assume that the particles have hard cores so that they cannot approach to a distance less than r_0. We model this by requiring that the potential satisfies

$$U(r_{jk}) = \infty \qquad \text{for } r_{jk} < r_0 \,. \tag{G.2}$$

But this has an obvious consequence: if the particles have finite size, then there is a maximum number of particles, N_V, that we can fit into a finite volume V. (Roughly this number is $N_V \sim V/r_0^3$). But that, in turn, means that the canonical partition function $Q(N,V,T) = 0$ for $N > N_V$, and the grand partition function Ξ is therefore a finite polynomial in the fugacity z, of order N_V. But if the partition function is a finite polynomial, there can't be any discontinuous behaviour associated with a phase transition. In particular, we can calculate

$$PV = k_B T \ln \Xi \,, \tag{G.3}$$

which gives us PV as a smooth function of z. We can also calculate

$$N = z \frac{\partial}{\partial z} \ln \Xi \,, \tag{G.4}$$

which gives us N as a function of z. Eliminating z between these two functions tells us that pressure p is a smooth function of density N/V. We're never going to get the behaviour that we derived from the Maxwell construction in which the plot of pressure against density shown in figure 3.7 exhibits a discontinuous derivative.

The discussion above is just re-iterating a statement that we've alluded to before: there are no phase transitions in a finite system. To see the discontinuous behaviour, we need to take the limit $V \to \infty$. A theorem due to Lee and Yang gives us a handle on the analytic properties of the partition function in this limit.

The surprising insight of Lee and Yang is that if you're interested in phase transitions, you should look at the zeros of Ξ in the complex z-plane.[22] Let's first look at these when V is finite. Importantly, at finite V there can be no zeros on the positive real axis, $z > 0$. This follows from the definition of Ξ where it is a sum of positive quantities. Moreover, from (G.4), we can see that Ξ is a monotonically increasing function of z because we necessarily have $N > 0$. Nonetheless, Ξ is a polynomial in z of order N_V so it certainly has N_V zeros somewhere in the complex z-plane. Since $\Xi^*(z) = \Xi(z^*)$, these zeros must either sit on the real negative axis or come in complex conjugate pairs.

However, the statements above rely on the fact that Ξ is a finite polynomial. As we take the limit $V \to \infty$, the maximum number of particles that we can fit in the system diverges, $N_V \to \infty$, and Ξ is now defined as an infinite series. But infinite series can do things that finite ones can't. The Lee-Yang theorem says that as long as the zeros of Ξ continue to stay away from the positive real axis as $V \to \infty$, then no phase transitions can happen. But if one or more zeros happen to touch the positive real axis, life gets more interesting.

Theorem (Lee–Yang Theorem). The quantity

$$\Theta := \lim_{V\to\infty} \left[\frac{1}{V} \ln \Xi(z, V, T) \right] \tag{G.5}$$

exists for all $z > 0$. The result is a continuous, non-decreasing function of z which is independent of the shape of the box (up to some sensible assumptions such as Surface Area$/V \sim V^{-1/3}$ which ensures that the box isn't some stupid fractal shape).

Moreover, let R be a fixed, volume independent, region in the complex z plane which contains part of the real, positive axis. If R contains no zero of $\Xi(z, V, T)$ for all $z \in R$, then Θ is an analytic function of z for all $z \in R$. In particular, all derivatives of Θ are continuous.

In other words, there can be no phase transitions in the region R even in the $V \to \infty$ limit. The last result means that, as long as we are safely in a region R, taking derivatives with respect to z commutes with taking the limit $V \to \infty$. In other words, we are allowed to use (G.4) to write the particle density $\rho = N/V$ as

$$\lim_{V\to\infty} \rho = \lim_{V\to\infty} z \frac{\partial}{\partial z} \left(\frac{P}{k_B T} \right) = z \frac{\partial \Theta}{\partial z} \,. \tag{G.6}$$

However, if we look at points z where zeros appear on the positive real axis, then Θ will generally not be analytic. If $\partial\Theta/\partial z$ is discontinuous, then the system is said to undergo a first order phase transition. More generally, if $\partial^m\Theta/\partial z$ is discontinuous for $m = n$, but continuous for all $m < n$, then the system undergoes an n^{th} order phase transition.

