Contents

1 Introduction (4)

1.1 Introduction: what it is all about?

Statistical physics might be one of the most conceptually difficult subjects you will encounter. Sorry about that! Don't worry though, we will (again) barely scratch the subject. The reason I think it is a difficult subject is that it touches and connects many scientific fields and each of those cases has its own specialities and nuances that are hard to grasp.

Notice that we said 'scientific fields' and not 'physical fields'. The topics we are going to discuss are far overreaching the fields of physics. We have chemists discussing the Gibbs energy, biologists talking about the effect of temperature, informaticians considering entropy and so on. In his famous lecture 'What is life?' Schrödinger talks about negative entropy as the fuel needed by the life.

Consider a simple example from signal analysis. You want to communicate a message – that means to transfer information to someone else. Is there any limit to this process or can you pack the information as densely as you please? You cannot as there exist thermal noise that can dominate over your message. How to overcome this issue? You can 'speak louder', that means spend more energy to communicate. Now this triangle seems familiar, doesn't it? Information (entropy), temperature, energy.

The goal of this course will be mostly to understand the physical aspects of statistical physics. However, now and then we will devote some time to pay homage to the vast of connections that stems from it.

1.1.1 Energy, entropy and temperature

At this point of your studies, you have perhaps noticed already: energy is everywhere, from the first course in mechanics to quantum field theory and everything between. Not only that, you perhaps understand why is this so, the reason is the famous Nöther theorem. As long as the laws of physics do not depend on time a quantity (that we call the energy) is conserved. The laws of micro and macro world seem to be both independent of time and therefore there is a quantity connecting the two worlds which share very few things otherwise. This is the reason why the energy plays a crucial role in statistical physics, perhaps even more than in any areas of physics and we will have to pay great attention to many details related to it.

A novel thing that appeared in study of microworld is called the entropy. Surprisingly, it appeared before we realised it was due to microscopic behaviour of matter constituents. So, what is the entropy? It describes the complexity of the microscopic world. How much we don't know when we know only macroscopic aspects of the system we study? It is the measure of missing information (and not only missing information, wait for chapter about the information theory). Some people claim that the entropy is the most important quantity in physics. It seems to be related to the arrow of time and dictates how will the final stage of our universe look like. Surprisingly, the entropy was discovered in a somewhat different context (as Clausius entropy) and was only later realised to be also related to the usual definition (and another definitions followed).

So, what is the relation between those two important concepts? Recall the Gibbs free energy $F = E - TS$ the system tries to minimize. This can be done either by minimizing the energy (which usually leads to very ordered system) or by maximizing the entropy (which leads to disordered system). Which of the forces dominate? That depends on the temperature. This quantity has been historically derived in a different way which leaves many with the impression that it is the measure of microscopic movement of particles while it is so much more. It is however the thing that relates energy with the entropy. A surprising aspect of this more general definition is that under very special circumstances the temperature can be negative.

1.1.2 Liouville flow and ergodicity

References: Reichl A2.3, C Statistics in general is about handling the uknown. Statistical physics is therefore about handling the uknown in physical systems. As it turns out, the world contains way too much information. We use physics to make predictions about the world. Given we know the state of the system right now and we know the laws it follows then we can extrapolate into the future. One of the problems is that the world is complicated – it contains way to much information. We don't care about most of it, that's the difference between microstates and macrostates. The first hint is that in statistical physics we are often interested in stationary states. There might be some small temporal fluctuations but they are averaged out.

$$\langle A\rangle_T = \lim_{T\to\infty}\frac{1}{T}\int_{t_0}^{T+t_0} A(t)dt. \tag{1}$$

This is on a paper, in a real experiment measuring the temperature in the room for a couple of seconds is sufficient! Ok, how is this helpful?

We use the notion of phase space. For N particles this space is $6N$ dimensional. If we know the exact position of a state $\mathbf{X}^N(\mathbf{p}^N, \mathbf{x}^N)$ we can evolve it using Hamiltonian equations. Sometimes, we are not so sure, we can only know a probability distribution $\rho(\mathbf{X}^N, t)$ that has $\int \rho(\mathbf{X}^N, t)d\mathbf{X}^N = 1$. Probability to find the system in a small phase region of volume V_0 is

$$P(R) = \int_R \rho(\mathbf{X}^N, t)d\mathbf{X}^N. \tag{2}$$

Now the system does move in the phase space which is described by $\dot{\mathbf{X}}^N = (\dot{\mathbf{p}}^N, \dot{\mathbf{x}}^N)$. By this the probability distribution flows like a liquid! And from Hamiltonian equations we know that this liquid is incompressible, $\nabla_{\mathbf{X}^N}\cdot\dot{\mathbf{X}}^N = 0$. One can then write down the e.o.m for $\rho(\mathbf{X}^N, t)$ as

$$i\frac{\partial\rho(\mathbf{X}^N,t)}{\partial t} = \hat{L}^N\rho(\mathbf{X}^N,t), \tag{3}$$

where $\hat{L}^N$ is the Liouville operator ($-i$ times the Poisson bracket operator), the equation is called the Liouville equation and its (formal) solution is

$$\rho(\mathbf{X}^N,t) = e^{-i\hat{L}^N t}\rho(\mathbf{X}^N,t). \tag{4}$$

So there is a good and well defined way how to handle this uncertainty. However, this still contains more detail that we are interested in. There is another way how to handle this.

A system that follows Hamiltonian equations may have many conserved quantities, such as the total energy. That means the system cannot move through entire phase space but only through some surface it it, let us denote it S_E. We will call the system ergodic if during the Hamilton flow it passes through a small neighbourhood of every point in S_E. We can define

$$\langle A\rangle_S = \frac{1}{\Sigma(E)}\int\limits_{S_E} A(\mathbf{X}^N)dS_E. \tag{5}$$

The ergodic theorem states that for an ergodic system $\langle A\rangle_S = \langle A\rangle_T$. This means we can obtain estimates about temporal averages (which is what we can measure) by calculation phase-space averages. And this is what we can done by figuring out how does the allowed phase space look like.

A necessary condition is that there can be no other constrains splitting the phase space.

1.1.3 Ensembles

The tricky thing, obviously, is to decide how to accurately describe 'all possible states of the system'. There are basically three common choices. The first is called the microcanonical ensemble – consider all possible states of the given energy, all of them with equal probability (indifference principle) which also maximizes the entropy under the fixed conditions.

The second choice is that we don't fix the total energy but only the number of constituents of the system. This is useful in the cases where the system is in thermal contact with the environment, that means it can exchange heat. Recall that due to thermal radiation it is impossible to have perfectly isolated system. The only way of having one is to consider everything to be a single system – then it has nothing to be isolated from. So our universe can (?) be considered a thermally isolated system with fixed (?) energy and we can describe it using the microcanonical ensemble. But what if we don't care about the entire universe but only about the balloon filled with gas in from of us? Than we can split the universe into the balloon part and everything else (which we call the reservoir). Now this everything else is of course very large but we still don't assume it to be unaffected by our system. When our system has more energy then the reservoir has less of it and its microcanonical ensemble description changes – which leads to different probabilities of different (energy) states of our tiny system. This leads to the Boltzman distribution.

The third choice is that our system can exchange both heat and particles, so called grand canonical ensemble. This isn't only useful when our balloon has holes in it and particles can move in and out. For example, some particles can just pop into existence if we give them the chance, for example photons in a box or "holes and anti-holes" it an atomic lattice. This shows that when we say particles we actually mean something more general – something that carries new degrees of freedom.

These are some restrictions but they are not overly restricted. Usually the task of performing a summation over all possible states is the only thing we need to do but it still is a bit too much so we need to simplify the model under consideration a bit. We will return to those problems in a couple of lectures, but before that, we will (re)visit some important topics in the probability theory. If time permits, we will try to finish the semester with a discussion of limits of limits of our knowledge. These are especially easy to put down in the theory of information and since statistical physics are so related to it, we can borrow some lessons.

1.2 Probability theory I

1.2.1 Generating random numbers

References: Cerny 2.5 and also for the final example `https://mathworld.wolfram.com/SpherePointPicking.html`
Usually we work this way: from a dataset we want to compute things like moments of the distribution to reproduce the underlying distribution. What if we need to do this in reverse: begin with a random distribution and generate a set of random numbers (or something derived from them) that has the same statistical properties. Even though most codes know to generate random numbers with normal probability distribution (and often some other common distributions), we will assume it only can do the uniform random distribution in the interval $0 \leq x \leq 1$.

It is easy to produce a different uniform for it, for example if we want random umbers in (a, b) interval one just needs to use $y = a+(b-a)*x$. That was easy, let us try something different. Assume we want to have a probability distribution function $f(x)$ and also the cumulative distribution function $F(x) = \int\limits_{-\infty}^{x} f(x)dx$. Now notice an interesting thing. We can generate numbers between 0 and 1 (as we assumed), but the cumulative function is also bounded the same way: $0 \leq F(x) \leq 1$. How can we utilise this? (Draw a simple cumulative function to figure this out.)

The thing to do is to invert the cumulative distribution function: $y = F^{-1}(x)$ with $0 \leq x \leq 1$ will have the distribution $f(x)$.

A formal proof goes like this

$$\langle q \rangle = \int\limits_{-\infty}^{\infty} q(p)f(p)dp = \int\limits_{0}^{1} q(F^{-1}(x))f(F^{-1}(x))\frac{1}{f(F^{-1}(x))}dx = \int\limits_{0}^{1} q(F^{-1}(x))dx. \tag{6}$$

Example: How to generate uniformly random points on a sphere if we can generate uniformly random numbers $0 \leq r_i \leq 1$? The naive choice to generate random $(\cos r_1 \sin r_2, \sin r_1 \sin r_2, \cos r_2)$ does not work. We have to take deformed geometry of the sphere into account. Every small surface element

dS needs to have the same probability $\frac{1}{4\pi}dS = f(\theta,\phi)d\theta d\phi$. Now since $dS = \sin(\theta)d\theta d\phi$ we have $f(\theta,\phi) = \frac{1}{4\pi}\sin\theta$. Marginalizing w.r.t. ϕ gives $f(\theta) = \sin(\theta)/2$. The c.d.f is $F(\phi) = \frac{1}{2}(1-\cos\theta)$. We need to invert this, which can be done as $F^{-1}(y) = \arccos(1-2y)$. Now y can be generated randomly using r_i.

1.2.2 Characteristic functions and cumulants

References: Reichl A 1.2.3

The characteristic function is defined as

$$\Phi_x(k) = \langle e^{ikx}\rangle = \int\limits_{-\infty}^{\infty} dx e^{ikx} f_x(x) = \sum_{n=0}^{\infty} \frac{(ik)^n \langle x^n\rangle}{n!}, \tag{7}$$

it is completely fixed by the moments and can be used to obtain distribution back from the moments by

$$f_x(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} dk e^{-ikx}\Phi_x(k). \tag{8}$$

The moments can be obtained by taking derivatives

$$\langle x^n\rangle = \lim_{k\to 0}(-i)^n \frac{d^n\Phi_X(n)}{dk^n}. \tag{9}$$

Characteristic functions has some important properties (closed under multiplication): $\Phi_X(0) = 1, |\Phi_X(k)| < 1$ and $\Phi_X(-k) = \Phi_X^*(k)$. The characteristic function can be used to defined cumulants (which can be obtained from the moments and vice versa):

$$\ln\Phi_X(k) = \sum_{i=1}^{\infty} \frac{(ik)^n}{n!} C_n(X). \tag{10}$$

Let us list the first couple of them

$$\begin{aligned} C_1(X) &= \langle x\rangle, \qquad (11)\\ C_2(X) &= \langle x^2\rangle - \langle x\rangle^2,\\ C_3(X) &= \langle x^3\rangle - 3\langle x\rangle\langle x^2\rangle + 2\langle x\rangle^3,\\ C_4(X) &= \langle x^4\rangle - 3\langle x^2\rangle^2 - 4\langle x\rangle\langle x^3\rangle + 12\langle x\rangle^2\langle x^2\rangle - 6\langle x\rangle^4. \end{aligned}$$

To appreciate the existence of cumulants, let us note that the first five moments of normal distribution are

$$\langle x^n\rangle = \big(\mu, \mu^2+\sigma^2, \mu(\mu^2+3\sigma^2), \mu^4+6\mu^2\sigma^2+3\sigma^4, \mu(\mu^4+10\mu^2\sigma^2+15\sigma^2)\big) \tag{12}$$

while the first five cumulants are

$$C_n = \big(\mu, \sigma^2, 0, 0, 0\big) \tag{13}$$

(and it is easy to guess the rest of the list).

(You can also use characteristic function to understand the Wick's theorem $\langle x_{k_1}...x_{k_l}\rangle = \sum_{\text{all possible pairings}} \langle x_{k_{P_1}} x_{k_{P_2}}\rangle ... \langle x_{k_{P_{l-1}}} x_{k_{P_l}}\rangle$. Details can be found here `http://ckw.phys.ncku.edu.tw/public/pub/Notes/PathIntegral/Zinn-Justin/01._GaussianIntegrals/1.2._GaussianExpectationValues-Wick'sTheorem.pdf`.)

1.2.3 Central limit theorem

References: Reichl A 1.4 Many things follow the normal probability distribution. This can be easily proven using the characteristic function. Consider a random number $Y_N = (1/N)(X_1 + ... + X_n) - \langle x\rangle$ where X are random numbers of unknown distribution. Can we tell something about the fluctuations of Y? Let us denote $Z_i = (1/N)(X_i - \langle x\rangle)$ so $Y_N = \sum_{i=1}^{N} Z_i$. The characteristic function for Z_i can be written as

$$\Phi_Z(k,N) = \int_{-\infty}^{\infty} e^{i(k/N)(x-\langle x\rangle))} P_X(x)dx = 1 - \frac{1}{2}\frac{k^2}{N^2}\sigma_X^2 + ... \tag{14}$$

The characteristic function of Y_N is

$$\Phi_{Y_N}(k) = \left(1 - \frac{1}{2}\frac{k^2}{N^2}\sigma_x^2 + ...\right)^N \to \exp\left(-\frac{k^2\sigma_X^2}{2N}\right) \tag{15}$$

in the $N \to \infty$ limit. From this we the distribution for Y_N

$$f_{Y_N}(y) \to \frac{1}{2\pi}\int_{-\infty}^{\infty} dk e^{iky} \exp\left(-\frac{k^2\sigma_X^2}{2N}\right) = \sqrt{\frac{N}{2\pi\sigma_x^2}} \exp\left(-\frac{Ny^2}{2\sigma_X^2}\right). \tag{16}$$

1.3 Probability theory II

1.3.1 Bayesian statistics

References (sec. 1-5) `https://arxiv.org/pdf/hep-ph/9512295.pdf`
Assume we have a fair coin and are about the flip it. What are the odds of getting head? Of course 50%. Now imagine I flipped the coin and dropped it, it is now under table and we don't see it. What are the odds of it being head? Now some of you might think it is still 50% and some of you find this question stupid: there is no uncertainty anymore, the coin has been flipped. If you are the first type of person, congratulation, you are a bayesian. If you are the second type, we are going to turn you now.

There are two other usual approaches to probability: combinatorial and frequentistic.

Combinatorial: out of all possible results, how many turn out in a certain way?

Frequentistic: if we repeat this experiment a very huge number of times, in what fraction of them would we get this result?

Bayesian: given my previous experience, what is my degree of belief in this given results? (How to measure degree of belief? Make a bet!)

The problem with the combinatorial is the question if all results are equally probable. The problem with frequentistic is the issue with the coin under table – the result would always be the same, is there any real uncertainty or only the lack of our knowledge? Finetti claimed that: probability does not exist. And that might be correct (depends on how does quantum theory really works). But probability is a good way to model our uncertainty which follows from our lack of knowledge.

Important concept is the conditioned probability. Assume we have a class of children. You can either ask 'what is the probability I'll randomly choose a kid of heigh larger than 160cm?' Another type of question is 'what is the probability I'll randomly choose a boy of heigh larger than 160cm?' In the second case, we have applied a condition: the selected pupil has to be a boy. This is usually denoted as $P(h > 160cm|\text{boy})$. Following the aforementioned references we will use E for event, H for a hypothesis, $\bar{H}$ for it's complement and Ω for every possibility. Trivially $P(E) = P(E|\Omega)$. A trivial thing holds

$$P(E) = P(E \cap H) + P(E \cap \bar{H}). \tag{17}$$

'The probability of detector producing a signal is the probability of producing the signal given an electron bumped into it plus the probability of producing the she signal given an electron didn't bump into it.' Still trivial.

Another trivial observations are

$$P(E \cap H) = P(E|H)P(H) = P(H|E)P(E). \tag{18}$$

which we usually present in the form $P(E|H) = \frac{P(E \cap H)}{P(H)}$ or

$$\frac{P(H|E)}{P(H)} = \frac{P(E|H)}{P(E)}. \tag{19}$$

Usually there is a set of different hypothesis $\bigcup_i H_i = \Omega$. From this we have

$$P(E) = \sum_i P(E|H_i)P(H_i). \tag{20}$$

We can plug this into the previous equation to obtain

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_j P(E|H_j)P(H_j)}. \tag{21}$$

This is the typical form of the Bayes theorem.

Example: You have a particle A detector that is triggered by particle A with 90% success rate. Unfortunately, it is also triggered by a particle B that is in the background with 2%. You know that their densities are in the ratio 9 : 1. When your detector triggers, what is the probability it was really triggered by A? And what is the signal-to-noise ratio (S/N)?

$$\begin{aligned} P(A|T) &= \frac{P(T|A)P(A)}{P(T|A)P(A) + P(T|B)P(B)} \approx 0.84 \\ S/N &= P(A|T)/P(B|T) \approx 5.3. \end{aligned} \tag{22}$$

Or in general:

$$S/N = \frac{P(E|S)}{P(E|N)}\frac{P(E)}{P(N)}. \tag{23}$$

So in noisy conditions $P(N) \gg P(E)$ you need to have a very selective detector $P(E|S) \gg P(E|N)$.

We can cast the theorem in yet another form!

$$\begin{aligned} P(H_i|E, H_0) &= \alpha P(H_i|H_0), \qquad (24)\\ \alpha &= \frac{P(E|H_i, H_0)}{\sum\limits_i P(E|H_i, H_0)P(H_i|H_0)}, \end{aligned}$$

where we have also included previous conditioning H_0 (sort of underlying assumptions about the theory and so on), they won't play any role in the calculations, it is just a reminder for us. Now this is an important formulation of the Bayes theorem, it is the ground for the approach/interpretation of 'learning by experience'. On the R.H.S. we have the *a priori* probability estimate (prior), the L.H.S. is the *a posteriori* probability after gaining new information. α is the coefficient that tells us how to adjust the probabilities. $P(E|H_i, H_0)$ is called the likelihood.

Example: How long does it take you are tossing a coin that has a head on both sides? You initially assume $P(head|fair) = P(tail|fair) = 1/2$. You have a suspicion that the coin might be fake and then $P(head|fake) = 1 - P(tail|fake) = 1$. You also have to assume some probabilities of the coin being a fake, lets say 5%. You now toss the coin N times and see all heads! How to adjust your initial assumptions ($95 - 5$ for the case of fair coin)?

$$\begin{aligned} P(fake|head^N) &= \frac{P(head^N|fake)P(fake)}{P(head^N|fake).P(fake) + P(head^N|fair).P(fair)} \qquad (25)\\ &= \frac{P(fake)}{P(fake) + 2^{-N}P(fair)} \to 1,\\ P(fair|head^N) &= \frac{P(fair^N|fake)P(fair)}{P(head^N|fake).P(fake) + P(head^N|fair).P(fair)}\\ &= \frac{2^{-N}P(fair)}{P(fake) + 2^{-N}P(fair)} \to 0. \end{aligned}$$

Our faith in the coin quickly decays!

We have complained about problems with combinatorial/frequentist approach to probability, now we overlooked the problem with Bayesian one. The tricky thing is that we have to make some initial guesses about the probabilities ($P(fake)$ and $P(fair)$) and this changes our results. A common thing is to assume as little as possible (the indifference principle or the maximum entropy principle). However, as we have seen, the initial probabilities are quickly replaced by our newly gained experience.

This is how hypothesis testing in science should work. One formulate a null hypothesis ('This effect doesn't exist') and a novel hypothesis ('It does exist and leads to this ...') and then we perform the experiment and ask: to what extent are the results consistent with null hypothesis and the novel one? Why is it important? Sometimes we are dealing with a complicated theory, make loads of calculations and then say: 'If my theory is right, we should see this!' But we also have to check if the same effect cannot be reproduced without this theory/effect, maybe only by a pure luck. And if so, how many positive results do we need to have to trust the new theory?

1.3.2 Markov chain

`https://www.stat.auckland.ac.nz/~fewster/325/notes/ch8.pdf`, *Reichl A.2*
Consider a random variable Y that can assume values $y_1, ..., y_N$ and moves between them in discrete time steps $s = 1, ..., \infty$. The probability to be in the state n at step s is denoted $P_n(s)$. Obviously

$$P_n(s+1) = \sum_{m=1}^{N} P_m(s) P(m, s|n, s+1), \tag{26}$$

where $P(m, s|n, q$ is the conditional probability also called the *transition probability.* In a more general form

$$P(n_0, s_0|n, s+1) = \sum_{m=1}^{N} P(n_0, s_0|m, s) P(m, s|n, s+1). \tag{27}$$

In some cases the transition probabilities do not change with time, in that case

$$P_n(s) = \sum_{m=1}^{N} P_m(0)(Q^s)_{m,n}. \tag{28}$$

The transition matrix Q carries all important information about the transition rates and one can solve the equations for any transition process in terms of its left and right eigenvectors and eigenvalues (it is not symmetric usually). We can rewrite the transition equation in the following form:

$$P(n, t+\Delta t) = \sum_{n=1}^{M} P(m, t) P(m, t|n, t+\Delta t). \tag{29}$$

We can derive the e.o.m. as

$$\frac{\partial P(n,t)}{\partial t} = \lim_{\Delta t \to 0} \left(\frac{P(n, t+\Delta t) - P(n,t)}{\Delta t} \right) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \sum_{n=1}^{M} P(m,t) \left(P(m, t|n, t+\Delta t) - \delta_{mn} \right) \tag{30}$$

which we can expand into

$$P(m, t|n, t+\Delta t) = \delta_{mn} \left(1 - \Delta t \sum_{l=1}^{M} w_{m,l} \right) + w_{m,n} \Delta t + ... \tag{31}$$

Here $w_{m,n}$ is called the probability of transition. Using it we have

$$\frac{\partial P(n,t)}{\partial t} = \sum_{m=1}^{M} \left(P(m,t) w_{m,n}(t) - P_1(n,t) w_{n,m}(t) \right) \tag{32}$$

which is called the master equation. Now this one is easy to interpret. From this we can see the condition for a stationary solution which is called the detailed balance

$$P^s(n)w_{n,m} = P^s(m)w_{m,n}, \tag{33}$$

where ^{s} is for stationary: $P^s(n) = \lim_{t\to\infty} P(n,t)$.

