# Statistical Physics II

sf2 Fall 2012

Helmut Schiessel

# 1 Introduction and reminder

## 1.1 The partition function

To explain what statistical physics is all about, we start from a simple example we are all familiar with since our childhood: a balloon filled with gas. The physical state of the gas in the balloon can be fully characterized by three physical quantities: (1) The volume $V$ of the balloon, that corresponds to the volume available for the gas. (2) The pressure $p$ that describes how hard one has to press to compress the gas. (3) The temperature $T$ of the gas. It has been known since a long time that these three quantities are related to each other. Robert Boyle found in 1692 that when a fixed amount of gas is kept at contant temperature, then the pressure and volume are inversely proportional, i.e., $p \sim 1/V$. Jacques Charles found in the 1780s that if the pressure of a fixed amount of gas is hold constant, then the volume is proportional to the temperature, i.e., $V \sim T$. And finally Joseph Louis Gay-Lussac stated in 1802 that the pressure of a fixed amount of gas in a fixed volume is proportional to the temperature, i.e., $p \sim T$. You can easily check that these three laws are all fulfilled if the ratio between the pressure-volume product and the temperature is constant:

$$\frac{pV}{T} = const. \tag{1}$$

That means if we look at a gas at two different states, $(p_1, V_1, T_1)$ and $(p_2, V_2, T_2)$, we always find $p_1V_1/T_1 = p_2V_2/T_2$. What is that value, i.e., the value of the constant on the right hand side (rhs) of Eq. 1? The value depends on the amount of gas inside the balloon. An amount of gas that occupies $V = 22.4$ litres at $T = 0°C = 273.15K$ and atmospheric pressure, $p = 1.013$ bar, is called one *mole*. The constant then takes the value

$$R = 8.31\frac{J}{K} \tag{2}$$

and is called the *universal gas constant*. If the amount of gas in the balloon is $n$ moles, then the constant in Eq. 1 has the value $nR$.

Equation 1, the so-called *combined gas law*, is an example of an empirical law that relates measurable physical quantities. Statistical physics is the theoretical framework that allows us to derive such laws from first principles. This is a quite daunting task. A gas is a collection of a huge number of particles. We nowadays know that one mole of gas contains $N_A = 6.02 \times 10^{23}$ particles (independent of the type of gas chosen; normal air, helium, etc.) where $N_A$ is called the *Avogadro*

*constant*. This is a rather mindblowing fact: a balloon which contains one mole of particles can be fully characterized by three so-called *macroscopic variables*, $p$,$V$ and $T$, yet it has a myriad of microstates characterized by the positions and velocities (both in $X$-, $Y$- and $Z$-direction) of $6 \times 10^{23}$ gas molecules!

Let us try to derive Eq. 1 from the microscopic structure of the gas. This will serve as a concrete example to introduce the methods of statistical physics. As a first step we introduce the very high-dimensional *phase space* of our system, that contains the positions and momenta of all the $N$ particles. A point in phase space is given by $(x_1, y_1, z_1, ..., x_N, y_N, z_N, p_1^x, p_1^y, p_1^z, ..., p_N^x, p_N^y, p_N^z)$, where e.g. $y_i$ and $p_i^y$ denote the position and momentum of the $i$th particle, both in the $Y$-direction. In short hand notation we can write $(\mathbf{q}, \mathbf{p})$ for such a point in phase space where $\mathbf{q}$ is a high dimensional vector that contains all the positions and $\mathbf{p}$ all the momenta of the $N$ particles. The amazing thing is that as the gas molecules move inside the balloon and bounce off its surface, i.e., as the point $(\mathbf{q}, \mathbf{p})$ races through the phase space, we can not see anything happen to the balloon in our hands, that stays quiet at a constant pressure, volume and temperature.

This suggests the following: To a given macrostate characterized by the triplet $(p, V, T)$ there is a myriad of microstates, each characterized by a high-dimensional vector $(\mathbf{q}, \mathbf{p})$. But not all possible microstates should have the same probability. As it is highly unlikely (but in principle not impossible) to throw a dice one billion times and to find a six each time, so it is highly unlikely that at a certain point in time all the $6 \times 10^{23}$ particles are in the left half of the balloon. We thus need to introduce the concept of probabilities by assigning to each microstate $(\mathbf{p}, \mathbf{q})$ a probability $\rho = \rho\,(\mathbf{q}, \mathbf{p})$.

We now present a line of argument that allows us to determine the form of $\rho$, namely Eq. 5 below. Please be warned that even though each of the steps looks rather compact, it is not easy to grasp them entirely. At this stage you might rather consider this is as a rough outline, providing you with a rather general view of things and allowing to quickly get to something concrete to work with. You do not have to feel too uncomfortable with this, since we shall later on provide a completely different argument that again leads us to Eq. 5.

In Fig. 1 we show the balloon with its gas molecules; we call this whole system $\Sigma$. We consider now two subsystem, $\Sigma_1$ and $\Sigma_2$, namely the molecules to the left and to the right of a virtual dividing plane as indicated in the figure by a dashed line. Real gas molecules have a very short range of interaction that is much shorter than the diameter of the balloon. This means that only a very tiny fraction of the molecules in $\Sigma_1$ feel molecules from $\Sigma_2$ and vice versa. Therefore, to a good approximation, the two subsystems can be considered as independent from each other. We can thus separately, for each subsystem, define probability densities $\rho_1$ and $\rho_2$ – without going here further into mathematical details. Now since $\Sigma_1$ and $\Sigma_2$ are independent the probability of the whole system is simply the product of the probabilities of its subsystems, $\rho = \rho_1 \rho_2$ (just as for two dice; the probability to throw a 6 amounts for each dice to 1/6 and the probability that both dice yield a 6 is $1/6 \times 1/6 = 1/36$). Using the functional property of

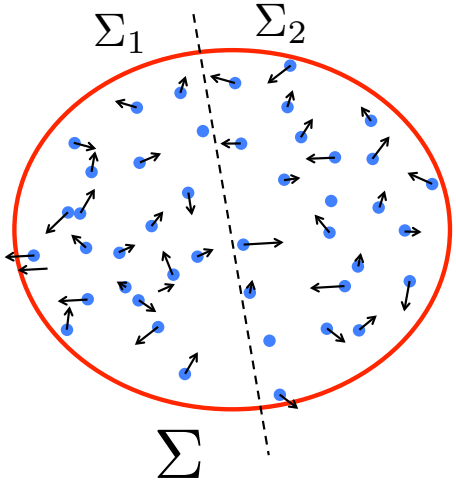



Figure 1: A balloon filled with gas molecules. The virtual line divides the whole balloon (system $\Sigma$) into two subsystems, its left half, $\Sigma_1$, and its right half, $\Sigma_2$. To a good approximation these two halves are independent from each other, as mathematically expressed in Eq. 3.

the (natural) logarithm, $\ln ab = \ln a + \ln b$, this can be rewritten as:

$$\ln \rho = \ln \rho_1 + \ln \rho_2. \tag{3}$$

This is one of the conditions that $\rho$ needs to fulfill. A second one is the following. Here and in the rest of this lecture we are considering systems in *equilibrium*. What we mean by this is that the system has evolved to a state where nothing changes anymore. For our balloon this means that the values of $p$, $V$ and $T$ stay constant in time (unlike e.g. a glass of water where all the water evaporates if you wait long enough). Likewise microscopically nothing changes anymore, i.e., the function $\rho = \rho(\mathbf{q}, \mathbf{p}, t)$ does not explicitly depend on time but is of the form $\rho = \rho(\mathbf{q}, \mathbf{p})$, as we had written it in the first place. In other words

$$\frac{\partial \rho}{\partial t} = 0. \tag{4}$$

Amazingly Eqs. 3 and 4 are enough to determine $\rho$. We know from Eq. 4 that $\rho$ is a conserved quantity, meaning a quantity that does not change in time. $\rho$ must thus be a function of a conserved physical quantity. Possible candidate quantities are: (a) the total energy $H$ of the system, (b) its total momentum $\mathbf{P}$ and (c) the particle number $N$ (for different types of particles the numbers $N_\alpha$ of each type). Most systems are confined by walls (e.g. a gas in a balloon). Whenever a gas molecule hits the balloon, it gets reflected and thereby transmits momentum to the balloon; thus $\mathbf{P}$ of the gas is not conserved. This means $\rho$ can only depend on $H$ and $N$. From Eq. 3 we know that $\ln \rho$ is an additive

quantity and so is the energy $H$, $H_\Sigma = H_{\Sigma_1} + H_{\Sigma_2}$ and the particle number $N$, $N_\Sigma = N_{\Sigma_1} + N_{\Sigma_2}$. This means we know more about how $\ln \rho$ should depend on $H$ and $N$, namely it must be a linear function of additive, conserved quantities.

This leaves several possibilities for $\ln \rho$ that depend on the concrete physical situation. For the balloon the number of particles inside the balloon is fixed since the gas molecules cannot pass through the balloon skin. However, energy can flow in and out of the balloon in the form of heat. In that case we should expect that $\ln \rho \sim -\beta H$ where $\beta$ is some constant. If in addition particles can move in and out, one should expect that $\ln \rho \sim +\alpha N - \beta H$ with $\alpha$ being yet another constant. The plus and minus signs here are just conventions and do not mean anything since we do not yet know the signs of $\alpha$ and $\beta$.

Let us begin with the first case, the one with $N$ fixed. Such a system is called the *canonical ensemble*. From above we know that $\rho$ must be of the form:

$$\rho\left(\mathbf{q}, \mathbf{p}\right) = \frac{1}{Z} e^{-\beta H(\mathbf{q},\mathbf{p})} \tag{5}$$

where the function $H = H\left(\mathbf{q}, \mathbf{p}\right)$ is the energy of the system that depends on the positions and momenta of all the particles. The role of the prefactor $1/Z$ is to normalize the probability distribution such that the sum over all different possible states of the system adds up to one. Surprisingly this seemingly harmless prefactor is the whole key to understand the properties of the system as we shall see below. As it turns out to be so important, it should not surprise you that it has a name: the *partition function*. We need to choose $Z$ such that

$$\frac{1}{N! h^{3N}} \int \rho\left(\mathbf{q}, \mathbf{p}\right) d^{3N} q \, d^{3N} p = 1 \tag{6}$$

and hence

$$Z = \frac{1}{N! h^{3N}} \int e^{-\beta H(\mathbf{q},\mathbf{p})} d^{3N} q \, d^{3N} p. \tag{7}$$

The prefactor $1/N!h^{3N}$ in front of the integrals in Eqs. 6 and 7 seems to be an unnecessary complication in the notation and needs some explanation. Let us start with the factor $1/N!$. This corresponds to the number of possible ways one could number the $N$ particles (we pick a particle and give it a number between 1 and $N$, then the second particle and give it one of the $N-1$ remaining tags and so on). If the microscopic world would behave classically (like the macroscopic world we are used to live in), we can give each of the $N$ gas molecules such an individual tag and follow its course in time. That way the two configurations shown in Fig. 2(a) are different from each other, since particles 1 and 2 are exchanged. However, the microscopic world of these particles is governed by the laws of quantum mechanics. One of these laws is that identical particles are indistinguishable, in other words the two conformations shown in Fig. 2(a) are identical and belong to exactly the same physical state, the one shown to the right, Fig. 2(b). When performing the integrals $\int d^{3N} q d^{3N} p$ in Eqs. 6 and 7 one would encounter $N!$ times such a configuration. The prefactor $1/N!$ prevents this overcounting.

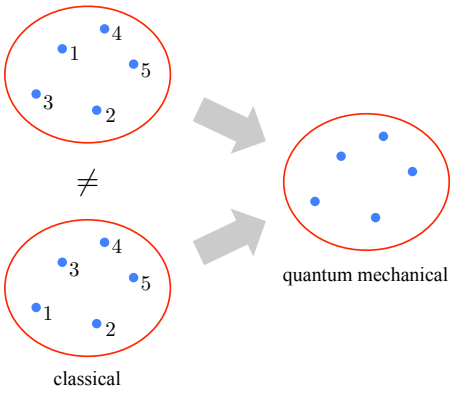



Figure 2: A balloon with $N = 5$ identical gas molecules. (a) In classical mechanics we can number the particles individually allowing us to distinguish between the configurations shown in the top and in the bottom. (b) In quantum mechanics identical particles are indistinguishable, which means that the two states that are shown on the left are one and the same, namely the configuration depicted at the right. This quantummechanical law is the cause of the $1/N!$ factor in Eq. 7.

Next we discuss the factor $h^{3N}$. This factor is introduced to make $Z$ dimensionless, i.e., no matter what units we use (e.g. meters or inches for length) $Z$ is always the same. $h$ is a quantity with the dimensions of length times momentum (or equivalently energy times time), namely

$$h = 6.626 \times 10^{-34} Js. \tag{8}$$

Even though this choice seems arbitrary from the viewpoint of classical mechanics it can be motivated to be the most logical choice in the realm of quantum mechanics. The quantity $h$ is the so-called *Planck constant* that appears in a famous relation in quantum mechanics: It is impossible to measure the position and momentum of a particle beyond a certain precison. According to the so-called *Heisenberg's uncertainty principle* the uncertainty in position, $\Delta x$, and in momentum, $\Delta p^x$, both in $X$-direction, obey the relation $\Delta x \Delta p^x \geq h/4\pi$ (more precisely $\Delta x$ and $\Delta p$ are the standard deviations found when the measurements are repeated again and again under identical conditions). So if one measures the position of a particle very precisely, there is a large uncertainty in its momentum and vice versa, a consequence of the particle-wave duality that we shall not discuss here further. Because of this it makes sense to divide our $6N$-dimensional space in small hypercubes of volume $h^{3N}$ which explains the choice of the prefactor in Eq. 7.

To give a concrete example we calculate the partition function of the gas in the balloon, Fig. 1. Before we can start to evaluate the integral, Eq. 7, we need to have an expression for the energy of the gas, $H = H(\mathbf{q}, \mathbf{p})$, its *Hamiltonian*. We consider here *ideal gas*, an idealization of a real gas. In this model the

interaction between different gas molecules is neglected altogether. This turns out to be an excellent approximation for most gases since the concentration of gas molecules is so low that they hardly ever feel each other's presence. This means that the energy is independent of the distribution of the molecules in space, i.e., $H = H(\mathbf{q}, \mathbf{p})$ does not depend on $\mathbf{q}$. This leaves us just with the kinetic energy of the particles (assumed here to all have the same mass $m$):

$$H = H(\mathbf{p}) = \sum_{i=1}^{N} \frac{p_i^2}{2m} = \frac{1}{2m} \sum_{i=1}^{N} \left[ (p_i^x)^2 + (p_i^y)^2 + (p_i^z)^2 \right] \tag{9}$$

with $p_i = |\mathbf{p}_i|$ being the length of the momentum vector $\mathbf{p}_i = (p_i^x, p_i^y, p_i^z)$. Plugging this into Eq. 7 we realize that we have Gaussian integrals. The momentum integration of each of the 3 components of each particle gives a factor $\sqrt{2\pi m/\beta}$. In addition each particle is allowed to move within the whole balloon so that its position integration gives a factor $V$. Altogether this leads to

$$Z = \frac{V^N}{N! h^{3N}} \int e^{-\frac{\beta}{2m} \sum_{i=1}^{N} p_i^2} d^{3N}p = \frac{V^N}{N! h^{3N}} \left( \frac{2\pi m}{\beta} \right)^{3N/2}. \tag{10}$$

It is customary to introduce a quantity called the *thermal de Broglie wave length*

$$\lambda_T = h \sqrt{\frac{\beta}{2\pi m}} \tag{11}$$

that allows us to write the partition function $Z$ of the ideal gas very compactly:

$$Z = \frac{1}{N!} \left( \frac{V}{\lambda_T^3} \right)^N. \tag{12}$$

We introduced the partition function in Eq. 5 merely as a prefactor necessary to normalize the probability distribution, but we mentioned already that then one can derive from $Z$ almost everything one would like to know about the macroscopic system. As a first example we show now that knowing $Z$ means that it is straightforward to determine $E = \langle H \rangle$, the average energy of the system:

$$\begin{aligned} \langle H \rangle &= \frac{\int H(\mathbf{q}, \mathbf{p})\, e^{-\beta H(\mathbf{q},\mathbf{p})} d^{3N}q\, d^{3N}p}{\int e^{-\beta H(\mathbf{q},\mathbf{p})} d^{3N}q\, d^{3N}p} \\ &= \frac{1}{Z} \frac{1}{N! h^{3N}} \int H(\mathbf{q}, \mathbf{p})\, e^{-\beta H(\mathbf{q},\mathbf{p})} d^{3N}q\, d^{3N}p. \end{aligned} \tag{13}$$

Here the denominator is necessary to normalize the canonical distribution and is, of course, again proportional to the partition function. It seems at first that the integral on the rhs of Eq. 13 needs to be evaluated all over again. However, the beauty of the partition function $Z$ is that it is of such a form that it allows expressions such as Eq. 13 to be obtained from it by straightforward differentiation. You can easily convince yourself that one has

$$E = \langle H \rangle = -\frac{\partial}{\partial \beta} \ln Z. \tag{14}$$

The differentiation of the ln-function produces the $1/Z$ prefactor on the rhs of Eq. 13 and the form of its integrand, $He^{-\beta H}$, follows simply from the differentiation, $-\partial e^{-\beta H}/\partial\beta$. This means all the hard work lies in calculating $Z$ through a high-dimensional integral, Eq. 7. Once this is done, the harvest consists of straightforward differentiation as in Eq. 14.

We can also calculate the variance of the energy fluctuations of the gas. These fluctuations result from the exchange of heat with the surrounding air outside the balloon that constitutes a so-called *heat bath*. This quantity is $\sigma_E^2 = \langle H^2 \rangle - \langle H \rangle^2$ and follows simply by differentiation of $\ln Z$ twice:

$$\begin{aligned} \sigma_E^2 &= \frac{\partial^2}{\partial\beta^2} \ln Z = -\frac{\partial}{\partial\beta} \langle H \rangle \\ &= \langle H^2 \rangle - Z \langle H \rangle \frac{\partial}{\partial\beta} \frac{1}{Z} = \langle H^2 \rangle - \langle H \rangle^2 . \end{aligned} \tag{15}$$

To arrive at the second line we used Eq. 13; the first term accounts for the $\beta$-dependence inside the integral, the second for that of the $Z^{-1}$ prefactor.

Since we already calculated the partition function of the ideal gas, Eq. 10, we can immediately obtain, via Eq. 14, its average energy:

$$E = \langle H \rangle = \frac{3}{2} \frac{N}{\beta}. \tag{16}$$

The energy is thus proportional to the particle number $N$, as one should expect for non-interacting particles, and inversely proportional to the quantity $\beta$. We still do not know the physical meaning of that quantity – even though, as we shall soon see, it is well-known to us; we even have a sensory organ for it. For now we can only give $\beta$ a rather technical meaning: It allows us, via Eq. 16, to set the average energy $\langle H \rangle$ of the gas to a given value.

We can now also calculate the typical relative deviation of the energy from its mean value $\langle H \rangle$. It follows from Eqs. 10 and 15 that

$$\frac{\sigma_E}{\langle H \rangle} = \sqrt{\frac{2}{3}} \frac{1}{\sqrt{N}}. \tag{17}$$

This means that for large systems the relative fluctuations around the mean value are so tiny that the system is, for any practical purposes, indistinguishable from a *microcanonical ensemble,* a system that is thermally isolated, i.e., that cannot exchange energy with the outside world.

Our aim is now to derive an equation for the pressure of the ideal gas and to check whether statistical mechanics allows us to derive from first principles the combined gas law, Eq. 1. To make the analysis more convenient we put the gas in a cylinder with a movable piston, Fig 3, instead of a balloon. If we apply a force $f$ on the piston, then the pressure on it is given by $p = f/A$ where $A$ is the area of the piston. The gas occupies a volume $V = Al$ with $l$ denoting the height of the piston above the bottom of the cylinder. To better understand how the gas can exert a force on the piston we add to the Hamiltonian $H(\mathbf{q}, \mathbf{p})$ a wall

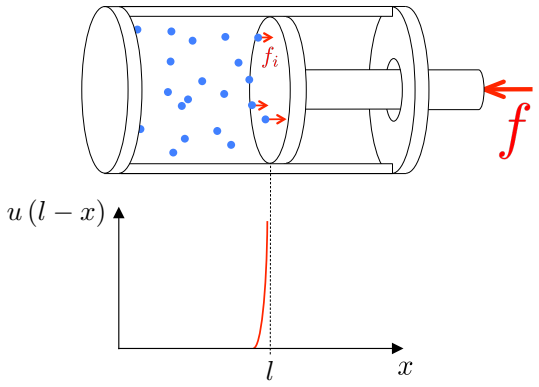



Figure 3: Gas in a cylinder. The piston is under an externally imposed force $f$ that counterbalances the individual forces $f_i$ of the gas molecules close to the surface of the piston. Each of these forces follows from a short ranged wall potential $u$ that smoothly goes to infinity as the gas molecule reaches the surface of the piston.

potential $U_{\text{wall}}(l, \mathbf{q})$ that depends on the positions of all the particles and on the height $l$ of the piston. We do not assume anything here about the form of the Hamiltonian $H(\mathbf{q}, \mathbf{p})$ so the following formulas are general. The wall potential $U_{\text{wall}}(l, \mathbf{q})$ takes an infinite value if any of the molecules is outside the allowed volume. This way the gas is forced to stay inside the cylinder. To calculate the force exerted by the gas molecules we assume that the potential goes smoothly to infinity over a microscopically short distance $\delta$ when a particle gets close to the surface of the piston (for the other confining walls we simply assume that the potential jumps right to infinity). More specifically, the wall potential is of the form

$$U_{\text{wall}}(l, \mathbf{q}) = \sum_{i=1}^{N} u(l - x_i) \tag{18}$$

as long as all particles are inside the cylinder and infinity otherwise. Most particles are far from the surface of the piston, $l - x_i > \delta$, and thus do not feel it, i.e., $u(l - x_i) = 0$. But a small fraction of them are closeby, $l - x_i < \delta$, and they are pushed to the left exerting a force on the piston. For a given configuration of particles, $\mathbf{q} = (x_1, y_1, z_1, ..., x_N, y_N, z_N)$ this force is given by

$$f = -\frac{\partial U_{\text{wall}}(l, \mathbf{q})}{\partial l} = -\sum_{i=1}^{N} \frac{\partial u(l - x_i)}{\partial l}. \tag{19}$$

We are, however, interested in the mean force $\langle f \rangle$ that is given by

$$\langle f \rangle = \frac{1}{Z} \frac{1}{N! h^{3N}} \int d^{3N} q d^{3N} p \left( -\frac{\partial U_{\text{wall}}(l, \mathbf{q})}{\partial l} \right) e^{-\beta [H(\mathbf{q}, \mathbf{p}) + U_{\text{wall}}(l, \mathbf{q})]}. \tag{20}$$

This expression might look complicated, but again it is just a simple derivative of the partition function, namely

$$\langle f \rangle = \frac{1}{\beta} \frac{\partial \ln Z}{\partial l}. \tag{21}$$

This is the average force that is exerted by the gas on the piston (and *vice versa*). Using the relations $p = f/A$ and $V = Al$ we can immediately write down the relation for the pressure:

$$\langle p \rangle = \frac{\langle f \rangle}{A} = \frac{1}{\beta} \frac{\partial \ln Z}{\partial V}. \tag{22}$$

We can now use Eq. 22 to determine the pressure of an ideal gas. When calculating its partition function in Eq. 10 we did not take account of a detailed wall potential. But since the wall potential increases over a microscopically small distance $\delta \ll l$, the partition function is not affected by such details. Using Eq. 12 we find

$$\langle p \rangle = \frac{N}{\beta V}. \tag{23}$$

Comparison with the combined gas law, Eq. 1, lets us finally understand the physical meaning of $\beta$: it is inversely proportional to the temperature:

$$\beta = \frac{1}{k_B T}. \tag{24}$$

The quantity $k_B$ is called the *Boltzmann constant*. From Eq. 1 together with Eq. 2 follows its value

$$k_B = \frac{R}{N_A} = 1.38 \times 10^{-23} \frac{J}{K}. \tag{25}$$

To summarize we have found two equations that characterize an ideal gas. From Eq. 16 we find for the energy

$$E = \frac{3}{2} N k_B T \tag{26}$$

and from Eq. 23 we obtain the ideal gas equation of state

$$pV = N k_B T. \tag{27}$$

The first relation, Eq. 26, states that each gas molecule has on average an energy of $(3/2)\, k_B T$, this is, as we can see from Eq. 9, its kinetic energy. The temperature of a gas is thus a measure of the average kinetic energy of its molecules that move on average faster inside a hotter gas. The second relation states how these molecules exert a force when they bounce off the inner side of the wall of the balloon, Fig. 1, or the piston, Fig. 3. The hotter the gas the faster the gas molecules and the larger the transferred momentum during

collision. The larger the volume, the longer the time before the molecule hits the wall again and thus the lower the average pressure.