H Landau–Ginzburg Theory

Landau's theory of phase transition focuses only on the average quantity, the order parameter. It ignores the fluctuations of the system, assuming that they are negligible. Here we sketch a

[22] This probably isn't too surprising if you stare at the definition of canonical partition function long enough and suddenly realise that $Q(N, V, T)$ is the (discrete) Laplace transform of $\Omega(N, V, E)$.

generalisation which attempts to account for these fluctuations. It is known as Landau–Ginzburg theory.

The idea is to stick with the concept of order parameter m, but now we allow the order parameter to vary in space so it becomes a function $m(\mathbf{r})$. Let us restrict ourselves to the situation where there is a symmetry $m \to -m$ so we only need to consider even powers in the expansion of the free energy. We add to these a gradient term whose role is to capture the fact that there is some stiffness in the system — it costs energy to vary the order parameter from one point to another. (For the example of the Ising model, this is simply the statement that nearby spins want to be aligned). The (truncated) free energy is then given by

$$A[m(\mathbf{r})] = \int \mathrm{d}^d\mathbf{r}\,[a(T)m^2 + b(T)m^4 + c(T)(\boldsymbol{\nabla} m)^2]\,. \tag{H.1}$$

Notice that we start with terms quadratic in the gradient: a term linear in the gradient would violate the rotational symmetry of the system.

We again require that the free energy is minimised. But now A is a functional — it is a function of the function $m(\mathbf{r})$. If you have learned variational principles, the stationary points will be given by the Euler–Lagrange equations. If you have not, long story short, we consider varying the function $m(\mathbf{r})$ by $\delta m(\mathbf{r})$. The variation of the free energy will then be

$$\begin{aligned}\delta A &= \int \mathrm{d}^d\mathbf{r}\,[2am\delta m + 4bm^3\delta m + 2c\boldsymbol{\nabla} m\cdot\boldsymbol{\nabla}\delta m] \\ &= \int \mathrm{d}^d\mathbf{r}\,[2am + 4bm^3 - 2c\nabla^2 m]\delta m\,,\end{aligned} \tag{H.2}$$

where we used integration by parts. The minimum energy occurs when $\delta A = 0$ for any δm to the first order, and so

$$c\nabla^2 m = am + 2bm^3\,. \tag{H.3}$$

The simplest solution to this equation have m constant, reducing us back to Landau theory. However, allowing for the possibility of spatial variation in m opens up the possibility for us to search for some more interesting solutions.

H.1 Domain Walls

Suppose we have $T < T_c$, the critical temperature, so there exists two degenerate ground states, $m = \pm m_0$. We could cook up a situation in which one half of space, say $x < 0$, lives in the ground state $m = -m_0$ while the other half of space, $x > 0$ lives in $m = +m_0$. This is exactly analogous to the liquid–gas phase separation where we have a half liquid and a half gas. The two regions in which spins point up or down are called *domains*. The place where these regions meet is called the *domain wall*.

We would like to understand the structure of the domain wall. How does the system interpolate between these two states? The transition can't happen instantaneously because that would result in the gradient term $(\boldsymbol{\nabla} m)^2$ giving an infinite contribution to the free energy. But neither can the transition linger too much because any point at which $m(\mathbf{r})$ differs significantly from the value $\pm m_0$ costs free energy from the m^2 and m^4 terms. There must be an equilibrium between these two effects.

To describe the system with two domains, $m(\mathbf{r})$ must vary, but it need only change in one direction: $m = m(x)$. The equation (H.3) then reduces to an ordinary differential equation

$$\frac{\mathrm{d}^2 m}{\mathrm{d}x^2} = \frac{am}{c} + \frac{2bm^3}{c}\,. \tag{H.4}$$

Remember we are below phase transition temperature, so $a < 0$. We then have

$$m = m_0 \tanh\left(\sqrt{\frac{-a}{2c}}x\right)\,, \tag{H.5}$$

where $m_0 = \sqrt{-a/2b}$ is the constant ground state solution for the spin. As $x \to \pm\infty$, the tanh function tends towards ± 1 which means that $m \to \pm m_0$, so this solution indeed interpolates between the two domains as required. We learn that the width of the domain wall is given by $\sqrt{-2c/a}$. Outside this region, the magnetisation relaxes exponentially quickly back to the ground state values.