1.4 Basics of information theory

`https://arxiv.org/abs/1805.11965` and also `https://www.fil.ion.ucl.ac.uk/~wpenny/course/info.pdf`

1.4.1 Information content

Information is a paradox. It cannot exist without a physical manifestation but it does not depend on a physical manifestation. The same information can have numerous forms. Information is what resolves uncertainty. The smallest amount of information is a single bit – yes or no answer. Claude Shannon was thinking about sending information, what is the information contained in a message?

His first clue was that the information should be (linearly) proportional to the message length. Two messages of equal length carry the same information as one information of double length. Also, the number of possible messages is a list of possible combinations of letters. Now the number of combinations with k letters is k^N and to make this linear in N we need to take the logarithm (any works, base of 2 is preferred in the information theory), that is

$$S \sim \log \Omega, \Omega = k^N. \tag{34}$$

We are on the right track, but we have to consider that not all messages are equally likely. Information in a message should be equal to surprise it makes. For example, imagine 0 stands for 'no fire' and 1 stands for 'fire'. Your alarm keeps sending you 000000000.... and if the next data bit is again 0 you obtain way less information (that means, it is easier to guess) that if you receive 1. This is also true for language. There are 26 letters in English alphabet but if you receive information $spaghet * ti$ with one letter missing, are you really missing any information at all? No, language is very redundant! You might think that the information per letter is $\log_2 26 \approx 4.7$, but actually it is somewhere between $2.5 - 3$. What is the point of redundancy? So it is easier to spot *ertors*!

Ok, imagine we have a long message with two letters appearing with probabilities p_1, p_2. Long enough message with contain NP_i letters of each type and there are

$$\frac{N!}{(p_1 N)!((p_2 N)!} = 2^{NS} \tag{35}$$

such messages. We take this to be the definition of $S = -p_1 \log p_1 - p_2 \log p_2$, the Shannon entropy (he chose this word deliberately). This can be generalised to k letters as

$$S_A = -\sum_{i=1}^{k} p_i \log p_i. \tag{36}$$

(The subscript is for this given probability distribution). This is called the information content of message, or the measure of average surprise the message contains. (If there is only one letter that has $p_1 = 1$, there is no surprise and no information gain.) Maximal information is obtained if $p_i = 1/k$, in that case the entropy per letter is

$$S_A = \sum_{i=1}^{k} (1/k) \log(1/k) = \log k. \tag{37}$$

1.4.2 Optimal coding

Cerny 2.11
Imagine we want to share an information, how to make (average) message as short as possible? Recall the Morse code, the shortest sequences are used to code the most frequently used letters. How could we translate this into two-digit code (without a break). We can use prefix code, that means no coded word is the beginning sequence of any other. For example, if we use 0010 to code the letter A, no other word beginning with this sequence is allowed. Assume we want to code m messages (for example $m = 26$ to code the alphabet or $m = 10$ to code a number) and denote s_i lengths of $i = 1, ..., m$ codes. For the length of all codes we have the Kraft equality

$$\sum_i 2^{-s_i} \leq 1. \tag{38}$$

(How to prove it: think of a tree-diagram where each branch corresponds to 0 or 1. Each point corresponds to a code: 0, 1, 00, ... If we pick a point to represent a prefix-free code, all points following in are to be discarded. To each code assign 2^{s_i}. Take s_m to be the longest code, there are 2^{s_m} points with corresponding length. We are either using all of them, then $\sum_i 2^{-s_i} = 1$ or don't need some of them, then $\sum_i 2^{-s_i} < 1$.

Optimal coding means we will assign shorter codes to more frequent words (given we know their probabilities p_i? This is an optimisation problem:

$$\mathcal{L} = \sum_i p_i s_i - \lambda \left(2^{L_i} - 1\right). \tag{39}$$

Now the solution is $s_i = -\log_2 p_i$. The optimal code length therefore is

$$L = -\sum_i p_i \log p_i. \tag{40}$$

Excercise: Choose and verify coding of two messages that appear with probabilities $p_1 = 0.1, \;\; p_2 = 0.9$ if you can use $s = 1, 4$.

1.4.3 Conditional entropy

Just the existence of a message does not mean that any information is being transferred as communication can be noisy. Assume Alice is sending random variable X that can take values x and Bob receives $Y = y$. There is a joint probability $P_{X,Y}(x_i, y_j)$ that Alice sends x_i and Bob receives y_j. Probability that Bob receives y_j is

$$P_Y(y_j) = \sum_i P_{X,Y}(x_i, y_j). \tag{41}$$

His estimate that Alice sent x_i is given by the conditional probability

$$P_{X|Y}(x_i|y_j) = \frac{P_{X,Y}(x_i, y_j)}{P_Y(y_j)}. \tag{42}$$

The Shannon entropy of this conditional probability is

$$S_{X|Y=y_j} = -\sum_i P_{X|Y}(x_i|y_j) \log(P_{X|Y}(x_i|y_j)). \tag{43}$$

If we average over all possible received messages we obtain

$$\begin{aligned} \sum_j P_Y(y_j) S_{X|Y} &= -\sum_{i,j} P_{X,Y}(x_i, y_j) \log P_{X,Y}(x_i, y_j) + \sum_{i,j} P_{X,Y}(x_i, y_j) \log P_Y(y_j) \\ &= S_{XY} - S_Y. \end{aligned} \tag{44}$$

This is called the conditional entropy $S_{X|Y} = S(X|Y)$, it is the entropy that remains in the probability distribution X once Y is known (the missing information). By construction, it is possitive $S_{XY} - S_Y \geq 0$. A very useful concept is the mutual information, that is the information Alice send minus what was lost in the process:

$$I(X;Y) = S_X - S_{XY} + S_Y. \tag{45}$$

It can be proven to be positive.

1.4.4 Relative entropy

How wrong are we? Suppose we are observing a random variable X with distribution P_X but think it is Q_X. We see s different outcomes with probabilities $q_i, i = 1, ..., s$ and after N experiments we estimate the probability to be

$$P = \prod_i q_i^{p_i N} \frac{N!}{\prod\limits_j (p_j N)!}, \tag{46}$$

where the second factor is combinatorial. For large N we have $\frac{N!}{\prod\limits_j (p_j N)!} \sim 2^{-N \sum\limits_i p_i \log p_i}$. Therefore

$$P \sim 2^{-N \sum\limits_i p_i (\log p_i - \log q_i)}. \tag{47}$$

From this we defined the relative entropy

$$S(P_X || Q_X) = \sum_i p_i (log p_i - log q_i). \tag{48}$$

This is the information we gain if we accept the new probability distribution P. We can be sure our initial distribution is wrong when $NS(P_X||Q_X) \gg 1$.

The entropy satisfies some important properties we will just list here

1. Subaditivity: $S_X + S_Y - S_{XY} \geq 0$.
2. Monoticity of rel. entropy: $S(P_{X,Y}||Q_{X,Y}) \geq S(P_X||Q_X)$.
3. Strong subaditivity: $S_{XY} + S_{YZ} \geq S_Y + S_{XYZ}$.

You don't need to learn them, they are here just to show you where this field is heading to.

1.4.5 Quantum information

There are two types of probabilities. We know statistical probabilities (lack of information about the system) and quantum probabilities (quantum superposition). How to mix them together? What can we do if we have a machine that randomly generates quantum states? This can be described using the density matrix

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|, \tag{49}$$

where $|\psi_j\rangle$ are pure quantum states and p_j are the corresponding probabilities. A mean value of an obsarvable A is obtained as

$$\langle A\rangle = \mathrm{Tr}(\rho A). \tag{50}$$

There is another way to derive this. Consider we have a system A and an environment B. The total Hilbert is of the form $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. The states are of the form

$$\psi_{AB} = \psi_A \otimes \psi_B, \tag{51}$$

Observables in A can be expressed as $\mathcal{O}_A \otimes 1_B$

$$\langle\psi_{AB}|\mathcal{O}_A \otimes 1_B|\psi_{AB}\rangle = \langle\psi_A|\mathcal{O}_A|\psi_A\rangle\langle\psi_B|1_B|\psi_B\rangle. \tag{52}$$

As an example think of $\mathcal{H}_A$ as being 2-dimensional and $\mathcal{H}_B$ as three dimensional. A general state (is pure but entangled and) is in the form

$$\psi_{AB} = \begin{pmatrix} * & * & * \\ * & * & * \end{pmatrix} \sim \begin{pmatrix} \sqrt{p_1} & 0 & 0 \\ 0 & \sqrt{p_2} & 0 \end{pmatrix} \tag{53}$$

where the later form can be obtained by unitary transformation (p_i sum up to one). In general any pure state can be written as

$$\psi_{AB} = \sum_i \sqrt{p_i}\psi_A^i \otimes \psi_B^i \tag{54}$$

What is the expectation value of $\mathcal{O}_A$? It is

$$\langle\mathcal{O}_A\rangle = \langle\psi_{AB}|\mathcal{O}_A \otimes 1_B|\psi_{AB}\rangle = \sum_i p_i\langle\psi_A^i|\mathcal{O}_A|\psi_A^i\rangle \tag{55}$$

which is exactly $\mathrm{Tr}\rho_A\mathcal{O}_A$ for $\rho_A = \sum_i p_i|\psi_A^i\rangle\langle\psi_A^i|$. A pure state is a state that has $\rho^2 = \rho$.

There are many definitions of quantum entropies but the most important is the von Neumann entropy

$$S(\rho_A) = -\mathrm{Tr}\rho_A \log \rho_A. \tag{56}$$

This, no surprise, yields now $S(\rho_A) = -\sum_i p_i \log p_i$.

2 Thermodynamics (2)

2.1 Thermodynamic potentials, response functions and stability conditions

References: Reichl 3.1 – 3.7.

2.1.1 The fundamental equation(s) of thermodynamics

The starting point of thermodynamics were the three (plus one) laws:

- **Zeroth law:** Two bodies in equilibrium with a third one are in equilibrium with each other.
- **First law:** Energy is conserved.
- **Second law:** Heat flows spontaneously from high temperature to low temperature.
- **Third law:** It is not possible to reach zero (absolute) temperature in a finite number of reversible steps.

Now imagine you knew nothing about thermodynamics, you wouldn't understand three of them, but you would agree with first law. There are start differences between the physics of a single or a few particles and the physics of $N \sim 10^{23}$ particles, new important concepts such as the entropy play a crucial law. However, there some concepts which are equally valuable for both microscopic and macroscopic systems – energy being a prime example.

What is the source of the omnipresence of energy? Why it seems to appear nearly everywhere? It is due to Nöether's theorem, as long as the law of physics (expressed in terms of Lagrangian) do not change in time there is conserved quantity we call energy. This makes the energy one of the bridges connecting the world of the *micro* and the *macro*.

The changes in internal energy is given by

$$dU = \delta Q - \delta W + \mu_j dN_j, \tag{57}$$

where δQ is the heat, δW and μ_j is the chemical potential for ν different types of particles. We use δ instead of d to underline that those quantities are small, but are not exact differentials (meaning there is no quantity such that $\delta Q \; = Q(s + d\ s) - Q(s)$. The work done upon (or by) the system can be of many forms

$$\delta W = PdV - \sigma dA - \phi de + ..., \tag{58}$$

where the PdV is well known, σ is the surface tension coupled to the surface area dA, ϕ is the electric potential coupled to the change of charge de and so on. In general we can write $\delta W = -Y dX$, where $Y = -P, \sigma, \phi, ...$ are the generalised forces and $dX = dV, dA, de, ...$ are generalised displacements (note a minus sign in front of P).

A novel variable in the context of statistical physics is the entropy defined as

$$S = k_B \log \Omega, \tag{59}$$

where Ω. Change of the entropy for a reversible process is a thermodynamic property

$$dS = \frac{\delta Q}{T} + d_i S, \tag{60}$$

where $d_i S$ denotes contribution from irreversible processes which is always positive; therefore $dS \geq 0$. Note that the entropy S is an extensive quantity

$$S\left(\lambda U, \lambda X, \{\lambda N_i\}\right) = \lambda\left(U, X, \{N_i\}\right). \tag{61}$$

Together we have

$$T dS \geq \delta Q = dU - Y dX - \sum \mu_j \, dN_j. \tag{62}$$

In thermodynamics quantities often depend on many others and one has to be careful when taking derivatives. Usually this is explicitly states as $\left(\frac{\partial y}{\partial x}\right)_z$ where the subscript means *while keeping z fixed.*

From (62) we have

$$\begin{aligned} \left(\frac{\partial S}{\partial U}\right)_{X,\{N_j\}} &= \frac{1}{T} = \beta \\ \left(\frac{\partial S}{\partial X}\right)_{U,\{N_j\}} &= -\frac{Y}{T} \\ \left(\frac{\partial S}{\partial N_j}\right)_{U,X} &= \; = \frac{\mu_j}{T} \end{aligned} \tag{63}$$

$$\tag{64}$$

These are called the thermal, mechanical and chemical equations of state. An important equation can be taking by taking the derivative $\frac{d(\lambda S)}{d\lambda}$. The by using (63) and setting $\lambda = 1$ one obtains the Euler equation (also sometimes called the fundamental equation of thermodynamics)

$$TS = U - XY - \mu_j N_j. \tag{65}$$

By taking its differential and substracting (62) one gets the Gibbs-Duhem equation which relates differentials of intensive variables

$$S dT + X dY + N_j d\mu_j = 0. \tag{66}$$

2.1.2 Potentials

We are interested in the amount of energy that can be stored and, let us be honest, extracted from a system. This all follows from (62) which different constrains applied and different set of variables.

Internal energy
This one we have seen already, it is $U(S, X, \{N_j\})$. The other quantities T, Y, μ are considered to depend on $S, X \text{and} N_j$ (see the equations of state).

The total differential dU can be obtained directly from (62) and so can be the fundamental equation, while the equation of states follows from (63). From (62) we can observe that equilibrium state (with fixed S, X, N_j is the one with minimal energy.

Useful relations can be obtained from the fact that the order of derivatives can be exchanged and that many quantities are related by derivatives. For example, take U and take derivatives w.r.t. S and X. Using this one gets $\left(\frac{\partial T}{\partial X}\right)_{S,\{N_j\}} = \left(\frac{\partial Y}{\partial S}\right)_{X,\{N_j\}}$, there are three more relations that can be obtained this way, jointly they are called Maxwell relations.

Enthalpy
The enthalpy $H(S, Y, N_j)$ describes a system that is thermally isolated and closed but mechanically coupled to the exterior. It has the following fundamental equation

$$H = U - YX, \tag{67}$$

where $-YX$ comes from the Legendre transformation from variables (S, X, N_j) to (S, Y, N_j). A similar set of equations can be constructed from (63), (62) and (65) (total differential, fundamental equation, equation of states, Maxwell equations and the equilibrium condition.)

Helmholtz Free Energy
For processes at fixed T, X, N_j the potential is the Helmholtz Free Energy with the fundamental equation

$$A = U - ST, \tag{68}$$

where the origin of $-ST$ is again from the Legendre transformation. Again, similar set of equations can be derived.

Gibbs Free Energy
For processes at fixed T, Y, N_j the potential is the Helmholtz Free Energy with the fundamental equation

$$G = U - ST - XY, \tag{69}$$

two new terms, but rest is the same.

Grand potential
For processes at fixed $T, X \text{and} \mu_j$ the potential is the Grand potential with the fundamental equation

$$\Omega = U - ST - \mu_j N_j = XY, \tag{70}$$

a very simple expression indeed!

2.1.3 Response functions and stability conditions

Response functions tell us how does observables of the system change when we act upon it. Usually we try to keep as many quantities fixed as possible and therefore we various types of the response functions. For example, the specific heat C relates the increase in temperature dT after adding amount of heat δQ, but the temperature increment is different if we kept the volume V fixed or the pressure P fixed and so on. Therefore, there we have various heat capacities C_P, C_V and so on. Response values are valuable as they can be easily measured and therefore connects the micro and macro worlds.

With generalised forces X and positions Y we can do the same exercise, one just need to utilise the equation (62) and possible apply a Maxwell relation to change a relation here and there. We have to heat capacities

$$\begin{aligned} C_{X,N_j} &= \left(\frac{\partial U}{\partial T}\right)_{X,N_j} = -T\left(\frac{\partial^2 A}{\partial T^2}\right)_{X,N_j} \\ C_{Y,N_j} &= C_{X,N_j} + \left[\left(\frac{\partial U}{\partial X}\right)_{T,N_j} - Y\right]\left(\frac{\partial X}{\partial T}\right)_{Y,N_j} = -T\left(\frac{\partial^2 G}{\partial T^2}\right)_{Y,N_j} \end{aligned} \tag{71}$$

these are the heat response functions. We also have mechanical response functions. For general case they are called isothermal susceptibility

$$\chi_{T,N_j} = \left(\frac{\partial X}{\partial Y}\right)_{T,N_j} = -\left(\frac{\partial^2 G}{\partial^2 Y}\right)_{T,N_j} \tag{72}$$

adiabatic susceptibility

$$\chi_{S,N_j} = \left(\frac{\partial X}{\partial Y}\right)_{S,N_j} = -\left(\frac{\partial^2 H}{\partial^2 Y}\right)_{S,N_j} \tag{73}$$

and thermal expansivity

$$\alpha_{Y,N_j} = \left(\frac{\partial X}{\partial T}\right)_{Y,N_j}. \tag{74}$$

These are the definitions for XYT systems, for PVT systems they include the factor $1/V$ in their definition.

An important fact is that using the expression for potentials, the response functions can be related to each other

$$\chi_{T,N_j}\left(C_{Y,N_j} - C_{X,N_j}\right) = T\alpha^2_{Y,N_j} \tag{75}$$

$$C_{Y,N_j}\left(\chi_{T,N_j} - \chi_{S,N_j}\right) = T\alpha^2_{Y,N_j} \tag{76}$$

$$\frac{C_{Y,N_j}}{C_{X,N_j}} = \frac{\chi_{T,N_j}}{\chi_{S,N_j}} \tag{77}$$

2.1.4 Stability conditions

For an isolated system to be in equilibrium it has have maximal entropy. This is basically a probability statement. However, this is true only in the limit of infinite number of degrees of freedom. A finite case has non-zero probability to fluctuation out of the equilibrium. For the equilibrium to be stable, the changes invoked by the fluctuation need to produce a response that drives the system back to the equilibrium state. This requirement restrict many important properties of the system, let us investigate them

Consider the system to be split into two parts, A, B, all of the extensive quantities are split as $V = V_A + V_B$ and so on for U, S, N. Change in the entropy in the first order is

$$\Delta S_T = \left(\beta_A^0 - \beta_B^0\right) \Delta U_A + \left(P_A^0\beta_A^0 - P_B^0\beta_B^0\right) \Delta V - \left(\mu_{j,A}^0\beta_A^0 - \mu_{j,B}^0\beta_B^0\right) \Delta_{j,A} + ... \tag{78}$$

as the energy is always allowed to flow we have $\beta_A^0 = \beta_B^0$, the temperature of both parts of the system need to be the same – entire system has a single energy – the first important consequence of equilibrium stability. If mechanical exchange is allowed as well, the system has to have a single equilibrium pressure. If particles can move between the subsystems, the chemical potentials has to be equal.

One can go one order further. For them to be strictly negative one needs to require that (assume a single particle PVT system)

$$C_{V,N} \geq 0, \kappa_{T,N} = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_{T,N} \geq 0. \tag{79}$$

These are examples of Le Chatelier principle: *If a system is in stable equilibrium, the any spontaneous change in its parameters must bring about processes which tend to restore the system to equilibrium.* For example, when we add some heat, the temperature of the heated subsystem has to increase so the heat moves to other parts. Or when we increase the volume, the pressure decreases so neighbouring subsystems counteract the change.

Recall that response functions are related to potentials. Therefore, the equilibrium condition teaches us something about (some of) them as well. The Helmholtz free energy is a concave function of temperature and a convex function of volume. The Gibbs free energy is a concave function of temperature and a concave function of pressure.

2.2 Throttling, osmotic pressure and chemical batteries

References: Reichl 3.8 - 3.11

The (pedagogical) problem with thermodynamic potentials is that they all look sort of the same: we fix something, does not fix something else and get some relations. To fight this problem we will have a look at three situations physics of which are greatly enlightened by thermodynamic potentials.

2.2.1 Throttling

Imagine a volume of gas separated by a movable porous (but thermally isolating) wall. Porous means particles can move through but the pressure does not equalize. Let denote the volumes and pressures in two parts as V_A, V_B and P_A, P_B. As we move the wall, V_A goes from V to 0 and vice versa for V_B. Each part is connected to a reservoir so the pressure is held fixed.

Which of the potentials is relevant for this situation? The system under consideration is thermally isolated but mechanically coupled (work can be done by moving something), so the potential to consider is the enthalpy. Note that during the process

$$\Delta W = \int_0^{V_f} P_f dV + \int_{V_i}^{0} P_i dV = P_f V_f - P_i V_i = -\Delta U \tag{80}$$

since $\delta Q = 0$. Therefore we have

$$H_i = U_i + P_i V_i = U_f + P_f V_f = H_f, \tag{81}$$

e.g. the enthalpy is constant during the process. Recall that $H = H(S, Y, N_j)$. If $dN = 0$ then

$$dH = TdS + VdP = 0. \tag{82}$$

Note that $dS = \left(\frac{\partial S}{\partial T}\right)_P dT + \left(\frac{\partial S}{\partial P}\right)_P dP$ and we have $C_P = T\left(\frac{\partial S}{\partial T}\right)_P$. Together we have

$$C_P dT + \left(V - T\left(\frac{\partial V}{\partial T}\right)_P\right) dP = 0 \tag{83}$$

where we have used the Maxwell relation $\left(\frac{\partial S}{\partial P}\right)_P = -\left(\frac{\partial V}{\partial T}\right)_P$.

$$\mu_{JT} = \left(\frac{\partial T}{\partial P}\right)_H = \frac{1}{C_P}\left(T\left(\frac{\partial V}{\partial T}\right)_P - V\right). \tag{84}$$

Here, μ_{JT} is called the Joule-Thomson coefficient that relates changes in temperature due to change in a pressure $dT = \mu_{JT} dP$. Why is it important? Imagine a long pipeline separated by a vent with different pressures on each side. Passing through the vent changes the temperature of the gas!