The quantity $k_BT$ is called the *thermal energy*. At room temperature, $T = 293K$, one has

$$k_BT = 4.1pNnm. \tag{28}$$

It is worthwhile to remember this formula by heart (instead of Eqs. 2 and 25).

Let us now come to the second case, the case of a system that exchanges energy and particles with its surroundings. In this case only the expectation values of the energy, $E = \langle H \rangle$, and the particle number, $N = \langle N \rangle$, can be given. This is the so-called *grandcanonical ensemble*. In that case we expect a density distribution $\rho$ of the form:

$$\rho = \frac{1}{Z_G} e^{\alpha N - \beta H}. \tag{29}$$

The grandcanonical partition function is a summation and integration over all possible states of the system, each state weighted with $\rho$. This means we have to sum over all particle numbers and then, for each number, over the positions and momenta of all the particles:

$$Z_G = \sum_{N=0}^{\infty} \frac{1}{h^{3N} N!} \int e^{\alpha N - \beta H(\mathbf{q},\mathbf{p})} d^{3N}q \, d^{3N}p. \tag{30}$$

This can be rewritten as

$$Z_G = \sum_{N=0}^{\infty} e^{\alpha N} Z_N = \sum_{N=0}^{\infty} z^N Z_N \tag{31}$$

where $Z_N$ is the canonical partiton function of a system of $N$ particles, i.e., the quantity that we called $Z$ in Eq. 7. On the rhs of Eq. 31 we introduced the so-called *fugacity* $z = e^{\alpha}$. It is straightforward to see, using similar arguments as the ones that led to Eqs. 14 and 15, that

$$\begin{aligned} E = \langle H \rangle &= -\frac{\partial}{\partial \beta} \ln Z_G, \qquad \sigma_E^2 = \frac{\partial^2}{\partial \beta^2} \ln Z_G \\ N = \langle N \rangle &= \frac{\partial}{\partial \alpha} \ln Z_G, \qquad \sigma_N^2 = \frac{\partial^2}{\partial \alpha^2} \ln Z_G. \end{aligned} \tag{32}$$

For large $N$ the relative fluctuations in energy and particle number, $\sigma_E/E$ and $\sigma_N/N$, become so small (just as in Eq. 17) that the grandcanonical ensemble with mean energy $E$ and mean particle number $N$ becomes physically equivalent to the canonical ensemble with mean energy $E$ and exact particle number $N$. It is thus just a matter of convenience which ensemble one chooses. Many calculations are more convenient in the grandcanonical ensemble since one does not have such a strict condition on $N$.

Let us again consider the ideal gas. Inserting Eq. 12 into Eq. 31 we find its grandcanonical partition function

$$Z_G = \sum_{N=0}^{\infty} z^N Z_N = \sum_{N=0}^{\infty} \frac{1}{N!} \left( \frac{zV}{\lambda_T^3} \right)^N = e^{\frac{zV}{\lambda_T^3}}. \tag{33}$$

The expectation value of the particle number follows from Eq. 32

$$N = \frac{\partial}{\partial \alpha} \ln Z_G = \frac{zV}{\lambda_T^3} \tag{34}$$

and that of the energy as well

$$E = -\frac{\partial}{\partial \beta} \ln Z_G = \frac{3}{2} k_B T \frac{zV}{\lambda_T^3} = \frac{3}{2} N k_B T. \tag{35}$$

This is equivalent to Eq. 26 but $N$ is now strictly speaking $\langle N \rangle$. The pressure formula, Eq. 27 follows even more directly from these relations as we shall see later below (cf. Eq. 62).

## 1.2 The entropy

In stf1 you learned about a quantity that is crucial for the understanding of macroscopic systems: the *entropy*. As we shall see, the concept of entropy allows for a different, more convincing argument for the Boltzmann distribution, Eq. 5. But before we come to that we start with a simple model system where it is quite straightforward to grasp the ideas behind entropy, especially the relation between a macroscopic state and its associated microscopic states.

The following system can be considered as an idealization of a so-called *paramagnet*. A paramagnet is a substance that consists of atoms that have magnetic dipole moment. The different dipoles do not feel each other and point in random directions. As a result such a system shows no net macroscopic magnetization. The model consists of a collection of microscopic so-called *spins* on a lattice as shown in Fig. 4. Each spin represents an atom sitting on the lattice of a solid – in contrast to a gas, Fig. 1, where the atoms can move freely in space. We call the spin at the $i$th site $s_i$ and assume that it can take either the value $+1$ or $-1$ with a corresponding magnetic moment $+\mu$ or $-\mu$. This leads to the overall *magnetization*

$$M = \mu \sum_{i=1}^{N} s_i. \tag{36}$$

We assume that the spins do not interact with each other. We also assume that there is no energy change involved when a spin flips from one value to the other. This means that all states have exactly the same energy. Therefore each microscopic state $\{s_1, s_2, ..., s_N\}$ is as good as any other. The spins in a paramagnet permanently flip back and forth due to the thermal environment.

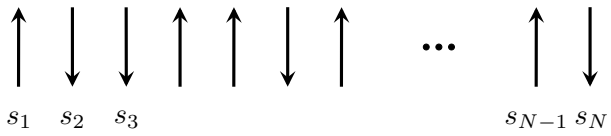



Figure 4: A system of $N$ noninteracting spins. Each spin can either point up or down.

We should thus expect when we look long enough at such a system to measure any value of $M$ between $-\mu N$ and $+\mu N$. However, for a large system, $N \gg 1$, a paramagnetic substance always ("always" not in the strict mathematical sense but almost always during the lifetime of the universe) shows an extremely small value, $|M| \ll \mu N$. How is this possible?

To understand this we have to look at the possible number of microstates that correspond to a given macrostate, i.e., a state with a given value $M$ of magnetization. If we find a macrostate $M$, then there must be $k$ spins pointing up (and hence $N-k$ spins pointing down) such that

$$M = \mu k - \mu\,(N-k) = \mu\,(2k-N)\,. \tag{37}$$

Let us determine the number of microstates that have this property. This is a simple problem in combinatorics. There are

$$\left(\begin{array}{c} N \\ k \end{array}\right) = \frac{N!}{k!\,(N-k)!} \tag{38}$$

possible combinations of spins where $k$ spins point up and $N-k$ point down. The point is now that for large $N$ there are overwhelmingly more configurations that lead to a vanishing $M$, $k = N/2$, then there are states for which $M$ takes its possible maximal value, $M = \mu N$. For the latter case there is obviously only one such state, namely all spins pointing up, whereas the former case can be achieved in $\left(\begin{array}{c} N \\ N/2 \end{array}\right)$ different ways. To get a better understanding of how big this number is, we employ *Stirling's formula* that gives the leading behavior of $N!$ for large values of $N$:

$$N! \overset{N\to\infty}{\to} \left(\frac{N}{e}\right)^N \sqrt{2\pi N}. \tag{39}$$

Equation 39 holds up to additional terms that are of the order $1/N$ smaller and can thus be neglected for large values of $N$. Combining Eqs. 38 and 39 it is straightforward to show that the number $N_{\max}$ of spin configurations that lead to $M = 0$ obeys

$$N_{\max} = \left(\begin{array}{c} N \\ N/2 \end{array}\right) \approx \sqrt{\frac{2}{\pi}}\frac{2^N}{N^{1/2}}. \tag{40}$$

As you can see $N_{\max}$ grows exponentially with $N$, $N_{max} \sim 2^N$. Macroscopic systems may contain something like $10^{23}$ spins which means that there is an

astronomically large number of states with $M = 0$ (namely a 1 with $10^{22}$ zeros), compared to one state with $M = \mu N$.

Let us call $N_{\text{micro}}(M)$ the number of microstates corresponding to a given macrostate characterized by $M$. One can show that to a good approximation

$$N_{\text{micro}}(M) = N_{\max} e^{-\frac{M^2}{2\mu^2 N}}. \tag{41}$$

This function is extremely peaked around $M = 0$ with the value $N_{\max}$ given by Eq. 40. It decays rapidly when one moves away from $M = 0$, e.g. it has decayed to $N_{\max}/e$ for $M = \pm\mu\sqrt{2N}$, a value much smaller than the maximal possible magnetization $\pm\mu N$. Suppose we could somehow start with some macroscopic state with a large value of $M$. Over the course of time the spins flip back and forth randomly. Given enough time it is overwhelmingly probable that $M$ will have values that stay in an extremely narrow range around $M = 0$, simply because there are so many more microstates available with tiny $M$-values than with larger $M$-values. Therefore it is just an effect of probabilities that a paramagnetic substance shows (close to) zero magnetization.

We can formulate this in a slightly different way. A macroscopic system will go to that state where there is the largest number of microstates to a given macrostate. This state is called the *equilibrium state* since once the system has reached this state it does not leave it anymore – not because this is in principle impossible, but because it is overwhelmingly improbable. We can also say the following: Of all the possible macroscopic states, the system chooses the one for which our ignorance of the microstate is maximal. If we measure $M = \mu N$ we would know for sure the microstate of the system, but if we measure $M = 0$ we only know that our system is in one of about $2^N$ (see Eq. 40) possible states.

We introduce a quantity that measures our ignorance about the microstate. If we require that this quantity is additive in the sense that if we have two independent (sub)systems our ignorance of the two systems is simply the sum of the two, then we should choose this quantity to be given by

$$S = k_B \ln N_{\text{micro}}. \tag{42}$$

The prefactor is in principle arbitrary, yet it is convention to choose it equal to the Boltzmann constant $k_B$, the quantity introduced in Eq. 25. A macroscopic system will always – given enough time – find the macroscopic state that maximizes its entropy. Let us reformulate Eq. 42. Suppose we know the macrostate of the system, namely that there are $k$ spins pointing up. Then each of the microstates corresponding to that macrostate has the same probability $p_k = 1/N_{\text{micro}}$. We can then rewrite Eq. 42 as follows

$$S = -k_B \ln p_k. \tag{43}$$

When $k$, and therefore $M$, changes, the entropy changes. Since the entropy is extremely sharply peaked around $k = N/2$, the system will spontaneously reach states around $k \approx N/2$ and never deviate from this anymore, not because it is forbidden, but because it is extremely improbable.

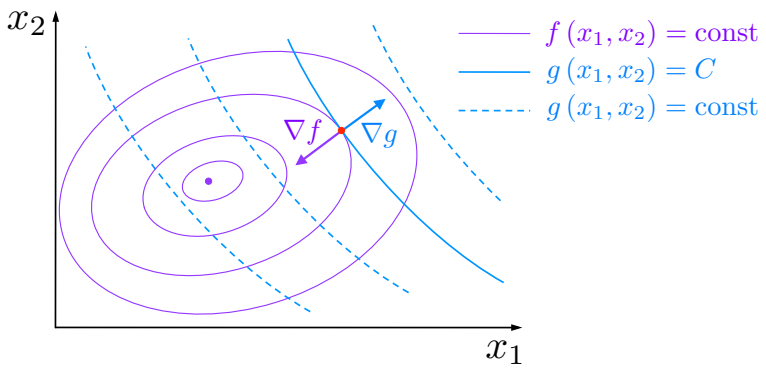



Figure 5: The method of the Lagrange multiplier. The objective is to find the maximum of the function $f(x_1, x_2)$ under the constraint $g(x_1, x_2) = C$. Shown are lines of equal height of $f$ (purple curves) and of $g$ (blue curves). The red point indicates the maximum of interest. It is the highest point of $f$ on the line defined by $g = C$. At this point the gradients of the two height profiles are parallel or antiparallel (case shown here). This means there exists a number $\lambda \neq 0$, called the Lagrange multiplier, for which $\nabla f = \lambda \nabla g$.

The goal in the following is to extend the concept of entropy to a system like our gas in a balloon. In such a case we also expect that the system goes to a macrostate with the largest number of microstates or, in other words, to the macrostate for which we know least about the microstate, the state of maximal entropy. In this case there is, however, a complication. We had required that the average energy has a certain value, $\langle H \rangle = E$, cf. Eq. 16. So we need to maximize the entropy with the constraint

$$\langle H \rangle = \sum_i p_i E_i = E. \tag{44}$$

Here we assume that the states are discrete, which – as outlined above – should in principle always be assumed due to the uncertainty principle. We already know from the previous section that the probabilities of states with different energies are different. Extending Eq. 43 we now define the entropy as our average ignorance about the system:

$$S = -k_B \sum_i p_i \ln p_i. \tag{45}$$

What we need to do is to maximize $S$, Eq. 45, under the constraint of having a certain average energy, Eq. 44. This can be achieved using the method of *Lagrange multipliers*. Suppose you want to maximize the function $f(x_1, ..., x_m)$. If this function has a maximum it must be one of the points where the function has zero slope, i.e., where its gradient vanishes: $\nabla f = 0$ with $\nabla = (\partial/\partial x_1, ..., \partial/\partial x_m)$. What do we have to do, however, if there is an additional

constraint, $g(x_1, ..., x_m) = C$ with $C$ some constant? This constraint defines an $(m-1)$-dimensional surface in the $m$-dimensional parameter space. Figure 5 explains the situation for $m = 2$. In that case $f(x_1, x_2)$ gives the height above (or below) the $(x_1, x_2)$-plane. As in a cartographic map we can draw contour lines for this function. The constraint $g(x_1, x_2) = C$ defines a single line $g_C$ (or combinations thereof) in the landscape. The line $g_C$ crosses contour lines of $f$. We are looking for the highest value of $f$ on $g_C$. It is straightforward to convince oneself that this value occurs when $g_C$ touches a contour line of $f$ (if it crosses a contour line one can always find a contour line with a higher value of $f$ that still crosses the $g_C$-line). Since $g_C$ and the particular contour line of $f$ touch tangentially, the gradients of the two functions at the touching point are parallel or antiparallel. In other words, at this point a number $\lambda$ exists (positive or negative), called the Lagrange multiplier, for which

$$\nabla (f - \lambda g) = 0. \tag{46}$$

Let us use this method in the context of the entropy. We want to find the maximum of $S/k_B$, a function depending on the parameters $(p_1, ..., p_{N_{\text{tot}}})$ where $p_i$ denotes the probability of the $i$th of the $N_{\text{tot}}$ microstates. In addition we need to fulfill the constraint 44. This leads to a condition equivalent to Eq. 46, namely

$$\nabla (S/k_B - \beta \langle H \rangle) = 0, \tag{47}$$

with $\nabla = (\partial/\partial p_1, ..., \partial/\partial p_{N_{\text{tot}}})$ and $\beta$ a Lagrange multiplier. For each $i$, $i = 1, ..., N_{\text{tot}}$, we find the condition

$$\frac{\partial}{\partial p_i} \left( \frac{S}{k_B} - \beta \langle H \rangle \right) = -\ln p_i - 1 - \beta E_i = 0. \tag{48}$$

This leads to $p_i \sim e^{-\beta E_i}$ which then still needs to be normalized to one, leading to

$$p_i = \frac{1}{Z} e^{-\beta E_i}. \tag{49}$$

This means that we again recover the Boltzmann distribution, Eq. 5, using a different line of argument. Whereas the previous argument combined the arguments concerning conserved physical quantities and independence of subsystems, the current argument simple looked for the macroscopic state where our ignorance about the microscopic state is maximal. The inverse temperature $\beta$ has now entered the scene as a Lagrange multiplier.

Inserting the Boltzmann distribution, Eq. 49, into the entropy, Eq. 45, one finds

$$S = -k_B \sum_i \frac{1}{Z} e^{-\beta E_i} (-\ln Z - \beta E_i) = k_B \ln Z + \frac{1}{T} \langle H \rangle . \tag{50}$$

Solving this relation for $-k_B T \ln Z$ leads to

$$F \equiv -k_B T \ln Z = E - TS. \tag{51}$$

From the partition function $Z$ follows thus immediately the difference between $E$, the *internal energy* of the system, and the entropy. The quantity $F$ is called *free energy*. Since in equilibrium the quantity $S - \beta E$ is maximized, cf. Eq. 47, the free energy has to be minimized to find the most probable macrostate characterized by the temperature, volume and number of particles. $F$ is thus a function of these quantities, i.e., $F = F(T, V, N)$. The free energy is an example of a so-called *thermodynamic potential*, a function from which one can find the equilibrium state of the system via minimization. Knowing $F$ allows to directly determine average quantities via differentiation, e.g. by combining Eq. 22 and 51 we find

$$p = -\frac{\partial F}{\partial V}. \tag{52}$$

As an example let us again consider the ideal gas. The free energy follows from Eq. 12:

$$F = -k_B T \ln\left(\frac{1}{N!}\left(\frac{V}{\lambda_T^3}\right)^N\right) \approx k_B T N\left(\ln\left(\frac{\lambda_T^3 N}{V}\right) - 1\right). \tag{53}$$

On the rhs we used Stirling's formula, Eq. 39, and then neglected the term $(k_B T/2)\ln(2\pi N)$ that is much smaller than the other terms. The pressure follows by differentiation of Eq. 53 with respect to $V$, see Eq. 52, leading again to $p = k_B T N/V$.

The method of Lagrange multipliers can also be used to derive the grandcanonical ensemble. Maximizing the entropy with two constraints, $E = \langle H\rangle$ and $N = \langle N\rangle$, can be done in an analogous way to the canonical case, Eq. 48, and leads to the condition

$$\nabla\left(\frac{S}{k_B} - \beta\langle H\rangle + \alpha\langle N\rangle\right) = 0. \tag{54}$$

The requirement is thus

$$\frac{\partial}{\partial p_i}\left(\frac{S}{k_B} - \beta\langle H\rangle + \alpha\langle N\rangle\right) = -\ln p_i - 1 - \beta E_i + \alpha N_i = 0. \tag{55}$$

This leads directly to the Boltzmann factor for the grandcanonical case, Eq. 29. Inserting this distribution into the entropy, Eq. 45, we obtain

$$S = \frac{k_B}{Z_G}\sum_i e^{-\beta E_i + \alpha N_i}\left(\ln Z_G + \beta E_i - \alpha N_i\right) = k_B \ln Z_G + \frac{1}{T}\langle H\rangle - k_B\alpha\langle N\rangle. \tag{56}$$

Solving this relation for $-k_B T \ln Z_G$ leads to

$$K = K(T, \mu, V) = -k_B T \ln Z_G = E - TS - \mu N \tag{57}$$

where we introduced the quantity $\mu = \alpha/\beta$, called the *chemical potential*. From $Z_G$ follows thus the difference between the internal energy $E$ and $TS - \mu N$.

The thermodynamic potential $K$ is called the *grandcanonical potential* or *Gibbs potential*.

Surprisingly the grandcanonical potential is directly related to the pressure of the system:

$$K = -pV. \tag{58}$$

To see this we start from the fact that $E$, $S$, $V$ and $N$ are so-called *extensive* quantities, i.e., quantities that are additive. For instance, let us look again at the gas-filled balloon, Fig. 1: The volume of the whole system $\Sigma$ is simply the sum of the volumes of the subsystems $\Sigma_1$ and $\Sigma_2$ and so are the energies, particle numbers and entropies. On the other hand, the temperature $T$, the pressure $p$ and the chemical potential $\mu$ are *intensive* quantities. For systems in equilibrium such quantities have the same value in the full system and in all its subsystems. Products of an intensive and an extensive quantity like $TS$ are thus also extensive. From this follows that the Gibbs potential $K$ is an extensive quantity since all of its terms, $E$, $-TS$ and $\mu N$, are extensive. This means that $K$ fulfills the relation

$$K\left(T, \mu, \lambda V\right) = \lambda K\left(T, \mu, V\right) \tag{59}$$

for any value of $\lambda > 0$. If we choose e.g. $\lambda = 1/2$, then the left hand side (lhs) of Eq. 59 gives the Gibbs potential of a subsystem with half the volume of the full system. Its Gibbs potential is half of that of the full system (rhs of Eq. 59). We now take the derivative with respect to $\lambda$ on both sides of Eq. 59 and then set $\lambda = 1$. This leads to

$$\frac{\partial K}{\partial V} V = K. \tag{60}$$

Now in complete analogy to the derivation of the relation for the free energy in Eq. 52 one can show that $p = -\partial K/\partial V$ and hence

$$pV = -K = k_B T \ln Z_G. \tag{61}$$

We can thus immediately obtain the pressure from $Z_G$. For instance, for the ideal gas we calculated $Z_G$ in Eq. 33 from which follows

$$pV = k_B T \frac{zV}{\lambda_T^3} = N k_B T \tag{62}$$

where we used Eq. 34. We thus rederived the ideal gas equation of state, Eq. 27.

Finally, let us take one more close look at the example discussed earlier, the gas in a cylinder, Fig. 3. We were a little bit sloppy since we said in the legend of that figure that "the piston is under an externally imposed force $f$", but then calculated instead the expectation value of the force for a given volume, cf. Eqs. 20 to 23. If we want to be formally correct, then we need to maximize the entropy under the two contraints $\langle V \rangle = V$ and $\langle H \rangle = E$. This is achieved by solving the following set of conditions

$$\frac{\partial}{\partial p_i}\left(\frac{S}{k_B} - \beta \left\langle E \right\rangle - \gamma \left\langle V \right\rangle\right) = -\ln p_i - 1 - \beta E_i - \gamma V_i = 0 \tag{63}$$

where we introduced the additional Lagrange multiplier $\gamma$. Along similar lines that led us to the grandcanonical potential, Eq. 57, we find a thermodynamic potential $G = E - TS + (\gamma/\beta) V$. The ratio of the two Lagrange parameters in front of $V$ is just the pressure, $p = \gamma/\beta$, as we shall see in a moment. The new thermodynamic potential

$$G(T, p, N) = F(T, V(T, p, N), N) + pV(T, p, N) \tag{64}$$

is called the *free enthalpy* $G$. We can immediately check

$$\frac{\partial G}{\partial p} = \frac{\partial F}{\partial V}\frac{\partial V}{\partial p} + V + p\frac{\partial V}{\partial p} = V \tag{65}$$

where we used Eq. 52.

For an ideal gas we find from its free energy, Eqs. 53 and $V(T, p, N) = Nk_BT/p$ (i.e., Eq. 23 solved for $V$):

$$G(T, p, N) = k_BTN \ln\left(\frac{\lambda_T^3 p}{k_BT}\right). \tag{66}$$

Inserting this into Eq. 65 one recovers indeed the ideal gas law, Eq. 27, but this time in the version $\langle V \rangle = k_BTN/p$.

The grandcanonical potential obeys a very simple relation, $K = -pV$ (cf. Eq. 61), and so does the free enthalpy. Using the same line of argument that led to Eq. 61 we find

$$G = \frac{\partial G}{\partial N} N = \mu N. \tag{67}$$

That $\partial G/\partial N$ is the chemical potential $\mu$ follows by comparing Eqs. 57 and 58 to Eq. 64.

Only in a few exceptional cases one can calculate $Z$, Eq. 7, or $Z_G$, Eq. 30, exactly, e.g. for the ideal gas. Such systems are often somewhat trivial and do not even show phase transitions as we know them for any real substance. Phase transitions can only come about if the molecules interact with each other but then $Z$ cannot be calculated anymore. Nevertheless, in many cases a real system is close to an exactly solvable case. The idea of approximate methods is usually to describe the deviations by a small parameter $\varepsilon$ and then to expand $Z$ in powers of $\varepsilon$ around the exacly solvable case:

$$Z = Z_{\text{exact}} + C_1\varepsilon + C_2\varepsilon^2 + ...$$

Here a few examples:

1. perturbation theory: $H = H_0 + \lambda W$, expansion in $\lambda$
2. quasiclassical approximation: expansion in $\hbar$
3. high temperature expansion in $T_0/T$
4. low temperature expansion in $T/T_0$

5. expansion around critical point: $\tau = \frac{T-T_c}{T_c}$
6. $1/N$-expansion with $N$ number of components of a suitable field
7. $\varepsilon$-expansion: $D = 4 - \varepsilon$ with $D$ space dimension
8. virial expansion in $n = N/V$

In addition there is meanfield theory that cannot be cast easily in the above scheme. The idea of meanfield theory is to replace the interaction of a particle with all the other particles by the interaction of that particle with a suitable meanfield.

In the following chapter we will consider the virial expansion and apply it to the gas-liquid transition. In Chapter 3 we use high- and low temperature expansions to prove the existence of a phase with spontaneous magnetization for the 2-dimensional Ising model (ferromagentism). In the Chapter 4 we use meanfield theory to get a very simple view on that phase transition. We also use this framework to study a system where the virial expansion fails, namely salt solutions.

# 2 Virial expansion

## 2.1 Virial expansion up to second order

The virial expansion is an expansion in the density $n = N/V$. It should be thus a good approximation for sufficiently dilute systems for the case that the particles interact with short-range interaction (as we shall see later - in Section 3.3 - it does not work for a salt solution where the ions experience a long-range $1/r$ interaction). We first study the expansion up to second order which is relatively straighforward. Then we also study the much more complex case of the virial expansion to arbitrary order.