We can also compute the cost in free energy due to the presence of the domain wall. To do this, we substitute the solution back into the free energy expression. The cost is not proportional to the volume of the system, but instead proportional to the area of the domain wall. This means that if the system has a linear size L then the free energy of the ground state scales as L^d while the free energy required by the wall scales only as L^{d-1}. It is simple to find the parametric dependence of this domain wall energy without doing any integrals; the energy per unit area scales as $\sqrt{-ca^3/b}$. Notice that as we approach the critical point, and $a \to 0$, the two vacua are closer, the width of the domain wall increases and its energy decreases.

H.2 Correlations

One of the most important applications of Landau–Ginzburg theory is to understand the correlations between fluctuations of the system at different points in space. Suppose that we know that the system has an unusually high fluctuation away from the average at some point in space, let's say the origin $\mathbf{r} = \mathbf{0}$. What is the effect of this on nearby points?

There is a simple way to answer this question that requires us only to solve the differential equation (H.3). However, there is also a more complicated way to derive the same result which has the advantage of stressing the underlying physics and the role played by fluctuations. We'll start by deriving the correlations in the simple manner. We'll then see how it can also be derived using the more technical machinery.

We assume that the system sits in a given ground state, say $m = +m_0$, and imagine a small perturbation around this. We write the magnetisation as

$$m(\mathbf{r}) = m_0 + \delta m(\mathbf{r})\,. \tag{H.6}$$

If we substitute this into equation (H.3) and keep only terms linear in δm, we find

$$c\nabla^2 \delta m + 2a\delta m = 0\,, \tag{H.7}$$

where we have substituted $m_0^2 = -a/2b$. We now perturb the system. This can be modelled by putting a delta-function source at the origin, so that the above equation becomes

$$c\nabla^2 \delta m + 2a\delta m = \frac{1}{2c}\delta(\mathbf{r})\,, \tag{H.8}$$

where the strength of the delta function is chosen merely to make the equation somewhat nicer. It is straightforward to solve the asymptotic behaviour of this equation.[23] Neglecting constant factors, it is

$$\delta m(\mathbf{r}) \sim \frac{\mathrm{e}^{-r/\xi}}{r^{(d-1)/2}}\,. \tag{H.9}$$

This tells us how the perturbation decays as we move away from the origin. This equation has several names, reflecting the fact that it arises in many contexts. In liquids, it is usually called the *Ornstein–Zernicke correlation*. It also arises in particle physics as the *Yukawa potential*. The length scale ξ is called the *correlation length*

$$\xi = \sqrt{\frac{-c}{2a}}\,. \tag{H.10}$$

It provides a measure of the distance it takes for correlations to decay. Notice that as we approach the critical point, $a \to 0$ and the correlation length diverges. This provides yet another hint that we need more powerful tools to understand the physics at the critical point. We will now take the first step towards developing these tools.

[23] The same type of equation will arise in the Debye–Hückel model of screening.

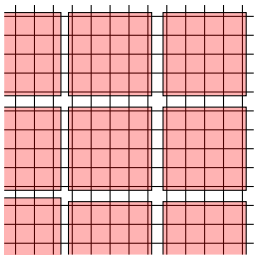


Figure H.1: Coarse graining in the Ising model lattice to define $m(\mathbf{x})$.

H.3 Fluctuations

The main motivation to allow the order parameter to depend on space is to take into account the effect of fluctuations. To see how we can do this, we first need to think a little more about the meaning of the quantity $A[m(\mathbf{r})]$ and what we can use it for.

To understand this point, it is best if we go back to basics. We know that the true free energy of the system can be equated with the log of the partition function. The Landau–Ginzburg functional $A[m(\mathbf{r})]$ is closely related to the true free energy, but it is not quite the same thing as we will see shortly. To avoid confusion, we will call the true free energy F. Since $F = -k_B T \ln Q$, we can write

$$\mathrm{e}^{-\beta F} = Q = \sum_n \mathrm{e}^{-\beta E_n} \,. \tag{H.11}$$

We would like to understand the right way to view the functional $A[m(\mathbf{r})]$ in this framework. Here we give a heuristic and fairly handwaving argument. A fuller treatment involves the idea of the renormalisation group, which you will learn if you decide to study statistical field theory.