For ideal gas we have $\mu_{JT} = 0$, but for more realistic van der Waals gas ($\left(P + a\frac{n^2}{V^2}\right)(V - nb) = nRT$) we have (in the low density limit)

$$\mu_{JT} \approx \frac{2}{5R}\left(\frac{2a}{RT} - b\right) \tag{85}$$

where a, b are parameters of van der Waals gas and R is the ideal gas constant. Note that the sign of μ_{JT} can change with temperature. This effect works for both gases and liquids.

2.2.2 Osmosis

Submerge a hollow cylinder with only-water-permeable membrane into a water bath. Now put sugary water into the cylinder. What happens? The height of (sugary) water increases by h. Why does this happen? Because of osmosis. As a result, there is a pressure difference $\pi = \rho_s h g$ across the membrane. How can we obtain π?

Differential of the Gibbs free energy $G(P, T, N_w, N_s)$ can be written as

$$dG = -SdT + VdP + \mu_w dn_w + \mu_s dn_s \tag{86}$$

where n are number of moles of either water or sugar. It is sometimes preferred to work with molar observables. It is useful to define $x = x_s + x_w$ where

$$\begin{aligned} x_w &= \frac{n_w}{n_w + n_s} \\ x_s &= \frac{n_s}{n_w + n_s} \end{aligned} \tag{87}$$

The equilibrium is reached when

$$\mu_w^0(P_0, T) = \mu_w(P, T, x_s), \tag{88}$$

where $P = P_0 + \pi$. There is a model for G that describes well the solution for $n_s \ll n_w$

$$G(P, T, n_s, n_w) = n_w \mu_w^0(P, T) + n_s \mu_s^0(P, T) - \lambda \frac{n_s n_w}{n} + n_w RT \log x_w + n_s RT \log x_s. \tag{89}$$

The middle term is due to interaction between water and sugar molecules. The last two terms are contributions due to mixing. Where do the log come from? They are from the Boltzman factor. A change in concentration is related to the change in energy $\Delta\varepsilon$ by $x = e^{-\Delta\varepsilon/k_B T}$. We are interested in the energy here so we take log of this equation. From this we have

$$\mu_w(P, T, x_s) = \left(\frac{\partial G}{\partial n_w}\right)_{P,T,n_s} \approx \mu_w^0(P, T) - x_s RT \overset{!}{==} \mu_w^0(P_0, T) \tag{90}$$

Now we have two expressions for μ_w, one at P and the other at P_0. As water is incompressible, the difference is small and their difference will be proportional to $P - P_0 = \pi$. The constant of proportionality is

$$\left(\frac{\partial \mu_w^0}{\partial P}\right)_{T,n_w} = \left(\frac{\partial V_0}{\partial n_w}\right)_{T,P} = \nu_w^0, \tag{91}$$

where $\nu_w^0 = V^0/n_W$ is the partial molar volume. And yes, we have used Maxwell relation, glad that you ask. Plugging the relations for $\mu_w^0(P_0, T)$ and $\mu_w(P, T, x_s)$ into the equilibrium condition we get

$$\pi \approx \frac{n_s RT}{V}. \tag{92}$$

That looks pretty much like the ideal gas law!

Further reading: `https://arxiv.org/pdf/1409.3985.pdf`

2.2.3 Electrolytes and batteries

Not part of the lectures.

Electrolytes are substances that ionize in certain solvents. The solution therefore consists of electrically charged particles. Why is it important? We have to take electric interaction into account as there might be charge-dependent energy needed to add/remove the particle

$$\mu_j^e = \mu_j + z_j \Phi F, \tag{93}$$

where F is the amount of charge in one mole of protons (called a Faraday, $1F \doteq 96.49 C/mol$). The solute (for example salt) splits into ν_c cations and ν_a anions with corresponding charges z_c and z_a, so we have $\nu_a z_a + \nu_c z_c = 0$ as long as the whole solution remains electrically neutral. The equilibrium is reached when

$$\mu_{ac} = \nu_a \mu_a + \nu_c \mu_c. \tag{94}$$

It is very difficult to obtain μ_{ac}, but experimental data suggest

$$\mu_{ac}(P,T,x_{ac}) = \mu^0_{ac}(P,T) + RT\log\alpha_{ac}, \tag{95}$$

where α_{ac} is called the *activity* and μ^0_{ac} is the chemical potential of the salt in aqueous solution at temperature T and pressure P in the infinite dilution limit (energy needed to add one salt particle to pure water). The activity can be related to activity of ions as

$$\log(\alpha_{ac}) = \nu_a \ln\alpha_a + \nu_c \ln\alpha_c. \tag{96}$$

In some cases (called ideal solutions) the activity is the concentration $\alpha \approx c$ (mol/volume), giving the equilibrium condition

$$\mu^e_j = \mu^0_j(P,T) + RT\log c_j + z_j F\Phi. \tag{97}$$

The setting for a simple battery goes like this. We have two metals submerged in solutions of their corresponding salts (for example silver electrode submerged in a dilute solution of $AGNO_3$. The metals are connected by a wire with a volt-meter and the solutions are connected by salt-bridge (potentials in solutions can equate). There are four different potentials, one for each metal and one for each solution.

The silver ions can move to the electrode (leading to positive holes) or it can dissolve and leave and leave excess electrons behind (recall that $AGNO_3$ dissociates into Ag^+ and NO_3^-).

We therefore have two equilibrium conditions

$$\begin{aligned}
\mu^{0,1}_{\mathrm{Ag}+}(s) + zF\Phi_1 &= \mu^0_{\mathrm{Ag}+}(l) + RT\ln c_1 + zF\phi_1 \qquad (98)\\
\mu^{0,2}_{\mathrm{Ag}+}(s) + zF\Phi_2 &= \mu^0_{\mathrm{Ag}+}(l) + RT\ln c_2 + zF\phi_2 \qquad (99)
\end{aligned}$$

Salt bridge makes $\phi_1 = \phi_2$. Also, $\mu^{0,1}_{\mathrm{Ag}+} = \mu^{0,2}_{\mathrm{Ag}+}$ as they are both silver. We can now subtract the equations to get

$$\Phi_1 - \Phi_2 = \frac{RT}{zF}\ln\left(\frac{c_1}{c_2}\right), \tag{100}$$

which is called the Nernst equation.

3 Phase transitions (2)

References: Reichl 4.1 – 4.3, 4.5, 4.6.1 and Tong 5.1.3

3.1 Thermodynamical phase transitions

3.1.1 Classification of phase transitions

Recall the Gibbs phase rule. For two phases to coexist, the corresponding chemical potentials has to be equal $\mu^I(Y,T) = \mu^{II}(Y,T)$ which limits the number of independent degrees of freedom. In general, consider n types of molecules in f possible phases. For each phase we have $n-1$ independent concentrations plus two thermodynamical quantities Y, T, therefore

$$\#d.o.f. = 2 + f(n-1). \tag{101}$$

For each phase boundary we have n restrictions due to chemical potentials, which gives

$$\#n.o.r. = n(f-1). \tag{102}$$

Subtracting the number of restrictions from the number of degrees of freedom we obtain

$$v = 2 + n - f \tag{103}$$

independent quantities. This for example tells us that for a single type system ($n = 1$) we can have coexistence of 3 phases at most.

The system with various component has Gibbs free energy of the form $G(Y, T, x_j)$ (x_j is the relative concentration. For processes at constant Y, T its change is given by

$$dG_{Y,T} = \mu_j dn_j. \tag{104}$$

We know that across the phase transition the chemical potential $\mu_j = \left(\frac{\partial G}{\partial n_j}\right)$ has to be equal. However, the other two quantities

$$\begin{aligned} X &= -\left(\frac{\partial G}{\partial P}\right)_{T,n_j} \\ S &= -\left(\frac{\partial G}{\partial T}\right)_{Y,n_j} \end{aligned} \tag{105}$$

are unrestricted. In principle, either them or their derivatives can be discontinuous which can be used to classify the order of phase transition. The old way (by Ehrenfest) was to define the order of transition by the first discontinuous derivative. So for example if there is a discontinuity in the first derivative and therefore $\Delta S \neq 0$ and $\Delta V \neq 0$ the transition would be called first-order. The modern classification takes first order transition to be one with latent heat $\Delta H = T\Delta S$ and higher-order transitions are called continuous.

3.1.2 Van der Waals gas

The chemical potential conditions gives us the coexistence curves defined by the Clausius-Clapeyron equation. Good example of phase transition is that of water analyzed by the means of the van der Waals equation. The crucial point was the Maxwell construction which followed from requirement of constant Gibbs free energy across the transition.

The equation of state for van den Waals gas is

$$p = \frac{k_b T}{v-b} - \frac{a}{v^2}, \tag{106}$$

where $v = V/N$ and a, b are factors due to interaction and non-zero particle size. At the critical point we have $p_c(v - v_c)^3 = 0$ which gives

$$k_B T_c = \frac{8a}{27b}, v_c = 3b, p_c = \frac{a}{27b^2}. \tag{107}$$

In physics we like to find that one system behaves in a similar way than something completely different – we like to search for universal behaviour. To do so, one needs to move to a universal/reduces variables which are in this case given by

$$\bar{T} = T/T_c, \bar{v} = v/v_c, \bar{p} = p/p_c. \tag{108}$$

This turns the vdW equation into the following form

$$\bar{p} = \frac{8}{3}\frac{\bar{T}}{\bar{v} - 1/3} - \frac{3}{\bar{v}^2}, \tag{109}$$

this now works for any gas. However, there are some limitations, our model is still a bit oversimplifying, the constant $p_c v_c/(k_b T_c) = 3/8 = 0.375$ is not matched perfectly by experiments (results in $0.28 - 0.3$ range). On the other hand, even if not perfect, the statement is very strong (and to a very good degree confirmed by experiments): all fluids described by their reduced quantities behave (nearly) the same. This is know as the law of corresponding states. Universality, here we come!

Universality: Universality is a topic that is often overlooked during lectures. The name is rather self-explanatory, but since it is such an important concept not only in physics, let us pause a moment to discuss it. We have seen a universal law before: the central limit theorem. No matter how are some quantities distributed statistically (as long as they are not correlated and there are enough of them), their mean is distributed normally. We have proved that it has to be so. There are other famous statistical rules of this type, for example the Benford law (statistics of digits in numbers) or the Zipf's law (rank-frequency distribution). While with the first one it is sort-of well understood why it works, not so much for the other one! This is the universality: the same law describing absolutely different things.

Let us have a look at the Zipf's law. To begin, make a sorted list of frequencies, for example the list of words in a language sorted by the frequency. The law says that the rankings of two different items on the list and their corresponding frequencies are in a fixed ration: $\frac{\mathrm{rank}_A}{\mathrm{rank}_B} = \left(\frac{\mathrm{freq}_A}{\mathrm{freq}_B}\right)^r$, so for example ($r = 1$) the second most frequent number appears in a language half as often as the first one. But that is not all! Ration of frequencies of the first and the second is the same as the hundredth and two-hundredth word! In principle the value of r can be different for different systems, dividing them into something called *universality class*.

We have a similar thing in physics. You can have two different systems, for example a hexagonal spin lattice and rectangular lattice with (un)activeted bonds. However, when close to ordered/disordered phase transitions cluster will form. The structure of cluster is the same across all scales and also the same for both systems. Actually phase transitions of physical systems breaks into a number of universality classes. (We will see shortly how to do the classification.)

The universality doesn't stop here! There seems to be a similarity between the behaviour of nuclear resonances, random matrices and prime numbers! That doesn't mean that they are deeply interconnected in a sense that when a neutron tries to penetrate a nucleus it has to consider all possible prime numbers. However, we have to admit that in many cases the reason why universality appears is not clear. This topic is still being thoroughly researched.

(Non-technical article on this subject by Terence Tao: `https://terrytao.wordpress.com/2010/09/14/a-second-draft-of-a-non-technical-article-on-universality/`).

3.1.3 Critical exponents

To compare transitions of various type a valuable thing to compare is how do various quantities change when the critical temperature is approached, for example, how does $v_{gas} - v_{liquid}$ change w.r.t. to $T_c - T$? The formal definition of a critical coefficient is that if we express the quantity f as a function of $\varepsilon = (T - T_c)/T_c$ it can be expanded, near the critical point, as $f(\varepsilon) = Ae^{\lambda}(1 + B\varepsilon^{y} + ...)$. Here the critical coefficient λ is

$$\lambda = \lim_{\varepsilon \to 0} \frac{\ln f(\varepsilon)}{\ln \varepsilon}. \tag{110}$$

From reduced vdW equation we get

$$\bar{T} \approx 1 - \frac{1}{16}\left(\bar{v}_{gas} - \bar{v}_{liquid}\right)^2 \tag{111}$$

which gives us

$$\delta\bar{v} \sim (T_c - T)^{\frac{1}{2}} \tag{112}$$

Similarly we have

$$p - p_c \sim (v - v_c)^3. \tag{113}$$

Another important factor is the compressibility

$$\kappa = -\frac{1}{v}\left(\frac{\partial v}{\partial p}\right)_T \sim (T - T_c)^{-1} \tag{114}$$

The critical coefficients/exponents have canonical names, in this case $\delta\bar{v} \sim \varepsilon^{\beta}$), $\kappa \sim \varepsilon^{-\gamma}$ and $\delta\bar{p} \sim \varepsilon^{\delta}$. Where is α? Usually in the definition of specific heat $C_V \sim \varepsilon^{\alpha}$.

Again, vdW equation gives fine but far from perfect prediction for experimental results ($(0.5, -1, 3)$ versus $(0.32, 1.2, 4.8)$ for (β, γ, δ).

By the way, we are used to the fact that the fluctuations are mild, $\Delta N/N \sim 1/\sqrt{N}$. However, the grand canonical partition function can be written as $\log Z = \beta V p(T, \mu)$ and the average particle number is

$$\langle N \rangle = V\left(\frac{\partial p}{\partial \mu}\right)_{T,V} \tag{115}$$

and

$$\Delta N^2 = T\left(\frac{\partial \langle N \rangle}{\partial \mu}\right)_{T,V} \tag{116}$$

by dividing those we get

$$\Delta N^2/N = \frac{1}{V\beta}\left(\frac{\partial \langle N \rangle}{\partial p}\right)_{T,V} \tag{117}$$

which by using the relation for partial derivatives $(\partial x/\partial y)_z(\partial y/\partial z)_x(\partial z/\partial x)_y = -1$ gives

$$\Delta N^2/N = -\frac{1}{\beta}\left(\frac{\partial \langle N \rangle}{\partial V}\right)_{p,T}\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_{N,T}. \tag{118}$$

On the l.h.s. we have the fluctuations on the r.h.s. there is the compressibility κ which as we have just found out diverges at the critical point. Therefore fluctuations grown indefinitely there as well.

3.1.4 Binary mixtures

Not part of the lectures.

Let us consider a mixture of two liquids, the Gibbs free energy is just

$$G(T, P, n_j) = n_1\mu_1 + n_2\mu_2. \tag{119}$$

We can use molar fractions x_i to write the differential for molar Gibbs free energy $g = G/n$ as

$$dg = -sdT + vdP + (\mu_1 - \mu_2)\, dx_1. \tag{120}$$

Using the stability condition $\left(\frac{\partial \mu_1}{\partial n_1}\right)_{P,T,n_2} > 0$ we can also show that $g(T, P, x_1)$ has to be a convex function of the mole fraction.

What are the equilibrium conditions? The chemical potentials $\mu_i = \left(\frac{\partial g}{\partial n_i}\right)_{P,T,n_{j\neq i}}$ for each particle type has to be equal in each of the phases

$$\begin{aligned} g^I + (1 - x_1^I)\left(\frac{\partial g}{\partial x_1}\right)^I_{P,T} &= g^{II} + (1 - x_1^{II})\left(\frac{\partial g}{\partial x_1}\right)^{II}_{P,T} \\ g^I - x_1^I\left(\frac{\partial g}{\partial x_2}\right)^I_{P,T} &= g^{II} - x_1^{II}\left(\frac{\partial g}{\partial x_2}\right)^{II}_{P,T} \end{aligned} \tag{121}$$

combing these equations gives us

$$\begin{aligned} \left(\frac{\partial g}{\partial x_1}\right)^I_{P,T} &= \left(\frac{\partial g}{\partial x_1}\right)^{II}_{P,T} \\ g^I - g^{II} &= (x_1^I - x_1^{II})\left(\frac{\partial g}{\partial x_1}\right)^I_{P,T} \end{aligned} \tag{122}$$

This tells us that at the equilibrium points g has the same points and there is tangent connecting those two points.

3.2 (Ginzburg-)Landau theory

References: Reichl 4.8, Tong 5.4

To study phase transition we want to identify an order parameter η. It can be, for example, the net magnetisation or the difference of the densities (liquid/gas transition). The discussion still can be very general.

The free energy ϕ is a function of $T, Y \text{and} f$, where f is a force conjugate to the order parameter

$$\phi(T, Y, f) = \phi_0(T, Y) + \alpha_2(T, Y)\eta^2 + \alpha_3(T, Y)\eta^3 + ... - f\eta. \tag{123}$$

Above the transition point we have $\langle\eta\rangle = 0$ so there can be no linear term. There are two important classes, depending on whether α_3 is present or not.

3.2.1 Continuous phase transition

In this case $\alpha_3 = 0$. Also, let us turn off the external force, $f = 0$. The question is, what value of η does minimise the free energy? Let us assume $\alpha_4 > 0$. If also $\alpha_2 > 0$ then the potential has a minimum at $\eta = 0$. However, if $\alpha_2 < 0$ the minimum is at $\eta \neq 0$. Therefore, the critical point is at

$$\alpha_2(T, Y) = 0. \tag{124}$$

Close to the critical point we can write $\alpha_2 = \alpha_0(T - T_c(Y))$, where α_0 is slowly varying function of T. There can be two equal minima, in this case the system chose one randomly (spontaneous symmetry breaking).

Minima of the order parameter are

$$\begin{aligned} \eta &= 0, T > T_c \\ \eta &= \pm\sqrt{\frac{\alpha_0}{2\alpha_4}(T_c - T)}, T < T_c. \end{aligned} \tag{125}$$

Plugging this back to the expression for ϕ we get that ϕ is equal to

$$\begin{aligned} \phi(T, Y, \eta) &= \phi_0(T, Y), T > T_c \\ \phi(T, Y, \eta) &= \phi_0(T, Y) - \frac{(\alpha_0^2(T_c - T))^2}{4\alpha_4}, T < T_c. \end{aligned} \tag{126}$$

Therefore there is no jump in the energy but there is in the specific heat $c_y = -T\left(\frac{\partial^2\phi}{\partial T^2}\right)_Y$

$$c_Y(T_c^-) - c_Y(T_c^+) = \frac{T_c\alpha_0^2}{2\alpha_4}. \tag{127}$$

If we turned on the force f this behaviour would change, continuous phase transition is destroyed. The equilibrium state is now give by

$$2\alpha_2\eta + 4\alpha_4\eta^3 - f = 0. \tag{128}$$

We can define the susceptibility

$$\chi(f) = \left(\frac{\partial\eta}{\partial f}\right)_{T,Y} = \frac{1}{2\alpha_2 + 12\alpha_4\eta^2}. \tag{129}$$

In the limit of $f \to 0$ we have discontinuity in the (diverging) susceptibility

$$\begin{aligned} \chi(f) = \lim_{f\to 0}\left(\frac{\partial\eta}{\partial f}\right)_{T,Y} &= \frac{1}{2\alpha_0(T - T_c)}, T > T_c \\ \chi(f) = \lim_{f\to 0}\left(\frac{\partial\eta}{\partial f}\right)_{T,Y} &= \frac{1}{4\alpha_0(T_c - T)}, T > T_c \end{aligned} \tag{130}$$

3.2.2 First-order transition

Now let us take $\alpha_3 \neq 0$. The potential now has two minima

$$\eta = 0 \text{ and } \eta = (-3\alpha_2 \pm \sqrt{9\alpha_3^2 - 32\alpha_2\alpha_4})/8\alpha_4. \tag{131}$$

When $9\alpha_3^2 - 32\alpha_2\alpha_4 > 0$ there are two minima, but the system prefers the lower one. Therefore, the transition happens at the point where the Gibbs free energy at the displaced minimum is the same as the one at the origin. By the way, for a finite size system the upper vacuum the system can move between two vacua. Minimum of a potential that is not the global minimum is called a false vacuum.

3.2.3 Magnetisation and superconductors

Magnetisation
Let us have a look at an example – a model of magnetization:

$$\Phi(T, \mathbf{H}) = \phi_0(T) - \mathbf{M} \cdot \mathbf{H} + \alpha_2 \mathbf{M}^2 + \alpha_4 \mathbf{M}^4 + ... \tag{132}$$

When we turn of the magnetic field $\mathbf{H}$, the magnetization $\mathbf{M} = 0$ above the critical point (called Curie temperature) and $\mathbf{M} = \pm\sqrt{(\alpha_0/2\alpha_4)(T_c - T)}\mathbf{e}_M$ below it. The specific heat curve (for Nickel for example) resembles the letter λ and therefore it is called the λ-curve, how clever!

Superconductors
So far we have been a little bit too oversimplifying (but glad it was working). One imporant problem: we assumed that fluctuations are not present, all physics are described by a single quantity η. However, as noted in the case of vdW equation, fluctuation tend to be infinite at the critical point. Some people call what have discussed so far only Landau theory and Ginzburg-Landau theory is the theory enhanced by fluctuations. We will not do this here, we will investigate fluctuations later in this course. At this moment, we will have a peak through the model of superconductors.