The Hamiltonian of a real gas is of the following form

$$H(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{N} \frac{p_i^2}{2m} + \sum_{i<j} w(|\mathbf{q}_i - \mathbf{q}_j|). \tag{68}$$

The first term represents the kinetic energy, the same as for the ideal gas, Eq. 9. The second term accounts for the interactions between the particles. The sum goes over all pairs of particles ("$i < j$" makes sure that each pair is only counted once) and we assume that the interaction potential $w$ depends only on the distances between the particles. It is now most convenient to use the grand canonical ensemble for which the partition function is of the form

$$Z_G = 1 + \sum_{N=1}^{\infty} z^N Z_N = 1 + \sum_{N=1}^{\infty} \frac{1}{N!} \left(\frac{z}{\lambda_T^3}\right)^N I_N. \tag{69}$$

The first step is just Eq. 31 where we wrote the $N = 0$ term separately. In the second step we inserted the explicit form of $Z_N$, Eq. 7, with $H(\mathbf{p}, \mathbf{q})$ given

by Eq. 68, and performed immediately the integration over the momenta. $I_N$ denotes thus the remaining integral

$$I_N = \int e^{-\beta \sum_{i<j} w(|\mathbf{q}_i - \mathbf{q}_j|)} d^3q_1 ... d^3q_N. \tag{70}$$

Let us first consider again the ideal gas. In this case $I_N = V^N$ and thus $Z_G = e^{zV/\lambda_T^3}$, Eq. 33. From that result we derived above, in Eq. 34, that $N = zV/\lambda_T^3$. In other words the quantity $z/\lambda_T^3$ that appears in Eq. 69 is in the case of the ideal gas precisely its density $n = N/V$. Now consider a real gas. If this gas is sufficiently dilute, then the interaction between its particles constitute only a small effect. The ratio $z/\lambda_T^3$ is then very close to its density. Since we assumed here the density to be small, the quantity $z/\lambda_T^3$ is small as well. We can thus interprete 69 as a series expansion in that small parameter. From this expansion we can learn how the interaction between the particles influences the macroscopic behavior of the system – at least in the regime of sufficiently dilute gas. In that regime it is then often sufficient to only account for the first or the first two correction terms since the higher order terms are negligibly small.

Unfortunately the quantity $z/\lambda_T^3$ has not such a clear physical meaning than the density $n$. But since both parameters are similar and small we can rewrite Eq. 69 to obtain a series expansion in $n$ instead of $z/\lambda_T^3$. This can be done in a few steps that we outline here for simplicity only to second order in $\zeta = z/\lambda_T^3$. We start from

$$Z_G = 1 + \zeta I_1 + \frac{\zeta^2}{2} I_2 + ... \tag{71}$$

To obtain the density $n = N/V$ we need to calculate the expectation value of $N$ that follows from $\ln Z_G$ via Eq. 32. We thus need next to find the expansion of $\ln Z_G$ starting from the expansion of $Z_G$. This is achieved by inserting $Z_G$ from Eq. 71 into $pV = k_B T \ln Z_G$, Eq. 61. To obtain again a series expansion in $\zeta$ we use the the series expansion of the logarithmus around $x = 1$, $\ln(1+x) = \sum_{k=1}^{\infty} (-1)^{k+1} x^k/k$. This leads to

$$\begin{aligned} \beta pV &= \ln Z_G = \ln\left(1 + \zeta I_1 + \frac{\zeta^2}{2} I_2 + ...\right) \\ &= \zeta I_1 + \frac{\zeta^2}{2} I_2 - \frac{\zeta^2 I_1^2}{2} + ... = \zeta I_1 + \frac{\zeta^2}{2}\left(I_2 - I_1^2\right) + ... \end{aligned} \tag{72}$$

When going from the first to the second line in Eq. 72 we neglected all terms higher than $\zeta^2$. The particle number follows by taking the derivative of $\ln Z_G$ with respect to $\alpha$, Eq. 32. Since $\zeta = e^{\alpha}/\lambda_T^3$ one has $\partial\zeta/\partial\alpha = \zeta$ and thus

$$N = \frac{\partial}{\partial\alpha} \ln Z_G = \zeta I_1 + \zeta^2\left(I_2 - I_1^2\right) + ... \tag{73}$$

We are now in the position to write an expansion in the density $n = N/V$ (instead of in $\zeta$) by substracting Eq. 73 from 72. This lead to

$$\beta p = \frac{N}{V} - \frac{\zeta^2}{2V}\left(I_2 - I_1^2\right) + ... \tag{74}$$

With this step we got rid of terms linear in $\zeta$ but there is still a $\zeta^2$-term. This term can now be easily replaced by using Eq. 73 that states that $N = \zeta I_1$ up to terms of the order $\zeta^2$. We can thus replace the $\zeta^2$-term in Eq. 74 by $(N/I_1)^2$ neglecting terms of the order $\zeta^3$. We arrive then at

$$\beta p = n - \left(\frac{N}{I_1}\right)^2 \frac{1}{2V}\left(I_2 - I_1^2\right) + ... \tag{75}$$

To see that Eq. 75 is indeed an expansion in $n$, we need to evaluate the integrals, $I_1$ and $I_2$, defined in Eq. 70. We find

$$I_1 = \int_V d^3q = V \tag{76}$$

and

$$\begin{aligned} I_2 &= \int\limits_V d^3q_1 d^3q_2 e^{-\beta w(|\mathbf{q}_1 - \mathbf{q}_2|)} = \int\limits_V d^3q_1 \int\limits_{"V-\mathbf{q}_1"} d^3\mathbf{r} e^{-\beta w(r)} \\ &\approx V \int_V d^3r e^{-\beta w(r).} \end{aligned} \tag{77}$$

The first step in Eq. 77 is simple the definition of $I_2$, Eq. 70. In the second step we substitute $\mathbf{q}_2$ by $\mathbf{r} = \mathbf{q}_2 - \mathbf{q}_1$, the distance vector between the two particles. The integration goes over all values of $\mathbf{r}$ such that $\mathbf{q}_2 = \mathbf{q}_1 + \mathbf{r}$ lies within the volume that we symbolically indicate by the shifted volume "$V - \mathbf{q}_1$". The last step where we replaced the shifted volume by the unshifted one involves an approximation. This can be done since the interaction between the particles, $w(r)$, decays to practically zero over microscopic small distances. Thus only a negligibly small fraction of configurations, namely where particle 1 has a distance to the wall below that microscopic small distance, is not properly accounted for.

Now we can finally write down the virial expansion to second order. Plugging the explicit forms of the integrals, Eqs. 76 and 77, into Eq. 75 we arrive at

$$\begin{aligned} \beta p &= n - \frac{n^2}{2}\int_V d^3r \left(e^{-\beta w(r)} - 1\right) + ... \approx n - \frac{n^2}{2}\int d^3r \left(e^{-\beta w(r)} - 1\right) + ... \\ &= n + n^2 B_2(T) + ... \end{aligned} \tag{78}$$

In the second step we replaced the integration over $V$ by an integration over the infinite space. This is again an excellent approximation for short-ranged $w(r)$ since $e^{-\beta w(r)} - 1$ vanishes for large $r$. The quantity $B_2(T)$ depends on the temperature via $\beta$ and is called the second *virial coefficient*. Introducing spherical coordinates $(r, \theta, \varphi)$ with $r_1 = r\sin\theta\cos\varphi$, $r_2 = r\sin\theta\sin\varphi$, and $r_3 = r\cos\theta$ we can write $B_2(T)$ as

$$\begin{aligned} B_2(T) &= -\frac{1}{2}\int\limits_0^{2\pi} d\varphi \int\limits_{-1}^{1} d\cos\theta \int\limits_0^{\infty} dr\, r^2 \left(e^{-\beta w(r)} - 1\right) \\ &= -2\pi \int\limits_0^{\infty} dr\, r^2 \left(e^{-\beta w(r)} - 1\right). \end{aligned} \tag{79}$$

## 2.2 Virial expansion to arbitrary order

We calculated the equation of state by reducing it to an effective two-particle system. We show here that the $l$th order in $n$, $n^l$, follows by accounting for interactions of complexes ("clusters") of $l$ interacting particles. We start again from

$$Z_G = \sum_{N=0}^{\infty} \frac{\zeta^N}{N!} I_N(T,V) \tag{80}$$

with $I_N$ in the classical case

$$I_N = \int e^{-\beta \sum_{i<j} w(\mathbf{x}_i - \mathbf{x}_j)} d^3x_1 ... d^3x_N \tag{81}$$

or in the quantummechanical case something like

$$I_N = \mathrm{Tr}\, e^{-\beta \mathrm{H}_N} = \int \langle \mathbf{x}_{1,...,}\mathbf{x}_N | \, e^{-\beta \mathrm{H}_N} \, | \mathbf{x}_{1,...,}\mathbf{x}_N \rangle \, d^3x_1 ... d^3x_N. \tag{82}$$

In general $I_N$ will be of the form

$$I_N = \int W_N(\mathbf{x}_1, ..., \mathbf{x}_N) \, d^3x_1 ... d^3x_N. \tag{83}$$

The quantity of interest is, however, not $Z_G$ but $\ln Z_G$. When one knows $Z_G$ exactly, this is a trivial step but here this turns out to be very challenging. $\ln Z_G$ can also be expanded in $\zeta$. It turns out to be of the form

$$\begin{aligned} \ln Z_G &= \sum_{r=1}^{\infty} \frac{1}{r!} \zeta^r J_r(T,V) \\ &= \sum_{r=1}^{\infty} \frac{1}{r!} \zeta^r \int U_r(\mathbf{x}_1, ..., \mathbf{x}_r) \, d^3x_1 ... d^3x_r. \end{aligned} \tag{84}$$

How can the functions $U_r$ be calculated? For small indices this can be done by hand by inserting the expansions, Eqs. 80 and 84 into the identity $Z_G = e^{\ln Z_G}$ and then by comparing the coefficients. This leads to

$$W_1(\mathbf{x}_1) = U_1(\mathbf{x}_1) = 1,$$

$$W_2(\mathbf{x}_1, \mathbf{x}_2) = U_2(\mathbf{x}_1, \mathbf{x}_2) + U_1(\mathbf{x}_1) \, U_1(\mathbf{x}_2),$$

$$\begin{aligned} W_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) &= U_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) + U_2(\mathbf{x}_1, \mathbf{x}_2) \, U_1(\mathbf{x}_3) + U_2(\mathbf{x}_2, \mathbf{x}_3) \, U_1(\mathbf{x}_1) \\ &+ U_2(\mathbf{x}_1, \mathbf{x}_3) \, U_1(\mathbf{x}_2) + U_1(\mathbf{x}_1) \, U_1(\mathbf{x}_2) \, U_1(\mathbf{x}_3). \end{aligned}$$

Inverting these relations allows to calculate the $U_r$'s from the $W_N$'s.

What we try to find now is a general relation between the $W_N$'s and the $U_r$'s. As a first step we need the following definition:

*Definition*: A partition $P$ of the set $\{1,2,...,N\}$ is the decomposition of the set in disjunct non-empty subsets $S_1,...,S_K$ $(1 \leq K \leq N)$ such that

$$\{1,2,...,N\} = \underset{S_k \in P}{\cup} S_k.$$

We call $\mathcal{P}^N$ the set of all such partitions, and $\mathcal{P}_K^N$ the set of all partitions made from $K$ subsets.

For a partition $P \in \mathcal{P}_K^N$ we denote by $r_i = |S_i|$ the cardinality (number of elements) of the set $S_i$ that belongs to $P$, and by $n_{r_i}$ the number of sets in $P$ of cardinality $r_i$. One can see immediately that there are

$$A_K\left(\{r_i\}\right) = \frac{N!}{r_1!....r_K!} \frac{1}{\prod n_{r_i!}} \tag{85}$$

different partions in $\mathcal{P}_K^N$ for a given set of $r_i$'s.

In our case it turns out that partitions of particles are important. Define for a given $P \in \mathcal{P}^N$ the function

$$U_{|S|}\left(S\right) = U_{|S|}\left(\mathbf{x}_{i_1},...,\mathbf{x}_{i_{|S|}}\right)$$

as the function for which the indices of the arguments $x_i$ are just the elements of $S$. One can then formulate the following important theorem:

*Theorem*: The function $W_N$ can be decomposed into its connected parts, the cluster functions $U_r$, as follows

$$W_N\left(\mathbf{x}_1,...,\mathbf{x}_N\right) = \sum_{P \in \mathcal{P}^N} \prod_{S \in P} U_{|S|}\left(S\right). \tag{86}$$

We can immediately check that the explicit cases $N = 1,2,3$ given above do indeed fulfill Eq. 86. The proof of this relation for general $N$ goes as follows. We start out from the equality $Z_G = e^{\ln Z_G}$:

$$Z_G = \sum_{N=0}^{\infty} \frac{\zeta^N}{N!} I_N = \sum_{k=0}^{\infty} \frac{1}{k!} \left( \sum_{r=1}^{\infty} \frac{1}{r!} \zeta^r J_r \right)^k.$$

Comparison of the cofficients for $\zeta^N$ with $N \geq 1$:

$$\frac{I_N}{N!} = \sum_{k=1}^{N} \frac{1}{k!} \sum_{\substack{r_1,...,r_k=1 \\ r_1+...+r_k=N}}^{\infty} \frac{1}{r_1!...r_k!} \prod_{i=1}^{k} J_{r_i}.$$

As a next step we order the $r_i$'s such that $r_1 \leq r_2 \leq ... \leq r_k$ and take this into account via the combinatorical factor $k!/\prod n_{r_i}!$:

$$\frac{I_N}{N!} = \sum_{k=1}^{N} \frac{1}{k!} \sum_{\{r_i\} \text{ ordered}} \frac{k!}{\prod n_{r_i}!} \frac{1}{r_1!...r_k!} \prod_{i=1}^{k} J_{r_i}.$$

Here the summation over "$\{r_i\}$ ordered" is a short-hand notation for the following summation: $\{\{r_i\}\,|\sum r_i = N,\, r_1 \leq r_2 \leq ... \leq r_k\}$. This can be written more compact by using $A_k(\{r_i\})$ from Eq. 85:

$$I_N = \sum_{k=1}^{N} \sum_{\{r_i\}\,\text{ordered}} A_k(\{r_i\}) \prod_{i=1}^{k} J_{r_i}.$$

Putting in the explicit forms of the $J_r$'s we arrive at

$$I_N = \sum_{k=1}^{N} \sum_{\{r_i\}\,\text{ordered}} A_k(\{r_i\}) \int U_{r_1}(S_1) ... U_{r_k}(S_1)\, d^3x_1 ... d^3x_N$$

which is simply

$$I_N = \sum_{k=1}^{N} \sum_{P \in \mathcal{P}_k^N} \int U_{r_1}(S_1) ... U_{r_k}(S_1)\, d^3x_1 ... d^3x_N.$$

From this follows finally

$$I_N = \int \sum_{P \in \mathcal{P}^N} \prod_{S \in P} U_{|S|}(S)\, d^3x_1 ... d^3x_N$$

from which follows directly Eq. 86.

What is the physical meaning of the connected parts? To a partition $P \in \mathcal{P}^N$ is associated a partitioning of the arguments $x_1, ..., x_N$ in disjunct subsets $S_i$ with $i = 1, ..., k$. We define the clusterlimit to a partition $P$ as the limit $|\mathbf{x}_a - \mathbf{x}_b| \to \infty$ for all $\mathbf{x}_a$, $\mathbf{x}_b$ with $\mathbf{x}_a \in S_i$ and $\mathbf{x}_b \in S_j$ with $S_i \cap S_j = \varnothing$ and $\mathbf{x}_a - \mathbf{x}_b$ fixed if $\mathbf{x}_a$ and $\mathbf{x}_b$ from the same subset. The functions $W_N(\mathbf{x}_1, ..., \mathbf{x}_N)$ have obviously the following cluster property: In the clusterlimit to a partition $P$, $W_N$ factorizes (asymptotic factorization):

$$W_N(\mathbf{x}_1, ..., \mathbf{x}_N) \to \prod_{S \in P} W_{|S|}(S) \tag{87}$$

if $w(r) \to 0$ for $r \to \infty$ fast enough. Then e.g.

$$\lim_{|\mathbf{x}_1 - \mathbf{x}_2| \to \infty} W_2(\mathbf{x}_1, \mathbf{x}_2) = W_1(\mathbf{x}_1)\, W_1(\mathbf{x}_2)$$

and hence (see above)

$$\lim_{|\mathbf{x}_1 - \mathbf{x}_2| \to \infty} U_2(\mathbf{x}_1, \mathbf{x}_2) = 0.$$

The latter property of $U_2$ can be generalized as follows:

All functions $U_r(\mathbf{x}_1, ..., \mathbf{x}_r)$ go to zero for each clusterlimit from $\mathcal{P}_k^r$ for any $k > 1$ if $W_N(\mathbf{x}_1, ..., \mathbf{x}_N)$ has the cluster property.

This can be proven by induction as follows. Suppose we have shown this for all $r \leq N-1$. Now consider a clusterlimit $P$ from $\mathcal{P}_k^N$ for any $k > 1$. Then $W_N$ takes asymptotically the form of Eq. 87. Using Eq. 86 we can write

$$W_N(\mathbf{x}_1, ..., \mathbf{x}_N) \to \prod_{S \in P} \sum_{\widetilde{P} \in \mathcal{P}^{|S|}} \prod_{\widetilde{S} \in \widetilde{P}} U_{|\widetilde{S}|}\left(\widetilde{S}\right) \tag{88}$$

On the other hand we can use directly Eq. 86 and then write the term containing $U_N$ separately. This leads to

$$W_N(\mathbf{x}_1, ..., \mathbf{x}_N) = U_N(\mathbf{x}_1, ..., \mathbf{x}_N) + \sum_{P \in \mathcal{P}_k^N, k>1} \prod_{S \in P} U_{|S|}(S) \tag{89}$$

Now we take again the cluster limit of this expression. We do not know yet anything about the first term, $U_N$, but we know that in the summation of the second term only those terms survive where there is no $U_{|S|}$ in the product that contains arguments from more than one cluster. This is exactly the expression above, Eq. 88. The only difference between Eq. 88 and the clusterlimit of Eq. 89 is the term $U_N$ which must thus be zero.

The property of $U_r$ that it vanishes for any non-trivial clusterlimit means that the quantity

$$\lim_{V \to \infty} \frac{1}{V} \int_V d^3x_1 ... d^3x_r U_r(\mathbf{x}_1, ..., \mathbf{x}_r)$$

exists since only conformations contribute to the integral where all particle are together in a cluster. This is in general not the case for the $W_N$'s.

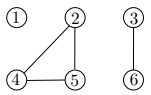
The calculation of the cluster functions can be simplified in the classical case:

$$f_{ij} = f_{ji} = e^{-\beta w(|\mathbf{x}_i - \mathbf{x}_j|)} - 1 \tag{90}$$

Then $W_N$ can be rewritten as

$$W_N(\mathbf{x}_1, ..., \mathbf{x}_N) = e^{-\beta \sum_{i<j} w(|\mathbf{x}_i - \mathbf{x}_j|)} = \prod_{i<j} (1 + f_{ij}) \tag{91}$$

On the rhs there are $2^{\binom{N}{2}}$ terms, each a product of $f_{ij}$'s and 1's. It is now convenient to represent each term by a numbered graph. As an example we give a graph of one of the terms occuring in Eq. 91 for $N = 6$:



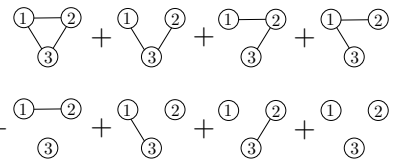

$$= \quad (f_{24} f_{45} f_{25})\, f_{36}$$

For $N = 1, 2, 3$ we find:

$W_1(\mathbf{x}_1)$ = ①

$W_2(\mathbf{x}_1, \mathbf{x}_2)$ = ①—② + ① ②

$W_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ =



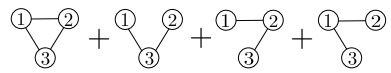
A graph is called connected if any two of its vertices are connected (direct or indirect) by an edge, otherwise not connected. The function of an unconnected graph is the product of its connected subgraphs.

We now formulate the following theorem: The function $U_r(\mathbf{x}_1, ..., \mathbf{x}_r)$ is given by the sum of all connected, numbered graphs with $r$ vertices.

Before we proove this theorem we give two examples:

$U_2(\mathbf{x}_1, \mathbf{x}_2)$ = ①—②

$U_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ =



From these examples we can immediately see that $U_r$ vanishes in every non-trivial clusterlimit.

Now we come to the proof of the theorem. The cases $r = 1, 2$ are clear. Suppose we have proven the theorem up to $r = N - 1$. We know that $W_N$ is of the form

$$W_N = \mathrm{Con}(\{1, 2, ..., N\}) + \sum_{k>1} \sum_{P \in \mathcal{P}_k^N} \mathrm{Con}(S_1) ... \mathrm{Con}(S_k)$$

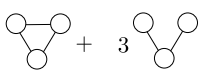
where $\mathrm{Con}(S)$ is the sum of all connected graphs to the set $S$. Using Eq. 86 we know that

$$W_N = U_N + \sum_{k>1} \sum_{P \in \mathcal{P}_k^N} U_{|S_1|}(S_1) ... U_{|S_k|}(S_k)$$

Thus $U_N = \mathrm{Con}(\{1, 2, ..., N\})$ as stated in the theorem above.

For the integrals $J_r(T, V) = \int_V d^3x_1 ... d^3x_r U_r(\mathbf{x}_1, ..., \mathbf{x}_r)$ each numbered graph that just has a different numbering gives the same contribution. The integrals can thus be represented by unnumbered graphs (Cluster integrals). To give examples:

$J_2(T, V) =$ ○—○ $= \int_V d^3x_1 d^3x_2 \underbrace{f(\mathbf{x}_1 - \mathbf{x}_2)}_{f(\mathbf{x}_1 - \mathbf{x}_2) = e^{-\beta w(|\mathbf{x}_1 - \mathbf{x}_2|)} - 1} = V \int_V d^3x f(\mathbf{x})$

$J_3(T, V) =$ 

with

$$= \int_V d^3x_1 d^3x_2 d^3x_3 f(\mathbf{x}_1 - \mathbf{x}_2) f(\mathbf{x}_2 - \mathbf{x}_3) f(\mathbf{x}_1 - \mathbf{x}_3)$$
$$= V \int_V d^3x d^3y f(\mathbf{x}) f(\mathbf{y}) f(\mathbf{x} + \mathbf{y})$$

$$= V \int_V d^3x d^3y f(\mathbf{x}) f(\mathbf{y}) = V \left\{ \int_V d^3x f(\mathbf{x}) \right\}^2$$

Clusterintegrals factorize if the graph can be disconnected by removal of a vertex as in the latter example.

Let us define

$$b_l(T,V) = \frac{1}{l!V} \int_V d^3x_1 ... d^3x_l U_l(\mathbf{x}_1, ..., \mathbf{x}_l) = \frac{J_l(T,V)}{l!V} \tag{92}$$

We expect that for $b_l(T,V)$ exists the limit

$$\lim_{V \to \infty} b_l(T,V) = b_l(T)$$

Then

$$\frac{1}{V} \ln Z_G = \frac{p}{k_B T} = \sum_{l=1}^{\infty} \zeta^l b_l(T,V) \tag{93}$$

and

$$\frac{N}{V} = n = +\frac{1}{V} \frac{\partial \ln Z_G}{\partial \alpha} = \sum_{l=1}^{\infty} l \zeta^l b_l(T,V) \tag{94}$$

Furthermore, the energy density follows from

$$\frac{E}{V} = -\frac{1}{V} \frac{\partial \ln Z_G}{\partial \beta} = -\frac{1}{V} \frac{\partial T}{\partial \beta} \frac{\partial \ln Z_G}{\partial T} = \frac{kT^2}{V} \frac{\partial \ln Z_G}{\partial T}$$

Using Eqs. 69 and 93 one finds

$$\frac{E}{V} = kT^2 \left( \sum_{l=1}^{\infty} \frac{3}{2} \frac{l}{T} \zeta^l b_l(T,V) + \sum_{l=1}^{\infty} \frac{\partial b_l(T,V)}{\partial T} \zeta^l \right)$$

With Eq. 94 this leads to

$$\frac{E}{V} = \frac{3}{2} n k_B T + kT^2 \sum_{l=1}^{\infty} b_l'(T,V) \zeta^l \tag{95}$$

with $b_l' = \partial b_l / \partial T$.

Equations 93 to 95 with the coefficients being the clusterintegrals (with their $U_r$ being the sums of connected graphs) are very elegant expressions. Unfortunately they are expansions in the activity $\zeta$ and not in the density $n$. To go to a general virial expansion

$$\frac{p}{k_B T} = \sum_{l=1}^{\infty} B_l(T)\, n^l \tag{96}$$

requires some extra work. The quantity $B_l(T)$ in Eq. 96 is called the $l$th virial coefficient.