The idea is that each microstate $|n\rangle$ of the system can be associated to some specific function of the spatially varying order parameter $m(\mathbf{r})$. To illustrate this, we'll talk in the language of the Ising model although the discussion generalises to any system. For the Ising model of N lattice points, we can divide the lattice into uniform blocks, each consisting of a number of lattice sites (say N' lattice points per box) but its scale should be smaller than any other length scales in our model (in particular, it should be smaller than the correlation length). This step is called *coarse graining*. For each box, we can evaluate its average magnetisation $m(\mathbf{r}) = \frac{1}{N'} \sum_i s_i$, where $\mathbf{r}$ is the centre of the box. Clearly this function is defined only at discrete points, if the number of boxes N/N' is big enough, we can effectively treat $\mathbf{r}$ as being continuous. Moreover, at each $\mathbf{r}$, $m(\mathbf{r})$ is quantised in units of $1/N'$, so we also need to take N' big enough so $m(\mathbf{r})$ can effectively take any value between 1 and -1.

In this way, we get a map from the microstates to the magnetisation function, $|n\rangle \to m(\mathbf{r})$. But due to the averaging process, this map is clearly not one-to-one. In this way, many microstates map onto the same average magnetisation. Summing over just these microstates provides a first principle construction of the $A[m(\mathbf{r})]$:

$$\mathrm{e}^{-\beta A[m(\mathbf{r})]} = \sum_{n|m(\mathbf{r})} \mathrm{e}^{-\beta E_n} \,. \tag{H.12}$$

Of course, we didn't actually perform this procedure to get to (H.1): we simply wrote it down the most general form in the vicinity of a critical point with a bunch of unknown coefficients $a(T)$, $b(T)$ and $c(T)$. But if we were up for a challenge, the above procedure tells us how we could go about figuring out those functions from first principles. More importantly, it also tells us what we should do with the Landau–Ginzburg free energy. Because in (H.12) we have only summed over those states that correspond to a particular value of $m(\mathbf{r})$. To compute the full partition function, we need to sum over all states. But we can do that by summing over all possible values of $m(\mathbf{r})$. In other words,

$$Q = \int \mathcal{D}m(\mathbf{r})\, \mathrm{e}^{-\beta A[m(\mathbf{r})]} \,. \tag{H.13}$$

This is a tricky beast: it is a functional integral. We are integrating over all possible functions $m(\mathbf{r})$, which is the same thing as performing an infinite number of integrations. (Actually because the order parameter $m(\mathbf{r})$ arose from an underlying lattice and are suitably smooth on short distance scales in our case, the problem is somewhat mitigated).

The result (H.13) is physically very nice, albeit mathematically somewhat daunting. It means that we should view the Landau–Ginzburg free energy as an effective Hamiltonian for a continuous variable $m(\mathbf{r})$, so that by integrating over all possible $m(\mathbf{r})$, we get the partition function. It arises from performing the partition function sum over much of the microscopic information, but still leaves us with a final sum, or integral, over fluctuations in an averaged quantity, namely the order parameter.

To complete the problem, we need to perform the functional integral (H.13). This is hard. Here "hard" means that the majority of the unsolved problems in theoretical physics can be boiled down to performing integrals of this type. Yet the fact it's hard shouldn't dissuade us, since there is a wealth of rich and beautiful physics hiding in the path integral, including the deep reason behind universality. You will explore this if you study statistical field theory.

I An Introduction to Linear-Response Theory

We have derived a number of links between equilibrium properties and transport coefficients, specifically diffusion. We investigated how the approach of a system to equilibrium (diffusion of a particle) and the fluctuations at equilibrium (the root mean square distance travelled by a particle) turns out to be very closely related (by the Einstein relation). In fact, this relation is not limited to diffusion: there is a general set of relations known as *Onsager reciprocal relations*, that express how ratios of flows and forces are related in near-equilibrium systems. We can extend out discussion slightly and formalise the link between equilibrium statistical mechanics and transport coefficients using what is known as the linear-response theory, and derive the relation between the mobility and the diffusion coefficient in another way still.