The condensed phase in a superconductor corresponds to a macroscopic quantum state (wave function) Ψ, the wave function can be used as the order parameter η, the free energy (per unit volume) is

$$\Phi(\mathbf{r}, T) = \phi_n(T) + \alpha_2(T)|\Psi(r)|^2 + \alpha_4(T)|\Psi(r)|^4 + \frac{1}{2m}|i\hbar\partial_i\Psi(\mathbf{r})|^2, \tag{133}$$

a very similar model can be written for magnetization with fluctuations allowed (actually the same, see Tong's lecture notes).The total energy can be found by integrating over the volume. We extremise by requiring $\int \delta(\Phi) d\mathbf{r} = 0$ (take $\Psi^* \to \Psi^* + \delta\Psi^*$ and require vanishing linear contribution) which gives us

$$\alpha_2\Psi + 2\alpha_4\Psi|\Psi|^2 - \frac{\hbar^2}{2m}\Delta\Psi(\mathbf{r}) = 0. \tag{134}$$

If we assume Ψ is real (fine as long there is no current of electron pairs) and rescale $f(z) = \Psi\sqrt{\alpha_4/2|\alpha_2|}$ we obtain

$$-\xi^2(T)f'' - f + f^3 = 0, \tag{135}$$

where $\xi(T)$ is the Ginzburg-Landau coherence length. We want to study spacial variations in the solution, how does it change from $f = 0$ (no condensate) in one region to $f = 1$ (condensate) in another. We will assume $f(z) = 0$ for $z \leq 0$ and $f(\infty) = 1$. To solve the (135) we first multiply it by f', rearrange and integrate to obtain

$$-\xi^2(T)(f')^2 = f^2 - \frac{1}{2}f^4 + c, \tag{136}$$

the integration constant can be fixed to $c = -1/2$ by requiring $f(\infty) = 1$ and $f(\infty)' = 0$. This gives

$$-\xi^2(T)(f')^2 = \frac{1}{2}(1 - f^2)^2, \tag{137}$$

which is solved by $f(z) = \tanh(z/\sqrt{2}\xi)$. Near the critical point, ξ becomes large – and with it the fluctuations.

3.2.4 Hagedorn phase transition

`http://cds.cern.ch/record/706020/files/0401103.pdf` and `https://cerncourier.com/a/the-tale-of-the-hagedorn-temperature/`.
There is an example of completely different type of phase transition. When people investigated statistics of particle jets they realised then the number of possible states grows with the energy of the system. A simple toy model is called the 'partitions thermodynamics'. Energy of the system is $E = E_0 n$ where n is an integer and the degeneracy of each state is 2^n. The basic idea is that the total energy of the system can be split/partitioned between different elements in $p(n)$ ways. For example, energy of $3E_0$ can be used to create 3 particles of energy E_0, or 1 particle of energy $3E_0$ and so on. This mimics energies of the integer tower of particle resonances. We can generalise this by either forbidding some partitions or adding more degrees of freedom, so $p(n) = \kappa^n$. The partition sum can be computed to be

$$Z = \kappa^n e^{\beta E_0 n} = \frac{\kappa}{\kappa - e^{\beta E_0}} \tag{138}$$

which diverges at $T_H = \frac{E_0}{\ln \kappa}$ which is called the Hagedorn temperature. At this temperature, adding more energy to the system does not increase average energy per particles but adds a new particle instead. Another point of view is that at this critical temperature the combinatorial factor is able to counter-act the Boltzman exponential factor and yield the partition function divergent.

4 Equilibrium statistical mechanics (8)

4.1 Canonical Ensemble, Debye solid

References: Reichl 5.1-5.3, 5.5

4.1.1 Microcanonical ensable

Let us start with the density operator $\hat{\rho}$. What can we tell about the equilibrium state of an isolated system? The density operator has to commute with the

Hamiltonian $\hat{H}$, therefore $\hat{\rho}(\hat{H})$. The Hilbert space of states with energy E is $|E, n\rangle$, where $n = 1, ..., N(E)$; $N(E)$ is the number of states with given energy. Probability to find system in a given state (one of many with the energy) is

$$P_n = \langle E, n|\hat{\rho}|E, n\rangle. \tag{139}$$

The Gibbs entropy

$$S = -k_B \mathrm{Tr}\left(\hat{\rho}\ln\hat{\rho}\right) = -k_B \sum_n P_n \ln P_n. \tag{140}$$

The question is, what distribution of probabilities (given the restriction $\sum_n P_n = 1$ extremize the entropy. This can by obtained by the method of Lagrange multipliers

$$\delta\left(\sum_n \left(\alpha_0 P_n - k_B P_n \ln P_n\right)\right) = \sum \left(\alpha_0 - k_B - k_B \ln P_n\right)\delta P_n = 0, \tag{141}$$

therefore $P_n = \exp(\alpha_0/k_B - 1) = \text{const.}$. The value of α_0 is obtained from the normalisation of probabilities which gives us

$$P_n = \frac{1}{N(E)}, S = k_B \ln N(E). \tag{142}$$

This is the microcanonical ensemble that holds for isolated system with fixed energy. Next we will move to closed systems (fixed number of particles but can exchange energy) and open systems (even particles can be exchanged).

4.1.2 Canonical ensemble

Now the energy is not fixed, but instead

$$\langle E\rangle = \mathrm{Tr}_N\left(\hat{H}_N\hat{\rho}\right), \tag{143}$$

the subscript $_N$ reminds us that we trace only over states with fixed number of particles N. Now we can do the same exercise as with microcanonical ensemble but with additional condition (143), yielding

$$\hat{\rho} = \exp\left(\left(\alpha_0/k_B\right)\hat{I} + \frac{\alpha_E}{k_B}\hat{H}_N\right). \tag{144}$$

which again can be fixed by the conditions.

$$\left(\alpha_0 - k_B\right)\hat{I} + \alpha_E\hat{H}_N - k_B\ln\hat{\rho} = 0. \tag{145}$$

Fixing α_0 is done by defining the partition function

$$Z = \exp\left(1 - \frac{\alpha_0}{k_B}\right) = \mathrm{Tr}\left(\exp\left(\alpha_e\hat{H}_N/k_B\right)\right). \tag{146}$$

Fixing α_E is done by plugging the expression for α_0 back into 145, multiplying with $\hat{\rho}$ and taking the trace

$$\mathrm{Tr}_N\left((\alpha_0 - k_B)\,\hat{\rho} + \alpha_E \hat{H}_N \hat{\rho} - k_B \hat{\rho} \ln \hat{\rho}\right) = -k_B \ln Z + \alpha_E \langle E \rangle + S = 0. \tag{147}$$

Now recall that $A = U - ST$ where $U = \langle E \rangle$. Now this looks the same if we identify $\alpha_E = -\beta = -1/T$ and take

$$A = -k_B T \ln Z. \tag{148}$$

The partition function takes the form

$$Z = e^{-\beta A} = \mathrm{Tr}_N\left(e^{-\beta \hat{H}}\right) \tag{149}$$

and the density operator is

$$\hat{\rho} = e^{-\beta\left(\hat{H} - A\right)} = \frac{e^{-\beta \hat{H}_N}}{\mathrm{Tr}_N\left(e^{-\beta \hat{H}_N}\right)}. \tag{150}$$

4.1.3 Energy fluctuations

We are starting to see some connections with thermodynamics. Let us follow through, we can differentiate the normalisation condition $\mathrm{Tr}\exp \beta(A - \hat{H}) = 1$ twice w.r.t. β to obtain

$$\mathrm{Tr}\left(\left(\frac{\partial^2 \beta A}{\partial \beta^2} + \left(-\hat{H}_N + \left(\frac{\partial \beta A}{\partial \beta}\right)\right)^2\right) e^{\beta(A - \hat{H})}\right) \tag{151}$$

or simply

$$\langle E^2 \rangle - \langle E \rangle^2 = -\left(\frac{\partial^2 \beta A}{\partial \beta^2}\right)_{X,N} = k_B T^2 C_{X,N}. \tag{152}$$

The fluctuations in energy are related to the specific heat! Also, since $\langle E \rangle \sim N \sim C_{X,N}$ we get that

$$\frac{\sqrt{\langle E^2 \rangle - \langle E \rangle^2}}{\langle E \rangle} \sim N^{-\frac{1}{2}}. \tag{153}$$

4.1.4 Semiclassical ideal gas

Imagine N indistinguishable particles, the trace is over all states

$$|k_1, k_2, ..., k_M\rangle^{\pm} = \sum_P (\pm 1)^P |k_1, k_2, ..., k_M\rangle \tag{154}$$

where we sum over all $N!$ permutations and (-1) is a factor depending on the corresponding statistics.

Let us consider only a single particle confined in a box with volume $V = L^3$. The momentum eigenvalues are $\mathbf{k} = \frac{2\pi \mathbf{n}}{L}$. The partition function becomes

$$Z_1(T) = \sum_{\mathbf{n}} \exp\left(-\beta \frac{4\pi^2 \hbar^2}{2mL^2} \mathbf{n}^2\right), \tag{155}$$

which assuming the volume is large enough can be replaced by integration over $\mathbf{p} = \frac{2\pi\mathbf{n}}{L}$, yielding

$$Z_1(T) = \frac{V}{\Lambda_T^3}, \tag{156}$$

where $\Lambda_T = \frac{h}{\sqrt{2\pi m k_B T}}$ is called the thermal wavelength. IF we add second particle, the result is

$$Z_2 = \frac{1}{2!} Z_1^2 \left(1 \pm O(\Lambda^3/V)\right) \tag{157}$$

in the high temperature/law density limit is just the same as for set of independent particles. In the same limit, N particles have the partition function

$$Z_N \approx \frac{1}{N!}\left(\frac{V}{\Lambda_T^3}\right) \approx \left(\frac{eV}{N\lambda_T^3}\right), \tag{158}$$

where we have used the Stirling's approximation $N! \approx (N/e)^N$. The Helmholtz free energy is

$$A = -k_B T \ln Z_N = -N k_B T - N k_B T \ln\left(\frac{V}{N}\left(\frac{h^2}{2\pi m k_B T}\right)^{-3/2}\right) \tag{159}$$

and the entropy is

$$S = -\left(\frac{\partial A}{\partial T}\right) = \frac{5}{2} N k_B + n k_B \ln\left(\frac{V}{N}\left(\frac{\hbar^2}{2\pi m k_b T}\right)^{-3/2}\right), \tag{160}$$

whick is the Sackur-Tetrode equation.

4.1.5 Debye lattice

Some addition notes `https://users.aber.ac.uk/ruw/teach/334/heatcap.php`
A simple model of atomic crystals has the Hamiltonian

$$\hat{H} = \sum_{i=1}^{3N} \frac{\hat{p}_i^2}{2m} + \sum_{(i,j)}^{N(N-1)/2} \kappa_{i,j} \hat{q}_i \hat{q}_j = \sum_{\alpha=1}^{3N} \left(\frac{P_\alpha}{2m} + \frac{m\omega_\alpha^2}{2} \hat{Q}_\alpha^2\right) \tag{161}$$

where we have used the canonical transformation. This is now an oscillator problem with

$$\omega_\alpha^2 = c^2 \left(\left(\frac{\pi l_{x\alpha}}{L_x}\right)^2 + \left(\frac{\pi l_{y\alpha}}{L_y}\right)^2 + \left(\frac{\pi l_{z\alpha}}{L_z}\right)^2\right) \tag{162}$$

and the Hamiltonian

$$\hat{H} = \sum_{\alpha=1}^{3N} \hbar\omega_\alpha \left(\hat{n}_\alpha + \frac{1}{2}\right) \tag{163}$$

then the partition function is

$$Z_N = \sum_{n_1} ... \sum_{n_{3N}} \exp -\beta \left(\sum_{\alpha=1}^{3N} \hbar\omega_\alpha \left(\hat{n}_\alpha + \frac{1}{2} \right) \right). \tag{164}$$

The average energy is

$$\langle E \rangle = \sum_{\alpha=1}^{3N} \left(\frac{\hbar\omega_\alpha}{2} + \frac{\hbar\omega_\alpha}{e^{\beta\hbar\omega_\alpha} - 1} \right). \tag{165}$$

Note that the average occupation numer $\langle \hat{n}_\alpha \rangle = \frac{1}{e^{\beta\hbar\omega_\alpha}-1}$ is the Planck's formula. By figuring out ω we can match this versus the experiment (for example measure C_V) and the results agree very well (better than the Einstein solid model).

Exercise: Compute heat capacity of Debye solid.

4.2 Spin lattices

References: Reichl 5.6.1,5.6.2, Tong 5.2

The spin lattice is one of the most physical systems to understand as it is clear to understand and demonstrate a plethora of important properties and effects. 1D model is easy to solve but has no phase transition. 2D model has a phase transition, is exactly solvable but the solution is complicated (first done by Onsager).

To begin with, consider a single spin to begin with. It can occupy two different states $s = \pm 1$ with corresponding energies $E = \mp \mu B s$. The partition function is

$$Z_1 = \sum_{i=1,2} \exp\left(\beta\mu B s\right) = 2\cosh(\beta\mu B). \tag{166}$$

This was rather simple! If we had N independent spins it is easy to see that $Z_N = Z_1^N$. We can now consider a 1D system of interacting spins. The easiest form of interaction is 'nearest-neighbour' that means that each spin only feels the one next to it, energy of the system has the form

$$E = \varepsilon \sum_{(i,j)} s_i s_j - \mu B \sum_i s_i. \tag{167}$$

What will the system look like? For large β, energy is preferred and all spins will be aligned (to minimise interaction energy) and pointing down (to minimise energy from the magnetic field). In small β limit, energy does not matter so much and the system tends to extremise the entropy (recall the free energy $F = U - TS$.

Before solving the problem in 1D, let us just define the magnetisation as

$$m = \frac{1}{N} \sum_i \langle s_i \rangle = \frac{1}{N\beta} \frac{\partial \ln Z}{\partial B}. \tag{168}$$

Also note that one can think of the Ising model as a gas particles living on lattice with occupation numbers $n_i = 0, 1$ and energy $E = -4J \sum n_i n_j - \mu \sum n_i$ where the first term is due to attraction. This is identical to the Ising model if we take $s_i = 2n_i - 1$. This may be useful in thinking about the problem.

Our goal is now to find the partition function Z by summing over all 2^N spin configurations with energy (167).

4.2.1 Transfer matrix method

(Also solved here `https://stanford.edu/~jeffjar/statmech/lec4.html#solving1`)

Exercise: Compute critical coefficients of 1D Ising model. We will consider a system with periodic boundary conditions (that is the chain is living on a circle). Let us begin with rewriting the partition function as

$$Z = \sum_{s_1} ... \sum_{s_N} \exp\left(\beta \sum_i \left(\varepsilon s_i s_{i+1} + \frac{1}{2}\mu\beta\left(s_i + s_{i+1}\right)\right)\right) \tag{169}$$

Let us now define the matrix

$$P = \begin{pmatrix} e^{\beta(\varepsilon+\mu B)} & e^{-\beta\varepsilon} \\ e^{-\beta\varepsilon} & e^{\beta(\varepsilon-\mu B)} \end{pmatrix} \tag{170}$$

defined as

$$\langle s_i|P|s_{i+1}\rangle = e^{\beta\left(\varepsilon s_i s_{i+1} + \frac{1}{2}\mu B(s_i+s_{i+1})\right)} \tag{171}$$

Now, we are nearly done!

$$\begin{aligned} Z &= \sum_{s_1} ... \sum_{s_N} \langle s_1|P|s_2\rangle\langle s_2|P|s_3\rangle ... \qquad (172) \\ &= \sum_{s_1} \langle s_1|P^N|s_1\rangle = \mathrm{Tr} P^N = \lambda_+^N + \lambda_-^N, \end{aligned}$$

where

$$\lambda_\pm = e^{\beta\varepsilon}\left(\cosh(\beta\mu B) \pm \sqrt{\cosh^2(\beta\mu B) - 2e^{-2\beta\varepsilon}\sinh(2\beta\varepsilon)}\right). \tag{173}$$

In the $N \to \infty$ limit only the larger eigenvalue contributes therefore we have the Gibbs free energy per site is $g(T, B) = -k_B T \ln \lambda_+$ and the order parameter (the magnetisation) is

$$\langle s\rangle = -\frac{\partial g}{\partial \mu B} = \frac{\sinh(\beta\mu B)}{\sqrt{\cosh^2(\beta\mu B) - 2e^{-2\beta\varepsilon}\sinh(\beta\mu B)}} \tag{174}$$

There is no phase transition as the magnetisation always disappear in the $B \to 0$ limit.

4.2.2 Mean field method

This method that works well for number of dimensions higher than 4. The idea is that in calculation of energy for each of the spins we take the value of all other spins to be the same as the mean value of spin. This simplifies the calculation (each spin is independent in the calculation) which we can use to compute the average value of spin (that in fact depends on itself). The result is obtained by the condition of self/consistency.

The Hamiltonian is taken to be

$$\hat{H} = -\sum_i E(\varepsilon, B) s_i, \tag{175}$$

where $E(\varepsilon, B) = \frac{1}{2}\nu\epsilon\langle s\rangle + \mu B$. Now the partition functions simply is

$$Z_N = (2\cosh(\beta E))^N \tag{176}$$

The average spin is

$$\langle s\rangle = Z^{-1}\sum_{s_i} s_i e^{\beta E s_i} = \tanh(\beta E) = \tanh\left(\beta\left(\frac{1}{2}\nu\epsilon\langle s\rangle + \mu B\right)\right) \tag{177}$$

This equation can be "solved" graphically for $B = 0$. The critical point is where the slopes of both curves match at $\langle s\rangle = 0$. This gives the critical temperature as

$$T_c = \frac{\nu\varepsilon}{2k_B}. \tag{178}$$

For $B \neq 0$ there is no phase transition and with increasing temperature the average spin value goes smoothly to zero.

4.3 Scaling and renormalisation

References: Reichl 5.7, 5.8.1

4.3.1 Homogeneous functions and widom scaling

Function is called homogeneous if it satisfies satisfy

$$F(\lambda x) = g(\lambda)F(x). \tag{179}$$

It can be shown that the only possible form is a power law $F(x) = F(1)x^p$. If such function depends on two variables

$$f(\lambda^p, \lambda^q) = \lambda f(x, y) \tag{180}$$

we can turn it into the form $f(x, y) = y^{1/q} f\left(x/y^{p/q}, 1\right)$.

The idea behind Widom scaling is that the free energy can be split into two parts: regular and singular w.r.t. to the critical point

$$g(\varepsilon, B) = g_r(\varepsilon, B) + g_s(\varepsilon, B) \tag{181}$$

where $\varepsilon = (T - T_c)/T_c$. We will assume that the singular part is homogeneous

$$g_s(\lambda^p\varepsilon, \lambda^q B) = \lambda g_s(\varepsilon, B). \tag{182}$$

Now recall the definition of critical coefficients, for example near the critical point the magnetisation goes as

$$M(\varepsilon, B = 0) \sim (-\varepsilon)^\beta. \tag{183}$$

Take (182) and differentiate w.r.t. B to obtain

$$\lambda^q M(\lambda^p \varepsilon, \lambda^q B) = \lambda M(\varepsilon, B) \tag{184}$$

and take $\lambda = (-\varepsilon)^{-1/p}$ and set $B = 0$. Voilá

$$M(\varepsilon, 0) = (-\varepsilon)^{(1-q)/p} M(-1, 0) \tag{185}$$

and therefore $\beta = \frac{1-q}{p}$. Actually, we can obtain all four critical coefficients $\alpha, \beta, \gamma, \delta$ using only two scaling coefficients p, q! Therefore they provide relations between the four coefficients that match the experiment very well (even better than the mean field calculations).

Exercise: Compute rest of the scaling coefficients.

4.3.2 Kadanoff scaling

We can now try to apply the idea of scaling to the Ising model. The Hamiltonian is

$$H = -K \sum_{(ij)}^{\Gamma N/2} S_i S_j - B \sum_{i=1}^{N} S_i, \tag{186}$$

where $S_i = \pm 1$. We will consider $d-$dimensional lattice and split it into blocks of with length of edge L (chosen to have $La \ll \xi$ where a is the length of the lattice site). We have NL^{-d} blocks, spin of each is

$$S_I' = \sum_{i \in I} S_i. \tag{187}$$

Since $La \ll \xi$ spins in a single block are strongly correlated and therefore we can represent them by

$$S_I' = Z S_I, \tag{188}$$

where $S_I = \pm 1$ and $Z = L^y$ where y will be obtained later. The Hamiltonian for the blocks is

$$H = -K_L \sum_{(IJ)}^{\Gamma L^{-d} N/2} S_I S_J - B_L \sum_{I=1}^{NL^{-d}} S_I, \tag{189}$$

which looks the same as before, but with a different range of summation and parameters. The free energy therefore the free energy has to have the form

$$g(\varepsilon_L, B_L) = L^d g(\varepsilon, B). \tag{190}$$

After such transformation the correlation length changes as $\xi_L(\varepsilon_L, B_L) = L^{-1}\xi(\varepsilon, B)$. Therefore, the critical temperature and magnetisation have to change as well. Let us assume $\varepsilon_L = \varepsilon L^x$. Also

$$B \sum_{i=1}^{N} S_i = B \sum_{I=1}^{NL^{-d}} \sum_{i \in I} S_i = B \sum_{I=1}^{NL^{-d}} S_I' = BZ \sum_{I=1}^{NL^{-d}} S_I, \tag{191}$$

therefore $B_L = BZ = L^y B$. The scaling equation therefore is

$$g(L^x \varepsilon, L^y B) = L^d g(\varepsilon, B) \tag{192}$$

which, comparing to Widom scaling, gives $x = pd$ and $y = qd$ (from which we see that $q < 1$).

This again allows us to related different scaling coefficients.

4.3.3 Renormalisation

Important, yet maybe unnoticed, idea of the previous section is that physics of the system as viewed on different scales can be related together. The special case is the critical point, where the self-similar behaviour of the system becomes manifest. The important aspect is how are the coupling coefficients on different scales related to each other. We will try to keep the discussion rather general. Consider the partition function

$$Z(\mathbf{K}, N) = \sum_{S_i} \exp\left(-H(\mathbf{K}, S_i, N)\right). \tag{193}$$

The effective Hamiltonian (with β included)

$$(\mathbf{K}, S_i, N) = K_0 + K_1 \sum_i S_i + K_2 \sum_{ij} S_i S_j + K_3 \sum_{ijk} S_i S_j S_k + ... \tag{194}$$

where K is a vector of coupling constants, for example in the case of the Ising model it is $\mathbf{K} = (0, -\beta B, -\beta J, 0, 0, ...)$. Now we again split the system into S_I blocks (that has spins σ_I inside)

$$Z(\mathbf{K}, N) = \sum_{S_I, \sigma_I} \exp\left(-H(\mathbf{K}, S_I, \sigma_I, N)\right) = \sum_{S_I} \exp\left(-H(\mathbf{K}, S_I, N)\right) = Z(\mathbf{K}_L, NL^{-d}) \tag{195}$$

The coupling coefficients are related by a transformation

$$\mathbf{K}_L = T(\mathbf{K}). \tag{196}$$

This transformation is generates the renormalisation group. Usually this drives us away from the critical point. But if we are sitting on the critical point we have

$$K^* = T(K^*). \tag{197}$$

Let us now simplify the discussion and assume only two non-zero coefficients $\mathbf{K}^* = (K_1^*, K_2^*)$. The transformation T moves the points in 2D space and we are interested how does it move points close to the critical point. Therefore we will consider the distance from it $\delta\mathbf{K}_L = (\mathbf{K}_L - \mathbf{K}^*)$, linearised transformation is

$$\delta\mathbf{K}_L = A\delta\mathbf{K}, \tag{198}$$

the matrix A can be diagonalised $\delta\mathbf{u}_L = \Lambda\delta\mathbf{u}$. The transformation acts on the eigenvectors as

$$\begin{aligned}\delta u_{nL,1} &= \lambda_1^n \delta u_1, \\ \delta u_{nL,2} &= \lambda_2^n \delta u_2.\end{aligned} \tag{199}$$

If $\lambda < 1$ the point flows toward the critical point, if $\lambda > 1$ it flows away. The eigencurves with $\lambda < 1$ will become critical after sufficient number of transformations, such system is said to exhibit "universality". The eigenvalue $\lambda > 1$ is called relevant and its eigenvector is identified as a physical quantity (ε or B for example) that measures the distance of the system from the critical point.