To calculate the virial expansion up to $l$th order, one has to do the following steps (here given for the case $l = 3$). We start with the power series in $\zeta$ for $n$, Eq. 94, up to third order:

$$n = \zeta + 2b_2\zeta^2 + 3b_3\zeta^3 + O\left(\zeta^4\right) \tag{97}$$

where we used the fact that $b_1 = 1$ (cf. Eq. 92). Now we have to solve for $\zeta$ up to terms of third order in $n$. We use the ansatz:

$$\zeta = a_1 n + a_2 n^2 + a_3 n^3 + O\left(n^4\right) \tag{98}$$

We plug Eq. 98 into Eq. 97 and find

$$n = a_1 n + \left(a_2 + 2b_2 a_1^2\right) n^2 + \left(a_3 + 4b_2 a_1 a_2 + 3b_3 a_1^3\right) n^3 + O\left(n^4\right)$$

For this equation to hold we need $a_1 = 1$ and the coefficients in front of $n^2$ and $n^3$ to vanish. This means that $a_2 = -2b_2$ and $a_3 = 8b_2^2 - 3b_3$. Thus

$$\zeta = n - 2b_2 n^2 + \left(8b_2^2 - 3b_3\right) n^3 + O\left(n^4\right)$$

Plugging this into Eq. 93 leads finally to

$$\frac{p}{k_B T} = n - b_2 n^2 + \left(4b_2^2 - 2b_3\right) n^3 + O\left(n^4\right) \tag{99}$$

We give here the first 4 virial coefficients

$$B_1 = b_1 = 1$$

$$B_2 = -b_2$$

$$B_3 = 4b_2^2 - 2b_3$$

$$B_4 = -20b_2^3 + 18b_2 b_3 - 3b_4$$

$$\vdots$$

Finally, let us give the expansions for the energy $E$ and the free energy $F$. Using $pV = k_B T \ln Z_G$ and Eq. 32 we find

$$\frac{E}{N} = \frac{3}{2} k_B T - k_B T^2 \sum_{l=2}^{\infty} \frac{n^{l-1}}{l-1} B_l'(T)\,. \tag{100}$$

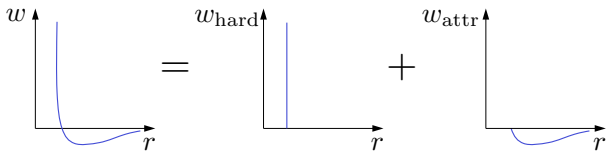



Figure 6: The typical interaction potential $w$ between two molecules as a function of their distance $r$. It typically is the sum of two contributions. The first one is a hardcore repulsive term, $w_{\text{hard}}$, that forbids particles to overlap (excluded volume). The second is a longer-ranged attractive potential, $w_{\text{attr}}$.

To derive $F$ remember that the pressure follows from $F$ by differentiation, $p = -\partial F/\partial V$ , Eq. 52. The free energy is thus obtained by integrating the pressure, Eq. 96, leading to

$$\beta F = N\left(\ln\left(\lambda_T^3 n\right) - 1\right) + V\sum_{l=2}^{\infty}\frac{n^l}{l-1}B_l\left(T\right). \tag{101}$$

The first term in Eq. 101 follows from integrating the $l = 1$ term that leads to $-N\ln V$. All the other contributions to the first term are just the integration constant that has to be chosen such that the result matches the ideal gas result, Eq. 53, for the case that all $B_l = 0$ for $l \geq 2$. You can convince yourself easily that one indeed obtains Eq. 96 from the virial expansion of $F$ by taking the derivative with respect to $V$, $p = -\partial F/\partial V$.

## 2.3 Van der Waals equation of state

The van der Waals equation of state is an ingeniously simple ad hoc approach that gives a qualitative idea of the equation of state of a real substance including its gas-liquid phase transition. It has been introduced by Johannes van der Waals in his thesis "Over de Continuïteit van den Gas- en Vloeistoftoestand" (Leiden University, 1873). Van der Waals assumed the existence of atoms (disputed at that time) and even more, that they have excluded volume and attract. Here we will discuss how his approach can be understood in a more systematic way as the first two (or three) terms in the virial expansion of a real gas.

As a start let us estimate from Eq. 79 the typical temperature dependence of the second virial coefficient $B_2\left(T\right)$. Figure 6 depicts the typical form of the interaction $w\left(r\right)$ between two molecules. For short distances $w\left(r\right)$ rises sharply, reflecting the fact that two molecules cannot overlap in space due to hardcore repulsion. For larger distances there is typically a weak attraction. As schematically indicated in the figure the total interaction potential can be written as the sum of these two contributions, $w\left(r\right) = w_{\text{hard}}\left(r\right) + w_{\text{attr}}\left(r\right)$. To a good approximation the hardcore term can be assumed to be infinite for $r \leq d$ and zero otherwise, where $d$ denotes the center-to-center distance of the touching particles, i.e., their diameter. The integral 79 can then be divided into

two terms accounting for the two contributions to the interaction:

$$\begin{aligned} B_2(T) &= -2\pi \int_0^d r^2 (-1)\, dr - 2\pi \int_d^\infty r^2 \left( e^{-\beta w_{\rm attr}(r)} - 1 \right) dr \\ &\approx 2\pi \frac{d^3}{3} + 2\pi \int_d^\infty r^2 \beta w_{\rm attr}(r)\, dr = v_0 - \frac{a}{k_B T}. \end{aligned} \tag{102}$$

The approximation involved by going to the second line is to replace $e^{-\beta w_{\rm attr}(r)}$ by $1 - \beta w_{\rm attr}(r)$ which is a good approximation if the attractive part is small compared to the thermal energy, i.e., $\beta w_{\rm hard}(r) \ll 1$ for all values of $r > d$.

In the final expression of Eq. 102 the volume $v_0 = 2\pi d^3/3$ accounts for the excluded volume of the particles. It is actually four times the volume $4\pi (d/2)^3/3$ of a particle. The factor 4 is the combination of two effects: $(i)$ A particle excludes for the other a volume $4\pi d^3/3$ that is eight times the eigenvolume. $(ii)$ An additional factor $1/2$ accounts for the implicit double counting of particle pairs by the $n^2$-term. The term $a = -2\pi \int_d^\infty r^2 w_{\rm attr}(r)\, dr$ is a positive quantity (assuming $w_{\rm attr}(r) \leq 0$ everywhere as is the case in Fig. 6). We thus find that with increasing temperature the attractive term becomes less and less important and the systems behaves more and more like a system with pure hardcore repulsion. There is a temperature $T^* = a/(k_B v_0)$ below which $B_2(T)$ becomes negative, i.e., the particles effectively start to attract each other.

We can now write the virial expansion up to second order in $n = 1/v$ ($v$ is the volume per particle). From Eq. 78

$$\frac{p}{k_B T} = \frac{1}{v} + B_2 \frac{1}{v^2} = \frac{1}{v}\left(1 + \frac{v_0}{v}\right) - \frac{a}{v^2 k_B T} \tag{103}$$

and from Eq. 100

$$\frac{E}{N} = \frac{3}{2} k_B T - k_B T^2 n B_2'(T) = \frac{3}{2} k_B T - \frac{a}{v} \tag{104}$$

Eq. 103 does not make sense for small values $v$ since then $p \to -\infty$ rather than the physically expecteed $p \to \infty$. This is not surprising since the virial expansion up to second order is not expected to hold in this regime. In fact, one can show that even the complete virial expansion, Eq. 96, will break down in that regime.

To come to at least a qualitative understanding of this system one can add various modifications to Eq. 103. The most natural thing would be to go in the virial expansion up to third order, i.e. adding an additional term of the form $B_3/v^3$ and simply to assume $B_3$ to be some positive constant. This is a procedure often done in the context of polymer physics (Flory, 1934). When thinking about a real gas, however, people typically tend to use the approximation introduced by van der Waals that constitute a rather arbitrary procedure. Replace in Eq. 103:

$$\frac{1}{v}\left(1 + \frac{v_0}{v}\right) \to \frac{1}{v - v_0}$$

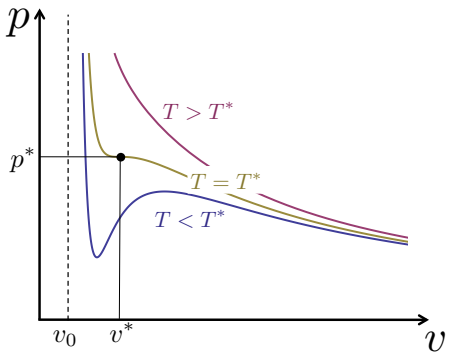



Figure 7: Isothermes of the van der Waals gas, Eq. 105.

Now this term goes to infinity at $v = v_0$ which makes sense because the pressure should become infinite once the molecules are densely packed. Even though the replacement seems completely arbitrary, note that for small $v_0/v$:

$$\frac{1}{v - v_0} = \frac{1}{v}\frac{1}{1 - v_0/v} = \frac{1}{v}\left(1 + \frac{v_0}{v} + O\left(\left(\frac{v_0}{v}\right)^2\right)\right)$$

Using this replacement we can write Eq. 103 as follows

$$\frac{p}{k_B T} = \frac{1}{v - v_0} - \frac{a}{v^2 k_B T}$$

which can be recast in the famous van der Waals equation

$$\left(p + \frac{a}{v^2}\right)(v - v_0) = k_B T \tag{105}$$

Originally this equation has been more phenomenologically introduced by modifying the ideal gas equation $pv = k_B T$ as follows: $v \to v - v_0$ as already discussed above and $p \to p + a/v^2$ which means that the pressure is effectively reduced due to the attraction between particles and that this reduction should be proportional to $n^2$.

Amazingly the isotherms of the van der Waals equation (or of the above mentioned virial expansion up to 3rd order) are qualitatively very similar to the isotherms that one measures for real substances - including even the liquid-gas phase transition, see Fig. 7. There is a so-called critical temperature $T^*$ such that for temperatures above $T^*$, $T > T^*$, the isothermes are similar to that of an ideal gas whereas for low temperatures, $T < T^*$, the isothermes feature a local minimum and maximum. The isotherm at $T = T^*$, the so-called *critical isotherm*, features an inflection point that follows from the conditions

$$\frac{\partial p}{\partial v}(T^*, v^*) = \frac{\partial^2 p}{\partial v^2}(T^*, v^*) = 0 \tag{106}$$

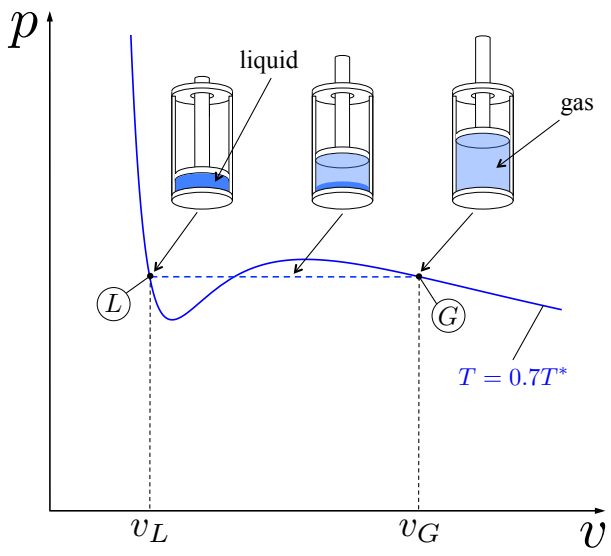



Figure 8: The Maxwell construction (see text). States where gas and liquid coexist (cylinder in the middle) lie on the line between points $G$ (pure gas, right cylinder) and $L$ (pure liquid, left cylinder).

From this follows

$$v^* = 3v_0,\ kT^* = \frac{8}{27}\frac{a}{v_0},\ p^* = \frac{1}{27}\frac{a}{v_0^2} \tag{107}$$

Note, however, that the isotherms for $T < T^*$ around $v^*$ are not physical since there the substance has a negative compressibility which makes the system instable. In fact, this hints at a first order phase transition between a liquid and a fluid. In the coexistence region (only present for $T < T^*$) the isotherm is supposed to be horizontal, i.e. one should have no increase in $p$ with decreasing $v$ but condensation of gas (low density state) into fluid (high density state). At $T = T^*$ the gas and the fluid have the same density and for $T > T^*$ there is no phase transition anymore (there is no distinction between gas and liquid anymore).

To make the unphysical isotherms of Eq. 84 that occur for $T < T^*$ physical, one needs to employ the *Maxwell construction*: One replaces a part of the isotherm by a horizontal line as shown in Fig. 8. The height of the horizonal line needs to be chosen such that the two areas that are enclosed between that line and the original isotherm are equal. We will give a justification of this in a moment, but for now let us discuss what happens to the system when it moves along the horizontal line.

Suppose we start at a very dilute system, i.e., at a large $v$-value. When we compress this system at constant temperature, then the pressure first rises. Once the volume $v_G$ per particle is reached something dramatic happens: the pressure remains constant under further compression, see Fig. 8. This signals the onset

of a phase transition. Whereas at point $G$ the cylinder is still completely filled with gas (right cylinder), as soon as we move along the Maxwell line there is also a second phase in the cylinder shown in darker blue in the middle cylinder. This is the liquid phase that has a higher density and it is thus found at the bottom of the cylinder. By compressing the volume further, more and more molecules in the gas phase enter into the liquid. Once the point $L$ is reached all the molecules have been transferred to the liquid phase, see the cylinder on the left. Upon further compression of the system the pressure rises sharply following again the original isotherm, reflecting now the compression of the liquid phase.

How is it possible that two phases coexist inside the cylinder? This is only possible if three conditions are fulfilled: (*i*) The temperatures in the two phases need to be the same since otherwise heat will flow from the hotter to the colder phase. This condition is fulfilled since both points, $L$ and $G$, lie on the same isotherm. (*ii*) The pressure in both phases needs to be the same since otherwise the phase with the higher pressure expands at the expense of the phase with the lower pressure. The horizontal line is by construction a line of constant pressure. (*iii*) Finally, the chemical potentials of the two phases need to be the same, i.e., the chemical potentials at points $L$ and $G$ in Fig. 8 have the same value:

$$\mu_G(T,p) = \mu_L(T,p). \tag{108}$$

Since this is the least intuitive condition we explain it here in more detail. We can think of each phase as a system under a given pressure $p$ at a given temperature $T$. The appropriate thermodynamic potential is thus the free enthalpy, Eq. 64. Using Eq. 67 the total free enthalpy of the two coexisting phases is given by

$$G = \mu_G N_G + \mu_L N_L = \mu_G N_G + \mu_L (N - N_G). \tag{109}$$

On the rhs we used the fact that the total number of particles, $N$, is the sum of the particles in the two phases, $N_G + N_L$. Suppose now that the two chemical potentials were different, e.g. $\mu_G > \mu_L$. In that case the free enthalpy can be lowered by transferring particles from the gas to the liquid phase. Equilibrium between the two phases, as shown inside the middle cylinder of Fig. 8, is thus only possible if the two chemical potentials are the same. Only then the free enthalpy is minimized: $\partial G/\partial N_G = \mu_G - \mu_L = 0$.

We now need to show that condition 108 is fulfilled when the equal area construction is obeyed. Combining Eqs. 64 and 67 we find for each phase the relation

$$\mu_k(T,p) = \frac{F_k + pV_k}{N_k} \tag{110}$$

with $k = G, L$. The coexistence condition, Eq. 108 together with the relation 110 then leads to the condition

$$f_L - f_G = p(v_G - v_L) \tag{111}$$

where $f_k$ denotes the free energy per molecule in the $k$th phase. Next we calculate the difference $f_L - f_G$ purely formally by integrating along the (unphysical)

isotherm:

$$f_L - f_G = \int\limits_{\text{isotherm } G\to L} df = -\int\limits_{v_G}^{v_L} p\,(T,v)\,dv. \tag{112}$$

In the second step we used the relation

$$df = f\,(T+dT, v+dv) - f\,(T,v) = \frac{\partial f}{\partial T}dT + \frac{\partial f}{\partial v}dv \overset{\text{isotherm}}{=} -pdv. \tag{113}$$

On the rhs we made use of the fact that per definition $dT \equiv 0$ along the isotherm and of the relation $\partial f/\partial v = \partial F/\partial V = -p$, Eq. 52. Combining Eqs. 111 and 112 we arrive at

$$\int\limits_{v_L}^{v_G} p\,(T,v)\,dv = p\,(v_G - v_L)\,. \tag{114}$$

This is just the mathematical formulation of the equal area requirement since only then the area under the isotherm between $v_L$ and $v_G$ equals the area of a rectangle of height $p$ and width $v_G - v_L$.

Next we point out that one can bring Eq. 105 in a universal form (i.e. make it independent of the specific values of $v_0$ and $a$) by making everything dimensionless. We introduce

$$\tilde{v} = \frac{v}{v^*},\ \tilde{p} = \frac{p}{p^*},\ \tilde{T} = \frac{T}{T^*},\ \tilde{e} = \frac{e}{e^*}$$

With Eq. 107 Eqs. 104 and 105 take the universal forms

$$\left(\tilde{p} + \frac{3}{\tilde{v}^2}\right)(3\tilde{v} - 1) = 8\tilde{T}, \quad \tilde{e} = 4\tilde{T} - \frac{3}{\tilde{v}} \tag{115}$$

Eq. 115 is qualitatively in good agreement with experiments but not quantitatively. For instance, Eq. 107 predicts

$$p^* v^* = \frac{3}{8}kT^* \tag{116}$$

whereas for real substances one finds typically $p^* v^* \approx 3.4kT^*$. Also the behavior close to the critical point is not correctly captured.

However, the existence of a universal equation of state

$$\tilde{p} = \tilde{p}\left(\tilde{T}, \tilde{v}\right), \quad \tilde{e} = \tilde{e}\left(\tilde{T}, \tilde{v}\right) \tag{117}$$

is quite well experimentally confirmed. Within classical statistical mechanics it can be understood as follows. Assume for all substances a qualitatively similar interaction profile $w\,(r)$ but of different depth $\varepsilon$ and range $\sigma$:

$$w\,(r) = \varepsilon\tilde{w}\left(\frac{r}{\sigma}\right)$$

This means for the partition function

$$\begin{aligned} Z_N &= \frac{1}{N!\lambda^{3N}} \int\limits_V d^{3N}x e^{-\beta\varepsilon \sum_{i<j} \tilde{w}\left(\frac{x}{\sigma}\right)} \\ &= \frac{\sigma^{3N}}{N!\lambda^{3N}} \int\limits_{\frac{1}{\sigma}V} d^{3N}x' e^{-(\beta\varepsilon) \sum_{i<j} \tilde{w}\left(x'\right)} \end{aligned} \tag{118}$$

From this we can directly see that $\sigma$ and $\varepsilon$ can be adsorbed in $V$ and $T$ which explains the univeral form of Eq. 117.

# 3 Low and high temperature expansion

In the following we will consider the Ising model to achieve some understanding of the phase transition in a ferromagnet. Here we consider the case of a vanishing external magnetic field. In the next chapter we shall also study the influence of a magnetic field within the meanfield approximation. In a ferromagnet one has an ordered phase with a non-vanishing magnetization below a critical temperature, the Curie temperature, and a disordered phase with zero magnetization above that temperature. The Ising model on a D dimensional lattice is a simple model for an uniaxial ferromagnet. On each lattice site $i$ site a spin that can assume the values $\sigma_i = \pm 1$. The Hamiltonian is given by

$$H\left(\{\sigma_i\}\right) = -h \sum_i \sigma_i - J \sum_{\mathrm{NN}} \sigma_i \sigma_j \tag{119}$$

Here $h = mB$ is the energy of the magnetic moment $m$ of the spin in the magentic field $B$, $J$ is the nearest neighbor coupling energy (spins prefer to align parallel) and $\sum_{\mathrm{NN}}$ means the summation over nearest neighbours. We will first study the one-dimensional case a warming-up exercise. There are many ways of calculating its partition function, here we use one more exotic one that sums over self-avoiding walks. Whereas the 1D case is trivial (no phase transition), the same method will be applied later on the 2D Ising model. This case is much more hard to deal with but it has the beautiful feature that it shows a phase transition to an ordered phase with non-zero magnetization.

## 3.1 The one-dimensional Ising model

We consider a one-dimensional lattice with the spins $i = 1, ..., N$. For simplicity we close the lattice into a ring, i.e., spin 1 and $N$ are considered to be near neighbors ($s_{N+1} = s_1$). We consider the case without external magentic field, $h = 0$. The partition function is then given by

$$
\begin{aligned}
Z &= \sum_{\{\sigma_i=\pm 1\}} e^{\beta J \sum_{i=1}^{N} \sigma_i \sigma_{i+1}} = \sum_{\{\sigma_i=\pm 1\}} \prod_{i=1}^{N} e^{\beta J \sigma_i \sigma_{i+1}} \\
&= \sum_{\{\sigma_i=\pm 1\}} \prod_{i=1}^{N} (\cosh \beta J + \sigma_i \sigma_{i+1} \sinh \beta J) \qquad (120)
\end{aligned}
$$

In the second line we used the identity

$$
e^{\beta J \sigma_i \sigma_{i+1}} = \cosh \beta J + \sigma_i \sigma_{i+1} \sinh \beta J = \begin{cases} e^{\beta J} & \text{for } \sigma_i \sigma_{i+1} = 1 \\ e^{-\beta J} & \text{for } \sigma_i \sigma_{i+1} = -1. \end{cases} \qquad (121)
$$

Equation 120 can be further rewritten as follows:

$$
\begin{aligned}
Z &= (\cosh \beta J)^N \sum_{\{\sigma_i=\pm 1\}} \prod_{i=1}^{N} (1 + \sigma_i \sigma_{i+1} \tanh \beta J) \\
&= (\cosh \beta J)^N \sum_{\{\sigma_i=\pm 1\}} \Bigg( 1 + \tanh \beta J \sum_i (\sigma_i \sigma_{i+1}) + \qquad (122) \\
&\qquad + (\tanh \beta J)^2 \sum_{i \neq j} (\sigma_i \sigma_{i+1}) (\sigma_j \sigma_{j+1}) + \ldots \Bigg)
\end{aligned}
$$

For the various terms of this expression we can introduce a graphical representation: Each pair $(\sigma_i \sigma_{i+1})$ is represented by a line that connects the lattice points $i$ and $i+1$. The first term in the expansion corresponds then to a graph of $N$ points without any connecting line. The second contains graphs in which exactly one pair of neighboring points is connected. The third point is made from graphs where 2 different pairs of points are connected by a line and so on. Finally, the last term consist of one graph where all lines are inscribed. Since

$$
\sum_{\sigma_i=\pm 1} \sigma_i^2 = 2 \quad \text{and} \quad \sum_{\sigma_i=\pm 1} \sigma_i = 0
$$

we see that only those terms contribute to the summation over $\{\sigma_i = \pm 1\}$ for which either $\sigma_i$ for each $i$ does not appear or where it appears quadratically. That means that only 2 terms contribute to the summation, namely the first term (no lines) and the last term (all lines present). Therefore

$$
Z = (\cosh \beta J)^N \sum_{\{\sigma_i=\pm 1\}} \left( 1 + (\tanh \beta J)^N \right) = (2 \cosh \beta J)^N \left( 1 + (\tanh \beta J)^N \right) \qquad (123)
$$

The righthand side can be interpreted (up to the factor $(2\cosh\beta J)^N$) as a summation over all closed walks on a lattice for which no line is used twice. Each walk has to be wighted by a factor $(\tanh\beta J)^l$ ($l$ : length of walk). There are two such self-avoiding walks: one of length 0 and one that goes all around the lattice. The latter contribution disappears in the thermodynamic limit $N \to \infty$.

The partition function is analytical for all temperatures. A finite temperature with zero magentization above and non-zero magnetization below a finite temperature can thus not exist and there is no phase transition.

## 3.2 The two-dimensional Ising model

First studied in Ising's PhD thesis in 1925, this model features a phase with spontaneous magnetization as proven by Peierls in 1936. Kramers and Wannier calculated in 1941 the exact expression for the temperature below which spontaneous magentization occurs (in the absence of a magnetic field). Onsager was the first to find the free energy of the 2D Ising model with algebraic methods (again in the absence of a magnetic field). We perform now high- and low temperature expansion of this model in the following subsection. In later subsection we will prove the existence of a phase transition and then determine the exact temperature where the phase transition occurs.