We first note that the thermal average of an observable X is

$$\langle X \rangle = \frac{\int \mathrm{d}\Gamma \, \mathrm{e}^{-\beta H(\Gamma) X(\Gamma)}}{\int \mathrm{d}\Gamma \, \mathrm{e}^{-\beta H(\Gamma)}} \,, \tag{I.1}$$

where we have denoted a point in the phase space by $\Gamma = (\mathbf{p}_1, \ldots, \mathbf{p}_N, \mathbf{r}_1, \ldots, \mathbf{r}_N)$, and $H = K + U$ is the total Hamiltonian. Now consider a small, linear perturbation λB, so that the perturbed Hamiltonian can be written as $H = H^{(0)} + \lambda B$, where B is some field to be defined later and λ is its strength. We can then decompose the thermal average of X into its unperturbed part $\langle X \rangle_0$ and a change due to the perturbation ΔX such that

$$\langle X \rangle = \langle X_0 \rangle + \langle \Delta X \rangle = \frac{\int \mathrm{d}\Gamma \, \mathrm{e}^{-\beta (H^{(0)} + \lambda B)} X}{\int \mathrm{d}\Gamma \, \mathrm{e}^{-\beta (H^{(0)} + \lambda B)}} \,. \tag{I.2}$$

Now suppose $\langle X \rangle$ is analytic in λ, so that we can write

$$\langle X \rangle = \langle X \rangle_0 + \left(\frac{\partial \langle X \rangle}{\partial \lambda} \right)_{\lambda = 0} \lambda + \frac{1}{2} \left(\frac{\partial^2 \langle X \rangle}{\partial \lambda^2} \right)_{\lambda = 0} \lambda^2 + \ldots \tag{I.3}$$

We will focus on the linear response, i.e. the first order correction in the language of perturbation theory, and write

$$\langle \Delta X \rangle = \left(\frac{\partial \langle X \rangle}{\partial \lambda} \right)_{\lambda = 0} \lambda + O(\lambda^2) \,. \tag{I.4}$$

We can find out an explicit expression for this simply by differentiating $\langle X\rangle$ and get

$$\begin{aligned}\frac{\partial\langle X\rangle}{\partial\lambda} &= \frac{\int \mathrm{d}\Gamma\, \frac{\partial}{\partial\lambda}\mathrm{e}^{-\beta(H^{(0)}+\lambda B)}X}{\int \mathrm{d}\Gamma\, \mathrm{e}^{-\beta(H^{(0)}+\lambda B)}} - \frac{\int \mathrm{d}\Gamma\, \mathrm{e}^{-\beta(H^{(0)}+\lambda B)}X}{[\int \mathrm{d}\Gamma\, \mathrm{e}^{-\beta(H^{(0)}+\lambda B)}]^2}\int \mathrm{d}\Gamma\, \frac{\partial}{\partial\lambda}\mathrm{e}^{-\beta(H^{(0)}+\lambda B)} \\ &= -\beta\frac{\int \mathrm{d}\Gamma\, \mathrm{e}^{-\beta(H^{(0)}+\lambda B)}BX}{\int \mathrm{d}\Gamma\, \mathrm{e}^{-\beta(H^{(0)}+\lambda B)}} + \beta\frac{\int \mathrm{d}\Gamma\, \mathrm{e}^{-\beta(H^{(0)}+\lambda B)}X}{\int \mathrm{d}\Gamma\, \mathrm{e}^{-\beta(H^{(0)}+\lambda B)}}\frac{\int \mathrm{d}\Gamma\, \mathrm{e}^{-\beta(H^{(0)}+\lambda B)}B}{\int \mathrm{d}\Gamma\, \mathrm{e}^{-\beta(H^{(0)}+\lambda B)}} \\ &= -\beta[\langle BX\rangle - \langle B\rangle\langle X\rangle]\,. \end{aligned} \tag{I.5}$$

We can evaluate this at $\lambda = 0$ to obtain the linear response to the perturbation

$$\langle\Delta X\rangle = -\beta\lambda[\langle BX\rangle_0 - \langle B\rangle_0\langle X\rangle_0] + O((\beta\lambda B)^2)\,. \tag{I.6}$$

As an example, let's consider the response of the polarisation $\mathbf{P}$ of a system to a weak static electric field E_x. The perturbative Hamiltonian is $-E_xP_x$, so

$$\langle\Delta P_x\rangle = -\beta E_x[-\langle P_xP_x\rangle_0 + \langle P_x\rangle_0\langle P_x\rangle_0] = \beta E_x[\langle P_x^2\rangle_0 - \langle P_x\rangle_0^2]\,. \tag{I.7}$$

The response to the external electric field therefore scales as $\langle P_x^2\rangle_0 - \langle P_x\rangle_0^2$. This is nothing other than the variance of the spontaneous polarisation without the electric field, which is a measure of the fluctuations of this property in the system at equilibrium. This is in fact quite a general result: the larger the spontaneous fluctuations a mechanical observable, the larger the response to an external field that is coupled to this observable.