Now the singular part of the free energy density can be written as

$$g_s(\delta \mathbf{u}) = L^{-d} g_s(\lambda_i \delta u_i) \tag{200}$$

this looks like the Widom scaling! For Ising system we expect two relevant parameters and therefore two constants $\lambda_{1,2} > 1$. Those can be observed to be $\delta u_1 = \varepsilon$ and $\delta u_2 = B$. Therefore,

$$\begin{aligned}\lambda_1 = (L^d)^p \quad &\rightarrow \quad p = \frac{\ln \lambda_1}{d \ln L} \\ \lambda_2 = (L^d)^q \quad &\rightarrow \quad q = \frac{\ln \lambda_1}{d \ln L}.\end{aligned} \tag{201}$$

Therefore we have related the critical exponents to the relevant eigenvalues.

4.4 Numerical solutions, Metropolis algorithm

Some textbooks `https://link.springer.com/content/pdf/10.1007%2Fb100712.pdf` This one is rather useful for high-energy physics `https://arxiv.org/abs/1808.08490`.

4.4.1 Importance sampling

We have studied various neat tricks and approaches to study the Ising model in various number of dimensions. In general, we relied (often implicitly) on some nice properties of the model. Analytical solution often work under restrictive assumptions and therefore might not describe the real situation properly. You may wonder: why bother? We know the expression for Z, why do not we just compute it directly? The problem is obvious, there are way too many configurations, 2^N in the case of the Ising model (just remember that $2^{10} \approx 10^3$).

However, consider the following challenge: sum this list of numbers within 1% precision:

$$\begin{aligned} S = \quad & 5.18471 \times 10^{21} + 1.06865 \times 10^{13} + 1.06865 \times 10^{13} \\ & +22026.5 + 1.06865 \times 10^{13} + 22026.5 + 22026.5 + 0.0000453999 \\ & +1.06865 \times 10^{13} + 22026.5 + 22026.5 + 0.0000453999 + 22026.5 \\ & +0.0000453999 + 0.0000453999 + 9.35762 \times 10^{-14} + 1.06865 \times 10^{13} \\ & +22026.5 + 22026.5 + 0.0000453999 + 22026.5 + 0.0000453999 + 0.0000453999 \\ & +9.35762 \times 10^{-14} + 22026.5 + 0.0000453999 + 0.0000453999 \\ & +9.35762 \times 10^{-14} + 0.0000453999 + 9.35762 \times 10^{-14} \\ & +9.35762 \times 10^{-14} + 1.92875 \times 10^{-22}. \end{aligned}$$

Easy, right? The answer is just the first term $S \approx 5.18471 \times 10^{21}$ (and the precision is way better than 1%). As it happens, those numbers correspond to the partition function of the noninteratcting Ising model with $N = 5$ spins with effective coupling constant (βB) equal to 10, that is $Z = \sum_{s_i} \exp(-10 s_i)$. This either means that the temperature is very low or that the coupling to the magnetic field is very large. We can obtain the value of partition sum only by looking at a very small (in this case one) state that dominates the summation.

The idea of importance sampling is to utilise this – can we scan the relevant part of configuration space in an accessible way? (That is within a reasonable amount of (computational) time?) Let us step back for a moment.

4.4.2 Monte Carlo methods

Numerical methods are computational methods that are often intended to be run on a computer. Some are deterministic, that means every time the code is run the same result is reached. For example, we can consider numerical integration of a differential equations. The other approach is to include some kind of randomness, therefore it is hard to reach the same result (unless we use the same set of random number) but results of separate run should agree reasonably well.

Imagine we want to compute the following integral

$$I = \int\limits_{-\inf}^{\inf} \frac{1}{1+x^2} dx \tag{202}$$

The first step is to replace the integration with a sum as computer can only deal with discrete numbers, therefore we take $x = i\varepsilon$ and let i run over all whole numbers. This is again problematic as computers do only have limited memory. Therefore we can take $i = -i_{min}, ..., i_{max}$, which approximates $\int\limits_{-i_{min}\varepsilon}^{i_{max}\varepsilon} \frac{1}{1+x^2} dx$. Now there are two sources of inaccuracies that need to be tuned individually (range of integration and integration step ε). Now we have, for example

$$I_N = \sum_{i=-10^4}^{i=10^4} \frac{1}{1+(i/10)^2} 10^{-1}, \tag{203}$$

where we took care of the fact that $dx \approx \varepsilon$. The exact result is $I = \pi$, this way we have obtained $I_N = 3.1395$, close enough!

Now consider a fact that looks deep but is rather trivial. One can think of any integral over finite region as computing a mean value with uniform probability, for example

$$I = \int\limits_{-L/2}^{L/2} \frac{L}{1+x^2} \frac{dx}{L} = \left\langle \frac{L}{1+x^2} \right\rangle_{uniform}. \tag{204}$$

So there is another way to compute the integral.

- Generate a list of random numbers $-L/2 < r_i < L/2, i = 1, ..., N$.

- Compute a new list $R_i = \frac{L}{1+r_i^2}, i = 1, ..., N$.
- Compute the mean value of this list $I = \frac{1}{N}\sum\limits_i R_i$.

With $L = 2000$ and $N = 10^5$ I've got $I = 3.143 \pm 0.019$. (I have ran the simulation 100 times to compute the error estimate). This would be called the Monte Carlo method (it is an umbrella term for various randomness-involving methods.)

In physics we often don't need to force the probability, it is inherently part of the problem. We often have to compute integrals of the form

$$\langle f \rangle = \int f(x)p(x)dx, \tag{205}$$

or their multidimensional variants.

Now we can do this numerically by changing the integral into a sum or we can generate uniformly-distributed random numbers r_i and then take mean over the list of $f(r_i)p(r_i)L$. The factor $p(r_i)$ just acts as a weighting factor. We could achieve the same result by generating random numbers $(r_p)_i$ with the distribution p and then compute only mean of the list $f((r_p)_i)$.

How to generate numbers with given distribution? On option is to know the inverse function p^{-1} and then generate random numbers $0 \leq r \leq 1$ and have $p^{-1}(r)$ with the desired distribution. However, inverting the probability function can be difficult. However, there is another – perfectly straightforward – way of achieving this.

4.4.3 Metropolis algorithm

The Metropolis agorithm is a simple way of obtaining a list of random numbers r_p with that have the probability distribution of the form $p(x) = Z^{-1}e^{-S(x)}$, where $S(x)$ is a function which we will call the action here. The algorithm starts with a random number x_1 and then goes over a loop

- store the current number $x_{old} := x_i$ and generate a new one x_{new},
- compute $\Delta S = S(x_{new} - x_{old})$,
- if $\delta S < 0$ we keep the new configurations, $x_{i+1} = x_{new}$ and go back to (1),
- If not, we give it a chance anyway and compare $e^{\Delta S}$ with a random number $0 \leq r \leq 1$,
- if it is larger we keep it anyway $x_{i+1} = x_{new}$, otherwise we restore the old value $x_{i+1} = x_{old}$,
- go to (1).

As a result we will obtain the list of random numbers (Markov chain) $(r_p)_i = \{x_1, ..., x_n\}$ with the desired probability distribution p. To compute $\langle f \rangle$ we just need to take the mean of $\{f(x_1), ..., f(x_n)\}$.

One can think of this procedure as a walk that is partially random but partially guided. We can interpret the action $S(x)$ as the energy of the system. During each loop, the system tries to move to a new position. If the energy goes down it can always do it. If it goes up it can do it, but larger the energy step the smaller the probability. We can interpret it as that the system needs a random 'kick' from the environment and large 'kicks' are very improbable.

We often think of the Markov chain as a temporal sequence and each step (loop) is referred to a single unit of Monte-Carlo time.

The image of a particle moving in a potential defined by $S(x)$ with the help of random (temperature) impulses from the environment is helpful in seeing the limitations of this method. Firstly, it can take some time for the system to walk over the important part of the configuration space. Also, if two regions are separated be either infinite or very high potential wall, it is (nearly) impossible for the system to get there.

On the other hand, using this method, we didn't have to specify anything (such the integration range and so on). There are very few parameters which need to be tuned (we often seek new states of the form $x_{i+1} = x_i + \varepsilon$. Small ε makes the walk slow, large ε makes the rejection rate high, one need to find a balance.

The great thing is that this method works equally well (yet slower) in higher dimensional cases and works for discrete system as well – for example the Ising model!

4.5 Grand Canonical Ensemble

References: Reichl 6.2, 6.3, 6.5

4.5.1 Grand Canonical Ensemble

We will begin with the same exercise as before, what probability distribution maximize the entropy given the restriction

$$\mathrm{Tr}\rho = \frac{\mathrm{Tr}\hat{H}\rho}{\langle E\rangle} = \frac{\mathrm{Tr}\hat{N}\rho}{\langle N\rangle} = 1. \tag{206}$$

First we have by by varying $\delta\rho$ the equation

$$(\alpha_0 - k_B) + \alpha_E \hat{H} + \alpha_N \hat{N} - k_B \ln\hat{\rho} = 0 \tag{207}$$

$$Z = \exp\left(\alpha_0/k_B - 1\right) = \mathrm{Tr}\left(\exp\left(\alpha_E \hat{H}/k_B + \alpha_N \hat{N}/k_B\right)\right). \tag{208}$$

From (207) by multiplying with ρ and taking trace we obtain

$$-k_B \ln Z + \alpha_E \langle E\rangle + \alpha_N \langle N\rangle + S = 0. \tag{209}$$

We can compare this with the grand potential $\Omega = U - TS - \mu N$ to make the identification

$$\Omega = -k_B T \ln Z \tag{210}$$

and

$$\rho = e^{-\beta\left(\hat{H}-\mu\hat{N}-\Omega\right)} = \frac{e^{-\beta\left(\hat{H}-\mu\hat{N}\right)}}{\mathrm{Tr}\left(e^{-\beta\left(\hat{H}-\mu\hat{N}\right)}\right)} \tag{211}$$

And as before, the fluctuations can be shown to behave as $\frac{\sqrt{\langle N^2\rangle - \langle N\rangle^2}}{\langle N\rangle} \sim N^{-1/2}$.

4.5.2 Chemical potential

Are you not perfectly happy with the notion of chemical potential? Try reading this `https://www.tf.uni-kiel.de/matwis/amat/def_en/kap_2/advanced/t2_4_1.pdf`

Important parameter that makes is the chemical potential μ. It is defined as the energy required/obtained to/by add/removing a particle into/from the system. How to compute it? Well, that depends on the situation. For example in semiconductor each particle can occupy a certain energy levels and μ is the cost to add that particle. For monatomic ideal gas we have computer the Sackur-Tetrode molar entropy to be

$$S = \frac{5}{2}R + R\ln\left(\frac{T^{5/2}}{P}\right) + R\ln\left(\left(\frac{2\pi m}{h^2}\right)^{3/2} k_B^{5/2}\right) \tag{212}$$

from Gibbs-Duhem equation $SdT + XdY + N_j d\mu_j = 0$ we have $(\partial\mu/\partial T)_P = -s$ and $(\partial u/\partial P)_T = \nu = RT/P$. We can integrate those to obtain

$$\mu_{\text{molar}} = -RT\ln\left(\frac{(k_BT)^{5/2}}{P}\left(\frac{2\pi m}{h^2}\right)^{3/2}\right) \tag{213}$$

Also recall that we have previous defined $\lambda_T = \frac{h}{\sqrt{2\pi m k_B T}}$.

$$\mu = -k_BT\ln\left(\frac{k_BT}{P_M\lambda_T^3}\right). \tag{214}$$

4.5.3 Ideal classical gas

Hamiltonian of a ideal gas is $\hat{H}_0 = \sum\limits_{j=1}^{N} \hbar^2\mathbf{k}_j^2/(2m)$. In the classical limit of very low density gas (each particle occupies its own state) we have for indistinguishable particles

$$\begin{aligned} Z &= \sum_N \frac{1}{N!}\sum_{\mathbf{k}_j}\langle\mathbf{k}_1,\mathbf{k}_2,...|e^{-\beta\left(\hat{H}_0-\mu\hat{N}\right)}|\mathbf{k}_1,\mathbf{k}_2,...\rangle \\ &= \sum_N \frac{e^{\beta\mu N}}{N!}\frac{V^N}{\lambda_T^{3N}} = \exp\left(e^{\beta\mu}\frac{V}{\lambda_T^3}\right). \end{aligned} \tag{215}$$

The pressure is obtained directly from the Grand sum

$$P = -\frac{\Omega}{V} = \frac{k_BT\ln Z}{V} = k_BT\frac{e^{\beta\mu}}{\lambda_T^3}. \tag{216}$$

The particle number can be also obtained from it as

$$\langle N\rangle = -\left(\frac{\partial\Omega}{\partial\mu}\right)_{T,V} = \frac{Ve^{-\beta\mu}}{\lambda_T^3}. \tag{217}$$

By removing the factor $\frac{Ve^{-\beta\mu}}{\lambda_T^3}$ we obtain

$$PV = \langle N\rangle k_BT. \tag{218}$$

4.5.4 Adsorption

Adsorption is the process when something (molecules) stick to the surface of something else (perhaps also molecules), for example pollutants stick to the surface of activated carbon (activated means it has a lot of pores and therefore large surface area). Another example is an oxygen molecule sticking to to myoglobin or hemoglobin (the first on is in muscles the other in blood).

Myoglobin binds one oxygen molecule, hemoglobin up to four, so we will discuss myoglobin for the sake of simplicity. Our atmosphere is 21% oxygen so its partial pressure is $P_{O_2} = 0.21\text{atm}$. When molecules join they release energy ε (the binding energy). The system (myoglobim + atmosphere) will be in equilibrium when their chemical potentials equal, therefore we can take $\mu = -k_B T \ln\left(\frac{k_B T}{P_m \lambda_T^3}\right)$.

The partition function for a system of N distinguishable myoglobin molecules is

$$Z(T,\mu) = \sum_{n=0}^{N} \frac{N!}{(N-n)!n!} \exp\left(-\beta\left(-n\varepsilon - n\mu\right)\right) = \left(1 + e^{\beta(\varepsilon+\nu)}\right)^N , \tag{219}$$

where we can identify the partition function for each single binding site. The fraction of (un)occupied sites is

$$f_0 = \frac{1}{1+e^{\beta(\varepsilon+\mu)}}, f_1 = \frac{e^{\beta(\varepsilon+\mu)}}{1+e^{\beta(\varepsilon+\mu)}}, \tag{220}$$

which after plugging in the expression for μ we obtain

$$f_{occupied} = \frac{P_{O_2}}{P_{O_2} + \frac{k_B T}{\lambda_T^3} e^{-\varepsilon/T}} \tag{221}$$

which is called the Langmuir adsorption equation (holds for general case, not only oxygen).

Exercise: How would the possibility of adsorbing two (or four) molecules change the result?

4.5.5 Blackbody radiation

Useful lecture on this topic `https://scholar.harvard.edu/files/schwartz/files/11-phononsphotons.pdf` We can think of the blackbody radiation in terms of grand canonical ensemble with $\mu = 0$. The partition function is for system in box $V = L^3$ and photon energies $\hbar\omega_i = \hbar c|\mathbf{k}|_i$ is

$$Z = \sum_{n_1=0}^{\infty} .... \sum_{n_\infty=0}^{\infty} \exp\left(-\beta \sum_{i=1}^{\infty} n_i \hbar\omega_i\right) = \prod_{i=1}^{\infty} \frac{1}{1-e^{-\beta\hbar\omega_i}} \tag{222}$$

so the grand potential is

$$\Omega = -PV = k_B T \sum_{i=1}^{\infty} \ln\left(1 - e^{-\beta\hbar\omega_i}\right) . \tag{223}$$

Note that for $\mu = 0$ we have $\Omega = U - TS$ where $S = -\left(\frac{\partial \Omega}{\partial T}\right)_P$ so we can write

$$U = \left(\frac{\beta \partial \Omega(T)}{\partial \beta}\right)_V = \sum_{i=1}^{\infty} \hbar \omega_i n(\omega_i) = \sum_{i=1}^{\infty} \frac{\hbar \omega_i}{e^{\beta \hbar \omega_i} - 1} \tag{224}$$

The allowed frequencies in a box are $\omega^2 = c^2 k^2 = c^2 \left(\left(\frac{n_x \pi}{L}\right)^2 + \left(\frac{n_y \pi}{L}\right)^2 + \left(\frac{n_z \pi}{L}\right)^2\right)$. We can think of states occupying a lattice with volume per point $(c\pi/L)^3$. Number of points with energy (frequency) small than some value is equal to the volume in the phase space times the density (one eight of a sphere)

$$\nu = \frac{1}{8}\frac{4}{3}\pi \omega^3 \left(\frac{L}{c\pi}\right)^3 = \frac{L^3 \omega^3}{6c^3 \pi^2} \tag{225}$$

if we assume that the lattice is dense enough and don't forget about two possible polarisations we can replace the sum over states with 2 times the integral to obtain the pressure

$$P = -\frac{k_B T}{\pi^2 c^3} \int_0^{\infty} \omega^2 d\omega \ln\left(1 - e^{-\beta \hbar \omega}\right) \tag{226}$$

which by integrating p.p. gives

$$P = \frac{a}{3} T^4, a = \frac{\pi^2 k_B^4}{15 c^3 \hbar^3}. \tag{227}$$

Similarly we can find the internal energy

$$U = \frac{L^3}{\pi^2 c^3} \int_0^{\infty} \omega^2 d\omega \frac{\hbar \omega}{(e^{\beta \hbar \omega} - 1)} = aVT^4. \tag{228}$$

The energy density (per volume) for each frequency component is

$$u(\omega) = \frac{\omega^2}{\pi^2 c^3} \frac{\hbar \omega}{(e^{\beta \hbar \omega} - 1)}. \tag{229}$$

4.6 Ideal Bose-Einstein quantum gases

4.6.1 Quantum gases

References: Reichl 6.6,6.7 Again, the same setting: particles in a box, now $V = L_x L_y L_z$, the wave numbers are $\mathbf{k}_\mathbf{l} = \sum_{i=1}^{3} \frac{2\pi l_i}{L_i} \mathbf{e}_i$ and $\mathbf{p} = \hbar \mathbf{k}$ (note that $-\infty < l_i < \infty$). The energies are

$$\varepsilon_l = \frac{\hbar^2}{2m} \sum_{\mathbf{l}} \frac{4\pi^2 l_i^2}{L_i^2}. \tag{230}$$

It is convenient to switch to the number representation where distinction between Bose and Fermi statistics is very clear

$$Z(T, V) = \prod_{\mathbf{l}} \left(\sum_{n_\mathbf{l}}\right) \langle \{n_\mathbf{l}\} | e^{-\beta\left(\hat{H}_0 - \mu \hat{N}\right)} | \{n_\mathbf{l}\} \rangle \tag{231}$$

Now we have two different cases. In case of Bose-Einstein system, any numbers of particles can occupy the same state. In case of Fermi-Dirac statics, only one at most. Therefore

$$\begin{aligned} Z_{\mathrm{BE}} &= \prod_{\mathbf{l}} \left(\sum_{n_{\mathbf{l}}=0}^{\infty} \exp\left(-\beta n_{\mathbf{l}}(\varepsilon_{\mathbf{l}} - \mu)\right) \right) \qquad (232) \\ Z_{\mathrm{FD}} &= \prod_{\mathbf{l}} \left(\sum_{n_{\mathbf{l}}=0}^{1} \exp\left(-\beta n_{\mathbf{l}}(\varepsilon_{\mathbf{l}} - \mu)\right) \right) \end{aligned}$$

We will begin with the first one.

4.6.2 Bose Einstein gas

Recall that $\sum\limits_{n} x^n = (1-x)^{-1}$. In dealing with statistical sums it always useful to keep in mind the huge list of *sum* identities present in mathematical literature. This gives the partition function

$$Z(T, V, \mu) = \prod_{\mathbf{l}} \frac{1}{1 - e^{-\beta(\varepsilon_l - \mu)}} \qquad (233)$$

and the grand potential

$$\Omega(T, V, \mu) = -k_B T \log Z = k_B T \sum_{\mathbf{l}} \ln\left(1 - e^{-\beta(\varepsilon_{\mathbf{l}} - \mu)}\right). \qquad (234)$$

From this we can obtain the mean value of N

$$\langle N \rangle = -\left(\frac{\partial \Omega}{\partial \mu}\right)_{T,V} = \sum_{\mathbf{l}} \frac{1}{e^{\beta(\varepsilon_l - \mu)} - 1} = \sum_{\mathbf{l}} \langle n_{\mathbf{l}} \rangle \qquad (235)$$

where

$$\langle n_{\mathbf{l}} \rangle = \frac{1}{e^{\beta(\varepsilon_l - \mu)} - 1} = \frac{z}{e^{\beta\varepsilon_l} - z} \qquad (236)$$

where the quantity $z = e^{\beta\mu}$ is called the fugacity. (Wiki: In chemical thermodynamics, the fugacity of a real gas is an effective partial pressure which replaces the mechanical partial pressure in an accurate computation of the chemical equilibrium constant. It is equal to the pressure of an ideal gas which has the same temperature and molar Gibbs free energy as the real gas.)