### 3.2.1 High- and low temperature expansions

We consider spins living on a two-dimensional quadratic lattice. As for the 1D Ising model we have the following identity for neighboring spins:

$$e^{\beta J\sigma_i\sigma_j} = \cosh\beta J + \sigma_i\sigma_j \sinh\beta J = \begin{cases} e^{\beta J} & \text{for } \sigma_i\sigma_j = 1 \\ e^{-\beta J} & \text{for } \sigma_i\sigma_j = -1. \end{cases} \tag{124}$$

leading to the partition function

$$Z = (\cosh\beta J)^{2N^2} \sum_{\{\sigma_i=\pm 1\}} \prod_{\langle i,j\rangle} (1 + \sigma_i\sigma_j \tanh\beta J)$$

$2N^2$ is number of lines on a quadratic $N \times N$ lattice when closing the lattice into a torus. As in the 1D case one can multiply all the factors in the product and represent them by graphs by connecting all spin pairs that occur in the corresponding term. Only such terms will contribute where a spin does not occur an odd number of times. This means that only such terms count where walks are closed. As no direction is associated to these self-avoiding walks, it is better to speak of polygons. Note that a term can correspond to a single or several closed polygons. The total length of those polygons, $l$, leads to a weight factor $(\tanh\beta J)^l$. The partition function is thus given by

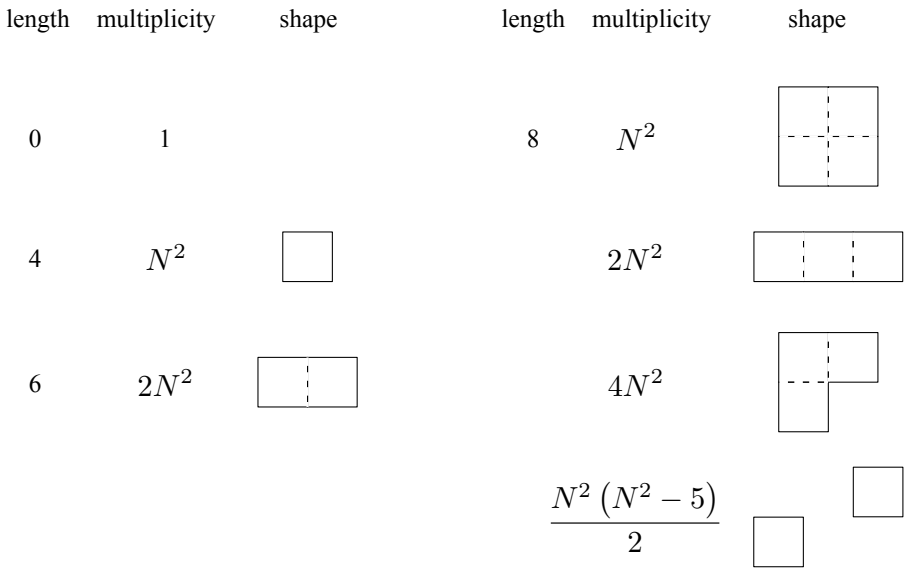



Figure 9: All closed polygons on a quadratic lattice up to length $l = 8$.

$$\begin{aligned} Z &= \left(\cosh\beta J\right)^{2N^2} 2^{N^2} \sum_{\text{Polygons}} \left(\tanh\beta J\right)^l \\ &= \left(2\cosh^2\beta J\right)^{N^2} \sum_l P\left(l\right)\left(\tanh\beta J\right)^l \end{aligned} \tag{125}$$

$P\left(l\right)$ is the number of closed polygons of length $l$. Note that this representation of the partition function of the Ising model is valid for any dimension of the lattice.

An expansion of $Z$ in powers of $\tau = \tanh\beta J$ is a high-temperature expansion (since $\tanh\beta J \ll 1$ for high temperatures). All the closed polygons of length $l \leq 8$ are depicted in Fig. 9 together with their multiplicity. From this follow the first terms of the partition function in the high temperature expansion

$$Z = \left(2\cosh^2\beta J\right)^{N^2}\left(1 + N^2\tau^4 + 2N^2\tau^6 + N^2\left(7 + \frac{1}{2}\left(N^2-5\right)\right)\tau^8 + ...\right) \tag{126}$$

and the corresponding free energy per spin

$$F_\infty = -\lim_{N\to\infty}\frac{1}{\beta N^2}\ln Z = -\frac{1}{\beta}\left(\ln\left(2\cosh^2\beta J\right) + \tau^4 + 2\tau^6 + \frac{9}{2}\tau^8 + ...\right) \tag{127}$$

where we used $\ln\left(1+x\right) \approx x - x^2/2$ for $x \ll 1$.

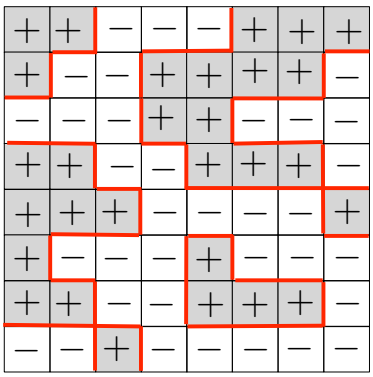


Figure 10: The dual lattice (see text for details).

As the next step we derive a low-temperature expansion ($\beta \to \infty$). As a first step we put a configuration-independent factor in front of the partition function:

$$Z = \sum_{\{\sigma_i = \pm 1\}} e^{\beta J \sum_{\langle i,j \rangle} (\sigma_i \sigma_j - 1 + 1)} = e^{2N^2 \beta J} \sum_{\{\sigma_i = \pm 1\}} e^{\beta J \sum_{\langle i,j \rangle} (\sigma_i \sigma_j - 1)} \tag{128}$$

Here $2N^2$ is the number of nearest neighbors on a quadratic lattice. The weight of a given configuration of spins is given by the number of lines that connect spins with opposite orientation. Each such line contributes a factor $e^{-2\beta J}$. It is then useful to go to the dual lattice where the roles of the plaquettes (the squares defined by 4 lines) and of the vertices are exchanged. The centers of the plaquettes of the original lattice are then the vertices of the dual lattice. The spins are now defined on the plaquettes of the dual lattice. We can now highlight the lines of the dual lattice bordering regions of spin $+1$ and spin $-1$, see Fig. 10. These border lines are dual to the lines that contribute a factor $e^{-2\beta}$ in the partition function. Each spin configuration corresponds to a configuration of border lines with the weight $e^{-2\beta l}$ where $l$ denotes the total length of the border between plus and minus spin regions.

Remarkably we encounter here again self-avoiding closed polygons as we did above in the high-temperature expansion. We therefore can write the partition function as follows:

$$Z = 2e^{2\beta J N^2} \sum_{\text{Polygons}} e^{-2\beta J l} \tag{129}$$

Again we find for the two-dimensional Ising model a representation as a summation over closed polygons but this time with a weight that vanishes for $\beta \to \infty$, a low-temperature expansion. The first few terms of the partition function are now given by

$$Z = 2e^{2\beta J N^2} \left( 1 + N^2 e^{-8\beta J} + 2N^2 e^{-12\beta J} + N^2 \left( 7 + \frac{1}{2} \left( N^2 - 5 \right) \right) e^{-16\beta J} + ... \right) \tag{130}$$

from which follow

$$-\beta F_\infty = \lim_{N\to\infty} \frac{1}{N^2} \ln Z = 2\beta J + e^{-8\beta J} + 2e^{-12\beta J} + \frac{9}{2} e^{-16\beta J} + ... \qquad (131)$$

The equilvalence between the high- and low-temperature expansion of the 2D Ising model is also called self-duality and will be useful later to determine the phase transition temperature. But first we need to proof that there is a phase transition at all.

### 3.2.2 Proof of the existence of a phase transition

The Ising model in two (and also in any higher) dimensions has a phase transition that separates a phase with a spontaneous magnetization from a phase without spontaneous magnetization. This can be described by an order parameter that is non-zero in the ordered phase and zero in the disordered phase. The most obvious candidate for such a quantity is the expectation value of the spin orientation. But the problem is that the Boltzmann factor is invariant under the transformation $\sigma_i \to -\sigma_i$ and thus the expectation value $\langle \sigma_i \rangle$ vanishes always.

To circumvent this problem we introduce a quantity that does not vanish in the ordered phase:

$$\mu = \lim_{r\to\infty} \lim_{N\to\infty} \langle \sigma_i \sigma_{i+r} \rangle \qquad (132)$$

This quantity determines the correlation between 2 spins that are infinitely far apart. In the case of spontaneous magnetization the probability to be parallel is larger than to be antiparallel and thus $\mu > 0$. In the disordered phase there are no correlation between spins that are infinitely far apart from each other and $\mu = 0$.

For $T = 0$ there are only 2 spin configurations that contribute, namely all spins have $+1$ or all spins have $-1$, both leading to $\mu = 1$. This, however, is also true for the one-dimensional Ising model, even though this system has no phase transition since the magetization vanishes for any $T \neq 0$. What needs to be shown here is that there is a finite temperature range above $T = 0$ where the magnetization is different from 0. The following proof is typical for an estimate of this kind in statistical mechanics. As a first step we rewrite $\mu$ as follows:

$$\begin{aligned} \mu &= 2 \lim_{r\to\infty} \lim_{N\to\infty} \frac{1}{Z} \left( \sum_{\{\sigma\}}^{+} e^{\beta J \sum_{\langle i,j \rangle} \sigma_i \sigma_j} - \sum_{\{\sigma\}}^{-} e^{\beta J \sum_{\langle i,j \rangle} \sigma_i \sigma_j} \right) \\ &= 2 \lim_{r\to\infty} \lim_{N\to\infty} \frac{e^{2N^2 \beta J}}{Z} \left( \sum_{\text{Polygons}}^{+} e^{-\beta J l} - \sum_{\text{Polygons}}^{-} e^{-\beta J l} \right) \end{aligned} \qquad (133)$$

The indices "+" and "−" indicate that the summations go only over those spin conformations where $\sigma_r = 1$ and $\sigma_0 = +1$ or $\sigma_0 = -1$, respectively. The factor 2 in front of the expressions accounts for the fact that we do not count the states with $\sigma_r = -1$ separately. The transition from the first to the second line is the same as the one from Eq. 128 to 129.

We have now to show that for sufficiently small temperatures the second sum (the one with $\sigma_0 = -1$) is smaller than the first one and hence $\mu > 0$. Since we only want to prove the existence of this ordered phase a rough estimate is sufficient. We do this using the formulation in terms of the polygons. The important point is here that contributions to the second sum need at least to have a closed border around the spin at $i = 0$ wheras this is not the case for the case of equal spins, $\sigma_0 = 1$. We use the inequality

$$\sum_{\text{Polygons}}^{+} e^{-\beta Jl} - \sum_{\text{Polygons}}^{-} e^{-\beta Jl} > \sum_{\text{Polygons}}^{+} e^{-\beta Jl} \left( 1 - 2 \sum_{\text{Polygons around 0}} e^{-2\beta Jl'} \right) \tag{134}$$

The summation "Polygons around 0" means a summation over all closed polygons that enclose the point 0, are simply connected and do not cross or touch themselves. That means for each term in that sum that there is an area with $-1$-spins around the point 0 and outside this area one has $+1$-spins. The reason why this is an inequality lies in the fact that not every polygon around 0 is allowed to transform a given conformation from the "+"-sum into a "−"-configuration because many such polygons around 0 would have common lines with polygons from that given "+"-configuration. The factor 2 accounts for the fact that we can have also put the boundary around the point at $r$. Even though we vastly overestimate the number of "−"-configurations on the righthand side of Eq. 134 we shall see now that the expression is still larger than zero for sufficiently small non-zero temperatures. In other words there is temperature range for large $\beta$-values where

$$\frac{1}{2} > \sum_{\text{Polygons around 0}} e^{-2\beta Jl} \tag{135}$$

A closed polygon of length $l$ encloses an area not larger than $(l/4)^2 = l^2/16$. This gives an upper estimate of the number of possibilities how to position such a polygon around 0. The number of shapes of polygons of length $l$ cannot be larger than $4 \times 3^{l-1}$ as for the first step there are 4 directions to go and for the following steps only 3 as the polygon is self-avoiding. This overcounts largely the number of possible shapes as they need to be closed after $l$ steps and cannot cross themselves. This very rough estimate sits in between the two quantities in Eq. 135 such that

$$\frac{1}{2} > \sum_{l=4,6,8,\ldots}^{\infty} \frac{l^2}{12} 3^l e^{-\beta Jl} > \sum_{\text{Polygons around 0}} e^{-\beta Jl} \tag{136}$$

It is the first inequality that remains to be proven. The summation can be performed exactly:

$$\frac{1}{2} > \sum_{l=4,6,8\ldots}^{\infty} \frac{l^2}{12} 3^l e^{-\beta Jl} = \frac{x^2 \left(4 - 3x + x^2\right)}{3 \left(1 - x\right)^3} \tag{137}$$

with $x = 9e^{-4\beta J}$. One can convince oneself that this inequality is fulfilled for large enough values of $\beta$ (larger than $\beta J \approx 0.8$).

To complete our proof of the existence of a phase transition we show now that there is a phase with vanishing magnetization for sufficiently large temperatures. The high-temperature expansion of the correlation function between 2 spins

$$\langle \sigma_i \sigma_j \rangle = \frac{1}{Z} \sum_{\{\sigma_i = \pm 1\}} \sigma_i \sigma_j e^{-\beta E} \tag{138}$$

can be written as the sum over all polygons where one polygon connects $i$ with $j$ whereas all the other polygons are closed (that way each spin occurs only in the form $\sigma_k^2$ or not at all and thus does not lead to a cancellation in the spin summation). For each term in the polygon summation of $Z$ one obtains a term in the summation of the nominator by adding an allowed connection between $i$ and $j$ weighted with $(\tanh \beta J)^l$ where $l$ denotes the length of that connection ("allowed" means that at each lattice point only 2 or 4 lines can come together). This leads to the inequality

$$\langle \sigma_i \sigma_j \rangle < \sum_{\text{connections } i \to j} (\tanh \beta J)^l \tag{139}$$

We take again the very rough estimate $4 \times 3^{l-1}$ from above for the number of connections of length $l$:

$$\sum_{\text{connections } i \to j} (\tanh \beta J)^l < \frac{4}{3} \sum_{l \geq |i-j|} (3 \tanh \beta J)^l = \frac{4}{3} \frac{(3 \tanh \beta J)^{|i-j|}}{1 - 3 \tanh \beta J} \tag{140}$$

Here $|i-j|$ denotes the length of the shortest connection between point $i$ and $j$. The geometric series converges for sufficiently small values of $\beta$. The value of $\mu$ that follows in the limit $|i-j| \to \infty$ vanishes, i.e. there is no spontaneous magnetization for sufficiently large temperatures.

### 3.2.3 Self-duality of the two-dimensional Ising model

We use now this self-duality to calculate exactly the temperature where the phase transition occurs (following Kramers and Wanniers). We define the dual temperature through

$$e^{-2\beta^* J} = \tanh \beta J \tag{141}$$

i.e.

$$\beta^* J = -\frac{1}{2} \ln \tanh \beta J \tag{142}$$

Comparing the low-temperature expansion

$$Z(\beta) = 2e^{2\beta J N^2} \sum_{l=0}^{\infty} P(l)\, e^{-2\beta J l} \tag{143}$$

and the high-temperature expansion

$$Z(\beta) = \left(2\cosh^2 \beta J\right)^{N^2} \sum_{l=0}^{\infty} P(l) (\tanh \beta J)^l \tag{144}$$

one finds the following relation between $Z(\beta)$ and $Z(\beta^*)$:

$$Z(\beta) = \frac{(2\cosh \beta J \sinh(\beta J))^{N^2}}{2} Z(\beta^*) \tag{145}$$

If we know $Z(\beta)$ we can calculate $Z(\beta^*)$. For large values of $\beta$ one has a small value of $\beta^*$ and vice versa. Equation 145 connects the partition function at low temperatures with the partition function at high temperatures. The duality transformation, Eq. 142 is an involution:

$$\beta^{**} J = -\frac{1}{2} \ln \tanh \beta^* J = -\frac{1}{2} \ln \left[\tanh\left(-\frac{1}{2} \ln \tanh \beta J\right)\right] = \beta J \tag{146}$$

In terms of the free energy per spin $F_\infty = -\lim_{N\to\infty} \left(1/\beta N^2\right) \ln Z$ the relation 145 is given by

$$F_\infty(\beta) = \frac{1}{\beta} \ln(\sinh 2\beta J) + F_\infty(\beta^*) \tag{147}$$

We assume now that the Ising model in 2 dimensions has only one phase transition, i.e., the free energy per spin has only one value where it is non-analytical. That means that this point must be a fixpoint of the duality transformation $\beta \to \beta^*$. This leads to the condition

$$e^{-2\beta J} = \tanh \beta J \tag{148}$$

which is solved for

$$\beta_c J = \frac{1}{2} \ln\left(1 + \sqrt{2}\right) \approx 0.440687 \tag{149}$$

# 4 Meanfield approximation

## 4.1 Introduction

To introduce the meanfield approximation somewhat general, let us first write down the exact probability to find a particle at position $\mathbf{x}_1$:

$$n_1(\mathbf{x}_1) = \frac{1}{Z} \frac{N}{\lambda^{3N} N!} \int d^3x_2 ... d^3x_N e^{-\beta V_N(\mathbf{x}_1, ..., \mathbf{x}_N)} \tag{150}$$

with

$$V_N(x_1, ..., x_N) = \frac{1}{2} \sum_{i \neq j} w(\mathbf{x}_i - \mathbf{x}_j) + \sum_{i=1}^{N} U(\mathbf{x}_i) \tag{151}$$

Here $w$ described the interaction between particles and $U$ is an external potential. Eq. 150 follows by integrating over all degrees of freedom one is not interested in (i.e. the positions of particle 2 to $N$ and all the $N$ momenta). The additional factor $N$ is chosen such that

$$\int d^3x\, n_1(\mathbf{x}) = N \tag{152}$$

Eq. 150 is in general too complicated to be solved explicitely. There are various approximation schemes taking correlations between particles into account up to a certain extent. The crudest approximation is to neglect correlations altogether which is the so-called *meanfield approximation*:

$$n_1(\mathbf{x}) = e^{\beta\left(\mu - U(\mathbf{x}) - \int d^3x' w(\mathbf{x}-\mathbf{x}') n_1(\mathbf{x}')\right)} \tag{153}$$

Here the influence of all the other particles onto a given particle which is given by the potential

$$V(\mathbf{x}) = U(\mathbf{x}) + \sum_{i=2}^{N} w(\mathbf{x}-\mathbf{x}_i) = U(\mathbf{x}) + \int d^3x' w(\mathbf{x}-\mathbf{x}')\, n(\mathbf{x}') \tag{154}$$

where $n(\mathbf{x}') = \sum_{i=2}^{N} \delta(\mathbf{x}-\mathbf{x}_i)$ is replaced by its average

$$U_{\text{eff}}(\mathbf{x}) = U(\mathbf{x}) + \int d^3x' w(\mathbf{x}-\mathbf{x}')\, n_1(\mathbf{x}') \tag{155}$$

Even though this is a very, very big approximation (throwing most details out of the window), the resulting expression, Eq. 153, is often not trivial to solve since it is a nonlinear selfconsistent equation for $n_1(\mathbf{x})$. The meanfield approximation is expected to be good if many particles contribute to $U_{\text{eff}}$. Thus it usually works better for a system in higher space dimensions. It also works well when the particle-particle interaction $w(\mathbf{r})$ becomes long-ranged (like for charged particles in an electrolyte or a plasma). In such a case the virial expansion does not work and in that sense the meanfield approximation can be complementary to the virial expansion. We first discuss the Weiss theory of ferromagnetism, a prototype meanfield theory that predicts a first order phase transition very similar to the one predicted by the van der Waals theory. We then discuss the Poisson-Boltzmann theory for electrolytes (salt solutions). We shall see that this case shows a low density behavior that is markably different from what a virial approximation can predict. We will in addition focus on the role of the nonlinearity and its physical interpretation and discuss the breakdown of meanfield theory in the case of strong ion-ion coupling where correlations dominate the behavior.

## 4.2 Ferromagnetism

We consider here the Ising model on a D dimensional lattice as a simple model for an uniaxial ferromagnet. On each lattice site $i$ site a spin that can assume

the values $\sigma_i = \pm 1$. The Hamiltonian is given by

$$H\left(\{\sigma_i\}\right) = -h\sum_i \sigma_i - J\sum_{\mathrm{NN}} \sigma_i\sigma_j \tag{156}$$

Here $h = mB$ is the energy of the magnetic moment $m$ of the spin in the magentic field $B$, $J$ is the nearest neighbor coupling energy and $\sum_{\mathrm{NN}}$ means the summation over nearest neighbours. The 1D case is trivial to solve but it shows no phase transition. The 2D case features a phase transition and Lars Onsager (1944) managed to calculate the free energy of the 2D Ising model exactly but only for $h = 0$. Nobody has been able so far to solve the 3D case analytically.

The meanfield theory of ferromagnetism provides a simple view (qualitative but not quantitative) of such systems. It predicts a phase transition irrespective of the dimensionality of the system and thus is obviously a bad approximation in 1D where we know that there is no phase transition. It works better and better at larger space dimensions and certain predictions of the theory become exact for $D \geq 4$. The meanfield Hamiltonian is assumed to be

$$H_{\mathrm{MF}}\left(\{\sigma_i\}\right) = -\left(h + Jz\left\langle\sigma\right\rangle\right)\sum_i \sigma_i \tag{157}$$

Here $z$ is the coordination number of the lattice (e.g. $z = 6$ for a three-dimensional cubic lattice). In Eq. 157 the interactions of the given spin $\sigma_i$ with its $z$ nearest neighbours (cf. Eq. 156) has been replaced by the interaction of that spin with the "meanfield" of the neighboring spins, assumed to be given by the mean magnetization per spin $\langle\sigma\rangle$. This approximation obviously neglects correlations like that the spin-spin coupling favors configurations where $\sigma_i$ is surrounded by spins with the same orientation. $\langle\sigma\rangle$ will be calculated below in a selfconsistant manner.

The partition function to the Hamiltian 157 can be calculated exactly:

$$\begin{aligned} Z_{\mathrm{MF}} &= \sum_{\{\sigma_i=\pm 1\}} e^{-\beta H_{\mathrm{MF}}(\{\sigma_i\})} = \sum_{\{\sigma_i=\pm 1\}} \prod_i e^{\beta(h+Jz\langle\sigma\rangle)\sigma_i} \\ &= \left(e^{\beta(h+Jz\langle\sigma\rangle)} + e^{-\beta(h+Jz\langle\sigma\rangle)}\right)^N = \left[2\cosh\left(\beta\left(Jz\left\langle\sigma\right\rangle + h\right)\right)\right]^N \end{aligned} \tag{158}$$

The free energy is thus

$$F_{\mathrm{MF}} = -k_B T \ln Z_{\mathrm{MF}} = -k_B T N \ln\left(2\cosh\left(\beta\left(Jz\left\langle\sigma\right\rangle + h\right)\right)\right) \tag{159}$$

Now we are in the position to calculate the mean-field magnetization per spin:

$$\left\langle\sigma\right\rangle = -\frac{1}{N}\frac{\partial F_{\mathrm{MF}}}{\partial h} = \tanh\left(\beta\left[Jz\left\langle\sigma\right\rangle + h\right]\right) \tag{160}$$

We arrived here is a selfconsistent, nonlinear equation for $\langle\sigma\rangle$, a typical feature of meanfield theories. Here the meanfield $\langle\sigma\rangle$ is simply a constant whereas in the general case Eq. 153 (that accounts also for the possible presence of a

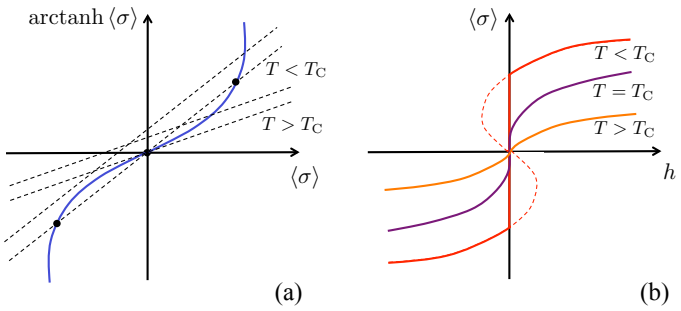



Figure 11: (a) Graphical representation of the selfconsistent equation for $\langle\sigma\rangle$, Eq. 161. The intersections between the lines are solutions of that equation. (b) From the intersections in (a) follow the magentization per particle as function of $h$. Note that for $T < T_C$ one has a first-order phase transition from a state with spontaneous negative magnetization to a state with spontaneous positive magnetization. To make the multivalued $T < T_C$-curve physical one needs to replace it by the curve with a jump, in very much the same way as the Maxwell construction for the liquid-gas transition, Fig. 8.

nonhomogeous external potential) one has to determine a function in space, $n_1(\mathbf{x})$.

We rewrite now Eq. 160 as follows

$$\operatorname{arctanh}\langle\sigma\rangle = \beta h + \beta J z \langle\sigma\rangle \tag{161}$$

In this form the selfconsistent equation can be solved graphically. In Fig.11 we plot the lhs of Eq. 161, namely $\operatorname{arctanh}\langle\sigma\rangle$ vs. $\langle\sigma\rangle$, as well as the rhs of that equation, $\beta h + \beta J z \langle\sigma\rangle$ vs. $\langle\sigma\rangle$. The solutions to Eq. 161 are the points where the curves cross. For $h = 0$ there are either one or three intersections between the arctanh and the linear function. This depends on whether the temperature is above or below a critical temperature

$$T_C = \frac{Jz}{k_B} \tag{162}$$

the so-called *Curie temperature*. If one is above $T = T_C$ there is no spontaneous magnetization, $\langle\sigma\rangle = 0$, below $T = T_C$ the system becomes ferromagnetic and features spontaneous magnetization (the solution with $\langle\sigma\rangle = 0$ is then irrelevant since it becomes a maximum of the free energy 159).