Let's now focus on a more general case of a system that we have equilibrated with an active perturbation with Hamiltonian $H = H^{(0)} + \lambda B$, but we switch off this perturbation at time $t = 0$. For notational simplicity, we will assume that in the absence of the field, $\langle X\rangle_0 = 0$. If this were not the case, we could simply define a new variable suitably shifted so that its average is zero, and we are not losing generality here. In classical mechanics, the motion is deterministic, and in the absence of an external force, the value of a mechanical observable at time t depends on the initial conditions, say at $t = 0$. Hence, we may evaluate the thermal average at time t using the integral of the phase space at $t = 0$,

$$\langle\Delta X(t)\rangle = \frac{\mathrm{d}\Gamma(0)\,\mathrm{e}^{-\beta(H^{(0)}+\lambda B)}X(t)}{\int \mathrm{d}\Gamma(0)\,\mathrm{e}^{-\beta(H^{(0)}+\lambda B)}}\,, \tag{I.8}$$

where the Hamiltonian is evaluated at $t = 0$, where $B = B(0)$ is at full strength. The remaining steps are exactly the same as above, and we get

$$\langle\Delta X(t)\rangle = -\beta\lambda[\langle B(0)X(t)\rangle_0 - \langle B(0)\rangle_0\langle X(t)\rangle_0]\,. \tag{I.9}$$

Since we have assumed that in the absence of the field, $\langle X\rangle_0 = 0$, we are left with

$$\langle\Delta X(t)\rangle = -\beta\lambda\langle B(0)X(t)\rangle_0 = -\beta\lambda C_{XB}(t)\,, \tag{I.10}$$

where $C_{XB}(t) \coloneqq \langle B(0)X(t)\rangle_0$ is the correlation function relating the mechanical property X at time t to the perturbative field B at times $t = 0$. For example, in the case of the weak electric field we considered above, when the field is on, there is a net polarisation $P_x(0)$, but the field is switched off, then polarisation decays to zero as $\langle P_x(t)\rangle = \beta E_x\langle P_x(0)P_x(t)\rangle$, where $\langle P_x(0)P_x(t)\rangle$ is the dipole autocorrelation function.

Finally, in order to make a link to the Einstein–Smoluchowski relation, we need to go one step further. Suppose that we write our perturbed Hamiltonian as $H = H^{(0)} + \lambda(t)B$, where $\lambda(t)$ is some function of time that governs the extent to which the perturbation is switched on. In general, we can write the response of the mechanical property X to the linear order in λ as

$$\langle\Delta X(t)\rangle = \int_{-\infty}^{\infty}\mathrm{d}t'\,\chi_{XB}(t,t')\lambda(t')\,, \tag{I.11}$$

where the response function $\chi_{XB}(t,t')$ is to be determined. There are two properties we require of this function:

- The response must be *causal*, i.e. the system cannot respond to the perturbation before it has been applied. Hence the correlation function must be zero if we are looking at correlations with some future behaviour of the system, i.e $\chi(t,t') = 0$ if $t < t'$. We can take this into account by integrating in t' from $-\infty$ to t only, so

$$\langle \Delta X(t) \rangle = \int_{-\infty}^{t} \mathrm{d}t' \, \chi_{XB}(t,t') \lambda(t') \,. \tag{I.12}$$

- The response must be *stationary*, i.e. if we apply the perturbation at a later stage, the response simply occurs correspondingly later. In other words, the response can only depend on the time difference $t - t'$, rather than the absolute time, so

$$\langle X(t) \rangle = \int_{-\infty}^{t} \mathrm{d}t' \, \chi_{XB}(t-t') \lambda(t') \,. \tag{I.13}$$