BE gas undergoes a phase transition at low temperatures, we should be able to understand it from we know so far. From (236) we see that $0 \leq z \leq 1$ (otherwise we could have infinite result). Therefore, $\mu = k_B T \ln(z)$ has to negative. As we decrease temperature, $e^{\beta\varepsilon}$ goes to infinite for all $\omega > 0$ and z goes to 1. However, for the state of zero energy $e^{\beta\varepsilon} = 1$ and therefore the expected occupation number becomes macroscopically large

$$\langle n_0 \rangle = \frac{z}{1 - z}. \qquad (237)$$

Now we want to compute the sums, $\langle N \rangle = \sum \langle n \rangle$. The idea is as usual, replace sums with integrals. However, we have to be careful about the zero-energy point, we will treat it separately ($\sum \to \int g(\varepsilon) d\varepsilon, g(\varepsilon) = \frac{m^{3/2} V \varepsilon^{1/2}}{2\pi^2 \hbar^3}$). We will assume $L_x > L_y > L_z$.

$$\begin{aligned}\langle N\rangle &= \frac{z}{1-z}+\frac{Vm^{3/2}}{\sqrt{2}\pi^2\hbar^3}\int\limits_{\frac{4\pi^2\hbar^2}{2mL_x^2}}^{\infty} d\varepsilon\varepsilon^{1/2}\frac{z}{\exp(\beta\varepsilon)-z} \\ &= \frac{z}{1-z}+\frac{4\pi V}{(2\pi\hbar)^3}\int\limits_{2\pi\hbar/L_x}^{\infty} p^2dp\frac{z}{\exp(\beta p^2/(2m))-z}\end{aligned} \tag{238}$$

in a similar manner

$$\Omega = k_BT\ln(1-z)+\frac{4\pi Vk_bT}{(2\pi\hbar)^3}\int\limits_{2\pi\hbar/L_x}^{\infty} p^2dp\ln(1-z\exp(-\beta p^2/(2m))). \tag{239}$$

Note that in the $L\to\infty$ limit we can move the lower integration bound to 0. It is convenient to move to dimensionless variables, $x^2=\beta p^2/(2m)$, then

$$P=\frac{-\Omega}{V}=-\frac{k_BT}{V}\ln(1-z)+\frac{k_BT}{\lambda_T^3}g_{5/2}(z) \tag{240}$$

where

$$g_{5/2}(z)=-\frac{4}{\sqrt{\pi}}\int\limits_0^{\infty} x^2dx\ln(1-ze^{-x^2})=\sum_{\alpha=1}^{\infty}\frac{z^\alpha}{\alpha^{5/2}}. \tag{241}$$

Similarly, for the average particle density

$$\langle n\rangle=\frac{\langle N\rangle}{V}=\frac{1}{V}\frac{z}{1-z}+\frac{1}{\lambda_T^3}g_{3/2}(z), \tag{242}$$

where

$$g_{3/2}(z)=\sum_{\alpha=1}^{\infty}\frac{z^\alpha}{\alpha^{3/2}}. \tag{243}$$

We are interested in the $V\to\infty, z\to 1$ limit. In this limit we can write $z=1-1/(n_0V)$. Two important functions in this limit are $\frac{\ln(1-z)}{V}\to 0$ and $\frac{1}{V}\frac{z}{1-z}\to n_0$. That means the contribution to the pressure can be neglected but the one to $\langle n\rangle$ cannot.

So in the infinite volume limit

$$P=\frac{k_BT}{\lambda_T^3}g_{5/2}(z) \tag{244}$$

for any value of z. However, the particle density goes as

$$\langle n\rangle=\begin{cases}\frac{k_BT}{\lambda_T^3}g_{3/2}(z) & \text{for } z<1\\ n_0+\frac{k_BT}{\lambda_T^3}g_{3/2}(1) & \text{for } z=1\end{cases} \tag{245}$$

The growth of $\langle n\rangle\lambda_T^3$ is driven by the $g_{3/2}$ until very close to $z=1$ where the n_0 becomes dominant, this is the BE condensation.

In the high temperature limit we can keep just a couple of few terms of expansion of $g_{3/2}$ and $g_{5/2}$ as z is small to obtain

$$\begin{aligned} \langle n \rangle &= \frac{1}{\lambda_T^3}\left(z + \frac{z^2}{2^{3/2}} + ...\right) \\ P &= \frac{k_B T}{\lambda_T^3}\left(z + \frac{z^2}{2^{5/2}} + ...\right) \end{aligned} \tag{246}$$

we can use this to compute $z(\langle n \rangle)$, for example keeping only the linear term to obtain $P = \langle n \rangle k_B T$.

The order parameter for this transition is the fraction

$$\eta = \frac{n_0}{\langle n \rangle}. \tag{247}$$

We can write $n_0 = \langle n \rangle - \frac{k_B T}{\lambda_T^3} g_{5/2}(1)$. The phase transition appears as $z \to 1$ where we have $\langle n \rangle \lambda_{T_c}^3 = g_{3/2}(1)$ were we have denoted the critical temperature T_c. Recall that $\lambda_T \sim T^{-1/2}$. From this we have

$$\eta = 1 - \frac{g_{3/2}(1)}{\langle n \rangle \lambda_T^3} = 1 - \left(\frac{\lambda_{T_c}}{\lambda_T}\right)^3 = 1 - \left(\frac{T}{T_c}\right)^{3/2} \tag{248}$$

The critical density as a function of temperature is

$$\langle n_c \rangle = \frac{1}{\langle v_c \rangle} = \frac{g_{3/2}(1)}{\lambda_T^3} \approx 2.612 \left(\frac{m k_B T}{2\pi\hbar^2}\right)^{3/2} \tag{249}$$

and the other way around, the critical temperature as a function of density is

$$T_c = \left(\frac{2\pi\hbar^2}{m K_B}\right)\left(\frac{\langle n \rangle}{2.612}\right)^{2/3} \tag{250}$$

where have used the fact that $g_{3/2} = \zeta(3/2)$ which is the well-known Riemann zeta function.

From the expression for pressure 244 we can obtain the entropy as

$$s = S/V = V^{-1}\left(\frac{\partial P}{\partial T}\right)_{V,\mu} = \begin{cases} k_B \frac{5}{2\lambda_T^3} g_{5/2}(z) - k_B \langle n \rangle \ln z & \text{for } z < 1 \\ k_B \frac{5}{2\lambda_T^3} g_{5/2}(1) & \text{for } z = 1 \end{cases} \tag{251}$$

To compute c_n we need to use $\left(\frac{\partial \beta\mu}{\partial T}\right)_n = -\frac{3}{2T}\frac{g_{3/2}(z)}{g_{5/2}(z)}$. From this we have

$$c_n = T\left(\frac{\partial s}{\partial T}\right)_n = \begin{cases} k_B \frac{15}{4\lambda_T^3} g_{5/2}(z) - k_B \langle n \rangle \frac{9}{4}\frac{g_{3/2}(z)}{g_{1/2}(z)} & \text{for } z < 1 \\ k_B \frac{15}{4\lambda_T^3} g_{5/2}(1) & \text{for } z = 1 \end{cases} \tag{252}$$

4.7 Bogoliubov mean field theory

References: Reichl 6.8

4.7.1 Short reminder of the scattering theory

In quantum mechanics (in c.o.m. system) the scattering Schrödinger equation for two identical particles takes the form

$$\left(\Delta + k^2\right)\psi_k(\mathbf{r}) = \frac{m}{\hbar^2}V(\mathbf{r})\psi_k(\mathbf{r}). \tag{253}$$

The solution is

$$\psi_k(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} + \frac{m}{\hbar^2}\int d\mathbf{r}' G_k(\mathbf{r},\mathbf{r}')V(\mathbf{r}')\psi_k(\mathbf{r}'), \tag{254}$$

where G_k is the green function now in the form $G_k(\mathbf{r},\mathbf{r}') = -\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}$. The solution is, in the asymptotic region where we can take $|\mathbf{r}-\mathbf{r}'| \approx r - \mathbf{r}\cdot\mathbf{r}'/r + ...$

Now we focus on case of two hard-sphere particles with radius a and energies low enough so that the wavelength is way larger than their radius. The energy eigenstates are a superposition of incoming and scattered waves

$$\psi_k(\mathbf{r}) = e^{ikz} + f(\theta)\frac{e^{ikr}}{r} \tag{255}$$

where $f(\theta)$ is the scattering amplitude $f(\theta) = -\frac{m}{4\pi\hbar^2}\int d\mathbf{r}' e^{-i\mathbf{k}'\cdot\mathbf{r}}V(\mathbf{r}')\Psi_k(\mathbf{r}')$. The boundary condition for a hard sphere is $\psi(a,\theta) = 0$ which fixes $f(\theta) = -a$.

Another useful potential is the contact potential $V(\mathbf{r}) = g\delta(\mathbf{r})$ which gives the scattering amplitude $f(\theta) = -\frac{m}{4\pi\hbar^2}g + ...$ Comparing with the hard-sphere potential we have $g = \frac{4\pi\hbar^2 a}{m}$. This works as a reasonably good simplification.

4.7.2 Bogoliubov mean field theory

The Hamiltonian for an interacting system takes in the number representation the following form

$$\hat{H} = \sum_{\mathbf{k}} \frac{\hbar^2 k^2}{2m}\hat{a}^+_{\mathbf{k}}\hat{a}_{\mathbf{k}} + \frac{1}{2V}\sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4} \langle \mathbf{k}_1,\mathbf{k}_2|\hat{V}|\mathbf{k}_3,\mathbf{k}_4\rangle \hat{a}^+_{\mathbf{k}_1}\hat{a}^+_{\mathbf{k}_2}\hat{a}_{\mathbf{k}_3}\hat{a}^+_{\mathbf{k}_4}. \tag{256}$$

Where $\hat{a}, \hat{a}^+$ are the usual creation/annihilation operators. In the case of BEC we can approximate the interaction to be contact, $V(\mathbf{r}) = g\delta(\mathbf{r})$ which makes the Hamiltonian

$$\hat{H} = \sum_{\mathbf{k}} \frac{\hbar^2 k^2}{2m}\hat{a}^+_{\mathbf{k}}\hat{a}_{\mathbf{k}} + \frac{g}{2V}\sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4} \delta_{\mathbf{k}_1+\mathbf{k}_2-\mathbf{k}_3-\mathbf{k}_4}\hat{a}^+_{\mathbf{k}_1}\hat{a}^+_{\mathbf{k}_2}\hat{a}_{\mathbf{k}_3}\hat{a}^+_{\mathbf{k}_4}. \tag{257}$$

This problem is difficult to solve. However, we are interested in BEC system at very love temperatures $T \approx 0K$. That means most of the particles are in the lowest energy state. Only $\Delta N = N - N_0$ are excited. To utilise this we will replace $\hat{a}_0 = \hat{a}^+_0 = \sqrt{N_0}$ and keep only up to quadratic terms in $\hat{a}_k, \hat{a}^+_k$ for $k > 1$.

This turns the Hamiltonian into

$$\hat{H} = \frac{gN_0^2}{2V} + \frac{1}{2}\sum_{\mathbf{k}\neq 0}\left((\varepsilon_k + \nu)\left(\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}} + \hat{a}_{-\mathbf{k}}^{+}\hat{a}_{-\mathbf{k}}\right) + \Delta\left(\hat{a}_{\mathbf{k}}^{+}\hat{a}_{-\mathbf{k}}^{+} + \hat{a}_{-\mathbf{k}}\hat{a}_{\mathbf{k}}\right)\right), \quad (258)$$

where $\varepsilon_k = \frac{\hbar^2 k^2}{2m}, \nu = 2gN_0/V, \Delta = gN_0/V$. (Recall that $N_0 \sim \Delta$ is the order parameter.) The particle number is $\hat{N} = N_0 + \sum\limits_{\mathbf{k}\neq 0}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}$.

The grand potential is

$$\Omega = -k_B T \ln\left(\mathrm{Tr}\left(e^{-\beta(K_0 + \hat{K})}\right)\right) \quad (259)$$

where $K_0 = gN_0^2/(2V) - \mu N_0$ and

$$\hat{K} = \sum_{\mathbf{k}}\left((\varepsilon_k + \nu - \mu)\,\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}} + \frac{\Delta}{2}\left(\hat{a}_{\mathbf{k}}^{+}\hat{a}_{-\mathbf{k}}^{+} + \hat{a}_{-\mathbf{k}}\hat{a}_{\mathbf{k}}\right)\right). \quad (260)$$

The usual effort is to try to diagonalize the operator, in this case this can be done using the Bogoliubov transformation

$$\hat{a}_{\mathbf{k}}^{+} = u_k \hat{b}_{\mathbf{k}}^{+} - v_k \hat{b}_{-\mathbf{k}} \quad (261)$$

where $\hat{b}, \hat{b}^{+}$ satisfy the ordinary oscillator relations and are called bogolon creation operators (create and annihilate excitations in BEC). This is allowed only if $u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2 = 1$. After the transformation we have

$$\hat{K} = \sum_{\mathbf{k}\neq 0} E_k \hat{b}_{\mathbf{k}}^{+}\hat{b}_{\mathbf{k}} + \frac{1}{2}\sum_{\mathbf{k}\neq 0}\left(E_k - (\varepsilon_k + v - \mu)\right). \quad (262)$$

where $u_k = \frac{1}{\sqrt{2}}\sqrt{\frac{\varepsilon_k + v - \mu}{E_k} + 1}$, $v_k = \frac{1}{\sqrt{2}}\sqrt{\frac{\varepsilon_k + v - \mu}{E_k} - 1}$ and $E_k = \sqrt{(\varepsilon_k + v - \mu)^2 - \Delta^2}$.

The BEC has no gap in the spectrum and for that to be true the chemical potential has to be $\mu = v - \Delta$ which makes the energy $E_k = \sqrt{(\varepsilon_k + \Delta)^2 - \Delta^2}$, also the expressions for u, v are simplified. Note that at low energies $E_k \sim k$ and at high energies $E_k \sim k^2$ (photon-like and particle-like). The grand potential now is

$$\Omega = -k_B T \ln\left(\mathrm{Tr}\left(e^{-\beta K_0}\right)\mathrm{Tr}\left(e^{-\beta \sum\limits_{\mathbf{k}\neq 0} E_k \hat{b}_k^{+}\hat{b}_k}\right)\mathrm{Tr}\left(e^{-(\beta/2)\sum\limits_{\mathbf{k}\neq 0} E_k - (\varepsilon_k + \Delta)}\right)\right). \quad (263)$$

Now the third term on the r.h.s. is problematic (UV divergence). Is is an artefact of our simplistic approach and would be taken care of in a proper treatment. Here we will just remove it by hand. Therefore, the grand potential is

$$\Omega = -\frac{gN_0^2}{2V} + k_B T \sum_{\mathbf{k}\neq 0}\ln(1 - e^{-\beta E_k}) \quad (264)$$

And now we are ready to roll! From the grand potential we can compute what we can possibly want to know about the system

$$
\begin{aligned}
\langle N\rangle &= N_0+\sum_{\mathbf{k}\neq 0}\frac{\varepsilon_k+\Delta}{E_k}\frac{1}{e^{\beta E_k}-1}\\
S &= -k_B\sum_{\mathbf{k}\neq 0}\ln(1-e^{\beta E_k})+\sum_{\mathbf{k}\neq 0}\frac{E_k}{T}\frac{1}{e^{\beta E_k}-1}\\
P &= \frac{gN_0^2}{2V^2}-\frac{k_BT}{V}\sum_{\mathbf{k}\neq 0}\ln(1-e^{-\beta E_k})\\
C_{V,N} &= \frac{1}{k_BT^2}\sum_{\mathbf{k}\neq 0}E_k^2\frac{e^{\beta E_k}}{(e^{\beta E_k}-1)^2} \qquad (265)
\end{aligned}
$$

This can again be simplified in the small temperature limit. Recall that $\Delta=(4\pi\hbar^2a/m)N_0/V$ where a is the effective size (wavelength) of the particles. It is also natural to express the wavenumber as $k=\frac{\kappa}{a}$ where κ is dimensionless and also to denote $d=8\pi a^3N_0/V$. Using this

$$E_k=\frac{\hbar^2}{2ma^2}\sqrt{(\kappa^2+d^2)^2-d^2} \qquad (266)$$

this can be expanded for $\kappa\ll d$ as

$$E_k=k_BT_a\sqrt{2d}\left(\kappa+\frac{\kappa^3}{4d}+...\right) \qquad (267)$$

where $T_a=\frac{\hbar^2}{k_B(2ma^2)}$.

Now we can plug this into the expression for the specific heat and the pressure. Also, the summation can be replaced by integration

$$C_{N,V}=\frac{1}{k_BT^2}\sum_{\mathbf{k}\neq 0}E_k^2\frac{e^{\beta E_k}}{(e^{\beta E_k}-1)^2}=\frac{1}{k_BT^2}4\pi\frac{V}{(2\pi)^3}\int_0^\infty dkk^2E_k^2\frac{e^{\beta E_k}}{(e^{\beta E_k}-1)^2}. \qquad (268)$$

By making the approximation $E_k=k_BT_a\sqrt{2d}\kappa$ we have

$$
\begin{aligned}
C_{N,V} &\approx \frac{4\pi}{k_BT^2}\frac{V}{(2\pi)^3}\frac{2dk_B^2T_a^2}{a^3}\int_0^\infty d\kappa\kappa^4\frac{e^{T_a\sqrt{2d}\kappa/T}}{(e^{T_a\sqrt{2d}\kappa/T}-1)^2}\\
&= k_BV\frac{\pi^2}{\sqrt{215}}\frac{1}{a^3d^{3/2}}\left(T/T_a\right)^3.\\
P &= \frac{g}{2}\left(\frac{N_0}{V}\right)^2+\frac{k_BT}{(2\pi)^3}\frac{4\pi}{a^3}\int d\kappa\kappa^2\ln(1-e^{-T_a\sqrt{2d}\kappa/T})\\
&= \frac{g}{2}\left(\frac{N_0}{V}\right)^2-\frac{k_B\pi^2T^4}{189\sqrt{a^3T_a^3}d^{3/2}}. \qquad (269)
\end{aligned}
$$

4.8 Statistical mechanics and path integration

`https://www-zeuthen.desy.de/~kjansen/lattice/qcd/miscellaneous/Welsing.pdf` and `https://www.zeuthen.desy.de/~kjansen/lattice/qcd/miscellaneous/CreutzFreedman.pdf`

The Wick rotation is a famous trick. It stems from the (rather annoying) fact that the temporal component of the metric tensor has a different sign. After taking $t \to i\tau$ we obtain $ds^2 = d\tau^2 + dx_i^2$, so we can turn to our comforting Euclidean metric.

It can connect relativistic field theory on Minkowski space $\mathbf{R}^{d,1}$ with at temperature T with field theory on Euclidean space $\mathbf{R}^d \times S^1_\beta$, where S^1_β is a circle of length $\beta = 1/T$. What does it means 'connect'? One can obtain the same partition function and therefore mean values of observables.

We will consider the example of farmionic oscillator. It is described by the Hamiltonian H described in terms of fermionic annihilation/creation operators $b, v^\dagger$.

$$H = \left(\psi^\dagger\psi - \frac{1}{2}\right)\omega\,. \tag{270}$$

The corresponding partition function is

$$Z = \text{Tr } e^{\beta H} = e^{\frac{\beta\omega}{2}} + e^{-\frac{\beta\omega}{2}} = 2\cosh\left(\beta\omega/2\right)\,. \tag{271}$$

The (Wick-rotated) action is on the lattice is

$$S = \left[\sum_{k=1}^{N} \bar{\psi}_k\left(\psi_k - \psi_{k-1}\right) + m\bar{\psi}_k\psi_k - \frac{1}{2}m\right] = \psi^+ M\psi - \frac{1}{2}\omega \tag{272}$$

with $m = a\omega$, $a = \beta/N$ and

$$M = \begin{pmatrix} (m+1) & 0 & 0 & 1 \\ -1 & (m+1) & 0 & 0 \\ & & & \\ 0 & -1 & \ddots & 0 \\ 0 & 0 & -1 & (m+1) \end{pmatrix} \tag{273}$$

Taking the eigenvector (for the Fourier transformation) as $\left(e^{iw}, e^{i2w}, ...\right)$ it can be quickly checked that the quantization condition is $e^{iNw} = -1 \to w = \frac{(2n-1)\pi}{N}$ and the eigenvalues are

$$\lambda_n = \left((m+1) - e^{-i\left(\frac{2n-1}{N}\pi\right)}\right). \tag{274}$$

Two comments are due here. Firstly, the frequencies in this case are called Matsubara frequencies. Secondly, the anti-periodic boundary condition is there only for fermions, bosons have the periodic condition. Now we have

$$Z = \int d\bar{\psi}d\psi e^{-S} = e^{\frac{1}{2}\omega\beta}\text{det } M = e^{\frac{1}{2}\omega}\prod_{n=1}^{N}\lambda_n \tag{275}$$

Using the fact that $\prod_{n=1}^{N}\left(A - Be^{-\frac{2\pi i}{N}n}\right) = A^n - B^n$ that follows from $\left(z^N - 1\right) = \prod_{n=1}^{N}\left(z - e^{\frac{2\pi i}{N}}\right)$, which can be believed after a couple of seconds of intensive gazing.

In our case $A = m+1,\ B = e^{i\pi/N}$ we get

$$\prod_{n=1}^{N}\left((m+1)-e^{-i\left(\frac{2n-1}{N}\pi\right)}\right)=(1+m)^N-e^{i\pi}=1+(1+m)^N \tag{276}$$

and after substituting $a=\beta/N, m=\omega\beta/N$ and remembering that $\lim\limits_{N\to\infty}(1+x/N)^N=e^x$

$$Z=e^{\frac{1}{2}\beta\omega}\left(1+e^{\omega\beta}\right) \tag{277}$$

which is the correct answer.

The procedure worked, because choosing asymetric derivative worked also a Wilson term (a term that removes fermion doublers from the lattice description). If we used a symmetrical one we would have m on the diagonal and ±1 above and below the diagonal. As a result, the eigenvalues would be $\lambda_n=\left(m+e^{-i\frac{2n-1}{N}}-e^{i\left(\frac{2n-1}{N}\pi\right)}\right)=m+2i\sin\frac{2n-1}{N}\pi$ showing doublers.

Exercise: repeat the same for the bosonic oscillator.