It is important to note that the van der Waals equation of state, Fig. 7, is very similar to the magnetic case, Fig. 11(a), if one identifies $p$ with $h$ and $\langle\sigma\rangle$ with $V$. In fact, one can show that lattice gases can be mapped one-to-one onto spin models. An important experimental difference is, however, that $p$, $V$ and $h$ can be imposed on a system but not $\langle\sigma\rangle$. Note that, as for the van der Waals equation, the meanfield model for ferromagnets works only qualitatively but not quantitatively. For instance, at $T = T_C$ one finds in the meanfield

approximation $\langle\sigma\rangle \sim h^{1/\delta}$ (cf. middle curve in Fig. 11(b)) with $\delta = 3$. But in 1D there is no phase transition at all, in $2D$ one finds such a power law but with $\delta = 15$, in 3D with $\delta \approx 4.8$ . The exponent $\delta$ takes the meanfield value only from 4 dimensions onwards.

## 4.3 Poisson-Boltzmann theory

A living cell is essentially a bag filled with charged objects. Besides the charged macromolecules (DNA, RNA and proteins; see Fig. 12) and the membranes (that also contain some charged lipids) there are lots of small ions. These ions are mostly *cations*, positively charged ions, compensating the overall negative charges of the macromolecules: 5-15mM sodium ions, $Na^+$, 140mM potassium ions, $K^+$, as well as smaller amounts of divalent ions, 0.5mM magnesium, $Mg^{2+}$, and $10^{-7}$mM calcium, $Ca^{2+}$. Here mM stands for *millimolar*, $10^{-3}$ moles of particles per liter. There are also small *anions*, mainly 5-15mM chloride ions, $Cl^-$. We know the forces between those charged object; in fact, basic electrostatics is even taught in school. But even if, for simplicity, we consider the macromolecules as fixed in space, a cell contains a huge number of mobile small ions that move according to the electrostatic forces acting on them which in turn modifies the fields around them and so on. This problem is far too complicated to allow an exact treatment. There is no straightforward statistical physics approach that can treat all kinds of charge-charge interactions occurring inside a cell. In other words, we have not yet a good handle on electrostatics, the major interaction force between molecules in the cell. And that despite many years of hard work. The current chapter tries to give you a feeling of what we understand well and what not.

The standard approach to theoretically describe the many-body problem of mobile charges in an aqueous solution in the presence of charged surfaces is the so-called Poisson Boltzmann (PB) theory. It is not an exact theory but is yet another example of the meanfield approximation. As I will argue, one needs to be quite careful when applying it to the highly charged molecules encountered in a cell.

To construct the PB theory one first distinguishes between mobile and fixed ions. This distinction comes very natural since the small ions move much more rapidly than the macromolecules. So it is usually reasonable to assume that at any given point in time the small ions have equilibrated in the field of the much slower moving macromolecules. Let us denote the concentration of small ions of charge $Z_i e$ by $c_i(\mathbf{x})$ where $e$ denotes the elementary charge and $|Z_i|$ the *valency* of the ion: $|Z_i| = 1$ for monovalent ions, $|Z_i| = 2$ for divalent ions and so on. The concentration of fixed charges, the "macromolecules", is denoted by $\rho_{\text{fixed}}(\mathbf{x})$. The total charge density at point $\mathbf{x}$ is then

$$\rho(\mathbf{x}) = \sum_i Z_i e c_i(\mathbf{x}) + \rho_{\text{fixed}}(\mathbf{x}). \tag{163}$$

From a given charge density $\rho(\mathbf{x})$ the *electrostatic potential* $\varphi(\mathbf{x})$ follows via

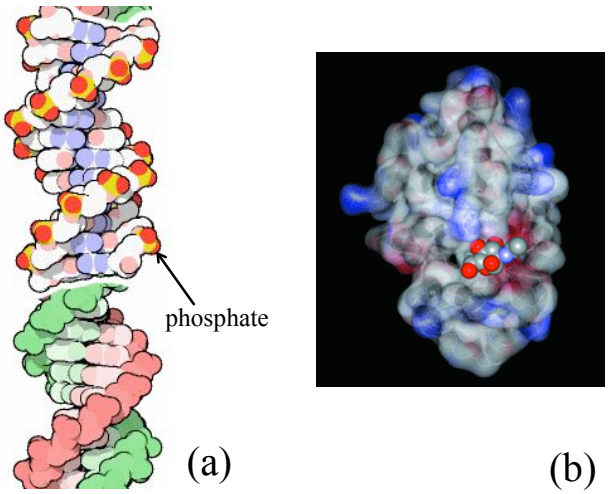



Figure 12: (a) The DNA double helix: each phosphate carries a negative charge. (b) The lysozyme (a protein): the colors indicate the electrostatic potential (blue: positive potential, red: negative)

the *Poisson equation*:

$$\nabla \cdot \nabla \varphi (\mathbf{x}) = \triangle \varphi (\mathbf{x}) = -\frac{4\pi}{\varepsilon} \rho (\mathbf{x}) . \tag{164}$$

Here $\varepsilon$ is the so-called *dielectric constant* that has the value $\varepsilon = 1$ in vacuum and the much larger value $\varepsilon \approx 80$ in water. That this value is so high in water, the main ingredient of the cell, is crucial since otherwise free charges would hardly exist, as we shall see below.

The Poisson equation, Eq. 164, is linear in $\varphi$ and $\rho$ so that it is straightforward to solve for any given charge density. First one needs to know the Green's function, i.e., the solution for a single point charge $e$ at position $\mathbf{x}'$, $\rho (\mathbf{x}) = e\delta (\mathbf{x} - \mathbf{x}')$. Since $\triangle (1/|\mathbf{x} - \mathbf{x}'|) = -4\pi\delta (\mathbf{x} - \mathbf{x}')$ this is given by

$$\varphi (\mathbf{x}) = eG (\mathbf{x}, \mathbf{x}') = \frac{e}{\varepsilon |\mathbf{x} - \mathbf{x}'|} . \tag{165}$$

Having the Green's function $G (\mathbf{x}, \mathbf{x}')$ of the Poisson equation, one can calculate the potential resulting from any given charge distribution $\rho (\mathbf{x})$ via integration:

$$\varphi (\mathbf{x}) = \int G (\mathbf{x}, \mathbf{x}') \rho (\mathbf{x}') d^3 x' = \int \frac{\rho (\mathbf{x}')}{\varepsilon |\mathbf{x} - \mathbf{x}'|} d^3 x' . \tag{166}$$

You can easily check that this solves indeed Eq. 164. Physically the integral in Eq. 166 can be interpreted as being a linear superposition of potentials of point charges, Eq. 165.

Unfortunately things are not as easy here since mobile ions are present. The potential produced by a given charge density is in general not flat so that the

mobile charges experience forces, i.e., they will move. If they move the charge density changes and thus also the potential and so on. What we are looking for is the thermodynamic equilibrium. In that case the charge density of each ion type is given by the Boltzmann distribution:

$$c_i(\mathbf{x}) = c_{0i} e^{-Z_i e\varphi(\mathbf{x})/k_B T} \tag{167}$$

with $c_{0i}$ denoting the charge density at places in space where $\varphi(\mathbf{x}) = 0$. Combining Eqs. 163, 164 and 167 leads to the *Poisson-Boltzmann equation*:

$$\triangle\varphi(\mathbf{x}) + \sum_i \frac{4\pi Z_i e c_{0i}}{\varepsilon} e^{-Z_i e\varphi(\mathbf{x})/k_B T} = -\frac{4\pi}{\varepsilon}\rho_{\text{fixed}}(\mathbf{x}). \tag{168}$$

This is an equation for $\varphi(\mathbf{x})$; the charge densities of the different mobile ion species are then given by Eq. 167. An additional contraint is that the total charge of the system needs to be zero:

$$\int_{\text{system}} \rho(\mathbf{x})\, d^3x = 0. \tag{169}$$

This condition can be understood as follows: If the system of size $R$ (here e.g. the whole cell) would carry a non-vanishing charge $Q$, then the energy that it costed to charge it would scale like $Q^2/(\varepsilon R)$. It is extremely unlikely that this energy would be much larger than the thermal energy and therefore $Q$ needs to stay very small. In other words, the huge positive and negative charges inside the cell need to cancel each other, leading to a total charge $Q$ that can be considered to be zero for any practical purposes.

There are two problems when dealing with a PB equation, one of more practical, the other of principal nature. The practical problem is that this is a non-linear differential equation for the potential $\varphi(\mathbf{x})$ that is usually very hard to solve analytically; there exist exact solutions only in a few special cases, two of which will be discussed below. That $\varphi(\mathbf{x})$ occurs at two different places in Eq. 168 just follows from the above mentioned fact that charges move in response to the potential and at the same time determine the potential. A solution needs to be self-consistent, i.e., the distribution of charges needs to induce an electrical potential in which they are Boltzmann distributed. The non-linearity makes it in many cases hard to understand how sensitive the solution is to details in the charge distribution.

What is, however, much more worrisome is the second problem. Solutions of Eq. 168 are usually smooth functions that look very different to the potentials featured by electrolyte solutions. Close to each ion the potential has very large absolute values that in the limit of point charges go even to infinity. Something has been lost on the way when we constructed the PB equation: Instead of looking at concrete realizations of ion distributions we consider averaged densities $c_i(\mathbf{x})$, Eq. 167. These averages create smooth potentials. This is a typical example of a meanfield approximation: the effect of ions on a given ion is replaced

by an averaged effect. *A priori* it is not clear at all whether such an approximation makes any sense when applied to the electrostatics of the cell. But it is intuitively clear that the field emerging from a solution of monovalent ions shows less dramatic variations than that of a solution of ions of higher valency. The question that we have to answer will be when PB works reasonably well, when it breaks down and what new phenomena might emerge in that case. As we shall see, this a fascinating topic with many surprising results.

**Electrostatics of charged surfaces**

We aim at understanding the electrostatic interactions between macromolecules. Especially we would like to know what happens if two DNA chains come close to each other or if a positively charged protein approaches a DNA chain. Usually the charges are not distributed homogeneously on the surface of a macromolecule. For instance, charges on the DNA double helix are located along the helical backbones and the distribution of charged groups on a protein is often rather complicated, see Fig. 12. Despite these complications, we shall see that one can learn a great deal about these systems by looking at much simpler geometries, especially by looking at the electrostatics of charged flat surfaces. The reason for this is that in many cases all the interesting electrostatics happens very close to the surface of a macromolecule. Essentially the ions experience then the macromolecules in a similar way as we experience our planet, namely as a flat disk. We shall see in the following section that this is indeed true; in this section we focus on charged planes.

To get started we rewrite the PB equation 168 in a more convenient form by multiplying it on both sites by $e/k_BT$:

$$\triangle\Phi\left(\mathbf{x}\right)+\sum_i 4\pi Z_i l_B c_{0i} e^{-Z_i\Phi(\mathbf{x})}=-4\pi l_B\left[\frac{\rho_{\text{fixed}}\left(\mathbf{x}\right)}{e}\right]. \tag{170}$$

Here $\Phi\left(\mathbf{x}\right)$ denotes the dimensionless potential $\Phi\left(\mathbf{x}\right)=e\varphi\left(\mathbf{x}\right)/k_BT$. In addition we introduced in Eq. 170 one of three important length scales in electrostatics, the so-called *Bjerrum length*

$$l_B=\frac{e^2}{\varepsilon k_BT}. \tag{171}$$

This is the length where two elementary charges feel an interaction energy $k_BT$: $e^2/\left(\varepsilon l_B\right)=k_BT$. In water with $\varepsilon=80$ one has $l_B=0.7nm$. This is small enough compared to atomic scales so that two oppositely charged ions "unbind". On the other hand, inside a protein core the dielectric constant is much smaller, roughly that of oil with $\varepsilon\approx 5$, and thus there are hardly any free charges inside the core. Inspecting again Eq. 171 one can see that another route to free charges is to heat a substance to extremely high temperatures. This leads to a so-called *plasma*, a state of matter of no biological relevance.

As warming up exercise let us first consider a simply special case, namely an infinite system without any fixed, only with mobile charges. Suppose we have

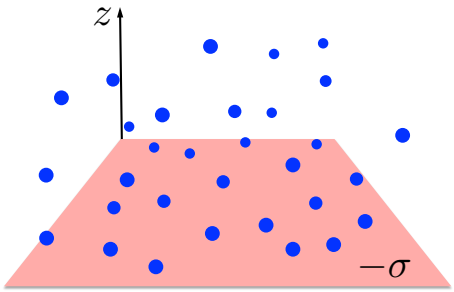



Figure 13: Atmosphere of positively charged counterions (blue) above a surface with a negative charge number density $-\sigma$ (light red) that is assumed to be homogeneously smeared out.

an equal number of positively and negatively charged ions of valency $Z$. In this case the PB equation 170 reduces to

$$-\triangle\Phi\left(\mathbf{x}\right)+8\pi l_{B}Zc_{\mathrm{salt}}\sinh\left(Z\Phi\left(\mathbf{x}\right)\right)=0 \tag{172}$$

with $c_{\mathrm{salt}}$ denoting the bulk ion density, the salt concentration. At first sight Eq. 172 might look difficult to solve but in fact the solution is as trivial as possible, namely

$$\Phi\left(\mathbf{x}\right)=0. \tag{173}$$

everywhere. This result is rather disappointing but not really surprising since the PB equation results from a mean-field approximation. And the mean electrical field of an overall neutral system of uniform positive and negative charges vanishes. In reality one has thermal fluctuations that lead locally to an imbalance between the two charge species. But such fluctuations are not captured in PB theory. So far it seems that PB produces nothing interesting. This is, however, not true: as soon as fixed charges are introduced one obtains non-trivial insights. As we shall see later on, even the fluctuations in a salt solution in the absence of fixed charges can be incorporated nicely in a linearized version of the PB theory, the Debye-Hückel theory, that we shall discuss later.

In the following we study the distribution of ions above a charged surface as depicted in Fig. 13. This is an exactly solvable case that provides crucial insight into the electrostatics of highly charged surfaces and – as we shall see later – of DNA itself. The system consists of the infinite half-space $z\geq 0$ and is bound by a homogeneously charged surface of surface charge number density $-\sigma$ at $z=0$. Above the surface, $z>0$, we assume to have only ions that carry charges of sign opposite to that of the surface, so-called *counterions*. The counterions can be interpreted to stem from a chemical dissociation at the surface, leaving behind the surface charges. These ions will make sure that the charge neutrality condition, Eq. 169, is respected. We assume that there is no added salt, i.e., there are no negatively charged ions present. The PB equation, Eq. 170, takes

now the following form:

$$\Phi''(z) + Ce^{-\Phi(z)} = 4\pi l_B \sigma \delta(z). \tag{174}$$

We replaced here the term $4\pi l_B Z c_0$ by the constant $C$ to be determined below and the primes denote differentiations with respect to $z$, $\Phi' = d\Phi/dz$. As a result of the symmetry of the problem, this is an equation for the $Z$-direction only since the potential is constant for directions parallel to the surface.

To solve Eq. 174 let us consider the space above the surface, $z > 0$. Due to the absence of fixed charges, we find

$$\Phi''(z) + Ce^{-\Phi(z)} = 0. \tag{175}$$

Multiplying this equation with $\Phi'$ and performing an integrating along $z$ leads to

$$E = \frac{1}{2}(\Phi')^2 - Ce^{-\Phi} \tag{176}$$

where $E$ denotes an integration constant. To solve Eq. 176 we use the trick of the *separation of variables*, here of $z$ and $\Phi$, i.e., we rewrite this equation as

$$dz = \pm \frac{d\Phi}{\sqrt{2E + 2Ce^{-\Phi}}}. \tag{177}$$

Integration yields

$$z - \bar{z} = \pm \int_{\bar{\Phi}}^{\Phi} \frac{d\Phi}{\sqrt{2E + 2Ce^{-\Phi}}} \tag{178}$$

where we start the integration at height $\bar{z}$ above the surface where $\Phi(\bar{z}) = \bar{\Phi}$. As we shall see *a posteriori* we obtain the solution with the right boundary conditions if we use the positive sign and set $E = 0$. This makes the integral in Eq. 178 trivial. If we set $\bar{z} = 0$ and choose $\bar{\Phi} = 0$ we find

$$z = \frac{1}{\sqrt{2C}} \int_0^{\Phi} e^{\Phi/2} d\Phi = \sqrt{\frac{2}{C}} \left( e^{\Phi/2} - 1 \right). \tag{179}$$

Solving this for $\Phi$ gives finally the potential as function of $z$:

$$\Phi = 2 \ln \left( 1 + \sqrt{\frac{C}{2}} z \right). \tag{180}$$

At a charged surface the *electrical field* $-d\varphi/dz$ makes a jump proportional to the surface charge density. It vanishes below the surface and attains just above the surface the value

$$\left. \frac{d\Phi}{dz} \right|_{z \downarrow 0} = 4\pi l_B \sigma = \sqrt{2C}. \tag{181}$$

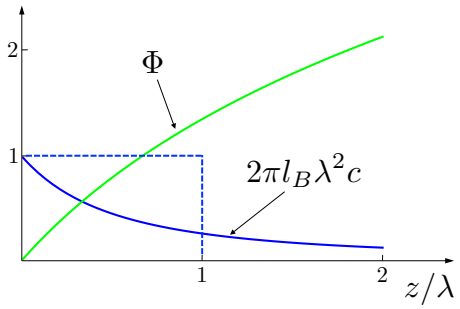



Figure 14: Potential $\Phi$, Eq. 183, and rescaled counterion density $2\pi l_B\lambda^2 c$, Eq. 186, as a function of the rescaled height $h/\lambda$ above a charged surface. The dashed lines indicate a simplified counterion profile where all counterions form an ideal gas inside a layer of thickness $\lambda$.

This sets $C$. In fact, $\sqrt{2/C}$ turns out to be the second important length scale in electrostatics, the *Gouy-Chapman length*:

$$\lambda = \frac{1}{2\pi l_B \sigma}. \tag{182}$$

The physical meaning of this length becomes clear further below. We can now rewrite Eq. 180 as

$$\Phi = 2\ln\left(1 + \frac{z}{\lambda}\right). \tag{183}$$

The atmosphere of counterions above the surface is then distributed according to Eq. 167:

$$c(z) = c_0 e^{-\Phi} = \frac{c_0\lambda^2}{(z+\lambda)^2}. \tag{184}$$

The prefactor $c_0$ in Eq. 184 has to be chosen such that the total charge of the counterions exactly compensates the charge of the surface, see Eq. 169:

$$\int_{-\infty}^{\infty} [c(z) - \sigma\delta(z)]\, dz = 0. \tag{185}$$

This sets $c_0$ to be $\sigma/\lambda$ and hence

$$c(z) = \frac{1}{2\pi l_B (z+\lambda)^2}. \tag{186}$$

This distribution is depicted together with the potential $\Phi$, Eq. 183, in Fig. 14.

The density of ions above the surface decays algebraically as $z^{-2}$ for distances larger than $\lambda$. This is somewhat surprising since we have seen that the distribution of gas molecules in a gravity field decays exponentially, namely as $c(z) \sim e^{-mgz/k_BT}$, the so-called *barometric formula*. The physical reason is

that the gas particles do not feel each other but the ions do. The higher the ions are above the surface, the less they "see" the original surface charge density since the atmosphere of ions below masks the surface charges. As a result the ions farer above the surface feel less strongly attracted which leads to a slower decay of the density with height.

We can now attach a physical meaning to the Gouy-Chapman length $\lambda$. First of all, $\lambda$ is the height up to which half of the counterions are found since $\int_0^\lambda c(z)\,dz = \sigma/2$. Secondly, if we take a counterion at the surface where $\Phi(0) = 0$ and move it up to the height $\lambda$ where $\Phi(\lambda) = 2\ln 2$ we have to do work on the order of the thermal energy, $e\varphi = 2\ln 2\, k_B T \approx k_B T$. One can say that the ions in the layer of thickness $\lambda$ above the surface form an ideal gas since the thermal energy overrules the electrostatic attraction to the surface. On the other hand, if an ion attempts to "break out" and escape to infinity, it will inevitably fail since it would have to pay an infinite price: $\Phi \to \infty$ for $z \to \infty$. That means that all the counterions are effectively bound to the surface. But half of the counterions, namely those close to the surface, are effectively not aware of their "imprisonment."

Based on these ideas let us now try to estimate the free energy $f_{\text{approx}}$ per area of this so-called *electrical double layer*. We assume that all the counterions form an ideal gas confined to a slab of thickness $\lambda$ above the surface as indicated in Fig. 14 by the dashed line. The density of the ions is thus $c = \sigma/\lambda$ that, according to Eq. 53, leads to the free energy density

$$\beta f_{\text{approx}} = c\left[\ln\left(c\lambda_T^3\right) - 1\right]\lambda = \sigma\left[\ln\left(\frac{\sigma\lambda_T^3}{\lambda}\right) - 1\right] \tag{187}$$

where $\lambda_T$ is the thermal de Broglie length, see Eq. 11.

We show now that this simple expression is astonishingly close to the exact (mean-field) expression. A more formal, less intuitive way of introducing the PB theory would have been to write down an appropriate free energy functional $F$ from which the PB equation follows via minimization. This functional is the sum of the electrostatic internal energy and the entropy of the ions in the solution:

$$\beta F = \frac{1}{8\pi l_B}\int \left(\nabla\Phi(\mathbf{r})\right)^2 d^3r + \int\left[\frac{\rho(\mathbf{r})}{e}\right]\left(\ln\left(\left[\frac{\rho(\mathbf{r})}{e}\right]\lambda_T^3\right) - 1\right)d^3r. \tag{188}$$

Replacing $\rho(\mathbf{r})$ in this functional by $\Phi(\mathbf{r})$ through the Poisson equation $\triangle\Phi = -4\pi l_B\rho/e$, Eq. 164, one finds that the Euler-Lagrange equation is indeed identical to the PB equation, Eq. 170, namely here $\triangle\Phi(\mathbf{x}) + 4\pi l_B c_0 e^{-\Phi(\mathbf{x})} = 0$. Inserting the PB solution for a charged surface, Eqs. 183 and 186 into the free energy functional, Eq. 188, we find the following free energy density per area:

$$\beta f = \sigma\left[\ln\left(\frac{\sigma\lambda_T^3}{\lambda}\right) - 2\right]. \tag{189}$$

The exact expression, Eq. 189, differs from the approximate one, Eq. 187, just by a term $-\sigma$. Given that agreement, it is fair to say that we have achieved

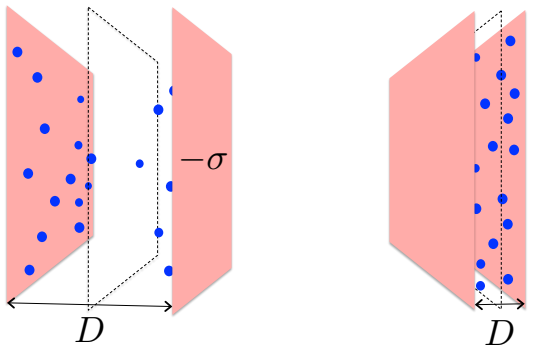



Figure 15: Two parallel, negatively charged surfaces and their counterions. (a) For large separations $D$ between the surfaces the two counterion clouds hardly interact. (b) If the planes are closeby the two clouds combine and form a dense "gas", homogeneously distributed across the gap.

a rather clear qualitative understanding of the physics of the electrical double layer.