Assuming that we are dealing with the same case as above, i.e. that $\lambda(t) = \lambda$ for $t < 0$ and 0 otherwise, we can shift the upper limit from t down to zero. Using the substitution $\tau = t - t'$ and swapping the limits of integration, we then have

$$\langle \Delta X(t) \rangle = \lambda \int_{t}^{\infty} \mathrm{d}\tau \, \chi_{XB}(\tau) \,. \tag{I.14}$$

We can equate this with (I.10) to get

$$\beta \, \langle B(0) X(t) \rangle_0 = - \int_{t}^{\infty} \mathrm{d}\tau \, \chi_{XB}(\tau) \,. \tag{I.15}$$

We can differentiate both sides with respect to t to get

$$\chi_{XB}(t) = \begin{cases} \beta \left\langle B(0) \dot{X}(t) \right\rangle_0 & \text{if } t > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{I.16}$$

Notice that since time-correlation functions are stationary,

$$0 = \frac{\partial}{\partial t} \langle B(t) X(t+t') \rangle = \left\langle \dot{B}(t) X(t+t') \right\rangle + \left\langle B(t) \dot{X}(t+t') \right\rangle \,, \tag{I.17}$$

we can alternatively write (I.16) as

$$\begin{aligned} \chi_{XB}(t) &= \begin{cases} -\beta \left\langle \dot{B}(0) X(t) \right\rangle_0 & \text{if } t > 0 \\ 0 & \text{otherwise.} \end{cases} \\ &= -\beta H(t) \left\langle \dot{B}(0) X(t) \right\rangle_0 \,, \end{aligned} \tag{I.18}$$

where $H(t)$ is the Heaviside step function that ensures the causality condition is satisfied.

Finally, we are in the position to consider the diffusion again. Suppose we have applied some force to move the molecules in a non-random direction. If a particle is moving under the influence of a force F_x, then $H = H^{(0)} - F_x x$, so in the notation used so far, $B = x$ and $\lambda = -F_x$. Suppose again that the force acts until a steady state is reached at $t = 0$, at which point it is switched off. The drift velocity $v_x(t)$ averages to zero ($\langle v_x \rangle_0 = 0$) in the absence of an external force, satisfying one of the assumptions we have previously made. Moreover, it is a mechanical observable, so we can write its average in terms of (I.14) as

$$\langle \Delta v_x(t) \rangle = \langle v_x(t) \rangle = -F_x \int_{t}^{\infty} \mathrm{d}\tau \, \chi_{v_x x}(\tau) \,. \tag{I.19}$$

Since $\chi_{v_x x} = -\beta \langle \dot{x}(0) v_x(t) \rangle_0 = -\beta \langle v_x(0) v_x(t) \rangle$ for $t > 0$, we have

$$\langle v_x(t) \rangle = \beta F_x \int_t^\infty \mathrm{d}\tau \, \langle v_x(0) v_x(\tau) \rangle \, . \tag{I.20}$$

Comparing this to the phenomenological steady-state velocity when a force is applied (i.e. at time $t = 0$), $\langle v_x(0) \rangle = u F_x$, where u is the mobility, gives

$$u = \beta \int_0^\infty \mathrm{d}\tau \, \langle v_x(0) v_x(\tau) \rangle \, . \tag{I.21}$$

Using the Green–Kubo relation, we obtain the Einstein–Smoluchowski relation $u = \beta D$.

We have thus derived the Einstein–Smoluchowski relation in another way, this time approaching the problem from an explicit out-of-equilibrium approach. In section 4.5, we obtained the same result from a consideration of steady-state fluxes. We thus have another, more formal way of justifying Onsager's regression hypothesis that spontaneous fluctuations and the approach from a non-equilibrium state towards equilibrium are indistinguishable.

The formalism introduced here entails many steps to obtain results that we were able to find considerably more simply in lectures. We shall leave our discussion of near-equilibrium behaviour at this stage, and so the machinery we have developed in this appendix will not be especially helpful to us in this course. More broadly, however, linear-response theory is a rather useful tool in statistical mechanics and allied fields. One could for example look at the energy dissipation caused by a perturbation, where it is possible to show that the shape of the infra-red absorption spectrum is determined by the dipole correlation function; in fact, every spectroscopic technique probes a correlation function. You will see numerous further examples of linear-response theory in Part III.