5 Out-of-equilibrium statistical mechanics

5.1 Fluctuations, correlation functions and the Wick's theorem

Systems are not always in equlibrium. So far we have mostly focused on the cases when they are, let us now investigate the other case. Imagine we have a system with energy E and macroscopic variables A_i, $i=1,...,n$. The number of such states is $\Gamma(E,\mathbf{A})$. Therefore, the probability to observe such a configuration is

$$P(E,\mathbf{A})=\frac{\Gamma(E,\mathbf{A})}{\Gamma(\mathbf{A})} \tag{278}$$

and the corresponding entropy is

$$S(E,\mathbf{A})=k_B\ln\left(\Gamma(E,\mathbf{A})\right) \tag{279}$$

and therefore

$$P(E,\mathbf{A})=\frac{1}{\Gamma(E)}\exp\left(\frac{1}{k_B}S(E,\mathbf{A})\right). \tag{280}$$

The variables A_i have equilibrium values of A_i^0 plus a possible fluctuation α_i

$$\alpha_i=A_i-A_i^0. \tag{281}$$

In the equilibrium the entropy assumes the minimal value and we can expand it as

$$S(E,\mathbf{A})=S(E,\mathbf{A}^0)-\frac{1}{2}\sum_{i,j=1}^{n}g_{ij}\alpha_i\alpha_j+... \tag{282}$$

where

$$g_{ij}=-\left(\frac{\partial^2 S}{\partial A_i\partial A_j}\right)_0. \tag{283}$$

Plugging this back we obtain we obtain

$$P(\alpha) = \sqrt{\frac{\det\, g}{(2\pi k_B)^n}} \exp\left(-\frac{1}{2k_B}\sum_{ij} g_{ij}\alpha_i\alpha_j\right). \tag{284}$$

where the normalisation is w.r.t. ($n-$dimensional integration)

$$\int\limits_{-\infty}^{\infty} d\alpha P(\alpha) = \int\limits_{-\infty}^{\infty} d\alpha_i \times ...P(\alpha) = 1. \tag{285}$$

For example for a CVT system spread into $i = 1, ..., l$ cells one can have

$$P(\{\Delta T_i, \Delta\rho_i\}) = \sqrt{\frac{\left(V_0 c_\rho/T_0^2\right)^l \left((V_0/T_0)(\partial\mu/\partial\rho)_0\right)^l}{(2\pi k_B)^{2l}}} \exp\left(-\frac{1}{2k_B}\sum_l \left(\frac{V_0 c_\rho}{T_0^2}\Delta T_i^2 + \frac{V_0}{T_0}\left(\frac{\partial\mu}{\partial\rho}\right)\Delta\rho_i^2\right)\right). \tag{286}$$

5.1.1 Wick's theorem

Let us now define the generating function

$$I(\beta) = \sqrt{\frac{\det\, g}{(2\pi k_B)^n}} \int\limits_{-\infty}^{\infty} d\alpha \exp\left(-\frac{1}{2k_B}\alpha^T \cdot g \cdot \alpha + \beta^T \cdot \alpha\right) = e^{\frac{1}{2}\beta^T \cdot g^{-1} \cdot \beta} \tag{287}$$

now using this we can see that the moments can be obtained by taking derivatives w.r.t β_i, for example

$$\langle \alpha_i \alpha_j \rangle = \lim_{\beta \to 0}\left(\frac{\partial^2}{\partial\beta_i\partial\beta_j} I(\beta)\right) = k_B g_{ij}^{-1}. \tag{288}$$

It is now a good time to recall the famous Wick theorem that states that we can express higher (than quadratic) moments in terms of the quadratic ones, for example

$$\langle \alpha_i\alpha_j\alpha_k\alpha_l \rangle = \langle \alpha_i\alpha_j \rangle\langle \alpha_k\alpha_l \rangle + \langle \alpha_i\alpha_k \rangle\langle \alpha_j\alpha_l \rangle + \langle \alpha_i\alpha_l \rangle\langle \alpha_j\alpha_k \rangle \tag{289}$$

5.2 Dynamical fluctuations

For the temporal dependence of fluctuations it follows that

$$\langle \alpha_i \alpha_j(\tau) \rangle = \langle \alpha_i(\tau)\alpha_j \rangle \tag{290}$$

which follows from time-reversibility of Newtonian dynamics (the correlation in that case depend only the difference between times t_1, t_2 on $|\tau| = |t_1 - t_2|$ which does not change with time shift and/or reversal). The correlation matrix $\langle \alpha_i\alpha_j(\tau) \rangle$

$$\langle \alpha_i\alpha_j(\tau) \rangle = \int\int d\alpha_i d\alpha_j \alpha_i\alpha_j P(\alpha_i, 0; \alpha_j, \tau) = \int\int d\alpha_i d\alpha_j \alpha_i\alpha_j P(\alpha_i)P(\alpha_i|\alpha_j, \tau), \tag{291}$$

where we have used the notion of joint probability. We already know that reasonable close to the equilibrium the fluctuations are governed by the equation

$$P(\alpha) = \sqrt{\frac{\det\, g}{(2\pi k_B)^n}} \exp\left(-\frac{1}{2k_B}\sum_{ij} g_{ij}\alpha_i\alpha_j\right). \tag{292}$$

The joint probability is (note that fluctuation is a macroscopic variable that has many microscopic realisations)

$$P(\alpha_i, 0; \alpha'_j, \tau) = P(\alpha_i)P(\alpha_i|\alpha'_j, \tau) \tag{293}$$

$$= \frac{1}{\Omega_{\Delta E}(E)}\int dq^N dp^N \int dq^N dp^N P(p^N, q^N|p'^N, q'^N, \tau) \tag{294}$$

where, since governed by Newtonian dynamics

$$P(p^N, q^N|p'^N, q'^N, \tau) = \delta\left(q'^N - q^N - \Delta q^N(p^N, q^N, \tau)\right)\delta\left(p'^N - p^N - \Delta p^N(p^N, q^N, \tau)\right) \tag{295}$$

where Δ denotes Hamiltonian evolution. Now it is obvious that

$$P(p^N, q^N|p'^N, q'^N, \tau) = P(-p'^N, q'^N| - p^N, q^N, \tau) \tag{296}$$

and joining those togather we have miscroscopic detailed balance

$$f(\alpha_i)P(\alpha_i|\alpha_j, \tau) = f(\alpha_j)P(\alpha_j|\alpha_i, \tau) \tag{297}$$

5.2.1 (Different) generalised forces and currents

The change in entropy due to the fluctuations is

$$\Delta S = -\frac{1}{2}\alpha_i g_{ij}\alpha_j. \tag{298}$$

We also know that when the system is pushed out of equilibrium, it tends to return there. We can therefore defined a generalised force

$$\mathcal{F}_i = g_{ij}\alpha_j = -\left(\frac{\partial \Delta S}{\partial \alpha_i}\right) \tag{299}$$

and also in analogy a generalised current

$$\mathcal{J}_i = \frac{d\alpha_i}{dt}. \tag{300}$$

Then we have

$$\frac{d\Delta S}{dt} = -\mathcal{F}_i\mathcal{J}_i \tag{301}$$

This all seems very familiar and basically the only thing we have done is that we have separated fluctuations from the equlibrium values of the variables (which can be seen as the background).

5.2.2 Regression of Fluctuations

Reichl and `https://www.iue.tuwien.ac.at/phd/mwagner/node44.html`
We can now ponder the time evolution of fluctuations

$$\langle\alpha(t)\rangle_{\alpha_0} = \int d\alpha\alpha P(\alpha_0|\alpha,t). \tag{302}$$

Onsagar assumed that on average the fluctuations decay according to *some* linear laws

$$\frac{d}{dt}\langle\alpha_i(t)\rangle_{\alpha_0} = -M_{ij}\langle\alpha_j(t)\rangle_{\alpha_0} \tag{303}$$

which can be solved easily

$$\langle\alpha(t)\rangle_{\alpha_0} = e^{-\bar{M}t}\alpha_0 \tag{304}$$

where bar is to stress that M is a matrix (α is a vector). We have to be careful about the meaning of this expression or the meaning of time-derivative.

$$\frac{d}{dt}\langle\alpha_i(t)\rangle_{\alpha_0} = \frac{\frac{d}{dt}\langle\alpha_i(t+\tau)\rangle_{\alpha_0} - \frac{d}{dt}\langle\alpha_i(t)\rangle_{\alpha_0}}{\tau}, \tag{305}$$

where $T_0 \ll \tau$
T, where T is the decay time of fluctuations and T_0 is the time between particle collisions. We can expand the solution as

$$\langle\alpha(t)\rangle_{\alpha_0} = \alpha_0 - t\bar{M}]\alpha_0 + O(t^2) \tag{306}$$

this can be plugged into (290) to obtain

$$\langle\alpha_0 \cdot \bar{M}\alpha_0\rangle = \langle\bar{M}\alpha_0 \cdot \alpha_0\rangle. \tag{307}$$

Now we can recall the relation $\langle\alpha_i\alpha_j\rangle = k_B g_{ij}^{-1}$ to define a new matrix

$$L = Mg^{-1}. \tag{308}$$

Using the notion of generalised force we can write down

$$\frac{d}{dt}\langle\alpha_i(t)\rangle_{\alpha_0} = -L \cdot \langle\mathcal{F}\rangle_{\alpha_0}. \tag{309}$$

This is the generalised Ohm's law ($j = \sigma E$). The crucial property of L is that $L_{ij} = L_{ji}$ which is called the Onsagar reciprocal relations. It states that if a force resulting from fluctuation α_i drives a flux of α_j (or A_j), the fluctuation of α_j drives the flux of α_i. For example, if the particle concentration current can drive a heat current then also a temperature gradient can drive particle current! This is very non-trivial!

For example we conservation laws for the flow of matter and energy

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{j}_\rho = 0 \tag{310}$$

$$\frac{\partial u}{\partial t} + \nabla\cdot\mathbf{j}_u = 0 \tag{311}$$

If the matter is not moving we have the Fourier's law

$$\mathbf{j}_u = -k\nabla T = kT^2\nabla(1/T). \tag{312}$$

On the contrary, in the absence of heat flows we have the Fick's law of diffusion

$$\mathbf{j}_\rho = -D\nabla\rho = D'\nabla(-\mu/T). \tag{313}$$

In a general case where both matter and heat can be transported we have

$$\begin{aligned}\mathbf{j}_u &= L_{uu}\nabla(1/T) + L_{u\rho}\nabla(-\mu/T) && (314)\\ \mathbf{j}_\rho &= L_{\rho u}\nabla(1/T) + L_{\rho\rho}\nabla(-\mu/T), && (315)\end{aligned}$$

or in compact form $\mathbf{j}_\alpha = L_{\alpha\beta} f_\beta$.

In this case the fluctuations are in u, ρ, the conjugate forces are $f_u = \nabla 1/T$ and $f_\rho = \nabla - \mu/T$. The Onsagar relation states that L is a symmetric matrix.

5.3 Wiener-Knitchine theorem, response matrix, fluctuation-dissipation Thoerem

Reichl 7.4.2, 7.5.

5.3.1 Wiener-Knitchine theorem

Our goal is now to seek connection between fluctuations of the system and the dissipation of energy in it. For example, if one forces a current in a circuit with a resistor, the current will decay. Also, if one does not force any current, there still will be some small (heat-forced) current fluctuations, so called Johnson noise. Connection between fluctuation and dissipation is very general.

To begin with, let us defined the correlation matrix

$$C_{ij}(\tau) = \langle \alpha_i(\tau)\alpha_j \rangle \tag{316}$$

which can be easily proven to be symmetric (follows from the microscopic reversibility discussed earlier). Recalling that $\langle \alpha_i \alpha_j \rangle = k_B g_{ij}^{-1} = C_{ij}(0)$ we can easily see that

$$C_{ij}(\tau) = \int d\alpha_0 P(\alpha_0)\alpha_0 \langle \alpha(\tau) \rangle_{\alpha_0} = k_B g^{-1} \cdot e^{-M|\tau|}. \tag{317}$$

We now define a modification of the time series $\alpha(t)$

$$\alpha(t; \mathcal{T}) = \alpha(t) \text{ for } |t| < \mathcal{T} \text{ and } 0 \text{ otherwise.} \tag{318}$$

We can now implement the Fourier transformation

$$\alpha(\omega, \mathcal{T}) = \int\limits_{-\mathcal{T}}^{\mathcal{T}} dt \alpha(t, \mathcal{T}) e^{i\omega\mathcal{T}}. \tag{319}$$

Note that since the fluctuations are real we have $\alpha^*(\omega, \mathcal{T}) = \alpha(-\omega, \mathcal{T})$. And now we can finally define the spectral density matrix

$$S_{ij}(\omega) = \lim_{\mathcal{T}\to\infty} \frac{1}{\mathcal{T}} \alpha_i^*(\omega, \mathcal{T})\alpha_j(\omega, \mathcal{T}). \tag{320}$$

So far only definitions and definitions! It will now escalate quickly

$$S_{ij}(\omega) = \int_{-\infty}^{\infty} \lim_{\mathcal{T}\to\infty} \int_{-\infty}^{\infty} dt \alpha_i(t, \mathcal{T}) \alpha(t+\tau, \mathcal{T}). \tag{321}$$

We can now employ the ergodic theorem (time averages equal averages over phase space averages)

$$\langle \alpha_i \alpha_j(\tau) \rangle = \lim_{\mathcal{T}\to\infty} \frac{1}{\mathcal{T}} \int_{-\infty}^{\infty} \alpha_i(t, \mathcal{T}) \alpha_j(t+\tau, \mathcal{T}) dt \tag{322}$$

from which we can see that

$$S_{ij} = \int_{-\infty}^{\infty} d\tau e^{i\omega\tau} \langle \alpha_i \alpha_j(\tau) \rangle = \int_{-\infty}^{\infty} d\tau e^{i\omega\tau} C_{ij}(\tau). \tag{323}$$

This is the celebrated Wiener-Khintchine theorem.

5.3.2 Linear response theory

As a starting step (and in our case the last as well) we can consider a linear response of α to external forces $F = (F_1, ..., F_n)$. When such forces are applied they lead to fluctuations

$$\langle \alpha(t) \rangle_F = \int_{-\infty}^{\infty} dt' K(t-t') \cdot F(t'). \tag{324}$$

K is called the *response matrix.* From causality it follows that $K(y) = 0$ for $y < 0$. Also, since finite force can only cause finite response we have $\int\limits_{0}^{\infty} dt < \infty$. We can define the Fourier transform of caused fluctuations as

$$\langle \alpha(t) \rangle_F = \frac{1}{2\pi} \lim_{-\infty}^{\infty} \langle \alpha(\omega) \rangle_F e^{-i\omega t} d\omega \tag{325}$$

and similarly for $F(t)$. From this we have

$$\langle \alpha(\omega) \rangle_F = \chi(\omega) \cdot F(\omega), \tag{326}$$

where $\chi(\omega) = \int\limits_{-\infty}^{\infty} K(t) e^{i\omega t} dt$ is called the dynamic susceptibility. This function is complex but from causality we can find a relation between its real and imaginary parts

$$\chi(t) = \frac{1}{\pi i} P \int_{-\infty}^{\infty} \frac{\chi(\omega) d\omega}{\omega - t} \tag{327}$$

where P is the Cauchy principal part.

5.3.3 Fluctuation-dissipation theorem

Let us begin by considering a force that is constant $F = F_0$ for $t < 0$ and $F = 0$ for $t > 0$. This leads to

$$\langle \alpha(t) \rangle_F = \chi(0) \cdot F_0 \text{ for } t < 0 \text{ and } \frac{1}{i\pi} P \int_{-\infty}^{\infty} d\omega \frac{\chi(\omega) F_0}{\omega} \cos(\omega t) \text{ for } t > 0. \tag{328}$$

We have established before that how quickly do fluctuations decay, so for $t > 0$ we can compute the fluctuations either using the decay expression or using the previous equation. Therefore we can establish that

$$e^{-Mt} \cdot \chi(0) \cdot F_0 = \frac{1}{i\pi} P \int_{-\infty}^{\infty} d\omega \cos\omega t \frac{\chi(\omega)}{\omega}. \tag{329}$$

Recalling that $C_{ij} = \langle \alpha_i(t) \alpha_j \rangle = e^{-M|t|} \langle \alpha_i \alpha_j \rangle$ we can obtain the famous fluctuation-dissipation theorem

$$C = \frac{1}{i\pi} P \int_{-\infty}^{\infty} d\omega \cos\omega t \frac{\chi(\omega)}{\omega} \cdot \chi^{-1}(0) \cdot \lambda\alpha\alpha\rangle. \tag{330}$$

One of examples it relates to: when we put a particle with velocity into a liquid, its velocity decays due to the drag. On the other hand, non-moving particle is forced to move – randomly fluctuate – due to interactions with surroundings, the so called Brownian motion.

The power absorbed by the system is simply obtained from

$$P(t) = \left\langle \frac{dW}{dt} \right\rangle_F = \left\langle \frac{-F \cdot d\alpha}{dt} \right\rangle_F = -F(t) \cdot \langle \dot{\alpha}(t) \rangle_F = -F(t) \cdot \frac{d}{dt} \int_{-\infty}^{\infty} dt' K(t-t') \cdot F(t'). \tag{331}$$

This can be evaluated for example for a force applied only at a certain moment $F(t) = F_0 \delta(t)$ to get

$$P(t) = i \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' \omega' F_0 \chi(\omega') F_0 e^{i(\omega+\omega')t}. \tag{332}$$

and

$$W_{\text{abs}} = \int_{-\infty}^{\infty} P(t) dt = -\frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \omega F_0 \chi''(\omega) F_0, \tag{333}$$

where $''$ denotes the imaginary part. Now this can be obtained experimentally, the real part can be computed from it from the causality condition.

6 Transport coefficients

6.1 Brownian motion, Langevin equation

(Reichl 7.2)
Movement of a particle undergoing a Brownian motion is described by the Langevin equation

$$\frac{dv(t)}{dt} = -\frac{\gamma}{m}v(t) + \frac{1}{m}\xi(t) \tag{334}$$

where $\xi(t)$ is a Gaussian white noise process that has $\langle \xi(t) \rangle_\xi = 0$ and

$$\langle \xi(t_1)\xi(t_2) \rangle_\xi = g\delta(t_2 - t_1). \tag{335}$$

Formal solution (that holds for a certain realisation of ξ) is

$$v(t) = v_0 e^{-\frac{\gamma}{m}t} + \frac{1}{m}\int_0^t ds e^{-(\gamma/m)(t-s)}\xi(s) \tag{336}$$

and

$$x(t) = x_0 + \frac{m}{\gamma}\left(1 - e^{-(\gamma/m)t}\right) + \frac{1}{\gamma}\int_0^t ds \left(1 - e^{-(\gamma/m)(t-s)}\right)\xi(s) \tag{337}$$

As with the 'drunken sailor' example, we can obtain some estimates despite knowing exact details of a process. Given that $\langle v_0 \xi(t) \rangle = 0$ we can obtain

$$\langle v(t_2)v(t_1) \rangle_\xi = \left(v_0^2 - \frac{g}{2m\gamma}\right) e^{-(\gamma/m)}(t_2 + t_1) + \frac{g}{2m\gamma}e^{-(\gamma/m)|t_2 - t_1|} \tag{338}$$

and similarly for the leading long-term displacement

$$\langle (x(t_2) - x_0)^2 \rangle_\xi \approx \left(\frac{g}{\gamma^2}\right) t \tag{339}$$

When we average those equations w.r.t. thermal fluctuations we can take $\langle v_0^2 \rangle_T = \frac{k_B T}{m}$. Also, a stationary solution can depend only on the difference $t_1 - t_2$ so the first term in the aforementioned equations has to be cancelled, this is achieved by taking $g = 2m\gamma v_0^2 = 2\gamma k_B T$, yield the correlation function

$$\langle \langle v(t_2)v(t_1) \rangle_\xi \rangle_T = \frac{k_B T}{m} e^{-(\gamma/m)|t_2 - t_1|} \tag{340}$$

and for the displacement

$$\langle \langle (x(t) - x_0)^2 \rangle_\xi \rangle_T = \frac{2k_B T}{\gamma}\left(1 - \frac{m}{\gamma}\left(1 - e^{-(\gamma/m)t}\right)\right) \tag{341}$$

where we assumed stationary particle at the origin. Thus after a long time $\langle \langle (x(t) - x_0)^2 \rangle_\xi \rangle_T = \frac{2k_B T}{\gamma} t$ from which we can read of the diffusion coefficient $D = k_B T/\gamma$ (from the diffusion equation $\partial_t f = D\partial_{xx} f$. For large Brownian particles we can use friction equation $\gamma = 6\pi\eta a$ to measure the size a of the particles.

The spectral density of the velocity auto-correlation function is

$$S_{v,v}(\omega) = \int\limits_{-\infty}^{\infty} d\tau e^{-i\omega\tau}\langle v(t_1+\tau)v(t_1)\rangle_{\xi,\tau} = \frac{2k_BT}{m}\frac{\gamma/m}{\omega^2+(\gamma/m)^2} = \int\limits_{-\infty}^{\infty} e^{-i\omega\tau}C_{v,v}(\tau), \tag{342}$$

where the last equation is from the Wiener-Khintchine theorem.

The spectral density of the white noise is

$$S_{\xi,\xi} = g = 2\gamma k_BT, \tag{343}$$

all frequencies are equally possible.

6.1.1 The Fokker-Planck equation

(Reichl 7.3)
We are interested in getting the probability to find a particle in a given portion of the phase space expressed in terms of the probability density ρ

$$P(x,v,t) = \langle\rho(x,v,t)\rangle_\xi. \tag{344}$$

We will denote the phase space coordinates $X = (x,v)$ with

$$\int\limits_{-\infty}^{\infty} dx \int\limits_{-\infty}^{\infty} dv\rho(x,v,t) = \int\limits_{-\infty}^{\infty} dX\rho(X,t) = 1. \tag{345}$$

In a usual way we can define $P(A_0) = \int_{A_0} dX\rho$. Probability current has to be conserved

$$\frac{\partial}{\partial t}P(A_0)\frac{\partial}{\partial t}\int\limits_{A_0} dX\rho(X,t) = -\oint_{L_0}\rho(X,t)\dot{X}\cdot dS_0 = -\int\limits_{A_0} dX\nabla_X\cdot(\dot{X}\rho). \tag{346}$$

As the area A_0 is arbitrary we can conclude that

$$\frac{\partial\rho}{\partial t} = -\nabla_X\cdot(\dot{X}\rho) = -\frac{\partial(\dot{x}\rho}{\partial x} - \frac{\partial(\dot{v}\rho}{\partial v}. \tag{347}$$

So far our discussion has been rather general. For the Brownian movement with a potential $V(x)$ we have

$$\dot{v} = -(\gamma/m)v + F/m + \xi/m; \quad \dot{x} = v, \tag{348}$$

where $F = -\partial V/\partial x$. We can plug those expressions into the equation (347) to obtain

$$\frac{\partial\rho}{\partial t} = -\hat{L}_0\rho - \hat{L}_1\rho \tag{349}$$

Where are interested in the observable probability $P = \langle\rho\rangle_\xi$, the equation governing its evolution is

$$\frac{\partial P}{\partial t} = -v\frac{\partial P}{\partial x} + \frac{\partial}{\partial v}\left(\left(\frac{\gamma}{m}v - \frac{1}{m}F(x)\right)P\right) + \frac{g}{m^2}\frac{\partial^2 P}{\partial v^2} \tag{350}$$

which is called the Fokker-Planck equation for the observed probability P. It can be cast in the form of continuity equation

$$\frac{\partial P}{\partial t} = -\nabla \cdot J \tag{351}$$

where $J = (j_x, j_v)$ is the probability flux. The Fokker-Planck equation can be solved exactly in the strong friction limit to obtain the familiar diffusion equation

$$\frac{\partial P}{\partial t} = D\frac{\partial^2 P}{\partial x^2}, \tag{352}$$

with $D = \frac{g}{2\gamma^2} = k_B T/\gamma$.