Since we are mainly interested in the interactions between macromolecules, especially between two DNA molecules and between a DNA molecule and a protein, we discuss now two model cases: the interaction between two negatively charged surfaces and the interaction between two oppositely charged surfaces. We begin with two negatively charged surfaces. The exact electrostatics can be worked out along the lines of Eqs. 174 to 180 using appropriate values of the integration constant. We prefer to give here a more physical line of argument. Suppose the two parallel walls, at distance $D$, carry exactly the same surface charge density $-\sigma$, see Fig. 15. Then due to the symmetry of the problem the electrical field in the midplane vanishes; this plane is indicated in the drawing, Fig. 15, by dashed lines. The *disjoining pressure* $\Pi$ between the two planes, i.e., the force per area with which they repel each other, can then be easily calculated since it must equal the pressure of the counterions in that midplane. Using the ideal gas law, Eq. 27, we find

$$\frac{\Pi}{k_B T} = c\left(\frac{D}{2}\right). \tag{190}$$

Without doing any extra work we can now predict the disjoining pressure between the two surfaces in two asymptotic cases. If the distance is much larger than the Gouy-Chapman length $\lambda$ of the planes, $D \gg \lambda$, we can assume that the two counterion clouds are independent from each other. The density in the midplane is then the sum of the two single-plane densities, see Fig. 15(a). From Eq. 186 we obtain

$$\frac{\Pi}{k_B T} \approx 2\frac{1}{2\pi l_B \frac{D^2}{4}} = \frac{4}{\pi l_B D^2}. \tag{191}$$

Remarkably the disjoining pressure is here independent of $\sigma$. This results from

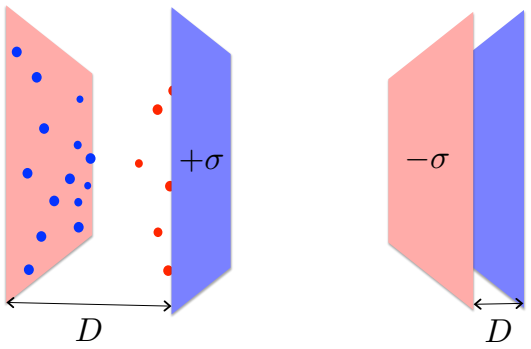



Figure 16: Two oppositely charged surfaces and their counterions. (a) For large separation $D$ between the surfaces the two counterion clouds hardly interact. (b) If the planes are closeby the counterions are not needed anymore. They gain entropy by escaping to infinity.

the fact that the single plane counterion density, Eq. 186, becomes independent of $\lambda$ (and thus $\sigma$) for $D \gg \lambda$. In the other limit, $D \ll \lambda$, the two counterion clouds are strongly overlapping and we expect a flat density profile, see Fig. 15(b). Hence

$$\frac{\Pi}{k_B T} \approx \frac{2\sigma}{D} = \frac{1}{\pi l_B \lambda D}. \tag{192}$$

The pressure is here linear in $\sigma$, reflecting the counterion density. Note that these results show that the situation is very different from how we are used to think about electrostatics, namely that the pressure results from the direct electrostatic repulsion of the two charged surfaces. In fact, in the absence of counterions the electrical field between the surfaces is constant and follows from the boundary condition, Eq. 181. This leads to

$$\frac{\Pi}{k_B T} = 4\pi l_B \sigma^2 \tag{193}$$

that is independent from the distance between the surfaces and, as a result of the pairwise interaction between surface charges, proportional to $\sigma^2$. Thus the counterions completely change and, in fact, "rule" the electrostatics.

This becomes even more evident when looking at the interaction between two oppositely charged surfaces, see Fig 16. Such a situation arises when a positively charged protein comes close to a negatively charged DNA molecule. For simplicity, let us assume that the number charge densities of the two surfaces are identical, $\sigma^+ = \sigma^- = \sigma$. If the two surfaces are very far from each other, we can assume that both form the usual electrical double layer of thickness $\lambda$, one with positive counterions, one with negative ones, see Fig 16(a). If the two surfaces come close to each other, Fig 16(b), there is, however, no need for counterions anymore since the two surfaces can neutralize each other. The counterions can therefore escape to infinity and gain translational entropy on

the order of $k_BT$. The binding energy per area of the two surfaces as a result of this *counterion release* should thus be something on the order of $k_BT\sigma$. If the surface charge densities are not the same, charge neutrality enforces that some of the counterions remain between the surfaces.

Note that our model system that assumes two infinitely large surfaces and no added salt is quite academic and that a precise calculation of this effect is not possible in the current framework. No matter how far the two surfaces are apart: if we look at length scales much larger than the surface separation of the two surfaces, they look together like a neutral plane. As a result, the counterions are never really bound. In the following sections we have to come up with slightly more realistic situations that allow better descriptions of the counterion release mechanism.

**Electrostatics of cylinders and spheres**

So far we have discussed planar charged surfaces. However, at length scales below its persistence length the DNA double helix looks more like a cylinder and the shapes of globular proteins might be better described by spheres. We ask here the question whether the basic physics that we described in the previous section still holds for such objects. As we shall see, this is actually a subtle problem that can be understood in beautiful physical terms.

Let us start with DNA. DNA is a charged cylinder with a diameter of $2nm$ and line charge density of $-2e/0.33nm$. The question that we like to answer is whether such a charged cylinder has its counterions effectively bound or whether they are free. The answer is surprising: Around three quarter of the DNA's counterions are indeed condensed but the rest is free and can go wherever they like. We give here a simple physical argument that goes back to the great Norwegian scientist Lars Onsager. For simplicity, we describe the DNA molecule as an infinitely long cylinder of line charge density $-e/b$ and diameter $2R$. The charges are assumed to be homogeneously smeared out on its surface. The dimensionless electrostatic potential of a cylinder is known to be

$$\Phi(r) = \frac{2l_B}{b} \ln\left(\frac{r}{R}\right) \tag{194}$$

where $r \geq R$ denotes the distance from the centerline of the cylinder. Suppose we start with a universe that consists only of one infinite cylinder. Now let us add one counterion. We ask ourselves whether this counterion will be bound to the cylinder or whether it is able to escape to infinity. In order to find out we introduce two arbitrary radii $r_1$ and $r_2$ with $r_2 \gg r_1 \gg R$ as depicted in Fig. 17. Now suppose the counterion tries to escape from the cylindrical region of radius $r_1$ to the larger cylindrical region of radius $r_2$. According to Eq. 194 the counterion has to pay a price, namely it has to move uphill in the electrostatic potential by an amount of the order of

$$\Delta\Phi = \Phi(r_2) - \Phi(r_1) = \frac{2l_B}{b} \ln\left(\frac{r_2}{r_1}\right). \tag{195}$$

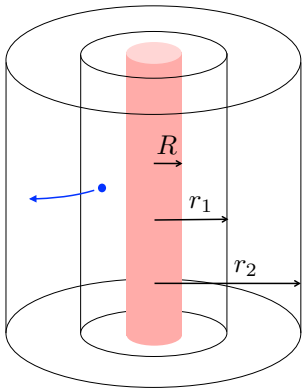



Figure 17: Onsager's argument for counterion condensation on charged cylinders is based on an estimate of the free energy change for a counterion that goes from a cylindrical region of radius $r_1$ to a larger region of radius $r_2$.

At the same time it has a much larger space at its disposal, i.e., it enjoys an entropy gain. The entropy of a single ion in a volume $V$ follows from the ideal gas entropy $S = k_B N \ln\left(V/\left(N\lambda_T^3\right) + 5/2\right)$ with $N = 1$. This equation follows from combining Eq. 53 with Eqs. 26 and 51. We assume here $V \gg \lambda_T^3$ so that we can neglect the 5/2-term. When the ion moves from the smaller to the larger region we find the following change in entropy:

$$\Delta S = S\left(r_2\right) - S\left(r_1\right) = k_B \ln\left(\frac{r_2^2}{r_1^2}\right) = 2k_B \ln\left(\frac{r_2}{r_1}\right). \tag{196}$$

Altogether this amounts to a change in the free energy of

$$\Delta F/k_B T = \Delta\Phi - \Delta S/k_B = 2\left(\frac{l_B}{b} - 1\right) \ln\left(\frac{r_2}{r_1}\right). \tag{197}$$

There are two possible cases. For *weakly charged* cylinders, $b > l_B$, the free energy change is negative, $\Delta F < 0$, and the counterion eventually escapes to infinity. For *highly charged* cylinders, $b < l_B$, one finds $\Delta F > 0$. In that case the energy cost is too high as compared to the entropy gain and the counterion stays always in the vicinity of the cylinder.

Now the same argument can be used for the rest of the counterions. What we have to do is simply to add, one by one, all the counterions. The non-trivial and thus interesting case is that of a highly charged cylinder with $b < l_B$. In the beginning all the counterions that we add condense, thereby reducing the effective line charge. This continues up to the point when the line charge density has been lowered to the value $-e/l_B$. All the following counterions that are added feel a cylinder that carries an effective line density that is just too weak to keep them sufficiently attracted allowing them to escape to infinity. To conclude, the interplay between entropy and energy regulates the charge density

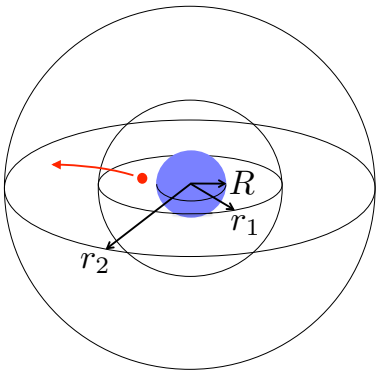



Figure 18: Onsager's argument applied to a charged sphere.

of a cylinder to the critical value $-e/l_B$. Cylinders with a higher effective charge density simply cannot exist.

According to the above given definition DNA is a highly charged cylinder. Counterion condensation reduces its bare charge density of $1/b = 2/(0.33nm)$ to the critical value $1/l_B = 1/(0.7nm)$. That means that a fraction

$$\frac{e/b - e/l_B}{e/b} = 1 - \frac{b}{l_B} = 1 - \frac{1}{\xi}, \tag{198}$$

i.e., about 76%, of the DNA's counterions are condensed. Counterion condensation on cylinders is called *Manning condensation* and is characterized by the dimensionless ratio $\xi = l_B/b$, the *Manning parameter*. Cylinders with $\xi > 1$ are highly charged and have condensed counterions. More precise treatments based on the PB equation show that this simple line of arguments is indeed correct.

There is another interesting interpretation for Manning condensation. We have seen above that all the counterions of an infinite, planar surface are condensed. Now a cylinder looks like a flat surface to a counterion if the Gouy-Chapman length, the typical height in which it lives above the surface, is much smaller than the radius of the cylinder, i.e., if $\lambda \ll R$. Using the definition of $\lambda$, Eq. 182, this leads to the condition $\xi \gg 1$. One can say that for $\xi > 1$ a counterion experiences the cylinder as a flat surface and thus stays bound to it.

Let us now study a model protein, a sphere of radius $R$ that carries a total charge $eZ$ homogeneously smeared out over its surface. We can again use an Onsager-like argument by adding a single counterion to a universe that consists only of that sphere. We estimate the change in free energy when the counterion moves from a spherical region of radius $r_1 \gg R$ around the sphere to a larger region of radius $r_2 \gg r_1$, see Fig. 18. The change in electrostatic energy is given by $\Delta\Phi = \Phi(r_2) - \Phi(r_1) = l_B Z\left(\left(r_2^{-1} - r_1^{-1}\right)\right)$ and that of the entropy by $\Delta S = 3\ln(r_2/r_1)$. We learn from this that the free energy change $\Delta F/k_BT = \Delta\Phi - \Delta S/k_B$ goes to $-\infty$ for $r_2 \to \infty$, no matter how highly the sphere is charged. This suggests that a charged sphere will always loose all its counterions.

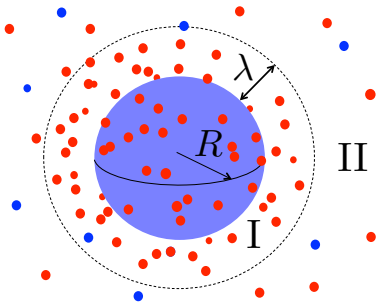



Figure 19: A highly charged sphere in a salt solution. To a good approximation ions can "live" in two zones. Zone I contains "condensed" counterions, zone II the bulk electrolyte solution.

Our results on counterion condensation that we have obtained so far can be summarized as follows. The fraction of condensed ions, $f_{\rm cond}$, depends on the shape of the charged object as follows:

plane: $f_{\rm cond} = 1$,

cylinder: $f_{\rm cond} = 1 - \xi^{-1}$ for $\xi > 1$, $f_{\rm cond} = 0$ otherwise,

sphere: $f_{\rm cond} = 0$.

It is important to realize that we have considered so far fairly academic special cases. First of all, we assumed infinitely extended planes and infinitely long cylinders but any real object is of finite size. Any object of finite extension looks from far apart like a point charge and will thus loose all its counterions, as a sphere does. One might therefore think that theorizing about counterion condensation is a purely academic exercise. This is luckily not the case since, as we shall see now, counterions might also condense on spheres. We came above to the conclusion that for the spherical case $f_{\rm cond} = 0$ by assuming that we had only one sphere in the universe. If there is a finite density of spheres, each with its counterions, the situation can be different. Also we assumed that there are no small ions present, except the counterions of the sphere. If we have a single sphere but a finite salt concentration, the situation can again be different from the above given academic case. In both cases, for a finite density of spheres or for a finite salt concentration, the entropy gain for a counterion to escape to infinity is not infinite anymore. Depending on the sphere charge and on the concentration of small ions in the bulk, there might be a free energy penalty instead.

We consider now a single sphere in a salt solution following the line of argument given by Alexander and coworkers (1984). For a highly charged sphere at moderate salt concentration $c_{\rm salt}$ they postulated two zones, see Fig. 19. Zone I is the layer of condensed counterions of thickness $\lambda$ and zone II is the bulk. When a counterion from the bulk, zone II, enters zone I it looses entropy since

it goes from the dilute salt solution of concentration $c_{\rm salt}$ to the dense layer of condensed counterions. The ion concentration of that layer can be estimated to be $c_{\rm cond} \approx \sigma/\lambda = 2\pi l_B \sigma^2$ where $\sigma$ denotes the surface charge number density of the sphere. We assume here that the sphere is so highly charged that most of its counterions are confined to zone I. The entropy loss is then given by

$$\Delta S = S_{\rm I} - S_{\rm II} \approx -k_B \ln \frac{c_{\rm cond}}{c_{\rm salt}} = -k_B \Omega. \tag{199}$$

The counterion also gains something by entering zone I. In zone II it does not feel the presence of the charged sphere since the electrostatic interaction is screened by the other small ions as shall become clear in the following section. On the other hand, in zone I it sees effectively a sphere of charge $Z^*$ where $Z^*$ denotes the sum of the actual sphere charge, $Z$, and the charges from the condensed counterions inside zone I. The gain in electrostatic energy is thus

$$\Delta \Phi \approx -\frac{l_B Z^*}{R}. \tag{200}$$

If we start with a system where all counterions are inside the bulk, counterions flow into zone I up to a point when there is no free energy gain anymore. This point is reached when the charge is renormalized to the value

$$eZ^* = e\Omega \frac{R}{l_B}. \tag{201}$$

To formulate it in a more elegant way: $Z^*$ is the point where the chemical potentials of zone I and II are identical.

Note, however, that in order to obtain Eq. 201 we cheated a bit since we assumed that $\Omega$ is a constant. This is not really the case since according to Eq. 199 $\Omega$ depends on $c_{\rm cond}$ and thus on $Z^*$. Since this dependence is logarithmic, i.e., very weak, this simplification is quite reasonable and one can assume $\Omega$ to be a constant with a value of around 5 for typical salt concentrations and surface charge densities encountered in cells. A more concise way of calculating the renormalized charge $Z^*$ is given in the next section.

We are now in the position to refine our argument on counterion release from above. Consider again the case of two oppositely charged surfaces as depicted in Fig. 16 but with additional salt. In the case of equal surface charge densities of the two surfaces all the counterions are released and the free energy gain reflects the change of concentration that the counterions experience. The free energy change per surface scales thus as

$$\frac{f}{k_B T} \approx \sigma \Omega \approx \sigma \ln \frac{2\pi l_B \sigma^2}{c_{\rm salt}}. \tag{202}$$

When discussing PB theory above – especially for spherical geometry where no analytical solutions exist – we had to rely on simplified arguments. It turns out that one can gain a great deal of insight by linearizing PB theory. Strictly speaking such a linearization makes only sense for weakly charged surfaces but we shall see that there is an elegant argument that allows us also to extend this framework to highly charged objects.

**Debye-Hückel theory**

As mentioned earlier, the PB equation is hard to handle since it is non-linear. Here we study its linearized version, the well-known *Debye-Hückel (DH) theory*. It provides an excellent approximation to PB theory for the case that the fixed charges are weak. Consider the PB equation of a salt solution of valency $Z = 1$ and concentration $c_{\text{salt}}$ in the presence of fixed charges of density $\rho_{\text{fixed}}$. The PB equation 170 takes then the form

$$\triangle\Phi + 4\pi l_B c_{\text{salt}} \left(e^{-\Phi} - e^{+\Phi}\right) = -4\pi l_B \frac{\rho_{\text{fixed}}}{e}. \tag{203}$$

Let us now assume that the electrostatic energy is small everywhere, i.e., that $\Phi(\mathbf{x}) \ll 1$ for all $\mathbf{x}$. In that case we can linearize the exponential functions, $e^{\Phi} \approx 1 + \Phi$ and $e^{-\Phi} \approx 1 - \Phi$. This results in the DH equation

$$-\Delta\Phi + \kappa^2\Phi = 4\pi l_B \frac{\rho_{\text{fixed}}}{e}. \tag{204}$$

We introduced here the final of the three length scales important in electrostatics, the *Debye screening length* $\kappa^{-1}$. For monovalent salt, as assumed here, this length is given by

$$\kappa^{-1} = \frac{1}{\sqrt{8\pi l_B c_{\text{salt}}}}. \tag{205}$$

Its physical meaning will become clear below.

We can now come back to the disappointing result we encountered earlier when we looked at a salt solution in the absence of fixed charges where the PB equation 172 is solved by $\Phi \equiv 0$. This has not changed here since also the DH equation produces the same trivial answer. But now we are in the position to go beyond this result and to include in our discussion correlations between salt ions. This would have been very difficult to do for the PB equation where no appropriate analytical solutions are available. Consider a point charge $+eZ$ at position $\mathbf{x}'$. The DH equation for such a test charge takes the form:

$$\left[-\Delta + \kappa^2\right] G(\mathbf{x}, \mathbf{x}') = 4\pi l_B Z \delta(\mathbf{x} - \mathbf{x}'). \tag{206}$$

Knowing $G(\mathbf{x}, \mathbf{x}')$, the Green's function, allows to calculate $\Phi$ for an arbitrary distribution of fixed charges:

$$\Phi(\mathbf{x}) = \int G(\mathbf{x}, \mathbf{x}') \frac{\rho_{\text{fixed}}(\mathbf{x}')}{eZ} d^3\mathbf{x}'. \tag{207}$$

The Green's function for Eq. 206 is given by

$$G(\mathbf{x}, \mathbf{x}') = \frac{l_B Z}{|\mathbf{x} - \mathbf{x}'|} e^{-\kappa|\mathbf{x}-\mathbf{x}'|}. \tag{208}$$

One calls this a *Yukawa-type potential*, referring to Yukawa's original treatment introduced to describe the nuclear interaction between protons and neutrons due to pion exchange. That this indeed solves Eq. 206 can be checked by letting

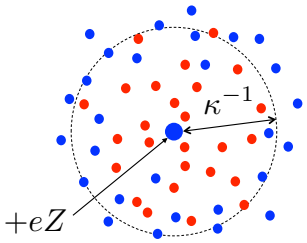



Figure 20: An ion of charge $+eZ$ is surrounded by an oppositely charged ion cloud of typical size $\kappa^{-1}$.

the Laplace operator in spherical coordinates act on the potential of a point charge at the origin:

$$-\left(\frac{\partial^2}{\partial r^2}+\frac{2}{r}\frac{\partial}{\partial r}\right)\frac{e^{-\kappa r}}{r}=-4\pi\delta\left(r\right)-\kappa^2\frac{e^{-\kappa r}}{r}. \tag{209}$$

To derive Eq. 209 we use the fact that $\triangle\left(1/\left|\mathbf{x}-\mathbf{x}'\right|\right)=-4\pi\delta\left(\mathbf{x}-\mathbf{x}'\right)$ as mentioned above Eq. 165.

What is the physical picture behind Eq. 208? In the absence of any salt ions one would have just the potential $\Phi\left(\mathbf{x}\right)=l_BZ/\left|\mathbf{x}-\mathbf{x}'\right|$ around our test charge, i.e., Eq. 208 without the exponential term or, if you prefer, the full Eq. 208 but with $\kappa=0$. In the presence of salt ions the test charge is surrounded by an oppositely charged ion cloud as schematically depicted in Fig. 20. This ion cloud effectively screens the test charge so that the potential decays faster than $1/r$, namely like $e^{-\kappa r}/r$. The screening length $\kappa^{-1}$ reflects the typical cloud size.

Having at hand an expression for the potential around an ion, we calculate now the free energy of a salt solution on the level of the DH theory. As a first step we determine the change of the self-energy of an ion that is brought from ion-free water to the salt solution. We consider the ion as a homogeneously charged ball of radius $a$ and charge density $\rho=3e/\left(4\pi a^3\right)$. We shall show below that the result will not depend on the radius so that we can take the limit $a\to0$. In an electrolyte free environment the self energy is

$$\lim_{a\to0}\frac{1}{2}\int d^3x'\int d^3x\frac{\rho\left(\mathbf{x}\right)\rho\left(\mathbf{x}'\right)}{\varepsilon\left|\mathbf{x}-\mathbf{x}'\right|}=\frac{1}{2}\left.\frac{e^2}{\varepsilon a}\right|_{a=0}=\infty. \tag{210}$$

On the right-hand side we assumed a point-like charge for which $\rho\left(\mathbf{x}\right)=\delta\left(\mathbf{x}\right)$. There is evidently a problem since the self-energy of the point charge is infinite. Let us nevertheless go ahead and calculate also the self-energy of the point charge inside an electrolyte solution:

$$\lim_{a\to0}\frac{1}{2}\int d^3x'\int d^3x\frac{\rho\left(\mathbf{x}\right)\rho\left(\mathbf{x}'\right)}{\varepsilon\left|\mathbf{x}-\mathbf{x}'\right|}e^{-\kappa\left|\mathbf{x}-\mathbf{x}'\right|}=\frac{1}{2}\left.\frac{e^2e^{-\kappa a}}{\varepsilon a}\right|_{a=0}=\infty. \tag{211}$$

Also here the self-energy is infinite. What saves us is that we are not interested what it costs to "form" a point ion. What we want to know instead is the change in the self-energy when the ion is transferred from ion-free water to the electrolyte solution. This change turns out to be finite:

$$\beta E_{\text{self}} = \frac{l_B}{2} \lim_{a \to 0} \left[ \frac{e^{-\kappa a}}{a} - \frac{1}{a} \right] = -\frac{l_B \kappa}{2}. \tag{212}$$

Each particle in the electrolyte contributes this value to the internal energy. This leads to the following change in the internal energy density:

$$\beta \Delta u = 2 c_{\text{salt}} E_{\text{self}} = -\frac{\kappa^3}{8\pi}. \tag{213}$$

Combining Eqs. 14 and 51 we know that the average internal energy density $\langle \Delta u \rangle$ follows from the free energy density $\triangle f$ via

$$\langle \Delta u \rangle = \frac{\partial}{\partial \beta} \left[ \beta \triangle f \right]. \tag{214}$$

This allows us to calculate the electrostatic contribution of the charge fluctuations to the free energy density:

$$\Delta f = -k_B T \frac{\kappa^3}{12\pi}. \tag{215}$$

This finding should surprise you. We discussed in Section 2.3 the impact of the interaction between particles of a real gas on its pressure and free energy. We found to lowest order in the density $n$ that the ideal gas expressions are changed by terms of the order $n^2$, see Eqs. 96 and 101. This reflects interactions between pairs of particles. Surprisingly, for the ion solution we find that interactions between ions lead to a free energy contribution proportional to $\kappa^3 \sim c_{\text{salt}}^{3/2}$ instead. How can one understand this discrepancy? The reason lies in the fact that the electrostatic interaction decays very slowly with distance. If one attempts to calculate the second virial coefficient $B_2$ for such a long-ranged $1/r$-potential one finds a diverging integral: According to Eq. 79 the integrand is proportional to $r^2 \left( e^{-\beta w(r)} - 1 \right)$ which scales then for large $r$ as $r^2 \left( 1/r \right) = r$.

We provide now a scaling argument that makes Eq. 215 transparent. Consider a very small volume $V$ inside the electrolyte solution. Ions can enter and leave this volume at will, as if they would be uncharged and as a result the volume displays random fluctuations in its net charge. According to the central limit theorem the net charge $Q$ can be estimated to be proportional to the square root of the number of ions $N_{\text{ion}}$ inside that volume, i.e.,

$$Q/e \approx \pm \sqrt{N_{\text{ion}}} = \pm \sqrt{c_{\text{salt}} V}. \tag{216}$$

The assumption that the ions are independent of each other is only true up to regions of size $L$ with volume $V = L^3$ for which the electrostatic self energy equals the thermal one

$$\frac{l_B Q^2}{L} = 1. \tag{217}$$

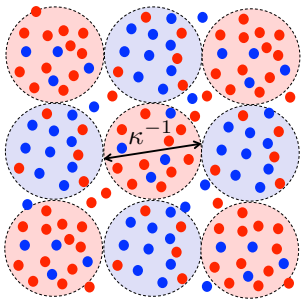



Figure 21: Schematic sketch of charge fluctuations inside an electrolyte solution. Regions of typical size $\kappa^{-1}$ with an excess of negative ions are surrounded by regions with positive net charge and vice versa.