6.2 Transport coefficients

The transport coefficients can be derived using a simple mean free path approach. We will consider a gas with particles well described by Maxwell-Boltzman distribution $F(v) = \left(\frac{m\beta}{2\pi}\right)\exp\left(-\frac{\beta m v^2}{2}\right)$. The mean free path λ is the average path that a typical particle travels between two collisions. Or expressed in other terms, number of collision after travelling a unit distance is $1/\lambda$.

Probability of having no collision satisfies

$$P_0(r + dr) = P_0(r)(1 - \frac{dr}{\lambda}). \tag{353}$$

A probability that satisfies this condition is

$$P_0(r) = e^{-r/\lambda}. \tag{354}$$

A simple sanity check: $\langle r\rangle = \int\limits_0^\infty rP_0(r)\frac{dr}{\lambda} = \lambda$.

How frequently do particles collide? To figure this out we need to compute the average relative velocity, which can be obtained from (double) Maxwell-Boltzman integral. Imagine having two types of particles, denoted A and B, their relative velocities are

$$\langle v_r\rangle_{AB} = \left(\frac{8k_B T}{\pi\mu_{AB}}\right) \tag{355}$$

where $\mu_{AB} = \frac{m_A m_B}{m_A + m_B}$.

Now knowing the relative velocity we can move to the rest frame of particle B and figure out how long will it take for a particle A to hit it. The collision parameter is $d_{AB} = (d_A + d_B)/2$ where d is the diameter. The particle A sweeps out a collision cylinder of volume

$$V = \pi d^2\langle v_r\rangle_{AB}t. \tag{356}$$

So the collision frequency is $f_{AB} = n_B\pi d_{ab}^2\langle v_r\rangle_{AB}$ and the number of collisions is

$$\nu_{AB} = n_A n_B \pi d_{ab}^2 \langle v_r \rangle_{AB} = n_A n_B \pi d_{ab}^2 \left(\frac{8 k_B T}{\pi \mu_{AB}} \right) \tag{357}$$

If the particles are of the same count we have to add a factor of 1/2 to prevent overcounting. So for a gas of identical particles we have

$$\lambda = \frac{\langle v \rangle}{f_{AA}} = \langle v \rangle \tau = \frac{1}{\sqrt{2} n_A \pi d_{AA}^2} \tag{358}$$

where τ is the mean time between collisions and $\langle v_r \rangle = \sqrt{2} \langle v \rangle$.

We can now imagine there are two types of particles, tracer particles we pay attention to and the rest of the stuff. If there is a gradient of the tracer particle concentration there will be some flow through a surface are dS. To compute it we have

$$\dot{N}_+ = \frac{\langle v \rangle}{4\pi\lambda} \int\limits_0^\infty r^2 dr \int\limits_0^{\pi/2} \sin\theta d\theta \int\limits_0^{2\pi} d\phi n_T(z) \cos\theta \frac{e^{-r/\lambda}}{r^2}. \tag{359}$$

a similarly for the particles coming from the other way. We can assume that the particles that passes the surface has collided last time at $z \approx \lambda$ so we can take $n_T(z) \approx n_T(0) + z \left(\frac{\partial n_T}{\partial z} \right)_0$. Using this we obtain

$$\dot{N}_+ - \dot{N}_- = \frac{\langle v \rangle \lambda}{3} \left(\frac{\partial n_T}{\partial z} \right)_0. \tag{360}$$

From this we have the diffusion equation

$$J_D(z) = -D \frac{\partial n_T(z)}{\partial z} \tag{361}$$

where $D = \frac{\langle v \rangle \lambda}{3}$ is the coefficient of self-diffusion.

If there is some property $A(z)$ that varies along the system (temperature, concentration, velocity in some preferred direction) we can compute how is this transported by the movement of particles.

$$J_A(z) = -b_A n \langle v \rangle \lambda \frac{\underset{\cdot}{\mathrm{A}}}{\underset{\cdot}{\mathrm{z}}}. \tag{362}$$

By comparing with the previous equation we can see the coefficient of proportionality is $b_A = 1/3$.

If the property is the velocity in a perpendicular direction $A(z) = m \langle v_y(z) \rangle$ we get

$$J_{zy} = -\eta \frac{d \langle v_y(z) \rangle}{dz} \tag{363}$$

where $\eta = \frac{1}{3} n m \langle v \rangle \lambda$ is the coefficient of shear viscosity.

If the property is $A(z) = \frac{1}{2} m \langle v^2(z) \rangle = \frac{3}{2} k_B T(z)$ we obtain the heat flow equation

$$J_Q = -K \frac{dT}{dz}, \tag{364}$$

with $K = \frac{1}{2} n \langle v \rangle \lambda k_B$.

6.3 Boltzman equation

Černý, Melo: Selected topics from statistical physics 8

6.3.1 Cross section

The scattering process is characterised by a cross section which is the ratio

$$d\sigma = dN/j. \tag{365}$$

We have the notion of an impact parameter $b(\theta)$ that states how far from the center (assuming at least axial but preferably rotational symmetry) needs the target to be hit to scatter the particle by the angle θ.

$$d\sigma = bdbd\varphi = b(\theta)|\frac{db(\theta)}{d\theta}|d\theta d\varphi. \tag{366}$$

Usually we are interested in the differential cross section

$$\frac{d\sigma}{d\Omega} = \frac{b(\theta)}{\sin\theta}|\frac{db(\theta)}{d\theta}|. \tag{367}$$

We can express the cross-section in terms of the undergone process $\sigma(\mathbf{v}_1, \mathbf{v}_2 \to \mathbf{v}_1', \mathbf{v}_2')$. The cross-section inherits the symmetries of the underlying theory

$$\begin{aligned} \sigma(\mathbf{v}_1, \mathbf{v}_2 \to \mathbf{v}_1', \mathbf{v}_2') &= \sigma(-\mathbf{v}_1', -\mathbf{v}_2' \to \mathbf{v}_1, \mathbf{v}_2) \qquad (368) \\ \sigma(\mathbf{v}_1, \mathbf{v}_2 \to \mathbf{v}_1', \mathbf{v}_2') &= \sigma(R\mathbf{v}_1, R\mathbf{v}_2 \to R\mathbf{v}_1', R\mathbf{v}_2') \\ \sigma(\mathbf{v}_1, \mathbf{v}_2 \to \mathbf{v}_1', \mathbf{v}_2') &= \sigma(\mathbf{v}_1', \mathbf{v}_2' \to \mathbf{v}_1, \mathbf{v}_2) \end{aligned}$$

where R denotes either the rotation or reflection.

The naive intuition is that the physical cross-section is the geometrical cross-section, which, however, is not true. This can be seen by considering that point particles (with for example EM interaction) have non-zero cross section. Also, the cross-section of a hard-sphere scattering $\sigma = \pi(2R)^2$ does not match its geometrical cross-section.

6.3.2 Boltzmann equation without collisions

We will consider the probability density $f(\mathbf{r}, \mathbf{v}, t)$. The particle density around the point $\mathbf{r}$ is

$$n(\mathbf{r}) = \int f(\mathbf{r}, \mathbf{v}, t) d^3v. \tag{369}$$

This function is related to a single particle density ρ_1 by $f = N\rho_1$ (this relation seems obvious but requires some work to prove).

Without collisions, under the influence of force F the positions and velocities evolve according to Newtonian laws. From this we have

$$f(\mathbf{r}_0, \mathbf{v}_0, 0) = f(\mathbf{r}_t, \mathbf{v}_t, t) \tag{370}$$

where the subscript $_t$ denotes the time-evolved quantities. From this we have the collision-free Boltzmann equation

$$\frac{d}{dt}f(\mathbf{r},\mathbf{v},t) = 0 = \left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\frac{\partial}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m}\cdot\frac{\partial}{\partial \mathbf{v}}\right) f(\mathbf{r},\mathbf{v},t) = Df(\mathbf{r},\mathbf{v},t) \tag{371}$$

The equation is simply $Df(\mathbf{r},\mathbf{v},t) = 0$, the collision version would have r.h.s. $Df(\mathbf{r},\mathbf{v},t) = D_c f(\mathbf{r},\mathbf{v},t)$.

6.3.3 Heat conduction in the approximation of relaxation time

In a similar manner as in our discussion of transport coefficients, let us take

$$f(x,\mathbf{v}) = \int \frac{dt'}{\tau} e^{-t'/\tau} f_0(\mathbf{r} - \mathbf{v}t, \mathbf{v}), \tag{372}$$

where we consider dependence only in the x direction and f_0 has the form of Maxwell distribution

$$f_0(x,\mathbf{v}) = n(x)\left(\frac{m\beta(x)}{2\pi}\right)^{3/2} \exp\left(-\frac{1}{2}\beta(x)\mathbf{v}^2\right). \tag{373}$$

The unknown functions n, β are related to f by

$$\begin{aligned} n(x) &= \int d^3v f(x,\mathbf{v},t), \\ \frac{3}{2\beta(x)} &= \int d^3v \frac{1}{2} m\mathbf{v}^2 f(x,\mathbf{v},t). \end{aligned} \tag{374}$$

We now apply per-partes to (372) to obtain

$$f(x,\mathbf{v}) = f_0(x,\mathbf{v}) + \int \frac{df_0(x,\mathbf{v})}{dt'} \exp\left(-t'/\tau\right) dt'. \tag{375}$$

Keeping in mind that only region $t \in (0,\tau)$ contributes to the integration we can plug in the expressions for n, β into f_0' with f_0 replacing f on the r.h.s. (which would be next-to-leading order correction in our small t expansion to obtain

$$f(x,\mathbf{v}) = f_0(x,\mathbf{v}) + \tau\left(-\frac{dn(x)}{n(x)dx}v_x - \frac{3}{2}\frac{d\beta(x)}{\beta(x)dx} + \frac{1}{2}m\mathbf{v}^2\frac{d\beta(x)}{dx}v_x f_0(x,\mathbf{v})\right) \tag{376}$$

We are interested in the stationary solution so $\langle v_x\rangle = 0$. As the distribution f_0 is Maxwell-like, so are the moments

$$\langle v_x^2\rangle_0 = frac1m\beta, \langle v_x^4\rangle_0 = \frac{3}{(m\beta)^2}, \cdots \tag{377}$$

and so on. Multiplying (376) with v_x and integrating over we obtain

$$0 = \left(-\frac{dn}{ndx}\langle v_x^2\rangle_0 - \frac{3}{2}\frac{d\beta}{\beta dx}\langle v_x^2\rangle_0 + \frac{1}{2}m\frac{d\beta}{dx}(\langle v_x^4 + v_x^2(v_y^2 + v_z^2)\rangle_0\right) \tag{378}$$

$$= \frac{\tau}{m\beta}\left(-\frac{dn}{ndx} + \frac{d\beta}{\beta dx}\right) \tag{379}$$

which gives

$$\frac{dn}{ndx} = \frac{d\beta}{\beta dx} \tag{380}$$

which tells us that $nkT = const$, in other words the pressure is constant. We can compute the flow of energy $j_E = \frac{1}{2}nm\left\langle \mathbf{v}^2 v_x \right\rangle$ using the relation for the moments

$$j_E = \frac{5\tau}{(m\beta)^2}\frac{d\beta}{dx} = \frac{5}{2}\frac{\tau k^2 Tn}{m}\left(-\frac{dT}{dx}\right) \tag{381}$$

which is (again) the heat flow equation with the coefficient $\kappa = \frac{5}{2}frac\tau k^2 Tnm$.

6.3.4 Collision term

If a volume in the phase space flows according to Newtonian dynamics, it is conserved. This can change if some particles interact, which can either cause them to be removed or added to the trajectory. The change rate of particles in infinitesimal phase space volume and a unit of time is

$$D_C^- f(\mathbf{r},\mathbf{v},t)d^3rd^3vdt = d^3rd^3vdt \iiint d^3v_1 d^3v' d^3v_1' |\mathbf{v}-\mathbf{v}_1| f(\mathbf{r},\mathbf{v},t) f(\mathbf{r},\mathbf{v}_1,t)\sigma(\mathbf{v},\mathbf{v}_1;\mathbf{v}',\mathbf{v}_1') \tag{382}$$

which is caused by a particle with any velocity $\mathbf{v}_1$ hitting the particle with $\mathbf{v}$ and scattering with velocities $\mathbf{v}_1'$ and $\mathbf{v}'$. The number of such interactions per unit time is proportional to the difference $|\mathbf{v}-\mathbf{v}_1|$. Similar process can result in a particle with the velocity $\mathbf{v}$ at the position $\mathbf{r}$.

$$D_C^+ f(\mathbf{r},\mathbf{v},t)d^3rd^3vdt = d^3rd^3vdt \iiint d^3v_1 d^3v' d^3v_1' |\mathbf{v}'-\mathbf{v}_1'| f(\mathbf{r},\mathbf{v}',t) f(\mathbf{r},\mathbf{v}_1',t)\sigma(\mathbf{v}',\mathbf{v}_1';\mathbf{v},\mathbf{v}_1). \tag{383}$$

The energy conservation gives $|\mathbf{v}-\mathbf{v}_1| = |\mathbf{v}'-\mathbf{v}_1'|$ while the PT invariance gives $\sigma(\mathbf{v},\mathbf{v}_1;\mathbf{v}',\mathbf{v}_1') = \sigma(\mathbf{v}',\mathbf{v}_1';\mathbf{v},\mathbf{v}_1)$.

The total collision term $D_C = D_C^+ - D_C^-$ is

$$D_C f(\mathbf{r},\mathbf{v},t) = \iiint d_1^v d^3v' d^3v_1' |\mathbf{v}'-\mathbf{v}_1'| \left(f'f_1' - ff_1\right)\sigma, \tag{384}$$

where we gave up on writing the arguments where they are obvious from the notation.

6.3.5 Conservation laws

Let us consider a general observable $\mathcal{O}$

$$\langle \mathcal{O} \rangle = \frac{1}{n(\mathbf{r},t)}\int d^3v f(\mathbf{r},\mathbf{v},t)\mathcal{O}(\mathbf{r},\mathbf{v},t). \tag{385}$$

We can now take the Boltzmann equation $Df = D_C f$, multiply it with $\mathcal{O}$ and integrate over d^3v. In each of the terms of Df we can use that $A\partial B =$

$\partial AB - (\partial A)B$ (and assuming forces being independent of the velocities) to obtain

$$\int d^3v\mathcal{O}Df(\mathbf{r},\mathbf{v},t) = n\left(D\mathcal{O}\right) - \frac{\partial}{\partial t}\langle n\mathcal{O}\rangle - \frac{\partial}{\partial x_a}\langle nv_a\mathcal{O}\rangle \tag{386}$$

Now for the r.h.s., in $\int d^3v\mathcal{O}D_Cf$ we integrate over 4 different velocities and can use the symmetries of σ to obtain that the integral is equal to

$$\int d^3v\mathcal{O}D_Cf = \frac{1}{4}\iiiint d^3vd^3v'd^3v'_1d^3v'(f'f'_1-ff_1)|\mathbf{v}_1-\mathbf{v}|\sigma\left(\mathcal{O}'+\mathcal{O}'_1-\mathcal{O}-\mathcal{O}_1\right). \tag{387}$$

So in the case of observed quantities $\mathcal{O}'+\mathcal{O}'_1 = \mathcal{O}+\mathcal{O}_1$ we have

$$n\left(D\mathcal{O}\right) = \frac{\partial}{\partial t}\langle n\mathcal{O}\rangle + \frac{\partial}{\partial x_a}\langle nv_a\mathcal{O}\rangle \tag{388}$$

If we take $\mathcal{O} = m$ the equation (388) yields

$$\frac{\partial}{\partial t}\langle nm\rangle + \frac{\partial}{\partial x_a}\langle nmv_a\rangle = 0 \tag{389}$$

which for the case of density being independent of velocity takes the form

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) = 0 \tag{390}$$

where we took $\mathbf{u} = \langle\mathbf{v}\rangle$ as the mean velocity to identify the continuity equation.

We can separate the mean velocity and the fluctuations $\mathbf{v} = \mathbf{u}+\mathbf{U}$, it is obvious to see that

$$\langle v_av_b\rangle = u_au_b + \langle U_aU_b\rangle. \tag{391}$$

Introducing the tension $P_{ab} = \rho\left\langle U_aU_b\right\rangle$ to see that the conservation of momentum v_am leads to Euler hydrodynamic equation.

$$\frac{\partial}{\partial t}\rho u_b + \frac{\partial}{\partial x_a}\rho u_au_b = -\frac{\partial P_{ab}}{\partial x_a} + F_b/m. \tag{392}$$

6.3.6 H-theorem

Boltzmann considered the time evolution of the quantity

$$H(t) = \int d^3vd^3f\ln f. \tag{393}$$

Why? I was known that microscopic world is reversible while macroscopic is not. Why are close to the border so we should be able to see the growth of the entropy. So he defined this term, coined it E, wrote it as the capital Greek letter Eta and the confusion began, turning it into H instead.

We are interested in $\frac{dH(t)}{dt}$ which can be – in a similar manner as before – turned into

$$\frac{dH(t)}{dt} = -\frac{1}{4}\int d^3r d^3v d^3v_1 d^3v' d^3v'_1 |\mathbf{v} - \mathbf{v}_1| f f_1 \ln\left(\frac{f' f'_1}{f f'}\right)\left(\frac{f' f'_1}{f f'} - \right)\sigma(\mathbf{v}, \mathbf{v}'; \mathbf{v}', \mathbf{v}'_1) \tag{394}$$

Most of the terms in the integrals are obviously positive. The $(x-1)\ln x \geq 0$ is less obvious but still firmly positive, therefore

$$\frac{d}{dt}H(t) \leq 0 \tag{395}$$

so the quantity tends to the minimum and stays there. The minimum is realised by f being Maxwell-Boltzman distribution.

7 Bonus lectures

7.1 Black hole entropy

Central point is the paper *Black Holes and Entropy* by Jacob Bekenstein. The central question is: what happens if we drop a system with some entropy into a black hole? Black hole is described by three parameters: M, J, Q and naively there is no microscopic description and therefore no entropy involved.

However, Hawking noticed that during any process, the total black area has to increase.

By differentiating the formula for the surface are and solving for dM one obtains

$$dM = \Theta d\alpha + \mathbf{J}\mathbf{dL} + \phi dQ, \tag{396}$$

where Θ is a function of the radius and $\alpha = A/4\pi$ where A is the surface area of the event horizon.

This seems even more suggestive, it seem that the surface area plays the role of the entropy. When you drop a particle into the black hole, your lose at least a bit of information – you don't know if the particle exists or not.

By considering the increase of surface area by dropping a harmonic oscillator (of know entropy) one obtains more detailed expression.

The conclusion of Bekenisten analysis is that the second law has to be generalized to include the black hole entropy (source of which is not understood even today)

$$dS_{universe} + dS_{BH} \geq 0, \tag{397}$$

where $S_{BH} = \frac{A}{4L_P^2}$ where L_P is the Planck length.

7.2 Maxwell demon, Szilard machine and Landauer heat

We know that the entropy of an isolated system cannot decrease. Imagine the situation of a room split into two parts, separated by a small door operated by a small creature – called the Maxwell demon. The demon can work as a filter: for example letting molecules of one type move from the left to right and the other way for the second type of the molecule. As a result, he decreases the entropy of the room. Is the second law violated?

Szilard and Brilloun thought that by obtaining information is the demon increasing the entropy by interacting with photons carrying the information about position of the particles – heating up by absorbing their entropy. This isn't the correct explanation.

We know that friction is not reversible. A ball sitting at one point does not tell you from which side did it come from. During this process, heat was produced. That makes sense, irreversibility means we are moving towards a larger portion of the phase space – which has larger entropy.

Landaur was considering the minimal heat that needs to be produced when manipulating with information. He found out that erasing a bit of information releases the heat $k_B T \ln 2$. Consider a system spit into two equal volumes, a particle can be in a single cell – denoting 0 or 1. How to reset this system to the state 0 lets say? Remove the wall (particle now moves in $2V$), place it on the right side and move it to the left. A work equal to $k_B T \ln 2$ has to be applied. Forgetting is costly. It was later realised that the demon returns the entropy when returning to his original state. However, what happens in between?

It was realised that there is another form of entropy involved. With the onset of computers Turing and Neumann considered the general idea of a computer. After passing a certain threshold, computers are equal in a sense that they can simulate each other (it can be very slow but reach the same result). The crossing point is called the Turing machine. A simple mechanism that can move along a line of bits and can 1. read, 2. write, 3. operates with an internal memory of at least 4 bits).

A computer program is something produces a string S of length L. How much information is stored in S? Some strings seem rather random 101000101001010101011 some seems to be easy to predict 000000000....0000. The definition is that the algorithmic entropy (or Kolmogorov complexity) of a string is proportional to the length of the shortest code $S*$ that produces it as out. (A random string is one for which $L_S \approx L_{S^*}$, there is no way of compressing the string.

The claim is that after obtaining the information and before realising it, demon's mind carries the algorithmic entropy. The problem is that the algorithmic entropy cannot be computed.

A known example is the halting problem – no program can predicting in general where other program will halt or not. If it existed, one could produce a computer version of the barber paradox – does a barber who shaves every man not shaving himself shave himself? One could make a program that halts if and only if checking a program that does not halt – and then let it act on itself.

Another known paradox is the one named after the librarian Berry. "A word that cannot be described by the twelve words or less" cannot exist since this sentence describes it. If we were able to compute the algorithmic entropy we could produce a computer version of this paradox, something along the lines "A shortest program that cannot be described using N bits or less".

7.3 Introduction to complexity theory

.

7.4 Quantum theory at finite temperature via path integration

.

7.5 Hagedorn phase transition

.

7.6 Statistical physics and brain

.