This condition can be rewritten as

$$L = \frac{1}{\sqrt{l_B c_{\text{salt}}}} \sim \kappa^{-1}, \tag{218}$$

i.e., the length scale up to which ions move independently from each other is just the Debye screening length, Eq. 205. For larger length scales an area $\kappa^{-3}$ that happens to carry a positive excess charge is typically surrounded by regions with negative excess charge as schematically indicated in Fig. 21. The interaction energy of two such neighboring, oppositely charged regions is on the order of $-k_B T$ as follows directly from Eq. 217. Thus we expect that the fluctuations in the charge distribution lead to a contribution to the free energy density that scales like $-k_B T/\kappa^3$. This is indeed what we found from the exact DH treatment, Eq. 215.

The DH equation can be solved analytically for various geometries. We present here the solutions for three standard geometries: a plane, a line and a charged ball. The DH equation for a plane of charge density $\sigma$ is given by

$$\left(-\frac{\partial^2}{\partial z^2} + \kappa^2\right) \Phi = 4\pi l_B \sigma \delta(z). \tag{219}$$

It is straightforward to check that this is solved by the potential

$$\Phi(z) = 4\pi l_B \sigma \kappa^{-1} e^{-\kappa z} \tag{220}$$

for $z \geq 0$ and $\Phi(z) = 0$ for $z < 0$. A corresponding DH equation in cylindrical symmetry for a charged line of line charge density $b^{-1}$ leads to the potential

$$\Phi(r) = -\frac{2l_B}{b} K_0(\kappa r) \approx \begin{cases} \frac{2l_B}{b} \ln \kappa r & \text{for } \kappa r \ll 1 \\ -\frac{l_B}{b}\sqrt{\frac{2\pi}{\kappa r}} e^{-\kappa r} & \text{for } \kappa r \gg 1. \end{cases} \tag{221}$$

The function $K_0$ is a *modified Bessel function* whose asymptotic behavior for small and large arguments has been used on the rhs of Eq. 221 to predict the

potential close to and far from the charged line. The short-distance behavior is identical to the one of a naked rod, Eq. 194, for larger distances the line charge is screened as $e^{-\kappa r}/\sqrt{r}$ (up to logarithmic corrections). Finally, for a charged sphere of radius $R$ and charge $Z$ one finds for $r > R$ the potential

$$\Phi(r) = \frac{l_B Z}{1+\kappa R} \frac{e^{-\kappa(r-R)}}{r}. \tag{222}$$

As for a point charge the potential decays proportional to $e^{-\kappa r}/r$. Here, however, for a sphere larger than the screening length, $\kappa R > 1$, the full charge can never be seen, not even close to its surface, since it is distributed in a volume larger than the screening length. $Z$ is then effectively reduced to $Z/(\kappa R)$.

The above given three potentials are not only exact solutions to the DH equation but also excellent approximations to the PB equation if the potential is everywhere much smaller than one, $\Phi \ll 1$. For a line charge this condition requires $l_B \ll b$, i.e., the Manning parameter $\xi$ needs to be much smaller than one. Hence DH theory works well if we do not have Manning condensation. In other words, counterion condensation is just a physical manifestation of the nonlinearity of the PB equation. For spheres the situation is similar. Assuming a sufficiently small sphere so that $\kappa R < 1$, the DH approximation works well if $l_B Z/R \ll 1$, see Eq. 222. This condition is fulfilled if the sphere charge is much smaller than the charge $Z^*$, Eq. 201, the value to which a highly charged sphere would be renormalized. In other words, DH can be used for weakly charged spheres that do not have charge renormalization.

But what can one do if surface charge densities are so high that $\Phi$ becomes larger than unity? Does one necessarily have to deal with the difficulties of nonlinear PB theory or can one somehow combine the insights into counterion condensation and DH theory to construct something that can be handled more easily? That this is indeed possible has been demonstrated by Alexander and coworkers (1984). The idea is that the nonlinearities of the PB equation cause the *charge renormalization* of highly charged surfaces. As a result the potential slightly away from such a surface is so small that DB theory can be used, but a DH theory with a properly reduced surface charge. Consider, for instance, a sphere with $\kappa R < 1$. If the sphere is weakly charged we can simply use Eq. 222. If the sphere is highly charged the nonlinearities of the PB theory predict a layer of condensed counterions of thickness $\lambda$ that effectively reduces the sphere charge $Z$ to a smaller value $Z^*$ as estimated in Eq. 201. We thus expect that the potential sufficiently away from the sphere's surface is given by

$$\Phi_{Z^*}(r) = l_B Z^* \frac{e^{-\kappa(r-R)}}{r}. \tag{223}$$

Note, however, that Eq. 201 is just a rough estimate of $Z^*$ based on an argument where the space around the sphere is artificially divided into two zones.

We are now in the position to give $Z^*$ a precise meaning by requiring that the renormalized DH solution $\Phi_{Z^*}$ and the exact PB solution $\Phi_{\mathrm{PB}}$ – that is here only known numerically – match asymptotically for large distances

$$\lim_{r\to\infty} \Phi_{Z^*}(r) = \lim_{r\to\infty} \Phi_{\mathrm{PB}}(r). \tag{224}$$

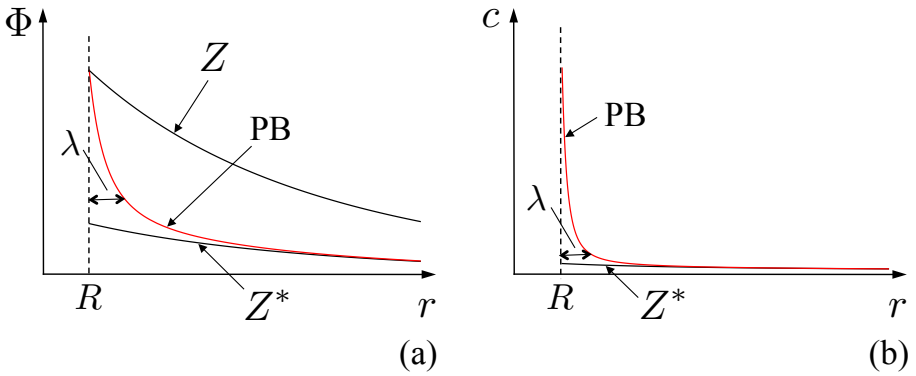



Figure 22: (a) Schematic sketch of the potential around a charged sphere for the full solution (PB), the DH solution ($Z$) and the DH solution with renormalized charge ($Z^*$). (b) Resulting counterion density for the PB solution and for the DH solution with renormalized charge. At large distances the densities are the same but closeby the sphere PB predicts a dense layer of condensed counterions.

That Eq. 224 has a precise mathematical meaning follows from two facts: (1) due to the symmetry of the problem the electrical field is radially symmetric and (2) the potential decays to zero away from the sphere. Therefore the potential must asymptotically look like the DH solution of a charged sphere. In Fig. 22(a) we sketch schematically the potential $\Phi(r)$ for the three solutions around a highly charged sphere: the full PB solution, the DH solution with the bare charge $Z$ and the DH solution with the renormalized charge $Z^*$. For a non-renormalized charge the DH solution overestimates the potential at large distances whereas the renormalized DH solution matches asymptotically the full PB solution. The resulting counterion density $c(r) \sim e^{\Phi(r)}$ for the full PB solution and the renormalized DH solution is depicted in Fig. 22(b).

You might be worried that all the details of the PB theory are lost since in this simple procedure everything is lumped together in one number, the renormalized charge. It is true that renormalized DH theory can only describe the electrostatics beyond the Gouy-Chapman length. It has nothing to say about the microscopic details inside the double layer. One can, however, argue that one does not really want to know about those microscopic details anyway. As a concrete example let us consider again DNA that has 2 elementary charges per $0.33nm$ and a radius of $R = 1nm$. This leads to the Gouy-Chapman length

$$\lambda = \frac{\sigma^{-1}}{2\pi l_B} = \frac{0.33nm \times R}{2 \times 0.7nm} \approx 0.24nm. \tag{225}$$

Up to now we assumed that the DNA charges are homogenously smeared out. In reality the DNA surface area per phosphate charge is given by

$$A = \frac{2\pi R \times 0.33nm}{2} \approx 1nm^2. \tag{226}$$

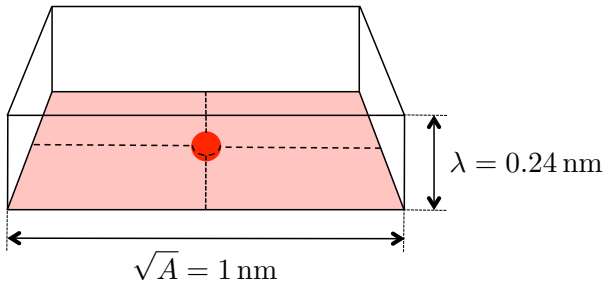



Figure 23: The area $A$ per charged group on a DNA chain is around $1nm^2$ but the Gouy-Chapman length $\lambda$ of a homogeneously charged surface with the same surface charge density is only $0.24nm$.

In other words, the layer of condensed counterions per surface charge is much thinner than it is wide. We must thus expect that the details of the charge distribution, namely its graininess has an effect on the counterion condensation. Smearing out the surface charges might create huge errors, e.g., in the value of the renormalized charge. It is, however, difficult to estimate the size of this error since the PB theory is extremely nonlinear close to the surface.

In principle it is, of course, possible to numerically solve the PB equation for any distribution of surface charges, but one has to ask oneself how meaningful that is. Typical ion radii are of the order of the $\lambda$-value of DNA and might have an effect that is again hard to determine due to the inherent nonlinearity of PB theory. And finally, there is yet another effect that we have brushed under the carpet: the difference in the dielectric constants between the inside of a macromolecule and the surrounding water. Since electrical field lines try to avoid regions of low dielectricity, e.g. the inside of a protein, ions feel an effective repulsion from such a region. In standard electrostatics such effects can be modelled via the introduction of so-called *image charges*, virtual charges that "live" inside regions of low dielectricity and repel real ions nearby. Again this is an effect where microscopic details matter and that can hardly be properly estimated. All what we can say is that all these effects act together in effectively reducing the charge densities of highly charged surfaces.

### Breakdown of mean-field theory

When discussing PB theory and its linearized version, DH theory, we might have given the impression that these theories always work in one way or another. We noted that the strong non-linearities close to highly charged surfaces are somewhat problematic but claimed that proper charge renormalization will always fix that problem. However, as we shall see now, electrostatics is not always as simple as that. Let us go back to the problem of two equally charged surfaces. PB theory predicts that two such surfaces repel, see the two expressions for the disjoining pressure at short and large separations, Eqs. 191 and 192. However,

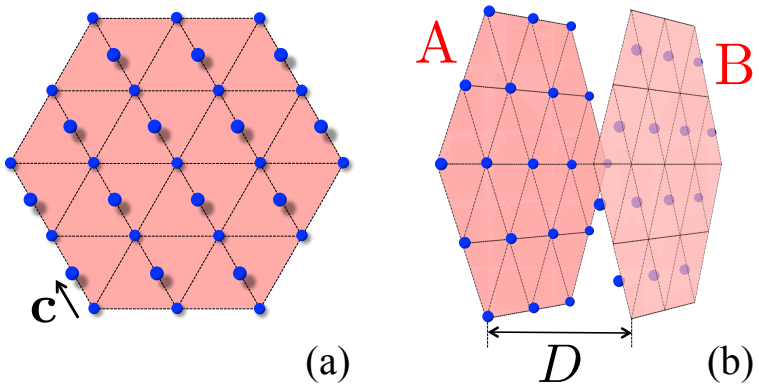



Figure 24: Two Wigner crystals formed by condensed counterions induce an attraction between two equally charged planes: (a) top view indicating the displacement vector **c** that leads to maximal attraction and (b) side view.

in many experiments it has been observed that equally charged objects attract, an effect that – as one can show strictly mathematically – can never be produced by PB theory. In other words, PB theory does sometimes not even get the sign of the force right. A well-known example is DNA. Under the right conditions a DNA molecule can condense onto itself. Such a condensed DNA molecule forms typically a toroid, thereby avoiding in its middle a region of too high curvature. How is it possible that a highly charged molecule like DNA attracts itself? In fact, this never happens inside monovalent salt solutions but when a sufficient amount of trivalent ions or ions of even higher valency is added, such a collapse is typically observed. It can be shown that the meanfield approximation becomes less and less accurate with increasing ion valency. We are lucky that monovalent ion charges are small enough that PB theory can be applied. In fact, one can go much further and not just worry about the applicability of that theory: if the smallest charge unit would be e.g. $4e$ instead of $e$, everything would glue together and there would be simply no life possible.

To come up with a very clean theory that describes the origins of this attraction is not straightforward. We give here a simple argument that goes back to Rouzina and Bloomfield (1996). Again we study the interaction between two identically charged surfaces with their counterions. We assume monovalent counterions but lower the temperature to zero, i.e., we study the ground state of the system. This is, of course, rather academic since water freezes long before but what we are aiming at is just a basic understanding of the principle. According to the so-called *Earnshaw's theorem* any electrostatic system collapses at a sufficiently low temperature. Two surfaces with their counterions should thus stuck on top of each other, $D = 0$, for zero temperature. We shall see that the two surfaces indeed attract in that case.

Let us start by first considering a single charged plane. For $T \to 0$ its Gouy-Chapman length goes to zero, $\lambda \to 0$, since the Bjerrum length goes to infinity,

$l_B \to \infty$. This means that all the counterions sit on the surface. In order to minimize their mutual repulsion, they form a two-dimensional triangular so-called *Wigner crystal* as depicted in Fig. 24. If we have now two such surfaces sufficiently far apart, then the counterions at both surfaces form such patterns independent from each other. When the two surfaces come closer, the counterions lower the electrostatic energy further by shifting their two Wigner crystals with respect to each other by a vector $\mathbf{c}$ as indicated Fig. 24(a). That way an ion in plane B is located above an ion-free area in plane A, namely above the center of a parallelogram with A-ions in its corners. In other words, the relative position of the two planes is shifted with respect to each other by half a lattice constant, so that the two Wigner crystals are out-of-register.

A counterion sitting on one plane, say plane A, feels then the following dimensionless potential resulting from the interaction with plane B and its counterions:

$$\Phi(D) = l_B \sum_l \frac{1}{\sqrt{|\mathbf{R}_l + \mathbf{c}|^2 + D^2}} - l_B \sigma \int \frac{d^2\mathbf{r}}{\sqrt{\mathbf{r}^2 + D^2}}. \tag{227}$$

The first term on the rhs describes the repulsion from the counterions condensed on surface B that are located at positions $\mathbf{R}_l + \mathbf{c}$ with $\mathbf{c}$ denoting the displacement vector between the two planes (both, $\mathbf{R}_l$ and $\mathbf{c}$, are in-plane vectors). The second term accounts for the attraction of the counterion to the homogeneous surface charge on plane B. Further terms do not appear in Eq. 227 since the attraction of the fixed charge of plane A to ions in plane B is exactly cancelled by the repulsion from the fixed charge of plane B. From Eq. 227 follows directly the pressure between the two surfaces:

$$\frac{\Pi(D)}{k_B T} = -\sigma \frac{\partial}{\partial D} \Phi(D) \approx -8\pi\sigma^2 l_B e^{-\frac{2\pi}{3^{1/4}}\sqrt{\sigma} D}. \tag{228}$$

This formula, derived in the Appendix, is accurate for distances $D$ much larger than the counterion spacing $\sim 1/\sqrt{\sigma}$. We thus find an attraction with a decay length proportional to the counterion-counterion spacing.

What is the condition that needs to be fulfilled to have attraction between equally charged surfaces? Above we argued that PB theory is not useful anymore if the Gouy-Chapman length becomes shorter than the distance between fixed charges on the surface, see Fig. 23. Here we use a similar argument, but this time we focus on the counterions in order to estimate when the alternative theory of correlated counterions becomes reasonable. If the counterions have valency $Z$, then the height up to which half of the counterions are found is $\lambda/Z$. On the other hand, the spacing $a$ between the counterions sitting in a Wigner crystal as shown in Fig. 24 is given by $\sqrt{3/2}a^2 = Z/\sigma$. The typical lateral distance between counterions is larger than the height of the counterion cloud if $a > \lambda/Z$. This leads to the condition

$$Z^3 > \sqrt{\frac{3}{2}} \frac{1}{4\pi^2 l_B^2 \sigma}. \tag{229}$$

From this follows that the cloud is essentially two-dimenional for large enough counterion valencies (note the cubic dependence) and for large enough surface charge densities. Remarkably, when condition 229 is fulfilled, one finds that – up to numerical prefactors – the spacing between counterions fulfills $a < Z^2 l_B$, i.e., the neighboring ions feel a mutual repulsion on the order of or larger than $k_B T$. Even though this is by far not strong enough to induce their ordering into a perfect Wigner crystal, the ions are correlated to some extent and can induce the attraction between the charged surfaces. For DNA one has $\sigma = 1nm^{-2}$ and condition 229 reads $Z^3 > 0.06$ or $Z > 0.4$. This seems to suggest that monovalent ions are already strong enough to cause attraction but the argument is evidently too simple to give a reliable quantitative estimate. In reality, ions with $Z = 3$ or larger cause attraction between DNA double helices.

**Appendix: Interaction between two charged plates at zero temperature**

Here we derive the pressure between two equally charged plates at zero temperature, Eq. 228. We begin by rewriting the sum over $l$ in Eq. 227:

$$I = l_B \sum_l \frac{1}{\sqrt{|\mathbf{R}_l + \mathbf{c}|^2 + D^2}} = l_B \int d\mathbf{r} \sum_l \frac{\delta(\mathbf{r} - \mathbf{R}_l)}{\sqrt{|\mathbf{r} + \mathbf{c}|^2 + D^2}}. \tag{230}$$

The $XY$-positions of the ions on one surface form a lattice given by the set of 2D vectors $\mathbf{R}_l$, whereas the ions on the other surface are shifted to the positions $\mathbf{R}_l + \mathbf{c}$. The integral introduced above is thus two dimensional. This leads to a sum over delta-functions that is a periodic function in two dimensions. Any such periodic function $f(\mathbf{r})$ can be written in the form of a *plane wave expansion*, a 2D version of the Fourier expansion introduced in Appendix 4.2. Here

$$f(\mathbf{r}) = \sum_l \delta(\mathbf{r} - \mathbf{R}_l) = \sum_{\mathbf{k}} f_{\mathbf{k}} e^{i\mathbf{kr}} \tag{231}$$

where the summation goes over all vectors $\mathbf{k}$ of the reciprocal lattice that is defined further below. The $f_{\mathbf{k}}$ are the Fourier coefficients that are given by

$$f_{\mathbf{k}} = \sigma \int_C e^{-i\mathbf{kr}} f(\mathbf{r})\, d\mathbf{r} \tag{232}$$

with $C$ denoting a *primitive cell* of the direct lattice, a minimum repeat unit containing one ion. Here $f_{\mathbf{k}} = \sigma$ and hence

$$I = l_B \sigma \sum_{\mathbf{k}} \int \frac{e^{i\mathbf{kr}}}{\sqrt{|\mathbf{r} + \mathbf{c}|^2 + D^2}} d\mathbf{r} = l_B \sigma \sum_{\mathbf{k}} e^{-i\mathbf{kc}} \int \frac{e^{i\mathbf{kr}}}{\sqrt{r^2 + D^2}} d\mathbf{r}. \tag{233}$$

We exchanged here the order of summation and integration; substituting $\mathbf{r} + \mathbf{c}$ by $\mathbf{r}$ yields the phase factor $e^{-i\mathbf{kc}}$. Note that the term with $\mathbf{k} = \mathbf{0}$ in the

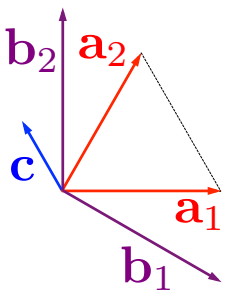



Figure 25: Primitive vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ that span the triangular lattice. Also indicated are the primitive vectors of the reciprocal lattice, $\mathbf{b}_1$ and $\mathbf{b}_2$, and the shiftvector $\mathbf{c}$ for maximal attraction between the two surfaces.

summuation corresponds exactly to the second term in Eq. 227. Hence we can write the dimensionless potential as

$$\Phi(D) = l_B\sigma \sum_{\mathbf{k}\neq\mathbf{0}} e^{-i\mathbf{kc}} \int \frac{e^{i\mathbf{kr}}}{\sqrt{r^2+D^2}} d\mathbf{r}. \tag{234}$$

Using Eq. 228 we calculate the pressure from the potential by differentiation

$$\begin{aligned} \frac{\Pi(D)}{k_BT} &= l_B\sigma^2 D \sum_{\mathbf{k}\neq\mathbf{0}} e^{-i\mathbf{kc}} \int d\mathbf{r} \frac{e^{i\mathbf{kr}}}{(r^2+D^2)^{3/2}} \\ &= l_B\sigma^2 D \sum_{\mathbf{k}\neq\mathbf{0}} e^{-i\mathbf{kc}} \int_0^{2\pi} d\phi \int_0^{\infty} dr \frac{re^{ikr\cos\phi}}{(r^2+D^2)^{3/2}}. \end{aligned}$$

We introduced here polar coordinates where $\phi$ denotes the angle between the respective $\mathbf{k}$-vector and $\mathbf{r}$. The double integral can be calculated analytically (first integrate over $\phi$, then over $r$) and yields $(2\pi/D)\, e^{-kD}$ with $k = |\mathbf{k}|$. This leads to

$$\frac{\Pi(D)}{k_BT} = 2\pi l_B\sigma^2 \sum_{\mathbf{k}\neq\mathbf{0}} e^{-i\mathbf{kc}} e^{-kD}. \tag{235}$$

We have thus expressed the interaction between the two surfaces as an infinite sum of exponentials. In the following we are interested in the leading terms of this sum for large distances. These will be the terms with the smallest value of $k$.

The ground state of a single plane is given by counterions that form a triangular Wigner crystal. We expect that each surface with its counterions still remains in this triangular groundstate as long as $D$ is much larger than the spacing between counterions within their planes. More specifically, the positions of

the counterions in one lattice are given by $n_1\mathbf{a}_1 + n_2\mathbf{a}_2$ with $n_i = 0, \pm 1, \pm 2, ...$, an example of a so-called *Bravais lattice*. The vectors $\mathbf{a}_i$ that span the lattice, the so-called *primitive vectors*, are given by

$$\mathbf{a}_1 = a\mathbf{e}_x,\ \mathbf{a}_2 = \frac{a}{2}\mathbf{e}_x + \frac{\sqrt{3}a}{2}\mathbf{e}_y \tag{236}$$

and are indicated in Fig. 25. The lattice spacing $a$ has to be chosen such to match the charge density $\sigma$, leading to $a = 2/\left(3^{1/4}\sigma^{1/2}\right)$. The *reciprocal lattice*, the set of all vectors $\mathbf{k}$ for which $e^{i\mathbf{kR}} = 1$ for all $\mathbf{R}$ in the Bravais lattice, is given by $\mathbf{k} = k_1\mathbf{b}_1 + k_2\mathbf{b}_2$, $k_i = 0, \pm 1, \pm 2, ...$, with

$$\mathbf{b}_1 = \frac{2\pi}{a}\left(\mathbf{e}_x - \frac{1}{\sqrt{3}}\mathbf{e}_y\right),\ \mathbf{b}_2 = \frac{4\pi}{\sqrt{3}a}\mathbf{e}_y. \tag{237}$$

The primitive vectors of the reciprocal lattice fulfill $\mathbf{b}_i\mathbf{a}_j = 2\pi\delta_{ij}$, see also Fig. 25. For large distances the leading terms in Eq. 235 are the ones with the smallest value of $k$, namely $(k_1, k_2) = (\pm 1, 0)$ and $(k_1, k_2) = (0, \pm 1)$. For distances $D$ with $D \gg a$ all higher order terms are negligible. The large distance pressure is thus to a very good approximation given by

$$\frac{\Pi(D)}{k_B T} \approx 4\pi\sigma^2 l_B \left(\cos(\mathbf{b}_1\mathbf{c}) + \cos(\mathbf{b}_2\mathbf{c})\right) e^{-\frac{4\pi}{\sqrt{3}a}D}. \tag{238}$$

For a vanishing length of $\mathbf{c}$ counterions of one surface are just on top of counterions of the other surface so that the two surfaces repel each other. One finds then $\cos(\mathbf{b}_1\mathbf{c}) + \cos(\mathbf{b}_2\mathbf{c}) = 2$ leading to maximal repulsion. If we, however, allow one of the plates with its counterions to move in the $XY$-plane relative to the other at a fixed value of $D$, the system can lower its energy. It reaches the groundstate when $\cos(\mathbf{b}_1\mathbf{c}) + \cos(\mathbf{b}_2\mathbf{c}) = -2$. This can be achieved by choosing e.g. the shift c such that $\mathbf{b}_1\mathbf{c} = -\pi$ and $\mathbf{b}_2\mathbf{c} = \pi$. This is achieved for

$$\mathbf{c} = -\frac{a}{4}\mathbf{e}_x + \frac{\sqrt{3}a}{4}\mathbf{e}_y, \tag{239}$$

as shown in Figs. 24 and 25. In that case we find Eq. 228 from Eq. 238.